Introduction to Longitudinal Data Analysis

Geert Molenberghs
Center for Statistics
Universiteit Hasselt, Belgium
geert.molenberghs@uhasselt.be
www.censtat.uhasselt.be

Geert Verbeke
Biostatistical Centre
K.U.Leuven, Belgium
geert.versebe@med.kuleuven.be
www.kuleuven.ac.be/biostat/

Master of Science in Biostatistics
Universiteit Hasselt
## Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Related References</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>Continuous Longitudinal Data</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Introduction</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>Cross-sectional versus Longitudinal Data</td>
<td>38</td>
</tr>
<tr>
<td>3</td>
<td>Simple Methods</td>
<td>51</td>
</tr>
<tr>
<td>4</td>
<td>The Multivariate Regression Model</td>
<td>61</td>
</tr>
<tr>
<td>5</td>
<td>A Model for Longitudinal Data</td>
<td>89</td>
</tr>
<tr>
<td>Chapter</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>6</td>
<td>Exploratory Data Analysis</td>
<td>120</td>
</tr>
<tr>
<td>7</td>
<td>Estimation of the Marginal Model</td>
<td>152</td>
</tr>
<tr>
<td>8</td>
<td>Inference for the Marginal Model</td>
<td>178</td>
</tr>
<tr>
<td>9</td>
<td>Inference for the Random Effects</td>
<td>233</td>
</tr>
<tr>
<td>10</td>
<td>General Guidelines for Model Building</td>
<td>257</td>
</tr>
<tr>
<td>11</td>
<td>Power Analyses under Linear Mixed Models</td>
<td>289</td>
</tr>
<tr>
<td>II</td>
<td>Marginal Models for Non-Gaussian Longitudinal Data</td>
<td>308</td>
</tr>
<tr>
<td>12</td>
<td>The Toenail Data</td>
<td>309</td>
</tr>
<tr>
<td>13</td>
<td>The Analgesic Trial</td>
<td>312</td>
</tr>
<tr>
<td>14</td>
<td>The National Toxicology Program (NTP) Data</td>
<td>315</td>
</tr>
<tr>
<td>15</td>
<td>Generalized Linear Models</td>
<td>318</td>
</tr>
<tr>
<td>16</td>
<td>Parametric Modeling Families</td>
<td>336</td>
</tr>
</tbody>
</table>
17  Conditional Models ........................................................................................................... 340
18  Full Marginal Models ........................................................................................................ 356
19  Generalized Estimating Equations ................................................................................. 368
20  A Family of GEE Methods .............................................................................................. 380

III  Generalized Linear Mixed Models for Non-Gaussian Longitudinal Data 400
21  The Beta-binomial Model .................................................................................................. 401
22  Generalized Linear Mixed Models (GLMM) ................................................................. 407
23  Fitting GLMM’s in SAS .................................................................................................... 425

IV  Marginal Versus Random-effects Models and Case Studies 435
24  Marginal Versus Random-effects Models ...................................................................... 436
25  Case Study: The NTP Data ............................................................................................ 449
26  Case Study: Binary Analysis of Analgesic Trial ............................................................. 471
V Incomplete Data

29 Setting The Scene ................................................................. 527
30 Proper Analysis of Incomplete Data ........................................... 540
31 Weighted Generalized Estimating Equations ............................... 563
32 Multiple Imputation .................................................................... 581
33 Creating Monotone Missingness .................................................. 596

VI Topics in Methods and Sensitivity Analysis for Incomplete Data

34 An MNAR Selection Model and Local Influence .......................... 599
35 Mechanism for Growth Data ......................................................... 613
36 Interval of Ignorance ................................................................ 616
37 Pattern-mixture Models

38 Concluding Remarks

Introduction to Longitudinal Data Analysis
Chapter 0
Related References


University Press.


New-York: Wiley.


Part I

Continuous Longitudinal Data
Chapter 1
Introduction

▷ Repeated Measures / Longitudinal data
▷ Examples
1.1 Repeated Measures / Longitudinal Data

Repeated measures are obtained when a response is measured repeatedly on a set of units

- Units:
  - Subjects, patients, participants, ...
  - Animals, plants, ...
  - Clusters: families, towns, branches of a company, ...
  - ...

- Special case: Longitudinal data
1.2 Captopril Data

- Taken from Hand, Daly, Lunn, McConway, & Ostrowski (1994)
- 15 patients with hypertension
- The response of interest is the supine blood pressure, before and after treatment with CAPTOPRIL

<table>
<thead>
<tr>
<th>Patient</th>
<th>Before SBP</th>
<th>Before DBP</th>
<th>After SBP</th>
<th>After DBP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>210</td>
<td>130</td>
<td>201</td>
<td>125</td>
</tr>
<tr>
<td>2</td>
<td>169</td>
<td>122</td>
<td>165</td>
<td>121</td>
</tr>
<tr>
<td>3</td>
<td>187</td>
<td>124</td>
<td>166</td>
<td>121</td>
</tr>
<tr>
<td>4</td>
<td>160</td>
<td>104</td>
<td>157</td>
<td>106</td>
</tr>
<tr>
<td>5</td>
<td>167</td>
<td>112</td>
<td>147</td>
<td>101</td>
</tr>
<tr>
<td>6</td>
<td>176</td>
<td>101</td>
<td>145</td>
<td>85</td>
</tr>
<tr>
<td>7</td>
<td>185</td>
<td>121</td>
<td>168</td>
<td>98</td>
</tr>
<tr>
<td>8</td>
<td>206</td>
<td>124</td>
<td>180</td>
<td>105</td>
</tr>
<tr>
<td>9</td>
<td>173</td>
<td>115</td>
<td>147</td>
<td>103</td>
</tr>
<tr>
<td>10</td>
<td>146</td>
<td>102</td>
<td>136</td>
<td>98</td>
</tr>
<tr>
<td>11</td>
<td>174</td>
<td>98</td>
<td>151</td>
<td>90</td>
</tr>
<tr>
<td>12</td>
<td>201</td>
<td>119</td>
<td>168</td>
<td>98</td>
</tr>
<tr>
<td>13</td>
<td>198</td>
<td>106</td>
<td>179</td>
<td>110</td>
</tr>
<tr>
<td>14</td>
<td>148</td>
<td>107</td>
<td>129</td>
<td>103</td>
</tr>
<tr>
<td>15</td>
<td>154</td>
<td>100</td>
<td>131</td>
<td>82</td>
</tr>
</tbody>
</table>
Research question:

How does treatment affect BP?

Remarks:

- Paired observations:
  Most simple example of longitudinal data

- Much variability between subjects
1.3 Growth Curves

- Taken from Goldstein 1979
- The height of 20 schoolgirls, with small, medium, or tall mothers, was measured over a 4-year period:

<table>
<thead>
<tr>
<th>Mothers height</th>
<th>Children numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small mothers &lt; 155 cm</td>
<td>1 → 6</td>
</tr>
<tr>
<td>Medium mothers [155cm; 164cm]</td>
<td>7 → 13</td>
</tr>
<tr>
<td>Tall mothers &gt; 164 cm</td>
<td>14 → 20</td>
</tr>
</tbody>
</table>

- Research question:
  Is growth related to height of mother?
• Individual profiles:

![Graphs showing height growth for different mother heights over age](image-url)
Remarks:

- Almost perfect linear relation between Age and Height
- Much variability between girls
- Little variability within girls
- Fixed number of measurements per subject
- Measurements taken at fixed time points
1.4 Growth Data

- Taken from Potthoff and Roy, Biometrika (1964)

- The distance from the center of the pituitary to the maxillary fissure was recorded at ages 8, 10, 12, and 14, for 11 girls and 16 boys

- Research question:

  Is dental growth related to gender?
- Individual profiles:
Remarks:

- Much variability between children
- Considerable variability within children
- Fixed number of measurements per subject
- Measurements taken at fixed time points
1.5 Rat Data

- Research question (Dentistry, K.U.Leuven):

  How does craniofacial growth depend on testosterone production?

- Randomized experiment in which 50 male Wistar rats are randomized to:
  - Control (15 rats)
  - Low dose of Decapeptyl (18 rats)
  - High dose of Decapeptyl (17 rats)
- Treatment starts at the age of 45 days; measurements taken every 10 days, from day 50 on.

- The responses are distances (pixels) between well defined points on x-ray pictures of the skull of each rat:
Measurements with respect to the roof, base and height of the skull. Here, we consider only one response, reflecting the height of the skull.

Individual profiles:
• Complication: Dropout due to anaesthesia (56%):

<table>
<thead>
<tr>
<th>Age (days)</th>
<th>Control</th>
<th>Low</th>
<th>High</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>15</td>
<td>18</td>
<td>17</td>
<td>50</td>
</tr>
<tr>
<td>60</td>
<td>13</td>
<td>17</td>
<td>16</td>
<td>46</td>
</tr>
<tr>
<td>70</td>
<td>13</td>
<td>15</td>
<td>15</td>
<td>43</td>
</tr>
<tr>
<td>80</td>
<td>10</td>
<td>15</td>
<td>13</td>
<td>38</td>
</tr>
<tr>
<td>90</td>
<td>7</td>
<td>12</td>
<td>10</td>
<td>29</td>
</tr>
<tr>
<td>100</td>
<td>4</td>
<td>10</td>
<td>10</td>
<td>24</td>
</tr>
<tr>
<td>110</td>
<td>4</td>
<td>8</td>
<td>10</td>
<td>22</td>
</tr>
</tbody>
</table>

• Remarks:
  ▶ Much variability between rats, much less variability within rats
  ▶ Fixed number of measurements scheduled per subject, but not all measurements available due to dropout, for known reason.
  ▶ Measurements taken at fixed time points
1.6 Toenail Data


- **Toenail Dermatophyte Onychomycosis**: Common toenail infection, difficult to treat, affecting more than 2% of population.

- Classical treatments with antifungal compounds need to be administered until the whole nail has grown out healthy.

- New compounds have been developed which reduce treatment to 3 months.

- Randomized, double-blind, parallel group, multicenter study for the comparison of two such new compounds ($A$ and $B$) for oral treatment.
• Research question:

Are both treatments equally effective for the treatment of TDO?

• $2 \times 189$ patients randomized, 36 centers

• 48 weeks of total follow up (12 months)

• 12 weeks of treatment (3 months)

• Measurements at months 0, 1, 2, 3, 6, 9, 12.
• Response considered here: Unaffected nail length (mm):

Please mark each infected nail with a "X". Indicate the TARGET nail, which you selected at baseline, for assessment with a circle.

RIGHT FOOT

LEFT FOOT

Please indicate on the diagram the margin of the unaffected target nail.
• As response is related to toe size, we restrict to patients with big toenail as target nail \( \rightarrow \) 150 and 148 subjects.

• 30 randomly selected profiles, in each group:

![Graphs showing longitudinal data for Treatment A and Treatment B](image-url)
• Complication: Dropout (24%):

<table>
<thead>
<tr>
<th>Time (months)</th>
<th>Treatment A</th>
<th>Treatment B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>150</td>
<td>148</td>
<td>298</td>
</tr>
<tr>
<td>1</td>
<td>149</td>
<td>142</td>
<td>291</td>
</tr>
<tr>
<td>2</td>
<td>146</td>
<td>138</td>
<td>284</td>
</tr>
<tr>
<td>3</td>
<td>140</td>
<td>131</td>
<td>271</td>
</tr>
<tr>
<td>6</td>
<td>131</td>
<td>124</td>
<td>255</td>
</tr>
<tr>
<td>9</td>
<td>120</td>
<td>109</td>
<td>229</td>
</tr>
<tr>
<td>12</td>
<td>118</td>
<td>108</td>
<td>226</td>
</tr>
</tbody>
</table>

• Remarks:
  - Much variability between subjects
  - Much variability within subjects
  - Fixed number of measurements scheduled per subject, but not all measurements available due to dropout, for unknown reason.
  - Measurements taken at fixed time points

Introduction to Longitudinal Data Analysis
1.7 Mastitis in Dairy Cattle

- Taken from Diggle & Kenward, Applied statistics (1994)

- Mastitis: Infectious disease, typically reducing milk yields

- Research question:

  Are high yielding cows more susceptible?

- Hence, is the probability of occurrence of mastitis related to the yield that would have been observed had mastitis not occurred?

- Hypothesis cannot be tested directly since ‘covariate is missing for all events’
• Individual profiles:

• Remarks:
  ▶ Paired observations: Most simple example of longitudinal data
  ▶ Much variability between cows
  ▶ Missingness process itself is of interest
1.8 The Baltimore Longitudinal Study of Aging (BLSA)


- BLSA: Ongoing, multidisciplinary observational study, started in 1958, with the study of normal human aging as primary objective

- Participants:
  - volunteers, predominantly white, well educated, and financially comfortable
  - return approximately every 2 years for 3 days of biomedical and psychological examinations
  - at first only males (over 1500 by now), later also females
  - an average of almost 7 visits and 16 years of follow-up
• The BLSA is a unique resource for rapidly evaluating longitudinal hypotheses:
  ▶ data from repeated clinical examinations
  ▶ a bank of frozen blood and urine samples

• Drawbacks of such observational studies:
  ▶ More complicated analyses needed (see later)
  ▶ Observed evolutions may be highly influenced by many covariates which may or may not be recorded in the study
1.8.1 Prostate Data

- References:

- Prostate disease is one of the most common and most costly medical problems in the United States

- Important to look for markers which can detect the disease at an early stage

- **Prostate-Specific Antigen** is an enzyme produced by both normal and cancerous prostate cells
• PSA level is related to the volume of prostate tissue.

• Problem: Patients with Benign Prostatic Hyperplasia also have an increased PSA level

• Overlap in PSA distribution for cancer and BPH cases seriously complicates the detection of prostate cancer.

• Research question (hypothesis based on clinical practice):

  Can longitudinal PSA profiles be used to detect prostate cancer in an early stage?
A retrospective case-control study based on frozen serum samples:

- 16 control patients
- 20 BPH cases
- 14 local cancer cases
- 4 metastatic cancer cases

Complication: No perfect match for age at diagnosis and years of follow-up possible

Hence, analyses will have to correct for these age differences between the diagnostic groups.
• Individual profiles:

\[ \ln(1 + \text{PSA}) \]

**Controls**

- \( N = 16 \)
- \( 4 \leq n_i \leq 10 \)
- \( 9.4 \leq t_{in} \leq 16.8 \)
- \( 56.7 \leq \text{age} \leq 80.5 \)

**BPH cases**

- \( N = 20 \)
- \( 5 \leq n_i \leq 11 \)
- \( 6.9 \leq t_{in} \leq 24.1 \)
- \( 64.6 \leq \text{age} \leq 86.7 \)

**L/R cancer cases**

- \( N = 14 \)
- \( 7 \leq n_i \leq 15 \)
- \( 10.6 \leq t_{in} \leq 24.9 \)
- \( 63.6 \leq \text{age} \leq 85.4 \)

**Metastatic cancer cases**

- \( N = 4 \)
- \( 7 \leq n_i \leq 12 \)
- \( 10 \leq t_{in} \leq 25.3 \)
- \( 62.7 \leq \text{age} \leq 82.8 \)
Remarks:
- Much variability between subjects
- Little variability within subjects
- Highly unbalanced data
1.8.2 Hearing Data

- References:

- Hearing thresholds, by means of sound proof chamber and Bekesy audiometer

- 11 frequencies: 125 → 8000 Hz, both ears

- Research question:
  
  How does hearing depend on aging?
• Data considered here:
  ▶ 500 Hz
  ▶ 6170 observations (3089 left ear, 3081 right ear) from 681 males without any otologic disease
  ▶ followed for up to 22 years, with a maximum of 15 measurements/subject

• 30 randomly selected profiles, for each ear:
Remarks:

- Much variability between subjects
- Much variability within subjects
- Highly unbalanced data
Chapter 2
Cross-sectional versus Longitudinal Data

- Introduction
- Paired versus unpaired $t$-test
- Cross-sectional versus longitudinal data
2.1 Introduction

- The examples have illustrated several aspects of longitudinal data structures:
  - Experimental and observational
  - Balanced and unbalanced
  - With or without missing data (dropout)

- Often, there is far more variability between subjects than within subjects.

- This is also reflected in correlation within units.
• For example, for the growth curves, the correlation matrix of the 5 repeated measurements equals

\[
\begin{pmatrix}
1.00 & 0.95 & 0.96 & 0.93 & 0.87 \\
0.95 & 1.00 & 0.97 & 0.96 & 0.89 \\
0.96 & 0.97 & 1.00 & 0.98 & 0.94 \\
0.93 & 0.96 & 0.98 & 1.00 & 0.98 \\
0.87 & 0.89 & 0.94 & 0.98 & 1.00 \\
\end{pmatrix}
\]

• This correlation structure cannot be ignored in the analyses (Section 2.2)

• The advantage however is that longitudinal data allow to study changes within subjects (Section 2.3).
2.2 Paired versus Unpaired $t$-test

2.2.1 Paired $t$-test

- The simplest case of longitudinal data are paired data

- We re-consider the diastolic blood pressures from the Captopril data

- The data can be summarized as:

![Descriptive Statistics Table](image)
• There is an average decrease of more than 9 mmHG

• The classical analysis of paired data is based on comparisons within subjects:

\[ \Delta_i = Y_{i1} - Y_{i2}, \quad i = 1, \ldots, 15 \]

• A positive \( \Delta_i \) corresponds to a decrease of the BP, while a negative \( \Delta_i \) is equivalent to an increase.

• Testing for treatment effect is now equivalent to testing whether the average difference \( \mu_\Delta \) equals zero.
● Statistica output:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std.Dv.</th>
<th>N</th>
<th>Diff.</th>
<th>Std.Dv. Diff.</th>
<th>t</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIA_NA</td>
<td>103.07</td>
<td>12.555</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Marked differences are significant at $p < .05000$

● Hence, the average change in BP is statistically, significantly different from zero ($p = 0.001$).
2.2.2 Unpaired, Two-sample, $t$-test

- What if we had ignored the paired nature of the data?

- We then could have used a two-sample (unpaired) $t$-test to compare the average BP of untreated patients (controls) with treated patients.

- We would still have found a significant difference ($p = 0.0366$), but the $p$-value would have been more than $30 \times$ larger compared to the one obtained using the paired $t$-test ($p = 0.001$).

- Conclusion:

$$15 \times 2 \neq 30 \times 1$$
• The two-sample \( t \)-test does not take into account the fact that the 30 measurements are not independent observations.

• This illustrates that classical statistical models which assume independent observations will not be valid for the analysis of longitudinal data.
2.3 Cross-sectional versus Longitudinal Data

- Suppose it is of interest to study the relation between some response $Y$ and age.
- A cross-sectional study yields the following data:

The graph suggests a negative relation between $Y$ and age.
• Exactly the same observations could also have been obtained in a longitudinal study, with 2 measurements per subject.

• First case:

Are we now still inclined to conclude that there is a negative relation between \( Y \) and Age?
• The graph suggests a negative cross-sectional relation but a positive longitudinal trend.

• Second case:

• The graph now suggests the cross-sectional as well as longitudinal trend to be negative.
• Conclusion:

Longitudinal data allow to distinguish differences between subjects from changes within subjects

• Application: Growth curves for babies (next page)
Chapter 3
Simple Methods

▷ Introduction
▷ Overview of frequently used methods
▷ Summary statistics
3.1 Introduction

- The reason why classical statistical techniques fail in the context of longitudinal data is that observations within subjects are correlated.

- In many cases the correlation between two repeated measurements decreases as the time span between those measurements increases.

- A correct analysis should account for this.

- The paired $t$-test accounts for this by considering subject-specific differences $\Delta_i = Y_{i1} - Y_{i2}$.

- This reduces the number of measurements to just one per subject, which implies that classical techniques can be applied again.
• In the case of more than 2 measurements per subject, similar simple techniques are often applied to reduce the number of measurements for the $i$th subject, from $n_i$ to 1.

• Some examples:
  ▶ Analysis at each time point separately
  ▶ Analysis of Area Under the Curve (AUC)
  ▶ Analysis of endpoints
  ▶ Analysis of increments
  ▶ Analysis of covariance
3.2 Overview of Frequently Used Methods

3.2.1 Analysis at Each Time Point

- The data are analysed at each occasion separately.

- Advantages:
  - Simple to interpret
  - Uses all available data

- Disadvantages:
  - Does not consider ‘overall’ differences
  - Does not allow to study evolution differences
  - Problem of multiple testing
3.2.2 Analysis of Area Under the Curve

• For each subject, the area under its curve is calculated:

\[ AUC_i = (t_{i2} - t_{i1}) \times (y_{i1} + y_{i2})/2 + (t_{i3} - t_{i2}) \times (y_{i2} + y_{i3})/2 + \ldots \]

• Afterwards, these \( AUC_i \) are analyzed.

• Advantages:

  ▶ No problems of multiple testing
  ▶ Does not explicitly assume balanced data
  ▶ Compares ‘overall’ differences

• Disadvantage: Uses only partial information: \( AUC_i \)
3.2.3 Analysis of Endpoints

• In randomized studies, there are no systematic differences at baseline.

• Hence, ‘treatment’ effects can be assessed by only comparing the measurements at the last occasion.

• Advantages:
  ▶ No problems of multiple testing
  ▶ Does not explicitly assume balanced data

• Disadvantages:
  ▶ Uses only partial information: $y_{in_i}$
  ▶ Only valid for large data sets
3.2.4 Analysis of Increments

- A simple method to compare evolutions between subjects, correcting for differences at baseline, is to analyze the subject-specific changes $y_{in_i} - y_{i1}$.

- Advantages:
  - No problems of multiple testing
  - Does not explicitly assume balanced data

- Disadvantage: Uses only partial information: $y_{in_i} - y_{i1}$
3.2.5 Analysis of Covariance

- Another way to analyse endpoints, correcting for differences at baseline, is to use analysis of covariance techniques, where the first measurement is included as covariate in the model.

- Advantages:
  - No problems of multiple testing
  - Does not explicitly assume balanced data

- Disadvantages:
  - Uses only partial information: $y_{i1}$ and $y_{in_i}$
  - Does not take into account the variability of $y_{i1}$
3.3 Summary Statistics

- The AUC, endpoints and increments are examples of summary statistics.

- Such summary statistics summarize the vector of repeated measurements for each subject separately.

- This leads to the following general procedure:
  - **Step 1**: Summarize data of each subject into one statistic, a summary statistic.
  - **Step 2**: Analyze the summary statistics, e.g. analysis of covariance to compare groups after correction for important covariates.

- This way, the analysis of longitudinal data is reduced to the analysis of independent observations, for which classical statistical procedures are available.
• However, all these methods have the disadvantage that (lots of) information is lost

• Further, they often do not allow to draw conclusions about the way the endpoint has been reached:

![Hypothetical average evolutions](image)
Chapter 4
The Multivariate Regression Model

▷ The general multivariate model
▷ Model fitting with SAS
▷ Model reduction
▷ Remarks
4.1 The General Multivariate Model

• We re-consider the growth data:
This is a completely balanced data set:
- 4 measurements for all subjects
- measurements taken at exactly the same time points

Let $Y_i$ be the vector of $n$ repeated measurements for the $i$th subject:

$$Y_i = \left( Y_{i1} \ Y_{i2} \ \ldots \ Y_{in} \right)'$$

The general multivariate model assumes that $Y_i$ satisfies a regression model

$$Y_i = X_i \beta + \varepsilon_i \quad \text{with} \quad \begin{cases} 
X_i : \text{matrix of covariates} \\
\beta : \text{vector of regression parameters} \\
\varepsilon_i : \text{vector of error components}, \varepsilon_i \sim N(0, \Sigma)
\end{cases}$$
• We then have the following distribution for $Y_i$:
  \[ Y_i \sim N(X_i \beta, \Sigma) \]

• The mean structure $X_i \beta$ is modelled as in classical linear regression and ANOVA models.

• Usually, $\Sigma$ is just a general $(n \times n)$ covariance matrix. However, special structures for $\Sigma$ can be assumed (see later).

• Assuming independence across individuals, $\beta$ and the parameters in $\Sigma$ can be estimated by maximizing

  \[
  L_{ML} = \prod_{i=1}^{N} \left\{ (2\pi)^{-n/2}\ |\Sigma|^{-\frac{1}{2}} \exp \left( -\frac{1}{2} (y_i - X_i \beta)' \Sigma^{-1} (y_i - X_i \beta) \right) \right\}
  \]
• Inference is based on classical maximum likelihood theory:
  ▶ LR tests
  ▶ Asymptotic WALD tests

• More details on inference will be discussed later
4.2 Model Fitting With SAS

4.2.1 Model Parameterization

• As an example, we fit a model with unstructured mean and unstructured covariance matrix to the growth data (Model 1).

• Let \( x_i \) be equal to 0 for a boy, and equal to 1 for a girl

• One possible parameterization of the model is

\[
Y_{i1} = \beta_{0,8}(1 - x_i) + \beta_{1,8}x_i + \varepsilon_{i1} \\
Y_{i2} = \beta_{0,10}(1 - x_i) + \beta_{1,10}x_i + \varepsilon_{i2} \\
Y_{i3} = \beta_{0,12}(1 - x_i) + \beta_{1,12}x_i + \varepsilon_{i3} \\
Y_{i4} = \beta_{0,14}(1 - x_i) + \beta_{1,14}x_i + \varepsilon_{i4}
\]
In matrix notation:

\[ Y_i = X_i \beta + \varepsilon_i, \]

with

\[
X_i = \begin{pmatrix}
(1 - x_i) & 0 & 0 & 0 & x_i & 0 & 0 & 0 \\
0 & (1 - x_i) & 0 & 0 & 0 & x_i & 0 & 0 \\
0 & 0 & (1 - x_i) & 0 & 0 & 0 & x_i & 0 \\
0 & 0 & 0 & (1 - x_i) & 0 & 0 & 0 & x_i \\
\end{pmatrix}
\]

and with

\[
\beta = (\beta_{0,8}, \beta_{0,10}, \beta_{0,12}, \beta_{0,14}, \beta_{1,8}, \beta_{1,10}, \beta_{1,12}, \beta_{1,14})'
\]
4.2.2 SAS Program

- SAS syntax:

```sas
proc mixed data = growth method = ml;
class idnr sex age;
model measure = age*sex / noint s;
repeated age / type = un subject = idnr;
run;
```

- Data structure:

<table>
<thead>
<tr>
<th>idnr</th>
<th>age</th>
<th>sex</th>
<th>measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0000</td>
<td>8.0000</td>
<td>1.0000</td>
<td>21.0000</td>
</tr>
<tr>
<td>1.0000</td>
<td>10.0000</td>
<td>1.0000</td>
<td>20.0000</td>
</tr>
<tr>
<td>1.0000</td>
<td>12.0000</td>
<td>1.0000</td>
<td>21.5000</td>
</tr>
<tr>
<td>1.0000</td>
<td>14.0000</td>
<td>1.0000</td>
<td>23.0000</td>
</tr>
<tr>
<td>2.0000</td>
<td>8.0000</td>
<td>1.0000</td>
<td>21.0000</td>
</tr>
<tr>
<td>2.0000</td>
<td>10.0000</td>
<td>1.0000</td>
<td>21.5000</td>
</tr>
</tbody>
</table>

......

| 26.0000 | 12.0000 | 0.0000 | 26.0000 |
| 26.0000 | 14.0000 | 0.0000 | 30.0000 |
| 27.0000 | 8.0000  | 0.0000 | 22.0000 |
| 27.0000 | 10.0000 | 0.0000 | 21.5000 |
| 27.0000 | 12.0000 | 0.0000 | 23.5000 |
| 27.0000 | 14.0000 | 0.0000 | 25.0000 |
• The mean is modeled in the MODEL statement, as in other SAS procedures for linear models

• The covariance matrix is modeled in the REPEATED statement:
  ▶ option ‘type=’ specifies covariance structure
  ▶ option ‘subject=idnr’ specifies the clusters in the data set
  ▶ the variable ‘age’ is used to order measurements within clusters
4.2.3 Results

- Maximized log-likelihood value: $\ell = -208.25$

- Estimates for parameters in mean structure, and implied fitted averages:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MLE</th>
<th>(s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{0,8}$</td>
<td>22.8750</td>
<td>(0.5598)</td>
</tr>
<tr>
<td>$\beta_{0,10}$</td>
<td>23.8125</td>
<td>(0.4921)</td>
</tr>
<tr>
<td>$\beta_{0,12}$</td>
<td>25.7188</td>
<td>(0.6112)</td>
</tr>
<tr>
<td>$\beta_{0,14}$</td>
<td>27.4688</td>
<td>(0.5371)</td>
</tr>
<tr>
<td>$\beta_{1,8}$</td>
<td>21.1818</td>
<td>(0.6752)</td>
</tr>
<tr>
<td>$\beta_{1,10}$</td>
<td>22.2273</td>
<td>(0.5935)</td>
</tr>
<tr>
<td>$\beta_{1,12}$</td>
<td>23.0909</td>
<td>(0.7372)</td>
</tr>
<tr>
<td>$\beta_{1,14}$</td>
<td>24.0909</td>
<td>(0.6478)</td>
</tr>
</tbody>
</table>

![Graph showing unstructured means and covariance with data points for girls and boys.](image)
• Fitted covariance and correlation matrices:

\[
\Sigma = \begin{pmatrix}
5.0143 & 2.5156 & 3.6206 & 2.5095 \\
2.5156 & 3.8748 & 2.7103 & 3.0714 \\
3.6206 & 2.7103 & 5.9775 & 3.8248 \\
2.5095 & 3.0714 & 3.8248 & 4.6164
\end{pmatrix} \quad \Rightarrow \quad \begin{pmatrix}
1.0000 & 0.5707 & 0.6613 & 0.5216 \\
0.5707 & 1.0000 & 0.5632 & 0.7262 \\
0.6613 & 0.5632 & 1.0000 & 0.7281 \\
0.5216 & 0.7262 & 0.7281 & 1.0000
\end{pmatrix}
\]
4.3 Model Reduction

- In many circumstances, one will be interested in reducing the model.

- For the growth data for example, one may be interested in finding out whether the fitted average profiles can be well described by straight lines.

- Also, the covariance matrix contained 10 parameters, not even of interest. If this can be reduced, one may gain efficiency for the mean structure.

- In practice, one therefore usually tries to reduce the mean and covariance structures, yielding more parsimonious models.

- This is now illustrated using the growth data.
4.3.1 Reduction of the Mean Structure

Model 2: Linear Average Trends

- Linear average trend within each group, unstructured $4 \times 4$ covariance matrix $\Sigma$

- Model 2 is given by ($x_i = 1$ for girls):

$$Y_{ij} = \beta_0 + \beta_{01}x_i + \beta_{10}t_j(1 - x_i) + \beta_{11}t_jx_i + \varepsilon_{ij},$$

- In matrix notation, this equals $Y_i = X_i\beta + \varepsilon_i$, with design matrix

$$X_i = \begin{pmatrix}
1 & x_i & 8(1 - x_i) & 8x_i \\
1 & x_i & 10(1 - x_i) & 10x_i \\
1 & x_i & 12(1 - x_i) & 12x_i \\
1 & x_i & 14(1 - x_i) & 14x_i
\end{pmatrix}.$$
Parameterization $\beta = (\beta_0, \beta_{01}, \beta_{10}, \beta_{11})'$:

- $\beta_0$ : intercept for boys
- $\beta_0 + \beta_{01}$ : intercept for girls
- $\beta_{10}$ : slope for boys
- $\beta_{11}$ : slope for girls

SAS program:

```sas
proc mixed data = growth method = ml;
class idnr sex ageclss;
model measure = sex age*sex / s;
repeated ageclss / type = un subject = idnr;
run;
```

The variable ageclss is a copy of the original variable age.
• LR test Model 2 versus Model 1:

<table>
<thead>
<tr>
<th>Mean</th>
<th>Covar</th>
<th>par</th>
<th>$-2\ell$</th>
<th>Ref</th>
<th>$G^2$</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 unstr.</td>
<td>unstr.</td>
<td>18</td>
<td>416.509</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 ≠ slopes</td>
<td>unstr.</td>
<td>14</td>
<td>419.477</td>
<td>1</td>
<td>2.968</td>
<td>4</td>
<td>0.5632</td>
</tr>
</tbody>
</table>

• Predicted trends:  
  girls : $\hat{Y}_j = 17.43 + 0.4764t_j$  
  boys : $\hat{Y}_j = 15.84 + 0.8268t_j$
Model 3: Parallel Average Profiles

- Linear average trend within each sex group, the same slope for both groups

- Unstructured $4 \times 4$ covariance matrix $\Sigma$

- Model 3 is given by:

$$Y_{ij} = \beta_0 + \beta_{01}x_i + \beta_1t_j + \varepsilon_{ij}.$$  

- In matrix notation, this equals $Y_i = X_i\beta + \varepsilon_i$, with design matrix

$$X_i = \begin{pmatrix}
1 & x_i & 8 \\
1 & x_i & 10 \\
1 & x_i & 12 \\
1 & x_i & 14
\end{pmatrix}$$
• Parameterization $\mathbf{\beta} = (\beta_0, \beta_{01}, \beta_1)'$:
  
  $\triangleright$ $\beta_0$: intercept for boys
  
  $\triangleright$ $\beta_0 + \beta_{01}$: intercept for girls
  
  $\triangleright$ $\beta_1$: common slope for boys and girls

• SAS program:

```sas
proc mixed data = growth method = ml;
class idnr sex ageclss;
model measure = sex age / s;
repeated ageclss / type = un subject = idnr;
run;
```

• LR test:

<table>
<thead>
<tr>
<th>Mean</th>
<th>Covar</th>
<th>par</th>
<th>$-2\ell$</th>
<th>Ref</th>
<th>$G^2$</th>
<th>df</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>unstr.</td>
<td>unstr.</td>
<td>18</td>
<td>416.509</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>$\neq$ slopes</td>
<td>unstr.</td>
<td>14</td>
<td>419.477</td>
<td>1</td>
<td>2.968</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>$=$ slopes</td>
<td>unstr.</td>
<td>13</td>
<td>426.153</td>
<td>2</td>
<td>6.676</td>
<td>1</td>
</tr>
</tbody>
</table>
• Predicted trends: girls: $\hat{Y}_j = 15.37 + 0.6747t_j$  boys: $\hat{Y}_j = 17.42 + 0.6747t_j$
4.3.2 Reduction of the Covariance Structure

- In order to reduce the number of parameters in the covariance structure, we can now fit models with more parsimonious structures.

- This often leads to more efficient inferences for the mean parameters.

- This is particularly useful when many repeated measurements are taken per subject.

- SAS includes a large variety of covariance structures (see SAS help function).
- Some examples:

<table>
<thead>
<tr>
<th>Structure</th>
<th>Example</th>
</tr>
</thead>
</table>
| Unstructured type=UN          | \(
|                              | \begin{pmatrix}
|                              | \sigma_1^2 & \sigma_{12} & \sigma_{13} \\
|                              | \sigma_{12} & \sigma_2^2 & \sigma_{23} \\
|                              | \sigma_{13} & \sigma_{23} & \sigma_3^2 \\
|                              | \end{pmatrix}
|                              |                                                          |
| Simple type=SIMPLE            | \(
|                              | \begin{pmatrix}
|                              | \sigma^2 & 0 & 0 \\
|                              | 0 & \sigma^2 & 0 \\
|                              | 0 & 0 & \sigma^2 \\
|                              | \end{pmatrix}
|                              |                                                          |
| Compound symmetry type=CS     | \(
|                              | \begin{pmatrix}
|                              | \sigma_1^2 + \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\
|                              | \sigma_1^2 & \sigma_1^2 + \sigma_1^2 & \sigma_1^2 + \sigma_1^2 \\
|                              | \sigma_1^2 & \sigma_1^2 & \sigma_1^2 + \sigma_1^2 \\
|                              | \end{pmatrix}
|                              |                                                          |
| Banded type=UN(2)             | \(
|                              | \begin{pmatrix}
|                              | \sigma_1^2 & \sigma_{12} & 0 \\
|                              | \sigma_{12} & \sigma_2^2 & \sigma_{23} \\
|                              | 0 & \sigma_{23} & \sigma_3^2 \\
|                              | \end{pmatrix}
|                              |                                                          |
| First-order autoregressive type=AR(1) | \(
|                              | \begin{pmatrix}
|                              | \sigma^2 & \rho \sigma^2 & \rho^2 \sigma^2 \\
|                              | \rho \sigma^2 & \sigma^2 & \rho \sigma^2 \\
|                              | \rho^2 \sigma^2 & \rho \sigma^2 & \sigma^2 \\
|                              | \end{pmatrix}
|                              |                                                          |

<table>
<thead>
<tr>
<th>Structure</th>
<th>Example</th>
</tr>
</thead>
</table>
| Toeplitz type=TOEP            | \(
|                              | \begin{pmatrix}
|                              | \sigma^2 & \sigma_{12} & \sigma_{13} \\
|                              | \sigma_{12} & \sigma_2^2 & \sigma_{12} \\
|                              | \sigma_{13} & \sigma_{12} & \sigma_3^2 \\
|                              | \end{pmatrix}
|                              |                                                          |
| Toeplitz (1) type=Toep(1)    | \(
|                              | \begin{pmatrix}
|                              | \sigma^2 & 0 & 0 \\
|                              | 0 & \sigma^2 & 0 \\
|                              | 0 & 0 & \sigma^2 \\
|                              | \end{pmatrix}
|                              |                                                          |
| Heterogeneous compound symmetry type=CSH | \(
|                              | \begin{pmatrix}
|                              | \sigma_1^2 & \rho \sigma_1 \sigma_2 & \rho \sigma_1 \sigma_3 \\
|                              | \rho \sigma_1 \sigma_2 & \sigma_2^2 & \rho \sigma_2 \sigma_3 \\
|                              | \rho \sigma_1 \sigma_3 & \rho \sigma_2 \sigma_3 & \sigma_3^2 \\
|                              | \end{pmatrix}
|                              |                                                          |
| Heterogeneous first-order autoregressive type=ARH(1) | \(
|                              | \begin{pmatrix}
|                              | \sigma_1^2 & \rho \sigma_1 \sigma_2 & \rho^2 \sigma_1 \sigma_3 \\
|                              | \rho \sigma_1 \sigma_2 & \sigma_2^2 & \rho \sigma_2 \sigma_3 \\
|                              | \rho^2 \sigma_1 \sigma_3 & \rho \sigma_2 \sigma_3 & \sigma_3^2 \\
|                              | \end{pmatrix}
|                              |                                                          |
| Heterogeneous Toeplitz type=TOEPH | \(
|                              | \begin{pmatrix}
|                              | \sigma_1^2 & \rho_1 \sigma_1 \sigma_2 & \rho_2 \sigma_1 \sigma_3 \\
|                              | \rho_1 \sigma_1 \sigma_2 & \sigma_2^2 & \rho_1 \sigma_2 \sigma_3 \\
|                              | \rho_2 \sigma_1 \sigma_3 & \rho_1 \sigma_2 \sigma_3 & \sigma_3^2 \\
|                              | \end{pmatrix}
|                              |                                                          |
Model 4: Toeplitz Covariance Structure

- Linear average trend within each sex group

- The estimated covariance matrix (s.e.) of the unstructured covariance matrix under Model 2 equals:

\[
\begin{pmatrix}
5.12(1.42) & 2.44(0.98) & 3.61(1.28) & 2.52(1.06) \\
2.44(0.98) & 3.93(1.08) & 2.72(1.07) & 3.06(1.01) \\
3.61(1.28) & 2.72(1.07) & 5.98(1.63) & 3.82(1.25) \\
2.52(1.06) & 3.06(1.01) & 3.82(1.25) & 4.62(1.26)
\end{pmatrix}
\]

- This suggests that a possible model reduction could consist of assuming equal variances, and banded covariances.
This is the so-called Toeplitz covariance matrix $\Sigma$, with elements of the form $\Sigma_{ij} = \alpha_{|i-j|}$:

$$
\begin{pmatrix}
\alpha_0 & \alpha_1 & \alpha_2 & \alpha_3 \\
\alpha_1 & \alpha_0 & \alpha_1 & \alpha_2 \\
\alpha_2 & \alpha_1 & \alpha_0 & \alpha_1 \\
\alpha_3 & \alpha_2 & \alpha_1 & \alpha_0 \\
\end{pmatrix}
$$

Note that this is only really meaningful when the time points at which measurements are taken are equally spaced, as in the current example.

SAS program:

```sas
proc mixed data = growth method = ml;
class sex idnr ageclss;
model measure = sex age*sex / s;
repeated ageclss / type = toep subject = idnr;
run;
```
• LR test Model 4 versus Model 2:

<table>
<thead>
<tr>
<th>Mean</th>
<th>Covar</th>
<th>par</th>
<th>$-2\ell$</th>
<th>Ref</th>
<th>$G^2$</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 unstr.</td>
<td>unstr.</td>
<td>18</td>
<td>416.509</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 $\neq$ slopes</td>
<td>unstr.</td>
<td>14</td>
<td>419.477</td>
<td>1</td>
<td>2.968</td>
<td>4</td>
<td>0.5632</td>
</tr>
<tr>
<td>4 $\neq$ slopes</td>
<td>banded</td>
<td>8</td>
<td>424.643</td>
<td>2</td>
<td>5.166</td>
<td>6</td>
<td>0.5227</td>
</tr>
</tbody>
</table>

• Fitted covariance and correlation matrices:

\[
\Sigma = \begin{pmatrix}
4.9439 & 3.0507 & 3.4054 & 2.3421 \\
3.0507 & 4.9439 & 3.0507 & 3.4054 \\
3.4054 & 3.0507 & 4.9439 & 3.0507 \\
2.3421 & 3.4054 & 3.0507 & 4.9439 \\
\end{pmatrix} \quad \Rightarrow \quad \begin{pmatrix}
1.0000 & 0.6171 & 0.6888 & 0.4737 \\
0.6171 & 1.0000 & 0.6171 & 0.6888 \\
0.6888 & 0.6171 & 1.0000 & 0.6171 \\
0.4737 & 0.6888 & 0.6171 & 1.0000 \\
\end{pmatrix}
\]
Model 5: AR(1) Covariance Structure

- Linear average trend within each sex group

- The AR(1) covariance structure assumes exponentially decaying correlations, i.e., elements of $\Sigma$ of the form $\Sigma_{ij} = \sigma^2 \rho^{|i-j|}$:

$$
\Sigma = \sigma^2 \begin{pmatrix}
1 & \rho & \rho^2 & \rho^3 \\
\rho & 1 & \rho & \rho^2 \\
\rho^2 & \rho & 1 & \rho \\
\rho^3 & \rho^2 & \rho & 1
\end{pmatrix}
$$

- Note that this is also only really meaningful when the time points at which measurements are taken are equally spaced.
• **SAS program:**

```
proc mixed data = growth method = ml;
class sex idnr ageclss;
model measure = sex age*sex / s;
repeated ageclss / type = AR(1) subject = idnr;
run;
```

• **LR test Model 5 versus Models 2 and 4 :**

<table>
<thead>
<tr>
<th>Mean</th>
<th>Covar</th>
<th>par</th>
<th>$-2\ell$</th>
<th>Ref</th>
<th>$G^2$</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 unstr.</td>
<td>unstr.</td>
<td>18</td>
<td>416.509</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 ≠ slopes</td>
<td>unstr.</td>
<td>14</td>
<td>419.477</td>
<td>1</td>
<td>2.968</td>
<td>4</td>
<td>0.5632</td>
</tr>
<tr>
<td>4 ≠ slopes</td>
<td>banded</td>
<td>8</td>
<td>424.643</td>
<td>2</td>
<td>5.166</td>
<td>6</td>
<td>0.5227</td>
</tr>
<tr>
<td>5 ≠ slopes</td>
<td>AR(1)</td>
<td>6</td>
<td>440.681</td>
<td>2</td>
<td>21.204</td>
<td>8</td>
<td>0.0066</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>16.038</td>
<td>4</td>
<td></td>
<td></td>
<td>0.0003</td>
</tr>
</tbody>
</table>
• Fitted covariance and correlation matrices:

\[
\begin{pmatrix}
4.8903 & 2.9687 & 1.8021 & 1.0940 \\
2.9687 & 4.8903 & 2.9687 & 1.8021 \\
1.8021 & 2.9687 & 4.8903 & 2.9687 \\
1.0940 & 1.8021 & 2.9687 & 4.8903
\end{pmatrix}
\Rightarrow
\begin{pmatrix}
1.0000 & 0.6070 & 0.3685 & 0.2237 \\
0.6070 & 1.0000 & 0.6070 & 0.3685 \\
0.3685 & 0.6070 & 1.0000 & 0.6070 \\
0.2237 & 0.3685 & 0.6070 & 1.0000
\end{pmatrix}
\]
4.4 Remarks

- The multivariate regression model is primarily suitable when measurements are taken at a relatively small number of fixed time points.

- Even if some measurements are missing, the multivariate regression model can be applied, as long as the software allows for unequal numbers of measurements per subject.

- In the SAS procedure MIXED, this is taken care of in the REPEATED statement:

  repeated ageclass / ;

  from which it can be derived which outcomes have been observed, and which ones are missing.
• In case of large numbers of repeated measurements:
  ▶ Multivariate regression models can only be applied under very specific mean and covariance structures, even in case of complete balance.
  ▶ For example, unstructured means and/or unstructured covariances require estimation of very many parameters

• In case of highly unbalanced data:
  ▶ Multivariate regression models can only be applied under very specific mean and covariance structures.
  ▶ For example, Toeplitz and AR(1) covariances are not meaningful since time points are not equally spaced.
  ▶ For example, compound symmetric covariances are meaningful, but based on very strong assumptions.
Chapter 5
A Model for Longitudinal Data

- Introduction
- The 2-stage model formulation
- Examples: Rat and prostate data
- The general linear mixed-effects model
- Hierarchical versus marginal model
- Examples: Rat and prostate data
- A model for the residual covariance structure
5.1 Introduction

- In practice: often unbalanced data:
  - unequal number of measurements per subject
  - measurements not taken at fixed time points
- Therefore, multivariate regression techniques are often not applicable
- Often, subject-specific longitudinal profiles can be well approximated by linear regression functions
- This leads to a 2-stage model formulation:
  - **Stage 1:** Linear regression model for each subject separately
  - **Stage 2:** Explain variability in the subject-specific regression coefficients using known covariates
5.2 A 2-stage Model Formulation

5.2.1 Stage 1

- Response $Y_{ij}$ for $i$th subject, measured at time $t_{ij}$, $i = 1, \ldots, N$, $j = 1, \ldots, n_i$

- Response vector $Y_i$ for $i$th subject: $Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{ini})'$

- Stage 1 model:

$$Y_i = Z_i \beta_i + \varepsilon_i$$
• $Z_i$ is a $(n_i \times q)$ matrix of known covariates

• $\beta_i$ is a $q$-dimensional vector of subject-specific regression coefficients

• $\epsilon_i \sim N(0, \Sigma_i)$, often $\Sigma_i = \sigma^2 I_{n_i}$

• Note that the above model describes the observed variability within subjects
5.2.2 Stage 2

- Between-subject variability can now be studied from relating the $\beta_i$ to known covariates

- Stage 2 model:

$$\beta_i = K_i \beta + b_i$$

- $K_i$ is a $(q \times p)$ matrix of known covariates

- $\beta$ is a $p$-dimensional vector of unknown regression parameters

- $b_i \sim N(0, D)$
5.3 Example: The Rat Data

- Individual profiles:

![Graphs showing individual profiles of rats in Control, Low dose, and High dose groups with response (pixels) against age (days).]
• Transformation of the time scale to linearize the profiles:

\[ \text{Age}_{ij} \longrightarrow t_{ij} = \ln[1 + (\text{Age}_{ij} - 45)/10] \]

• Note that \( t = 0 \) corresponds to the start of the treatment (moment of randomization)

• Stage 1 model:

\[ Y_{ij} = \beta_1 + \beta_2 t_{ij} + \varepsilon_{ij}, \quad j = 1, \ldots, n_i \]

• Matrix notation:

\[ Y_i = Z_i \beta_i + \varepsilon_i \quad \text{with} \quad Z_i = \begin{pmatrix} 1 & t_{i1} \\ 1 & t_{i2} \\ \vdots & \vdots \\ 1 & t_{in_i} \end{pmatrix} \]
• In the second stage, the subject-specific intercepts and time effects are related to the treatment of the rats

• Stage 2 model:

\[
\begin{align*}
\beta_{1i} &= \beta_0 + b_{1i}, \\
\beta_{2i} &= \beta_1 L_i + \beta_2 H_i + \beta_3 C_i + b_{2i},
\end{align*}
\]

• \(L_i, H_i, \) and \(C_i\) are indicator variables:

\[
\begin{align*}
L_i &= \begin{cases} 
1 & \text{if low dose} \\
0 & \text{otherwise}
\end{cases}, &
H_i &= \begin{cases} 
1 & \text{if high dose} \\
0 & \text{otherwise}
\end{cases}, &
C_i &= \begin{cases} 
1 & \text{if control} \\
0 & \text{otherwise}
\end{cases}
\]
Parameter interpretation:

- $\beta_0$: average response at the start of the treatment (independent of treatment)
- $\beta_1$, $\beta_2$, and $\beta_3$: average time effect for each treatment group
5.4 Example: The Prostate Data

• Individual profiles:

![Graphs showing longitudinal data](image-url)
• Transformation of the response:

\[ \text{PSA}_{ij} \rightarrow Y_{ij} = \ln(\text{PSA}_{ij} + 1) \]

• Stage 1 model:

\[ Y_{ij} = \beta_1 i + \beta_2 t_{ij} + \beta_3 t_{ij}^2 + \varepsilon_{ij}, \quad j = 1, \ldots, n_i \]

• Matrix notation:

\[ Y_i = Z_i \beta_i + \varepsilon_i \quad \text{with} \quad Z_i = \begin{pmatrix} 1 & t_{i1} & t_{i1}^2 \\ 1 & t_{i2} & t_{i2}^2 \\ \vdots & \vdots & \vdots \\ 1 & t_{ini} & t_{ini}^2 \end{pmatrix} \]

• In the second stage, the subject-specific intercepts and time effects are related to the age (at diagnosis) and disease status
• Stage 2 model:

\[
\begin{align*}
\beta_{1i} &= \beta_1 \text{Age}_i + \beta_2 C_i + \beta_3 B_i + \beta_4 L_i + \beta_5 M_i + b_{1i}, \\
\beta_{2i} &= \beta_6 \text{Age}_i + \beta_7 C_i + \beta_8 B_i + \beta_9 L_i + \beta_{10} M_i + b_{2i}, \\
\beta_{3i} &= \beta_{11} \text{Age}_i + \beta_{12} C_i + \beta_{13} B_i + \beta_{14} L_i + \beta_{15} M_i + b_{3i}
\end{align*}
\]

• \(C_i, B_i, L_i\) and \(M_i\) are indicator variables:

\[
\begin{align*}
C_i &= \begin{cases} 
1 & \text{if Control} \\
0 & \text{otherwise}
\end{cases} \\
B_i &= \begin{cases} 
1 & \text{if BPH case} \\
0 & \text{otherwise}
\end{cases} \\
L_i &= \begin{cases} 
1 & \text{if L/R cancer case} \\
0 & \text{otherwise}
\end{cases} \\
M_i &= \begin{cases} 
1 & \text{if Metastatic cancer case} \\
0 & \text{otherwise}
\end{cases}
\]
• Parameter interpretation:
  ▶ $\beta_2$, $\beta_3$, $\beta_4$, and $\beta_5$: average intercepts after correction for age
  ▶ $\beta_7$, $\beta_8$, $\beta_9$, and $\beta_{10}$: average linear time effects after correction for age.
  ▶ $\beta_{12}$, $\beta_{13}$, $\beta_{14}$, and $\beta_{15}$: average quadratic time effects after correction for age.
5.5 The General Linear Mixed-effects Model

- A 2-stage approach can be performed explicitly in the analysis

- However, this is just another example of the use of summary statistics:
  - $Y_i$ is summarized by $\bar{\beta}_i$
  - summary statistics $\bar{\beta}_i$ analysed in second stage

- The associated drawbacks can be avoided by combining the two stages into one model:

$$
\begin{align*}
\{ & Y_i = Z_i \beta_i + \varepsilon_i \\
& \beta_i = K_i \beta + b_i \}
\Rightarrow
Y_i = \underbrace{Z_i K_i \beta}_{X_i} + Z_i b_i + \varepsilon_i = X_i \beta + Z_i b_i + \varepsilon_i
\end{align*}
$$
• General linear mixed-effects model:

\[
Y_i = X_i\beta + Z_i b_i + \varepsilon_i
\]

\[
b_i \sim N(0, D), \quad \varepsilon_i \sim N(0, \Sigma_i),
\]

\[
b_1, \ldots, b_N, \varepsilon_1, \ldots, \varepsilon_N \text{ independent}
\]

• Terminology:
  ▶ Fixed effects: \( \beta \)
  ▶ Random effects: \( b_i \)
  ▶ Variance components: elements in \( D \) and \( \Sigma_i \)
5.6 Hierarchical versus Marginal Model

- The general linear mixed model is given by:

\[
\begin{align*}
Y_i &= X_i \beta + Z_i b_i + \varepsilon_i \\
\begin{cases}
  b_i &\sim N(0, D), \\
  \varepsilon_i &\sim N(0, \Sigma_i), \\
  b_1, \ldots, b_N, \varepsilon_1, \ldots, \varepsilon_N &\text{independent}
\end{cases}
\end{align*}
\]

- It can be rewritten as:

\[
Y_i | b_i \sim N(X_i \beta + Z_i b_i, \Sigma_i), \quad b_i \sim N(0, D)
\]
• It is therefore also called a hierarchical model:
  ▶ A model for $Y_i$ given $b_i$
  ▶ A model for $b_i$

• Marginally, we have that $Y_i$ is distributed as:

$$Y_i \sim N(X_i\beta, Z_iDZ_i' + \Sigma_i)$$

• Hence, very specific assumptions are made about the dependence of mean and covariance on the covariates $X_i$ and $Z_i$:
  ▶ **Implied mean**: $X_i\beta$
  ▶ **Implied covariance**: $V_i = Z_iDZ_i' + \Sigma_i$

• Note that the hierarchical model implies the marginal one, **NOT** vice versa
5.7 Example: The Rat Data

- **Stage 1 model:** 
  \[ Y_{ij} = \beta_{1i} + \beta_{2i} t_{ij} + \varepsilon_{ij}, \quad j = 1, \ldots, n_i \]

- **Stage 2 model:** 
  \[ \begin{aligned} 
  \beta_{1i} &= \beta_0 + b_{1i}, \\
  \beta_{2i} &= \beta_1 L_i + \beta_2 H_i + \beta_3 C_i + b_{2i}, 
  \end{aligned} \]

- **Combined:** 
  \[ Y_{ij} = (\beta_0 + b_{1i}) + (\beta_1 L_i + \beta_2 H_i + \beta_3 C_i + b_{2i}) t_{ij} + \varepsilon_{ij} \]

  \[ = \begin{cases} 
  \beta_0 + b_{1i} + (\beta_1 + b_{2i}) t_{ij} + \varepsilon_{ij}, & \text{if low dose} \\
  \beta_0 + b_{1i} + (\beta_2 + b_{2i}) t_{ij} + \varepsilon_{ij}, & \text{if high dose} \\
  \beta_0 + b_{1i} + (\beta_3 + b_{2i}) t_{ij} + \varepsilon_{ij}, & \text{if control.} 
  \end{cases} \]
• Implied marginal mean structure:
  ▶ Linear average evolution in each group
  ▶ Equal average intercepts
  ▶ Different average slopes

• Implied marginal covariance structure ($\Sigma_i = \sigma^2 I_{n_i}$):

\[
\text{Cov}(Y_i(t_1), Y_i(t_2)) = \begin{pmatrix} 1 & t_1 \\ t_2 \end{pmatrix} D \begin{pmatrix} 1 \\ t_2 \end{pmatrix} + \sigma^2 \delta\{t_1,t_2\}
\]

\[
= d_{22} t_1 t_2 + d_{12} (t_1 + t_2) + d_{11} + \sigma^2 \delta\{t_1,t_2\}.
\]

• Note that the model implicitly assumes that the variance function is quadratic over time, with positive curvature $d_{22}$. 
A model which assumes that all variability in subject-specific slopes can be ascribed to treatment differences can be obtained by omitting the random slopes $b_{2i}$ from the above model:

$$Y_{ij} = (\beta_0 + b_{1i}) + (\beta_1 L_i + \beta_2 H_i + \beta_3 C_i) t_{ij} + \varepsilon_{ij}$$

$$= \begin{cases} 
\beta_0 + b_{1i} + \beta_1 t_{ij} + \varepsilon_{ij}, & \text{if low dose} \\
\beta_0 + b_{1i} + \beta_2 t_{ij} + \varepsilon_{ij}, & \text{if high dose} \\
\beta_0 + b_{1i} + \beta_3 t_{ij} + \varepsilon_{ij}, & \text{if control.} 
\end{cases}$$

This is the so-called random-intercepts model.

The same marginal mean structure is obtained as under the model with random slopes.
• Implied marginal covariance structure \((\Sigma_i = \sigma^2 I_{n_i})\):

\[
\text{Cov}(Y_i(t_1), Y_i(t_2)) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} D \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \sigma^2 \delta_{\{t_1, t_2\}}
\]

\[
= d_{11} + \sigma^2 \delta_{\{t_1, t_2\}}.
\]

• Hence, the implied covariance matrix is compound symmetry:
  - constant variance \(d_{11} + \sigma^2\)
  - constant correlation \(\rho_I = d_{11}/(d_{11} + \sigma^2)\) between any two repeated measurements within the same rat
5.8 Example: The Prostate Data

- Stage 1 model: 
  \[ Y_{ij} = \beta_{1i} + \beta_{2i}t_{ij} + \beta_{3i}t_{ij}^2 + \varepsilon_{ij}, \quad j = 1, \ldots, n_i \]

- Stage 2 model: 
  \[
  \begin{aligned}
  \beta_{1i} &= \beta_{11} \text{Age}_i + \beta_{21} C_i + \beta_{31} B_i + \beta_{41} L_i + \beta_{51} M_i + b_{1i}, \\
  \beta_{2i} &= \beta_{61} \text{Age}_i + \beta_{71} C_i + \beta_{81} B_i + \beta_{91} L_i + \beta_{101} M_i + b_{2i}, \\
  \beta_{3i} &= \beta_{111} \text{Age}_i + \beta_{121} C_i + \beta_{131} B_i + \beta_{141} L_i + \beta_{151} M_i + b_{3i},
  \end{aligned}
  \]

- Combined: 
  \[ Y_{ij} = \beta_{1} \text{Age}_i + \beta_{2} C_i + \beta_{3} B_i + \beta_{4} L_i + \beta_{5} M_i + \left(\beta_{6} \text{Age}_i + \beta_{7} C_i + \beta_{8} B_i + \beta_{9} L_i + \beta_{10} M_i\right) t_{ij} + \left(\beta_{11} \text{Age}_i + \beta_{12} C_i + \beta_{13} B_i + \beta_{14} L_i + \beta_{15} M_i\right) t_{ij}^2 + b_{1i} + b_{2i} t_{ij} + b_{3i} t_{ij}^2 + \varepsilon_{ij}. \]
• Implied marginal mean structure:
  - Quadratic average evolution in each group
  - Average intercept and linear as well as quadratic slopes corrected for age differences

• Implied marginal covariance structure ($\Sigma_i = \sigma^2 I_{n_i}$):

\[
\text{Cov}(Y_i(t_1), Y_i(t_2)) = \begin{pmatrix} 1 & t_1 & t_1^2 \end{pmatrix} D \begin{pmatrix} t_2 \\ t_2^2 \\ t_2^2 \end{pmatrix} + \sigma^2 \delta_{\{t_1, t_2\}} = d_{33} t_1^2 t_2^2 + d_{23}(t_1^2 t_2 + t_1 t_2^2) + d_{22} t_1 t_2 + d_{13}(t_1^2 + t_2^2) + d_{12}(t_1 + t_2) + d_{11} + \sigma^2 \delta_{\{t_1, t_2\}}.
\]

• The implied variance function is now a four-degree polynomial over time.
5.9 Example: Bivariate Observations

- Balanced data, two measurements per subject \((n_i = 2)\), two models:

**Model 1:**
Random intercepts
+ heterogeneous errors

\[
V = \begin{pmatrix} 1 \\ 1 \end{pmatrix} (d) \begin{pmatrix} 1 & 1 \end{pmatrix} + \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} d + \sigma_1^2 & d \\ d & d + \sigma_2^2 \end{pmatrix}
\]

**Model 2:**
Uncorrelated intercepts and slopes
+ measurement error

\[
V = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} = \begin{pmatrix} d_1 + \sigma^2 & d_1 \\ d_1 & d_1 + d_2 + \sigma^2 \end{pmatrix}
\]
• Different hierarchical models can produce the same marginal model

• Hence, a good fit of the marginal model cannot be interpreted as evidence for any of the hierarchical models.

• A satisfactory treatment of the hierarchical model is only possible within a Bayesian context.
5.10 A Model for the Residual Covariance Structure

- Often, $\Sigma_i$ is taken equal to $\sigma^2 I_{n_i}$

- We then obtain conditional independence:
  Conditional on $b_i$, the elements in $Y_i$ are independent

- In the presence of no, or little, random effects, conditional independence is often unrealistic

- For example, the random intercepts model not only implies constant variance, it also implicitly assumes constant correlation between any two measurements within subjects.
• Hence, when there is no evidence for (additional) random effects, or if they would have no substantive meaning, the correlation structure in the data can be accounted for in an appropriate model for $\Sigma_i$

• Frequently used model:

$$Y_i = X_i\beta + Z_i b_i + \varepsilon_{(1)i} + \varepsilon_{(2)i}$$

$$\varepsilon_i$$

• 3 stochastic components:
  - $b_i$: between-subject variability
  - $\varepsilon_{(1)i}$: measurement error
  - $\varepsilon_{(2)i}$: serial correlation component
• \( \varepsilon_{(2)i} \) represents the belief that part of an individual’s observed profile is a response to time-varying stochastic processes operating within that individual.

• This results in a correlation between serial measurements, which is usually a decreasing function of the time separation between these measurements.

• The correlation matrix \( H_i \) of \( \varepsilon_{(2)i} \) is assumed to have \((j, k)\) element of the form

\[
h_{ijk} = g(|t_{ij} - t_{ik}|)
\]

for some decreasing function \( g(\cdot) \) with \( g(0) = 1 \)

• Frequently used functions \( g(\cdot) \):
  
  ▶ Exponential serial correlation: \( g(u) = \exp(-\phi u) \)
  
  ▶ Gaussian serial correlation: \( g(u) = \exp(-\phi u^2) \)
• Graphically, for $\phi = 1$:

$$g(u) = \exp(-u)$$

*Exponential: $g(u) = \exp(-u)$*

$$g(u) = \exp(-u^2)$$

*Gaussian: $g(u) = \exp(-u^2)$*

• Extreme cases:

  $\triangleright\phi = +\infty$: components in $\varepsilon_{(2)i}$ independent

  $\triangleright\phi = 0$: components in $\varepsilon_{(2)i}$ perfectly correlated
• In general, the smaller $\phi$, the stronger is the serial correlation.

• Resulting final linear mixed model:

\[
Y_i = X_i \beta + Z_i b_i + \varepsilon_{(1)i} + \varepsilon_{(2)i}
\]

\[
\begin{align*}
    b_i &\sim N(0, D) \\
    \varepsilon_{(1)i} &\sim N(0, \sigma^2 I_{n_i}) \\
    \varepsilon_{(2)i} &\sim N(0, \tau^2 H_i)
\end{align*}
\]

\{ independent \}
- Graphical representation of all 4 components in the model:
Chapter 6
Exploratory Data Analysis

▷ Introduction
▷ Mean structure
▷ Variance function
▷ Correlation structure
▷ Individual profiles
6.1 Introduction

• A linear mixed model makes assumptions about:
  ▶ mean structure: (non-)linear, covariates, . . .
  ▶ variance function: constant, quadratic, . . .
  ▶ correlation structure: constant, serial, . . .
  ▶ subject-specific profiles: linear, quadratic, . . .

• In practice, linear mixed models are often obtained from a two-stage model formulation

• However, this may or may not imply a valid marginal model
• As an example, reconsider the growth curves:
• The individual profiles support a random-intercepts model

• However, the estimated covariance matrix suggests non-constant variance function:

\[
\begin{pmatrix}
6.11 & 6.88 & 8.26 & 7.44 & 7.18 \\
6.88 & 8.53 & 9.78 & 9.01 & 8.70 \\
7.44 & 9.01 & 10.99 & 10.42 & 10.56 \\
7.18 & 8.70 & 10.96 & 10.56 & 11.24
\end{pmatrix}.
\]

• Data exploration is therefore extremely helpful as additional tool in the selection of appropriate models
6.2 Exploring the Mean Structure

- For balanced data, averages can be calculated for each occasion separately, and standard errors for the means can be added.

- Example: rat data:

  ▶ SAS program:

  ```sas
  filename fig1 'd:\path\file.eps';
goptions reset=all ftext=swiss device=psepsf gsfname=fig1 gsfmode=replace
   rotate=landscape;
proc gplot data=test;
plot y*age / haxis=axis1 vaxis=axis2;
symbol c=red  i=std1mjt w=2 mode=include;
axis1 label=(h=2 'Age (days)') value=(h=1.5) order=(40 to 120 by 10) minor=none;
axis2 label=(h=2 A=90 'Response (pixels)') value=(h=1.5) order=(70 to 85 by 5)
   minor=none;
title h=3 'Average evolution, with standard errors of means';
run;quit;
```
SAS output:

Average evolution, with standard errors of means

Conclusion: non-linear average trend, increasing standard errors due to dropout
For unbalanced data:

- Discretize the time scale and use simple averaging within intervals
- Smoothing techniques to estimate the average evolution nonparametrically

Example: prostate data:

- SAS program for loess smoothing:

```sas
proc loess data=test;
ods output scorerresults=out;
model lnpsa=time;
score data=test;
run;

proc sort data=out;
   by time;
run;

filename fig1 'd:\path\file.eps';
goptions reset=all ftext=swiss device=psepsf
   gsfname=fig1 gsfmode=replace rotate=landscape;
proc gplot data=out;
   plot lnpsa*time=1 p_lnpsa*time=2
      / overlay haxis=axis1 vaxis=axis2;
   symbol1 c=red v=dot h=0.2 mode=include;
   symbol2 c=black i=join w=2 mode=include;
   axis1 label=(h=2 'Years before diagnosis')
      value=(h=1.5) order=(0 to 30 by 5) minor=none;
   axis2 label=(h=2 A=90 'ln(PSA+1)') value=(h=1.5)
      order=(0 to 4 by 1) minor=none;
   title h=3 'Loess smoothing';
run;quit;
```
SAS output:

Loess smoothing

Years before diagnosis

ln(PSA+1)
• If (important) covariates or factors are known, similar plots can be constructed for subgroups with different values for these covariates or factors.

• Example for the rat data:
• Example for the prostate data:
6.3 Exploring the Variance Function

- The variance function equals

\[ \sigma^2(t) = E[Y(t) - \mu(t)]^2 \]

- Hence, an estimate for \( \sigma^2(t) \) can be obtained from applying any of the techniques described for exploring the mean structure to squared residuals \( r_{ij}^2 \).
• Example for the rat data (averages with standard deviations):

![Graph showing average evolution with standard deviations](image-url)
Example for the prostate data (based on group-specific smoothing of averages):
6.4 Exploring the Correlation Structure

6.4.1 Scatterplot and Correlation Matrix

- For balanced longitudinal data, the correlation structure can be studied through the correlation matrix, or a scatterplot matrix

\[
\begin{pmatrix}
1.00 & 0.63 & 0.71 & 0.60 \\
0.63 & 1.00 & 0.63 & 0.76 \\
0.71 & 0.63 & 1.00 & 0.80 \\
0.60 & 0.76 & 0.80 & 1.00
\end{pmatrix}
\]

- Correlation matrix for the growth data:

- Graphically, pairwise scatterplots can be used for exploring the correlation between any two repeated measurements
• Scatterplot matrix for the growth data:
6.4.2 Semi-variogram

- For unbalanced data, the same approach can be used, after discretizing the time scale.

- An alternative method, in case the variance function suggests constant variance is the semi-variogram

- Re-consider the general linear mixed model:

\[
Y_i = X_i \beta + Z_i b_i + \varepsilon_{(1)i} + \varepsilon_{(2)i}
\]

\[
\begin{align*}
 b_i &\sim N(0, D) \\
 \varepsilon_{(1)i} &\sim N(0, \sigma^2 I_{n_i}) \\
 \varepsilon_{(2)i} &\sim N(0, \tau^2 H_i)
\end{align*}
\]

\text{independent}
Based on a mean function exploration, residuals $r_{ij} = y_{ij} - \mu(t_{ij})$ can be obtained.

These residuals are assumed to follow the model:

\[ r_i = Z_i b_i + \epsilon_{(1)i} + \epsilon_{(2)i} \]

The semi-variogram assumes constant variance, which implies that the only random effects in the model will at most be intercepts, i.e., $Z_i = \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix}'$.

We will denote the variance of the random intercepts by $\nu^2$.

The covariance matrix is then of the form

\[ V_i = \text{Var}(Y_i) = \text{Var}(r_i) = \nu^2 Z_i Z_i' + \sigma^2 I_{n_i} + \tau^2 H_i \]

The residuals $r_{ij}$ have constant variance $\nu^2 + \sigma^2 + \tau^2$.
• The correlation between any two residuals \( r_{ij} \) and \( r_{ik} \) from the same subject \( i \) is given by

\[
\rho(|t_{ij} - t_{ik}|) = \frac{\nu^2 + \tau^2}{\nu^2 + \sigma^2 + \tau^2} g(|t_{ij} - t_{ik}|)
\]

• One can show that, for \( j \neq k \),

\[
\frac{1}{2} E (r_{ij} - r_{ik})^2 = \sigma^2 + \tau^2 (1 - g(|t_{ij} - t_{ik}|)) = v(u_{ijk})
\]

• The function \( v(u) \) is called the semi-variogram, and it only depends on the time points \( t_{ij} \) through the time lags \( u_{ijk} = |t_{ij} - t_{ik}| \).

• Decreasing serial correlation functions \( g(\cdot) \) yield increasing semi-variograms \( v(u) \), with \( v(0) = \sigma^2 \), which converge to \( \sigma^2 + \tau^2 \) as \( u \) grows to infinity.
- Semi-variograms for exponential and Gaussian serial correlation functions $g(\cdot)$, $\sigma^2 = 0.7$, $\tau^2 = 1.3$, and $\nu^2 = 1$, $\phi = 1$:
• Obviously, an estimate of $v(u)$ can be used to explore the relative importance of the stochastic components $b_i$, $\varepsilon_{(1)i}$, and $\varepsilon_{(2)i}$, as well as the nature of the serial correlation function $g(\cdot)$.

• An estimate of $v(u)$ is obtained from smoothing the scatter plot of the
$\sum_{i=1}^{N} n_i (n_i - 1)/2$ half-squared differences $v_{ijk} = (r_{ij} - r_{ik})^2/2$ between pairs of residuals within subjects versus the corresponding time lags $u_{ijk} = |t_{ij} - t_{ik}|$.

• One can also show that, for $i \neq k$:
$\frac{1}{2}E[r_{ij} - r_{kl}]^2 = \sigma^2 + \tau^2 + \nu^2$

• Hence, the total variability in the data (assumed constant) can be estimated by

$$\hat{\sigma}^2 + \hat{\tau}^2 + \hat{\nu}^2 = \frac{1}{2N^*} \sum_{i \neq k} \sum_{j=1}^{n_i} \sum_{l=1}^{n_l} (r_{ij} - r_{kl})^2,$$

where $N^*$ is the number of terms in the sum.
• Example: prostate data

▶ We now consider the control group only:

![Graph showing control data with annotations]

▶ Assuming constant variability, the variogram can be constructed to explore the 3 stochastic components.
SAS program for loess smoothing:

/* Calculation of residuals, linear average trend */
proc glm data=prostate;
  model lnp = time;
  output out=out r=residual;
run;

/* Calculation of the variogram */
proc variogram data=out outpair=out;
  coordinates xc=time yc=id;
  compute robust novariogram;
  var residual;
run;

data variogram;set out;
  if y1=y2; vario=(v1-v2)**2/2; run;
data variance;set out;
  if y1<y2; vario=(v1-v2)**2/2; run;

/* Calculation of the total variance (=0.148) */
proc means data=variance mean;
  var vario;
run;
/* Loess smoothing of the variogram */
proc loess data=variogram;
ods output scoreresults=out;
model vario=distance;
score data=variogram;
run;

proc sort data=out;by distance;run;

filename fig1 'd:\path\file.eps';
goptions reset=all ftext=swiss device=psepsf gsfname=fig1
gsfmode=replace rotate=landscape;
proc gplot data=out;
plot vario*distance=1 p_vario*distance=2
   / overlay haxis=axis1 vaxis=axis2 vref=0.148 lvref=3;
symbol1 c=red v=dot h=0.2 mode=include;
symbol2 c=black i=join w=2 mode=include;
axis1 label=(h=2 'Time lag') value=(h=1.5)
   order=(0 to 20 by 5) minor=none;
axis2 label=(h=2 A=90 'v(u)') value=(h=1.5)
   order=(0 to 0.4 by 0.1) minor=none;
title h=3 'Semi-variogram';
run;quit;
The total variability is estimated to be 0.148.

Random intercepts represent most of the variability, while there is very little evidence for the presence of serial correlation.
6.5 Exploring the Individual Profiles

6.5.1 Introduction

• As discussed before, linear mixed models are often obtained from a two-stage model formulation

• This is based on a good approximation of the subject-specific profiles by linear regression models

• This requires methods for the exploration of longitudinal profiles
6.5.2 Graphical Exploration

- An natural way to explore longitudinal profiles is by plotting them

- Example: Prostate data:
  ▶ SAS program:

```sas
proc sort data=prostate;
by id time;
run;

filename fig1 'd:\path\file.eps';
goptions reset=all ftext=swiss device=psepsf gsfname=fig1
gsfmode=replace rotate=landscape i=join;
proc gplot data=test;plot lnpsa*time=id / haxis=axis1 vaxis=axis2 nolegend;
axis1 label=(h=2 'Years before diagnosis') value=(h=1.5)
    order=(0 to 30 by 5) minor=none;
axis2 label=(h=2 A=90 'ln(PSA+1)') value=(h=1.5)
    order=(0 to 4 by 1) minor=none;
title h=3 'Individual profiles';
run;quit;
```
• In case of large data sets:
  ▶ Randomly select some profiles
  ▶ Order subjects according to a specific profile characteristic (mean, variability, …) and plot profiles for some profiles
6.5.3 Exploring Subject-specific Regression Model

- Some ad hoc statistical procedures for checking the linear regression models
  \[ Y_i = Z_i \beta_i + \epsilon_i \]
  used in the first stage of the model formulation.

- Extensions of classical linear regression techniques:
  ▶ Coefficient $R^2$ of multiple determination
  ▶ Formal test for the need of a model extension
Coefficients of Multiple Determination

- In linear regression: \( R^2 = \frac{\text{SSTO} - \text{SSE}}{\text{SSTO}} \)

- Subject-specific coefficients: \( R^2_i = \frac{\text{SSTO}_i - \text{SSE}_i}{\text{SSTO}_i} \)

- Histogram of \( R^2_i \) or scatterplot of \( R^2_i \) versus \( n_i \)

- Overall \( R^2 \):

\[
R^2_{\text{meta}} = \frac{\sum_{i=1}^{N} (\text{SSTO}_i - \text{SSE}_i)}{\sum_{i=1}^{N} \text{SSTO}_i},
\]

- SAS macro available
Test for Model Extension

• Test for the need to extend the linear regression model $Y = X\beta + \varepsilon$ with additional covariates in $X^*$:

$$F = \frac{(SSE(R) - SSE(F))/p^*}{SSE(F)/(N - p - p^*)}$$

• Overall test for the need to extend the stage 1 model:

$$F_{\text{meta}} = \frac{\sum_{\{i: n_i \geq p + p^*\}} (SSE_i(R) - SSE_i(F)) \bigg/ \sum_{\{i: n_i \geq p + p^*\}} p^*}{\sum_{\{i: n_i \geq p + p^*\}} SSE_i(F) \bigg/ \sum_{\{i: n_i \geq p + p^*\}} (n_i - p - p^*)}$$

• Null-distribution is $F$ with $\sum_{\{i: n_i \geq p + p^*\}} p^*$ and $\sum_{\{i: n_i \geq p + p^*\}} (n_i - p - p^*)$ degrees of freedom

• SAS macro available
Example: Prostate Data

- Scatterplots of $R_i^2$ under linear and quadratic model:

![Linear model](image1)

![Quadratic model](image2)

Introduction to Longitudinal Data Analysis
• Linear model:
  \[ R^2_{\text{meta}} = 0.8188 \]
  \[ F\text{-test linear vs. quadratic: } F_{54,301} = 6.2181 \ (p < 0.0001) \]

• Quadratic model:
  \[ R^2_{\text{meta}} = 0.9143 \]
  \[ F\text{-test quadratic vs. cubic: } F_{54,247} = 1.2310 \ (p = 0.1484) \]
Chapter 7
Estimation of the Marginal Model

- Introduction
- Maximum likelihood estimation
- Restricted maximum likelihood estimation
- Fitting linear mixed models in SAS
- Negative variance components
7.1 Introduction

- Recall that the general linear mixed model equals

\[ Y_i = X_i \beta + Z_i b_i + \varepsilon_i \]

\[ b_i \sim N(0, D) \]
\[ \varepsilon_i \sim N(0, \Sigma_i) \quad \text{independent} \]

- The implied marginal model equals \( Y_i \sim N(X_i \beta, Z_i D Z_i' + \Sigma_i) \)

- Note that inferences based on the marginal model do not explicitly assume the presence of random effects representing the natural heterogeneity between subjects
- Notation:
  - $\beta$: vector of fixed effects (as before)
  - $\alpha$: vector of all variance components in $D$ and $\Sigma_i$
  - $\theta = (\beta', \alpha')'$: vector of all parameters in marginal model

- Marginal likelihood function:
  \[
  L_{ML}(\theta) = \prod_{i=1}^{N} \left\{ (2\pi)^{-n_i/2} \left| V_i(\alpha) \right|^{-\frac{1}{2}} \exp \left( -\frac{1}{2} (Y_i - X_i\beta)' V_i^{-1}(\alpha) (Y_i - X_i\beta) \right) \right\}
  \]

- If $\alpha$ were known, MLE of $\beta$ equals
  \[
  \hat{\beta}(\alpha) = \left( \sum_{i=1}^{N} X_i'W_iX_i \right)^{-1} \sum_{i=1}^{N} X_i'W_iy_i,
  \]
  where $W_i$ equals $V_i^{-1}$.  

---

Introduction to Longitudinal Data Analysis
• In most cases, $\alpha$ is not known, and needs to be replaced by an estimate $\hat{\alpha}$

• Two frequently used estimation methods for $\alpha$:
  ▶ Maximum likelihood
  ▶ Restricted maximum likelihood
7.2 Maximum Likelihood Estimation (ML)

- \( \hat{\alpha}_{\text{ML}} \) obtained from maximizing
  \[ L_{\text{ML}}(\alpha, \hat{\beta}(\alpha)) \]
  with respect to \( \alpha \)

- The resulting estimate \( \hat{\beta}(\hat{\alpha}_{\text{ML}}) \) for \( \beta \) will be denoted by \( \hat{\beta}_{\text{ML}} \)

- \( \hat{\alpha}_{\text{ML}} \) and \( \hat{\beta}_{\text{ML}} \) can also be obtained from maximizing \( L_{\text{ML}}(\theta) \) with respect to \( \theta \), i.e., with respect to \( \alpha \) and \( \beta \) simultaneously.
7.3 Restricted Maximum Likelihood Estimation (REML)

7.3.1 Variance Estimation in Normal Populations

- Consider a sample of \( N \) observations \( Y_1, \ldots, Y_N \) from \( N(\mu, \sigma^2) \)

- For known \( \mu \), MLE of \( \sigma^2 \) equals: 
  \[
  \hat{\sigma}^2 = \frac{\sum_i (Y_i - \mu)^2}{N}
  \]

- \( \hat{\sigma}^2 \) is unbiased for \( \sigma^2 \)

- When \( \mu \) is not known, MLE of \( \sigma^2 \) equals: 
  \[
  \tilde{\sigma}^2 = \frac{\sum_i (Y_i - \bar{Y})^2}{N}
  \]

- Note that \( \tilde{\sigma}^2 \) is biased for \( \sigma^2 \): 
  \[
  \mathbb{E}(\tilde{\sigma}^2) = \frac{N - 1}{N} \sigma^2
  \]
• The bias expression tells us how to derive an unbiased estimate:

\[ S^2 = \frac{\sum_i (Y_i - \bar{Y})^2}{(N - 1)} \]

• Apparently, having to estimate \( \mu \) introduces bias in MLE of \( \sigma^2 \)

• How to estimate \( \sigma^2 \), without estimating \( \mu \) first?

• The model for all data simultaneously:

\[
\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_N \end{pmatrix} \sim N \left( \begin{pmatrix} \mu \\ \vdots \\ \mu \end{pmatrix}, \sigma^2 I_N \right)
\]
• We transform $\mathbf{Y}$ such that $\mu$ vanishes from the likelihood:

$$
\mathbf{U} = \begin{pmatrix}
Y_1 - Y_2 \\
Y_2 - Y_3 \\
\vdots \\
Y_{N-2} - Y_{N-1} \\
Y_{N-1} - Y_N
\end{pmatrix} = \mathbf{A}'\mathbf{Y} \sim N(\mathbf{0}, \sigma^2 \mathbf{A}'\mathbf{A})
$$

• MLE of $\sigma^2$, based on $\mathbf{U}$, equals:

$$
S^2 = \frac{1}{N - 1} \sum_i (Y_i - \bar{Y})^2
$$

• $\mathbf{A}$ defines a set of $N - 1$ linearly independent ‘error contrasts’

• $S^2$ is called the REML estimate of $\sigma^2$, and $S^2$ is independent of $\mathbf{A}$

---

Introduction to Longitudinal Data Analysis
7.3.2 Estimation of Residual Variance in Linear Regression Model

- Consider a sample of $N$ observations $Y_1, \ldots, Y_N$ from a linear regression model:

\[
Y = \begin{pmatrix}
Y_1 \\
\vdots \\
Y_N
\end{pmatrix} \sim N(X\beta, \sigma^2 I)
\]

- MLE of $\sigma^2$:

\[
\hat{\sigma}^2 = (Y - X\hat{\beta})'(Y - X\hat{\beta})/N,
\]

- Note that $\hat{\sigma}^2$ is biased for $\sigma^2$:

\[
E(\hat{\sigma}^2) = \frac{N - p}{N} \sigma^2
\]
• The bias expression tells us how to derive an unbiased estimate:

\[ \text{MSE} = (\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta})/(N - p), \]

• The MSE can also be obtained from transforming the data orthogonal to \( \mathbf{X} \):

\[ \mathbf{U} = \mathbf{A}'\mathbf{Y} \sim N(0, \sigma^2\mathbf{A}'\mathbf{A}) \]

• The MLE of \( \sigma^2 \), based on \( \mathbf{U} \), now equals the mean squared error, MSE

• The MSE is again called the REML estimate of \( \sigma^2 \)
7.3.3 REML for the Linear Mixed Model

- We first combine all models

\[ Y_i \sim N(X_i \beta, V_i) \]

into one model

\[ Y \sim N(X \beta, V) \]

in which

\[
Y = \begin{pmatrix}
Y_1 \\
\vdots \\
Y_N
\end{pmatrix},
X = \begin{pmatrix}
X_1 \\
\vdots \\
X_N
\end{pmatrix},
V(\alpha) = \begin{pmatrix}
V_1 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & V_N
\end{pmatrix}
\]

- Again, the data are transformed orthogonal to \( X \):

\[ U = A'Y \sim N(0, A'V(\alpha)A) \]
• The MLE of $\alpha$, based on $U$ is called the REML estimate, and is denoted by $\hat{\alpha}_{REML}$.

• The resulting estimate $\hat{\beta}(\hat{\alpha}_{REML})$ for $\beta$ will be denoted by $\hat{\beta}_{REML}$.

• $\hat{\alpha}_{REML}$ and $\hat{\beta}_{REML}$ can also be obtained from maximizing

\[
L_{REML}(\theta) = \left| \sum_{i=1}^{N} X_i' W_i(\alpha) X_i \right|^{-\frac{1}{2}} L_{ML}(\theta)
\]

with respect to $\theta$, i.e., with respect to $\alpha$ and $\beta$ simultaneously.

• $L_{REML}(\alpha, \hat{\beta}(\alpha))$ is the likelihood of the error contrasts $U$, and is often called the REML likelihood function.

• Note that $L_{REML}(\theta)$ is NOT the likelihood for our original data $Y$. 

---

Introduction to Longitudinal Data Analysis

163
Reconsider the model for the prostate data:

\[
\ln(\text{PSA}_{ij} + 1) = \beta_1 \text{Age}_i + \beta_2 C_i + \beta_3 B_i + \beta_4 L_i + \beta_5 M_i \\
+ (\beta_6 \text{Age}_i + \beta_7 C_i + \beta_8 B_i + \beta_9 L_i + \beta_{10} M_i) t_{ij} \\
+ (\beta_{11} \text{Age}_i + \beta_{12} C_i + \beta_{13} B_i + \beta_{14} L_i + \beta_{15} M_i) t_{ij}^2 \\
+ b_{1i} + b_{2i} t_{ij} + b_{3i} t_{ij}^2 + \varepsilon_{ij}.
\]

Factor *group* defined by:

- control : *group* = 1
- BPH : *group* = 2
- local cancer : *group* = 3
- metastatic cancer : *group* = 4
We will assume $\Sigma_i = \sigma^2 I_{n_i}$

- **time** and **timeclss** are time, expressed in decades before diagnosis

- **age** is age at the time of diagnosis

- $lnpsa = \ln(PSA + 1)$

- **SAS program:**

```sas
proc mixed data=prostate method=reml;
class id group timeclss;
model lnpsa = group age group*time age*time2 / noint solution;
random intercept time time2 / type=un subject=id g gcorr v vcorr;
repeated timeclss / type=simple subject=id r rcorr;
run;
```
• PROC MIXED statement:
  ▶ calls procedure MIXED
  ▶ specifies data-set (records correspond to occasions)
  ▶ estimation method: ML, REML (default), ...

• CLASS statement: definition of the factors in the model

• MODEL statement:
  ▶ response variable
  ▶ fixed effects
  ▶ options similar to SAS regression procedures
• **RANDOM** statement:
  ▶ definition of random effects (including intercepts !)
  ▶ identification of the ‘subjects’ : independence across subjects
  ▶ type of random-effects covariance matrix $D$
  ▶ options ‘g’ and ‘gcorr’ to print out $D$ and corresponding correlation matrix
  ▶ options ‘v’ and ‘vcorr’ to print out $V_i$ and corresponding correlation matrix

• **REPEATED** statement:
  ▶ ordering of measurements within subjects
  ▶ the effect(s) specified must be of the factor-type
  ▶ identification of the ‘subjects’ : independence across subjects
  ▶ type of residual covariance matrix $\Sigma_i$
  ▶ options ‘r’ and ‘rcorr’ to print out $\Sigma_i$ and corresponding correlation matrix
Some frequently used covariance structures available in RANDOM and REPEATED statements:

<table>
<thead>
<tr>
<th>Structure</th>
<th>Example</th>
<th>Structure</th>
<th>Example</th>
</tr>
</thead>
</table>
| Unstructured               | \[
\begin{pmatrix}
\sigma_1^2 & \sigma_{12} & \sigma_{13} \\
\sigma_{12} & \sigma_2^2 & \sigma_{23} \\
\sigma_{13} & \sigma_{23} & \sigma_3^2
\end{pmatrix}
\] | Toeplitz                   | \[
\begin{pmatrix}
\sigma_1^2 & \sigma_{12} & \sigma_{13} \\
\sigma_{12} & \sigma_2^2 & \sigma_{12} \\
\sigma_{13} & \sigma_{12} & \sigma_3^2
\end{pmatrix}
\] |
| Simple                     | \[
\begin{pmatrix}
\sigma^2 & 0 & 0 \\
0 & \sigma^2 & 0 \\
0 & 0 & \sigma^2
\end{pmatrix}
\] | Toeplitz (1)               | \[
\begin{pmatrix}
\sigma^2 & 0 & 0 \\
0 & \sigma^2 & 0 \\
0 & 0 & \sigma^2
\end{pmatrix}
\] |
| Compound symmetry          | \[
\begin{pmatrix}
\sigma_1^2 + \sigma^2 & \sigma_1^2 & \sigma_1^2 \\
\sigma_1^2 & \sigma_1^2 + \sigma^2 & \sigma_1^2 \\
\sigma_1^2 & \sigma_1^2 & \sigma_1^2 + \sigma^2
\end{pmatrix}
\] | Heterogeneous compound symmetry | \[
\begin{pmatrix}
\sigma_1^2 & \rho \sigma_1 \sigma_2 & \rho \sigma_1 \sigma_3 \\
\rho \sigma_1 \sigma_2 & \sigma_2^2 & \rho \sigma_2 \sigma_3 \\
\rho \sigma_1 \sigma_3 & \rho \sigma_2 \sigma_3 & \sigma_3^2
\end{pmatrix}
\] |
| Banded                     | \[
\begin{pmatrix}
\sigma_1^2 & \sigma_{12} & 0 \\
\sigma_{12} & \sigma_2^2 & \sigma_{23} \\
0 & \sigma_{23} & \sigma_3^2
\end{pmatrix}
\] | Heterogeneous first-order autoregressive | \[
\begin{pmatrix}
\sigma_1^2 & \rho \sigma_1 \sigma_2 & \rho^2 \sigma_1 \sigma_3 \\
\rho \sigma_1 \sigma_2 & \sigma_2^2 & \rho \sigma_2 \sigma_3 \\
\rho^2 \sigma_1 \sigma_3 & \rho \sigma_2 \sigma_3 & \sigma_3^2
\end{pmatrix}
\] |
| First-order autoregressive | \[
\begin{pmatrix}
\sigma^2 & \rho \sigma^2 & \rho^2 \sigma^2 \\
\rho \sigma^2 & \sigma^2 & \rho \sigma^2 \\
\rho^2 \sigma^2 & \rho \sigma^2 & \sigma^2
\end{pmatrix}
\] | Heterogeneous Toeplitz      | \[
\begin{pmatrix}
\sigma_1^2 & \rho_1 \sigma_1 \sigma_2 & \rho_2 \sigma_1 \sigma_3 \\
\rho_1 \sigma_1 \sigma_2 & \sigma_2^2 & \rho_1 \sigma_2 \sigma_3 \\
\rho_2 \sigma_1 \sigma_3 & \rho_1 \sigma_2 \sigma_3 & \sigma_3^2
\end{pmatrix}
\] |
When serial correlation is to be fitted, it should be specified in the \texttt{REPEATED} statement, and the option ‘local’ can then be added to also include measurement error, if required.

Some frequently used serial correlation structures available in \texttt{RANDOM} and \texttt{REPEATED} statements:

<table>
<thead>
<tr>
<th>Structure</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power</td>
<td>$\sigma^2 \begin{pmatrix} 1 &amp; \rho_{12}^{d_{12}} &amp; \rho_{13}^{d_{13}} \ \rho_{12}^{d_{12}} &amp; 1 &amp; \rho_{23}^{d_{23}} \ \rho_{13}^{d_{13}} &amp; \rho_{23}^{d_{23}} &amp; 1 \end{pmatrix}$</td>
</tr>
<tr>
<td>Exponential</td>
<td>$\sigma^2 \begin{pmatrix} 1 &amp; \exp(-d_{12}/\rho) &amp; \exp(-d_{13}/\rho) \ \exp(-d_{12}/\rho) &amp; 1 &amp; \exp(-d_{23}/\rho) \ \exp(-d_{13}/\rho) &amp; \exp(-d_{23}/\rho) &amp; 1 \end{pmatrix}$</td>
</tr>
<tr>
<td>Gaussian</td>
<td>$\sigma^2 \begin{pmatrix} 1 &amp; \exp(-d_{12}^2/\rho^2) &amp; \exp(-d_{13}^2/\rho^2) \ \exp(-d_{12}^2/\rho^2) &amp; 1 &amp; \exp(-d_{23}^2/\rho^2) \ \exp(-d_{13}^2/\rho^2) &amp; \exp(-d_{23}^2/\rho^2) &amp; 1 \end{pmatrix}$</td>
</tr>
</tbody>
</table>
• ML and REML estimates for fixed effects:

<table>
<thead>
<tr>
<th>Effect</th>
<th>Parameter</th>
<th>MLE (s.e.)</th>
<th>REMLE (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age effect</td>
<td>$\beta_1$</td>
<td>0.026 (0.013)</td>
<td>0.027 (0.014)</td>
</tr>
<tr>
<td>Intercepts</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>$\beta_2$</td>
<td>−1.077 (0.919)</td>
<td>−1.098 (0.976)</td>
</tr>
<tr>
<td>BPH</td>
<td>$\beta_3$</td>
<td>−0.493 (1.026)</td>
<td>−0.523 (1.090)</td>
</tr>
<tr>
<td>L/R cancer</td>
<td>$\beta_4$</td>
<td>0.314 (0.997)</td>
<td>0.296 (1.059)</td>
</tr>
<tr>
<td>Met. cancer</td>
<td>$\beta_5$</td>
<td>1.574 (1.022)</td>
<td>1.549 (1.086)</td>
</tr>
<tr>
<td>Age×time effect</td>
<td>$\beta_6$</td>
<td>−0.010 (0.020)</td>
<td>−0.011 (0.021)</td>
</tr>
<tr>
<td>Time effects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>$\beta_7$</td>
<td>0.511 (1.359)</td>
<td>0.568 (1.473)</td>
</tr>
<tr>
<td>BPH</td>
<td>$\beta_8$</td>
<td>0.313 (1.511)</td>
<td>0.396 (1.638)</td>
</tr>
<tr>
<td>L/R cancer</td>
<td>$\beta_9$</td>
<td>−1.072 (1.469)</td>
<td>−1.036 (1.593)</td>
</tr>
<tr>
<td>Met. cancer</td>
<td>$\beta_{10}$</td>
<td>−1.657 (1.499)</td>
<td>−1.605 (1.626)</td>
</tr>
<tr>
<td>Age×time$^2$ effect</td>
<td>$\beta_{11}$</td>
<td>0.002 (0.008)</td>
<td>0.002 (0.009)</td>
</tr>
<tr>
<td>Time$^2$ effects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>$\beta_{12}$</td>
<td>−0.106 (0.549)</td>
<td>−0.130 (0.610)</td>
</tr>
<tr>
<td>BPH</td>
<td>$\beta_{13}$</td>
<td>−0.119 (0.604)</td>
<td>−0.158 (0.672)</td>
</tr>
<tr>
<td>L/R cancer</td>
<td>$\beta_{14}$</td>
<td>0.350 (0.590)</td>
<td>0.342 (0.656)</td>
</tr>
<tr>
<td>Met. cancer</td>
<td>$\beta_{15}$</td>
<td>0.411 (0.598)</td>
<td>0.395 (0.666)</td>
</tr>
</tbody>
</table>
ML and REML estimates for variance components:

<table>
<thead>
<tr>
<th>Effect</th>
<th>Parameter</th>
<th>MLE (s.e.)</th>
<th>REMLE (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariance of ( b_i ):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{var}(b_{1i}) )</td>
<td>( d_{11} )</td>
<td>0.398 (0.083)</td>
<td>0.452 (0.098)</td>
</tr>
<tr>
<td>( \text{var}(b_{2i}) )</td>
<td>( d_{22} )</td>
<td>0.768 (0.187)</td>
<td>0.915 (0.230)</td>
</tr>
<tr>
<td>( \text{var}(b_{3i}) )</td>
<td>( d_{33} )</td>
<td>0.103 (0.032)</td>
<td>0.131 (0.041)</td>
</tr>
<tr>
<td>( \text{cov}(b_{1i}, b_{2i}) )</td>
<td>( d_{12} = d_{21} )</td>
<td>−0.443 (0.113)</td>
<td>−0.518 (0.136)</td>
</tr>
<tr>
<td>( \text{cov}(b_{2i}, b_{3i}) )</td>
<td>( d_{23} = d_{32} )</td>
<td>−0.273 (0.076)</td>
<td>−0.336 (0.095)</td>
</tr>
<tr>
<td>( \text{cov}(b_{3i}, b_{1i}) )</td>
<td>( d_{13} = d_{31} )</td>
<td>0.133 (0.043)</td>
<td>0.163 (0.053)</td>
</tr>
<tr>
<td>Residual variance:</td>
<td>( \text{var}(\varepsilon_{ij}) )</td>
<td>( \sigma^2 )</td>
<td>( \text{var}(\varepsilon_{ij}) )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.028 (0.002)</td>
<td>0.028 (0.002)</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td></td>
<td>−1.788</td>
<td>−31.235</td>
</tr>
</tbody>
</table>
- Fitted average profiles at median age at diagnosis:
7.5 Negative Variance Components

- Reconsider the model for the rat data:

\[ Y_{ij} = (\beta_0 + b_{1i}) + (\beta_1 L_i + \beta_2 H_i + \beta_3 C_i + b_{2i})t_{ij} + \varepsilon_{ij} \]

- REML estimates obtained from SAS procedure MIXED:

<table>
<thead>
<tr>
<th>Effect</th>
<th>Parameter</th>
<th>REMLE (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>( \beta_0 )</td>
<td>68.606 (0.325)</td>
</tr>
<tr>
<td>Time effects:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low dose</td>
<td>( \beta_1 )</td>
<td>7.503 (0.228)</td>
</tr>
<tr>
<td>High dose</td>
<td>( \beta_2 )</td>
<td>6.877 (0.231)</td>
</tr>
<tr>
<td>Control</td>
<td>( \beta_3 )</td>
<td>7.319 (0.285)</td>
</tr>
<tr>
<td>Covariance of ( b_i ):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{var}(b_{1i}) )</td>
<td>( d_{11} )</td>
<td>3.369 (1.123)</td>
</tr>
<tr>
<td>( \text{var}(b_{2i}) )</td>
<td>( d_{22} )</td>
<td>0.000 (0)</td>
</tr>
<tr>
<td>( \text{cov}(b_{1i}, b_{2i}) )</td>
<td>( d_{12} = d_{21} )</td>
<td>0.090 (0.381)</td>
</tr>
<tr>
<td>Residual variance:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{var}(\varepsilon_{ij}) )</td>
<td>( \sigma^2 )</td>
<td>1.445 (0.145)</td>
</tr>
<tr>
<td>REML log-likelihood</td>
<td></td>
<td>−466.173</td>
</tr>
</tbody>
</table>
• This suggests that the REML likelihood could be further increased by allowing negative estimates for $d_{22}$

• In SAS, this can be done by adding the option ‘nobound’ to the PROC MIXED statement.

### Results:

<table>
<thead>
<tr>
<th>Effect</th>
<th>Parameter</th>
<th>Parameter restrictions for $\alpha$</th>
<th>REMLE (s.e.)</th>
<th>REMLE (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>$\beta_0$</td>
<td>$d_{ii} \geq 0, \sigma^2 \geq 0$</td>
<td>68.606 (0.325)</td>
<td>68.618 (0.313)</td>
</tr>
<tr>
<td>Time effects:</td>
<td></td>
<td>$d_{ii} \in \mathbb{R}, \sigma^2 \in \mathbb{R}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low dose</td>
<td>$\beta_1$</td>
<td></td>
<td>7.503 (0.228)</td>
<td>7.475 (0.198)</td>
</tr>
<tr>
<td>High dose</td>
<td>$\beta_2$</td>
<td></td>
<td>6.877 (0.231)</td>
<td>6.890 (0.198)</td>
</tr>
<tr>
<td>Control</td>
<td>$\beta_3$</td>
<td></td>
<td>7.319 (0.285)</td>
<td>7.284 (0.254)</td>
</tr>
<tr>
<td>Covariance of $b_i$:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{var}(b_{1i})$</td>
<td>$d_{11}$</td>
<td></td>
<td>3.369 (1.123)</td>
<td>2.921 (1.019)</td>
</tr>
<tr>
<td>$\text{var}(b_{2i})$</td>
<td>$d_{22}^*$</td>
<td></td>
<td>0.000 (——)</td>
<td>−0.287 (0.169)</td>
</tr>
<tr>
<td>$\text{cov}(b_{1i}, b_{2i})$</td>
<td>$d_{12} = d_{21}$</td>
<td></td>
<td>0.090 (0.381)</td>
<td>0.462 (0.357)</td>
</tr>
<tr>
<td>Residual variance:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{var}(\varepsilon_{ij})$</td>
<td>$\sigma^2$</td>
<td></td>
<td>1.445 (0.145)</td>
<td>1.522 (0.165)</td>
</tr>
<tr>
<td>REML log-likelihood</td>
<td></td>
<td></td>
<td>−466.173</td>
<td>−465.193</td>
</tr>
</tbody>
</table>
• Note that the REML log-likelihood value has been further increased and a negative estimate for $d_{22}$ is obtained.


---

**Negative variance components**

The usual action when a negative variance component estimate is obtained for a random coefficient would be to refit the model with the random coefficient removed. However, the user should be warned that not all packages will produce a negative variance component estimate. For example, in PROC MIXED we have found that non-convergence or a message stating that the $G$ matrix is not positive semi-definite are usually indications of a negative variance component. (A matrix, $A$, is positive semi-definite if $x'Ax$ is a non-negative number for all vectors, $x$.)

The recommended action is then to remove the random coefficients one by one in decreasing order of complexity until all variance components become positive.
• Meaning of negative variance component?

▷ Fitted variance function:

\[
\text{Var}(Y_i(t)) = \begin{pmatrix} 1 & t \end{pmatrix} \hat{D} \begin{pmatrix} 1 \\ t \end{pmatrix} + \hat{\sigma}^2
\]

\[
= \hat{d}_{22} t^2 + 2\hat{d}_{12} t + \hat{d}_{11} + \hat{\sigma}^2 = -0.287 t^2 + 0.924 t + 4.443
\]

▷ The suggested negative curvature in the variance function is supported by the sample variance function:
This again shows that the hierarchical and marginal models are not equivalent:

- The marginal model allows negative variance components, as long as the marginal covariances $V_i = Z_i D Z_i' + \sigma^2 I_{n_i}$ are positive definite
- The hierarchical interpretation of the model does not allow negative variance components
Chapter 8
Inference for the Marginal Model

▷ Inference for fixed effects:
  * Wald test
  * $t$-test and $F$-test
  * Robust inference
  * LR test

▷ Inference for variance components:
  * Wald test
  * LR test

▷ Information criteria
8.1 Inference for the Fixed Effects

- Estimate for $\beta$:

$$\hat{\beta}(\alpha) = \left( \sum_{i=1}^{N} X_i'W_iX_i \right)^{-1} \sum_{i=1}^{N} X_i'W_iy_i$$

with $\alpha$ replaced by its ML or REML estimate

- Conditional on $\alpha$, $\hat{\beta}(\alpha)$ is multivariate normal with mean $\beta$ and covariance

$$\text{Var}(\hat{\beta}) = \left( \sum_{i=1}^{N} X_i'W_iX_i \right)^{-1} \left( \sum_{i=1}^{N} X_i'W_i\text{var}(Y_i)W_iX_i \right) \left( \sum_{i=1}^{N} X_i'W_iX_i \right)^{-1}$$

$$= \left( \sum_{i=1}^{N} X_i'W_iX_i \right)^{-1}$$

- In practice one again replaces $\alpha$ by its ML or REML estimate
8.1.1 Approximate Wald Test

- For any known matrix \( L \), consider testing

\[
H_0 : L\beta = 0, \quad \text{versus} \quad H_A : L\beta \neq 0
\]

- Wald test statistic:

\[
G = \hat{\beta}' L' \left[ L \left( \sum_{i=1}^{N} X_i' V_i^{-1}(\hat{\alpha}) X_i \right)^{-1} L' \right]^{-1} L\hat{\beta}
\]

- Asymptotic null distribution of \( G \) is \( \chi^2 \) with rank(\( L \)) degrees of freedom
8.1.2 Approximate $t$-test and $F$-test

- Wald test based on

$$\text{Var}(\hat{\beta}) = \left( \sum_{i=1}^{N} X_i'W_i(\alpha)X_i \right)^{-1}$$

- Variability introduced from replacing $\alpha$ by some estimate is not taken into account in Wald tests

- Therefore, Wald tests will only provide valid inferences in sufficiently large samples

- In practice, this is often resolved by replacing the $\chi^2$ distribution by an appropriate $F$-distribution (are the normal by a $t$).

- For any known matrix $L$, consider testing

$$H_0 : L\beta = 0, \quad \text{versus} \quad H_A : L\beta \neq 0$$
• $F$ test statistic:

\[
F = \frac{\tilde{\beta}'L' \left[ L \left( \sum_{i=1}^{N} X_i'V_i^{-1}(\tilde{\alpha})X_i \right)^{-1} L' \right]^{-1} L\tilde{\beta}}{\text{rank}(L)}.
\]

• Approximate null-distribution of $F$ is $F$ with numerator degrees of freedom equal to $\text{rank}(L)$

• Denominator degrees of freedom to be estimated from the data:
  ▶ Containment method
  ▶ Satterthwaite approximation
  ▶ Kenward and Roger approximation
  ▶ ...
In the context of longitudinal data, all methods typically lead to large numbers of degrees of freedom, and therefore also to very similar \( p \)-values.

For univariate hypotheses \((\text{rank}(L) = 1)\) the \( F \)-test reduces to a \( t \)-test.
8.1.3 Example: The Prostate Data

- Linear hypotheses of the form
  \[ H_0 : L\beta = 0, \quad \text{versus} \quad H_A : L\beta \neq 0 \]
  can be tested in SAS using a CONTRAST statement

- As an example, reconsider the model for the prostate data:
  \[
  \ln(\text{PSA}_{ij} + 1) = \beta_1 \text{Age}_i + \beta_2 C_i + \beta_3 B_i + \beta_4 L_i + \beta_5 M_i \\
  + (\beta_6 \text{Age}_i + \beta_7 C_i + \beta_8 B_i + \beta_9 L_i + \beta_{10} M_i) t_{ij} \\
  + (\beta_{11} \text{Age}_i + \beta_{12} C_i + \beta_{13} B_i + \beta_{14} L_i + \beta_{15} M_i) t_{ij}^2 \\
  + b_1 + b_2 t_{ij} + b_3 t_{ij}^2 + \varepsilon_{ij}.
  \]

- We now test whether the local cancer cases evolve different from the metastatic cancer cases.
• The null-hypothesis is specified by
\[
H_0 : \begin{cases}
\beta_4 = \beta_5 \\
\beta_9 = \beta_{10} \\
\beta_{14} = \beta_{15},
\end{cases}
\]

• This is equivalent with testing
\[
H_0 : \left( \begin{array}{ccccccc}
0 & 0 & 0 & 1 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & -1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array} \right) \beta = 0,
\]

which is of the form \( L\beta = 0 \)
• Related statements in SAS:

```sas
model lnpsa = group age group*time age*time
       group*time2 age*time2
/ noint ddfm=satterth;

contrast 'L/R can = Met can' group 0 0 1 -1,
       group*time 0 0 1 -1,
       group*time2 0 0 1 -1 / chisq;
```

• Remarks:

▷ The Satterthwaite approximation is used for the denominator degrees of freedom
▷ The option ‘chisq’ in CONTRAST statement is needed in order also to obtain a Wald test
• Additional table in the output:

<table>
<thead>
<tr>
<th>Source</th>
<th>NDF</th>
<th>DDF</th>
<th>ChiSq</th>
<th>F</th>
<th>Pr &gt; ChiSq</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>L/R can = Met can</td>
<td>3</td>
<td>24.4</td>
<td>17.57</td>
<td>5.86</td>
<td>0.0005</td>
<td>0.0037</td>
</tr>
</tbody>
</table>

• Several CONTRAST statements can now be used to reduce the model, in a stepwise procedure.

• This leads to the following simplifications:
  ▶ no interaction \( age \times time^2 \)
  ▶ no interaction \( age \times time \)
  ▶ quadratic time effect the same for both cancer groups
  ▶ the quadratic time effect is not significant for the non-cancer groups
  ▶ the linear time effect is not significant for the controls
• Simultaneous testing of all these hypotheses is testing the null hypothesis

\[ H_0 : \begin{cases} 
\beta_6 = 0 & \text{(no age by time interaction)} \\
\beta_7 = 0 & \text{(no linear time effect for controls)} \\
\beta_{11} = 0 & \text{(no age}\times\text{time}^2\text{interaction)} \\
\beta_{12} = 0 & \text{(no quadratic time effect for controls)} \\
\beta_{13} = 0 & \text{(no quadratic time effect for BPH)} \\
\beta_{14} = \beta_{15} & \text{(equal quadratic time effect for both cancer groups).}
\end{cases} \]

• This hypothesis is of the form

\[ H_0 : \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
\end{pmatrix} \begin{pmatrix} \beta \\ \beta \end{pmatrix} = 0 \]
The hypothesis can be tested with the following statements:

```sas
model lnpsa = group age group*time age*time group*time2 age*time2 / noint ddfm=satterth;
contrast 'Final model' age*time 1,
    group*time 1 0 0 0,
    age*time2 1,
    group*time2 1 0 0 0,
    group*time2 0 1 0 0,
    group*time2 0 0 1 -1 / chisq;
```

This results in the following table in the output (Satterthwaite approximation):

<table>
<thead>
<tr>
<th>Source</th>
<th>NDF</th>
<th>DDF</th>
<th>ChiSq</th>
<th>F</th>
<th>Pr &gt; ChiSq</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final model</td>
<td>6</td>
<td>46.7</td>
<td>3.39</td>
<td>0.56</td>
<td>0.7587</td>
<td>0.7561</td>
</tr>
</tbody>
</table>
• The simplified model is now given by:

\[
\ln(PSA_{ij} + 1) = \beta_1 \text{Age}_i + \beta_2 C_i + \beta_3 B_i + \beta_4 L_i + \beta_5 M_i \\
+ (\beta_8 B_i + \beta_9 L_i + \beta_{10} M_i) t_{ij} \\
+ \beta_{14} (L_i + M_i) t_{ij}^2 \\
+ b_{1i} + b_{2i} t_{ij} + b_{3i} t_{ij}^2 + \varepsilon_{ij},
\]

• SAS procedure MIXED also allows using an ESTIMATE statement to estimate and test linear combinations of the elements of \( \beta \)

• Using similar arguments as for the approximate Wald-test, \( t \)-test, and \( F \)-test, approximate confidence intervals can be obtained for such linear combinations, also implemented in the ESTIMATE statement.

• Specification of \( L \) remains the same as for the CONTRAST statement, but \( L \) can now only contain one row.
8.1.4 Robust Inference

- Estimate for $\beta$:
  \[
  \hat{\beta}(\alpha) = \left( \sum_{i=1}^{N} X'_i W_i X_i \right)^{-1} \sum_{i=1}^{N} X'_i W_i Y_i
  \]
  with $\alpha$ replaced by its ML or REML estimate

- Conditional on $\alpha$, $\hat{\beta}$ has mean
  \[
  E[\hat{\beta}(\alpha)] = \left( \sum_{i=1}^{N} X'_i W_i X_i \right)^{-1} \sum_{i=1}^{N} X'_i W_i E(Y_i)
  \]
  \[
  = \left( \sum_{i=1}^{N} X'_i W_i X_i \right)^{-1} \sum_{i=1}^{N} X'_i W_i X_i \beta = \beta
  \]
  provided that $E(Y_i) = X_i \beta$

- Hence, in order for $\hat{\beta}$ to be unbiased, it is sufficient that the mean of the response is correctly specified.
Conditional on $\alpha$, $\hat{\beta}$ has covariance

$$\text{Var}(\hat{\beta}) = \left( \sum_{i=1}^{N} X'_i W_i X_i \right)^{-1} \left( \sum_{i=1}^{N} X'_i W_i \text{Var}(Y_i) W_i X_i \right) \left( \sum_{i=1}^{N} X'_i W_i X_i \right)^{-1}$$

$$= \left( \sum_{i=1}^{N} X'_i W_i X_i \right)^{-1}$$

Note that this assumes that the covariance matrix $\text{Var}(Y_i)$ is correctly modelled as $V_i = Z_i D Z'_i + \Sigma_i$

This covariance estimate is therefore often called the ‘naive’ estimate.

The so-called ‘robust’ estimate for $\text{Var}(\hat{\beta})$, which does not assume the covariance matrix to be correctly specified is obtained from replacing $\text{Var}(Y_i)$ by $[Y_i - X_i\hat{\beta}] [Y_i - X_i\hat{\beta}]'$ rather than $V_i$
• The only condition for \( (Y_i - X_i\hat{\beta}) (Y_i - X_i\hat{\beta})' \) to be unbiased for \( \text{Var}(Y_i) \) is that the mean is again correctly specified.

• The so-obtained estimate is called the ‘robust’ variance estimate, also called the sandwich estimate:

\[
\text{Var}(\hat{\beta}) = \left( \sum_{i=1}^{N} X_i'W_iX_i \right)^{-1} \left( \sum_{i=1}^{N} X_i'W_i\text{Var}(Y_i)W_iX_i \right) \left( \sum_{i=1}^{N} X_i'W_iX_i \right)^{-1}
\]

\[\downarrow\]

BREAD

\[\downarrow\]

MEAT

\[\downarrow\]

BREAD

• Based on this sandwich estimate, robust versions of the Wald test as well as of the approximate \( t \)-test and \( F \)-test can be obtained.
• Note that this suggests that as long as interest is only in inferences for the mean structure, little effort should be spent in modeling the covariance structure, provided that the data set is sufficiently large.

• Extreme point of view: OLS with robust standard errors.

• Appropriate covariance modeling may still be of interest:
  ▶ for the interpretation of random variation in data
  ▶ for gaining efficiency
  ▶ in presence of missing data, robust inference only valid under very severe assumptions about the underlying missingness process (see later).
8.1.5 Example: Prostate Data

• We reconsider the reduced model for the prostate data:

\[
\ln(\text{PSA}_{ij} + 1) = \beta_1 \text{Age}_i + \beta_2 C_i + \beta_3 B_i + \beta_4 L_i + \beta_5 M_i \\
+ (\beta_8 B_i + \beta_9 L_i + \beta_{10} M_i) t_{ij} \\
+ \beta_{14} (L_i + M_i) t_{ij}^2 \\
+ b_{1i} + b_{2it_{ij}} + b_{3it_{ij}^2} + \varepsilon_{ij},
\]

• Robust inferences for the fixed effects can be obtained from adding the option ‘empirical’ to the PROC MIXED statement:

```plaintext
proc mixed data=prostate method=reml empirical;
```
• Comparison of naive and robust standard errors (only fixed effects!):

<table>
<thead>
<tr>
<th>Effect</th>
<th>Parameter</th>
<th>Estimate (s.e.(^{(1)}), s.e.(^{(2)}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age effect</td>
<td>(\beta_1)</td>
<td>0.016 (0.006;0.006)</td>
</tr>
<tr>
<td>Intercepts:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>(\beta_2)</td>
<td>−0.564 (0.428;0.404)</td>
</tr>
<tr>
<td>BPH</td>
<td>(\beta_3)</td>
<td>0.275 (0.488;0.486)</td>
</tr>
<tr>
<td>L/R cancer</td>
<td>(\beta_4)</td>
<td>1.099 (0.486;0.499)</td>
</tr>
<tr>
<td>Met. cancer</td>
<td>(\beta_5)</td>
<td>2.284 (0.531;0.507)</td>
</tr>
<tr>
<td>Time effects:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BPH</td>
<td>(\beta_8)</td>
<td>−0.410 (0.068;0.067)</td>
</tr>
<tr>
<td>L/R cancer</td>
<td>(\beta_9)</td>
<td>−1.870 (0.233;0.360)</td>
</tr>
<tr>
<td>Met. cancer</td>
<td>(\beta_{10})</td>
<td>−2.303 (0.262;0.391)</td>
</tr>
<tr>
<td>Time(^2) effects:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cancer</td>
<td>(\beta_{14} = \beta_{15})</td>
<td>0.510 (0.088;0.128)</td>
</tr>
</tbody>
</table>

s.e.\(^{(1)}\): Naive, s.e.\(^{(2)}\): Robust

• For some parameters, the robust standard error is smaller than the naive, model-based one. For other parameters, the opposite is true.
8.1.6 Example: Growth Data

- Comparison of naive and robust standard errors under Model 1 (unstructured mean as well as covariance), for the orthodontic growth data:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MLE</th>
<th>(naive s.e.)</th>
<th>(robust s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{0,8}$</td>
<td>22.8750</td>
<td>(0.5598)</td>
<td>(0.5938)</td>
</tr>
<tr>
<td>$\beta_{0,10}$</td>
<td>23.8125</td>
<td>(0.4921)</td>
<td>(0.5170)</td>
</tr>
<tr>
<td>$\beta_{0,12}$</td>
<td>25.7188</td>
<td>(0.6112)</td>
<td>(0.6419)</td>
</tr>
<tr>
<td>$\beta_{0,14}$</td>
<td>27.4688</td>
<td>(0.5371)</td>
<td>(0.5048)</td>
</tr>
<tr>
<td>$\beta_{1,8}$</td>
<td>21.1818</td>
<td>(0.6752)</td>
<td>(0.6108)</td>
</tr>
<tr>
<td>$\beta_{1,10}$</td>
<td>22.2273</td>
<td>(0.5935)</td>
<td>(0.5468)</td>
</tr>
<tr>
<td>$\beta_{1,12}$</td>
<td>23.0909</td>
<td>(0.7372)</td>
<td>(0.6797)</td>
</tr>
<tr>
<td>$\beta_{1,14}$</td>
<td>24.0909</td>
<td>(0.6478)</td>
<td>(0.7007)</td>
</tr>
</tbody>
</table>

- How could the covariance structure be improved?
• We fit a model with a separate covariance structure for each group (Model 0)

• SAS program:

```sas
proc mixed data=test method=ml ;
class idnr sex age;
model measure = age*sex / noint s;
repeated age / type=un subject=idnr r rcorr group=sex;
run;
```

• LR test for Model 1 versus Model 0 : $p = 0.0082$

• The fixed-effects estimates remain unchanged.

• The naive standard errors under Model 0 are exactly the same as the sandwich estimated standard errors under Model 1.
8.1.7 Likelihood Ratio Test

- Comparison of nested models with different mean structures, but equal covariance structure

- Null hypothesis of interest equals $H_0 : \beta \in \Theta_{\beta,0}$, for some subspace $\Theta_{\beta,0}$ of the parameter space $\Theta_{\beta}$ of the fixed effects $\beta$.

- Notation:
  - $L_{ML}$: ML likelihood function
  - $\hat{\theta}_{ML,0}$: MLE under $H_0$
  - $\hat{\theta}_{ML}$: MLE under general model
• Test statistic:

\[-2 \ln \lambda_N = -2 \ln \left[ \frac{L_{ML}(\hat{\theta}_{ML,0})}{L_{ML}(\hat{\theta}_{ML})} \right]\]

• Asymptotic null distribution: $\chi^2$ with d.f. equal to the difference in dimension of $\Theta_\beta$ and $\Theta_{\beta,0}$. 
8.1.8 Example: Prostate Data

- We reconsider the reduced model:
  \[
  \ln(\text{PSA}_{ij} + 1) = \beta_1 \text{Age}_i + \beta_2 C_i + \beta_3 B_i + \beta_4 L_i + \beta_5 M_i + (\beta_8 B_i + \beta_9 L_i + \beta_{10} M_i) t_{ij} \\
  + \beta_{14} (L_i + M_i) t^2_{ij} + b_1_i + b_2_i t_{ij} + b_3_i t^2_{ij} + \varepsilon_{ij},
  \]

- Testing for the need of age correction, i.e., \( H_0 : \beta_1 = 0 \)

- Results under ML estimation:

<table>
<thead>
<tr>
<th></th>
<th>ML estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Under ( \beta_1 \in \mathbb{R} )</td>
<td>( L_{\text{ML}} = -3.575 )</td>
</tr>
<tr>
<td>Under ( H_0 : \beta_1 = 0 )</td>
<td>( L_{\text{ML}} = -6.876 )</td>
</tr>
<tr>
<td>(-2 \ln \lambda_N)</td>
<td>6.602</td>
</tr>
<tr>
<td>degrees of freedom</td>
<td>1</td>
</tr>
<tr>
<td>( p)-value</td>
<td>0.010</td>
</tr>
</tbody>
</table>
Results under REML estimation:

<table>
<thead>
<tr>
<th></th>
<th>REML estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Under $\beta_1 \in \mathbb{R}$</td>
<td>$L_{\text{REML}} = -20.165$</td>
</tr>
<tr>
<td>Under $H_0 : \beta_1 = 0$</td>
<td>$L_{\text{REML}} = -19.003$</td>
</tr>
<tr>
<td>$-2 \ln \lambda_N$</td>
<td>$-2.324$</td>
</tr>
<tr>
<td>degrees of freedom</td>
<td>—</td>
</tr>
<tr>
<td>$p$-value</td>
<td>—</td>
</tr>
</tbody>
</table>

Negative LR test statistic!
8.1.9 LR Test for Fixed Effects Under REML

- How can the negative LR test statistic be explained?

- Under REML, the response $Y$ is transformed into error contrasts $U = A'Y$, for some matrix $A$ with $A'X = 0$.

- Afterwards, ML estimation is performed based on the error contrasts.

- The reported likelihood value, $L_{REML}(\hat{\theta})$ is the likelihood at maximum for the error contrasts $U$.

- Models with different mean structures lead to different sets of error contrasts.

- Hence, the corresponding REML likelihoods are based on different observations, which makes them no longer comparable.
**Conclusion:**

LR tests for the mean structure are not valid under REML.
8.2 Inference for the Variance Components

- Inference for the mean structure is usually of primary interest.

- However, inferences for the covariance structure is of interest as well:
  - interpretation of the random variation in the data
  - overparameterized covariance structures lead to inefficient inferences for mean
  - too restrictive models invalidate inferences for the mean structure
8.2.1 Approximate Wald Test

- Asymptotically, ML and REML estimates of $\alpha$ are normally distributed with correct mean and inverse Fisher information matrix as covariance

- Hence approximate s.e.’s and Wald tests can easily be obtained
8.2.2 Example: Prostate Data

- We reconsider the reduced model:

\[
\ln(\text{PSA}_{ij} + 1) = \beta_1 \text{Age}_i + \beta_2 C_i + \beta_3 B_i + \beta_4 L_i + \beta_5 M_i + (\beta_8 B_i + \beta_9 L_i + \beta_{10} M_i) t_{ij} \\
+ \beta_{14} (L_i + M_i) t_{ij}^2 + b_{1i} + b_{2i} t_{ij} + b_{3i} t_{ij}^2 + \varepsilon_{ij},
\]

- Standard errors and approximate Wald tests for variance components can be obtained in PROC MIXED from adding the option ‘covtest’ to the PROC MIXED statement:

\[
\text{proc mixed data=prostate method=reml covtest;}
\]
- **Related output:**

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Subject</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Z</th>
<th>Pr Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>UN(1,1)</td>
<td>XRAY</td>
<td>0.4432</td>
<td>0.09349</td>
<td>4.74</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>UN(2,1)</td>
<td>XRAY</td>
<td>-0.4903</td>
<td>0.1239</td>
<td>-3.96</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>UN(2,2)</td>
<td>XRAY</td>
<td>0.8416</td>
<td>0.2033</td>
<td>4.14</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>UN(3,1)</td>
<td>XRAY</td>
<td>0.1480</td>
<td>0.04702</td>
<td>3.15</td>
<td>0.0017</td>
</tr>
<tr>
<td>UN(3,2)</td>
<td>XRAY</td>
<td>-0.3000</td>
<td>0.08195</td>
<td>-3.66</td>
<td>0.0003</td>
</tr>
<tr>
<td>UN(3,3)</td>
<td>XRAY</td>
<td>0.1142</td>
<td>0.03454</td>
<td>3.31</td>
<td>0.0005</td>
</tr>
<tr>
<td>timeclss</td>
<td>XRAY</td>
<td>0.02837</td>
<td>0.002276</td>
<td>12.47</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

- The reported *p*-values often do not test meaningful hypotheses
- The reported *p*-values are often wrong
8.2.3 Caution with Wald Tests for Variance Components

Marginal versus Hierarchical Model

• One of the Wald tests for the variance components in the reduced model for the prostate data was

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Subject</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Z</th>
<th>Pr Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>UN(3,3)</td>
<td>XRAY</td>
<td>0.1142</td>
<td>0.03454</td>
<td>3.31</td>
<td>0.0005</td>
</tr>
</tbody>
</table>

• This presents a Wald test for $H_0 : d_{33} = 0$

• However, under the hierarchical model interpretation, this null-hypothesis is not of any interest, as $d_{23}$ and $d_{13}$ should also equal zero whenever $d_{33} = 0$.

• Hence, the test is meaningful under the marginal model only, i.e., when no underlying random effects structure is believed to describe the data.
**Boundary Problems**

- The quality of the normal approximation for the ML or REML estimates strongly depends on the true value $\alpha$.

- Poor normal approximation if $\alpha$ is relatively close to the boundary of the parameter space.

- If $\alpha$ is a boundary value, the normal approximation completely fails.

- One of the Wald tests for the variance components in the reduced model for the prostate data was:

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Subject</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Z Value</th>
<th>Pr Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>UN(3,3)</td>
<td>XRAY</td>
<td>0.1142</td>
<td>0.03454</td>
<td>3.31</td>
<td>0.0005</td>
</tr>
</tbody>
</table>

- This presents a Wald test for $H_0 : d_{33} = 0$. 
• Under the hierarchical model interpretation, \( d_{33} = 0 \) is a boundary value, implying the calculation of the above \( p \)-value is based on an incorrect null-distribution for the Wald test statistic.

• Indeed, how could ever, under \( H_0 \), \( \hat{d}_{33} \) be normally distributed with mean 0, if \( d_{33} \) is estimated under the restriction \( d_{33} \geq 0 \)?

• Hence, the test is only correct, when the null-hypothesis is not a boundary value (e.g., \( H_0 : d_{33} = 0.1 \)).

• Note that, even under the hierarchical model interpretation, a classical Wald test is valid for testing \( H_0 : d_{23} = 0 \).
8.2.4 Likelihood Ratio Test

- Comparison of nested models with equal mean structures, but different covariance structure

- Null hypothesis of interest equals $H_0: \alpha \in \Theta_{\alpha,0}$, for some subspace $\Theta_{\alpha,0}$ of the parameter space $\Theta_{\alpha}$ of the variance components $\alpha$.

- Notation:
  - $L_{ML}$: ML likelihood function
  - $\hat{\theta}_{ML,0}$: MLE under $H_0$
  - $\hat{\theta}_{ML}$: MLE under general model

- Test statistic: $-2 \ln \lambda_N = -2 \ln \left[ \frac{L_{ML}(\hat{\theta}_{ML,0})}{L_{ML}(\hat{\theta}_{ML})} \right]$
Asymptotic null distribution: $\chi^2$ with d.f. equal to the difference in dimension of $\Theta_\alpha$ and $\Theta_{\alpha,0}$.

Note that, as long as models are compared with the same mean structure, a valid LR test can be obtained under REML as well.

Indeed, both models can be fitted using the same error contrasts, making the likelihoods comparable.

Note that, if $H_0$ is a boundary value, the classical $\chi^2$ approximation may not be valid.

For some very specific null-hypotheses on the boundary, the correct asymptotic null-distribution has been derived.
8.2.5 Marginal Testing for the Need of Random Effects

- Under a hierarchical model interpretation, the asymptotic null-distribution for the LR test statistic for testing significance of all variance components related to one or multiple random effects, can be derived.

- Example: for the prostate model, testing whether the variance components associated to the quadratic random time effect are equal to zero, is equivalent to testing

\[ H_0 : d_{13} = d_{23} = d_{33} = 0 \]

- Note that, under the hierarchical interpretation of the model, \( H_0 \) is on the boundary of the parameter space
Case 1: No Random Effects versus one Random Effect

- Hypothesis of interest:

\[ H_0 : D = 0 \quad \text{versus} \quad H_A : D = d_{11} \]

for some non-negative scalar \( d_{11} \)

- Asymptotic null-distribution equals \(-2 \ln \lambda_N \overset{\text{d}}{\longrightarrow} \chi^2_{0:1}\), the mixture of \( \chi^2_0 \) and \( \chi^2_1 \) with equal weights 0.5:
• Under $H_0$, $-2 \ln \lambda_N$ equals 0 in 50% of the cases

• Intuitive explanation:
  ▶ consider the extended parameter space $\mathbb{IR}$ for $d_{11}$
  ▶ under $H_0$, $\hat{d}_{11}$ will be negative in 50% of the cases
  ▶ under the restriction $d_{11} \geq 0$, these cases lead to $\hat{d}_{11} = 0$
  ▶ hence, $L_{ML}(\hat{\theta}_{ML,0}) = L_{ML}(\hat{\theta}_{ML})$ in 50% of the cases

• Graphically ($\tau^2 = d_{11}$):
Case 2: One versus two Random Effects

- Hypothesis of interest:
  \[ H_0 : D = \begin{pmatrix} d_{11} & 0 \\ 0 & 0 \end{pmatrix}, \]
  for \( d_{11} > 0 \), versus \( H_A \) that \( D \) is \((2 \times 2)\) positive semidefinite

- Asymptotic null-distribution:
  \[ -2 \ln \lambda_N \longrightarrow \chi^2_{1:2}, \] the mixture of \( \chi^2_1 \) and \( \chi^2_2 \) with equal weights 0.5:
Case 3: $q$ versus $q+1$ Random Effects

- Hypothesis of interest:
  \[ H_0 : D = \begin{pmatrix} D_{11} & 0 \\ 0' & 0 \end{pmatrix}, \]
  for $D_{11}$ $(q \times q)$ positive definite, versus $H_A$ that $D$ is $((q + 1) \times (q + 1))$ positive semidefinite.

- Asymptotic null-distribution: $-2 \ln \lambda_N \rightarrow \chi^2_{q,q+1}$, the mixture of $\chi^2_q$ and $\chi^2_{q+1}$ with equal weights 0.5.
Case 4: $q$ versus $q + k$ Random Effects

- Hypothesis of interest:

$$H_0 : D = \begin{pmatrix} D_{11} & 0 \\ 0 & 0 \end{pmatrix},$$

for $D_{11} (q \times q)$ positive definite, versus $H_A$ that $D$ is $((q + k) \times (q + k))$ positive semidefinite.

- Simulations needed to derive asymptotic null distribution
Conclusions

• Correcting for the boundary problem reduces \( p \)-values

• Thus, ignoring the boundary problem too often leads to over-simplified covariance structures

• Hence, ignoring the boundary problem may invalidate inferences, even for the mean structure
8.2.6 Example: Rat Data

- We reconsider the model with random intercepts and slopes for the rat data:

\[ Y_{ij} = (\beta_0 + b_{1i}) + (\beta_1 L_i + \beta_2 H_i + \beta_3 C_i + b_{2i})t_{ij} + \varepsilon_{ij} \]

in which \( t_{ij} \) equals \( \ln[1 + (\text{Age}_{ij} - 45)/10] \)

- The marginal model assumes linear average trends with common intercept for the 3 groups, and covariance structure:

\[
\text{Cov}(Y_i(t_1), Y_i(t_2)) = \begin{pmatrix} 1 & t_1 \end{pmatrix} D \begin{pmatrix} 1 \\ t_2 \end{pmatrix} + \sigma^2 \delta_{\{t_1, t_2\}} \\
= d_{22}t_1 t_2 + d_{12}(t_1 + t_2) + d_{11} + \sigma^2 \delta_{\{t_1, t_2\}}. 
\]
• Exploring the variance function yields:
  
  - This suggested earlier that the above random-effects model might not be valid, as it does not allow negative curvature in the variance function.
  
  - It is therefore of interest to test whether the random slopes $b_{2i}$ may be left out of the model.
  
  • Interpretation:
    - On hierarchical level: all rats receiving the same treatment have the same slope.
    - On marginal level: constant variance, constant correlation.
- Null-hypothesis to be tested: \( H_0 : d_{12} = d_{22} = 0 \)

- REML estimates under hierarchical and marginal interpretation, as well as under \( H_0 \):

<table>
<thead>
<tr>
<th>Effect</th>
<th>Parameter</th>
<th>Parameter restrictions for ( \alpha )</th>
<th>REMLE (s.e.)</th>
<th>REMLE (s.e.)</th>
<th>REMLE (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>( \beta_0 )</td>
<td>( d_{ii} \geq 0, \sigma^2 \geq 0 )</td>
<td>68.606 (0.325)</td>
<td>68.618 (0.313)</td>
<td>68.607 (0.331)</td>
</tr>
<tr>
<td>Time effects:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low dose</td>
<td>( \beta_1 )</td>
<td>( d_{ii} \in \mathbb{R}, \sigma^2 \in \mathbb{R} )</td>
<td>7.503 (0.228)</td>
<td>7.475 (0.198)</td>
<td>7.507 (0.225)</td>
</tr>
<tr>
<td>High dose</td>
<td>( \beta_2 )</td>
<td></td>
<td>6.877 (0.231)</td>
<td>6.890 (0.198)</td>
<td>6.871 (0.228)</td>
</tr>
<tr>
<td>Control</td>
<td>( \beta_3 )</td>
<td></td>
<td>7.319 (0.285)</td>
<td>7.284 (0.254)</td>
<td>7.507 (0.225)</td>
</tr>
<tr>
<td>Covariance of ( b_i ):</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{var}(b_{1i}) )</td>
<td>( d_{11} )</td>
<td>( d_{ii} \geq 0, \sigma^2 \geq 0 )</td>
<td>3.369 (1.123)</td>
<td>2.921 (1.019)</td>
<td>3.565 (0.808)</td>
</tr>
<tr>
<td>( \text{var}(b_{2i}) )</td>
<td>( d_{22} )</td>
<td></td>
<td>0.000 (-----)</td>
<td>-0.287 (0.169)</td>
<td>---- (-----)</td>
</tr>
<tr>
<td>( \text{cov}(b_{1i}, b_{2i}) )</td>
<td>( d_{12} = d_{21} )</td>
<td>( d_{ii} \geq 0, \sigma^2 \geq 0 )</td>
<td>0.090 (0.381)</td>
<td>0.462 (0.357)</td>
<td>---- (-----)</td>
</tr>
<tr>
<td>Residual variance:</td>
<td>( \text{var}(\varepsilon_{ij}) )</td>
<td>( d_{ii} \geq 0, \sigma^2 \geq 0 )</td>
<td>1.445 (0.145)</td>
<td>1.522 (0.165)</td>
<td>1.445 (0.145)</td>
</tr>
<tr>
<td>REML log-likelihood</td>
<td></td>
<td></td>
<td>( -466.173 )</td>
<td>( -465.193 )</td>
<td>( -466.202 )</td>
</tr>
</tbody>
</table>
Test Under Marginal Interpretation

- Unrestricted parameter space for $\alpha$, no boundary problem

- Wald test:
  
  - Test statistic:
    
    $$(\tilde{d}_{12} \quad \tilde{d}_{22}) \left( \begin{array}{cc} \text{Var}(\tilde{d}_{12}) & \text{Cov}(\tilde{d}_{12}, \tilde{d}_{22}) \\ \text{Cov}(\tilde{d}_{12}, \tilde{d}_{22}) & \text{Var}(\tilde{d}_{22}) \end{array} \right)^{-1} \left( \begin{array} c \tilde{d}_{12} \\ \tilde{d}_{22} \end{array} \right)$$

  
  $$= \left( \begin{array} {cc} 0.462 & -0.287 \end{array} \right) \left( \begin{array} {cc} 0.127 & -0.038 \\ -0.038 & 0.029 \end{array} \right)^{-1} \left( \begin{array} c 0.462 \\ -0.287 \end{array} \right) = 2.936,$$

  
  - $p$-value:
    
    $$P(\chi^2_2 \geq 2.936 \mid H_0) = 0.2304$$
• LR test:
  ▶ Test statistic:

\[-2 \ln \lambda_N = -2(-466.202 + 465.193) = 2.018\]

▶ p-value:

\[P(\chi^2_2 \geq 2.018 \mid H_0) = 0.3646\]
Test Under Hierarchical Interpretation

- Restricted parameter space for \( \alpha \) (positive semi-definite \( D \)), boundary problem.

- LR test statistic:

\[
-2 \ln \lambda_N = -2(-466.202 + 466.173) = 0.058
\]

- \( p \)-value:

\[
P(\chi^2_{1:2} \geq 0.058 \mid H_0)
= 0.5 \ P(\chi^2_1 \geq 0.058 \mid H_0) 
+ 0.5 \ P(\chi^2_2 \geq 0.058 \mid H_0) 
= 0.8906
\]

- Note that the naive \( p \)-value, obtained from ignoring the boundary problem is indeed larger:

\[
P(\chi^2_2 \geq 0.058 \mid H_0) = 0.9714
\]
Reduced Model

• Under both model interpretations, $H_0$ was accepted, leading to the reduced model:

$$Y_{ij} = (\beta_0 + b_{1i}) + (\beta_1 L_i + \beta_2 H_i + \beta_3 C_i)t_{ij} + \varepsilon_{ij}$$

• Marginal interpretation:
  ▶ linear average trends with common intercept for the 3 groups
  ▶ constant variance estimated to be
    $$\bar{d}_{11} + \sigma^2 = 3.565 + 1.445 = 5.010$$
  ▶ constant (intraclass) correlation
    $$\bar{\rho}_i = \frac{\bar{d}_{11}}{\bar{d}_{11} + \sigma^2} = 0.712$$

• The hierarchical interpretation, possible since $\bar{d}_{11} = 3.565 > 0$, is that heterogeneity between rats is restricted to differences in starting values, not slopes.
8.3 Information Criteria

8.3.1 Definition of Information Criteria

• LR tests can only be used to compare nested models

• How to compare non-nested models?

• The general idea behind the LR test for comparing model $A$ to a more extensive model $B$ is to select model $A$ if the increase in likelihood under model $B$ is small compared to increase in complexity

• A similar argument can be used to compare non-nested models $A$ and $B$
• One then selects the model with the largest (log-)likelihood provided it is not too complex.

• The model is selected with the highest penalized log-likelihood $\ell - \mathcal{F}(\#\theta)$ for some function $\mathcal{F}(\cdot)$ of the number $\#\theta$ of parameters in the model.

• Different functions $\mathcal{F}(\cdot)$ lead to different criteria:

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Definition of $\mathcal{F}(\cdot)$*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akaike (AIC)</td>
<td>$\mathcal{F}(#\theta) = #\theta$</td>
</tr>
<tr>
<td>Schwarz (SBC)</td>
<td>$\mathcal{F}(#\theta) = (#\theta \ln n^*)/2$</td>
</tr>
<tr>
<td>Hannan and Quinn (HQIC)</td>
<td>$\mathcal{F}(#\theta) = #\theta \ln(\ln n^*)$</td>
</tr>
<tr>
<td>Bozdogan (CAIC)</td>
<td>$\mathcal{F}(#\theta) = #\theta (\ln n^* + 1)/2$</td>
</tr>
</tbody>
</table>

*: $n^* = n = \sum_{i=1}^{N} n_i$ under ML

*: $n^* = n - p$ under REML
• Information criteria are no formal testing procedures!

• For the comparison of models with different mean structures, information criteria should be based on ML rather than REML, as otherwise the likelihood values would be based on different sets of error contrasts, and therefore would no longer be comparable.
8.3.2 Example: Rat Data

- Consider the random-intercepts model for the rat data:

\[ Y_{ij} = (\beta_0 + b_{1i}) + (\beta_1 L_i + \beta_2 H_i + \beta_3 C_i)t_{ij} + \varepsilon_{ij} \]

in which \( t_{ij} \) equals \( \ln[1 + (\text{Age}_{ij} - 45)/10]) \)

- We now want to compare this model with a model which assumes common average slope for the 3 treatments.

- Information criteria can be obtained in SAS from adding the option ‘ic’ to the PROC MIXED statement:

```sas
proc mixed data=rats method=ml ic;
```
• Summary of results:

<table>
<thead>
<tr>
<th>Mean structure</th>
<th>$\ell_{ML}$</th>
<th>$#\theta$</th>
<th>AIC</th>
<th>SBC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Separate average slopes</td>
<td>$-464.326$</td>
<td>$6$</td>
<td>$-470.326$</td>
<td>$-480.914$</td>
</tr>
<tr>
<td>Common average slope</td>
<td>$-466.622$</td>
<td>$4$</td>
<td>$-470.622$</td>
<td>$-477.681$</td>
</tr>
</tbody>
</table>

• Selected models:
  ▶ AIC: model with separate slopes
  ▶ SBC: model with common slopes

• Based on Wald test, the average slopes are found not to be significantly different from each other ($p = 0.0987$)
Chapter 9
Inference for the Random Effects

▷ Empirical Bayes inference
▷ Best linear unbiased prediction
▷ Example: Prostate data
▷ Shrinkage
▷ Example: Random-intercepts model
▷ Example: Prostate data
▷ Normality assumption for random effects
9.1 Empirical Bayes Inference

- Random effects $b_i$ reflect how the evolution for the $i$th subject deviates from the expected evolution $X_i\beta$.

- Estimation of the $b_i$ helpful for detecting outlying profiles

- This is only meaningful under the hierarchical model interpretation:

$$Y_i|b_i \sim N(X_i\beta + Z_i b_i, \Sigma_i) \quad b_i \sim N(0, D)$$

- Since the $b_i$ are random, it is most natural to use Bayesian methods

- Terminology: prior distribution $N(0, D)$ for $b_i$
• Posterior density:

\[ f(b_i | y_i) \equiv f(b_i | Y_i = y_i) = \frac{f(y_i | b_i) f(b_i)}{\int f(y_i | b_i) f(b_i) \, db_i} \]

\[ \propto f(y_i | b_i) f(b_i) \]

\[ \propto \ldots \]

\[ \propto \exp \left\{-\frac{1}{2} (b_i - DZ_i'W_i(y_i - X_i\beta))' \Lambda_i^{-1} (b_i - DZ_i'W_i(y_i - X_i\beta)) \right\} \]

for some positive definite matrix \( \Lambda_i \).

• Posterior distribution:

\[ b_i \mid y_i \sim N(DZ_i'W_i(y_i - X_i\beta), \Lambda_i) \]
• Posterior mean as estimate for $b_i$:

$$
\hat{b}_i(\theta) = E [b_i \mid Y_i = y_i] = \int b_i f(b_i \mid y_i) \, db_i = DZ_i'W_i(\alpha)(y_i - X_i\beta)
$$

• $\hat{b}_i(\theta)$ is normally distributed with covariance matrix

$$
\text{var}(\hat{b}_i(\theta)) = DZ_i' \left\{ W_i - W_iX_i \left( \sum_{i=1}^{N} X'_i W_i X_i \right)^{-1} X'_i W_i \right\} Z_i D
$$

• Note that inference for $b_i$ should account for the variability in $b_i$

• Therefore, inference for $b_i$ is usually based on

$$
\text{var}(\hat{b}_i(\theta) - b_i) = D - \text{var}(\hat{b}_i(\theta))
$$
• Wald tests can be derived

• Parameters in $\theta$ are replaced by their ML or REML estimates, obtained from fitting the marginal model.

• $\hat{b}_i = \hat{b}_i(\hat{\theta})$ is called the Empirical Bayes estimate of $b_i$.

• Approximate $t$- and $F$-tests to account for the variability introduced by replacing $\theta$ by $\hat{\theta}$, similar to tests for fixed effects.
9.2 Best Linear Unbiased Prediction (BLUP)

- Often, parameters of interest are linear combinations of fixed effects in $\beta$ and random effects in $b_i$.

- For example, a subject-specific slope is the sum of the average slope for subjects with the same covariate values, and the subject-specific random slope for that subject.

- In general, suppose $u = \lambda'_\beta \beta + \lambda'_b b_i$ is of interest.

- Conditionally on $\alpha$, $\hat{u} = \lambda'_\beta \hat{\beta} + \lambda'_b \hat{b}_i$ is BLUP:
  - linear in the observations $Y_i$
  - unbiased for $u$
  - minimum variance among all unbiased linear estimators.
9.3 Example: Prostate Data

- We reconsider the reduced model:

\[
\ln(PSA_{ij} + 1) = \beta_1 \text{Age}_i + \beta_2 C_i + \beta_3 B_i + \beta_4 L_i + \beta_5 M_i + (\beta_8 B_i + \beta_9 L_i + \beta_{10} M_i) t_{ij} \\
+ \beta_{14} (L_i + M_i) t_{ij}^2 + b_{1i} + b_{2i} t_{ij} + b_{3i} t_{ij}^2 + \varepsilon_{ij}
\]

- In SAS the estimates can be obtained from adding the option ‘solution’ to the random statement:

```sas
random intercept time time2
   / type=un subject=id solution;
```

```sas
ods listing exclude solutionr;
ods output solutionr=out;
```
• The ODS statements are used to write the EB estimates into a SAS output data set, and to prevent SAS from printing them in the output window.

• In practice, histograms and scatterplots of certain components of $\bar{b}_i$ are used to detect model deviations or subjects with ‘exceptional’ evolutions over time.
- Strong negative correlations in agreement with correlation matrix corresponding to fitted $\hat{D}$:

$$\hat{D}_{corr} = \begin{pmatrix}
1.000 & -0.803 & 0.658 \\
-0.803 & 1.000 & -0.968 \\
0.658 & -0.968 & 1.000
\end{pmatrix}$$

- Histograms and scatterplots show outliers

- Subjects #22, #28, #39, and #45, have highest four slopes for time$^2$ and smallest four slopes for time, i.e., with the strongest (quadratic) growth.

- Subjects #22, #28 and #39 have been further examined and have been shown to be metastatic cancer cases which were misclassified as local cancer cases.

- Subject #45 is the metastatic cancer case with the strongest growth
9.4 Shrinkage Estimators $\hat{b}_i$

- Consider the prediction of the evolution of the $i$th subject:

$$\hat{Y}_i \equiv X_i\hat{\beta} + Z_i\hat{b}_i$$

$$= X_i\hat{\beta} + Z_iDZ_i'V_i^{-1}(y_i - X_i\hat{\beta})$$

$$= (I_{n_i} - Z_iDZ_i'V_i^{-1})X_i\hat{\beta} + Z_iDZ_i'V_i^{-1}y_i$$

$$= \sum_i V_i^{-1}X_i\hat{\beta} + (I_{n_i} - \sum_i V_i^{-1})y_i,$$

- Hence, $\hat{Y}_i$ is a weighted mean of the population-averaged profile $X_i\hat{\beta}$ and the observed data $y_i$, with weights $\sum_i \hat{V}_i^{-1}$ and $I_{n_i} - \sum_i \hat{V}_i^{-1}$ respectively.
• Note that $X_i \hat{\beta}$ gets much weight if the residual variability is ‘large’ in comparison to the total variability.

• This phenomenon is usually called shrinkage:

> The observed data are shrunk towards the prior average profile $X_i \beta$.

• This is also reflected in the fact that for any linear combination $\lambda' b_i$ of random effects,

$$\text{var}(\lambda' \hat{b}_i) \leq \text{var}(\lambda' b_i).$$
9.5 Example: Random-intercepts Model

- Consider the random-intercepts model, without serial correlation:
  - $Z_i = 1_{n_i}$, vector of ones
  - $D = \sigma_b^2$, scalar
  - $\Sigma_i = \sigma^2 I_{n_i}$

- The EB estimate for the random intercept $b_i$ then equals

$$
\hat{b_i} = \sigma_b^2 1_{n_i}' \left( \sigma_b^2 1_{n_i} 1_{n_i}' + \sigma^2 I_{n_i} \right)^{-1} (y_i - X_i \beta)
$$

$$
= \frac{\sigma_b^2}{\sigma^2} 1_{n_i}' \left( I_{n_i} - \frac{\sigma_b^2}{\sigma^2 + n_i \sigma_b^2} 1_{n_i} 1_{n_i}' \right) (y_i - X_i \beta)
$$

$$
= \frac{n_i \sigma_b^2}{\sigma^2 + n_i \sigma_b^2} \left( \frac{1}{n_i} \sum_{j=1}^{n_i} (y_{ij} - X_i^{[j]} \beta) \right)
$$
Remarks:

- $\hat{b}_i$ is weighted average of 0 (prior mean) and the average residual for subject $i$
- less shrinkage the larger $n_i$
- less shrinkage the smaller $\sigma^2$ relative to $\sigma^b$
9.6 Example: Prostate Data

- Comparison of predicted, average, and observed profiles for the subjects #15 and #28, obtained under the reduced model:
- Illustration of the shrinkage effect:

\[
\tilde{\text{Var}}(\tilde{b}_i) = \begin{pmatrix}
0.403 & -0.440 & 0.131 \\
-0.440 & 0.729 & -0.253 \\
0.131 & -0.253 & 0.092
\end{pmatrix}, \quad \tilde{D} = \begin{pmatrix}
0.443 & -0.490 & 0.148 \\
-0.490 & 0.842 & -0.300 \\
0.148 & -0.300 & 0.114
\end{pmatrix}
\]
9.7 The Normality Assumption for Random Effects

- In practice, histograms of EB estimates are often used to check the normality assumption for the random effects.

- However, since

\[ \hat{b}_i = D Z'_i W_i (y_i - X_i \beta) \]

\[ \text{var}(\hat{b}_i) = D Z'_i \left\{ W_i - W_i X_i \left( \sum_{i=1}^{N} X'_i W_i X_i \right)^{-1} X'_i W_i \right\} Z_i D \]

one should at least first standardize the EB estimates.

- Further, due to the shrinkage property the EB estimates do not fully reflect the heterogeneity in the data.
• Small simulation example:
  ▶ 1000 profiles with 5 measurements, balanced
  ▶ 1000 random intercepts sampled from
    \[ \frac{1}{2} N(-2, 1) + \frac{1}{2} N(2, 1) \]
  ▶ \( \Sigma_i = \sigma^2 I_{n_i} \), \( \sigma^2 = 30 \)
  ▶ Data analysed assuming normality for the intercepts
Histogram of sampled intercepts and empirical Bayes estimates:

Clearly, severe shrinkage forces the estimates $\hat{b}_i$ to satisfy the normality assumption.
• Conclusion:

     EB estimates obtained under normality cannot be used to check normality

• This suggests that the only possibility to check the normality assumption is to fit a more general model, with the classical linear mixed model as special case, and to compare both models using LR methods.
9.8 The Heterogeneity Model

- One possible extension of the linear mixed model is to assume a finite mixture as random-effects distribution:

\[ b_i \sim \sum_{j=1}^{g} p_j N(\mu_j, D), \quad \text{with} \quad \sum_{j=1}^{g} p_j = 1 \text{ and } \sum_{j=1}^{g} p_j \mu_j = 0 \]

- Interpretation:
  - Population consists of \( g \) subpopulations
  - Each subpopulation contains fraction \( p_j \) of total population
  - In each subpopulation, a linear mixed model holds

- The classical model is a special case: \( g = 1 \)
- Very flexible class of parametric models for random-effects distribution:

\[ \mu_1 = -2, \mu_2 = 2, \sigma^2_b = 1, p = 0.5 \]

\[ \mu_1 = -2, \mu_2 = 2, \sigma^2_b = 1, p = 0.7 \]

\[ \mu_1 = -2, \mu_2 = 2, \sigma^2_b = 4, p = 0.5 \]

\[ \mu_1 = -2, \mu_2 = 2, \sigma^2_b = 4, p = 0.7 \]
• Fitting of the model is based on the EM algorithm

• SAS macro available

• EB estimates can be calculated under the heterogeneity model

• Small simulation example:
  ▶ 1000 profiles with 5 measurements, balanced
  ▶ 1000 random intercepts sampled from

  \[
  \frac{1}{2}N(-2, 1) + \frac{1}{2}N(2, 1)
  \]

  ▶ \( \Sigma_i = \sigma^2 I_i \), \( \sigma^2 = 30 \)

  ▶ Data analysed under heterogeneity model
Histogram of sampled intercepts and empirical Bayes estimates:

The correct random-effects distribution is (much) better reflected, than before under the assumption of normality.
Chapter 10
General Guidelines for Model Building

▷ Introduction
▷ General strategy
▷ Example: The prostate data
10.1 Introduction

- Marginal linear mixed model:

\[ Y_i \sim N(X_i \beta, Z_i D Z_i' + \sigma^2 I_{n_i} + \tau^2 H_i) \]

- Fitting a linear mixed model requires specification of a mean structure, as well as covariance structure

- Mean structure:
  - Covariates
  - Time effects
  - Interactions

- Covariance structure:
  - Random effects
  - Serial correlation
Both components affect each other:

- Mean structure $X_i\beta$
- Covariance structure $V_i$

- Estimation of $\theta$
- Covariance matrix for $\hat{\theta}$
- t-tests and F-tests
- Confidence intervals
- Efficiency
- Prediction
• When most variability is due to between-subject variability, the two-stage approach will often lead to acceptable marginal models.

• In the presence of a lot within-subject variability, the two-stage approach is less straightforward.

• Also, a two-stage approach may imply unrealistic marginal models.
For example, reconsider the growth curves:

- Individual profiles:

  - Short mother
  - Medium mother
  - Tall mother

- A random-intercepts model seems reasonable
However, the covariance matrix equals

\[
\begin{pmatrix}
6.11 & 6.88 & 8.26 & 7.44 & 7.18 \\
6.88 & 8.53 & 9.78 & 9.01 & 8.70 \\
7.44 & 9.01 & 10.99 & 10.42 & 10.56 \\
7.18 & 8.70 & 10.96 & 10.56 & 11.24
\end{pmatrix}.
\]

- The aim of this chapter is to discuss some general guidelines for model building.
10.2 General Strategy

\[ Y_i = X_i \beta + Z_i b_i + \varepsilon_i \]

1. Preliminary mean structure \( X_i \beta \)
2. Preliminary random-effects structure \( Z_i b_i \)
3. Residual covariance structure \( \Sigma_i \)
4. Reduction of the random-effects structure \( Z_i b_i \)
5. Reduction of the mean structure \( X_i \beta \)
10.3 Preliminary Mean Structure

10.3.1 Strategy

- Remove all systematic trends from the data, by calculating OLS residual profiles:

\[ r_i = y_i - X_i \hat{\beta}_{\text{OLS}} \approx Z_i b_i + \varepsilon_i \]

- For balanced designs with few covariates:

Saturated mean structure
• For balanced designs with many covariates, or for highly unbalanced data sets:

The most elaborate model one is prepared to consider for the mean structure

• Selection of preliminary mean structures will be based on exploratory tools for the mean.

• Note that the calculation of $\hat{\beta}_{\text{OLS}}$ ignores the longitudinal structure, and can be obtained in any regression module.

• Provided the preliminary mean structure is ‘sufficiently richt’, consistency of $\hat{\beta}_{\text{OLS}}$ follows from the theory on robust inference for the fixed effects.
10.3.2 Example: Prostate Data

- Smoothed average trend within each group:

![Diagrams showing trend in PSA levels before diagnosis for different groups]
• Quadratic function over time, within each diagnostic group

• Correction for age, via the inclusion of $age$, $age \times time$ and $age \times time^2$.

• Note that this yields the same model as the model originally obtained from a two-stage approach, containing 15 fixed effects
10.4 Preliminary Random-effects Structure

10.4.1 Strategy

\[ r_i \approx Z_i b_i + \varepsilon_i \]

- Explore the residual profiles
- Any structure left, may indicate the presence of subject-specific regression coefficients
- Try to describe each residual profile with a (relatively) simple model.
• Do not include covariates in $Z_i$ which are not included in $X_i$. Otherwise, it is not justified to assume $E(b_i) = 0$.

• Use ‘well-formulated’ models: Do not include higher-order terms unless all lower-order terms are included as well.

• Compare implied variance and covariance functions with results from exploratory tools for covariance structure
10.4.2 Example: Prostate Data

• OLS residual profiles and smoothed average of squared OLS residuals:

• We assume a quadratic function for each residual profile

• This results in a model with random intercepts, and random slopes for the linear as well as quadratic time effect.
• Variance function:

\[
\begin{pmatrix} 1 & t & t^2 \end{pmatrix} D \begin{pmatrix} 1 \\ t \\ t^2 \end{pmatrix} + \sigma^2
\]

• Comparison of smoothed average of squared OLS residuals and fitted variance function:
Possible explanation for observed differences:

- Small $t$: some subjects have extremely large responses close to diagnosis. This may have inflated the fitted variance

- Large $t$: few observations available: only 24 out of 463 measurements taken earlier than 20 years prior to diagnosis.
10.5 Residual Covariance Structure

10.5.1 Strategy

\[ r_i \approx Z_i b_i + \varepsilon_i \]

- Which covariance matrix \( \Sigma_i \) for \( \varepsilon_i \)?

- In many applications, random effects explain most of the variability.

- Therefore, in the presence of random effects other than intercepts, often \( \Sigma_i = \sigma^2 I_{n_i} \) is assumed.

- However, many other covariance structures can be specified as well.
A special class of parametric models for $\Sigma_i$ is obtained from splitting $\varepsilon_i$ into a measurement error component $\varepsilon_{(1)i}$ and a serial correlation component $\varepsilon_{(2)i}$:

$$Y_i = X_i \beta + Z_i b_i + \varepsilon_{(1)i} + \varepsilon_{(2)i}$$

$$b_i \sim N(0, D)$$

$$\varepsilon_{(1)i} \sim N(0, \sigma^2 I_{n_i})$$

$$\varepsilon_{(2)i} \sim N(0, \tau^2 H_i)$$

Only the correlation matrix $H_i$ then still needs to be specified

$H_i$ is assumed to have $(j, k)$ element of the form $h_{ijk} = g(|t_{ij} - t_{ik}|)$ for some decreasing function $g(\cdot)$ with $g(0) = 1$
• Frequently used functions $g(\cdot)$:
  ▶ Exponential serial correlation: $g(u) = \exp(-\phi u)$
  ▶ Gaussian serial correlation: $g(u) = \exp(-\phi u^2)$

• Graphical representation ($\phi = 1$):

Exponential: $g(u) = \exp(-u)$

Gaussian: $g(u) = \exp(-u^2)$
• When only random intercepts are included, the semi-variogram can be used to explore the presence and the nature of serial correlation.

• When other random effects are present as well, an extension of the variogram is needed.

• Also, a variety of serial correlation functions can be fitted and compared.
10.5.2 Example: Prostate Data

- Based on the preliminary mean and random-effects structures, several serial correlation functions can be fitted.

- For example, a model with Gaussian serial correlation can be fitted in SAS using the following program:

```sas
proc mixed data=prostate method=reml;
class id group timeclss;
model lnpsa = group age group*time age*time group*time2 age*time2 / noint solution;
random intercept time time2 / type=un subject=id g gcorr v vcorr;
repeated timeclss / type=sp(gau)(time) local subject=id r rcorr;
run;
```

- **REPEATED statement:**
  - the serial correlation model is specified in the ‘type’ option
  - ‘local’ is added to include measurement error
• Summary of model fits:

<table>
<thead>
<tr>
<th>Residual covariance structure</th>
<th>REML log-likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measurement error</td>
<td>−31.235</td>
</tr>
<tr>
<td>Measurement error + Gaussian</td>
<td>−24.787</td>
</tr>
<tr>
<td>Measurement error + exponential</td>
<td>−24.266</td>
</tr>
</tbody>
</table>

• The presence of serial correlation is clearly detected

• However, there seems to be little information in the data to distinguish between different serial correlation structures

• Practical experience suggests that including serial correlation, if present, is far more important than correctly specifying the serial correlation function.
• Variance function: 
\[
\begin{pmatrix}
1 & t & t^2
\end{pmatrix}
D
\begin{pmatrix}
1 \\
t \\
 t^2
\end{pmatrix}
+ \sigma^2 + \tau^2
\]

• Comparison of smoothed average of squared OLS residuals and fitted variance function:

![Variance function graph]
• Inclusion of serial correlation leads to different estimates for the variance components in $D$.

• Therefore, the fitted variance function differs from the one obtained before without serial correlation.

• The deviation for small values of $t$ remains, but the functions coincide better for large $t$. 
10.6 Reduction of Preliminary Random-effects Structure

- Once an appropriate residual covariance model is obtained, one can try to reduce the number of random effects in the preliminary random-effects structure.

- This is done based on inferential tools for variance components.
10.7 Reduction of Preliminary Mean Structure

- Once an appropriate covariance model is obtained, one can try to reduce the number of covariates in the preliminary mean structure.

- This is done based on inferential tools for fixed effects.

- In case there is still some doubt about the validity of the marginal covariance structure, robust inference can be used to still obtain correct inferences.
10.8 Example: Prostate Data

- Fixed effects estimates from the final model, under Gaussian serial correlation, and without serial correlation:

<table>
<thead>
<tr>
<th>Effect</th>
<th>Parameter</th>
<th>Serial corr. Estimate (s.e.)</th>
<th>No serial corr. Estimate (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age effect</td>
<td>$\beta_1$</td>
<td>0.015 (0.006)</td>
<td>0.016 (0.006)</td>
</tr>
<tr>
<td>Intercepts:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>$\beta_2$</td>
<td>$-0.496$ (0.411)</td>
<td>$-0.564$ (0.428)</td>
</tr>
<tr>
<td>BPH</td>
<td>$\beta_3$</td>
<td>0.320 (0.470)</td>
<td>0.275 (0.488)</td>
</tr>
<tr>
<td>L/R cancer</td>
<td>$\beta_4$</td>
<td>1.216 (0.469)</td>
<td>1.099 (0.486)</td>
</tr>
<tr>
<td>Met. cancer</td>
<td>$\beta_5$</td>
<td>2.353 (0.518)</td>
<td>2.284 (0.531)</td>
</tr>
<tr>
<td>Time effects:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BPH</td>
<td>$\beta_8$</td>
<td>$-0.376$ (0.070)</td>
<td>$-0.410$ (0.068)</td>
</tr>
<tr>
<td>L/R cancer</td>
<td>$\beta_9$</td>
<td>$-1.877$ (0.210)</td>
<td>$-1.870$ (0.233)</td>
</tr>
<tr>
<td>Met. cancer</td>
<td>$\beta_{10}$</td>
<td>$-2.274$ (0.244)</td>
<td>$-2.303$ (0.262)</td>
</tr>
<tr>
<td>Time$^2$ effects:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cancer</td>
<td>$\beta_{14} = \beta_{15}$</td>
<td>0.484 (0.073)</td>
<td>0.510 (0.088)</td>
</tr>
</tbody>
</table>
- Variance components estimates from the final model, under Gaussian serial correlation, and without serial correlation:

<table>
<thead>
<tr>
<th>Effect</th>
<th>Parameter</th>
<th>Serial corr. Estimate (s.e.)</th>
<th>No serial corr. Estimate (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariance of $b_i$:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{var}(b_{1i})$</td>
<td>$d_{11}$</td>
<td>0.393 (0.093)</td>
<td>0.443 (0.093)</td>
</tr>
<tr>
<td>$\text{var}(b_{2i})$</td>
<td>$d_{22}$</td>
<td>0.550 (0.187)</td>
<td>0.842 (0.203)</td>
</tr>
<tr>
<td>$\text{var}(b_{3i})$</td>
<td>$d_{33}$</td>
<td>0.056 (0.028)</td>
<td>0.114 (0.035)</td>
</tr>
<tr>
<td>$\text{cov}(b_{1i}, b_{2i})$</td>
<td>$d_{12} = d_{21}$</td>
<td>$-0.382$ (0.114)</td>
<td>$-0.490$ (0.124)</td>
</tr>
<tr>
<td>$\text{cov}(b_{2i}, b_{3i})$</td>
<td>$d_{23} = d_{32}$</td>
<td>$-0.170$ (0.070)</td>
<td>$-0.300$ (0.082)</td>
</tr>
<tr>
<td>$\text{cov}(b_{3i}, b_{1i})$</td>
<td>$d_{13} = d_{31}$</td>
<td>0.098 (0.039)</td>
<td>0.148 (0.047)</td>
</tr>
<tr>
<td>Measurement error variance:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{var}(\varepsilon_{(wij)})$</td>
<td>$\sigma^2$</td>
<td>0.023 (0.002)</td>
<td>0.028 (0.002)</td>
</tr>
<tr>
<td>Gaussian serial correlation:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{var}(\varepsilon_{(wij)})$</td>
<td>$\tau^2$</td>
<td>0.029 (0.018)</td>
<td>—— (——)</td>
</tr>
<tr>
<td>Rate of exponential decrease</td>
<td>$1/\sqrt{\phi}$</td>
<td>0.599 (0.192)</td>
<td>—— (——)</td>
</tr>
<tr>
<td>REML log-likelihood</td>
<td></td>
<td>$-13.704$</td>
<td>$-20.165$</td>
</tr>
</tbody>
</table>
• Many standard errors are smaller under the model which includes the Gaussian serial correlation component

• Hence, adding the serial correlation leads to more efficient inferences for most parameters in the marginal model.
10.9 Random-effects Structure versus Residual Covariance Structure

- The marginal covariance structure equals

\[ V_i = Z_iDZ_i' + \Sigma_i \]

- Hence, the residual covariance \( \Sigma_i \) models all variation not yet been accounted for by random effects.

- In practice, one therefore often observes strong competition between these two sources of stochastic variation.

- This is also reflected in substantial correlations between the variance components estimates.
As an example, consider the final model for the prostate data, with Gaussian serial correlation

Estimated correlation matrix for variance components estimates:

\[
\text{Corr} \left( \hat{d}_{11}, \hat{d}_{12}, \hat{d}_{22}, \hat{d}_{13}, \hat{d}_{23}, \hat{d}_{33}, \hat{\tau}^2, 1/\sqrt{\hat{\phi}}, \hat{\sigma}^2 \right) =
\frac{1}{0.18} \begin{pmatrix}
1.00 & -0.87 & 0.62 & 0.70 & -0.49 & 0.39 & -0.18 & -0.10 & -0.00 \\
-0.87 & 1.00 & -0.85 & -0.94 & 0.75 & -0.63 & 0.21 & 0.08 & -0.03 \\
0.62 & -0.85 & 1.00 & 0.88 & -0.97 & 0.91 & -0.46 & -0.29 & 0.02 \\
0.70 & -0.94 & 0.88 & 1.00 & -0.82 & 0.72 & -0.22 & -0.06 & 0.05 \\
-0.49 & 0.75 & -0.97 & -0.82 & 1.00 & -0.97 & 0.51 & 0.33 & -0.02 \\
0.39 & -0.63 & 0.91 & 0.72 & -0.97 & 1.00 & -0.57 & -0.38 & 0.01 \\
-0.18 & 0.21 & -0.46 & -0.22 & 0.51 & -0.57 & 1.00 & 0.81 & 0.04 \\
-0.10 & 0.08 & -0.29 & -0.06 & 0.33 & -0.38 & 0.81 & 1.00 & 0.32 \\
-0.00 & -0.03 & 0.02 & 0.05 & -0.02 & 0.01 & 0.04 & 0.32 & 1.00
\end{pmatrix}
\]
• Relatively large correlations between $\hat{\tau}^2$ and the estimates of some of the parameters in $D$

• Small correlations between $\hat{\sigma}^2$ and the other estimates, except for $1/\sqrt{\hat{\phi}}$.

• Indeed, the serial correlation component vanishes for $\phi$ becoming infinitely large.
Chapter 11
Power Analyses under Linear Mixed Models

- $F$ test for fixed effects
- Calculation in SAS
- Examples
11.1 $F$ Statistics for Fixed Effects

- Consider a general linear hypothesis

\[ H_0 : L\beta = 0, \quad \text{versus} \quad H_A : L\beta \neq 0 \]

- $F$ test statistic:

\[
F = \frac{\hat{\beta}'L' \left[ L \left( \sum_{i=1}^{N} X_i'V_i^{-1}(\alpha)X_i \right)^{-1} L' \right]^{-1} L\hat{\beta}}{\text{rank}(L)}.
\]

- Approximate null-distribution of $F$ is $F$ with numerator degrees of freedom equal to $\text{rank}(L)$
- Denominator degrees of freedom to be estimated from the data:
  - Containment method
  - Satterthwaite approximation
  - Kenward and Roger approximation
  - ... 

- In general (not necessarily under $H_0$), $F$ is approximately $F$ distributed with the same numbers of degrees of freedom, but with non-centrality parameter

$$
\phi = \beta' L' \left[ L \left( \sum_{i=1}^{N} X_i' V_i^{-1} (\bar{\alpha}) X_i \right)^{-1} L' \right]^{-1} L \beta
$$

which equals 0 under $H_0$.

- This can be used to calculate powers under a variety of models, and under a variety of alternative hypotheses
• Note that $\phi$ is equal to $\text{rank}(L) \times F$, and with $\widehat{\beta}$ replaced by $\beta$.

• The SAS procedure MIXED can therefore be used for the calculation of $\phi$ and the related numbers of degrees of freedom.
11.2 Calculation in SAS

- Construct a data set of the same dimension and with the same covariates and factor values as the design for which power is to be calculated.

- Use as responses $y_i$ the average values $X_i\beta$ under the alternative model.

- The fixed effects estimate will then be equal to

$$\hat{\beta}(\alpha) = \left( \sum_{i=1}^{N} X_i' W_i(\alpha) X_i \right)^{-1} \sum_{i=1}^{N} X_i' W_i(\alpha) y_i$$

$$= \left( \sum_{i=1}^{N} X_i' W_i(\alpha) X_i \right)^{-1} \sum_{i=1}^{N} X_i' W_i(\alpha) X_i \beta = \beta$$

- Hence, the $F$-statistic reported by SAS will equal $\phi/\text{rank}(L)$.
This calculated $F$ value, and the associated numbers of degrees of freedom can be saved and used afterwards for calculation of the power.

Note that this requires keeping the variance components in $\alpha$ fixed, equal to the assumed population values.

Steps in calculations:
- Use PROC MIXED to calculate $\phi$, and degrees of freedom $\nu_1$ and $\nu_2$
- Calculate critical value $F_c$:
  $$P(F_{\nu_1,\nu_2,0} > F_c) = \text{level of significance}$$
- Calculate power:
  $$\text{power} = P(F_{\nu_1,\nu_2,\phi} > F_c)$$

The SAS functions ‘finv’ and ‘probf’ are used to calculated $F_c$ and the power
11.3 Example 1

- Re-consider the random-intercepts model previously discussed for the rat data:

\[ Y_{ij} = (\beta_0 + b_{1i}) + (\beta_1 L_i + \beta_2 H_i + \beta_3 C_i) t_{ij} + \varepsilon_{ij} \]

in which \( t_{ij} \) equals \( \ln[1 + (\text{Age}_{ij} - 45)/10] \)

- This model is fitted in SAS as follows:

```sas
proc mixed data = test;
class treat rat;
model y = treat*t / solution ddfm=kr;
random intercept / subject=rat;
contrast 'Equal slopes' treat*t 1 -1 0, treat*t 1 0 -1;
run;
```
• The CONTRAST statement is added to test equality of the average slopes.

• Suppose a new experiment is to be designed, to test the above hypothesis, when the true parameter values are given by:

<table>
<thead>
<tr>
<th>Effect</th>
<th>Parameter</th>
<th>True value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>$\beta_0$</td>
<td>68</td>
</tr>
<tr>
<td>Time effects:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low dose</td>
<td>$\beta_1$</td>
<td>7</td>
</tr>
<tr>
<td>High dose</td>
<td>$\beta_2$</td>
<td>7.5</td>
</tr>
<tr>
<td>Control</td>
<td>$\beta_3$</td>
<td>6.5</td>
</tr>
<tr>
<td>Covariance of $b_i$:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{var}(b_{1i})$</td>
<td>$d_{11}$</td>
<td>3.6</td>
</tr>
<tr>
<td>Residual variance:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{var}(\varepsilon_{ij})$</td>
<td>$\sigma^2$</td>
<td>1.4</td>
</tr>
</tbody>
</table>
The power of a design with 10 rats per treatment group is calculated as follows:

Construction of data set with expected averages as response values:

data power;
do treat=1 to 3;
  do rat=1 to 10;
    do age=50 to 110 by 10;
      t=log(1+(age-45)/10);
      if treat=1 then y=68 + 7.5*t;
      if treat=2 then y=68 + 7.0*t;
      if treat=3 then y=68 + 6.5*t;
      output;
    end;
  end;
end;
end;
Fit model, keeping the variance components equal to their true values:

```plaintext
proc mixed data = power noprofile;
class treat rat;
model y = treat*t ;
random intercept / subject=rat(treat);
parms (3.6) (1.4) / noiter;
contrast 'Equal slopes' treat*t 1 -1 0, 
              treat*t 1 0 -1;
ods output contrasts=c;
run;
```

- PARMS statement to specify starting values for the variance components.
- The ‘noiter’ and ‘noprofile’ options request that no iterations be performed and that inferences are based on the specified values.
- ODS statement needed to save $F$, $\nu_1$ and $\nu_2$. 
Calculation of $\phi$, $F_c$ and power:

```sas
data power;
set c;
alpha=0.05;
ncparm=numdf*fvalue;
fc=finv(1-alpha,numdf,dendf,0);
power=1-probf(fc,numdf,dendf,ncparm);
run;

proc print;run;
```

Output:

<table>
<thead>
<tr>
<th>Label</th>
<th>Num</th>
<th>Den</th>
<th>FValue</th>
<th>ProbF</th>
<th>alpha</th>
<th>ncparm</th>
<th>fc</th>
<th>power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equal slopes</td>
<td>2</td>
<td>177</td>
<td>4.73</td>
<td>0.0100</td>
<td>0.05</td>
<td>9.46367</td>
<td>3.04701</td>
<td>0.78515</td>
</tr>
</tbody>
</table>
• Hence, there is a power of 78.5% to detect the prespecified differences at the 5% level of significance.

• Increasing the number of rats yields the following powers:

<table>
<thead>
<tr>
<th>Group size</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>78.5%</td>
</tr>
<tr>
<td>11</td>
<td>82.5%</td>
</tr>
<tr>
<td>12</td>
<td>85.9%</td>
</tr>
<tr>
<td>13</td>
<td>88.7%</td>
</tr>
<tr>
<td>14</td>
<td>91.0%</td>
</tr>
<tr>
<td>15</td>
<td>92.9%</td>
</tr>
<tr>
<td>20</td>
<td>97.9%</td>
</tr>
</tbody>
</table>
11.4 Example 2

- We continue the previous random-intercepts model and study the effect of varying the variance components values.

- Results (10 rats per group):

<table>
<thead>
<tr>
<th>$d_{11}$</th>
<th>3.2</th>
<th>3.6</th>
<th>4.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2 = 1.0$</td>
<td>89.3%</td>
<td>88.5%</td>
<td>87.9%</td>
</tr>
<tr>
<td>$\sigma^2 = 1.4$</td>
<td>79.8%</td>
<td>78.5%</td>
<td>77.4%</td>
</tr>
<tr>
<td>$\sigma^2 = 1.8$</td>
<td>71.9%</td>
<td>70.3%</td>
<td>68.9%</td>
</tr>
</tbody>
</table>

- Conclusions:
  - The power decreases as the total variance increases.
  - Keeping the total variance constant, the power increases as the intraclass correlation $\rho_I = d_{11}/(d_{11} + \sigma^2)$ increases.
11.5 Example 3

11.5.1 Introduction

• Experiment for the comparison of two treatments $A$ and $B$

• A total of $N$ general practitioners (GP’s) involved

• Each GP treats $n$ subjects

• $Y_{ij}$ is the response for subject $j$ treated by GP $i$

• The analysis should account for the variability between GP’s
• We use the following random-intercepts model, where the random intercepts reflect random GP effects:

\[
Y_{ij} = \begin{cases} 
\beta_1 + b_{1i} + \varepsilon_{ij} & \text{if treatment } A \\
\beta_2 + b_{1i} + \varepsilon_{ij} & \text{if treatment } B
\end{cases}
\]

• Assumed true parameter values:

<table>
<thead>
<tr>
<th>Effect</th>
<th>Parameter</th>
<th>True value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed effects:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average treatment $A$</td>
<td>$\beta_1$</td>
<td>1</td>
</tr>
<tr>
<td>Average treatment $B$</td>
<td>$\beta_2$</td>
<td>2</td>
</tr>
<tr>
<td>Variance components:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{var}(b_{1i})$</td>
<td>$d_{11}$</td>
<td>?</td>
</tr>
<tr>
<td>$\text{var}(\varepsilon_{ij})$</td>
<td>$\sigma^2$</td>
<td>?</td>
</tr>
<tr>
<td>$d_{11} + \sigma^2$</td>
<td></td>
<td>4</td>
</tr>
</tbody>
</table>
• Hence, the individual variance components are unknown. Only the total variability is known to equal 4.

• Power analyses will be performed for several values for the intraclass correlation \( \rho_I = d_{11}/(d_{11} + \sigma^2) \)
11.5.2 Case 1: Treatments Assigned to GP’s

- We now consider the situation in which the treatments will be randomly assigned to GP’s, and all subjects with the same GP will be treated identically.

- Powers for $2 \times 25 = 50$ GP’s, each treating 10 subjects ($\alpha = 0.05$):

<table>
<thead>
<tr>
<th>$\rho_I$</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>86%</td>
</tr>
<tr>
<td>0.50</td>
<td>65%</td>
</tr>
<tr>
<td>0.75</td>
<td>50%</td>
</tr>
</tbody>
</table>

- The power decreases as the intraclass correlation increases
11.5.3 Case 2: Treatments Assigned to Subjects

- We now consider the situation in which the treatments will be randomly assigned to subjects within GP’s, with the same number \( n/2 \) of subjects assigned to both treatments.

- Powers for \( 2 \times 5 = 10 \) subjects within 10 GP’s (\( \alpha = 0.05 \)): 

<table>
<thead>
<tr>
<th>( \rho_I )</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>81%</td>
</tr>
<tr>
<td>0.50</td>
<td>94%</td>
</tr>
<tr>
<td>0.75</td>
<td>100%</td>
</tr>
</tbody>
</table>

- The power increases as the intraclass correlation increases.

- Note also that Case 2 requires many less observations than Case 1.
11.5.4 Conclusion

Within-‘subject’ correlation

increases power for inferences on within-‘subject’ effects,

but decreases power for inferences on between-‘subject’ effects
Part II

Marginal Models for Non-Gaussian Longitudinal Data
Chapter 12
The Toenail Data

- **Toenail Dermatophyte Onychomycosis**: Common toenail infection, difficult to treat, affecting more than 2% of population.

- Classical treatments with antifungal compounds need to be administered until the whole nail has grown out healthy.

- New compounds have been developed which reduce treatment to 3 months.

- Randomized, double-blind, parallel group, multicenter study for the comparison of two such new compounds (A and B) for oral treatment.
Research question:

Severity relative to treatment of TDO?

- 2 x 189 patients randomized, 36 centers
- 48 weeks of total follow up (12 months)
- 12 weeks of treatment (3 months)
- measurements at months 0, 1, 2, 3, 6, 9, 12.
• Frequencies at each visit (both treatments):
Chapter 13
The Analgesic Trial

- single-arm trial with 530 patients recruited (491 selected for analysis)
- analgesic treatment for pain caused by chronic nonmalignant disease
- treatment was to be administered for 12 months
- we will focus on Global Satisfaction Assessment (GSA)
- GSA scale goes from 1=very good to 5=very bad
- GSA was rated by each subject 4 times during the trial, at months 3, 6, 9, and 12.
• Research questions:
  ▶ Evolution over time
  ▶ Relation with baseline covariates: age, sex, duration of the pain, type of pain, disease progression, Pain Control Assessment (PCA), . . .
  ▶ Investigation of dropout

• Frequencies:

<table>
<thead>
<tr>
<th>GSA</th>
<th>Month 3</th>
<th>Month 6</th>
<th>Month 9</th>
<th>Month 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>55</td>
<td>38</td>
<td>40</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>14.3%</td>
<td>12.6%</td>
<td>17.6%</td>
<td>13.5%</td>
</tr>
<tr>
<td>2</td>
<td>112</td>
<td>84</td>
<td>67</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>29.1%</td>
<td>27.8%</td>
<td>29.5%</td>
<td>29.6%</td>
</tr>
<tr>
<td>3</td>
<td>151</td>
<td>115</td>
<td>76</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>39.2%</td>
<td>38.1%</td>
<td>33.5%</td>
<td>43.5%</td>
</tr>
<tr>
<td>4</td>
<td>52</td>
<td>51</td>
<td>33</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>13.5%</td>
<td>16.9%</td>
<td>14.5%</td>
<td>12.1%</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>14</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>3.9%</td>
<td>4.6%</td>
<td>4.9%</td>
<td>1.4%</td>
</tr>
<tr>
<td>Tot</td>
<td>385</td>
<td>302</td>
<td>227</td>
<td>223</td>
</tr>
</tbody>
</table>
### Missingness:

<table>
<thead>
<tr>
<th>Measurement occasion</th>
<th>Month 3</th>
<th>Month 6</th>
<th>Month 9</th>
<th>Month 12</th>
<th>Number</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completers</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>163</td>
<td>41.2</td>
</tr>
<tr>
<td>Dropouts</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>M</td>
<td>51</td>
<td>12.91</td>
</tr>
<tr>
<td></td>
<td>O</td>
<td>O</td>
<td>M</td>
<td>M</td>
<td>51</td>
<td>12.91</td>
</tr>
<tr>
<td></td>
<td>O</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>63</td>
<td>15.95</td>
</tr>
<tr>
<td>Non-monotone missingness</td>
<td>O</td>
<td>O</td>
<td>M</td>
<td>O</td>
<td>30</td>
<td>7.59</td>
</tr>
<tr>
<td></td>
<td>O</td>
<td>M</td>
<td>O</td>
<td>O</td>
<td>7</td>
<td>1.77</td>
</tr>
<tr>
<td></td>
<td>O</td>
<td>M</td>
<td>O</td>
<td>M</td>
<td>2</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>O</td>
<td>M</td>
<td>M</td>
<td>O</td>
<td>18</td>
<td>4.56</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>2</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>O</td>
<td>O</td>
<td>M</td>
<td>1</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>O</td>
<td>M</td>
<td>O</td>
<td>1</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>O</td>
<td>M</td>
<td>M</td>
<td>3</td>
<td>0.76</td>
</tr>
</tbody>
</table>
Chapter 14
The National Toxicology Program (NTP) Data

Developmental Toxicity Studies

- Research Triangle Institute

- The effect in mice of 3 chemicals:
  - DEHP: di(2-ethylhexyl)-phtalate
  - EG: ethylene glycol
  - DYME: diethylene glycol dimethyl ether
• Implanted fetuses:
  ▶ death/resorbed
  ▶ viable:
    * weight
    * malformations: visceral, skeletal, external

• Data structure:

```
            dam
           /         \
          /           /
   implant (m_i) viable (n_i) non-viable (r_i)
           /     \        /     \        /     \
          /       /       /       /       /     /
   malf. (z_i) weight death resorption
            /   \
           /     /
          / 1   K
```
<table>
<thead>
<tr>
<th>Exposure</th>
<th>Dose</th>
<th>Impl.</th>
<th>Viab.</th>
<th>Live</th>
<th>Size (mean)</th>
<th>Malformations</th>
</tr>
</thead>
<tbody>
<tr>
<td>EG</td>
<td>0</td>
<td>25</td>
<td>25</td>
<td>297</td>
<td>11.9</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>750</td>
<td>24</td>
<td>24</td>
<td>276</td>
<td>11.5</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>1500</td>
<td>23</td>
<td>22</td>
<td>229</td>
<td>10.4</td>
<td>1.7</td>
</tr>
<tr>
<td></td>
<td>3000</td>
<td>23</td>
<td>23</td>
<td>226</td>
<td>9.8</td>
<td>7.1</td>
</tr>
<tr>
<td>DEHP</td>
<td>0</td>
<td>30</td>
<td>30</td>
<td>330</td>
<td>13.2</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>44</td>
<td>26</td>
<td>26</td>
<td>288</td>
<td>11.1</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>91</td>
<td>26</td>
<td>26</td>
<td>277</td>
<td>10.7</td>
<td>5.4</td>
</tr>
<tr>
<td></td>
<td>191</td>
<td>24</td>
<td>17</td>
<td>137</td>
<td>8.1</td>
<td>17.5</td>
</tr>
<tr>
<td></td>
<td>292</td>
<td>25</td>
<td>9</td>
<td>50</td>
<td>5.6</td>
<td>54.0</td>
</tr>
<tr>
<td>DYME</td>
<td>0</td>
<td>21</td>
<td>21</td>
<td>282</td>
<td>13.4</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>62.5</td>
<td>20</td>
<td>20</td>
<td>225</td>
<td>11.3</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>125</td>
<td>24</td>
<td>24</td>
<td>290</td>
<td>12.1</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>23</td>
<td>23</td>
<td>261</td>
<td>11.3</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>22</td>
<td>22</td>
<td>141</td>
<td>6.1</td>
<td>66.0</td>
</tr>
</tbody>
</table>
Chapter 15
Generalized Linear Models

▷ The model
▷ Maximum likelihood estimation
▷ Examples
▷ McCullagh and Nelder (1989)
15.1 The Generalized Linear Model

- Suppose a sample $Y_1, \ldots, Y_N$ of independent observations is available

- All $Y_i$ have densities $f(y_i|\theta_i, \phi)$ which belong to the exponential family:
  \[
  f(y|\theta_i, \phi) = \exp \left\{ \phi^{-1}[y\theta_i - \psi(\theta_i)] + c(y, \phi) \right\}
  \]

- $\theta_i$ the natural parameter

- Linear predictor: $\theta_i = x_i'\beta$

- $\theta$ is the scale parameter (overdispersion parameter)

- $\psi(.)$ is a function to be discussed next
15.2 Mean and Variance

• We start from the following general propterty:

\[
\int f(y|\theta, \phi) dy = \int \exp \{\phi^{-1}[y\theta - \psi(\theta)] + c(y, \phi)\} \ dy = 1
\]

• Taking first and second-order derivatives with respect to \( \theta \) yields

\[
\begin{align*}
\frac{\partial}{\partial \theta} \int f(y|\theta, \phi) \ dy &= 0 \\
\frac{\partial^2}{\partial \theta^2} \int f(y|\theta, \phi) \ dy &= 0
\end{align*}
\]
\[
\begin{align*}
\int [y - \psi'(\theta)] f(y|\theta, \phi) \, dy &= 0 \\
\int [\phi^{-1}(y - \psi'(\theta))^2 - \psi''(\theta)] f(y|\theta, \phi) \, dy &= 0
\end{align*}
\]

\[
\begin{align*}
E(Y) &= \psi'(\theta) \\
\text{Var}(Y) &= \phi \psi''(\theta)
\end{align*}
\]

- Note that, in general, the mean \( \mu \) and the variance are related:

\[
\text{Var}(Y) = \phi \psi''[\psi^{-1}(\mu)] = \phi v(\mu)
\]
• The function \( v(\mu) \) is called the variance function.

• The function \( \psi'^{-1} \) which expresses \( \theta \) as function of \( \mu \) is called the link function.

• \( \psi' \) is the inverse link function
15.3 Examples

15.3.1 The Normal Model

- Model:
  \[ Y \sim N(\mu, \sigma^2) \]

- Density function:
  \[
  f(y|\theta, \phi) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{\sigma^2} (y - \mu)^2 \right\}
  \]
  \[
  = \exp \left\{ \frac{1}{\sigma^2} \left( y\mu - \frac{\mu^2}{2} \right) + \left( \frac{\ln(2\pi\sigma^2)}{2} - \frac{y^2}{2\sigma^2} \right) \right\}
  \]
• Exponential family:
  ▶ $\theta = \mu$
  ▶ $\phi = \sigma^2$
  ▶ $\psi(\theta) = \theta^2/2$
  ▶ $c(y, \phi) = \frac{\ln(2\pi\phi)}{2} - \frac{y^2}{2\phi}$

• Mean and variance function:
  ▶ $\mu = \theta$
  ▶ $v(\mu) = 1$

• Note that, under this normal model, the mean and variance are not related:

$$\phi v(\mu) = \sigma^2$$

• The link function is here the identity function: $\theta = \mu$
15.3.2 The Bernoulli Model

- Model:
  \[ Y \sim \text{Bernoulli}(\pi) \]

- Density function:
  \[
  f(y|\theta, \phi) = \pi^y(1-\pi)^{1-y} \\
  = \exp \left\{ y \ln \pi + (1-y) \ln(1-\pi) \right\} \\
  = \exp \left\{ y \ln \left( \frac{\pi}{1-\pi} \right) + \ln(1-\pi) \right\}
  \]
• Exponential family:
  \[ \theta = \ln \left( \frac{\pi}{1-\pi} \right) \]
  \[ \phi = 1 \]
  \[ \psi(\theta) = \ln(1 - \pi) = \ln(1 + \exp(\theta)) \]
  \[ c(y, \phi) = 0 \]

• Mean and variance function:
  \[ \mu = \frac{\exp \theta}{1 + \exp \theta} = \pi \]
  \[ v(\mu) = \frac{\exp \theta}{(1 + \exp \theta)^2} = \pi(1 - \pi) \]

• Note that, under this model, the mean and variance are related:
  \[ \phi v(\mu) = \mu(1 - \mu) \]

• The link function here is the logit link: \[ \theta = \ln \left( \frac{\mu}{1-\mu} \right) \]
15.3.3 The Poisson Model

- Model:
  \[ Y \sim \text{Poisson}(\lambda) \]

- Density function:

  \[
  f(y|\theta, \phi) = \frac{e^{-\lambda} \lambda^y}{y!} = \exp\{y \ln \lambda - \lambda - \ln y!\}
  \]
• Exponential family:
  \[ \theta = \ln \lambda \]
  \[ \phi = 1 \]
  \[ \psi(\theta) = \lambda = \exp \theta \]
  \[ c(y, \phi) = -\ln y! \]

• Mean and variance function:
  \[ \mu = \exp \theta = \lambda \]
  \[ v(\mu) = \exp \theta = \lambda \]

• Note that, under this model, the mean and variance are related:
  \[ \phi v(\mu) = \mu \]

• The link function is here the log link: \[ \theta = \ln \mu \]
15.4 Generalized Linear Models (GLM)

- Suppose a sample $Y_1, \ldots, Y_N$ of independent observations is available.

- All $Y_i$ have densities $f(y_i | \theta_i, \phi)$ which belong to the exponential family.

- In GLM's, it is believed that the differences between the $\theta_i$ can be explained through a linear function of known covariates:

$$\theta_i = \mathbf{x}_i \beta$$

- $\mathbf{x}_i$ is a vector of $p$ known covariates.

- $\beta$ is the corresponding vector of unknown regression parameters, to be estimated from the data.
15.5 Maximum Likelihood Estimation

- Log-likelihood:

\[ \ell(\beta, \phi) = \frac{1}{\phi} \sum_i [y_i \theta_i - \psi(\theta_i)] + \sum_i c(y_i, \phi) \]

- First order derivative with respect to \( \beta \):

\[ \frac{\partial \ell(\beta, \phi)}{\partial \beta} = \frac{1}{\phi} \sum_i \frac{\partial \theta_i}{\partial \beta} \left[ y_i - \psi'(\theta_i) \right] \]

- The score equations for \( \beta \) to be solved:

\[ S(\beta) = \sum_i \frac{\partial \theta_i}{\partial \beta} \left[ y_i - \psi'(\theta_i) \right] = 0 \]
• Since $\mu_i = \psi'(\theta_i)$ and $v_i = v(\mu_i) = \psi''(\theta_i)$, we have that

$$\frac{\partial \mu_i}{\partial \beta} = \psi''(\theta_i) \frac{\partial \theta_i}{\partial \beta} = v_i \frac{\partial \theta_i}{\partial \beta}$$

• The score equations now become

$$S(\beta) = \sum_i \frac{\partial \mu_i}{\partial \beta} v_i^{-1} (y_i - \mu_i) = 0$$

• Note that the estimation of $\beta$ depends on the density only through the means $\mu_i$ and the variance functions $v_i = v(\mu_i)$. 
The score equations need to be solved numerically:
- iterative (re-)weighted least squares
- Newton-Raphson
- Fisher scoring

Inference for $\beta$ is based on classical maximum likelihood theory:
- asymptotic Wald tests
- likelihood ratio tests
- score tests
In some cases, $\phi$ is a known constant, in other examples, estimation of $\phi$ may be required to estimate the standard errors of the elements in $\beta$.

Estimation can be based on $\text{Var}(Y_i) = \phi v_i$:

$$\hat{\phi} = \frac{1}{N - p} \sum_i (y_i - \hat{\mu}_i)^2 / v_i(\hat{\mu}_i)$$

For example, under the normal model, this would yield:

$$\hat{\sigma}^2 = \frac{1}{N - p} \sum_i (y_i - x_i'\hat{\beta})^2,$$

the mean squared error used in linear regression models to estimate the residual variance.
15.6 Illustration: The Analgesic Trial

- Early dropout (did the subject drop out after the first or the second visit)?
- Binary response
- PROC GENMOD can fit GLMs in general
- PROC LOGISTIC can fit models for binary (and ordered) responses
- SAS code for logit link:

```sas
proc genmod data=earlydrp;
  model earlydrp = pca0 weight psychiat physfct / dist=b;
run;

proc logistic data=earlydrp descending;
  model earlydrp = pca0 weight psychiat physfct;
run;
```
• SAS code for probit link:

    proc genmod data=earlydrp;
    model earlydrp = pca0 weight psychiat physfct / dist=b link=probit;
    run;

    proc logistic data=earlydrp descending;
    model earlydrp = pca0 weight psychiat physfct / link=probit;
    run;

• Selected output:

    Analysis Of Parameter Estimates

    Parameter  DF  Estimate  Standard Error  Wald 95% Confidence Limits  Chi-Square  Pr > ChiSq
    Intercept  1  -1.0673    0.7328    -2.5037    0.3690        2.12        0.1453
    PCA0       1   0.3981    0.1343     0.1349    0.6614        8.79        0.0030
    WEIGHT     1  -0.0211    0.0072    -0.0353    -0.0070       8.55        0.0034
    PSYCHIAT   1   0.7169    0.2871    0.1541    1.2796       6.23        0.0125
    PHYSFCT    1   0.0121    0.0050    0.0024    0.0219        5.97        0.0145
    Scale      0   1.0000    0.0000    1.0000    1.0000

    NOTE: The scale parameter was held fixed.
Chapter 16
Parametric Modeling Families

▷ Continuous outcomes
▷ Longitudinal generalized linear models
▷ Notation
16.1 Continuous Outcomes

- **Marginal Models:**
  \[
  E(Y_{ij} | x_{ij}) = x'_{ij} \beta
  \]

- **Random-Effects Models:**
  \[
  E(Y_{ij} | b_i, x_{ij}) = x'_{ij} \beta + z'_{ij} b_i
  \]

- **Transition Models:**
  \[
  E(Y_{ij} | Y_{i,j-1}, \ldots, Y_{i1}, x_{ij}) = x'_{ij} \beta + \alpha Y_{i,j-1}
  \]
16.2 Longitudinal Generalized Linear Models

- Normal case: easy transfer between models
- Also non-normal data can be measured repeatedly (over time)
- Lack of key distribution such as the normal $\rightarrow$
  - A lot of modeling options
  - Introduction of non-linearity
  - No easy transfer between model families

<table>
<thead>
<tr>
<th></th>
<th>cross-sectional</th>
<th>longitudinal</th>
</tr>
</thead>
<tbody>
<tr>
<td>normal outcome</td>
<td>linear model</td>
<td>LMM</td>
</tr>
<tr>
<td>non-normal outcome</td>
<td>GLM</td>
<td>?</td>
</tr>
</tbody>
</table>
16.3 Notation

- Let the outcomes for subject $i = 1, \ldots, N$ be denoted as $(Y_{i1}, \ldots, Y_{in_i})$.

- Group into a vector $Y_i$:
  - Binary data: each component is either 0 or 1.
  - (Binary data: each component is either 1 or 2.)
  - (Binary data: each component is either $-1$ or $+1$.)
  - (Categorical data: $Y_{ij} \in \{1, \ldots, c\}$.)

- The corresponding covariate vector is $x_{ij}$.

- It is convenient to use (binary 0/1 data):

\[
E(Y_{ij}) = \Pr(Y_{ij} = 1) = \mu_{ij} \quad \text{and} \quad \mu_{ijk} = E(Y_{ij}Y_{ik}) = \Pr(Y_{ij} = 1, Y_{ik} = 1)
\]
Chapter 17
Conditional Models

- A log-linear model
- Quadratic version of the model
- Linear version of the model
- Clustered-data versions of the model
- Transition models
17.1 A Log-linear Model

- Cox (1972)

- Joint distribution of $Y_i$ in terms of a multivariate exponential family:

$$f(y_i, \theta_i) = \exp \left( \sum_{j=1}^{n} \theta_{ij} y_{ij} + \sum_{j_1 < j_2} \theta_{i_{j_1 j_2}} y_{i_{j_1} j_1} y_{i_{j_2} j_2} + \ldots + \theta_{i_1 \ldots n} y_{i_1} \ldots y_{in} - A(\theta_i) \right)$$

$$= c(\theta_i) \exp \left( \sum_{j=1}^{n} \theta_{ij} y_{ij} + \sum_{j_1 < j_2} \theta_{i_{j_1 j_2}} y_{i_{j_1} j_1} y_{i_{j_2} j_2} + \ldots + \theta_{i_1 \ldots n} y_{i_1} \ldots y_{in} \right)$$

- $A(\theta_i)$ [equivalently, $c(\theta_i)$] is the normalizing constant

- $\theta_i$ is the canonical parameter, consisting of first, second, up to $n$th order components.
• Interpretation of Parameters:

▷ The parameters have a conditional interpretation:

\[
\theta_{ij} = \ln \left( \frac{\Pr(Y_{ij} = 1 | Y_{ik} = 0 ; k \neq j)}{\Pr(Y_{ij} = 0 | Y_{ik} = 0 ; k \neq j)} \right)
\]

▷ \( \Rightarrow \) the first order parameters (main effects) are interpreted as conditional logits.

▷ Similarly,

\[
\theta_{ijk} = \ln \left( \frac{\Pr(Y_{ij} = 1, Y_{ik} = 1 | Y_{i\ell} = 0 ; k, j \neq \ell) \Pr(Y_{ij} = 0, Y_{ik} = 0 | Y_{i\ell} = 0 ; k, j \neq \ell)}{\Pr(Y_{ij} = 1, Y_{ik} = 0 | Y_{i\ell} = 0 ; k, j \neq \ell) \Pr(Y_{ij} = 0, Y_{ik} = 1 | Y_{i\ell} = 0 ; k, j \neq \ell)} \right)
\]

▷ These are conditional log odds ratios.
• **Advantages:**
  
  ▶ The parameter vector is not constrained. All values of $\theta \in \mathbb{R}$ yield nonnegative probabilities.
  
  ▶ Calculation of the joint probabilities is fairly straightforward:
    * ignore the normalizing constant
    * evaluate the density for all possible sequences $y$
    * sum all terms to yield $c(\theta)^{-1}$

• **Drawbacks:**
  
  ▶ Due to above conditional interpretation, the models are less useful for regression.
    The dependence of $E(Y_{ij})$ on covariates involves all parameters, not only the main effects.
  
  ▶ The interpretation of the parameters depends on the length $n_i$ of a sequence.
    These drawbacks make marginal models or models that combine marginal and conditional features better suited.
17.2 Quadratic and Linear Versions

- Cox (1972) and others suggest that often the higher order interactions can be neglected. This claim is supported by empirical evidence.

- **The quadratic exponential model:**

  \[
  f(y_i, \theta_i) = \exp\left( \sum_{j=1}^{n} \theta_{ij}y_{ij} + \sum_{j_1 < j_2} \theta_{ij_1j_2}y_{ij_1}y_{ij_2} - A(\theta_i) \right) \\
  = c(\theta_i) \exp\left( \sum_{j=1}^{n} \theta_{ij}y_{ij} + \sum_{j_1 < j_2} \theta_{ij_1j_2}y_{ij_1}y_{ij_2} \right).
  \]

- **The linear exponential model:**

  \[
  f(y_i, \theta_i) = \exp\left( \sum_{j=1}^{n} \theta_{ij}y_{ij} - A(\theta_i) \right)
  \]

  then this model reflects the assumption of independence.

- The linear model equals logistic regression.
17.3 A Version for Clustered Binary Data

- NTP data: $Y_{ij}$ is malformation indicator for fetus $j$ in litter $i$

- Code $Y_{ij}$ as $-1$ or $1$

- $d_i$ is dose level at which litter $i$ is exposed

- Simplification: $\theta_{ij} = \theta_i = \beta_0 + \beta_d d_i$ and $\theta_{ij_1j_2} = \beta_a$

- Using $Z_i = \sum_{j=1}^{n_i} Y_{ij}$

  we obtain

  $$f(z_i|\theta_i, \beta_a) = \binom{n_i}{z_i} \exp\{\theta_i z_i + \beta_a z_i(n_i - z_i) - A(\theta_i)\}$$
17.4 Transition Models

- Molenberghs and Verbeke (2005, Section 11.5)

- Outcome \( Y_{ij} \) or error term \( \varepsilon_{ij} \) is a function of history \( h_{ij} = (Y_{i1}, \ldots, Y_{i,j-1}) \)

- Order of transition model: \# of previous outcomes in regression

- Stationary model: functional form of dependence independent of occurrence time

- A stationary first-order autoregressive model for continuous data is:

\[
Y_{i1} = \mathbf{x}_{i1}' \beta + \varepsilon_{i1}
\]

\[
Y_{ij} = \mathbf{x}_{ij}' \beta + \alpha Y_{i,j-1} + \varepsilon_{ij}
\]
• Assume
\[
\varepsilon_{i1} \sim N(0, \sigma^2) \quad \text{and} \quad \varepsilon_{ij} \sim N(0, \sigma^2(1 - \alpha^2))
\]
then
\[
\text{cov}(Y_{ij}, Y_{ij'}) = \alpha^{|j' - j|} \sigma^2
\]
implies a marginal multivariate normal model with AR(1) variance-covariance matrix.

• For non-Gaussian outcomes, first write
\[
Y_{ij} = \mu_{ij}^c + \varepsilon_{ij}^c
\]
and then
\[
\mu_{ij}^c = E(Y_{ij}|h_{ij})
\]
\[
\phi v^c(\mu_{ij}^c) = \text{var}(Y_{ij}|h_{ij})
\]
• Example of a linear predictor:

\[ \eta_{ij}(\mu_{ij}^c) = x'_ij \beta + \kappa(h_{ij}, \beta, \alpha) \]

• \( \kappa \) is a function of the history.

• This model is easy to fit since it leads to independent GLM contributions:

\[
f(y_{i1}, \ldots, y_{in_i}) = f(y_{i1}) \cdot f(y_{i2}|y_{i1}) \cdot f(y_{i3}|y_{i1}, y_{i2}) \cdot f(y_{in_i}|y_{i1}, \ldots, y_{i,n_i-1})
\]

\[= f(y_{i1}) \cdot \prod_{j=2}^{n_i} f(y_{ij}|h_{ij}) = f(y_{i1}, \ldots, y_{iq}) \cdot \prod_{j=q+1}^{n_i} f(y_{ij}|h_{ij})\]

• This product yields \( n_i - q \) independent univariate GLM contributions.

• A separate model may need to be considered for the first \( q \) measurements.
• A logistic-regression type example:

$$\text{logit}[P(Y_{ij} = 1|\mathbf{x}_{ij}, Y_{i,j-1} = y_{i,j-1}, \beta, \alpha)] = \mathbf{x}_{ij}'\beta + \alpha y_{i,j-1}.$$ 

• The marginal means and variances do not follow easily, except in the normal case.

• Recursive formulas are:

$$\mu_{ij} = \mu_{ij}^c(0)[1 - \mu_{i,j-1}] + \mu_{ij}^c(1)\mu_{i,j-1}$$

$$v_{ij} = [\mu_{ij}^c(1) - \mu_{ij}^c(0)]^2v_{i,j-1} + v_{ij}^c(0)[1 - \mu_{i,j-1}] + v_{ij}^c(1)\mu_{i,j-1}$$
17.4.1 Analysis of the Toenail Data

- Formulate a transition model (Model I):
  \[ Y_{ij} \sim \text{Bernoulli}(\mu_{ij}) \]
  \[ \text{logit} \left( \frac{\mu_{ij}}{1 - \mu_{ij}} \right) = \beta_0 + \beta_1 T_i + \beta_2 t_{ij} + \beta_3 T_i t_{ij} + \alpha_1 y_{i,j-1} \]

- To account for unequal spacing (Model II):
  - \( \alpha_1 \) describes the transition effect for the later measurements
  - \( \alpha_{1a} \) is the ‘excess’ during the first quarter
  - hence: autoregressive effect at months 1, 2, and 3 is \( \alpha_1 + \alpha_{1a} \)

- Alternatively: dependence on two prior occasions:
  \[ \text{logit} \left( \frac{\mu_{ij}}{1 - \mu_{ij}} \right) = \beta_0 + \beta_1 T_i + \beta_2 t_{ij} + \beta_3 T_i t_{ij} + \alpha_1 y_{i,j-1} + \alpha_2 y_{i,j-2} \]
- Fitted models:

<table>
<thead>
<tr>
<th>Effect</th>
<th>Par.</th>
<th>I</th>
<th>II</th>
<th>Second order</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>( \beta_0 )</td>
<td>-3.14 (0.27)</td>
<td>-3.77 (0.34)</td>
<td>-3.28 (0.34)</td>
</tr>
<tr>
<td>( T_i )</td>
<td>( \beta_1 )</td>
<td>0.00 (0.31)</td>
<td>-0.08 (0.32)</td>
<td>0.13 (0.39)</td>
</tr>
<tr>
<td>( t_{ij} )</td>
<td>( \beta_2 )</td>
<td>-0.09 (0.04)</td>
<td>0.03 (0.05)</td>
<td>-0.05 (0.04)</td>
</tr>
<tr>
<td>( T_i \cdot t_{ij} )</td>
<td>( \beta_3 )</td>
<td>-0.08 (0.06)</td>
<td>-0.06 (0.06)</td>
<td>-0.09 (0.07)</td>
</tr>
<tr>
<td>Dep. on ( Y_{i,j-1} )</td>
<td>( \alpha_1 )</td>
<td>4.48 (0.22)</td>
<td>3.59 (0.29)</td>
<td>4.01 (0.39)</td>
</tr>
<tr>
<td>Dep. on ( Y_{i,j-1} )</td>
<td>( \alpha_{1a} )</td>
<td></td>
<td>1.56 (0.35)</td>
<td></td>
</tr>
<tr>
<td>Dep. on ( Y_{i,j-2} )</td>
<td>( \alpha_2 )</td>
<td></td>
<td></td>
<td>0.25 (0.38)</td>
</tr>
</tbody>
</table>
• Two separate models, depending on the level of the previous outcome:

\[
\text{logit} \left( \frac{\mu_{ij}}{1 - \mu_{ij}} \right) = (\beta_{00} + \beta_{10}T_i + \beta_{20}t_{ij} + \beta_{30}T_i t_{ij})I_{Y_{i,j-1}=0} + (\beta_{01} + \beta_{11}T_i + \beta_{21}t_{ij} + \beta_{31}T_i t_{ij})I_{Y_{i,j-1}=1}
\]

• Fitted model:

<table>
<thead>
<tr>
<th>Effect</th>
<th>Par.</th>
<th>Estimate (s.e.)</th>
<th>Par.</th>
<th>Estimate (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>$\beta_{00}$</td>
<td>-3.92 (0.56)</td>
<td>$\beta_{01}$</td>
<td>1.56 (1.26)</td>
</tr>
<tr>
<td>$T_i$</td>
<td>$\beta_{10}$</td>
<td>0.45 (0.70)</td>
<td>$\beta_{11}$</td>
<td>-0.01 (0.37)</td>
</tr>
<tr>
<td>$t_{ij}$</td>
<td>$\beta_{20}$</td>
<td>-0.06 (0.09)</td>
<td>$\beta_{21}$</td>
<td>-0.20 (0.06)</td>
</tr>
<tr>
<td>$T_i \cdot t_{ij}$</td>
<td>$\beta_{30}$</td>
<td>0.07 (0.10)</td>
<td>$\beta_{31}$</td>
<td>0.04 (0.07)</td>
</tr>
</tbody>
</table>
17.4.2 Transition Model in SAS

- Prepare models so that the previous outcome can be used as a covariate (using the same code as used to fit a model for dropout – see Part V)

```sas
%dropout(data=test,id=idnum,time=time,response=onyresp,out=test2);

data test2a;
set test2;
prenv1=prev;
drop prev;
run;

%dropout(data=test2a,id=idnum,time=time,response=prev1,out=test3);

data test3a;
set test3;
prenv2=prev;
drop prev;
run;
```
The result for the first subject is

<table>
<thead>
<tr>
<th>Obs</th>
<th>idnum</th>
<th>time</th>
<th>treatn</th>
<th>onyresp</th>
<th>prev1</th>
<th>prev2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>9</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>12</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Code to fit a transition model:

```plaintext
proc genmod data=test3a descending;
model onyresp = treatn time treatn*time prev1 / dist=binomial;
run;
```
• When both predecessors are used, one merely adds ‘prev2’ to MODEL statement:

\[
\text{model onyresp = prev1 treatn*prev1 time*prev1 treatn*time*prev1} \\
/ \text{ noint dist=binomial;}
\]

• To fit Model II, an additional variable ‘prev1a’ needs to be created:

\begin{verbatim}
data test3b;
set test3a;
prev1a=prev1;
if time>3 then prev1a=0;
run;
\end{verbatim}

which is then added to the logistic regression, next to ‘prev1.’
Chapter 18
Full Marginal Models

▶ Introduction
▶ Link functions
▶ Associations
▶ Bahadur model
▶ Multivariate Probit model
▶ Example: POPS data
18.1 Introduction

- Choices to make:
  - Description of mean profiles (univariate parameters) and of association (bivariate and higher order parameters)
  - Degree of modeling:
    * joint distribution fully specified $\Rightarrow$ likelihood procedures
    * only a limited number of moments $\Rightarrow$ e.g., generalized estimating equations

- Minimally, one specifies:
  - $\eta_i(\mu_i) = \{\eta_{i1}(\mu_{i1}), \ldots, \eta_{in}(\mu_{in})\}$
  - $E(Y_i) = \mu_i$ and $\eta_i(\mu_i) = X_i\beta$
  - $\text{var}(Y_i) = \phi v(\mu_i)$ where $v(.)$ is a known variance function
  - $\text{corr}(Y_i) = R(\alpha)$
18.2 Univariate Link Functions

- The marginal **logit link**:

  \[ \eta_{ij} = \ln(\mu_{ij}) - \ln(1 - \mu_{ij}) = \text{logit}(\mu_{ij}). \]

- The **probit link**:

  \[ \eta_{ij} = \Phi^{-1}(\mu_{ij}). \]

- The complementary log-log link

  \[ \cdots \]
18.3 Pairwise Association

- **Success probability approach.** (Ekholm 1991)
  
  Logit link for two-way probabilities

  \[ \eta_{ijk} = \ln(\mu_{ijk}) - \ln(1 - \mu_{ijk}) = \logit(\mu_{ijk}), \]

- **Marginal correlation coefficient.** (Bahadur model)

  \[ \rho_{ijk} = \frac{\mu_{ijk} - \mu_{ij}\mu_{ik}}{\sqrt{\mu_{ij}(1 - \mu_{ij})\mu_{ik}(1 - \mu_{ik})}} \]

  \[ \eta_{ijk} = \ln(1 + \rho_{ijk}) - \ln(1 - \rho_{ijk}) \quad (\text{Fisher’s } z \text{ transform}) \]
• **Marginal odds ratio.** (Dale model)

\[
\psi_{ijk} = \frac{(\mu_{ijk})(1 - \mu_{ij} - \mu_{ik} + \mu_{ijk})}{(\mu_{ik} - \mu_{ijk})(\mu_{ij} - \mu_{ijk})} \\
= \frac{\Pr(Y_{ij} = 1, Y_{ik} = 1)\Pr(Y_{ij} = 0, Y_{ik} = 0)}{\Pr(Y_{ij} = 0, Y_{ik} = 1)\Pr(Y_{ij} = 1, Y_{ik} = 0)}
\]

\[
\eta_{ijk} = \ln(\psi_{ijk}) \quad \text{(log odds ratio)}
\]

• Higher order association defined similarly

• Calculations can become cumbersome
18.4 The Bahadur Model

- **Univariate:** \( E(Y_{ij}) = P(Y_{ij} = 1) \equiv \pi_{ij} \).

- **Bivariate:** \( E(Y_{ij}Y_{ik}) = P(Y_{ij} = 1, Y_{ik} = 1) \equiv \pi_{ijk} \).

- **Correlation structure:**
  \[
  \text{Corr}(Y_{ij}, Y_{ik}) \equiv \rho_{ijk} = \frac{\pi_{ijk} - \pi_{ij}\pi_{ik}}{[\pi_{ij}(1 - \pi_{ij})\pi_{ik}(1 - \pi_{ik})]^{1/2}}.
  \]

- This yields expression for pairwise probabilities:
  \[
  \pi_{ijk} = \pi_{ij}\pi_{ik} + \rho_{ijk}[\pi_{ij}(1 - \pi_{ij})\pi_{ik}(1 - \pi_{ik})]^{1/2}.
  \]

- Similarly for the full joint distribution \( f(y) \).
Let

\[ \varepsilon_{ij} = \frac{Y_{ij} - \pi_{ij}}{\sqrt{\pi_{ij}(1 - \pi_{ij})}} \quad \text{and} \quad e_{ij} = \frac{y_{ij} - \pi_{ij}}{\sqrt{\pi_{ij}(1 - \pi_{ij})}}, \]

and

\[ \rho_{ijk} = E(\varepsilon_{ij}\varepsilon_{ik}), \]

\[ \rho_{ijkl} = E(\varepsilon_{ij}\varepsilon_{ik}\varepsilon_{il}), \]

\[ \vdots \]

\[ \rho_{i12\ldots ni} = E(\varepsilon_{i1}\varepsilon_{i2}\ldots \varepsilon_{in_i}). \]
A general expression:
\[ f(y_i) = f_1(y_i)c(y_i), \]
with
\[ f_1(y_i) = \prod_{j=1}^{n_i} \pi_{ij}^{y_{ij}}(1 - \pi_{ij})^{1-y_{ij}} \]
and
\[ c(y_i) = 1 + \sum_{j<k} \rho_{ijk} e_{ij} e_{ik} + \sum_{j<k<\ell} \rho_{ijk\ell} e_{ij} e_{ik} e_{i\ell} + \ldots + \rho_{i12...ni} e_{i1} e_{i2} \ldots e_{in_i}. \]
18.5 The Multivariate Probit Model

- E.g., $4 \times 3$ categorical outcome arises from underlying bivariate normal

- Covariate effects $\equiv$ shift of cut off points

- Correlation $=$ polychoric correlation: allowed to depend on covariates
18.6 The POPS Data

- **Project On Preterm and Small for Gestational Age Infants**

- 1530 Dutch children (1983)

- Collected data:

  - **Perinatal information:**
    - Bilirubin value
    - Neonatal seizures
    - Congenital malformations

  - **Ability scores at the age of 2:**
    - Are the child’s movements natural?
    - Can the child pile three bricks?
    - Can the child put a ball in a boxed when asked to?
## 18.7 Application to POPS Data

<table>
<thead>
<tr>
<th></th>
<th>Bahad</th>
<th>Probit</th>
<th>Dale-Norm</th>
<th>Dale-Logist</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First Ability Score</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>3.67(0.49)</td>
<td>2.01(0.26)</td>
<td>2.03(0.27)</td>
<td>3.68(0.52)</td>
</tr>
<tr>
<td>Neonatal seiz.</td>
<td>-1.94(0.42)</td>
<td>-1.12(0.26)</td>
<td>-1.16(0.26)</td>
<td>-2.06(0.44)</td>
</tr>
<tr>
<td>Congenital malf.</td>
<td>-1.21(0.31)</td>
<td>-0.61(0.18)</td>
<td>-0.62(0.18)</td>
<td>-1.17(0.33)</td>
</tr>
<tr>
<td>100× Bilirubin</td>
<td>-0.69(0.25)</td>
<td>-0.32(0.14)</td>
<td>-0.32(0.14)</td>
<td>-0.64(0.27)</td>
</tr>
<tr>
<td><strong>Second Ability Score</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>4.03(0.51)</td>
<td>2.19(0.27)</td>
<td>2.21(0.27)</td>
<td>4.01(0.54)</td>
</tr>
<tr>
<td>Neonatal seiz.</td>
<td>-2.26(0.43)</td>
<td>-1.27(0.26)</td>
<td>-1.29(0.26)</td>
<td>-2.28(0.44)</td>
</tr>
<tr>
<td>Congenital malf.</td>
<td>-1.08(0.32)</td>
<td>-0.56(0.19)</td>
<td>-0.59(0.19)</td>
<td>-1.11(0.34)</td>
</tr>
<tr>
<td>100× Bilirubin</td>
<td>-0.85(0.26)</td>
<td>-0.42(0.14)</td>
<td>-0.41(0.14)</td>
<td>-0.80(0.27)</td>
</tr>
<tr>
<td><strong>Third Ability Score</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>3.32(0.50)</td>
<td>1.84(0.27)</td>
<td>1.91(0.27)</td>
<td>3.49(0.54)</td>
</tr>
<tr>
<td>Neonatal seiz.</td>
<td>-1.55(0.44)</td>
<td>-0.88(0.27)</td>
<td>-0.93(0.27)</td>
<td>-1.70(0.46)</td>
</tr>
<tr>
<td>Congenital malf.</td>
<td>-0.96(0.32)</td>
<td>-0.47(0.19)</td>
<td>-0.49(0.19)</td>
<td>-0.96(0.35)</td>
</tr>
<tr>
<td>100× Bilirubin</td>
<td>-0.44(0.26)</td>
<td>-0.21(0.14)</td>
<td>-0.24(0.14)</td>
<td>-0.49(0.28)</td>
</tr>
</tbody>
</table>
## Association parameters

<table>
<thead>
<tr>
<th></th>
<th>Bahad</th>
<th>Probit</th>
<th>Dale-Norm</th>
<th>Dale-Logist</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1,2): $\rho$ or $\psi$</td>
<td>0.27(0.05)</td>
<td>0.73(0.05)</td>
<td>17.37(5.19)</td>
<td>17.35(5.19)</td>
</tr>
<tr>
<td>(1,2): $z(\rho)$ or $\ln \psi$</td>
<td>0.55(0.11)</td>
<td>1.85(0.23)</td>
<td>2.85(0.30)</td>
<td>2.85(0.30)</td>
</tr>
<tr>
<td>(1,3): $\rho$ or $\psi$</td>
<td>0.39(0.05)</td>
<td>0.81(0.04)</td>
<td>30.64(9.78)</td>
<td>30.61(9.78)</td>
</tr>
<tr>
<td>(1,3): $z(\rho)$ or $\ln \psi$</td>
<td>0.83(0.12)</td>
<td>2.27(0.25)</td>
<td>3.42(0.32)</td>
<td>3.42(0.32)</td>
</tr>
<tr>
<td>(2,3): $\rho$ or $\psi$</td>
<td>0.23(0.05)</td>
<td>0.72(0.05)</td>
<td>17.70(5.47)</td>
<td>17.65(5.47)</td>
</tr>
<tr>
<td>(2,3): $z(\rho)$ or $\ln \psi$</td>
<td>0.47(0.10)</td>
<td>1.83(0.23)</td>
<td>2.87(0.31)</td>
<td>2.87(0.31)</td>
</tr>
<tr>
<td>(1,2,3): $\rho$ or $\psi$</td>
<td>—</td>
<td>—</td>
<td>0.91(0.69)</td>
<td>0.92(0.69)</td>
</tr>
<tr>
<td>(1,2,3): $z(\rho)$ or $\ln \psi$</td>
<td>—</td>
<td>—</td>
<td>-0.09(0.76)</td>
<td>-0.09(0.76)</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-598.44</td>
<td>-570.69</td>
<td>-567.11</td>
<td>-567.09</td>
</tr>
</tbody>
</table>
Chapter 19
Generalized Estimating Equations

▷ General idea
▷ Asymptotic properties
▷ Working correlation
▷ Special case and application
▷ SAS code and output
19.1 General Idea

• Univariate GLM, score function of the form (scalar $Y_i$):

$$S(\beta) = \sum_{i=1}^{N} \frac{\partial \mu_i}{\partial \beta} v_i^{-1} (y_i - \mu_i) = 0 \quad \text{with} \quad v_i = \text{Var}(Y_i)$$

• In longitudinal setting: $Y = (Y_1, \ldots, Y_N)$:

$$S(\beta) = \sum_i \sum_j \frac{\partial \mu_{ij}}{\partial \beta} v_{ij}^{-1} (y_{ij} - \mu_{ij}) = \sum_{i=1}^{N} D_i' [V_i(\alpha)]^{-1} (y_i - \mu_i) = 0$$

where

- $D_i$ is an $n_i \times p$ matrix with $(i, j)$th elements $\frac{\partial \mu_{ij}}{\partial \beta}$
- $y_i$ and $\mu_i$ are $n_i$-vectors with elements $y_{ij}$ and $\mu_{ij}$
- Is $V_i$ $n_i \times n_i$ diagonal or more complex?
• $V_i = \text{Var}(Y_i)$ is more complex since it involves a set of nuisance parameters $\alpha$, determining the covariance structure of $Y_i$: 

$$V_i(\beta, \alpha) = \phi A_i^{1/2}(\beta) R_i(\alpha) A_i^{1/2}(\beta)$$

in which

$$A_i^{1/2}(\beta) = \begin{pmatrix} \sqrt{v_{i1}(\mu_{i1}(\beta))} & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \sqrt{v_{ini}(\mu_{ini}(\beta))} \end{pmatrix}$$

and $R_i(\alpha)$ is the correlation matrix of $Y_i$, parameterized by $\alpha$.

• Same form as for full likelihood procedure, **but** we restrict specification to the first moment only

• Liang and Zeger (1986)
19.2 Large Sample Properties

As \( N \to \infty \)

\[
\sqrt{N}(\hat{\beta} - \beta) \sim N(0, I_0^{-1})
\]

where

\[
I_0 = \sum_{i=1}^{N} D_i' [V_i(\alpha)]^{-1} D_i
\]

- (Unrealistic) Conditions:
  - \( \alpha \) is known
  - the parametric form for \( V_i(\alpha) \) is known

- This is the naive\( \equiv \)purely model based variance estimator

- Solution: working correlation matrix
19.3 Unknown Covariance Structure

Keep the score equations

\[ S(\beta) = \sum_{i=1}^{N} [D_i]' [V_i(\alpha)]^{-1} (y_i - \mu_i) = 0 \]

BUT

• suppose \( V_i(.) \) is not the true variance of \( Y_i \) but only a plausible guess, a so-called working correlation matrix

• specify correlations and not covariances, because the variances follow from the mean structure

• the score equations are solved as before
• The asymptotic normality results change to

$$\sqrt{N}(\hat{\beta} - \beta) \sim N(0, I_0^{-1}I_1I_0^{-1})$$

$$I_0 = \sum_{i=1}^{N} D_i'[\text{Var}Y_i(V_i(\alpha))]^{-1}D_i$$

$$I_1 = \sum_{i=1}^{N} D_i'[\text{Var}Y_i(V_i(\alpha))]^{-1}\text{Var}Y_i(V_i(\alpha))]^{-1}D_i.$$ 

• This is the robust≈empirically corrected≈ sandwich variance estimator
  ▶ $I_0$ is the bread
  ▶ $I_1$ is the filling (ham or cheese)

• Correct guess $\implies$ likelihood variance
• The estimators $\hat{\beta}$ are consistent even if the working correlation matrix is incorrect.

• An estimate is found by replacing the unknown variance matrix $\text{Var}(Y_i)$ by

$$
(Y_i - \hat{\mu}_i)(Y_i - \hat{\mu}_i)'.
$$

• Even if this estimator is bad for $\text{Var}(Y_i)$ it leads to a good estimate of $I_1$, provided that:
  - replication in the data is sufficiently large
  - same model for $\mu_i$ is fitted to groups of subjects
  - observation times do not vary too much between subjects

• A bad choice of working correlation matrix can affect the efficiency of $\hat{\beta}$

• Care needed with incomplete data (see Part V)
19.4 The Working Correlation Matrix

\[ V_i = V_i(\beta, \alpha, \phi) = \phi A_i^{1/2}(\beta) R_i(\alpha) A_i^{1/2}(\beta) \]

- **Variance function:** \( A_i \) is \((n_i \times n_i)\) diagonal with elements \( v(\mu_{ij}) \), the known GLM variance function.

- **Working correlation:** \( R_i(\alpha) \) possibly depends on a different set of parameters \( \alpha \).

- **Overdispersion parameter:** \( \phi \), assumed 1 or estimated from the data.

- **The unknown quantities are expressed in terms of the Pearson residuals**

\[ e_{ij} = \frac{y_{ij} - \mu_{ij}}{\sqrt{v(\mu_{ij})}}. \]

Note that \( e_{ij} \) depends on \( \beta \).
19.5 Estimation of Working Correlation

Liang and Zeger (1986) proposed moment-based estimates for the working correlation.

<table>
<thead>
<tr>
<th>Corr($Y_{ij}, Y_{ik}$)</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independence</td>
<td>0</td>
</tr>
<tr>
<td>Exchangeable</td>
<td>$\alpha$</td>
</tr>
<tr>
<td>AR(1)</td>
<td>$\alpha$</td>
</tr>
<tr>
<td>Unstructured</td>
<td>$\alpha_{jk}$</td>
</tr>
</tbody>
</table>

Dispersion parameter:

$$\hat{\phi} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{n_i} \sum_{j=1}^{n_i} e_{ij}^2.$$
19.6 Fitting GEE

The standard procedure, implemented in the SAS procedure GENMOD.

1. Compute initial estimates for $\beta$, using a univariate GLM (i.e., assuming independence).
2. ▶ Compute Pearson residuals $e_{ij}$.
   ▶ Compute estimates for $\alpha$ and $\phi$.
   ▶ Compute $R_i(\alpha)$ and $V_i(\beta, \alpha) = \phi A_i^{1/2}(\beta) R_i(\alpha) A_i^{1/2}(\beta)$.
3. Update estimate for $\beta$:
   $$\beta^{(t+1)} = \beta^{(t)} - \left[ \sum_{i=1}^{N} D_i' V_i^{-1} D_i \right]^{-1} \left[ \sum_{i=1}^{N} D_i' V_i^{-1} (y_i - \mu_i) \right].$$

Estimates of precision by means of $I_0^{-1}$ and/or $I_0^{-1} I_1 I_0^{-1}$. 
19.7 Special Case: Linear Mixed Models

- Estimate for $\beta$

$$\hat{\beta}(\alpha) = \left( \sum_{i=1}^{N} X_i'W_iX_i \right)^{-1} \sum_{i=1}^{N} X_i'W_iY_i$$

with $\alpha$ replaced by its ML or REML estimate

- Conditional on $\alpha$, $\hat{\beta}$ has mean

$$E[\hat{\beta}(\alpha)] = \left( \sum_{i=1}^{N} X_i'W_iX_i \right)^{-1} \sum_{i=1}^{N} X_i'W_iX_i\beta = \beta$$

provided that $E(Y_i) = X_i\beta$

- Hence, in order for $\hat{\beta}$ to be unbiased, it is sufficient that the mean of the response is correctly specified.
Conditional on $\alpha$, $\hat{\beta}$ has covariance

$$\text{Var}(\hat{\beta}) = \left( \sum_{i=1}^{N} X'_i W_i X_i \right)^{-1} \left( \sum_{i=1}^{N} X'_i W_i \text{Var}(Y_i) W_i X_i \right) \left( \sum_{i=1}^{N} X'_i W_i X_i \right)^{-1} = \left( \sum_{i=1}^{N} X'_i W_i X_i \right)^{-1}$$

Note that this model-based version assumes that the covariance matrix $\text{Var}(Y_i)$ is correctly modelled as $V_i = Z_i D Z'_i + \Sigma_i$.

An empirically corrected version is:

$$\text{Var}(\hat{\beta}) = \left( \sum_{i=1}^{N} X'_i W_i X_i \right)^{-1} \left( \sum_{i=1}^{N} X'_i W_i \text{Var}(Y_i) W_i X_i \right) \left( \sum_{i=1}^{N} X'_i W_i X_i \right)^{-1}$$

\[\text{BREAD} \quad \text{MEAT} \quad \text{BREAD}\]
Chapter 20
A Family of GEE Methods

▷ Classical approach
▷ Prentice’s two sets of GEE
▷ Linearization-based version
▷ GEE2
▷ Alternating logistic regressions
20.1 Prentice’s GEE

\[
\sum_{i=1}^{N} D_i V_i^{-1} (Y_i - \mu_i) = 0, \quad \sum_{i=1}^{N} E_i W_i^{-1} (Z_i - \delta_i) = 0
\]

where

\[
Z_{ijk} = \frac{(Y_{ij} - \mu_{ij})(Y_{ik} - \mu_{ik})}{\sqrt{\mu_{ij}(1 - \mu_{ij})\mu_{ik}(1 - \mu_{ik})}}, \quad \delta_{ijk} = E(Z_{ijk})
\]

The joint asymptotic distribution of \(\sqrt{N}(\hat{\beta} - \beta)\) and \(\sqrt{N}(\hat{\alpha} - \alpha)\) normal with variance-covariance matrix consistently estimated by

\[
N \begin{pmatrix}
A & 0 \\
B & C
\end{pmatrix}
\begin{pmatrix}
\Lambda_{11} & \Lambda_{12} \\
\Lambda_{21} & \Lambda_{22}
\end{pmatrix}
\begin{pmatrix}
A & B' \\
0 & C
\end{pmatrix}
\]
where

\[
A = \left( \sum_{i=1}^{N} D_i' V_i^{-1} D_i \right)^{-1}, \\
B = \left( \sum_{i=1}^{N} E_i' W_i^{-1} E_i \right)^{-1} \left( \sum_{i=1}^{N} E_i' W_i^{-1} \frac{\partial Z_i}{\partial \beta} \right) \left( \sum_{i=1}^{N} D_i' V_i^{-1} D_i \right)^{-1}, \\
C = \left( \sum_{i=1}^{N} E_i' W_i^{-1} E_i \right)^{-1},
\]

\[
\Lambda_{11} = \sum_{i=1}^{N} D_i' V_i^{-1} \text{Cov}(Y_i) V_i^{-1} D_i, \\
\Lambda_{12} = \sum_{i=1}^{N} D_i' V_i^{-1} \text{Cov}(Y_i, Z_i) W_i^{-1} E_i, \\
\Lambda_{21} = \Lambda_{12}, \\
\Lambda_{22} = \sum_{i=1}^{N} E_i' W_i^{-1} \text{Cov}(Z_i) W_i^{-1} E_i,
\]

and

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td>\text{Var}(Y_i)</td>
<td>( (Y_i - \mu_i)(Y_i - \mu_i)' )</td>
</tr>
<tr>
<td>\text{Cov}(Y_i, Z_i)</td>
<td>( (Y_i - \mu_i)(Z_i - \delta_i)' )</td>
</tr>
<tr>
<td>\text{Var}(Z_i)</td>
<td>( (Z_i - \delta_i)(Z_i - \delta_i)' )</td>
</tr>
</tbody>
</table>
20.2 GEE Based on Linearization

20.2.1 Model formulation

- Previous version of GEE are formulated directly in terms of binary outcomes

- This approach is based on a linearization:

\[ y_i = \mu_i + \epsilon_i \]

with

\[ \eta_i = g(\mu_i), \quad \eta_i = X_i\beta, \quad \text{Var}(y_i) = \text{Var}(\epsilon_i) = \Sigma_i. \]

- \( \eta_i \) is a vector of linear predictors,

- \( g(.) \) is the (vector) link function.
20.2.2 Estimation (Nelder and Wedderburn 1972)

- Solve iteratively:
  \[
  \sum_{i=1}^{N} X_i'W_iX_i\beta = \sum_{i=1}^{N} W_i y_i^*,
  \]
  where
  \[
  W_i = F_i \Sigma_i^{-1} F_i, \quad y_i^* = \hat{\eta}_i + (y_i - \hat{\mu}_i)F_i^{-1},
  \]
  \[
  F_i = \frac{\partial \mu_i}{\partial \eta_i}, \quad \Sigma_i = \text{Var}(\varepsilon), \quad \mu_i = E(y_i).
  \]

- Remarks:
  - $y_i^*$ is called ‘working variable’ or ‘pseudo data’.
  - Basis for SAS macro and procedure GLIMMIX
  - For linear models, $D_i = I_{n_i}$ and standard linear regression follows.
20.2.3 The Variance Structure

\[ \Sigma_i = \phi A_i^{1/2}(\beta) R_i(\alpha) A_i^{1/2}(\beta) \]

- \( \phi \) is a scale (overdispersion) parameter,

- \( A_i = v(\mu_i) \), expressing the mean-variance relation (this is a function of \( \beta \)),

- \( R_i(\alpha) \) describes the correlation structure:
  - If independence is assumed then \( R_i(\alpha) = I_{n_i} \).
  - Other structures, such as compound symmetry, AR(1),... can be assumed as well.
20.3 GEE2

- **Model:**
  - Marginal mean structure
  - Pairwise association:
    - Odds ratios
    - Correlations

- **Working assumptions:** Third and fourth moments

- **Estimation:**
  - Second-order estimating equations
  - Likelihood (assuming 3rd and 4th moments are correctly specified)
20.4 Alternating Logistic Regression

- Diggle, Heagerty, Liang, and Zeger (2002) and Molenberghs and Verbeke (2005)

- When marginal odds ratios are used to model association, $\alpha$ can be estimated using ALR, which is
  - almost as efficient as GEE2
  - almost as easy (computationally) than GEE1

- $\mu_{ijk}$ as before and $\alpha_{ijk} = \ln(\psi_{ijk})$ the marginal log odds ratio:

\[
\begin{align*}
\text{logit Pr}(Y_{ij} = 1|\mathbf{x}_{ij}) &= \mathbf{x}_{ij}\beta \\
\text{logit Pr}(Y_{ij} = 1|Y_{ik} = y_{ik}) &= \alpha_{ijk}y_{ik} + \ln \left( \frac{\mu_{ij} - \mu_{ijk}}{1 - \mu_{ij} - \mu_{ik} + \mu_{ijk}} \right)
\end{align*}
\]
• \( \alpha_{ijk} \) can be modelled in terms of predictors

• the second term is treated as an offset

• the estimating equations for \( \beta \) and \( \alpha \) are solved in turn, and the ‘alternating’ between both sets is repeated until convergence.

• this is needed because the offset clearly depends on \( \beta \).
20.5 Application to the Toenail Data

20.5.1 The model

- Consider the model:

\[ Y_{ij} \sim \text{Bernoulli}(\mu_{ij}), \quad \log \left( \frac{\mu_{ij}}{1 - \mu_{ij}} \right) = \beta_0 + \beta_1 T_i + \beta_2 t_{ij} + \beta_3 T_i t_{ij} \]

- \( Y_{ij} \): severe infection (yes/no) at occasion \( j \) for patient \( i \)

- \( t_{ij} \): measurement time for occasion \( j \)

- \( T_i \): treatment group
20.5.2 Standard GEE

- **SAS Code:**

```
proc genmod data=test descending;
  class idnum timeclss;
  model onyresp = treatn time treatn*time
      / dist=binomial;
  repeated subject=idnum / withinsubject=timeclss
      type=exch covb corrw modelse;
run;
```

- **SAS statements:**
  - The `REPEATED` statements defines the GEE character of the model.
  - `type=`: working correlation specification (UN, AR(1), EXCH, IND,...)
  - `modelse`: model-based s.e.’s on top of default empirically corrected s.e.’s
  - `corrw`: printout of working correlation matrix
  - `withinsubject=`: specification of the ordering within subjects
• **Selected output:**

  ▶ **Regression parameters:**

  Analysis Of Initial Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-0.5571</td>
<td>0.1090</td>
<td>-0.7708 -0.3433</td>
<td>26.10</td>
</tr>
<tr>
<td>treatn</td>
<td>1</td>
<td>0.0240</td>
<td>0.1565</td>
<td>-0.2827 0.3307</td>
<td>0.02</td>
</tr>
<tr>
<td>time</td>
<td>1</td>
<td>-0.1769</td>
<td>0.0246</td>
<td>-0.2251 -0.1288</td>
<td>51.91</td>
</tr>
<tr>
<td>treatn*time</td>
<td>1</td>
<td>-0.0783</td>
<td>0.0394</td>
<td>-0.1556 -0.0010</td>
<td>3.95</td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.0000</td>
<td>0.0000</td>
<td>1.0000 1.0000</td>
<td></td>
</tr>
</tbody>
</table>

  ▶ Estimates from fitting the model, ignoring the correlation structure, i.e., from fitting a classical GLM to the data, using proc GENMOD.

  ▶ The reported log-likelihood also corresponds to this model, and therefore should not be interpreted.

  ▶ The reported estimates are used as starting values in the iterative estimation procedure for fitting the GEE’s.
### Analysis Of GEE Parameter Estimates

#### Empirical Standard Error Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Error</th>
<th>95% Confidence Limits</th>
<th>Z</th>
<th>Pr &gt;</th>
<th>Z</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.5840</td>
<td>0.1734</td>
<td>-0.9238 -0.2441</td>
<td>-3.37</td>
<td>0.0008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>treatn</td>
<td>0.0120</td>
<td>0.2613</td>
<td>-0.5001 0.5241</td>
<td>0.05</td>
<td>0.9633</td>
<td></td>
<td></td>
</tr>
<tr>
<td>time</td>
<td>-0.1770</td>
<td>0.0311</td>
<td>-0.2380 -0.1161</td>
<td>-5.69</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>treatn*time</td>
<td>-0.0886</td>
<td>0.0571</td>
<td>-0.2006 0.0233</td>
<td>-1.55</td>
<td>0.1208</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Analysis Of GEE Parameter Estimates

#### Model-Based Standard Error Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Error</th>
<th>95% Confidence Limits</th>
<th>Z</th>
<th>Pr &gt;</th>
<th>Z</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.5840</td>
<td>0.1344</td>
<td>-0.8475 -0.3204</td>
<td>-4.34</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>treatn</td>
<td>0.0120</td>
<td>0.1866</td>
<td>-0.3537 0.3777</td>
<td>0.06</td>
<td>0.9486</td>
<td></td>
<td></td>
</tr>
<tr>
<td>time</td>
<td>-0.1770</td>
<td>0.0209</td>
<td>-0.2180 -0.1361</td>
<td>-8.47</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>treatn*time</td>
<td>-0.0886</td>
<td>0.0362</td>
<td>-0.1596 0.0177</td>
<td>-2.45</td>
<td>0.0143</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

▷ **The working correlation:**

#### Exchangeable Working Correlation

| Correlation | 0.420259237 |
20.5.3 Alternating Logistic Regression

- ‘type=exch’ → ‘logor=exch’

- Note that $\alpha$ now is a genuine parameter
### Analysis Of GEE Parameter Estimates
#### Empirical Standard Error Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Error</th>
<th>95% Confidence Limits</th>
<th>Z</th>
<th>Pr &gt;</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.5244</td>
<td>0.1686</td>
<td>-0.8548</td>
<td>-0.1940</td>
<td>-3.11</td>
<td>0.0019</td>
</tr>
<tr>
<td>treatn</td>
<td>0.0168</td>
<td>0.2432</td>
<td>-0.4599</td>
<td>0.4935</td>
<td>0.07</td>
<td>0.9448</td>
</tr>
<tr>
<td>time</td>
<td>-0.1781</td>
<td>0.0296</td>
<td>-0.2361</td>
<td>-0.1200</td>
<td>-6.01</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>treatn*time</td>
<td>-0.0837</td>
<td>0.0520</td>
<td>-0.1856</td>
<td>0.0182</td>
<td>-1.61</td>
<td>0.1076</td>
</tr>
<tr>
<td>Alpha1</td>
<td>3.2218</td>
<td>0.2908</td>
<td>2.6519</td>
<td>3.7917</td>
<td>11.08</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

### Analysis Of GEE Parameter Estimates
#### Model-Based Standard Error Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Error</th>
<th>95% Confidence Limits</th>
<th>Z</th>
<th>Pr &gt;</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.5244</td>
<td>0.1567</td>
<td>-0.8315</td>
<td>-0.2173</td>
<td>-3.35</td>
<td>0.0008</td>
</tr>
<tr>
<td>treatn</td>
<td>0.0168</td>
<td>0.2220</td>
<td>-0.4182</td>
<td>0.4519</td>
<td>0.08</td>
<td>0.9395</td>
</tr>
<tr>
<td>time</td>
<td>-0.1781</td>
<td>0.0233</td>
<td>-0.2238</td>
<td>-0.1323</td>
<td>-7.63</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>treatn*time</td>
<td>-0.0837</td>
<td>0.0392</td>
<td>-0.1606</td>
<td>-0.0068</td>
<td>-2.13</td>
<td>0.0329</td>
</tr>
</tbody>
</table>
20.5.4 Linearization Based Method

- **GLIMMIX macro:**

```sas
%glimmix(data=test, procopt=%str(method=ml empirical),
         stmts=%str(
             class idnum timeclss;
             model onyresp = treatn time treatn*time / solution;
             repeated timeclss / subject=idnum type=cs rcorr;
         ),
         error=binomial,
         link=logit);
```

- **GLIMMIX procedure:**

```sas
proc glimmix data=test method=RSPL empirical;
   class idnum;
   model onyresp (event='1') = treatn time treatn*time
       / dist=binary solution;
   random _residual_ / subject=idnum type=cs;
run;
```
• Both produce the same results

• The GLIMMIX macro is a MIXED core, with GLM-type surrounding statements

• The GLIMMIX procedure does not call MIXED, it has its own engine

• PROC GLIMMIX combines elements of MIXED and of GENMOD

• RANDOM _residual_ is the PROC GLIMMIX way to specify residual correlation
### 20.5.5 Results of Models Fitted to Toenail Data

<table>
<thead>
<tr>
<th>Effect</th>
<th>Par.</th>
<th>IND</th>
<th>EXCH</th>
<th>UN</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEE1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Int.</td>
<td>$\beta_0$</td>
<td>-0.557(0.109;0.171)</td>
<td>-0.584(0.134;0.173)</td>
<td>-0.720(0.166;0.173)</td>
</tr>
<tr>
<td>$T_i$</td>
<td>$\beta_1$</td>
<td>0.024(0.157;0.251)</td>
<td>0.012(0.187;0.261)</td>
<td>0.072(0.235;0.246)</td>
</tr>
<tr>
<td>$t_{ij}$</td>
<td>$\beta_2$</td>
<td>-0.177(0.025;0.030)</td>
<td>-0.177(0.021;0.031)</td>
<td>-0.141(0.028;0.029)</td>
</tr>
<tr>
<td>$T_i \cdot t_{ij}$</td>
<td>$\beta_3$</td>
<td>-0.078(0.039;0.055)</td>
<td>-0.089(0.036;0.057)</td>
<td>-0.114(0.047;0.052)</td>
</tr>
<tr>
<td>ALR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Int.</td>
<td>$\beta_0$</td>
<td>0.524(0.157;0.169)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T_i$</td>
<td>$\beta_1$</td>
<td>0.017(0.222;0.243)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_{ij}$</td>
<td>$\beta_2$</td>
<td>-0.178(0.023;0.030)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T_i \cdot t_{ij}$</td>
<td>$\beta_3$</td>
<td>-0.084(0.039;0.052)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ass.</td>
<td>$\alpha$</td>
<td>3.222(       )</td>
<td>0.291</td>
<td></td>
</tr>
</tbody>
</table>

Linearization based method

<table>
<thead>
<tr>
<th>Effect</th>
<th>Par.</th>
<th>IND</th>
<th>EXCH</th>
<th>UN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Int.</td>
<td>$\beta_0$</td>
<td>-0.557(0.112;0.171)</td>
<td>-0.585(0.142;0.174)</td>
<td>-0.630(0.171;0.172)</td>
</tr>
<tr>
<td>$T_i$</td>
<td>$\beta_1$</td>
<td>0.024(0.160;0.251)</td>
<td>0.011(0.196;0.262)</td>
<td>0.036(0.242;0.242)</td>
</tr>
<tr>
<td>$t_{ij}$</td>
<td>$\beta_2$</td>
<td>-0.177(0.025;0.030)</td>
<td>-0.177(0.022;0.031)</td>
<td>-0.204(0.038;0.034)</td>
</tr>
<tr>
<td>$T_i \cdot t_{ij}$</td>
<td>$\beta_3$</td>
<td>-0.078(0.040;0.055)</td>
<td>-0.089(0.038;0.057)</td>
<td>-0.106(0.058;0.058)</td>
</tr>
</tbody>
</table>

estimate (model-based s.e.; empirical s.e.)
20.5.6 Discussion

• GEE1: All empirical standard errors are correct, but the efficiency is higher for the more complex working correlation structure, as seen in p-values for $T_i \cdot t_{ij}$ effect:

<table>
<thead>
<tr>
<th>Structure</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>IND</td>
<td>0.1515</td>
</tr>
<tr>
<td>EXCH</td>
<td>0.1208</td>
</tr>
<tr>
<td>UN</td>
<td>0.0275</td>
</tr>
</tbody>
</table>

Thus, opting for reasonably adequate correlation assumptions still pays off, in spite of the fact that all are consistent and asymptotically normal.

• Similar conclusions for linearization-based method
• Model-based s.e. and empirically corrected s.e. in reasonable agreement for UN

• Typically, the model-based standard errors are much too small as they are based on the assumption that all observations in the data set are independent, hereby overestimating the amount of available information, hence also overestimating the precision of the estimates.

• ALR: similar inferences but now also $\alpha$ part of the inferences
Part III

Generalized Linear Mixed Models for Non-Gaussian Longitudinal Data
Chapter 21
The Beta-binomial Model

▷ Genesis of the model
▷ Implied marginal distribution
21.1 Genesis of the Beta-binomial Model

- Skellam (1948), Kleinman (1973)

- Let $Y_i$ be a $n_i$-dimensional vector of Bernoulli-distributed outcomes, with success probability $b_i$.

- Assume the elements in $Y_i$ to be independent, conditionally on $b_i$.

- Then, the conditional density of $Y_i$, given $b_i$ is proportional to the density of

$$Z_i = \sum_{j=1}^{n_i} Y_{ij}$$

- The density of $Z_i$, given $b_i$ is binomial with $n_i$ trials and success probability $b_i$. 
• The beta-binomial model assumes the $b_i$ to come from a beta distribution with parameters $\alpha$ and $\beta$:

$$f(b_i|\alpha, \beta) = \frac{b_i^{\alpha-1}(1 - b_i)^{\beta-1}}{B(\alpha, \beta)}$$

$B(., .)$: the beta function

• $\alpha$ and $\beta$ can depend on covariates, but this dependence is temporarily dropped from notation
21.2 Implied Marginal Model

- The marginal density of $Z_i$ is the so-called beta-binomial density:

$$f_i(z_i|\alpha, \beta) = \int \binom{n_i}{z_i} b_i^{z_i} (1 - b_i)^{n_i - z_i} f(b_i|\alpha, \beta) \, db_i$$

$$= \binom{n_i}{z_i} \frac{B(z_i + \alpha, n_i - z_i + \beta)}{B(\alpha, \beta)}$$
• Useful moments and relationships ($\pi = \mu_i/n_i$):

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha = \pi(\rho^{-1} - 1)$</td>
</tr>
<tr>
<td></td>
<td>$\beta = (1 - \pi)(\rho^{-1} - 1)$</td>
</tr>
<tr>
<td>Mean</td>
<td>$\mu_i = \mathbb{E}(Z_i) = n_i \frac{\alpha}{\alpha + \beta}$</td>
</tr>
<tr>
<td>Correlation</td>
<td>$\rho = \text{Corr}(Y_{ij}, Y_{ik}) = \frac{1}{\alpha + \beta + 1}$</td>
</tr>
<tr>
<td>Variance</td>
<td>$\text{Var}(Z_i) = n_i \pi(1 - \pi)[1 + (n_i - 1)\rho]$</td>
</tr>
</tbody>
</table>
• The density can now be written as:

\[ f_i(z_i | \pi, \rho) = \binom{n_i}{z_i} \frac{B[z_i + \pi(\rho^{-1} - 1), n_i - z_i + (1 - \pi)(\rho^{-1} - 1)]}{B[\pi(\rho^{-1} - 1), (1 - \pi)(\rho^{-1} - 1)]} \]

• When there are covariates (e.g., sub-populations, dose groups), rewrite \( \pi \) and/or \( \rho \) as \( \pi_i \) and/or \( \rho_i \), respectively.

• It is then easy to formulate a model through the marginal parameters \( \pi_i \) and \( \rho_i \):
  \( \pi_i \) can be modeled through, e.g., a logit link
  \( \rho_i \) can be modeled through, e.g., Fisher’s \( z \) transformation

• In Part IV, the NTP data will be analyzed using the beta-binomial model
Introduction: LMM Revisited

Generalized Linear Mixed Models (GLMM)

Fitting Algorithms

Example
22.1 Introduction: LMM Revisited

• We re-consider the linear mixed model:

\[ Y_i | b_i \sim N(X_i \beta + Z_i b_i, \Sigma_i), \quad b_i \sim N(0, D) \]

• The implied marginal model equals \( Y_i \sim N(X_i \beta, Z_i D Z'_i + \Sigma_i) \)

• Hence, even under conditional independence, i.e., all \( \Sigma_i \) equal to \( \sigma^2 I_{n_i} \), a marginal association structure is implied through the random effects.

• The same ideas can now be applied in the context of GLM’s to model association between discrete repeated measures.
22.2 Generalized Linear Mixed Models (GLMM)

- Given a vector \( b_i \) of random effects for cluster \( i \), it is assumed that all responses \( Y_{ij} \) are independent, with density

\[
f(y_{ij}|\theta_{ij}, \phi) = \exp \left\{ \phi^{-1}[y_{ij}\theta_{ij} - \psi(\theta_{ij})] + c(y_{ij}, \phi) \right\}
\]

- \( \theta_{ij} \) is now modelled as \( \theta_{ij} = x_{ij}'\beta + z_{ij}'b_i \)

- As before, it is assumed that \( b_i \sim N(0, D) \)

- Let \( f_{ij}(y_{ij}|b_i, \beta, \phi) \) denote the conditional density of \( Y_{ij} \) given \( b_i \), the conditional density of \( Y_i \) equals

\[
f_i(y_i|b_i, \beta, \phi) = \prod_{j=1}^{n_i} f_{ij}(y_{ij}|b_i, \beta, \phi)
\]
• The marginal distribution of $Y_i$ is given by

$$f_i(y_i|\beta, D, \phi) = \int f_i(y_i|b_i, \beta, \phi) f(b_i|D) \, db_i$$

$$= \int \prod_{j=1}^{n_i} f_{ij}(y_{ij}|b_i, \beta, \phi) f(b_i|D) \, db_i$$

where $f(b_i|D)$ is the density of the $N(0, D)$ distribution.

• The likelihood function for $\beta$, $D$, and $\phi$ now equals

$$L(\beta, D, \phi) = \prod_{i=1}^{N} f_i(y_i|\beta, D, \phi)$$

$$= \prod_{i=1}^{N} \int \prod_{j=1}^{n_i} f_{ij}(y_{ij}|b_i, \beta, \phi) f(b_i|D) \, db_i$$
Under the normal linear model, the integral can be worked out analytically.

In general, approximations are required:
- Approximation of integrand
- Approximation of data
- Approximation of integral

Predictions of random effects can be based on the posterior distribution

\[ f(b_i|Y_i = y_i) \]

‘Empirical Bayes (EB) estimate’:
Posterior mode, with unknown parameters replaced by their MLE
22.3 Laplace Approximation of Integrand

- Integrals in $L(\beta, D, \phi)$ can be written in the form $I = \int e^{Q(b)} db$

- Second-order Taylor expansion of $Q(b)$ around the mode yields
  \[ Q(b) \approx Q(\hat{b}) + \frac{1}{2}(b - \hat{b})'Q''(\hat{b})(b - \hat{b}), \]

- Quadratic term leads to re-scaled normal density. Hence,
  \[ I \approx (2\pi)^{q/2} \left| -Q''(\hat{b}) \right|^{-1/2} e^{Q(\hat{b})}. \]

- Exact approximation in case of normal kernels

- Good approximation in case of many repeated measures per subject
22.4 Approximation of Data

22.4.1 General Idea

- Re-write GLMM as:

\[ Y_{ij} = \mu_{ij} + \varepsilon_{ij} = h(x_{ij}' \beta + z_{ij}' b_i) + \varepsilon_{ij} \]

with variance for errors equal to \( \text{Var}(Y_{ij}|b_i) = \phi(v(\mu_{ij})) \)

- Linear Taylor expansion for \( \mu_{ij} \):
  - Penalized quasi-likelihood (PQL): Around current \( \hat{\beta} \) and \( \hat{b}_i \)
  - Marginal quasi-likelihood (MQL): Around current \( \hat{\beta} \) and \( b_i = 0 \)
22.4.2 Penalized quasi-likelihood (PQL)

- Linear Taylor expansion around current $\hat{\beta}$ and $\hat{b}_i$:

\[
Y_{ij} \approx h(x'_{ij}\hat{\beta} + z'_{ij}\hat{b}_i) + h'(x'_{ij}\hat{\beta} + z'_{ij}\hat{b}_i)x'_{ij}(\beta - \hat{\beta}) + h'(x'_{ij}\hat{\beta} + z'_{ij}\hat{b}_i)z'_{ij}(b_i - \hat{b}_i) + \varepsilon_{ij}
\]

\[
\approx \hat{\mu}_{ij} + v(\hat{\mu}_{ij})x'_{ij}(\beta - \hat{\beta}) + v(\hat{\mu}_{ij})z'_{ij}(b_i - \hat{b}_i) + \varepsilon_{ij}
\]

- In vector notation: $Y_i \approx \hat{\mu}_i + \hat{V}_iX_i(\beta - \hat{\beta}) + \hat{V}_iZ_i(b_i - \hat{b}_i) + \varepsilon_i$

- Re-ordering terms yields:

\[
Y^*_i \equiv \hat{V}_i^{-1}(Y_i - \hat{\mu}_i) + X_i\hat{\beta} + Z_i\hat{b}_i \approx X_i\beta + Z_i\hat{b}_i + \varepsilon^*_i,
\]

- Model fitting by iterating between updating the pseudo responses $Y^*_i$ and fitting the above linear mixed model to them.
22.4.3 Marginal quasi-likelihood (MQL)

- Linear Taylor expansion around current $\hat{\beta}$ and $b_i = 0$:

  $$Y_{ij} \approx h(x'_{ij}\hat{\beta}) + h'(x'_{ij}\hat{\beta})x'_{ij}(\beta - \hat{\beta}) + h'(x'_{ij}\hat{\beta})z'_ij b_i + \epsilon_{ij}$$

  $$\approx \hat{\mu}_{ij} + v(\hat{\mu}_{ij})x'_{ij}(\beta - \hat{\beta}) + v(\hat{\mu}_{ij})z'ij b_i + \epsilon_{ij}$$

- In vector notation: $Y_i \approx \hat{\mu}_i + V_iX_i(\beta - \hat{\beta}) + V_iZ_ib_i + \epsilon_i$

- Re-ordering terms yields:

  $$Y_i^* \equiv V_i^{-1}(Y_i - \hat{\mu}_i) + X_i\hat{\beta} \approx X_i\beta + Z_ib_i + \epsilon_i^*$$

- Model fitting by iterating between updating the pseudo responses $Y_i^*$ and fitting the above linear mixed model to them.
22.4.4 PQL versus MQL

- MQL only performs reasonably well if random-effects variance is (very) small
- Both perform bad for binary outcomes with few repeated measurements per cluster
- With increasing number of measurements per subject:
  - MQL remains biased
  - PQL consistent
- Improvements possible with higher-order Taylor expansions
22.5 Approximation of Integral

- The likelihood contribution of every subject is of the form
  \[ \int f(z)\phi(z)dz \]
  where \( \phi(z) \) is the density of the (multivariate) normal distribution

- Gaussian quadrature methods replace the integral by a weighted sum:
  \[ \int f(z)\phi(z)dz \approx \sum_{q=1}^{Q} w_q f(z_q) \]
  \( Q \) is the order of the approximation. The higher \( Q \) the more accurate the approximation will be
• The nodes (or quadrature points) \( z_q \) are solutions to the \( Q \)th order Hermite polynomial

• The \( w_q \) are well-chosen weights

• The nodes \( z_q \) and weights \( w_q \) are reported in tables. Alternatively, an algorithm is available for calculating all \( z_q \) and \( w_q \) for any value \( Q \).

• With **Gaussian quadrature**, the nodes and weights are fixed, independent of \( f(z)\phi(z) \).

• With **adaptive Gaussian quadrature**, the nodes and weights are adapted to the ‘support’ of \( f(z)\phi(z) \).
Graphically ($Q = 10$):
• Typically, adaptive Gaussian quadrature needs (much) less quadrature points than classical Gaussian quadrature.

• On the other hand, adaptive Gaussian quadrature is much more time consuming.

• Adaptive Gaussian quadrature of order one is equivalent to Laplace transformation.

• Ample detail can be found in Molenberghs and Verbeke (2005, Sections 14.3–14.5)
22.6 Example: Toenail Data

- $Y_{ij}$ is binary severity indicator for subject $i$ at visit $j$.

- Model:

$$Y_{ij} | b_i \sim \text{Bernoulli}(\pi_{ij}), \quad \log \left( \frac{\pi_{ij}}{1 - \pi_{ij}} \right) = \beta_0 + b_i + \beta_1 T_i + \beta_2 t_{ij} + \beta_3 T_i t_{ij}$$

- Notation:
  - $T_i$: treatment indicator for subject $i$
  - $t_{ij}$: time point at which $j$th measurement is taken for $i$th subject

- Adaptive as well as non-adaptive Gaussian quadrature, for various $Q$. 
Results:

<table>
<thead>
<tr>
<th></th>
<th>Gaussian quadrature</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$Q = 3$</td>
<td>$Q = 5$</td>
<td>$Q = 10$</td>
<td>$Q = 20$</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>-1.52 (0.31)</td>
<td>-2.49 (0.39)</td>
<td>-0.99 (0.32)</td>
<td>-1.54 (0.69)</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.39 (0.38)</td>
<td>0.19 (0.36)</td>
<td>0.47 (0.36)</td>
<td>-0.43 (0.80)</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>-0.32 (0.03)</td>
<td>-0.38 (0.04)</td>
<td>-0.38 (0.05)</td>
<td>-0.40 (0.05)</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>-0.09 (0.05)</td>
<td>-0.12 (0.07)</td>
<td>-0.15 (0.07)</td>
<td>-0.14 (0.07)</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>2.26 (0.12)</td>
<td>3.09 (0.21)</td>
<td>4.53 (0.39)</td>
<td>3.86 (0.33)</td>
</tr>
<tr>
<td>$-2\ell$</td>
<td>1344.1</td>
<td>1259.6</td>
<td>1254.4</td>
<td>1249.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Adaptive Gaussian quadrature</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$Q = 3$</td>
<td>$Q = 5$</td>
<td>$Q = 10$</td>
<td>$Q = 20$</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>-2.05 (0.59)</td>
<td>-1.47 (0.40)</td>
<td>-1.65 (0.45)</td>
<td>-1.63 (0.43)</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.16 (0.64)</td>
<td>-0.09 (0.54)</td>
<td>-0.12 (0.59)</td>
<td>-0.11 (0.59)</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>-0.42 (0.05)</td>
<td>-0.40 (0.04)</td>
<td>-0.41 (0.05)</td>
<td>-0.40 (0.05)</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>-0.17 (0.07)</td>
<td>-0.16 (0.07)</td>
<td>-0.16 (0.07)</td>
<td>-0.16 (0.07)</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>4.51 (0.62)</td>
<td>3.70 (0.34)</td>
<td>4.07 (0.43)</td>
<td>4.01 (0.38)</td>
</tr>
<tr>
<td>$-2\ell$</td>
<td>1259.1</td>
<td>1257.1</td>
<td>1248.2</td>
<td>1247.8</td>
</tr>
</tbody>
</table>
Conclusions:

- (Log-)likelihoods are not comparable
- Different $Q$ can lead to considerable differences in estimates and standard errors
- For example, using non-adaptive quadrature, with $Q = 3$, we found no difference in time effect between both treatment groups ($t = -0.09/0.05, p = 0.0833$).
- Using adaptive quadrature, with $Q = 50$, we find a significant interaction between the time effect and the treatment ($t = -0.16/0.07, p = 0.0255$).
- Assuming that $Q = 50$ is sufficient, the ‘final’ results are well approximated with smaller $Q$ under adaptive quadrature, but not under non-adaptive quadrature.
Comparison of fitting algorithms:
- Adaptive Gaussian Quadrature, $Q = 50$
- MQL and PQL

Summary of results:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>QUAD</th>
<th>PQL</th>
<th>MQL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept group A</td>
<td>-1.63 (0.44)</td>
<td>-0.72 (0.24)</td>
<td>-0.56 (0.17)</td>
</tr>
<tr>
<td>Intercept group B</td>
<td>-1.75 (0.45)</td>
<td>-0.72 (0.24)</td>
<td>-0.53 (0.17)</td>
</tr>
<tr>
<td>Slope group A</td>
<td>-0.40 (0.05)</td>
<td>-0.29 (0.03)</td>
<td>-0.17 (0.02)</td>
</tr>
<tr>
<td>Slope group B</td>
<td>-0.57 (0.06)</td>
<td>-0.40 (0.04)</td>
<td>-0.26 (0.03)</td>
</tr>
<tr>
<td>Var. random intercepts ($\tau^2$)</td>
<td>15.99 (3.02)</td>
<td>4.71 (0.60)</td>
<td>2.49 (0.29)</td>
</tr>
</tbody>
</table>

Severe differences between QUAD (gold standard?) and MQL/PQL.

MQL/PQL may yield (very) biased results, especially for binary data.
Chapter 23
Fitting GLMM’s in SAS

▷ Proc GLIMMIX for PQL and MQL
▷ Proc NLMIXED for Gaussian quadrature
23.1 Procedure GLIMMIX for PQL and MQL

- Re-consider logistic model with random intercepts for toenail data

- SAS code (PQL):

  ```sas
  proc glimmix data=test method=RSPL ;
  class idnum;
  model onyresp (event='1') = treatn time treatn*time
       / dist=binary solution;
  random intercept / subject=idnum;
  run;
  ```

- MQL obtained with option ‘method=RMPL’

- Inclusion of random slopes:

  ```sas
  random intercept time / subject=idnum type=un;
  ```
Selected SAS output (PQL):

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Subject</th>
<th>Estimate</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>idnum</td>
<td>4.7095</td>
<td>0.6024</td>
</tr>
</tbody>
</table>

Solutions for Fixed Effects

| Effect      | Estimate | Error  | DF  | t Value | Pr > |t| |
|-------------|----------|--------|-----|---------|------|---|
| Intercept   | -0.7204  | 0.2370 | 292 | -3.04   | 0.0026 |
| treatn      | -0.02594 | 0.3360 | 1612| -0.08   | 0.9385 |
| time        | -0.2782  | 0.03222| 1612| -8.64   | <.0001 |
| treatn*time | -0.09583 | 0.05105| 1612| -1.88   | 0.0607 |
23.2 Procedure NLMIXED for Gaussian Quadrature

- Re-consider logistic model with random intercepts for toenail data

- SAS program (non-adaptive, \( Q = 3 \)):

```sas
proc nlmixed data=test noad qpoints=3;
parms beta0=-1.6 beta1=0 beta2=-0.4 beta3=-0.5 sigma=3.9;
teta = beta0 + b + beta1*treatn + beta2*time + beta3*timetr;
expteta = exp(teta);
p = expteta/(1+expteta);
model onyresp ~ binary(p);
random b ~ normal(0,sigma**2) subject=idnum;
run;
```

- Adaptive Gaussian quadrature obtained by omitting option ‘noad’
• Automatic search for ‘optimal’ value of $Q$ in case of no option ‘qpoints=’

• Selected SAS output (non-adaptive, $Q = 3$):

| Parameter | Estimate | Error  | DF  | t Value | Pr > |t| | Alpha | Lower | Upper | Gradient |
|-----------|----------|--------|-----|---------|-------|---|-------|-------|-------|----------|
| beta0     | -1.5311  | 0.2961 | 293 | -5.17   | <.0001| 0.05 | -2.1139| -0.9483|       | 2.879E-7 |
| beta1     | -0.4294  | 0.3728 | 293 | -1.15   | 0.2503| 0.05 | -1.1631| 0.3043|       | -2.11E-6 |
| beta2     | -0.3107  | 0.03373| 293 | -9.21   | <.0001| 0.05 | -0.3771| -0.2443|       | -0.00003|
| beta3     | -0.07539 | 0.04998| 293 | -1.51   | 0.1325| 0.05 | -0.1738| 0.02298|       | -0.00003|
| sigma     | 2.2681   | 0.1220 | 293 | 18.58   | <.0001| 0.05 | 2.0279 | 2.5083|       | -3.6E-6  |

• Good starting values needed!
The inclusion of random slopes can be specified as follows:

```plaintext
proc nlmixed data=test noad qpoints=3;
parms beta0=-1.6 beta1=0 beta2=-0.4 beta3=-0.5
d11=3.9 d12=0 d22=0.1;
teta = beta0 + b1 + beta1*treatn + beta2*time
    + b2*time + beta3*timetr;
expteta = exp(teta);
p = expteta/(1+expteta);
model onyresp ~ binary(p);
random b1 b2 ~ normal([0, 0] , [d11, d12, d22])
    subject=idnum;
run;
```
23.2.1 Some Comments on the NLMIXED Procedure

- Different optimization algorithms are available to carry out the maximization of the likelihood.

- Constraints on parameters are also allowed in the optimization process.

- The conditional distribution (given the random effects) can be specified as Normal, Binomial, Poisson, or as any distribution for which you can specify the likelihood by programming statements.

- E-B estimates of the random effects can be obtained.

- Only one RANDOM statement can be specified.

- Only normal random effects are allowed.
• Does not calculate automatic initial values.

• Make sure your data set is sorted by cluster ID!

• PROC NLMIXED can perform Gaussian quadrature by using the options NOAD and NOADSCALE. The number of quadrature points can be specified with the option QPOINTS=m.

• PROC NLMIXED can maximize the marginal likelihood using the Newton-Raphson algorithm by specifying the option TECHNIQUE=NEWRAP.
23.2.2 The Main Statements

- NLMIXED statement:
  - option ‘noad’ to request no adaptive quadrature
  - by default, adaptive Gaussian quadrature is used
  - the option ‘qpoints’ specifies the number of quadrature points
  - by default, the number of quadrature points is selected adaptively by evaluating the log-likelihood function at the starting values of the parameters until two successive evaluations show sufficiently small relative change.

- PARMS statement:
  - starting values for all parameters in the model
  - by default, parameters not listed in the PARMS statement are given an initial value of 1
- **MODEL statement:**
  - conditional distribution of the data, given the random effects
  - valid distributions:
    - normal$(m,v)$: Normal with mean $m$ and variance $v$
    - binary$(p)$: Bernoullie with probability $p$
    - binomial$(n,p)$: Binomial with count $n$ and probability $p$
    - poisson$(m)$: Poisson with mean $m$
    - general$(ll)$: General model with log-likelihood $ll$
  - since no factors can be defined, explicit creation of dummies is required

- **RANDOM statement:**
  - specification of the random effects
  - the procedure requires the data to be ordered by subject!
  - empirical Bayes estimates can be obtained by adding `out=eb`
Part IV

Marginal Versus Random-effects Models and Case Studies
Chapter 24
Marginal Versus Random-effects Models

▷ Interpretation of GLMM parameters
▷ Marginalization of GLMM
▷ Conclusion
24.1 Interpretation of GLMM Parameters: Toenail Data

- We compare our GLMM results for the toenail data with those from fitting GEE’s (unstructured working correlation):

<table>
<thead>
<tr>
<th>Parameter</th>
<th>GLMM</th>
<th>GEE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept group A</td>
<td>$-1.6308$ (0.4356)</td>
<td>$-0.7219$ (0.1656)</td>
</tr>
<tr>
<td>Intercept group B</td>
<td>$-1.7454$ (0.4478)</td>
<td>$-0.6493$ (0.1671)</td>
</tr>
<tr>
<td>Slope group A</td>
<td>$-0.4043$ (0.0460)</td>
<td>$-0.1409$ (0.0277)</td>
</tr>
<tr>
<td>Slope group B</td>
<td>$-0.5657$ (0.0601)</td>
<td>$-0.2548$ (0.0380)</td>
</tr>
</tbody>
</table>
The strong differences can be explained as follows:

- Consider the following GLMM:

  \[ Y_{ij} | b_i \sim \text{Bernoulli}(\pi_{ij}), \quad \log \left( \frac{\pi_{ij}}{1 - \pi_{ij}} \right) = \beta_0 + b_i + \beta_1 t_{ij} \]

- The conditional means \( E(Y_{ij} | b_i) \), as functions of \( t_{ij} \), are given by

\[
E(Y_{ij} | b_i) = \frac{\exp(\beta_0 + b_i + \beta_1 t_{ij})}{1 + \exp(\beta_0 + b_i + \beta_1 t_{ij})}
\]
The marginal average evolution is now obtained from averaging over the random effects:

\[
E(Y_{ij}) = E[E(Y_{ij} | b_i)] = E \left[ \frac{\exp(\beta_0 + b_i + \beta_1 t_{ij})}{1 + \exp(\beta_0 + b_i + \beta_1 t_{ij})} \right]
\]

\[
\ne \neq \frac{\exp(\beta_0 + \beta_1 t_{ij})}{1 + \exp(\beta_0 + \beta_1 t_{ij})}
\]
• Hence, the parameter vector $\beta$ in the GEE model needs to be interpreted completely different from the parameter vector $\beta$ in the GLMM:
  
  ▶ GEE: marginal interpretation
  ▶ GLMM: conditional interpretation, conditionally upon level of random effects

• In general, the model for the marginal average is not of the same parametric form as the conditional average in the GLMM.

• For logistic mixed models, with normally distributed random random intercepts, it can be shown that the marginal model can be well approximated by again a logistic model, but with parameters approximately satisfying

$$\frac{\hat{\beta}^{\text{RE}}}{\hat{\beta}^{\text{M}}} = \sqrt{c^2 \sigma^2 + 1} > 1, \quad \sigma^2 = \text{variance random intercepts}$$

$$c = \frac{16\sqrt{3}}{(15\pi)}$$
• For the toenail application, $\sigma$ was estimated as 4.0164, such that the ratio equals $\sqrt{c^2\sigma^2 + 1} = 2.5649$.

• The ratio’s between the GLMM and GEE estimates are:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>GLMM Estimate (s.e.)</th>
<th>GEE Estimate (s.e.)</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept group A</td>
<td>−1.6308 (0.4356)</td>
<td>−0.7219 (0.1656)</td>
<td>2.2590</td>
</tr>
<tr>
<td>Intercept group B</td>
<td>−1.7454 (0.4478)</td>
<td>−0.6493 (0.1671)</td>
<td>2.6881</td>
</tr>
<tr>
<td>Slope group A</td>
<td>−0.4043 (0.0460)</td>
<td>−0.1409 (0.0277)</td>
<td>2.8694</td>
</tr>
<tr>
<td>Slope group B</td>
<td>−0.5657 (0.0601)</td>
<td>−0.2548 (0.0380)</td>
<td>2.2202</td>
</tr>
</tbody>
</table>

• Note that this problem does not occur in linear mixed models:
  ▶ Conditional mean: $E(Y_i|b_i) = X_i\beta + Z_i b_i$
  ▶ Specifically: $E(Y_i|b_i = 0) = X_i\beta$
  ▶ Marginal mean: $E(Y_i) = X_i\beta$
The problem arises from the fact that, in general,

\[ E[g(Y)] \neq g[E(Y)] \]

So, whenever the random effects enter the conditional mean in a non-linear way, the regression parameters in the marginal model need to be interpreted differently from the regression parameters in the mixed model.

In practice, the marginal mean can be derived from the GLMM output by integrating out the random effects.

This can be done numerically via Gaussian quadrature, or based on sampling methods.
24.2 Marginalization of GLMM: Toenail Data

- As an example, we plot the average evolutions based on the GLMM output obtained in the toenail example:

\[
P(Y_{ij} = 1) = \begin{cases} 
E \left[ \frac{\exp(-1.6308 + b_i - 0.4043t_{ij})}{1 + \exp(-1.6308 + b_i - 0.4043t_{ij})} \right], \\
E \left[ \frac{\exp(-1.7454 + b_i - 0.5657t_{ij})}{1 + \exp(-1.7454 + b_i - 0.5657t_{ij})} \right],
\end{cases}
\]
SAS code (averaging over 1000 draws):

```sas
data h;
  do treat=0 to 1 by 1;
    do subject=1 to 1000 by 1;
      b=4.0164*rannor(-1);
      do t=0 to 12 by 0.1;
        if treat=0 then y=exp(-1.6308 + b -0.4043*t)/(1+ exp(-1.6308 + b -0.4043*t));
        else y=exp(-1.7454 + b -0.5657*t)/(1+ exp(-1.7454 + b -0.5657*t));
        output;
      end;
    end;
  end;
proc sort data=h;
  by t treat;run;
proc means data=h;
  var y;
  by t treat;
  output out=out;
run;
proc gplot data=out;
  plot y*t=treat / haxis=axis1 vaxis=axis2 legend=legend1;
  axis1 label=(h=2 'Time') value=(h=1.5) order=(0 to 14 by 1) minor=none;
  axis2 label=(h=2 A=90 'P(Y=1)') value=(h=1.5) order=(0 to 0.4 by 0.1) minor=none;
  legend1 label=(h=1.5 'Treatment: ') value=(h=1.5 'A' 'B');
title h=2.5 'Marginal average evolutions (GLMM)';
symbol1 c=black i=join w=5 l=1 mode=include;
symbol2 c=black i=join w=5 l=2 mode=include;
where _stat_='MEAN';
run;quit;run;
```
• Average evolutions obtained from the GEE analyses:

\[ P(Y_{ij} = 1) = \begin{cases} \frac{\exp(-0.7219 - 0.1409t_{ij})}{1 + \exp(-0.7219 - 0.1409t_{ij})} \\ \frac{\exp(-0.6493 - 0.2548t_{ij})}{1 + \exp(-0.6493 - 0.2548t_{ij})} \end{cases} \]
• In a GLMM context, rather than plotting the marginal averages, one can also plot the profile for an ‘average’ subject, i.e., a subject with random effect $b_i = 0$:

$$P(Y_{ij} = 1|b_i = 0) = \begin{cases} 
\frac{\exp(-1.6308 - 0.4043t_{ij})}{1 + \exp(-1.6308 - 0.4043t_{ij})} \\
\frac{\exp(-1.7454 - 0.5657t_{ij})}{1 + \exp(-1.7454 - 0.5657t_{ij})}
\end{cases}$$
24.3 Example: Toenail Data Revisited

- Overview of all analyses on toenail data:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>QUAD</th>
<th>PQL</th>
<th>MQL</th>
<th>GEE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept group A</td>
<td>−1.63 (0.44)</td>
<td>−0.72 (0.24)</td>
<td>−0.56 (0.17)</td>
<td>−0.72 (0.17)</td>
</tr>
<tr>
<td>Intercept group B</td>
<td>−1.75 (0.45)</td>
<td>−0.72 (0.24)</td>
<td>−0.53 (0.17)</td>
<td>−0.65 (0.17)</td>
</tr>
<tr>
<td>Slope group A</td>
<td>−0.40 (0.05)</td>
<td>−0.29 (0.03)</td>
<td>−0.17 (0.02)</td>
<td>−0.14 (0.03)</td>
</tr>
<tr>
<td>Slope group B</td>
<td>−0.57 (0.06)</td>
<td>−0.40 (0.04)</td>
<td>−0.26 (0.03)</td>
<td>−0.25 (0.04)</td>
</tr>
<tr>
<td>Var. random intercepts ($\tau^2$)</td>
<td>15.99 (3.02)</td>
<td>4.71 (0.60)</td>
<td>2.49 (0.29)</td>
<td></td>
</tr>
</tbody>
</table>

- Conclusion:

\[ |\text{GEE}| < |\text{MQL}| < |\text{PQL}| < |\text{QUAD}| \]
Model Family

marginal random-effects model

inference inference

likelihood GEE marginal hierarchical

\( \beta^M \) \( \beta^M \) \( \beta^{RE} \) \( (\beta^{RE}, b_i) \)

\( \beta^M \)
Chapter 25
Case Study: The NTP Data

▷ Research question
▷ Conditional model
▷ Bahadur model
▷ GEE1 analyses
▷ GEE2 analysis
▷ Alternating logistic regressions
▷ Beta-binomial model
▷ Generalized linear mixed model
▷ Discussion
25.1 Research Question

- Dose-response relationship: effect of dose on malformations

- Regression relationship:

\[
\text{logit}[P(Y_{ij} = 1 | d_i, \ldots)] = \beta_0 + \beta_d d_i
\]

- Association parameter: \(\beta_d\) Precise meaning is model-dependent:
  - Transformed conditional odds ratio
  - Transformed correlation
  - Transformed marginal odds ratio
25.2 Conditional Model

- **Regression relationship:**
  \[
  \text{logit}[P(Y_{ij} = 1|d_i, Y_{ik} = 0, k \neq j)] = \beta_0 + \beta_d d_i
  \]

- \(\delta_i = \beta_a\) is conditional log odds ratio

- Quadratic loglinear model

- Maximum likelihood estimates (model based standard errors; empirically corrected standard errors)
<table>
<thead>
<tr>
<th>Outcome</th>
<th>Par.</th>
<th>DEHP</th>
<th>EG</th>
<th>DYME</th>
</tr>
</thead>
<tbody>
<tr>
<td>External</td>
<td>$\beta_0$</td>
<td>-2.81(0.58;0.52)</td>
<td>-3.01(0.79;1.01)</td>
<td>-5.78(1.13;1.23)</td>
</tr>
<tr>
<td></td>
<td>$\beta_d$</td>
<td>3.07(0.65;0.62)</td>
<td>2.25(0.68;0.85)</td>
<td>6.25(1.25;1.41)</td>
</tr>
<tr>
<td></td>
<td>$\beta_a$</td>
<td>0.18(0.04;0.04)</td>
<td>0.25(0.05;0.06)</td>
<td>0.09(0.06;0.06)</td>
</tr>
<tr>
<td>Visceral</td>
<td>$\beta_0$</td>
<td>-2.39(0.50;0.52)</td>
<td>-5.09(1.55;1.51)</td>
<td>-3.32(0.98;0.89)</td>
</tr>
<tr>
<td></td>
<td>$\beta_d$</td>
<td>2.45(0.55;0.60)</td>
<td>3.76(1.34;1.20)</td>
<td>2.88(0.93;0.83)</td>
</tr>
<tr>
<td></td>
<td>$\beta_a$</td>
<td>0.18(0.04;0.04)</td>
<td>0.23(0.09;0.09)</td>
<td>0.29(0.05;0.05)</td>
</tr>
<tr>
<td>Skeletal</td>
<td>$\beta_0$</td>
<td>-2.79(0.58;0.77)</td>
<td>-0.84(0.17;0.18)</td>
<td>-1.62(0.35;0.48)</td>
</tr>
<tr>
<td></td>
<td>$\beta_d$</td>
<td>2.91(0.63;0.82)</td>
<td>0.98(0.20;0.20)</td>
<td>2.45(0.51;0.82)</td>
</tr>
<tr>
<td></td>
<td>$\beta_a$</td>
<td>0.17(0.04;0.05)</td>
<td>0.20(0.02;0.02)</td>
<td>0.25(0.03;0.03)</td>
</tr>
<tr>
<td>Collapsed</td>
<td>$\beta_0$</td>
<td>-2.04(0.35;0.42)</td>
<td>-0.81(0.16;0.16)</td>
<td>-2.90(0.43;0.51)</td>
</tr>
<tr>
<td></td>
<td>$\beta_d$</td>
<td>2.98(0.51;0.66)</td>
<td>0.97(0.20;0.20)</td>
<td>5.08(0.74;0.96)</td>
</tr>
<tr>
<td></td>
<td>$\beta_a$</td>
<td>0.16(0.03;0.03)</td>
<td>0.20(0.02;0.02)</td>
<td>0.19(0.03;0.03)</td>
</tr>
</tbody>
</table>
25.3 The Bahadur Model

- Regression relationship:

\[
\text{logit}[P(Y_{ij} = 1|d_i)] = \beta_0 + \beta_d d_i
\]

- \(\beta_a\): Fisher’s \(z\) transformed correlation

- \(\rho\): correlation
<table>
<thead>
<tr>
<th>Outcome</th>
<th>Parameter</th>
<th>DEHP</th>
<th>EG</th>
<th>DYME</th>
</tr>
</thead>
<tbody>
<tr>
<td>External</td>
<td>$\beta_0$</td>
<td>-4.93(0.39)</td>
<td>-5.25(0.66)</td>
<td>-7.25(0.71)</td>
</tr>
<tr>
<td></td>
<td>$\beta_d$</td>
<td>5.15(0.56)</td>
<td>2.63(0.76)</td>
<td>7.94(0.77)</td>
</tr>
<tr>
<td></td>
<td>$\beta_a$</td>
<td>0.11(0.03)</td>
<td>0.12(0.03)</td>
<td>0.11(0.04)</td>
</tr>
<tr>
<td></td>
<td>$\rho$</td>
<td>0.05(0.01)</td>
<td>0.06(0.01)</td>
<td>0.05(0.02)</td>
</tr>
<tr>
<td>Visceral</td>
<td>$\beta_0$</td>
<td>-4.42(0.33)</td>
<td>-7.38(1.30)</td>
<td>-6.89(0.81)</td>
</tr>
<tr>
<td></td>
<td>$\beta_d$</td>
<td>4.38(0.49)</td>
<td>4.25(1.39)</td>
<td>5.49(0.87)</td>
</tr>
<tr>
<td></td>
<td>$\beta_a$</td>
<td>0.11(0.02)</td>
<td>0.05(0.08)</td>
<td>0.08(0.04)</td>
</tr>
<tr>
<td></td>
<td>$\rho$</td>
<td>0.05(0.01)</td>
<td>0.02(0.04)</td>
<td>0.04(0.02)</td>
</tr>
<tr>
<td>Skeletal</td>
<td>$\beta_0$</td>
<td>-4.67(0.39)</td>
<td>-2.49(0.11)</td>
<td>-4.27(0.61)</td>
</tr>
<tr>
<td></td>
<td>$\beta_d$</td>
<td>4.68(0.56)</td>
<td>2.96(0.18)</td>
<td>5.79(0.80)</td>
</tr>
<tr>
<td></td>
<td>$\beta_a$</td>
<td>0.13(0.03)</td>
<td>0.27(0.02)</td>
<td>0.22(0.05)</td>
</tr>
<tr>
<td></td>
<td>$\rho$</td>
<td>0.06(0.01)</td>
<td>0.13(0.01)</td>
<td>0.11(0.02)</td>
</tr>
<tr>
<td>Collapsed</td>
<td>$\beta_0$</td>
<td>-3.83(0.27)</td>
<td>-2.51(0.09)</td>
<td>-5.31(0.40)</td>
</tr>
<tr>
<td></td>
<td>$\beta_d$</td>
<td>5.38(0.47)</td>
<td>3.05(0.17)</td>
<td>8.18(0.69)</td>
</tr>
<tr>
<td></td>
<td>$\beta_a$</td>
<td>0.12(0.03)</td>
<td>0.28(0.02)</td>
<td>0.12(0.03)</td>
</tr>
<tr>
<td></td>
<td>$\rho$</td>
<td>0.06(0.01)</td>
<td>0.14(0.01)</td>
<td>0.06(0.01)</td>
</tr>
</tbody>
</table>
25.4 GEE1

- Regression relationship:
  \[ \text{logit}[P(Y_{ij} = 1|d_i)] = \beta_0 + \beta_d d_i \]

- \( \phi \): overdispersion parameter

- \( \rho \): working correlation

- Parameter estimates (model-based standard errors; empirically corrected standard errors)

- Two sets of working assumptions:
  - Independence working assumptions
  - Exchangeable working assumptions
<table>
<thead>
<tr>
<th>Outcome</th>
<th>Par.</th>
<th>Standard</th>
<th>Prentice</th>
<th>Linearized</th>
</tr>
</thead>
<tbody>
<tr>
<td>External</td>
<td>$\beta_0$</td>
<td>$-5.06(0.30;0.38)$</td>
<td>$-5.06(0.33;0.38)$</td>
<td>$-5.06(0.28;0.38)$</td>
</tr>
<tr>
<td></td>
<td>$\beta_d$</td>
<td>$5.31(0.44;0.57)$</td>
<td>$5.31(0.48;0.57)$</td>
<td>$5.31(0.42;0.57)$</td>
</tr>
<tr>
<td></td>
<td>$\phi$</td>
<td>0.90</td>
<td></td>
<td>0.74</td>
</tr>
<tr>
<td>Visceral</td>
<td>$\beta_0$</td>
<td>$-4.47(0.28;0.36)$</td>
<td>$-4.47(0.28;0.36)$</td>
<td>$-4.47(0.28;0.36)$</td>
</tr>
<tr>
<td></td>
<td>$\beta_d$</td>
<td>$4.40(0.43;0.58)$</td>
<td>$4.40(0.43;0.58)$</td>
<td>$4.40(0.43;0.58)$</td>
</tr>
<tr>
<td></td>
<td>$\phi$</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
</tr>
<tr>
<td>Skeletal</td>
<td>$\beta_0$</td>
<td>$-4.87(0.31;0.47)$</td>
<td>$-4.87(0.31;0.47)$</td>
<td>$-4.87(0.32;0.47)$</td>
</tr>
<tr>
<td></td>
<td>$\beta_d$</td>
<td>$4.89(0.46;0.65)$</td>
<td>$4.90(0.47;0.65)$</td>
<td>$4.90(0.47;0.65)$</td>
</tr>
<tr>
<td></td>
<td>$\phi$</td>
<td>0.99</td>
<td></td>
<td>1.02</td>
</tr>
<tr>
<td>Collapsed</td>
<td>$\beta_0$</td>
<td>$-3.98(0.22;0.30)$</td>
<td>$-3.98(0.22;0.30)$</td>
<td>$-3.98(0.22;0.30)$</td>
</tr>
<tr>
<td></td>
<td>$\beta_d$</td>
<td>$5.56(0.40;0.61)$</td>
<td>$5.56(0.40;0.61)$</td>
<td>$5.56(0.41;0.61)$</td>
</tr>
<tr>
<td></td>
<td>$\phi$</td>
<td>0.99</td>
<td></td>
<td>1.04</td>
</tr>
<tr>
<td>Outcome</td>
<td>Par.</td>
<td>Standard</td>
<td>Prentice</td>
<td>Linearized</td>
</tr>
<tr>
<td>--------------</td>
<td>------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>External</td>
<td>$\beta_0$</td>
<td>-4.98(0.40;0.37)</td>
<td>-4.99(0.46;0.37)</td>
<td>-5.00(0.36;0.37)</td>
</tr>
<tr>
<td></td>
<td>$\beta_d$</td>
<td>5.33(0.57;0.55)</td>
<td>5.32(0.65;0.55)</td>
<td>5.32(0.51;0.55)</td>
</tr>
<tr>
<td></td>
<td>$\phi$</td>
<td>0.88</td>
<td>0.65</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\rho$</td>
<td>0.11</td>
<td>0.11(0.04)</td>
<td>0.06</td>
</tr>
<tr>
<td>Visceral</td>
<td>$\beta_0$</td>
<td>-4.50(0.37;0.37)</td>
<td>-4.51(0.40;0.37)</td>
<td>-4.50(0.36;0.37)</td>
</tr>
<tr>
<td></td>
<td>$\beta_d$</td>
<td>4.55(0.55;0.59)</td>
<td>4.59(0.58;0.59)</td>
<td>4.55(0.54;0.59)</td>
</tr>
<tr>
<td></td>
<td>$\phi$</td>
<td>1.00</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\rho$</td>
<td>0.08</td>
<td>0.11(0.05)</td>
<td>0.08</td>
</tr>
<tr>
<td>Skeletal</td>
<td>$\beta_0$</td>
<td>-4.83(0.44;0.45)</td>
<td>-4.82(0.47;0.44)</td>
<td>-4.82(0.46;0.45)</td>
</tr>
<tr>
<td></td>
<td>$\beta_d$</td>
<td>4.84(0.62;0.63)</td>
<td>4.84(0.67;0.63)</td>
<td>4.84(0.65;0.63)</td>
</tr>
<tr>
<td></td>
<td>$\phi$</td>
<td>0.98</td>
<td>0.86</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\rho$</td>
<td>0.12</td>
<td>0.14(0.06)</td>
<td>0.13</td>
</tr>
<tr>
<td>Collapsed</td>
<td>$\beta_0$</td>
<td>-4.05(0.32;0.31)</td>
<td>-4.06(0.35;0.31)</td>
<td>-4.04(0.33;0.31)</td>
</tr>
<tr>
<td></td>
<td>$\beta_d$</td>
<td>5.84(0.57;0.61)</td>
<td>5.89(0.62;0.61)</td>
<td>5.82(0.58;0.61)</td>
</tr>
<tr>
<td></td>
<td>$\phi$</td>
<td>1.00</td>
<td>0.96</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\rho$</td>
<td>0.11</td>
<td>0.15(0.05)</td>
<td>0.11</td>
</tr>
</tbody>
</table>
• Regression relationship:

\[
\text{logit}[P(Y_{ij} = 1|d_i)] = \beta_0 + \beta_d d_i
\]

• \(\beta_a\): Fisher’s \(z\) transformed correlation

• \(\rho\): correlation

• Working assumption: third- and fourth-order correlations are zero

• Parameter estimates (empirically corrected standard errors)
<table>
<thead>
<tr>
<th>Outcome</th>
<th>Parameter</th>
<th>DEHP</th>
<th>EG</th>
<th>DYME</th>
</tr>
</thead>
<tbody>
<tr>
<td>External</td>
<td>$\beta_0$</td>
<td>-4.98(0.37)</td>
<td>-5.63(0.67)</td>
<td>-7.45(0.73)</td>
</tr>
<tr>
<td></td>
<td>$\beta_d$</td>
<td>5.29(0.55)</td>
<td>3.10(0.81)</td>
<td>8.15(0.83)</td>
</tr>
<tr>
<td></td>
<td>$\beta_a$</td>
<td>0.15(0.05)</td>
<td>0.15(0.05)</td>
<td>0.13(0.05)</td>
</tr>
<tr>
<td></td>
<td>$\rho$</td>
<td>0.07(0.02)</td>
<td>0.07(0.02)</td>
<td>0.06(0.02)</td>
</tr>
<tr>
<td>Visceral</td>
<td>$\beta_0$</td>
<td>-4.49(0.36)</td>
<td>-7.50(1.05)</td>
<td>-6.89(0.75)</td>
</tr>
<tr>
<td></td>
<td>$\beta_d$</td>
<td>4.52(0.59)</td>
<td>4.37(1.14)</td>
<td>5.51(0.89)</td>
</tr>
<tr>
<td></td>
<td>$\beta_a$</td>
<td>0.15(0.06)</td>
<td>0.02(0.02)</td>
<td>0.11(0.07)</td>
</tr>
<tr>
<td></td>
<td>$\rho$</td>
<td>0.07(0.03)</td>
<td>0.01(0.01)</td>
<td>0.05(0.03)</td>
</tr>
<tr>
<td>Skeletal</td>
<td>$\beta_0$</td>
<td>-5.23(0.40)</td>
<td>-4.05(0.33)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta_d$</td>
<td>5.35(0.60)</td>
<td>4.77(0.43)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta_a$</td>
<td>0.18(0.02)</td>
<td>0.30(0.03)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\rho$</td>
<td>0.09(0.01)</td>
<td>0.15(0.01)</td>
<td></td>
</tr>
<tr>
<td>Collapsed</td>
<td>$\beta_0$</td>
<td>-5.23(0.40)</td>
<td>-4.07(0.71)</td>
<td>-5.75(0.48)</td>
</tr>
<tr>
<td></td>
<td>$\beta_d$</td>
<td>5.35(0.60)</td>
<td>4.89(0.90)</td>
<td>8.82(0.91)</td>
</tr>
<tr>
<td></td>
<td>$\beta_a$</td>
<td>0.18(0.02)</td>
<td>0.26(0.14)</td>
<td>0.18(0.12)</td>
</tr>
<tr>
<td></td>
<td>$\rho$</td>
<td>0.09(0.01)</td>
<td>0.13(0.07)</td>
<td>0.09(0.06)</td>
</tr>
</tbody>
</table>
25.6 Alternating Logistic Regressions

• Regression relationship:

\[ \logit[P(Y_{ij} = 1|d_i)] = \beta_0 + \beta_d d_i \]

• Exchangeable association structure

• \( \alpha \): log odds ratio

• \( \psi \): odds ratio

• Parameter estimates (empirically corrected standard errors)
<table>
<thead>
<tr>
<th>Outcome</th>
<th>Parameter</th>
<th>DEHP</th>
<th>EG</th>
<th>DYME</th>
</tr>
</thead>
<tbody>
<tr>
<td>External</td>
<td>$\beta_0$</td>
<td>-5.16(0.35)</td>
<td>-5.72(0.64)</td>
<td>-7.48(0.75)</td>
</tr>
<tr>
<td></td>
<td>$\beta_d$</td>
<td>5.64(0.52)</td>
<td>3.28(0.72)</td>
<td>8.25(0.87)</td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>0.96(0.30)</td>
<td>1.45(0.45)</td>
<td>0.79(0.31)</td>
</tr>
<tr>
<td></td>
<td>$\psi$</td>
<td>2.61(0.78)</td>
<td>4.26(1.92)</td>
<td>2.20(0.68)</td>
</tr>
<tr>
<td>Visceral</td>
<td>$\beta_0$</td>
<td>-4.54(0.36)</td>
<td>-7.61(1.06)</td>
<td>-7.24(0.88)</td>
</tr>
<tr>
<td></td>
<td>$\beta_d$</td>
<td>4.72(0.57)</td>
<td>4.50(1.13)</td>
<td>6.05(1.04)</td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>1.12(0.30)</td>
<td>0.49(0.42)</td>
<td>1.76(0.59)</td>
</tr>
<tr>
<td></td>
<td>$\psi$</td>
<td>3.06(0.92)</td>
<td>1.63(0.69)</td>
<td>5.81(3.43)</td>
</tr>
<tr>
<td>Skeletal</td>
<td>$\beta_0$</td>
<td>-4.87(0.49)</td>
<td>-3.28(0.22)</td>
<td>-4.92(0.34)</td>
</tr>
<tr>
<td></td>
<td>$\beta_d$</td>
<td>4.90(0.70)</td>
<td>3.85(0.39)</td>
<td>6.73(0.65)</td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>1.05(0.40)</td>
<td>1.43(0.22)</td>
<td>1.62(0.37)</td>
</tr>
<tr>
<td></td>
<td>$\psi$</td>
<td>2.86(1.14)</td>
<td>4.18(0.92)</td>
<td>5.05(1.87)</td>
</tr>
<tr>
<td>Collapsed</td>
<td>$\beta_0$</td>
<td>-4.04(0.31)</td>
<td>-3.19(0.22)</td>
<td>-5.08(0.37)</td>
</tr>
<tr>
<td></td>
<td>$\beta_d$</td>
<td>5.93(0.63)</td>
<td>3.86(0.40)</td>
<td>7.98(0.75)</td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>1.17(0.29)</td>
<td>1.40(0.22)</td>
<td>1.26(0.31)</td>
</tr>
<tr>
<td></td>
<td>$\psi$</td>
<td>3.22(0.93)</td>
<td>4.06(0.89)</td>
<td>3.53(1.09)</td>
</tr>
</tbody>
</table>
25.7 Beta-binomial Model

- Regression relationship:
  \[
  \text{logit}[P(Y_{ij} = 1|d_i)] = \beta_0 + \beta_d d_i
  \]

- \(\beta_a\): Fisher’s \(z\) transformed correlation

- \(\rho\): correlation

- Parameter estimates (standard errors)
<table>
<thead>
<tr>
<th>Outcome</th>
<th>Parameter</th>
<th>DEHP</th>
<th>EG</th>
<th>DYME</th>
</tr>
</thead>
<tbody>
<tr>
<td>External</td>
<td>$\beta_0$</td>
<td>-4.91(0.42)</td>
<td>-5.32(0.71)</td>
<td>-7.27(0.74)</td>
</tr>
<tr>
<td></td>
<td>$\beta_d$</td>
<td>5.20(0.59)</td>
<td>2.78(0.81)</td>
<td>8.01(0.82)</td>
</tr>
<tr>
<td></td>
<td>$\beta_a$</td>
<td>0.21(0.09)</td>
<td>0.28(0.14)</td>
<td>0.21(0.12)</td>
</tr>
<tr>
<td></td>
<td>$\rho$</td>
<td>0.10(0.04)</td>
<td>0.14(0.07)</td>
<td>0.10(0.06)</td>
</tr>
<tr>
<td>Visceral</td>
<td>$\beta_0$</td>
<td>-4.38(0.36)</td>
<td>-7.45(1.17)</td>
<td>-6.21(0.83)</td>
</tr>
<tr>
<td></td>
<td>$\beta_d$</td>
<td>4.42(0.54)</td>
<td>4.33(1.26)</td>
<td>4.94(0.90)</td>
</tr>
<tr>
<td></td>
<td>$\beta_a$</td>
<td>0.22(0.09)</td>
<td>0.04(0.09)</td>
<td>0.45(0.21)</td>
</tr>
<tr>
<td></td>
<td>$\rho$</td>
<td>0.11(0.04)</td>
<td>0.02(0.04)</td>
<td>0.22(0.10)</td>
</tr>
<tr>
<td>Skeletal</td>
<td>$\beta_0$</td>
<td>-4.88(0.44)</td>
<td>-2.89(0.27)</td>
<td>-5.15(0.47)</td>
</tr>
<tr>
<td></td>
<td>$\beta_d$</td>
<td>4.92(0.63)</td>
<td>3.42(0.40)</td>
<td>6.99(0.71)</td>
</tr>
<tr>
<td></td>
<td>$\beta_a$</td>
<td>0.27(0.11)</td>
<td>0.54(0.09)</td>
<td>0.61(0.14)</td>
</tr>
<tr>
<td></td>
<td>$\rho$</td>
<td>0.13(0.05)</td>
<td>0.26(0.04)</td>
<td>0.30(0.06)</td>
</tr>
<tr>
<td>Collapsed</td>
<td>$\beta_0$</td>
<td>-3.83(0.31)</td>
<td>-2.51(0.09)</td>
<td>-5.42(0.45)</td>
</tr>
<tr>
<td></td>
<td>$\beta_d$</td>
<td>5.59(0.56)</td>
<td>3.05(0.17)</td>
<td>8.29(0.79)</td>
</tr>
<tr>
<td></td>
<td>$\beta_a$</td>
<td>0.32(0.10)</td>
<td>0.28(0.02)</td>
<td>0.33(0.10)</td>
</tr>
<tr>
<td></td>
<td>$\rho$</td>
<td>0.16(0.05)</td>
<td>0.14(0.01)</td>
<td>0.16(0.05)</td>
</tr>
</tbody>
</table>
25.8 Generalized Linear Mixed Model

- Regression relationship:
  \[
  \text{logit}[P(Y_{ij} = 1|d_i, b_i)] = \beta_0 + b_i + \beta_d d_i, \quad b_i \sim N(0, \tau^2)
  \]

- External malformation in DEHP study

- Four ways of dealing with the integral: Laplace, adaptive Gaussian quadrature, PQL, and MQL

- Two versions of PQL and MQL: ML and REML

- Parameter estimates (standard errors)
<table>
<thead>
<tr>
<th>Effect</th>
<th>Parameter</th>
<th>Laplace</th>
<th>QUAD</th>
<th>PQL (REML)</th>
<th>PQL (ML)</th>
<th>MQL (REML)</th>
<th>MQL (ML)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>$\beta_0$</td>
<td>-6.02 (0.59)</td>
<td>-5.97 (0.57)</td>
<td>-5.32 (0.40)</td>
<td>-5.30 (0.40)</td>
<td>-5.18 (0.40)</td>
<td>-5.17 (0.39)</td>
</tr>
<tr>
<td>Dose effect</td>
<td>$\beta_d$</td>
<td>6.50 (0.86)</td>
<td>6.45 (0.84)</td>
<td>5.73 (0.65)</td>
<td>5.71 (0.64)</td>
<td>5.70 (0.66)</td>
<td>5.67 (0.65)</td>
</tr>
<tr>
<td>Intercept var.</td>
<td>$\tau^2$</td>
<td>1.42 (0.70)</td>
<td>1.27 (0.62)</td>
<td>0.95 (0.40)</td>
<td>0.89 (0.38)</td>
<td>1.20 (0.53)</td>
<td>1.10 (0.50)</td>
</tr>
</tbody>
</table>
25.9 Summary Table

- External malformation in DEHP study

- All conditional, marginal, and random-effects models considered

- Parameter estimates (standard errors)

- For non-likelihood methods, the empirically corrected standard errors are reported
<table>
<thead>
<tr>
<th>Family</th>
<th>Model</th>
<th>$\beta_0$</th>
<th>$\beta_d$</th>
<th>Association</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Conditional</strong></td>
<td>Quadr. loglin. (ML)</td>
<td>-2.81(0.58)</td>
<td>3.07(0.65)</td>
<td>LOG OR 0.18(0.04)</td>
</tr>
<tr>
<td></td>
<td>Quadr. loglin. (PL)</td>
<td>-2.85(0.53)</td>
<td>3.24(0.60)</td>
<td>LOG OR 0.18(0.04)</td>
</tr>
<tr>
<td><strong>Marginal</strong></td>
<td>Lik. Bahadur</td>
<td>-4.93(0.39)</td>
<td>5.15(0.56)</td>
<td>$\rho$ 0.05(0.01)</td>
</tr>
<tr>
<td></td>
<td>St. GEE1 (exch)</td>
<td>-4.98(0.37)</td>
<td>5.33(0.55)</td>
<td>$\rho$ 0.11</td>
</tr>
<tr>
<td></td>
<td>St. GEE1 (ind)</td>
<td>-5.06(0.38)</td>
<td>5.31(0.57)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Prent. GEE1 (exch)</td>
<td>-4.99(0.37)</td>
<td>5.32(0.55)</td>
<td>$\rho$ 0.11 (0.04)</td>
</tr>
<tr>
<td></td>
<td>Prent. GEE1 (ind)</td>
<td>-5.06(0.38)</td>
<td>5.31(0.57)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lin. based (exch)</td>
<td>-5.00(0.37)</td>
<td>5.32(0.55)</td>
<td>$\rho$ 0.06</td>
</tr>
<tr>
<td></td>
<td>Lin. based (ind)</td>
<td>-5.06(0.38)</td>
<td>5.31(0.57)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GEE2</td>
<td>-4.98(0.37)</td>
<td>5.29(0.55)</td>
<td>$\rho$ 0.07(0.02)</td>
</tr>
<tr>
<td></td>
<td>ALR</td>
<td>-.516(0.35)</td>
<td>5.64(0.52)</td>
<td>$\beta_a$ 0.96(0.30)</td>
</tr>
<tr>
<td><strong>Random-effects</strong></td>
<td>Beta-binomial</td>
<td>-4.91(0.42)</td>
<td>5.20(0.59)</td>
<td>$\rho$ 0.10(0.04)</td>
</tr>
<tr>
<td></td>
<td>GLLM (MQL)</td>
<td>-5.18(0.40)</td>
<td>5.70(0.66)</td>
<td>Int. var $\tau^2$ 1.20(0.53)</td>
</tr>
<tr>
<td></td>
<td>GLMM (PQL)</td>
<td>-5.32(0.40)</td>
<td>5.73(0.65)</td>
<td>Int. var $\tau^2$ 0.95(0.40)</td>
</tr>
<tr>
<td></td>
<td>GLMM (QUAD)</td>
<td>-5.97(0.57)</td>
<td>6.45(0.84)</td>
<td>Int. var $\tau^2$ 1.27(0.62)</td>
</tr>
</tbody>
</table>
25.10 Discussion

• Relationship between regression model parameters:

  \[ |\text{conditional}| < |\text{marginal}| < |\text{random-effects}| \]

• Beta-binomial model behaves like a marginal model (similar to the linear mixed model)

• Marginal model parameters:
  ▶ Mean function parameters: very similar
  ▶ Correlation parameters:

  \[ |\text{Bahadur}| < |\text{GEE2}| < |\text{GEE1}| < |\text{beta-binomial}| \]
Reason: strength of constraints:

* Bahadur model valid if all higher order probabilities are valid
* GEE2 valid if probabilities of orders 1, 2, 3, and 4 are valid
* GEE1 valid if probabilities of orders 1 and 2 are valid
* beta-binomial model is unconstrained of correlations in $[0, 1]$

Correlation in Bahadur model really highly constrained:

For instance, the allowable range of $\beta_a$ for the external outcome in the DEHP data is $(-0.0164; 0.1610)$ when $\beta_0$ and $\beta_d$ are fixed at their MLE. This range excludes the MLE under a beta-binomial model. It translates to $(-0.0082; 0.0803)$ on the correlation scale.

- Additional conditional and marginal approaches can be based on **pseudo-likelihood** (Molenberghs and Verbeke 2005, Chapters 9 and 12, in particular pages 200 and 246)

- Programs: Molenberghs and Verbeke (2005, p. 219ff)
The random effects in generalized linear mixed models

- enter linearly on the logit scale:
  \[
  \text{logit}[P(Y_{ij} = 1|d_i, b_i)] = \beta_0 + b_i + \beta_1 d_i
  \]
  * mean of random intercepts is 0
  * mean of average over litters is \(-3.8171\)
  * mean of predicted value over litters is \(-3.8171\)

- enter non-linearly on the probability scale:
  \[
  P(Y_{ij} = 1|d_i, b_i) = \frac{\exp(\beta_0 + b_i + \beta_1 d_i)}{1 + \exp(\beta_0 + b_i + \beta_1 d_i)}
  \]
  * mean of random effect is 0.0207
  * mean of average probabilities over litters is 0.0781
  * mean of predicted probabilities over litters is 0.0988
Chapter 26
Case Study: Binary Analysis of Analgesic Trial

▷ Research question
▷ GEE
▷ Alternating logistic regressions
▷ Further GEE analyses
▷ Generalized linear mixed model
▷ Discussion
26.1 Research Question

- Binary version of Global Satisfaction Assessment
  \[
  \text{GSABIN} = \begin{cases} 
  1 & \text{if } \text{GSA} \leq 3 \text{ (‘Very Good’ to ‘Moderate’)}, \\
  0 & \text{otherwise.}
  \end{cases}
  \]

- Marginal regression relationship:
  \[
  \text{logit}[P(Y_{ij} = 1|t_{ij}, X_i)] = \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + \beta_3 X_i.
  \]

- GLMM regression relationship:
  \[
  \text{logit}[P(Y_{ij} = 1|t_{ij}, X_i, b_i)] = \beta_0 + b_i + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + \beta_3 X_i.
  \]

- \(X_i\): baseline pain control assessment (PCA0)

- Association parameters: correlation or marginal odds ratio
26.2 GEE1

- Parameter estimates (model-based standard errors; empirically corrected standard errors)

- Four sets of working assumptions:
  - Independence
  - Exchangeable
  - AR(1)
  - Unstructured
<table>
<thead>
<tr>
<th>Effect</th>
<th>Parameter</th>
<th>IND</th>
<th>EXCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>$\beta_1$</td>
<td>2.80(0.49;0.47)</td>
<td>2.92(0.49;0.46)</td>
</tr>
<tr>
<td>Time</td>
<td>$\beta_2$</td>
<td>-0.79(0.39;0.34)</td>
<td>-0.83(0.34;0.33)</td>
</tr>
<tr>
<td>Time$^2$</td>
<td>$\beta_3$</td>
<td>0.18(0.08;0.07)</td>
<td>0.18(0.07;0.07)</td>
</tr>
<tr>
<td>Basel. PCA</td>
<td>$\beta_4$</td>
<td>-0.21(0.09;0.10)</td>
<td>-0.23(0.10;0.10)</td>
</tr>
<tr>
<td>Correlation</td>
<td>$\rho$</td>
<td>—</td>
<td>0.22</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effect</th>
<th>Parameter</th>
<th>AR</th>
<th>UN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>$\beta_1$</td>
<td>2.94(0.49;0.47)</td>
<td>2.87(0.48;0.46)</td>
</tr>
<tr>
<td>Time</td>
<td>$\beta_2$</td>
<td>-0.90(0.35;0.33)</td>
<td>-0.78(0.33;0.32)</td>
</tr>
<tr>
<td>Time$^2$</td>
<td>$\beta_3$</td>
<td>0.20(0.07;0.07)</td>
<td>0.17(0.07;0.07)</td>
</tr>
<tr>
<td>Basel. PCA</td>
<td>$\beta_4$</td>
<td>-0.22(0.10;0.10)</td>
<td>-0.23(0.10;0.10)</td>
</tr>
<tr>
<td>Correlation</td>
<td>$\rho$</td>
<td>0.25</td>
<td>—</td>
</tr>
<tr>
<td>Correlation (1,2)</td>
<td>$\rho_{12}$</td>
<td></td>
<td>0.18</td>
</tr>
<tr>
<td>Correlation (1,3)</td>
<td>$\rho_{13}$</td>
<td></td>
<td>0.25</td>
</tr>
<tr>
<td>Correlation (1,4)</td>
<td>$\rho_{14}$</td>
<td></td>
<td>0.20</td>
</tr>
<tr>
<td>Correlation (2,3)</td>
<td>$\rho_{23}$</td>
<td></td>
<td>0.18</td>
</tr>
<tr>
<td>Correlation (2,4)</td>
<td>$\rho_{24}$</td>
<td></td>
<td>0.18</td>
</tr>
<tr>
<td>Correlation (3,4)</td>
<td>$\rho_{34}$</td>
<td></td>
<td>0.46</td>
</tr>
</tbody>
</table>
• Fitted working correlation matrices:

\[
R_{EXCH} = \begin{pmatrix}
1 & 0.22 & 0.22 & 0.22 \\
0 & 1 & 0.22 & 0.22 \\
0 & 0 & 1 & 0.22 \\
0 & 0 & 0 & 1
\end{pmatrix}
\]

\[
R_{AR} = \begin{pmatrix}
1 & 0.25 & 0.06 & 0.02 \\
0 & 1 & 0.25 & 0.06 \\
0 & 0 & 1 & 0.25 \\
0 & 0 & 0 & 1
\end{pmatrix}
\]

\[
R_{UN} = \begin{pmatrix}
1 & 0.18 & 0.25 & 0.20 \\
0 & 1 & 0.18 & 0.18 \\
0 & 0 & 1 & 0.46 \\
0 & 0 & 0 & 1
\end{pmatrix}
\]
26.3 Alternating Logistic Regressions

- Parameter estimates (empirically corrected standard errors)

- Three sets of odds ratio structures:
  - Exchangeable
  - Unstructured $\equiv$ full clustering (FULLCLUST)
  - User-defined design (ZREP)
<table>
<thead>
<tr>
<th>Effect</th>
<th>Parameter</th>
<th>EXCH</th>
<th>FULLCLUST</th>
<th>ZREP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>$\beta_1$</td>
<td>2.98(0.46)</td>
<td>2.92(0.46)</td>
<td>2.92(0.46)</td>
</tr>
<tr>
<td>Time</td>
<td>$\beta_2$</td>
<td>-0.87(0.32)</td>
<td>-0.80(0.32)</td>
<td>-0.80(0.32)</td>
</tr>
<tr>
<td>Time$^2$</td>
<td>$\beta_3$</td>
<td>0.18(0.07)</td>
<td>0.17(0.06)</td>
<td>0.17(0.07)</td>
</tr>
<tr>
<td>Basel. PCA</td>
<td>$\beta_4$</td>
<td>-0.23(0.22)</td>
<td>-0.24(0.10)</td>
<td>-0.24(0.10)</td>
</tr>
<tr>
<td>Log OR</td>
<td>$\alpha$</td>
<td>1.43(0.22)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log OR(1,2)</td>
<td>$\alpha_{12}$</td>
<td></td>
<td>1.13(0.33)</td>
<td></td>
</tr>
<tr>
<td>Log OR(1,3)</td>
<td>$\alpha_{13}$</td>
<td></td>
<td>1.56(0.39)</td>
<td></td>
</tr>
<tr>
<td>Log OR(1,4)</td>
<td>$\alpha_{14}$</td>
<td></td>
<td>1.60(0.42)</td>
<td></td>
</tr>
<tr>
<td>Log OR(2,3)</td>
<td>$\alpha_{23}$</td>
<td></td>
<td>1.19(0.37)</td>
<td></td>
</tr>
<tr>
<td>Log OR(2,4)</td>
<td>$\alpha_{24}$</td>
<td></td>
<td>0.93(0.42)</td>
<td></td>
</tr>
<tr>
<td>Log OR(3,4)</td>
<td>$\alpha_{34}$</td>
<td></td>
<td>2.44(0.48)</td>
<td></td>
</tr>
<tr>
<td>Log OR par.</td>
<td>$\alpha_0$</td>
<td></td>
<td></td>
<td>1.26(0.23)</td>
</tr>
<tr>
<td>Log OR par.</td>
<td>$\alpha_1$</td>
<td></td>
<td></td>
<td>1.17(0.47)</td>
</tr>
</tbody>
</table>
• In the FULLCLUST structure, there is a hint that $\alpha_{34}$ is different from the others, with all others being equal.

• To confirm this, a Wald test can be used for the null hypothesis:

$$H_0 : \alpha_{12} = \alpha_{13} = \alpha_{14} = \alpha_{23} = \alpha_{24}$$

• Details on the test: Molenberghs and Verbeke (2005, pp. 312–313)

• The reduced structure, fitted with ZREP, is:

$$\alpha_{12} = \alpha_{13} = \alpha_{14} = \alpha_{23} = \alpha_{24} = \alpha_0,$$

$$\alpha_{34} = \alpha_0 + \alpha_1$$

• At the odds ratio level, with fitted values:

$$\hat{\psi}_{12} = \hat{\psi}_{13} = \hat{\psi}_{14} = \hat{\psi}_{23} = \hat{\psi}_{24} = \hat{\psi}_0 = 3.53,$$

$$\hat{\psi}_{34} = \hat{\psi}_0 \cdot \hat{\psi}_1 = 11.36.$$
“Odds ratio matrices”:

\[
\Psi_{\text{EXCH}} = \begin{pmatrix}
1 & 4.18 & 4.18 & 4.18 \\
1 & 4.18 & 4.18 \\
1 & 4.18 \\
1 & 1
\end{pmatrix}
\]

\[
\Psi_{\text{UN}} = \begin{pmatrix}
1 & 3.10 & 4.76 & 4.95 \\
1 & 3.29 & 2.53 \\
1 & 11.47
\end{pmatrix}
\]

\[
\Psi_{\text{ZREP}} = \begin{pmatrix}
1 & 3.53 & 3.53 & 3.53 \\
1 & 3.53 & 3.53 \\
1 & 11.36 \\
1 & 1
\end{pmatrix}
\]
26.4 A Variety of GEE Methods

- Methods used:
  - Ordinary logistic regression
  - Standard GEE
  - Prentice’s GEE
  - The linearization-based method
  - Alternating logistic regression

- Exchangeably working assumption (except for logistic regression)

- Parameter estimates (empirically corrected standard errors, unless for logistic regression)
<table>
<thead>
<tr>
<th>Effect</th>
<th>Parameter</th>
<th>Log. regr.</th>
<th>Standard</th>
<th>Prentice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>$\beta_1$</td>
<td>2.80(0.49)</td>
<td>2.92(0.46)</td>
<td>2.94(0.46)</td>
</tr>
<tr>
<td>Time</td>
<td>$\beta_2$</td>
<td>-0.79(0.39)</td>
<td>-0.83(0.33)</td>
<td>-0.84(0.33)</td>
</tr>
<tr>
<td>Time$^2$</td>
<td>$\beta_3$</td>
<td>0.18(0.08)</td>
<td>0.18(0.07)</td>
<td>0.18(0.07)</td>
</tr>
<tr>
<td>Basel. PCA</td>
<td>$\beta_4$</td>
<td>-0.21(0.09)</td>
<td>-0.23(0.10)</td>
<td>-0.23(0.10)</td>
</tr>
<tr>
<td>Correlation</td>
<td>$\rho$</td>
<td></td>
<td>0.21</td>
<td>0.26(0.05)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effect</th>
<th>Parameter</th>
<th>Lineariz.</th>
<th>ALR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>$\beta_1$</td>
<td>2.94(0.46)</td>
<td>2.98(0.46)</td>
</tr>
<tr>
<td>Time</td>
<td>$\beta_2$</td>
<td>-0.84(0.33)</td>
<td>-0.87(0.32)</td>
</tr>
<tr>
<td>Time$^2$</td>
<td>$\beta_3$</td>
<td>0.18(0.07)</td>
<td>0.18(0.07)</td>
</tr>
<tr>
<td>Basel. PCA</td>
<td>$\beta_4$</td>
<td>-0.23(0.10)</td>
<td>-0.23(0.10)</td>
</tr>
<tr>
<td>Corr.</td>
<td>$\rho$</td>
<td>0.26(0.04)</td>
<td></td>
</tr>
<tr>
<td>Log OR</td>
<td>$\alpha$</td>
<td></td>
<td>1.43(0.22)</td>
</tr>
</tbody>
</table>
26.5 Generalized Linear Mixed Models

- Four tools:
  - SAS procedure GLIMMIX:
    - MQL (\(\equiv\) MQL1)
    - PQL (\(\equiv\) PQL1)
  - MLwiN:
    - PQL1
    - PQL2
  - SAS procedure NLMIXED:
    - I: non-adaptive (\(Q = 10\))
    - II: non-adaptive (\(Q = 10\))
      \(\equiv\) adaptive (\(Q = 10\))
      \(\equiv\) adaptive (\(Q = 20\))
  - MIXOR

- Parameter estimates (standard errors)
<table>
<thead>
<tr>
<th>Effect</th>
<th>Par.</th>
<th>SAS GLIMMIX MQL</th>
<th>SAS GLIMMIX PQL1</th>
<th>MLwiN PQL1</th>
<th>MLwiN PQL2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>$\beta_1$</td>
<td>2.91(0.53)</td>
<td>3.03(0.55)</td>
<td>3.02(0.55)</td>
<td>4.07(0.70)</td>
</tr>
<tr>
<td>Time</td>
<td>$\beta_2$</td>
<td>-0.83(0.39)</td>
<td>-0.87(0.41)</td>
<td>-0.87(0.41)</td>
<td>-1.17(0.48)</td>
</tr>
<tr>
<td>Time$^2$</td>
<td>$\beta_3$</td>
<td>0.18(0.08)</td>
<td>0.19(0.08)</td>
<td>0.19(0.08)</td>
<td>0.25(0.10)</td>
</tr>
<tr>
<td>Basel. PCA</td>
<td>$\beta_4$</td>
<td>-0.22(0.11)</td>
<td>-0.22(0.11)</td>
<td>-0.22(0.11)</td>
<td>-0.31(0.15)</td>
</tr>
<tr>
<td>Rand. int s.d.</td>
<td>$\tau$</td>
<td>1.06(0.25)</td>
<td>1.04(0.23)</td>
<td>1.01(0.12)</td>
<td>1.61(0.15)</td>
</tr>
<tr>
<td>Rand. int var.</td>
<td>$\tau^2$</td>
<td>1.12(0.53)</td>
<td>1.08(0.48)</td>
<td>1.02(0.25)</td>
<td>2.59(0.47)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effect</th>
<th>Par.</th>
<th>SAS NLMIXED I</th>
<th>SAS NLMIXED II</th>
<th>MIXOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>$\beta_1$</td>
<td>4.07(0.71)</td>
<td>4.05(0.71)</td>
<td>4.05(0.55)</td>
</tr>
<tr>
<td>Time</td>
<td>$\beta_2$</td>
<td>-1.16(0.47)</td>
<td>-1.16(0.47)</td>
<td>-1.16(0.45)</td>
</tr>
<tr>
<td>Time$^2$</td>
<td>$\beta_3$</td>
<td>0.25(0.09)</td>
<td>0.24(0.09)</td>
<td>0.24(0.10)</td>
</tr>
<tr>
<td>Basel. PCA</td>
<td>$\beta_4$</td>
<td>-0.30(0.14)</td>
<td>-0.30(0.14)</td>
<td>-0.30(0.15)</td>
</tr>
<tr>
<td>Rand. int s.d.</td>
<td>$\tau$</td>
<td>1.60(0.22)</td>
<td>1.59(0.21)</td>
<td>1.59(0.21)</td>
</tr>
<tr>
<td>Rand. int var.</td>
<td>$\tau^2$</td>
<td>2.56(0.70)</td>
<td>2.53(0.68)</td>
<td>2.53(0.67)</td>
</tr>
</tbody>
</table>
26.6 Discussion

- Results are very similar, due to a relatively weak random-effects variance
- PQL1 and MQL1 perform relatively poorly
- The ratio between the RE and marginal parameters now is 1.37
- Programs: Molenberghs and Verbeke (2005, p. 219ff)
Chapter 27
Case Study: Ordinal Analysis of Analgesic Trial

▷ Proportional odds logistic regression
▷ Generalized estimating equations
▷ Generalized linear mixed models
▷ Analysis of the analgesic trial
27.1 Proportional Odds Logistic Regression

- Standard logistic regression for binary data:
  \[ \logit[P(Y_i = 1|x_i)] = \alpha + \beta x_i \]

- An extension to ordinal data: proportional odds logistic regression
  \[ \logit[P(Y_i \leq k|x_i)] = \alpha_k + \beta x_i, \quad (k = 1, \ldots, c - 1) \]

- A further extension poses problems with range-preserving restrictions:
  \[ \logit[P(Y_i \leq k|x_i)] = \alpha_k + \beta o x_i, \quad (k = 1, \ldots, c - 1) \]
  and is usually not considered

- An alternative model for ordinal data is the continuation-ratio model:
  \[ \logit[P(Y_i > k|Y_i \geq k, x_i)] = \alpha_k + \beta_k x_i, \quad (k = 1, \ldots, c - 1) \]
It is of use only when there is one natural directionality in the data: subjects go from the lowest category to higher categories, without ever returning. This is often not satisfied.

- Proportional-odds model for the 5-point GSA outcome in the analgesic trial:
  \[
  \text{logit}(P(Y_{ij} \leq k|t_{ij}, X_i)) = \alpha_k + \beta_2 t_{ij} + \beta_3 t_{ij}^2 + \beta_4 X_i, \quad (k = 1, \ldots, 4)
  \]

- SAS code:

  ```sas
  proc genmod data=m.gsa2;
  title 'Analgesic, logistic regression, Ordinal';
  class patid timecls;
  model gsa = time|time pca0 / dist=multinomial link=cumlogit;
  run;
  ```

- Note that the ‘dist’ and ‘link’ options have been adapted
• Selected output:

The GENMOD Procedure

Analysis Of Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept1</td>
<td>1</td>
<td>-1.0048</td>
<td>0.3437</td>
<td>-1.6785</td>
<td>8.55</td>
<td>0.0035</td>
</tr>
<tr>
<td>Intercept2</td>
<td>1</td>
<td>0.5225</td>
<td>0.3407</td>
<td>-0.1452</td>
<td>2.35</td>
<td>0.1251</td>
</tr>
<tr>
<td>Intercept3</td>
<td>1</td>
<td>2.3171</td>
<td>0.3481</td>
<td>1.6349</td>
<td>44.31</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Intercept4</td>
<td>1</td>
<td>4.0525</td>
<td>0.3754</td>
<td>3.3166</td>
<td>116.51</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>TIME</td>
<td>1</td>
<td>-0.2027</td>
<td>0.2706</td>
<td>-0.7330</td>
<td>0.56</td>
<td>0.4539</td>
</tr>
<tr>
<td>TIME*TIME</td>
<td>1</td>
<td>0.0479</td>
<td>0.0545</td>
<td>-0.0590</td>
<td>0.77</td>
<td>0.3798</td>
</tr>
<tr>
<td>PCA0</td>
<td>1</td>
<td>-0.2141</td>
<td>0.0622</td>
<td>-0.3361</td>
<td>11.84</td>
<td>0.0006</td>
</tr>
</tbody>
</table>

• There are $5 - 1 = 4$ intercepts, as it should.
27.2 Generalized Estimating Equations

- The same regression model as in the PO logistic regression case is used:

\[
\text{logit}[P(Y_{ij} \leq k|t_{ij}, X_i)] = \alpha_k + \beta_2 t_{ij} + \beta_3 t_{ij}^2 + \beta_4 X_i, \quad (k = 1, \ldots, 4)
\]

- This model is supplemented with working assumptions to obtain GEE

- In the SAS procedure GENMOD, only independence working assumptions are implemented for ordinal outcomes:

```sas
proc genmod data=m.gsa2;
  title 'Analgesic, GEE, Ordinal';
  class patid timecls;
  model gsa = time|time pca0 / dist=multinomial link=cumlogit;
  repeated subject=patid / type=ind covb corrw within=timecls modelse;
run;
```
The output is structured in the same way as for PO logistic regression:

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

| Parameter | Estimate | Standard Error | 95% Confidence Limits | Z | Pr > |Z| |
|-----------|----------|----------------|-----------------------|---|------|---|
| Intercept1 | -1.0048  | 0.3549         | -1.7004 -0.3092       | -2.83 | 0.0046 |
| Intercept2 | 0.5225   | 0.3568         | -0.1767 1.2218        | 1.46  | 0.1430 |
| Intercept3 | 2.3171   | 0.3669         | 1.5980 3.0363         | 6.31  | <.0001 |
| Intercept4 | 4.0525   | 0.3938         | 3.2807 4.8243         | 10.29 | <.0001 |
| TIME      | -0.2027  | 0.2028         | -0.6001 0.1948        | -1.00 | 0.3176 |
| TIME*TIME | 0.0479   | 0.0399         | -0.0304 0.1261        | 1.20  | 0.2304 |
| PCA0      | -0.2141  | 0.0911         | -0.3927 -0.0356       | -2.35 | 0.0187 |
27.3 Generalized Linear Mixed Models

- A generalized linear mixed model for ordinal data:
  \[
  \text{logit}[P(Y_{ij} \leq k|X_i, Z_i)] = \alpha_k + x_{ij}'\beta + z_{ij}'b_i, \quad (k = 1, \ldots, c - 1)
  \]

- This is the obvious counterpart for the PO logistic and GEE marginal models considered above.

- For the case of the 5-point GSA outcome in the analgesic study:
  \[
  \text{logit}[P(Y_{ij} \leq k|t_{ij}, X_i, b_i)] = \alpha_k + b_i + \beta_2 t_{ij} + \beta_3 t_{ij}^2 + \beta_4 X_i, \quad (k = 1, \ldots, 4)
  \]
Code for the SAS procedure GLIMMIX:

```sas
proc glimmix data=m.gsa2 method=RSPL;
title 'PROC GLIMMIX analysis, ordinal, RSPL (PQL, REML)';
class patid timecls;
    nloptions maxiter=50;
model gsa = time|time pca0 / dist=multinomial link=cumlogit solution;
random intercept / subject=patid type=un;
run;
```

Also here, the ‘dist’ and ‘link’ functions have to be adapted to the ordinal setting.
Selected output:

Covariance Parameter Estimates

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Subject</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>UN(1,1)</td>
<td>PATID</td>
<td>3.5348</td>
<td>0.4240</td>
</tr>
</tbody>
</table>

Solutions for Fixed Effects

| Effect      | GSA | Estimate | Standard Error | DF | t Value | Pr > |t| |
|-------------|-----|----------|----------------|----|---------|------|---|
| Intercept 1 | 1   | -1.4352  | 0.5033         | 393| -2.85   | 0.0046 |
| Intercept 2 | 2   | 0.9101   | 0.4999         | 393| 1.82    | 0.0694 |
| Intercept 3 | 3   | 3.4720   | 0.5084         | 393| 6.83    | <.0001|
| Intercept 4 | 4   | 5.6263   | 0.5358         | 393| 10.50   | <.0001|
| TIME       |     | -0.4825  | 0.2958         | 737| -1.63   | 0.1033|
| TIME*TIME  |     | 0.1009   | 0.05972        | 737| 1.69    | 0.0916|
| PCA0       |     | -0.2843  | 0.1249         | 737| -2.28   | 0.0231|
• In case the procedure NLMIXED is used, more drastic changes are needed:

```plaintext
proc nlmixed data=m.gsa2 qpoints=20;
  title 'Analgesic, PROC NLMIXED, ordinal, adaptive, q=20';
  parms int1=-1.5585 int2=1.0292 int3=3.8916 int4=6.2144
             beta1=0.5410 beta2=-0.1123 beta3=0.3173 d=2.1082;
  eta = beta1*time + beta2*time*time + beta3*pca0 + b1;
  if gsa=1 then z = 1/(1+exp(-(int1-eta)));
  else if gsa=2 then z = 1/(1+exp(-(int2-eta))) - 1/(1+exp(-(int1-eta)));
  else if gsa=3 then z = 1/(1+exp(-(int3-eta))) - 1/(1+exp(-(int2-eta)));
  else if gsa=4 then z = 1/(1+exp(-(int4-eta))) - 1/(1+exp(-(int3-eta)));
  else z = 1 - 1/(1+exp(-(int4-eta)));
  if z > 1e-8 then ll = log(z);
  else ll = -1e100;
  model gsa ~ general(ll);
  random b1 ~ normal(0,d*d) subject=patid;
  estimate 'var(d)' d*d;
run;
```
• Now, the general likelihood is used: a fully user-defined likelihood function.

• The probabilities are obtained as differences between cumulative probabilities:

\[ P(Y_{ij} = k) = P(Y_{ij} <= k) - P(Y_{ij} <= k - 1), \quad (k = 1, \ldots, 5) \]

with

\[ P(Y_{ij} <= 0) = 0 \]

\[ P(Y_{ij} <= 5) = 1 \]

• \( \eta \) is the part of the linear predictor excluding the intercept
Selected output:

### Parameter Estimates

| Parameter | Estimate | Error  | DF  | t Value | Pr > |t|  | Alpha | Lower | Upper | Gradient |
|-----------|----------|--------|-----|---------|-------|-----|-------|-------|-------|----------|
| int1      | -1.5585  | 0.5481 | 394 | -2.84   | 0.0047| 0.05| -2.6360| -0.4810| 0.000235|
| int2      | 1.0292   | 0.5442 | 394 | 1.89    | 0.0593| 0.05| -0.04063| 2.0991| -0.00004|
| int3      | 3.8916   | 0.5624 | 394 | 6.92    | <.0001| 0.05| 2.7860 | 4.9973| -0.00017|
| int4      | 6.2144   | 0.5990 | 394 | 10.37   | <.0001| 0.05| 5.0368 | 7.3920| -0.00004|
| beta1     | 0.5410   | 0.3078 | 394 | 1.76    | 0.0796| 0.05| -0.06421| 1.1462| -0.00008|
| beta2     | -0.1123  | 0.06187| 394 | -1.82   | 0.0702| 0.05| -0.2340| 0.009311| 0.000019|
| beta3     | 0.3173   | 0.1386 | 394 | 2.29    | 0.0226| 0.05| 0.04475| 0.5898| 0.000013|
| d         | 2.1082   | 0.1412 | 394 | 14.94   | <.0001| 0.05| 1.8307 | 2.3858| 0.000331|

### Additional Estimates

| Label | Estimate | Error  | DF  | t Value | Pr > |t|  | Alpha | Lower | Upper |
|-------|----------|--------|-----|---------|-------|-----|-------|-------|-------|
| var(d)| 4.4447   | 0.5952 | 394 | 7.47    | <.0001| 0.05| 3.2746| 5.6148|
27.4 Analysis of the Analgesic Trial

- Three approaches:
  - Logistic regression
  - GEE
  - GLMM

- For GEE: (model based standard errors; empirically corrected standard errors)

- MQL performs again rather poorly
## Marginal models

<table>
<thead>
<tr>
<th>Effect</th>
<th>Parameter</th>
<th>OLR</th>
<th>GEE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept 1</td>
<td>$\alpha_1$</td>
<td>-1.00(0.34)</td>
<td>-1.00(0.34;0.35)</td>
</tr>
<tr>
<td>Intercept 2</td>
<td>$\alpha_2$</td>
<td>0.52(0.34)</td>
<td>0.52(0.34;0.36)</td>
</tr>
<tr>
<td>Intercept 3</td>
<td>$\alpha_3$</td>
<td>2.32(0.35)</td>
<td>2.32(0.34;0.37)</td>
</tr>
<tr>
<td>Intercept 4</td>
<td>$\alpha_4$</td>
<td>4.05(0.38)</td>
<td>4.05(0.37;0.39)</td>
</tr>
<tr>
<td>Time</td>
<td>$\beta_2$</td>
<td>-0.20(0.27)</td>
<td>-0.20(0.27;0.20)</td>
</tr>
<tr>
<td>$\text{Time}^2$</td>
<td>$\beta_3$</td>
<td>0.05(0.05)</td>
<td>0.05(0.05;0.04)</td>
</tr>
<tr>
<td>Basel. PCA</td>
<td>$\beta_4$</td>
<td>-0.21(0.06)</td>
<td>-0.21(0.06;0.09)</td>
</tr>
</tbody>
</table>
## Random-effects models

<table>
<thead>
<tr>
<th>Effect</th>
<th>Parameter</th>
<th>MQL</th>
<th>PQL</th>
<th>N.Int.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept 1</td>
<td>$\alpha_1$</td>
<td>-0.93(0.40)</td>
<td>-1.44(0.50)</td>
<td>-1.56(0.55)</td>
</tr>
<tr>
<td>Intercept 2</td>
<td>$\alpha_2$</td>
<td>0.60(0.39)</td>
<td>0.51(0.50)</td>
<td>1.03(0.54)</td>
</tr>
<tr>
<td>Intercept 3</td>
<td>$\alpha_3$</td>
<td>2.39(0.40)</td>
<td>3.47(0.51)</td>
<td>3.89(0.56)</td>
</tr>
<tr>
<td>Intercept 4</td>
<td>$\alpha_4$</td>
<td>4.13(0.42)</td>
<td>5.63(0.54)</td>
<td>6.21(0.60)</td>
</tr>
<tr>
<td>Time</td>
<td>$\beta_2$</td>
<td>-0.30(0.28)</td>
<td>-0.48(0.30)</td>
<td>0.54(0.31)</td>
</tr>
<tr>
<td>Time&lt;sup&gt;2&lt;/sup&gt;</td>
<td>$\beta_3$</td>
<td>0.06(0.06)</td>
<td>0.10(0.06)</td>
<td>-0.11(0.06)</td>
</tr>
<tr>
<td>Basel. PCA</td>
<td>$\beta_4$</td>
<td>-0.21(0.09)</td>
<td>-0.28(0.12)</td>
<td>0.32(0.14)</td>
</tr>
<tr>
<td>Rand. int s.d.</td>
<td>$\tau$</td>
<td>1.06(0.08)</td>
<td>1.88(0.11)</td>
<td>2.11(0.14)</td>
</tr>
<tr>
<td>Rand. int var.</td>
<td>$\tau^2$</td>
<td>1.13(0.16)</td>
<td>3.53(0.42)</td>
<td>4.44(0.60)</td>
</tr>
</tbody>
</table>
Chapter 28
Count Data: The Epilepsy Study

▷ The epilepsy data
▷ Poisson regression
▷ Generalized estimating equations
▷ Generalized linear mixed models
▷ Overview of analyses of the epilepsy study
▷ Marginalization of the GLMM
28.1 The Epilepsy Data

- Consider the epilepsy data:
• We want to test for a treatment effect on number of seizures, correcting for the average number of seizures during the 12-week baseline phase, prior to the treatment.

• The response considered now is the total number of seizures a patient experienced, i.e., the sum of all weekly measurements.

• Let $Y_i$ now be the total number of seizures for subject $i$:

$$Y_i = \sum_{i=1}^{n_i} Y_{ij}$$

where $Y_{ij}$ was the original (longitudinally measured) weekly outcome.
• Histogram:

As these sums are not taken over an equal number of visits for all subjects, the above histogram is not a ‘fair’ one as it does not account for differences in $n_i$ for this.
• We will therefore use the following Poisson model:

\[ Y_i \sim \text{Poisson}(\lambda_i) \]

\[
\ln(\frac{\lambda_i}{n_i}) = x_i'\beta
\]

• Note that the regression model is equivalent to

\[ \lambda_i = n_i \exp(x_i'\beta) = \exp(x_i'\beta + \ln n_i) \]

• Since \( n_i \) is the number of weeks for which the number of seizures was recorded for subject \( i \), \( \exp(x_i'\beta) \) is the average number of seizures per week.

• \( \ln n_i \) is called an offset in the above model.

• In our application, the covariates in \( x_i \) are the treatment as well as the baseline seizure rate.
SAS statements for the calculation of outcome, offset, and for fitting the Poisson model:

```sas
proc sort data=test;
by id studyweek;
run;

proc means data=test sum n nmiss;
var nseizw;
by id;
output out=result
   n=n
   nmiss=nmiss
   sum=sum;
run;

data result;
set result;
offset=log(n-nmiss);
keep id offset sum;
run;

data first;
set test;
by id;
if first.id;
keep id bserate trt;
run;

data result;
merge result first;
by id;
run;

proc genmod data=result;
model sum=bserate trt
   / dist=poisson offset=offset;
run;
```
• The treatment variable $\text{trt}$ is coded as 0 for placebo and 1 for treated

• Output from the GENMOD procedure:

```
Analysis Of Parameter Estimates

Parameter DF Estimate    Error  Confidence Limits  Chi-Square
Intercept 1 0.8710 0.0218  0.8283  0.9138 1596.16
bserate 1 0.0172 0.0002  0.0167  0.0177  4826.14
trt 1 -0.4987 0.0341 -0.5655 -0.4319  214.18
Scale 0 1.0000 0.0000  1.0000  1.0000
```

• We obtain a highly significant reduction in the average number of seizures in the treated group, in comparison to the placebo group.
• A more general model would allow the treatment effect to depend on the baseline average number of seizures:

```plaintext
proc genmod data=result;
model sum=bserate trt bserate*trt
   / dist=poisson offset=offset;
run;
```

• Relevant part of the output:

```
Analysis Of Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>0.2107</td>
<td>0.0353</td>
<td>0.1415 - 0.2799</td>
<td>35.60</td>
</tr>
<tr>
<td>bserate</td>
<td>1</td>
<td>0.0450</td>
<td>0.0009</td>
<td>0.0432 - 0.0469</td>
<td>2286.94</td>
</tr>
<tr>
<td>trt</td>
<td>1</td>
<td>0.2938</td>
<td>0.0454</td>
<td>0.2047 - 0.3829</td>
<td>41.81</td>
</tr>
<tr>
<td>bserate*trt</td>
<td>1</td>
<td>-0.0295</td>
<td>0.0010</td>
<td>-0.0314 - -0.0276</td>
<td>911.43</td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.0000</td>
<td>0.0000</td>
<td>1.0000 - 1.0000</td>
<td></td>
</tr>
</tbody>
</table>
```
• We get a significant interaction.

• In order to explore the nature of this interaction, we estimate the treatment effect when the baseline average number of seizures equals 6, 10.5, as well as 21 (quartiles).

• This is possible via inclusion of estimate statements:

```bash
proc genmod data=result;
model sum=bserate trt bserate*trt
   / dist=poisson offset=offset;
estimate 'trt, bserate=6' trt 1 bserate*trt 6;
estimate 'trt, bserate=10.5' trt 1 bserate*trt 10.5;
estimate 'trt, bserate=21' trt 1 bserate*trt 21;
run;
```
• Additional output:

Contrast Estimate Results

<table>
<thead>
<tr>
<th>Label</th>
<th>Estimate</th>
<th>Error</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt, bserate=6</td>
<td>0.1167</td>
<td>0.0415</td>
<td>0.05</td>
</tr>
<tr>
<td>trt, bserate=10.5</td>
<td>-0.0161</td>
<td>0.0388</td>
<td>0.05</td>
</tr>
<tr>
<td>trt, bserate=21</td>
<td>-0.3260</td>
<td>0.0340</td>
<td>0.05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Label</th>
<th>Confidence Limits</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt, bserate=6</td>
<td>0.0355, 0.1980</td>
<td>7.93</td>
<td>0.0049</td>
</tr>
<tr>
<td>trt, bserate=10.5</td>
<td>-0.0921, 0.0600</td>
<td>0.17</td>
<td>0.6786</td>
</tr>
<tr>
<td>trt, bserate=21</td>
<td>-0.3926, -0.2593</td>
<td>91.86</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

• On average, there are more seizures in the treatment group when there are few seizures at baseline. The opposite is true for patients with many seizures at baseline.
28.2 Generalized Estimating Equations

- Poisson regression models will be used to describe the marginal distributions, i.e., the distribution of the outcome at each time point separately:

\[
Y_{ij} = \text{Poisson}(\lambda_{ij})
\]

\[
\log(\lambda_{ij}) = \beta_0 + \beta_1 T_i + \beta_2 t_{ij} + \beta_3 T_i t_{ij}
\]

- Notation:
  - $T_i$: treatment indicator for subject $i$
  - $t_{ij}$: time point at which $j$th measurement is taken for $i$th subject

- Note that, again, the randomization would allow to set $\beta_1$ equal to 0.
• More complex mean models can again be considered (e.g. including polynomial time effects, or including covariates).

• As the response is now the number of seizures during a fixed period of one week, we do not need to include an offset, as was the case in the GLM fitted previously to the epilepsy data, not in the context of repeated measurements.

• Given the long observation period, an unstructured working correlation would require estimation of many correlation parameters.

• Further, the long observation period makes the assumption of an exchangeable correlation structure quite unrealistic.

• We therefore use the AR(1) working correlation structure, which makes sense since we have equally spaced time points at which measurements have been taken.
• SAS code:

```sas
proc genmod data=test;
class id timeclss;
model nseizw = trt time trt*time / dist=poisson;
repeated subject=id / withinsubject=timeclss type=AR(1) corrw modelse;
run;
```

• Relevant SAS output:

```
Working Correlation Matrix

<table>
<thead>
<tr>
<th></th>
<th>Col1</th>
<th>Col2</th>
<th>Col3</th>
<th>.....</th>
<th>Col26</th>
<th>Col27</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row1</td>
<td>1.0000</td>
<td>0.5946</td>
<td>0.3535</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Row2</td>
<td>0.5946</td>
<td>1.0000</td>
<td>0.5946</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Row3</td>
<td>0.3535</td>
<td>0.5946</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Row26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.0000</td>
<td>0.5946</td>
</tr>
<tr>
<td>Row27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.5946</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
```
### Analysis Of GEE Parameter Estimates

**Empirical Standard Error Estimates**

| Parameter | Estimate | Error   | Lower Limit | Upper Limit | Z      | Pr > |Z| |
|-----------|----------|---------|-------------|-------------|--------|-------|---|
| Intercept | 1.2259   | 0.1778  | 0.8774      | 1.5743      | 6.90   | <.0001|
| trt       | 0.1681   | 0.2785  | -0.3777     | 0.7138      | 0.60   | 0.5461|
| time      | -0.0071  | 0.0229  | -0.0519     | 0.0378      | -0.31  | 0.7574|
| trt*time  | -0.0183  | 0.0279  | -0.0730     | 0.0364      | -0.66  | 0.5124|

### Analysis Of GEE Parameter Estimates

**Model-Based Standard Error Estimates**

| Parameter | Estimate | Error   | Lower Limit | Upper Limit | Z      | Pr > |Z| |
|-----------|----------|---------|-------------|-------------|--------|-------|---|
| Intercept | 1.2259   | 0.2349  | 0.7655      | 1.6862      | 5.22   | <.0001|
| trt       | 0.1681   | 0.3197  | -0.4585     | 0.7947      | 0.53   | 0.5991|
| time      | -0.0071  | 0.0230  | -0.0521     | 0.0380      | -0.31  | 0.7585|
| trt*time  | -0.0183  | 0.0310  | -0.0790     | 0.0425      | -0.59  | 0.5553|
• The AR(1) correlation coefficient is estimated to be equal to 0.5946.

• There is no difference in average evolution between both treatment groups ($p = 0.5124$).

• Note also the huge discrepancies between the results for the initial parameter estimates and the final results based on the GEE analysis.
28.3 Random-effects Model

- Conditionally on a random intercept $b_i$, Poisson regression models will be used to describe the marginal distributions, i.e., the distribution of the outcome at each time point separately:

$$ Y_{ij} = \text{Poisson}(\lambda_{ij}) $$

$$ \log(\lambda_{ij}) = \beta_0 + b_i + \beta_1 T_i + \beta_2 t_{ij} + \beta_3 T_i t_{ij} $$

- Notation:
  - $\Delta T_i$: treatment indicator for subject $i$
  - $\Delta t_{ij}$: time point at which $j$th measurement is taken for $i$th subject

- Similar as in our GEE analysis, we do not need to include an offset, because the response is now the number of seizures during a fixed period of one week.
Two equivalent SAS programs:

```sas
proc nlmixed data=test;
parms int0=0.5 slope0=-0.1 int1=1 slope1=0.1 sigma=1;
if (trt = 0) then eta = int0 + b + slope0*time;
else if (trt = 1) then eta = int1 + b + slope1*time;
lambda = exp(eta);
model nseizw ~ poisson(lambda);
random b ~ normal(0,sigma**2) subject = id;
estimate 'difference in slope' slope1-slope0;
run;

proc nlmixed data=test;
parms int0=0.5 slope0=-0.1 int1=1 slope1=0.1 sigma=1;
eta = (1-trt)*int0 + trt*int1 + b
      + (1-trt)*slope0*time + trt*slope1*time;
lambda = exp(eta);
model nseizw ~ poisson(lambda);
random b ~ normal(0,sigma**2) subject = id;
estimate 'difference in slope' slope1-slope0;
run;
```
As in the MIXED procedure, CONTRAST and ESTIMATE statements can be specified as well. However, under PROC NLMIXED, one is no longer restricted to linear functions of the parameters in the mean structure only.

For example, estimation of the ratio of both slopes, as well as of the variance of the random intercepts is achieved by adding the following ESTIMATE statements:

```plaintext
estimate 'ratio of slopes' slope1/slope0;
estimate 'variance RIs' sigma**2;
```

Inference for such functions of parameters is based on the so-called ‘delta-method’:

- Let $\psi$ be the vector of all parameters in the marginal model.
- Let $\hat{\psi}$ be the MLE of $\psi$
- $\hat{\psi}$ is asymptotically normally distributed with mean $\psi$ and covariance matrix $\text{var}(\hat{\psi})$ (inverse Fisher information matrix).
The 'delta-method' then implies that any function $F(\hat{\psi})$ of $\hat{\psi}$ is asymptotically normally distributed with mean $F(\psi)$ and covariance matrix equal to

$$\text{var}(F(\hat{\psi})) = \frac{\partial F(\psi)}{\partial \psi} \text{var}(\hat{\psi}) \frac{\partial F'(\psi)}{\partial \psi}$$

Hence, a Wald-type test can be constructed, replacing the parameters in $\text{var}(F(\hat{\psi}))$ by their estimates.

- Relevant SAS output:

| Parameter | Estimate | Standard Error | DF | t Value | Pr > |t| | Alpha | Lower | Upper | Gradient |
|-----------|----------|----------------|----|---------|-------|---------|--------|--------|--------|----------|
| int0      | 0.8180   | 0.1675         | 88 | 4.88    | <.0001| 0.05    | 0.4852 | 1.1509 | 0.006008|
| slope0    | -0.01429 | 0.004404       | 88 | -3.24   | 0.0017| 0.05    | -0.02304| -0.00554| 0.022641|
| int1      | 0.6478   | 0.1699         | 88 | 3.81    | 0.0003| 0.05    | 0.3101 | 0.9855 | 0.010749|
| slope1    | -0.01200 | 0.004318       | 88 | -2.78   | 0.0067| 0.05    | -0.02058| -0.00342| -0.04858|
| sigma     | 1.0742   | 0.08556        | 88 | 12.55   | <.0001| 0.05    | 0.9042 | 1.2442 | 0.009566|
Additional Estimates

<table>
<thead>
<tr>
<th>Label</th>
<th>Estimate</th>
<th>Error</th>
<th>DF</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>t</th>
<th></th>
<th>Alpha</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>difference in slope</td>
<td>0.002287</td>
<td>0.006167</td>
<td>88</td>
<td>0.37</td>
<td>0.7116</td>
<td>0.05</td>
<td>-0.00997</td>
<td>0.01454</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ratio of slopes</td>
<td>0.8399</td>
<td>0.3979</td>
<td>88</td>
<td>2.11</td>
<td>0.0376</td>
<td>0.05</td>
<td>0.04923</td>
<td>1.6306</td>
<td></td>
<td></td>
</tr>
<tr>
<td>variance RIs</td>
<td>1.1539</td>
<td>0.1838</td>
<td>88</td>
<td>6.28</td>
<td>&lt;.0001</td>
<td>0.05</td>
<td>0.7886</td>
<td>1.5192</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- The number of quadrature points was not specified, and therefore was selected adaptively, and set equal to only one.

- In order to check whether $Q = 1$ is sufficient, we refitted the model, prespecifying $Q = 20$. This produced essentially the same output.
• Corresponding code for the GLIMMIX procedure is:

```sas
proc glimmix data=test method=RSPL;
class id trt;
model nseizw = trt*time / dist=poisson solution;
random intercept time / type=UNR subject=id;
estimate 'diff slopes' trt*time 1 -1;
run;
```
28.4 Overview of Epilepsy Data Analyses

- GEE analysis (empirically corrected s.e.; model based s.e.)

<table>
<thead>
<tr>
<th>Effect</th>
<th>Parameter</th>
<th>Estimate (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common intercept</td>
<td>$\beta_0$</td>
<td>1.3140 (0.1435; 0.1601)</td>
</tr>
<tr>
<td>Slope placebo</td>
<td>$\beta_1$</td>
<td>$-0.0142$ (0.0234; 0.0185)</td>
</tr>
<tr>
<td>Slope treatment</td>
<td>$\beta_2$</td>
<td>$-0.0192$ (0.0178; 0.0174)</td>
</tr>
</tbody>
</table>

- Various GLMM analyses:
  - MQL
  - PQL
  - Laplace
  - Gaussian quadrature
<table>
<thead>
<tr>
<th>Effect</th>
<th>Parameter</th>
<th>MQL Estimate (s.e.)</th>
<th>MQL Estimate (s.e.)</th>
<th>PQL Estimate (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common intercept</td>
<td>$\beta_0$</td>
<td>1.3525 (0.1492)</td>
<td>0.8079 (0.1261)</td>
<td></td>
</tr>
<tr>
<td>Slope placebo</td>
<td>$\beta_1$</td>
<td>$-0.0180$ (0.0144)</td>
<td>$-0.0242$ (0.0094)</td>
<td></td>
</tr>
<tr>
<td>Slope treatment</td>
<td>$\beta_2$</td>
<td>$-0.0151$ (0.0144)</td>
<td>$-0.0191$ (0.0094)</td>
<td></td>
</tr>
<tr>
<td>Variance of intercepts</td>
<td>$d_{11}$</td>
<td>1.9017 (0.2986)</td>
<td>1.2510 (0.2155)</td>
<td></td>
</tr>
<tr>
<td>Variance of slopes</td>
<td>$d_{22}$</td>
<td>0.0084 (0.0014)</td>
<td>0.0024 (0.0006)</td>
<td></td>
</tr>
<tr>
<td>Correlation rand.eff.</td>
<td>$\rho$</td>
<td>$-0.3268$ (0.1039)</td>
<td>$-0.3394$ (0.1294)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effect</th>
<th>Parameter</th>
<th>Laplace Estimate (s.e.)</th>
<th>QUAD Estimate (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common intercept</td>
<td>$\beta_0$</td>
<td>0.7740 (0.1291)</td>
<td>0.7739 (0.1293)</td>
</tr>
<tr>
<td>Slope placebo</td>
<td>$\beta_1$</td>
<td>$-0.0244$ (0.0096)</td>
<td>$-0.0245$ (0.0096)</td>
</tr>
<tr>
<td>Slope treatment</td>
<td>$\beta_2$</td>
<td>$-0.0193$ (0.0096)</td>
<td>$-0.0193$ (0.0097)</td>
</tr>
<tr>
<td>Variance of intercepts</td>
<td>$d_{11}$</td>
<td>1.2814 (0.2220)</td>
<td>1.2859 (0.2231)</td>
</tr>
<tr>
<td>Variance of slopes</td>
<td>$d_{22}$</td>
<td>0.0024 (0.0006)</td>
<td>0.0024 (0.0006)</td>
</tr>
<tr>
<td>Correlation rand.eff.</td>
<td>$\rho$</td>
<td>$-0.3347$ (0.1317)</td>
<td>$-0.3349$ (0.1318)</td>
</tr>
</tbody>
</table>
28.5 Marginalization of the Random-effects Model

- Regression coefficients in GLMM need to be interpreted conditionally on the random effects $b_i$.

- Additional computations are needed for the population-averaged evolutions.

- The marginal expectation of $Y_{ij}$ measured at $t_{ij}$ in the placebo group is

$$E[Y_{ij}] = E[E[Y_{ij}|b_i]]$$

$$= E[\exp[(\beta_0 + b_{i1}) + (\beta_1 + b_{i2})t_{ij}]]$$

$$\neq \exp[\beta_0 + \beta_1 t_{ij}]$$

- Calculations can be done using numerical integration or numerical averaging.

- SAS code and computation: Molenberghs and Verbeke (2005, pp. 343–344)
Marginal evolutions (GEE)  

Marginal evolutions (integrated GLMM)

Sampled predicted profiles for 20 placebo patients & marginal evolution (bold)

Evolutions average subjects \((b_i = 0)\)  

Introduction to Longitudinal Data Analysis
• Curvature different in GEE and GLMM

• Ordering of treatment groups different in GEE and GLMM (although none significant)

• Watch out for the effects of missingness: many patients leave the study after week 16

• The evolution of an ‘average’ patient is completely different from the population-averaged evolution
Part V

Incomplete Data
Chapter 29
Setting The Scene

▷ Orthodontic growth data
▷ Depression trial
▷ Age-related macular degeneration trial
▷ Notation
▷ Taxonomies
29.1 Growth Data

- Taken from Potthoff and Roy, Biometrika (1964)

- Research question:

  Is dental growth related to gender?

- The distance from the center of the pituitary to the maxillary fissure was recorded at ages 8, 10, 12, and 14, for 11 girls and 16 boys
Individual profiles:

- Much variability between girls / boys
- Considerable variability within girls / boys
- Fixed number of measurements per subject
- Measurements taken at fixed time points
29.2 The Depression Trial

- Clinical trial: experimental drug *versus* standard drug
- 170 patients
- Response: change versus baseline in $HAMD_{17}$ score
- 5 post-baseline visits: 4–8
29.3 Age-related Macular Degeneration Trial

- Pharmacological Therapy for Macular Degeneration Study Group (1997)
- An ocular pressure disease which makes patients progressively lose vision
- 240 patients enrolled in a multi-center trial (190 completers)
- Treatment: Interferon-$\alpha$ (6 million units) versus placebo
- Visits: baseline and follow-up at 4, 12, 24, and 52 weeks
- Continuous outcome: visual acuity: \# letters correctly read on a vision chart
- Binary outcome: visual acuity versus baseline $\geq 0$ or $\leq 0$
### Missingness:

<table>
<thead>
<tr>
<th>Measurement occasion</th>
<th>Completers</th>
<th>Dropouts</th>
<th>Non-monotone missingness</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 wks 12 wks 24 wks 52 wks</td>
<td>Number</td>
<td>%</td>
<td>Number</td>
</tr>
<tr>
<td>Completers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O O O O O</td>
<td>188</td>
<td>78.33</td>
<td></td>
</tr>
<tr>
<td>Dropouts</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O O O M</td>
<td>24</td>
<td>10.00</td>
<td></td>
</tr>
<tr>
<td>O O M M</td>
<td>8</td>
<td>3.33</td>
<td></td>
</tr>
<tr>
<td>O M M M</td>
<td>6</td>
<td>2.50</td>
<td></td>
</tr>
<tr>
<td>M M M M</td>
<td>6</td>
<td>2.50</td>
<td></td>
</tr>
<tr>
<td>Non-monotone missingness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O O M O</td>
<td>4</td>
<td>1.67</td>
<td></td>
</tr>
<tr>
<td>O M M O</td>
<td>1</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td>M O O O</td>
<td>2</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td>M O M M</td>
<td>1</td>
<td>0.42</td>
<td></td>
</tr>
</tbody>
</table>
29.4 Incomplete Longitudinal Data
29.5 Scientific Question

- In terms of entire longitudinal profile
- In terms of last planned measurement
- In terms of last observed measurement
29.6 Notation

• Subject $i$ at occasion (time) $j = 1, \ldots, n_i$

• Measurement $Y_{ij}$

• Missingness indicator $R_{ij} = \begin{cases} 1 & \text{if $Y_{ij}$ is observed,} \\ 0 & \text{otherwise.} \end{cases}$

• Group $Y_{ij}$ into a vector $Y_i = (Y_{i1}, \ldots, Y_{ini})' = (Y^o_i, Y^m_i)$

\begin{align*}
Y^o_i & \text{ contains } Y_{ij} \text{ for which } R_{ij} = 1, \\
Y^m_i & \text{ contains } Y_{ij} \text{ for which } R_{ij} = 0.
\end{align*}

• Group $R_{ij}$ into a vector $R_i = (R_{i1}, \ldots, R_{ini})'$

• $D_i$: time of dropout: $D_i = 1 + \sum_{j=1}^{n_i} R_{ij}$
29.7 Framework

\[ f(Y_i, D_i | \theta, \psi) \]

**Selection Models:**

\[ f(Y_i | \theta) \quad f(D_i | Y_i^o, Y_i^m, \psi) \]

**MCAR \quad \rightarrow \quad MAR \quad \rightarrow \quad MNAR**

\[ f(D_i | \psi) \quad f(D_i | Y_i^o, \psi) \quad f(D_i | Y_i^o, Y_i^m, \psi) \]

**Pattern-Mixture Models:**

\[ f(Y_i | D_i, \theta) f(D_i | \psi) \]

**Shared-Parameter Models:**

\[ f(Y_i | b_i, \theta) f(D_i | b_i, \psi) \]
\[ f(Y_i, D_i | \theta, \psi) \]

**Selection Models:**

\[
\begin{align*}
    f(Y_i | \theta) \\
    f(D_i | Y_i^0, Y_i^m, \psi)
\end{align*}
\]

- MCAR \[\rightarrow\] MAR \[\rightarrow\] MNAR

- CC ?
- LOCF ?
- imputation ?

- direct likelihood !
- expectation-maximization (EM).
- multiple imputation (MI).
- (weighted) GEE !

joint model !?

sensitivity analysis ?!
29.8 Selection Models versus Pattern-mixture Models: A Paradox!

- Glynn, Laird and Rubin (1986)

- Two measurements \((Y_1, Y_2)\)

- \(Y_1\) always observed.

- \(Y_2\) observed \((R = 1)\) or missing \((R = 0)\).
Selection model versus pattern-mixture model

\[ f(y_1, y_2)g(r = 1|y_1, y_2) = f_1(y_1, y_2)p(r = 1) \]
\[ f(y_1, y_2)g(r = 0|y_1, y_2) = f_0(y_1, y_2)p(r = 0) \]

or

\[ f(y_1, y_2)g(y_1, y_2) = f_1(y_1, y_2)p \]
\[ f(y_1, y_2)[1 - g(y_1, y_2)] = f_0(y_1, y_2)[1 - p] \]

of which the ratio yields:

\[ f_0(y_1, y_2) = \frac{1 - g(y_1, y_2)}{g(y_1, y_2)} \cdot \frac{p}{1 - p} \cdot f_1(y_1, y_2) \]

The right hand side is identifiable \( \iff \) the left hand side is not...
Chapter 30
Proper Analysis of Incomplete Data

▷ Simple methods
▷ Bias for LOCF and CC
▷ Direct likelihood inference
▷ Weighted generalized estimating equations
30.1 Incomplete Longitudinal Data
Data and Modeling Strategies

[Graph showing various lines and annotations, including Inc.Obs., LOCF 'data', MAR, LOCF, AC, CC, and Comp.Obs.]

Introduction to Longitudinal Data Analysis
Modeling Strategies
30.2 Simple Methods

MCAR

Complete case analysis:
⇒ delete incomplete subjects
- Standard statistical software
- Loss of information
- Impact on precision and power
- Missingness \( \neq \) MCAR \( \Rightarrow \) bias

Last observation carried forward:
⇒ impute missing values
- Standard statistical software
- Increase of information
- Constant profile after dropout: unrealistic
- Usually bias
Quantifying the Bias

**Dropouts**  \( t_{ij} = 0 \)

- **Probability**  \( p_0 \)
- **Treatment indicator**  \( T_i = 0, 1 \)

\[
E(Y_{ij}) = \beta_0 + \beta_1 T_i + \beta_2 t_{ij} + \beta_3 T_i t_{ij}
\]

**Completers**  \( t_{ij} = 0, 1 \)

- **Probability**  \( 1 - p_0 = p_1 \)
- **Treatment indicator**  \( T_i = 0, 1 \)

\[
E(Y_{ij}) = \gamma_0 + \gamma_1 T_i + \gamma_2 t_{ij} + \gamma_3 T_i t_{ij}
\]

<table>
<thead>
<tr>
<th></th>
<th>CC</th>
<th>LOCF</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MCAR</strong></td>
<td>0</td>
<td>((p_1 - p_0)\beta_2 - (1 - p_1)\beta_3)</td>
</tr>
<tr>
<td><strong>MAR</strong></td>
<td>(-\sigma[(1 - p_1)(\beta_0 + \beta_1 - \gamma_0 - \gamma_1) - (1 - p_0)(\beta_0 - \gamma_0)])</td>
<td>(p_1(\gamma_0 + \gamma_1 + \gamma_2 + \gamma_3) + (1 - p_1)(\beta_0 + \beta_1) - p_0(\gamma_0 + \gamma_2) - (1 - p_0)\beta_0 - \gamma_1 - \gamma_3 - \sigma[(1 - p_1)(\beta_0 + \beta_1 - \gamma_0 - \gamma_1) - (1 - p_0)(\beta_0 - \gamma_0)])</td>
</tr>
</tbody>
</table>
Let us decide to use likelihood based estimation.

The full data likelihood contribution for subject $i$:

\[
L^*(\theta, \psi|Y_i, D_i) \propto f(Y_i, D_i|\theta, \psi).
\]

Base inference on the observed data:

\[
L(\theta, \psi|Y_i, D_i) \propto f(Y_i^o, D_i|\theta, \psi)
\]

with

\[
f(Y_i^o, D_i|\theta, \psi) = \int f(Y_i, D_i|\theta, \psi)dY_i^m
\]

\[
= \int f(Y_i^o, Y_i^m|\theta)f(D_i|Y_i^o, Y_i^m, \psi)dY_i^m.
\]
• Under a MAR process:

\[
f(Y_i^o, D_i|\theta, \psi) = \int f(Y_i^o, Y_i^m|\theta)f(D_i|Y_i^o, \psi)dY_i^m \\
= f(Y_i^o|\theta)f(D_i|Y_i^o, \psi),
\]

• The likelihood factorizes into two components.
30.3.1 Ignorability: Summary

Likelihood/Bayesian + MAR

&

Frequentist + MCAR
30.4 Direct Likelihood Maximization

\[
\text{MAR} : f(Y_i^0 | \theta) f(D_i | Y_i^0, \psi)
\]

Mechanism is MAR

\[\theta \text{ and } \psi \text{ distinct}\]

Interest in \(\theta\)

Use observed information matrix

\[\Rightarrow \text{ Likelihood inference is valid}\]

<table>
<thead>
<tr>
<th>Outcome type</th>
<th>Modeling strategy</th>
<th>Software</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>Linear mixed model</td>
<td>SAS proc MIXED</td>
</tr>
<tr>
<td>Non-Gaussian</td>
<td>Generalized linear mixed model</td>
<td>SAS proc GLIMMIX, NLMIXED</td>
</tr>
</tbody>
</table>
### 30.4.1 Original, Complete Orthodontic Growth Data

<table>
<thead>
<tr>
<th>Mean</th>
<th>Covar</th>
<th># par</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>unstructured</td>
<td>18</td>
</tr>
<tr>
<td>2</td>
<td>≠ slopes</td>
<td>unstructured</td>
</tr>
<tr>
<td>3</td>
<td>= slopes</td>
<td>unstructured</td>
</tr>
<tr>
<td>7</td>
<td>≠ slopes</td>
<td>CS</td>
</tr>
</tbody>
</table>

[Graphs showing growth data for different models with unstructured means and covariances, CS covariance structure.]

---

*Introduction to Longitudinal Data Analysis* 550
### 30.4.2 Trimmed Growth Data: Simple Methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Model</th>
<th>Mean</th>
<th>Covar</th>
<th># par</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete case</td>
<td>7a</td>
<td>= slopes</td>
<td>CS</td>
<td>5</td>
</tr>
<tr>
<td>LOCF</td>
<td>2a</td>
<td>quadratic</td>
<td>unstructured</td>
<td>16</td>
</tr>
<tr>
<td>Unconditional mean</td>
<td>7a</td>
<td>= slopes</td>
<td>CS</td>
<td>5</td>
</tr>
<tr>
<td>Conditional mean</td>
<td>1</td>
<td>unstructured</td>
<td>unstructured</td>
<td>18</td>
</tr>
</tbody>
</table>
30.4.3 Trimmed Growth Data: Direct Likelihood

<table>
<thead>
<tr>
<th>Mean</th>
<th>Covar</th>
<th># par</th>
</tr>
</thead>
<tbody>
<tr>
<td>7 ≠ slopes</td>
<td>CS</td>
<td>6</td>
</tr>
</tbody>
</table>

Introduction to Longitudinal Data Analysis

Growth Data, Model 1
Missing At Random
Unstructured Means, Unstructured Covariance

Growth Data, Model 2
Missing At Random
Two Lines, Unstructured Covariance

Growth Data, Model 3
Missing At Random
Parallel Lines, Unstructured Covariance

Growth Data, Model 7
Missing At Random
Two Lines, Compound Symmetry
30.4.4 Growth Data: Comparison of Analyses

**Data**
- Complete cases
- LOCF imputed data
- All available data

**Analysis methods**
- Direct likelihood
  - ML
  - REML
- MANOVA
- ANOVA per time point

**Model**
- Unstructured group by time mean
- Unstructured covariance matrix
<table>
<thead>
<tr>
<th>Principle</th>
<th>Method</th>
<th>Boys at Age 8</th>
<th>Boys at Age 10</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Original</strong></td>
<td><em>Direct likelihood, ML</em></td>
<td>22.88 (0.56)</td>
<td>23.81 (0.49)</td>
</tr>
<tr>
<td></td>
<td><em>Direct likelihood, REML ≡ MANOVA</em></td>
<td>22.88 (0.58)</td>
<td>23.81 (0.51)</td>
</tr>
<tr>
<td></td>
<td><em>ANOVA per time point</em></td>
<td>22.88 (0.61)</td>
<td>23.81 (0.53)</td>
</tr>
<tr>
<td><strong>Direct Lik.</strong></td>
<td><em>Direct likelihood, ML</em></td>
<td>22.88 (0.56)</td>
<td>23.17 (0.68)</td>
</tr>
<tr>
<td></td>
<td><em>Direct likelihood, REML</em></td>
<td>22.88 (0.58)</td>
<td>23.17 (0.71)</td>
</tr>
<tr>
<td></td>
<td><em>MANOVA</em></td>
<td>24.00 (0.48)</td>
<td>24.14 (0.66)</td>
</tr>
<tr>
<td></td>
<td><em>ANOVA per time point</em></td>
<td>22.88 (0.61)</td>
<td>24.14 (0.74)</td>
</tr>
<tr>
<td><strong>CC</strong></td>
<td><em>Direct likelihood, ML</em></td>
<td>24.00 (0.45)</td>
<td>24.14 (0.62)</td>
</tr>
<tr>
<td></td>
<td><em>Direct likelihood, REML ≡ MANOVA</em></td>
<td>24.00 (0.48)</td>
<td>24.14 (0.66)</td>
</tr>
<tr>
<td></td>
<td><em>ANOVA per time point</em></td>
<td>24.00 (0.51)</td>
<td>24.14 (0.74)</td>
</tr>
<tr>
<td><strong>LOCF</strong></td>
<td><em>Direct likelihood, ML</em></td>
<td>22.88 (0.56)</td>
<td>22.97 (0.65)</td>
</tr>
<tr>
<td></td>
<td><em>Direct likelihood, REML ≡ MANOVA</em></td>
<td>22.88 (0.58)</td>
<td>22.97 (0.68)</td>
</tr>
<tr>
<td></td>
<td><em>ANOVA per time point</em></td>
<td>22.88 (0.61)</td>
<td>22.97 (0.72)</td>
</tr>
</tbody>
</table>
30.4.5  Growth Data: Graphical Comparison of Analyses

Introduction to Longitudinal Data Analysis
30.4.6 Behind the Scenes

- $R$ completers $\leftrightarrow N - R$ “incompleters”

\[
\begin{pmatrix}
Y_{i1} \\
Y_{i2}
\end{pmatrix} \sim N
\begin{pmatrix}
\mu_1 \\
\mu_2
\end{pmatrix},
\begin{pmatrix}
\sigma_{11} & \sigma_{12} \\
\sigma_{12} & \sigma_{22}
\end{pmatrix}
\]

- Conditional density

\[
Y_{i2} | y_{i1} \sim N(\beta_0 + \beta_1 y_{i1}, \sigma_{22.1})
\]

<table>
<thead>
<tr>
<th>$\mu_1$</th>
<th>freq. &amp; lik.</th>
<th>$\bar{\mu}<em>1 = \frac{1}{N} \sum</em>{i=1}^{N} y_{i1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_2$</td>
<td>frequentist</td>
<td>$\bar{\mu}<em>2 = \frac{1}{R} \sum</em>{i=1}^{R} y_{i2}$</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>likelihood</td>
<td>$\bar{\mu}<em>2 = \frac{1}{N} \left{ \sum</em>{i=1}^{R} y_{i2} + \sum_{i=R+1}^{N} \left[ \bar{y}_2 + \bar{\beta}<em>1(y</em>{i1} - \bar{y}_1) \right] \right}$</td>
</tr>
</tbody>
</table>
30.4.7 Growth Data: Further Comparison of Analyses

<table>
<thead>
<tr>
<th>Principle</th>
<th>Method</th>
<th>Boys at Age 8</th>
<th>Boys at Age 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>Direct likelihood, ML</td>
<td>22.88 (0.56)</td>
<td>23.81 (0.49)</td>
</tr>
<tr>
<td>Direct Lik.</td>
<td>Direct likelihood, ML</td>
<td>22.88 (0.56)</td>
<td>23.17 (0.68)</td>
</tr>
<tr>
<td>CC</td>
<td>Direct likelihood, ML</td>
<td>24.00 (0.45)</td>
<td>24.14 (0.62)</td>
</tr>
<tr>
<td>LOCF</td>
<td>Direct likelihood, ML</td>
<td>22.88 (0.56)</td>
<td>22.97 (0.65)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data</th>
<th>Mean</th>
<th>Covariance</th>
<th>Boys at Age 8</th>
<th>Boys at Age 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete</td>
<td>Unstructured</td>
<td>Unstructured</td>
<td>22.88</td>
<td>23.81</td>
</tr>
<tr>
<td></td>
<td>Unstructured</td>
<td>CS</td>
<td>22.88</td>
<td>23.81</td>
</tr>
<tr>
<td></td>
<td>Unstructured</td>
<td>Independence</td>
<td>22.88</td>
<td>23.81</td>
</tr>
<tr>
<td>Incomplete</td>
<td>Unstructured</td>
<td>Unstructured</td>
<td>22.88</td>
<td>23.17</td>
</tr>
<tr>
<td></td>
<td>Unstructured</td>
<td>CS</td>
<td>22.88</td>
<td>23.52</td>
</tr>
<tr>
<td></td>
<td>Unstructured</td>
<td>Independence</td>
<td>22.88</td>
<td>24.14</td>
</tr>
</tbody>
</table>
Growth Data, Models 1, 7b, and 8b
Missing At Random

Distance

Age in Years

Introduction to Longitudinal Data Analysis
### 30.4.8 Growth Data: SAS Code for Model 1

<table>
<thead>
<tr>
<th>IDNR</th>
<th>AGE</th>
<th>SEX</th>
<th>MEASURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>2</td>
<td>21.0</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>2</td>
<td>20.0</td>
</tr>
<tr>
<td>1</td>
<td>12</td>
<td>2</td>
<td>21.5</td>
</tr>
<tr>
<td>1</td>
<td>14</td>
<td>2</td>
<td>23.0</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>2</td>
<td>20.5</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>2</td>
<td>24.5</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
<td>2</td>
<td>26.0</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**SAS code:**

```sas
proc mixed data = growth method = ml;
  class sex idnr age;
  model measure = sex age*sex / s;
  repeated age / type = un subject = idnr;
run;
```

- Subjects in terms of IDNR blocks
- age ensures proper ordering of observations within subjects!
### 30.4.9 Growth Data: SAS Code for Model 2

<table>
<thead>
<tr>
<th>IDNR</th>
<th>AGE</th>
<th>SEX</th>
<th>MEASURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>2</td>
<td>21.0</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>2</td>
<td>20.0</td>
</tr>
<tr>
<td>1</td>
<td>12</td>
<td>2</td>
<td>21.5</td>
</tr>
<tr>
<td>1</td>
<td>14</td>
<td>2</td>
<td>23.0</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>2</td>
<td>20.5</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>2</td>
<td>24.5</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
<td>2</td>
<td>26.0</td>
</tr>
</tbody>
</table>

- **SAS code:**

```sas
data help;
  set growth;
  agecat = age;
  run;

proc mixed data = growth method = ml;
  class sex idnr agecat;
  model measure = sex age*sex / s;
  repeated agecat / type = un
    subject = idnr;
  run;
```

- Time ordering variable needs to be categorical
30.5 Analysis of the Depression Trial

- **Complete case analysis:**
  
  \[ %cc(data=depression, id=patient, time=visit, response=change, out={cc}); \]

  ⇒ performs analysis on CC data set

- **LOCF analysis:**
  
  \[ %locf(data=depression, id=patient, time=visit, response=change, out={locf}); \]

  ⇒ performs analysis on LOCF data

- **Direct-likelihood analysis:** ⇒ fit linear mixed model to incomplete data
- Treatment effect at visit 8 (last follow-up measurement):

<table>
<thead>
<tr>
<th>Method</th>
<th>Estimate (s.e.)</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>-1.94 (1.17)</td>
<td>0.0995</td>
</tr>
<tr>
<td>LOCF</td>
<td>-1.63 (1.08)</td>
<td>0.1322</td>
</tr>
<tr>
<td>MAR</td>
<td>-2.38 (1.16)</td>
<td>0.0419</td>
</tr>
</tbody>
</table>

Observe the slightly significant $p$-value under the MAR model
Chapter 31
Weighted Generalized Estimating Equations

▷ General Principle
▷ Analysis of the analgesic trial
▷ Analysis of the ARMD trial
▷ Analysis of the depression trial
31.1 General Principle

- Standard GEE inference correct only under MCAR

- Under MAR: weighted GEE
  
  \[ D_i = (R_{i1}, \ldots, R_{in}) = (1, \ldots, 1, 0, \ldots, 0) \]

  Robins, Rotnitzky & Zhao (JASA, 1995)

  Fitzmaurice, Molenberghs & Lipsitz (JRSSB, 1995)

- Decompose dropout time

MAR and non-ignorable!
• Weigh a contribution by inverse dropout probability

\[ \nu_{id_i} \equiv P[D_i = d_i] = \frac{d_i - 1}{\prod_{k=2}^{d_i - 1} (1 - P[R_{ik} = 0|R_{i2} = \ldots = R_{i,k-1} = 1])} \times P[R_{id_i} = 0|R_{i2} = \ldots = R_{i,d_i-1} = 1]I\{d_i \leq T\} \]

• Adjust estimating equations

\[ \sum_{i=1}^{N} \frac{1}{\nu_{id_i}} \cdot \frac{\partial \mu_i}{\partial \beta} V_i^{-1}(y_i - \mu_i) = 0 \]
31.2 Computing the Weights

- Predicted values from (PROC GENMOD) output

- The weights are now defined at the individual measurement level:
  - At the first occasion, the weight is $w = 1$
  - At other than the last occasion, the weight is the already accumulated weight, multiplied by $1 - \text{the predicted probability}$
  - At the last occasion within a sequence where dropout occurs the weight is multiplied by the predicted probability
  - At the end of the process, the weight is inverted
31.3 Analysis of the Analgesic Trial

- A logistic regression for the dropout indicator:

$$\text{logit}[P(D_i = j|D_i \geq j, \cdot)] = \psi_0 + \psi_{11}I(GSA_{i,j-1} = 1) + \psi_{12}I(GSA_{i,j-1} = 2) + \psi_{13}I(GSA_{i,j-1} = 3) + \psi_{14}I(GSA_{i,j-1} = 4) + \psi_2\text{PCA}_{0i} + \psi_3\text{PF}_i + \psi_4\text{GD}_i$$

with

- $GSA_{i,j-1}$ the 5-point outcome at the previous time
- $I(\cdot)$ is an indicator function
- $\text{PCA}_{0i}$ is pain control assessment at baseline
- $\text{PF}_i$ is physical functioning at baseline
- $\text{GD}_i$ is genetic disorder at baseline (are used)
<table>
<thead>
<tr>
<th>Effect</th>
<th>Par.</th>
<th>Estimate (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>$\psi_0$</td>
<td>-1.80 (0.49)</td>
</tr>
<tr>
<td>Previous GSA = 1</td>
<td>$\psi_{11}$</td>
<td>-1.02 (0.41)</td>
</tr>
<tr>
<td>Previous GSA = 2</td>
<td>$\psi_{12}$</td>
<td>-1.04 (0.38)</td>
</tr>
<tr>
<td>Previous GSA = 3</td>
<td>$\psi_{13}$</td>
<td>-1.34 (0.37)</td>
</tr>
<tr>
<td>Previous GSA = 4</td>
<td>$\psi_{14}$</td>
<td>-0.26 (0.38)</td>
</tr>
<tr>
<td>Basel. PCA</td>
<td>$\psi_2$</td>
<td>0.25 (0.10)</td>
</tr>
<tr>
<td>Phys. func.</td>
<td>$\psi_3$</td>
<td>0.009 (0.004)</td>
</tr>
<tr>
<td>Genetic disfunc.</td>
<td>$\psi_4$</td>
<td>0.59 (0.24)</td>
</tr>
</tbody>
</table>

- There is some evidence for MAR: $P(D_i = j | D_i \geq j)$ depends on previous GSA.
- Furthermore: baseline PCA, physical functioning and genetic/congenital disorder.
- GEE and WGEE:

\[
\text{logit}[P(Y_{ij} = 1|t_j, \text{PCA0}_i)] = \beta_1 + \beta_2 t_j + \beta_3 t_j^2 + \beta_4 \text{PCA0}_i
\]

<table>
<thead>
<tr>
<th>Effect</th>
<th>Par.</th>
<th>GEE</th>
<th>WGEE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>(\beta_1)</td>
<td>2.95 (0.47)</td>
<td>2.17 (0.69)</td>
</tr>
<tr>
<td>Time</td>
<td>(\beta_2)</td>
<td>-0.84 (0.33)</td>
<td>-0.44 (0.44)</td>
</tr>
<tr>
<td>Time(^2)</td>
<td>(\beta_3)</td>
<td>0.18 (0.07)</td>
<td>0.12 (0.09)</td>
</tr>
<tr>
<td>Basel. PCA</td>
<td>(\beta_4)</td>
<td>-0.24 (0.10)</td>
<td>-0.16 (0.13)</td>
</tr>
</tbody>
</table>

- A hint of potentially important differences between both
● Working correlation matrices:

\[
R_{\text{UN, GEE}} = \begin{pmatrix}
1 & 0.173 & 0.246 & 0.201 \\
1 & 1 & 0.177 & 0.113 \\
1 & 0.456 & 1 & 1 \\
\end{pmatrix}
\]

\[
R_{\text{UN, WGEE}} = \begin{pmatrix}
1 & 0.215 & 0.253 & 0.167 \\
1 & 1 & 0.196 & 0.113 \\
1 & 0.409 & 1 & 1 \\
\end{pmatrix}
\]
31.4 Analgesic Trial: Steps for WGEE in SAS

1. Preparatory data manipulation:
   
   %dropout(...)

2. Logistic regression for weight model:
   
   ```sas
   proc genmod data=gsac;
   class prevgsa;
   model dropout = prevgsa pca0 physfct gendis / pred dist=b;
   ods output obstats=pred;
   run;
   ```

3. Conversion of predicted values to weights:
   
   ```sas
   ...
   %dropwgt(...)
   ```
4. Weighted GEE analysis:

    proc genmod data=repbin.gsam;
    scwgt wi;
    class patid timecls;
    model gsabin = time|time pca0 / dist=b;
    repeated subject=patid / type=un corrw within=timecls;
    run;
31.5 Analysis of the ARMD Trial

• Model for the weights:

\[
\text{logit}[P(D_i = j | D_i \geq j)] = \psi_0 + \psi_1 y_{i,j-1} + \psi_2 T_i + \psi_{31} L_{1i} + \psi_{32} L_{2i} + \psi_{34} L_{3i} \\
+ \psi_{41} I(t_j = 2) + \psi_{42} I(t_j = 3)
\]

with

▷ \( y_{i,j-1} \) the binary outcome at the previous time \( t_{i,j-1} = t_{j-1} \) (since time is common to all subjects)

▷ \( T_i = 1 \) for interferon-α and \( T_i = 0 \) for placebo

▷ \( L_{ki} = 1 \) if the patient’s eye lesion is of level \( k = 1, \ldots, 4 \) (since one dummy variable is redundant, only three are used)

▷ \( I(\cdot) \) is an indicator function
Results for the weights model:

<table>
<thead>
<tr>
<th>Effect</th>
<th>Parameter</th>
<th>Estimate (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>$\psi_0$</td>
<td>0.14 (0.49)</td>
</tr>
<tr>
<td>Previous outcome</td>
<td>$\psi_1$</td>
<td>0.04 (0.38)</td>
</tr>
<tr>
<td>Treatment</td>
<td>$\psi_2$</td>
<td>-0.86 (0.37)</td>
</tr>
<tr>
<td>Lesion level 1</td>
<td>$\psi_{31}$</td>
<td>-1.85 (0.49)</td>
</tr>
<tr>
<td>Lesion level 2</td>
<td>$\psi_{32}$</td>
<td>-1.91 (0.52)</td>
</tr>
<tr>
<td>Lesion level 3</td>
<td>$\psi_{33}$</td>
<td>-2.80 (0.72)</td>
</tr>
<tr>
<td>Time 2</td>
<td>$\psi_{41}$</td>
<td>-1.75 (0.49)</td>
</tr>
<tr>
<td>Time 3</td>
<td>$\psi_{42}$</td>
<td>-1.38 (0.44)</td>
</tr>
</tbody>
</table>
GEE:

\[ \text{logit}[P(Y_{ij} = 1|T_i, t_j)] = \beta_{j1} + \beta_{j2}T_i \]

with

- \( T_i = 0 \) for placebo and \( T_i = 1 \) for interferon-\( \alpha \)
- \( t_j \ (j = 1, \ldots, 4) \) refers to the four follow-up measurements
- Classical GEE and linearization-based GEE
- Comparison between CC, LOCF, and GEE analyses

SAS code: Molenberghs and Verbeke (2005, Section 32.5)

Results:
<table>
<thead>
<tr>
<th>Effect</th>
<th>Par.</th>
<th>CC</th>
<th>LOCF</th>
<th>Observed data</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Unweighted</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>WGEE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Standard GEE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Int.4</td>
<td>$\beta_{11}$</td>
<td>-1.01(0.24;0.24)</td>
<td>-0.87(0.20;0.21)</td>
<td>-0.87(0.21;0.21)</td>
<td>-0.98(0.10;0.44)</td>
</tr>
<tr>
<td>Int.12</td>
<td>$\beta_{21}$</td>
<td>-0.89(0.24;0.24)</td>
<td>-0.97(0.21;0.21)</td>
<td>-1.01(0.21;0.21)</td>
<td>-1.78(0.15;0.38)</td>
</tr>
<tr>
<td>Int.24</td>
<td>$\beta_{31}$</td>
<td>-1.13(0.25;0.25)</td>
<td>-1.05(0.21;0.21)</td>
<td>-1.07(0.22;0.22)</td>
<td>-1.11(0.15;0.33)</td>
</tr>
<tr>
<td>Int.52</td>
<td>$\beta_{41}$</td>
<td>-1.64(0.29;0.29)</td>
<td>-1.51(0.24;0.24)</td>
<td>-1.71(0.29;0.29)</td>
<td>-1.72(0.25;0.39)</td>
</tr>
<tr>
<td>Tr.4</td>
<td>$\beta_{12}$</td>
<td>0.40(0.32;0.32)</td>
<td>0.22(0.28;0.28)</td>
<td>0.22(0.28;0.28)</td>
<td>0.80(0.15;0.67)</td>
</tr>
<tr>
<td>Tr.12</td>
<td>$\beta_{22}$</td>
<td>0.49(0.31;0.31)</td>
<td>0.55(0.28;0.28)</td>
<td>0.61(0.29;0.29)</td>
<td>1.87(0.19;0.61)</td>
</tr>
<tr>
<td>Tr.24</td>
<td>$\beta_{32}$</td>
<td>0.48(0.33;0.33)</td>
<td>0.42(0.29;0.29)</td>
<td>0.44(0.30;0.30)</td>
<td>0.73(0.20;0.52)</td>
</tr>
<tr>
<td>Tr.52</td>
<td>$\beta_{42}$</td>
<td>0.40(0.38;0.38)</td>
<td>0.34(0.32;0.32)</td>
<td>0.44(0.37;0.37)</td>
<td>0.74(0.31;0.52)</td>
</tr>
<tr>
<td>Corr.</td>
<td>$\rho$</td>
<td>0.39</td>
<td>0.44</td>
<td>0.39</td>
<td>0.33</td>
</tr>
</tbody>
</table>
## Linearization-based GEE

<table>
<thead>
<tr>
<th>Effect</th>
<th>Par.</th>
<th>CC</th>
<th>LOCF</th>
<th>Unweighted</th>
<th>WGEE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Int.4</td>
<td>( \beta_{11} )</td>
<td>-1.01(0.24;0.24)</td>
<td>-0.87(0.21;0.21)</td>
<td>-0.87(0.21;0.21)</td>
<td>-0.98(0.18;0.44)</td>
</tr>
<tr>
<td>Int.12</td>
<td>( \beta_{21} )</td>
<td>-0.89(0.24;0.24)</td>
<td>-0.97(0.21;0.21)</td>
<td>-1.01(0.22;0.21)</td>
<td>-1.78(0.26;0.42)</td>
</tr>
<tr>
<td>Int.24</td>
<td>( \beta_{31} )</td>
<td>-1.13(0.25;0.25)</td>
<td>-1.05(0.21;0.21)</td>
<td>-1.07(0.23;0.22)</td>
<td>-1.19(0.25;0.38)</td>
</tr>
<tr>
<td>Int.52</td>
<td>( \beta_{41} )</td>
<td>-1.64(0.29;0.29)</td>
<td>-1.51(0.24;0.24)</td>
<td>-1.71(0.29;0.29)</td>
<td>-1.81(0.39;0.48)</td>
</tr>
<tr>
<td>Tr.4</td>
<td>( \beta_{12} )</td>
<td>0.40(0.32;0.32)</td>
<td>0.22(0.28;0.28)</td>
<td>0.22(0.29;0.29)</td>
<td>0.80(0.26;0.67)</td>
</tr>
<tr>
<td>Tr.12</td>
<td>( \beta_{22} )</td>
<td>0.49(0.31;0.31)</td>
<td>0.55(0.28;0.28)</td>
<td>0.61(0.28;0.29)</td>
<td>1.85(0.32;0.64)</td>
</tr>
<tr>
<td>Tr.24</td>
<td>( \beta_{32} )</td>
<td>0.48(0.33;0.33)</td>
<td>0.42(0.29;0.29)</td>
<td>0.44(0.30;0.30)</td>
<td>0.98(0.33;0.60)</td>
</tr>
<tr>
<td>Tr.52</td>
<td>( \beta_{42} )</td>
<td>0.40(0.38;0.38)</td>
<td>0.34(0.32;0.32)</td>
<td>0.44(0.37;0.37)</td>
<td>0.97(0.49;0.65)</td>
</tr>
</tbody>
</table>

\( \sigma^2 \) = 0.62  0.57  0.62  1.29  
\( \tau^2 \) = 0.39  0.44  0.39  1.85  
Corr.  \( \rho \) = 0.39  0.44  0.39  0.59
31.6 Analysis of the Depression Trial

- Response: create binary indicator $y_{bin}$ for $HAMD_{17} > 7$

- Model for dropout:

$$\text{logit}[P(D_i = j|D_i \geq j)] = \psi_0 + \psi_1 y_{i,j-1} + \gamma T_i$$

with

- $y_{i,j-1}$: the binary indicator at the previous occasion
- $T_i$: treatment indicator for patient $i$
• SAS code:
  ▶ Preparing the dataset:
  \%
  dropout(data=depression, id=patient, time=visit, response=ybin, out=dropout);
  producing:
  ▶ \texttt{dropout} indicates whether missingness at a given time occurs
  ▶ \texttt{prev} contains outcome at the previous occasion
  ▶ The logistic model for dropout:
  \%
  proc genmod data=dropout descending;
   class trt;
   model dropout = prev trt / pred dist=b;
   output out=pred p=pred;
  run;
  ▶ The weights can now be included in the GENMOD program which specifies the
  GEE, through the \texttt{WEIGHT} or \texttt{SCWGT} statements:
proc genmod data=study descending;
  weight wi;
  class patient visitclass trt;
  model ybin = trt visit trt*visit basval basval*visit / dist=bin;
  repeated subject=patient / withinsubject=visitclass type=cs corrw;
run;

• Results:

<table>
<thead>
<tr>
<th>Effec</th>
<th>WGEE</th>
<th>GEE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>est.</td>
<td>s.e.</td>
</tr>
<tr>
<td>Treatment at visit 4</td>
<td>-1.57</td>
<td>0.99</td>
</tr>
<tr>
<td>Treatment at visit 5</td>
<td>-0.67</td>
<td>0.65</td>
</tr>
<tr>
<td>Treatment at visit 6</td>
<td>0.62</td>
<td>0.56</td>
</tr>
<tr>
<td>Treatment at visit 7</td>
<td>-0.57</td>
<td>0.37</td>
</tr>
<tr>
<td>Treatment at visit 8</td>
<td>-0.84</td>
<td>0.39</td>
</tr>
</tbody>
</table>
Chapter 32
Multiple Imputation

▷ General idea
▷ Estimation
▷ Hypothesis testing
▷ Use of MI in practice
▷ Analysis of the growth data
▷ Analysis of the ARMD trial
▷ Creating monotone missingness
32.1 General Principles

- Valid under MAR

- An alternative to direct likelihood and WGEE

- Three steps:
  1. The missing values are filled in $M$ times $\Rightarrow M$ complete data sets
  2. The $M$ complete data sets are analyzed by using standard procedures
  3. The results from the $M$ analyses are combined into a single inference

32.1.1 Informal Justification

- We need to estimate \( \theta \) from the data (e.g., from the complete cases)

- Plug in the estimated \( \hat{\theta} \) and use

\[
f(y_i^m | y_i^o, \hat{\theta})
\]

to impute the missing data.

- We need to acknowledge that \( \hat{\theta} \) is a random variable; its uncertainty needs to be included in the imputation process

- Given this distribution we:
  - draw a random \( \theta^* \) from the distribution of \( \hat{\theta} \)
  - put this \( \theta^* \) in to draw a random \( Y_i^m \) from

\[
f(y_i^m | y_i^o, \theta^*).
\]
32.1.2 The Algorithm

1. Draw $\theta^*$ from its posterior distribution

2. Draw $Y_{i}^{m*}$ from $f(y_{i}^{m}|y_{i}^{o}, \theta^*)$.

3. To estimate $\beta$, then calculate the estimate of the parameter of interest, and its estimated variance, using the completed data, $(Y^o, Y^{m*})$:

$$\hat{\beta} = \hat{\beta}(Y) = \hat{\beta}(Y^o, Y^{m*})$$

The within imputation variance is

$$U = \text{Var}(\hat{\beta})$$

4. Repeat steps 1, 2 and 3 a number of $M$ times

$$\Rightarrow \quad \hat{\beta}^{m} \quad \& \quad U^{m} \quad (m = 1, \ldots, M)$$
32.1.3 Pooling Information

• With $M$ imputations, the estimate of $\beta$ is

$$\hat{\beta}^* = \frac{\sum_{m=1}^{M} \hat{\beta}^m}{M}.$$ 

• Further, one can make normally based inferences for $\beta$ with

$$(\beta - \hat{\beta}^*) \sim N(0, V),$$

where

$$V = W + \left( \frac{M + 1}{M} \right) B$$

and

$$W = \frac{\sum_{m=1}^{M} U^m}{M}$$

$$B = \frac{\sum_{m=1}^{M} (\hat{\beta}^m - \hat{\beta}^*) (\hat{\beta}^m - \hat{\beta}^*)'}{M - 1}$$
32.1.4 Hypothesis Testing

- Two “sample sizes”:
  - $N$: The sample size of the data set
  - $M$: The number of imputations

- Both play a role in the asymptotic distribution (Li, Raghunathan, and Rubin 1991)

\[
H_0 : \theta = \theta_0 \\
\downarrow \\
p = P(F_{k,w} > F)
\]
where

\[ k : \text{ length of the parameter vector } \theta \]

\[
F_{k,w} \sim F
\]

\[
F = \frac{(\theta^* - \theta_0)'W^{-1}(\theta^* - \theta_0)}{k(1 + r)}
\]

\[
w = 4 + (\tau - 4) \left[ 1 + \frac{(1 - 2\tau^{-1})}{r} \right]^2
\]

\[
r = \frac{1}{k} \left( 1 + \frac{1}{M} \right) \text{tr}(BW^{-1})
\]

\[
\tau = k(M - 1)
\]

- Limiting behavior:

\[
F \xrightarrow{M \to \infty} F_{k,\infty} = \chi^2/k
\]
32.2 Use of MI in Practice

- Many analyses of the same incomplete set of data
- A combination of missing outcomes and missing covariates
- As an alternative to WGEE: MI can be combined with classical GEE
- MI in SAS:
  
  - **Imputation Task:**
    
    PROC MI
  
  - **Analysis Task:**
    
    PROC “MYFAVORITE”
  
  - **Inference Task:**
    
    PROC MIANALYZE
32.2.1 MI Analysis of the Orthodontic Growth Data

- The same Model 1 as before

- Focus on boys at ages 8 and 10

- Results

<table>
<thead>
<tr>
<th></th>
<th>Boys at Age 8</th>
<th>Boys at Age 10</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Original Data</strong></td>
<td>22.88 (0.58)</td>
<td>23.81 (0.51)</td>
</tr>
<tr>
<td><strong>Multiple Imputation</strong></td>
<td>22.88 (0.66)</td>
<td>22.69 (0.81)</td>
</tr>
</tbody>
</table>

- Between-imputation variability for age 10 measurement

- Confidence interval for Boys at age 10: [21.08, 24.29]
32.3 MI Analysis of the ARMD Trial

- $M = 10$ imputations

- GEE:

$$\text{logit}[P(Y_{ij} = 1|T_i, t_j)] = \beta_{j1} + \beta_{j2}T_i$$

- GLMM:

$$\text{logit}[P(Y_{ij} = 1|T_i, t_j, b_i)] = \beta_{j1} + b_i + \beta_{j2}T_i, \quad b_i \sim N(0, \tau^2)$$

- $T_i = 0$ for placebo and $T_i = 1$ for interferon-$\alpha$

- $t_j (j = 1, \ldots, 4)$ refers to the four follow-up measurements

- Imputation based on the continuous outcome
Results:

<table>
<thead>
<tr>
<th>Effect</th>
<th>Par.</th>
<th>GEE</th>
<th>GLMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Int.4</td>
<td>$\beta_{11}$</td>
<td>-0.84(0.20)</td>
<td>-1.46(0.36)</td>
</tr>
<tr>
<td>Int.12</td>
<td>$\beta_{21}$</td>
<td>-1.02(0.22)</td>
<td>-1.75(0.38)</td>
</tr>
<tr>
<td>Int.24</td>
<td>$\beta_{31}$</td>
<td>-1.07(0.23)</td>
<td>-1.83(0.38)</td>
</tr>
<tr>
<td>Int.52</td>
<td>$\beta_{41}$</td>
<td>-1.61(0.27)</td>
<td>-2.69(0.45)</td>
</tr>
<tr>
<td>Trt.4</td>
<td>$\beta_{12}$</td>
<td>0.21(0.28)</td>
<td>0.32(0.48)</td>
</tr>
<tr>
<td>Trt.12</td>
<td>$\beta_{22}$</td>
<td>0.60(0.29)</td>
<td>0.99(0.49)</td>
</tr>
<tr>
<td>Trt.24</td>
<td>$\beta_{32}$</td>
<td>0.43(0.30)</td>
<td>0.67(0.51)</td>
</tr>
<tr>
<td>Trt.52</td>
<td>$\beta_{42}$</td>
<td>0.37(0.35)</td>
<td>0.52(0.56)</td>
</tr>
<tr>
<td>R.I. s.d.</td>
<td>$\tau$</td>
<td></td>
<td>2.20(0.26)</td>
</tr>
<tr>
<td>R.I. var.</td>
<td>$\tau^2$</td>
<td></td>
<td>4.85(1.13)</td>
</tr>
</tbody>
</table>
32.4 SAS Code for MI

1. Preparatory data analysis so that there is one line per subject

2. The imputation task:

   proc mi data=armd13 seed=486048 out=armd13a simple nimpute=10 round=0.1;
   var lesion diff4 diff12 diff24 diff52;
   by treat;
   run;

   Note that the imputation task is conducted on the continuous outcome ‘diff’, indicating the difference in number of letters versus baseline

3. Then, data manipulation takes place to define the binary indicators and to create a longitudinal version of the dataset
4. The analysis task (GEE):

```
proc genmod data=armd13c;
   class time subject;
   by _imputation_;
   model bindif = time1 time2 time3 time4
                trttime1 trttime2 trttime3 trttime4
       / noint dist=binomial covb;
   repeated subject=subject / withinsubject=time type=exch modelse;
   ods output ParameterEstimates=gmparms parminfo=gmpinfo CovB=gmcovb;
run;
```
5. The analysis task (GLMM):

```plaintext
proc nlmixed data=armd13c qpoints=20 maxiter=100 technique=newrap cov ecov;
   by _imputation_
   eta = beta11*time1+beta12*time2+beta13*time3+beta14*time4+b
       +beta21*trttime1+beta22*trttime2+beta23*trttime3+beta24*trttime4;
p = exp(eta)/(1+exp(eta));
model bindif ~ binary(p);
random b ~ normal(0,tau*tau) subject=subject;
estimate 'tau2' tau*tau;
ods output ParameterEstimates=nlparms CovMatParmEst=nlcovb
      AdditionalEstimates=nlparmsa CovMatAddEst=nlcovba;
run;
```
6. The inference task (GEE):

```sas
proc mianalyze parms=gmparms covb=gmcovb parminfo=gmpinfo wcov bcov tcov;
modeleffects time1 time2 time3 time4 trttime1 trttime2 trttime3 trttime4;
run;
```

7. The inference task (GLMM):

```sas
proc mianalyze parms=nlparsms covb=nlcovb wcov bcov tcov;
modeleffects beta11 beta12 beta13 beta14 beta21 beta22 beta23 beta24;
run;
```
Chapter 33
Creating Monotone Missingness

- When missingness is non-monotone, one might think of several mechanisms operating simultaneously:
  - A simple (MCAR or MAR) mechanism for the intermittent missing values
  - A more complex (MNAR) mechanism for the missing data past the moment of dropout

- Analyzing such data are complicated, especially with methods that apply to dropout only
Solution:

- Generate multiple imputations that render the datasets monotone missing, by including into the MI procedure:

  `mcmc impute = monotone;`

- Apply method of choice to the so-completed multiple sets of data

Note: this is different from the monotone method in PROC MI, intended to fully complete already monotone sets of data
Part VI

Topics in Methods and Sensitivity Analysis for Incomplete Data
Chapter 34
An MNAR Selection Model and Local Influence

▷ The Diggle and Kenward selection model
▷ Mastitis in dairy cattle
▷ An informal sensitivity analysis
▷ Local influence to conduct sensitivity analysis
34.1 A Full Selection Model

MNAR: \[ \int f(Y_i | \theta) f(D_i | Y_i, \psi) dY_i^m \]

- Linear mixed model
  \[ Y_i = X_i \beta + Z_i b_i + \varepsilon_i \]

- Logistic regressions for dropout
  \[
  \text{logit} \left[ P(D_i = j \mid D_i \geq j, Y_{i,j-1}, Y_{ij}) \right] = \psi_0 + \psi_1 Y_{i,j-1} + \psi_2 Y_{ij}
  \]

Diggle and Kenward (JRSSC 1994)
34.2 Mastitis in Dairy Cattle

- Infectious disease of the udder
- Leads to a reduction in milk yield
- High yielding cows more susceptible?
- **But** this cannot be measured directly because of the effect of the disease: *evidence is missing* since infected cause have no reported milk yield
• **Model for milk yield:**

\[
\begin{pmatrix}
Y_{i1} \\
Y_{i2}
\end{pmatrix} \sim N\left[\begin{pmatrix}
\mu \\
\mu + \Delta
\end{pmatrix}, \begin{pmatrix}
\sigma_1^2 & \rho \sigma_1 \sigma_2 \\
\rho \sigma_1 \sigma_2 & \sigma_1^2
\end{pmatrix}\right]
\]

• **Model for mastitis:**

\[
\text{logit} [P(R_i = 1|Y_{i1}, Y_{i2})] = \psi_0 + \psi_1 Y_{i1} + \psi_2 Y_{i2}
\]

\[
= 0.37 + 2.25 Y_{i1} - 2.54 Y_{i2}
\]

\[
= 0.37 - 0.29 Y_{i1} - 2.54 (Y_{i2} - Y_{i1})
\]

• **LR test for** \(H_0 : \psi_2 = 0 : G^2 = 5.11\)
34.3 Criticism → Sensitivity Analysis

“. . ., estimating the ‘unestimable’ can be accomplished only by making modelling assumptions, . . . The consequences of model misspecification will ( . . . ) be more severe in the non-random case.” (Laird 1994)

- Change distributional assumptions (Kenward 1998)
- Local and global influence methods
- Pattern-mixture models
- Several plausible models or ranges of inferences
- Semi-parametric framework (Scharfstein et al 1999)
34.4 Kenward’s Sensitivity Analysis

- Deletion of #4 and #5 $\Rightarrow G^2$ for $\psi_2$: 5.11 $\rightarrow$ 0.08

- Cows #4 and #5 have unusually large increments

- Kenward conjectures: #4 and #5 ill during the first year

- Kenward (SiM 1998)
34.5 Local Influence

- Verbeke, Thijs, Lesaffre, Kenward (Bcs 2001)

- Perturbed MAR dropout model:

\[
\text{logit } [ P(D_i = 1|Y_{i1}, Y_{i2}) ] = \psi_0 + \psi_1 Y_{i1} + \omega_i Y_{i2}
\]

- Likelihood displacement:

\[
LD(\omega) = 2 \left[ L_{\omega=0}(\hat{\theta}, \bar{\psi}) - L_{\omega=0}(\hat{\theta}_\omega, \bar{\psi}_\omega) \right] \geq 0
\]
34.5 Local Influence

- Verbeke, Thijs, Lesaffre, Kenward (Bcs 2001)

- Perturbed MAR dropout model:

\[
\text{logit } \left[ P(D_i = 1|Y_{i1}, Y_{i2}) \right] = \psi_0 + \psi_1 Y_{i1} + \omega_i Y_{i2} \\
\text{or } \psi_0 + \psi_1 Y_{i1} + \omega_i (Y_{i2} - Y_{i1})
\]

- Likelihood displacement:

\[
LD(\omega) = 2 \left[ L_{\omega=0}(\theta, \psi) - L_{\omega=0}(\hat{\theta}_\omega, \hat{\psi}_\omega) \right] \geq 0
\]
34.5.1 Likelihood Displacement

Local influence direction $h$

$\uparrow$

normal curvature $C_h$

- Local influence for $\theta$ and $\psi$:

$$C_h = C_h(\theta) + C_h(\psi)$$
34.5.2 Computational Approaches

Measuring local influence:

- Expression for $C_h$:

$$C_h = 2 \mid h' \Delta' \hat{L}^{-1} \Delta h \mid$$

- Choices for $h$
  - Direction of the $i$th subject $\Rightarrow C_i$
  - Direction $h_{\text{max}}$ of maximal curvature $C_{\text{max}}$

Fit for continuous outcomes:

- Fit MAR model:
  - linear mixed model for outcomes
  - logistic regression for dropout

- evaluate closed-form expressions for local influence
34.6 Application to Mastitis Data

- Removing #4, #5 and #66

\[ G^2 = 0.005 \]

- \( h_{\text{max}} \): different signs for (#4,#5) and #66
34.6.1 Interpretable Components of $C_i(\psi)$

\[ P(R_i = 1) [1 - P(R_i = 1)] \]

\[ Y_{i2} - Y_{i1} \]

or

\[ E(Y_{i2} \mid Y_{i1}) - Y_{i1} \]

\[ V_i \left[ \begin{array}{c} 1 \\ Y_{i1} \end{array} \right] \left\{ \sum_j V_j \left[ \begin{array}{c} 1 \\ Y_{j1} \end{array} \right] \left[ \begin{array}{c} 1 \\ Y_{j1} \end{array} \right] \right\}^{-1} \times \left[ \begin{array}{c} 1 \\ Y_{i1} \end{array} \right] \]
34.7 Global Influence Analysis

- MAR versus MNAR model

- For a variety of subsets:
  - All data
  - Removal of:
    - *(53,54,66,69)*: from local influence on \( Y_{i2} \)
    - *(4,5)*: from Kenward’s informal analysis
    - *(66)*: additional one identified from local influence on \( Y_{i2} - Y_{i1} \)
    - *(4,5,66)*: from local influence on \( Y_{i2} - Y_{i1} \)
### MAR

<table>
<thead>
<tr>
<th>Effect</th>
<th>Parameter</th>
<th>all</th>
<th>(53,54,66,69)</th>
<th>(4,5)</th>
<th>(66)</th>
<th>(4,5,66)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Measurement model:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>$\mu$</td>
<td>5.77(0.09)</td>
<td>5.69(0.09)</td>
<td>5.81(0.08)</td>
<td>5.75(0.09)</td>
<td>5.80(0.09)</td>
</tr>
<tr>
<td>Time effect</td>
<td>$\Delta$</td>
<td>0.72(0.11)</td>
<td>0.70(0.11)</td>
<td>0.64(0.09)</td>
<td>0.68(0.10)</td>
<td>0.60(0.08)</td>
</tr>
<tr>
<td>First variance</td>
<td>$\sigma^2_1$</td>
<td>0.87(0.12)</td>
<td>0.76(0.11)</td>
<td>0.77(0.11)</td>
<td>0.86(0.12)</td>
<td>0.76(0.11)</td>
</tr>
<tr>
<td>Second variance</td>
<td>$\sigma^2_2$</td>
<td>1.30(0.20)</td>
<td>1.08(0.17)</td>
<td>1.30(0.20)</td>
<td>1.10(0.17)</td>
<td>1.09(0.17)</td>
</tr>
<tr>
<td>Correlation</td>
<td>$\rho$</td>
<td>0.58(0.07)</td>
<td>0.45(0.08)</td>
<td>0.72(0.05)</td>
<td>0.57(0.07)</td>
<td>0.73(0.05)</td>
</tr>
<tr>
<td><strong>Dropout model:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>$\psi_0$</td>
<td>-2.65(1.45)</td>
<td>-3.69(1.63)</td>
<td>-2.34(1.51)</td>
<td>-2.77(1.47)</td>
<td>-2.48(1.54)</td>
</tr>
<tr>
<td>First measurement</td>
<td>$\psi_1$</td>
<td>0.27(0.25)</td>
<td>0.46(0.28)</td>
<td>0.22(0.25)</td>
<td>0.29(0.24)</td>
<td>0.24(0.26)</td>
</tr>
<tr>
<td>Second measurement</td>
<td>$\omega = \psi_2$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>-2 loglikelihood</td>
<td></td>
<td>280.02</td>
<td>246.64</td>
<td>237.94</td>
<td>264.73</td>
<td>220.23</td>
</tr>
</tbody>
</table>

### MNAR

<table>
<thead>
<tr>
<th>Effect</th>
<th>Parameter</th>
<th>all</th>
<th>(53,54,66,69)</th>
<th>(4,5)</th>
<th>(66)</th>
<th>(4,5,66)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Measurement model:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>$\mu$</td>
<td>5.77(0.09)</td>
<td>5.69(0.09)</td>
<td>5.81(0.08)</td>
<td>5.75(0.09)</td>
<td>5.80(0.09)</td>
</tr>
<tr>
<td>Time effect</td>
<td>$\Delta$</td>
<td>0.33(0.14)</td>
<td>0.35(0.14)</td>
<td>0.40(0.18)</td>
<td>0.34(0.14)</td>
<td>0.63(0.29)</td>
</tr>
<tr>
<td>First variance</td>
<td>$\sigma^2_1$</td>
<td>0.87(0.12)</td>
<td>0.76(0.11)</td>
<td>0.77(0.11)</td>
<td>0.86(0.12)</td>
<td>0.76(0.11)</td>
</tr>
<tr>
<td>Second variance</td>
<td>$\sigma^2_2$</td>
<td>1.61(0.29)</td>
<td>1.29(0.25)</td>
<td>1.39(0.25)</td>
<td>1.34(0.25)</td>
<td>1.10(0.20)</td>
</tr>
<tr>
<td>Correlation</td>
<td>$\rho$</td>
<td>0.48(0.09)</td>
<td>0.42(0.10)</td>
<td>0.67(0.06)</td>
<td>0.48(0.09)</td>
<td>0.73(0.05)</td>
</tr>
<tr>
<td><strong>Dropout model:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>$\psi_0$</td>
<td>0.37(2.33)</td>
<td>-0.37(2.65)</td>
<td>-0.77(2.04)</td>
<td>0.45(2.35)</td>
<td>-2.77(3.52)</td>
</tr>
<tr>
<td>First measurement</td>
<td>$\psi_1$</td>
<td>2.25(0.77)</td>
<td>2.11(0.76)</td>
<td>1.61(1.13)</td>
<td>2.06(0.76)</td>
<td>0.07(1.82)</td>
</tr>
<tr>
<td>Second measurement</td>
<td>$\omega = \psi_2$</td>
<td>-2.54(0.83)</td>
<td>-2.22(0.86)</td>
<td>-1.66(1.29)</td>
<td>-2.33(0.86)</td>
<td>0.20(2.09)</td>
</tr>
<tr>
<td>-2 loglikelihood</td>
<td></td>
<td>274.91</td>
<td>243.21</td>
<td>237.86</td>
<td>261.15</td>
<td>220.23</td>
</tr>
<tr>
<td>$G^2$ for MNAR</td>
<td></td>
<td>5.11</td>
<td>3.43</td>
<td>0.08</td>
<td>3.57</td>
<td>0.005</td>
</tr>
</tbody>
</table>
Chapter 35
Mechanism for Growth Data

By the way, how did Little and Rubin delete data from the growth data set?
35.1 Modeling Missingness

- Candidate model for missingness:

\[
\text{logit}[P(R_i = 0 | y_i)] = \psi_0 + \psi_1 y_{ij}, \quad \text{with } j = 1, 2, 3, \text{ or } 4
\]

- When \( j = 2 \), then MNAR, else MAR.

- Results:

<table>
<thead>
<tr>
<th>Mechanism</th>
<th>Effects</th>
<th>Deviance</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAR</td>
<td>( Y_{i1} )</td>
<td>19.51</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>MAR</td>
<td>( Y_{i3} )</td>
<td>7.43</td>
<td>0.0064</td>
</tr>
<tr>
<td>MAR</td>
<td>( Y_{i4} )</td>
<td>2.51</td>
<td>0.1131</td>
</tr>
<tr>
<td>MNAR</td>
<td>( Y_{i2} )</td>
<td>2.55</td>
<td>0.1105</td>
</tr>
</tbody>
</table>
• Including covariates:

Boys: \( \text{logit}[P(R_i = 0|y_{i1}, x_i = 0)] = \infty(22 - y_{i1}) \)

Girls: \( \text{logit}[P(R_i = 0|y_{i1}, x_i = 1)] = \infty(20.75 - y_{i1}) \)

• These models are interpreted as follows:

Boys: \( P(R_i = 0|y_{i1}, x_i = 0) = \begin{cases} 
1 & \text{if } y_{i1} < 22, \\
0.5 & \text{if } y_{i1} = 22, \\
0 & \text{if } y_{i1} > 22.
\end{cases} \)

Girls: \( P(R_i = 0|y_{i1}, x_i = 1) = \begin{cases} 
1 & \text{if } y_{i1} < 20.75, \\
0 & \text{if } y_{i1} > 20.75.
\end{cases} \)
Chapter 36
Interval of Ignorance

▷ The Fluvoxamine Study
▷ The Slovenian Public Opinion Survey
▷ MAR and MNAR analyses
▷ Informal sensitivity analysis
▷ Interval of ignorance & interval of uncertainty
36.1 Fluvoxamine Trial: Side Effects

- Post-marketing study of fluvoxamine in psychiatric patients
- Absence versus presence of side effects
- Two measurement occasions
- 315 subjects:
  - 224 completers, 75 drop out after first, 2 non-monotone, 14 without follow up
- Questions:
  - Do side effects evolve over time?
  - Are both measurements dependent?
36.2 The Slovenian Plebiscite

- Slovenian Public Opinion (SPO) Survey
- Four weeks prior to decisive plebiscite
- Three questions:
  1. Are you in favor of Slovenian independence?
  2. Are you in favor of Slovenia’s secession from Yugoslavia?
  3. Will you attend the plebiscite?
- Political decision: ABSENCE ≡ NO
- Primary Estimand: \( \theta \): Proportion in favor of independence
### Slovenian Public Opinion Survey Data:

<table>
<thead>
<tr>
<th>Secession</th>
<th>Attendance</th>
<th>Independence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>1191</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>8  0  21</td>
</tr>
<tr>
<td></td>
<td>*</td>
<td>107 3 9</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>158 68 29</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>7 14 3</td>
</tr>
<tr>
<td></td>
<td>*</td>
<td>18 43 31</td>
</tr>
<tr>
<td>*</td>
<td>Yes</td>
<td>90 2 109</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>1 2 25</td>
</tr>
<tr>
<td></td>
<td>*</td>
<td>19 8 96</td>
</tr>
</tbody>
</table>
36.3 Slovenian Public Opinion: 1st Analysis

- **Pessimistic**: All who *can* say NO will say NO

  \[ \hat{\theta} = \frac{1439}{2074} = 0.694 \]

- **Optimistic**: All who *can* say YES will say YES

  \[ \hat{\theta} = \frac{1439 + 159 + 144 + 136}{2074} = \frac{1878}{2076} = 0.904 \]

- **Resulting Interval**: 

  \[ \theta \in [0.694; 0.904] \]
• Resulting Interval:

\[ \theta \in [0.694; 0.904] \]

• Complete cases: All who answered on 3 questions

\[ \hat{\theta} = \frac{1191 + 158}{1454} = 0.928 \]

• Available cases: All who answered on both questions

\[ \hat{\theta} = \frac{1191 + 158 + 90}{1549} = 0.929 \]
36.4 Slovenian Public Opinion: 2nd Analysis

- **Missing at Random:**
  Non-response is allowed to depend on observed, but not on unobserved outcomes:
  - Based on two questions: \( \hat{\theta} = 0.892 \)
  - Based on three questions: \( \hat{\theta} = 0.883 \)

- **Missing Not at Random (NI):**
  Non-response is allowed to depend on unobserved measurements:
  \( \hat{\theta} = 0.782 \)
36.5 Slovenian Public Opinion Survey

<table>
<thead>
<tr>
<th>Estimator</th>
<th>( \hat{\theta} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pessimistic bound</td>
<td>0.694</td>
</tr>
<tr>
<td>Optimistic bound</td>
<td>0.904</td>
</tr>
<tr>
<td>Complete cases</td>
<td>0.928</td>
</tr>
<tr>
<td>Available cases</td>
<td>0.929</td>
</tr>
<tr>
<td>MAR (2 questions)</td>
<td>0.892</td>
</tr>
<tr>
<td>MAR (3 questions)</td>
<td>0.883</td>
</tr>
<tr>
<td>MNAR</td>
<td>0.782</td>
</tr>
</tbody>
</table>
36.6 Slovenian Plebiscite: The Truth?

\[ \theta = 0.885 \]

<table>
<thead>
<tr>
<th>Estimator</th>
<th>( \hat{\theta} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pessimistic bound</td>
<td>0.694</td>
</tr>
<tr>
<td>Optimistic bound</td>
<td>0.904</td>
</tr>
<tr>
<td>Complete cases</td>
<td>0.928</td>
</tr>
<tr>
<td>Available cases</td>
<td>0.929</td>
</tr>
<tr>
<td>MAR (2 questions)</td>
<td>0.892</td>
</tr>
<tr>
<td>MAR (3 questions)</td>
<td>0.883</td>
</tr>
<tr>
<td>MNAR</td>
<td>0.782</td>
</tr>
</tbody>
</table>
36.7 Did “the” MNAR model behave badly?

Consider a family of MNAR models

- Baker, Rosenberger, and DerSimonian (1992)
- Counts $Y_{r_1r_2jk}$
- $j, k = 1, 2$ indicates YES/NO
- $r_1, r_2 = 0, 1$ indicates MISSING/OBSERVED
36.7.1 Model Formulation

\[ E(Y_{11jk}) = m_{jk}, \]
\[ E(Y_{10jk}) = m_{jk}\beta_{jk}, \]
\[ E(Y_{01jk}) = m_{jk}\alpha_{jk}, \]
\[ E(Y_{00jk}) = m_{jk}\alpha_{jk}\beta_{jk}\gamma_{jk}, \]

**Interpretation:**

- \( \alpha_{jk} \): models non-response on independence question
- \( \beta_{jk} \): models non-response on attendance question
- \( \gamma_{jk} \): interaction between both non-response indicators (cannot depend on \( j \) or \( k \))
36.7.2 Identifiable Models

<table>
<thead>
<tr>
<th>Model</th>
<th>Structure</th>
<th>d.f.</th>
<th>loglik</th>
<th>θ</th>
<th>C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRD1</td>
<td>(α, β)</td>
<td>6</td>
<td>-2495.29</td>
<td>0.892</td>
<td>[0.878;0.906]</td>
</tr>
<tr>
<td>BRD2</td>
<td>(α, β_j)</td>
<td>7</td>
<td>-2467.43</td>
<td>0.884</td>
<td>[0.869;0.900]</td>
</tr>
<tr>
<td>BRD3</td>
<td>(α_k, β)</td>
<td>7</td>
<td>-2463.10</td>
<td>0.881</td>
<td>[0.866;0.897]</td>
</tr>
<tr>
<td>BRD4</td>
<td>(α, β_k)</td>
<td>7</td>
<td>-2467.43</td>
<td>0.765</td>
<td>[0.674;0.856]</td>
</tr>
<tr>
<td>BRD5</td>
<td>(α_j, β)</td>
<td>7</td>
<td>-2463.10</td>
<td>0.844</td>
<td>[0.806;0.882]</td>
</tr>
<tr>
<td>BRD6</td>
<td>(α_j, β_j)</td>
<td>8</td>
<td>-2431.06</td>
<td>0.819</td>
<td>[0.788;0.849]</td>
</tr>
<tr>
<td>BRD7</td>
<td>(α_k, β_k)</td>
<td>8</td>
<td>-2431.06</td>
<td>0.764</td>
<td>[0.697;0.832]</td>
</tr>
<tr>
<td>BRD8</td>
<td>(α_j, β_k)</td>
<td>8</td>
<td>-2431.06</td>
<td>0.741</td>
<td>[0.657;0.826]</td>
</tr>
<tr>
<td>BRD9</td>
<td>(α_k, β_j)</td>
<td>8</td>
<td>-2431.06</td>
<td>0.867</td>
<td>[0.851;0.884]</td>
</tr>
</tbody>
</table>
36.7.3 An “Interval” of MNAR Estimates

\[ \theta = 0.885 \]

<table>
<thead>
<tr>
<th>Estimator</th>
<th>( \hat{\theta} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Pessimistic; optimistic]</td>
<td>[0.694; 0.904]</td>
</tr>
<tr>
<td>Complete cases</td>
<td>0.928</td>
</tr>
<tr>
<td>Available cases</td>
<td>0.929</td>
</tr>
<tr>
<td>MAR (2 questions)</td>
<td>0.892</td>
</tr>
<tr>
<td>MAR (3 questions)</td>
<td>0.883</td>
</tr>
<tr>
<td>MNAR</td>
<td>0.782</td>
</tr>
<tr>
<td>MNAR “interval”</td>
<td>[0.741; 0.892]</td>
</tr>
</tbody>
</table>
36.8 A More Formal Look

Statistical Uncertainty

Statistical Imprecision

Statistical Ignorance
**Statistical Imprecision:** *Due to finite sampling*

- Fundamental concept of mathematical statistics
- Consistency, efficiency, precision, testing,…
- Disappears as sample size increases

**Statistical Ignorance:** *Due to incomplete observations*

- Received less attention
- Can invalidate conclusions
- Does not disappear with increasing sample size

Kenward, Goetghebeur, and Molenberghs (StatMod 2001)
### 36.8.1 Monotone Patterns

<table>
<thead>
<tr>
<th>$R = 1$</th>
<th>$R = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_{1,11}$</td>
<td>$Y_{0,1}$</td>
</tr>
<tr>
<td>$Y_{1,12}$</td>
<td>$Y_{0,2}$</td>
</tr>
<tr>
<td>$Y_{1,21}$</td>
<td>$Y_{1,22}$</td>
</tr>
</tbody>
</table>

↑

<table>
<thead>
<tr>
<th>$R = 1$</th>
<th>$R = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_{1,11}$</td>
<td>$Y_{0,11}$</td>
</tr>
<tr>
<td>$Y_{1,12}$</td>
<td>$Y_{0,12}$</td>
</tr>
<tr>
<td>$Y_{1,21}$</td>
<td>$Y_{0,21}$</td>
</tr>
<tr>
<td>$Y_{1,22}$</td>
<td>$Y_{0,22}$</td>
</tr>
</tbody>
</table>
36.8.2 Models for Monotone Patterns

\[
\mu_{r,ij} = P_{ij} q_r | ij, \quad (i,j=1,2; r=0,1)
\]

| Model         | \(q_r | ij\)                        | \# Par. | Observed d.f.  | Complete d.f. |
|---------------|---------------------------------|--------|----------------|---------------|
| 1. MCAR       | \(q_r\)                         | 4      | Non-saturated  | Non-saturated |
| 2. MAR        | \(q_r | i\)                        | 5      | Saturated      | Non-saturated |
| 3. MNAR(0)    | \(q_r | j\)                        | 5      | Saturated      | Non-saturated |
| 4. MNAR(1)    | \(\text{logit}(q_r | ij) = \alpha + \beta_i + \gamma_j\) | 6      | Overspecified  | Non-saturated |
| 5. MNAR(2)    | \(q_r | ij\)                        | 7      | Overspecified  | Saturated     |
36.8.3 Sensitivity Parameter Method

**Sensitivity Parameter:** A minimal set $\eta$

**Estimable Parameter:** $\mu$, estimable, given $\eta$

**Procedure:**
- Given $\eta$, calculate parameter and C.I. for $\mu$
- Set of parameter estimates: *region of ignorance*
- Set of interval estimates: *region of uncertainty*
- Single parameter case: ‘region’ becomes ‘interval’
## 36.9 Side Effects: Monotone Patterns

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Model 1/2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Margin</td>
<td>II 0.43</td>
<td>0.43</td>
<td>0.43</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>IU [0.37;0.48]</td>
<td>[0.37;0.48]</td>
<td>[0.37;0.48]</td>
<td>[0.37;0.48]</td>
</tr>
<tr>
<td>2nd Margin</td>
<td>II 0.64</td>
<td>0.59</td>
<td>[0.49;0.74]</td>
<td>[0.49;0.74]</td>
</tr>
<tr>
<td></td>
<td>IU [0.58;0.70]</td>
<td>[0.53;0.65]</td>
<td>[0.43;0.79]</td>
<td>[0.43;0.79]</td>
</tr>
<tr>
<td>Log O.R.</td>
<td>II 2.06</td>
<td>2.06</td>
<td>[1.52;2.08]</td>
<td>[0.41;2.84]</td>
</tr>
<tr>
<td></td>
<td>IU [1.37;2.74]</td>
<td>[1.39;2.72]</td>
<td>[1.03;2.76]</td>
<td>[0.0013;2.84]</td>
</tr>
<tr>
<td>O.R.</td>
<td>II 7.81</td>
<td>7.81</td>
<td>[4.57;7.98]</td>
<td>[1.50;17.04]</td>
</tr>
<tr>
<td></td>
<td>IU [3.95;15.44]</td>
<td>[4.00;15.24]</td>
<td>[2.79;15.74]</td>
<td>[1.0013;32.89]</td>
</tr>
</tbody>
</table>
### 36.10 Side Effects: Non-Monotone Patterns

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>First</td>
<td>Second</td>
</tr>
<tr>
<td>BRD1</td>
<td>6</td>
<td>4.5</td>
<td>0.104</td>
<td>0.43[0.37;0.49]</td>
<td>0.64[0.58;0.71]</td>
</tr>
<tr>
<td>BRD2</td>
<td>7</td>
<td>1.7</td>
<td>0.192</td>
<td>0.43[0.37;0.48]</td>
<td>0.64[0.58;0.70]</td>
</tr>
<tr>
<td>BRD3</td>
<td>7</td>
<td>2.8</td>
<td>0.097</td>
<td>0.44[0.38;0.49]</td>
<td>0.66[0.60;0.72]</td>
</tr>
<tr>
<td>BRD4</td>
<td>7</td>
<td>1.7</td>
<td>0.192</td>
<td>0.43[0.37;0.48]</td>
<td>0.58[0.49;0.68]</td>
</tr>
<tr>
<td>BRD7</td>
<td>8</td>
<td>0.0</td>
<td>-</td>
<td>0.44[0.38;0.49]</td>
<td>0.61[0.53;0.69]</td>
</tr>
<tr>
<td>BRD9</td>
<td>8</td>
<td>0.0</td>
<td>-</td>
<td>0.43[0.38;0.49]</td>
<td>0.66[0.60;0.72]</td>
</tr>
<tr>
<td>Model 10:II</td>
<td>9</td>
<td>0.0</td>
<td>-</td>
<td>[0.425;0.429]</td>
<td>[0.47;0.75]</td>
</tr>
<tr>
<td>Model 10:IU</td>
<td>9</td>
<td>0.0</td>
<td>-</td>
<td>[0.37;0.49]</td>
<td>[0.41;0.80]</td>
</tr>
</tbody>
</table>
## 36.11 Slovenian Public Opinion: 3rd Analysis

<table>
<thead>
<tr>
<th>Model</th>
<th>Structure</th>
<th>d.f.</th>
<th>loglik</th>
<th>$\theta$</th>
<th>C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRD1</td>
<td>($\alpha, \beta$)</td>
<td>6</td>
<td>-2495.29</td>
<td>0.892</td>
<td>[0.878;0.906]</td>
</tr>
<tr>
<td>BRD2</td>
<td>($\alpha, \beta_j$)</td>
<td>7</td>
<td>-2467.43</td>
<td>0.884</td>
<td>[0.869;0.900]</td>
</tr>
<tr>
<td>BRD3</td>
<td>($\alpha_k, \beta$)</td>
<td>7</td>
<td>-2463.10</td>
<td>0.881</td>
<td>[0.866;0.897]</td>
</tr>
<tr>
<td>BRD4</td>
<td>($\alpha, \beta_k$)</td>
<td>7</td>
<td>-2467.43</td>
<td>0.765</td>
<td>[0.674;0.856]</td>
</tr>
<tr>
<td>BRD5</td>
<td>($\alpha_j, \beta$)</td>
<td>7</td>
<td>-2463.10</td>
<td>0.844</td>
<td>[0.806;0.882]</td>
</tr>
<tr>
<td>BRD6</td>
<td>($\alpha_j, \beta_j$)</td>
<td>8</td>
<td>-2431.06</td>
<td>0.819</td>
<td>[0.788;0.849]</td>
</tr>
<tr>
<td>BRD7</td>
<td>($\alpha_k, \beta_j$)</td>
<td>8</td>
<td>-2431.06</td>
<td>0.764</td>
<td>[0.697;0.832]</td>
</tr>
<tr>
<td>BRD8</td>
<td>($\alpha_j, \beta_k$)</td>
<td>8</td>
<td>-2431.06</td>
<td>0.741</td>
<td>[0.657;0.826]</td>
</tr>
<tr>
<td>BRD9</td>
<td>($\alpha_k, \beta_j$)</td>
<td>8</td>
<td>-2431.06</td>
<td>0.867</td>
<td>[0.851;0.884]</td>
</tr>
<tr>
<td>Model 10</td>
<td>($\alpha_k, \beta_{jk}$)</td>
<td>9</td>
<td>-2431.06</td>
<td>[0.762;0.893]</td>
<td>[0.744;0.907]</td>
</tr>
<tr>
<td>Model 11</td>
<td>($\alpha_{jk}, \beta_j$)</td>
<td>9</td>
<td>-2431.06</td>
<td>[0.766;0.883]</td>
<td>[0.715;0.920]</td>
</tr>
<tr>
<td>Model 12</td>
<td>($\alpha_{jk}, \beta_{jk}$)</td>
<td>10</td>
<td>-2431.06</td>
<td>[0.694;0.904]</td>
<td></td>
</tr>
</tbody>
</table>
36.12 Every MNAR Model Has Got a MAR Bodyguard

- Fit an MNAR model to a set of incomplete data
- Change the conditional distribution of the unobserved outcomes, given the observed ones, to comply with MAR
- The resulting new model will have exactly the same fit as the original MNAR model
- The missing data mechanism has changed
- This implies that definitively testing for MAR versus MNAR is not possible
36.13 Slovenian Public Opinion: 4rd Analysis

<table>
<thead>
<tr>
<th>Model</th>
<th>Structure</th>
<th>d.f.</th>
<th>loglik</th>
<th>$\hat{\theta}$</th>
<th>C.I.</th>
<th>$\hat{\theta}_{MAR}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRD1</td>
<td>$(\alpha, \beta)$</td>
<td>6</td>
<td>-2495.29</td>
<td>0.892</td>
<td>[0.878;0.906]</td>
<td>0.8920</td>
</tr>
<tr>
<td>BRD2</td>
<td>$(\alpha, \beta_j)$</td>
<td>7</td>
<td>-2467.43</td>
<td>0.884</td>
<td>[0.869;0.900]</td>
<td>0.8915</td>
</tr>
<tr>
<td>BRD3</td>
<td>$(\alpha_k, \beta)$</td>
<td>7</td>
<td>-2463.10</td>
<td>0.881</td>
<td>[0.866;0.897]</td>
<td>0.8915</td>
</tr>
<tr>
<td>BRD4</td>
<td>$(\alpha, \beta_k)$</td>
<td>7</td>
<td>-2467.43</td>
<td>0.765</td>
<td>[0.674;0.856]</td>
<td>0.8915</td>
</tr>
<tr>
<td>BRD5</td>
<td>$(\alpha_j, \beta)$</td>
<td>7</td>
<td>-2463.10</td>
<td>0.844</td>
<td>[0.806;0.882]</td>
<td>0.8915</td>
</tr>
<tr>
<td>BRD6</td>
<td>$(\alpha_j, \beta_j)$</td>
<td>8</td>
<td>-2431.06</td>
<td>0.819</td>
<td>[0.788;0.849]</td>
<td>0.8919</td>
</tr>
<tr>
<td>BRD7</td>
<td>$(\alpha_k, \beta_k)$</td>
<td>8</td>
<td>-2431.06</td>
<td>0.764</td>
<td>[0.697;0.832]</td>
<td>0.8919</td>
</tr>
<tr>
<td>BRD8</td>
<td>$(\alpha_j, \beta_k)$</td>
<td>8</td>
<td>-2431.06</td>
<td>0.741</td>
<td>[0.657;0.826]</td>
<td>0.8919</td>
</tr>
<tr>
<td>BRD9</td>
<td>$(\alpha_k, \beta_j)$</td>
<td>8</td>
<td>-2431.06</td>
<td>0.867</td>
<td>[0.851;0.884]</td>
<td>0.8919</td>
</tr>
<tr>
<td>Model 10</td>
<td>$(\alpha_k, \beta_{jk})$</td>
<td>9</td>
<td>-2431.06</td>
<td>[0.762;0.893]</td>
<td>[0.744;0.907]</td>
<td>0.8919</td>
</tr>
<tr>
<td>Model 11</td>
<td>$(\alpha_{jk}, \beta_j)$</td>
<td>9</td>
<td>-2431.06</td>
<td>[0.766;0.883]</td>
<td>[0.715;0.920]</td>
<td>0.8919</td>
</tr>
<tr>
<td>Model 12</td>
<td>$(\alpha_{jk}, \beta_{jk})$</td>
<td>10</td>
<td>-2431.06</td>
<td>[0.694;0.904]</td>
<td></td>
<td>0.8919</td>
</tr>
</tbody>
</table>
\[ \theta = 0.885 \]

<table>
<thead>
<tr>
<th>Estimator</th>
<th>( \hat{\theta} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Pessimistic; optimistic]</td>
<td>[0.694; 0.904]</td>
</tr>
<tr>
<td>MAR (3 questions)</td>
<td>0.883</td>
</tr>
<tr>
<td>MNAR</td>
<td>0.782</td>
</tr>
<tr>
<td>MNAR “interval”</td>
<td>[0.753; 0.891]</td>
</tr>
<tr>
<td>Model 10</td>
<td>[0.762; 0.893]</td>
</tr>
<tr>
<td>Model 11</td>
<td>[0.766; 0.883]</td>
</tr>
<tr>
<td>Model 12</td>
<td>[0.694; 0.904]</td>
</tr>
</tbody>
</table>
Chapter 37
Pattern-mixture Models

▷ A selection model for the vorozole study
▷ Initial pattern-mixture models for the vorozole study
▷ Principles of pattern-mixture models
▷ Connection between selection models and pattern-mixture models
37.1 The Vorozole Study

- open-label study in 67 North American centers
- postmenopausal women with metastatic breast cancer
- 452 patients, followed until disease progression/death
- two groups: vorozole 2.5 mg × 1 ↔ megestrol acetate 40 mg × 4
- several outcomes: response rate, survival, safety,…
- focus: quality of life: total Function Living Index: Cancer (FLIC)  
  a higher score is more desirable
37.2 A Selection Model for the Vorozole Study

<table>
<thead>
<tr>
<th>Effect</th>
<th>Parameter</th>
<th>Estimate (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed-Effect Parameters:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>time</td>
<td>( \beta_0 )</td>
<td>7.78 (1.05)</td>
</tr>
<tr>
<td>time*baseline</td>
<td>( \beta_1 )</td>
<td>-0.065 (0.009)</td>
</tr>
<tr>
<td>time*treatment</td>
<td>( \beta_2 )</td>
<td>0.086 (0.157)</td>
</tr>
<tr>
<td>time(^2)</td>
<td>( \beta_3 )</td>
<td>-0.30 (0.06)</td>
</tr>
<tr>
<td>time(^2)*baseline</td>
<td>( \beta_4 )</td>
<td>0.0024 (0.0005)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variance Parameters:</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Random intercept</td>
<td>( d )</td>
<td>105.42</td>
</tr>
<tr>
<td>Serial variance</td>
<td>( \tau^2 )</td>
<td>77.96</td>
</tr>
<tr>
<td>Serial association</td>
<td>( \lambda )</td>
<td>7.22</td>
</tr>
<tr>
<td>Measurement error</td>
<td>( \sigma^2 )</td>
<td>77.83</td>
</tr>
</tbody>
</table>

Treatment effect: \( p = 0.5822 \)

Fitted Mean Profiles
37.2.1 The Dropout Model

<table>
<thead>
<tr>
<th></th>
<th>MAR</th>
<th>MNAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>$0.080(0.341) - 0.014(0.003)\text{base}<em>i - 0.033(0.004)y</em>{i,j-1}$</td>
<td>$0.53 - 0.015\text{base}<em>i - 0.076y</em>{i,j-1} + 0.057y_{ij}$</td>
</tr>
<tr>
<td>Extended</td>
<td>$0.033(0.401) - 0.013(0.003)\text{base}<em>i - 0.023(0.005)\frac{y</em>{i,j-2} + y_{i,j-1}}{2}$</td>
<td>$1.38 - 0.021\text{base}<em>i - 0.0027y</em>{i,j-2}$</td>
</tr>
<tr>
<td></td>
<td>$-0.047(0.010)\frac{y_{i,j-1} - y_{i,j-2}}{2}$</td>
<td>$-0.064y_{i,j-1} + 0.035y_{ij}$</td>
</tr>
</tbody>
</table>

Dropout increases with:

- low score
- negative trend
- lower baseline value
37.3 Pattern-mixture Analysis of the Vorozole Study: Profiles
37.3.1 Two Pattern-Mixture Models

Includes: time*treatment

Standard

New

Assessment of Treatment Effect

Selection model: \( p = 0.5822 \) (1 df; output)

PMM1: \( p = 0.6868 \) (1 df; output)

PMM2: \( p = 0.2403 \) (13 df; output)
\( p = 0.3206 \) (1 df; delta method)
37.3.2 Estimating Marginal Effects From PMM

- Pattern-membership probabilities:
  \[ \pi_1, \ldots, \pi_t, \ldots, \pi_T. \]

- The marginal effects:
  \[ \beta_\ell = \sum_{t=1}^{n} \beta_\ell t \pi_t, \quad \ell = 1, \ldots, g \]

- Their variance:
  \[ \text{Var}(\beta_1, \ldots, \beta_g) = AV A' \]

  where

  \[ V = \begin{pmatrix} \text{Var}(\beta_\ell t) & 0 \\ 0 & \text{Var}(\pi_t) \end{pmatrix} \]

  and

  \[ A = \frac{\partial (\beta_1, \ldots, \beta_g)}{\partial (\beta_{11}, \ldots, \beta_{ng}, \pi_1, \ldots, \pi_n)} \]
37.3.3 Considerations

- Models fitted over the observation period within a certain pattern

- How do we extrapolate beyond dropout time?

- Making the model simple enough?

- Formal identifying restrictions?

- ...

37.4 PMM: Three Strategies

(1a) **Simple model per pattern:**

\[ Y_i = X_i \beta(d_i) + Z_i b_i + \varepsilon_i \]
\[ b_i \sim N(0, D(d_i)) \]
\[ \varepsilon_i \sim N(0, \Sigma_i(d_i)) \]

(1b) **Pattern as covariate:**

\[ Y_i = X_i \beta + Z_i b_i + d_i \theta + \varepsilon_i \]

(2) **Identifying restrictions:**

**CCMV:** Complete Case Missing Values

**ACMV:** Available Case Missing Values

**NCMV:** Neighbouring Case Missing Values
37.4.1 Identifying Restrictions

Pattern 1

Pattern 2

Pattern 3
### Pattern 1:

<table>
<thead>
<tr>
<th>Effect</th>
<th>Initial</th>
<th>CCMV</th>
<th>NCMV</th>
<th>ACMV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>3.40(3.94)</td>
<td>13.21(15.91)</td>
<td>7.56(16.45)</td>
<td>4.43(18.78)</td>
</tr>
<tr>
<td>Time+base</td>
<td>-0.11(0.13)</td>
<td>-0.16(0.16)</td>
<td>-0.14(0.16)</td>
<td>-0.11(0.17)</td>
</tr>
<tr>
<td>Time+treat</td>
<td>0.33(3.91)</td>
<td>-2.09(2.19)</td>
<td>-1.20(1.93)</td>
<td>-0.41(2.52)</td>
</tr>
<tr>
<td>Time²</td>
<td>-0.84(4.21)</td>
<td>-2.12(4.24)</td>
<td>-0.70(4.22)</td>
<td></td>
</tr>
<tr>
<td>Time²+base</td>
<td>0.01(0.04)</td>
<td>0.03(0.04)</td>
<td>0.02(0.04)</td>
<td></td>
</tr>
<tr>
<td>σ₁₁</td>
<td>131.09(31.34)</td>
<td>151.91(42.34)</td>
<td>134.54(32.85)</td>
<td>137.33(34.18)</td>
</tr>
<tr>
<td>σ₁₂</td>
<td>59.84(40.46)</td>
<td>119.76(40.38)</td>
<td>97.86(38.65)</td>
<td></td>
</tr>
<tr>
<td>σ₂₂</td>
<td>201.54(65.38)</td>
<td>257.07(86.05)</td>
<td>201.87(80.02)</td>
<td></td>
</tr>
<tr>
<td>σ₁₃</td>
<td>55.12(58.03)</td>
<td>49.88(44.16)</td>
<td>61.87(43.22)</td>
<td></td>
</tr>
<tr>
<td>σ₂₃</td>
<td>84.99(48.54)</td>
<td>99.97(57.47)</td>
<td>110.42(87.95)</td>
<td></td>
</tr>
<tr>
<td>σ₃₃</td>
<td>245.06(75.56)</td>
<td>241.99(79.79)</td>
<td>286.16(117.90)</td>
<td></td>
</tr>
</tbody>
</table>

### Pattern 2:

<table>
<thead>
<tr>
<th>Effect</th>
<th>Initial</th>
<th>CCMV</th>
<th>NCMV</th>
<th>ACMV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>53.85(14.12)</td>
<td>29.78(10.43)</td>
<td>33.74(11.11)</td>
<td>28.69(11.37)</td>
</tr>
<tr>
<td>Time+base</td>
<td>-0.46(0.12)</td>
<td>-0.29(0.09)</td>
<td>-0.33(0.10)</td>
<td>-0.29(0.10)</td>
</tr>
<tr>
<td>Time+treat</td>
<td>-0.95(1.86)</td>
<td>-1.68(1.21)</td>
<td>-1.56(2.47)</td>
<td>-2.12(1.36)</td>
</tr>
<tr>
<td>Time²</td>
<td>-18.91(6.36)</td>
<td>-4.45(2.87)</td>
<td>-7.00(3.80)</td>
<td>-4.22(4.20)</td>
</tr>
<tr>
<td>Time²+base</td>
<td>0.15(0.05)</td>
<td>0.04(0.02)</td>
<td>0.07(0.03)</td>
<td>0.05(0.04)</td>
</tr>
<tr>
<td>σ₁₁</td>
<td>170.77(26.14)</td>
<td>175.59(27.53)</td>
<td>176.49(27.65)</td>
<td>177.86(28.19)</td>
</tr>
<tr>
<td>σ₁₂</td>
<td>151.84(29.19)</td>
<td>147.14(29.39)</td>
<td>149.05(29.77)</td>
<td>146.98(29.63)</td>
</tr>
<tr>
<td>σ₂₂</td>
<td>292.32(44.61)</td>
<td>297.38(46.04)</td>
<td>299.40(47.22)</td>
<td>297.39(46.04)</td>
</tr>
<tr>
<td>σ₁₃</td>
<td>57.22(37.96)</td>
<td>89.10(34.07)</td>
<td>99.18(35.07)</td>
<td></td>
</tr>
<tr>
<td>σ₂₃</td>
<td>71.58(36.73)</td>
<td>107.62(47.59)</td>
<td>166.64(66.45)</td>
<td></td>
</tr>
<tr>
<td>σ₃₃</td>
<td>212.68(101.31)</td>
<td>264.57(76.73)</td>
<td>300.78(77.97)</td>
<td></td>
</tr>
</tbody>
</table>

### Pattern 3:

<table>
<thead>
<tr>
<th>Effect</th>
<th>Initial</th>
<th>CCMV</th>
<th>NCMV</th>
<th>ACMV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>29.91(9.08)</td>
<td>29.91(9.08)</td>
<td>29.91(9.08)</td>
<td>29.91(9.08)</td>
</tr>
<tr>
<td>Time+base</td>
<td>-0.26(0.08)</td>
<td>-0.26(0.08)</td>
<td>-0.26(0.08)</td>
<td>-0.26(0.08)</td>
</tr>
<tr>
<td>Time+treat</td>
<td>0.82(0.95)</td>
<td>0.82(0.95)</td>
<td>0.82(0.95)</td>
<td>0.82(0.95)</td>
</tr>
<tr>
<td>Time²</td>
<td>-6.42(2.23)</td>
<td>-6.42(2.23)</td>
<td>-6.42(2.23)</td>
<td>-6.42(2.23)</td>
</tr>
<tr>
<td>Time²+base</td>
<td>0.05(0.02)</td>
<td>0.05(0.02)</td>
<td>0.05(0.02)</td>
<td>0.05(0.02)</td>
</tr>
<tr>
<td>σ₁₁</td>
<td>206.73(35.86)</td>
<td>206.73(35.86)</td>
<td>206.73(35.86)</td>
<td>206.73(35.86)</td>
</tr>
<tr>
<td>σ₁₂</td>
<td>96.97(26.57)</td>
<td>96.97(26.57)</td>
<td>96.97(26.57)</td>
<td>96.97(26.57)</td>
</tr>
<tr>
<td>σ₂₂</td>
<td>174.12(31.10)</td>
<td>174.12(31.10)</td>
<td>174.12(31.10)</td>
<td>174.12(31.10)</td>
</tr>
<tr>
<td>σ₁₃</td>
<td>87.38(30.66)</td>
<td>87.38(30.66)</td>
<td>87.38(30.66)</td>
<td>87.38(30.66)</td>
</tr>
<tr>
<td>σ₂₃</td>
<td>91.66(28.86)</td>
<td>91.66(28.86)</td>
<td>91.66(28.86)</td>
<td>91.66(28.86)</td>
</tr>
<tr>
<td>σ₃₃</td>
<td>262.16(44.70)</td>
<td>262.16(44.70)</td>
<td>262.16(44.70)</td>
<td>262.16(44.70)</td>
</tr>
</tbody>
</table>
### 37.4.2 Pattern As Covariate

<table>
<thead>
<tr>
<th>Effect</th>
<th>Pattern</th>
<th>Estimate (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>7.29(15.69)</td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>37.05(7.67)</td>
</tr>
<tr>
<td>Time</td>
<td>3</td>
<td>39.40(9.97)</td>
</tr>
<tr>
<td>Time*treat</td>
<td>1</td>
<td>5.25(6.41)</td>
</tr>
<tr>
<td>Time*treat</td>
<td>2</td>
<td>3.48(5.46)</td>
</tr>
<tr>
<td>Time*treat</td>
<td>3</td>
<td>3.44(6.04)</td>
</tr>
<tr>
<td>Time*base</td>
<td>1</td>
<td>-0.21(0.15)</td>
</tr>
<tr>
<td>Time*base</td>
<td>2</td>
<td>-0.34(0.06)</td>
</tr>
<tr>
<td>Time*base</td>
<td>3</td>
<td>-0.36(0.08)</td>
</tr>
<tr>
<td>Time<em>treat</em>base</td>
<td></td>
<td>-0.06(0.04)</td>
</tr>
<tr>
<td>Time^2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Time^2</td>
<td>2</td>
<td>-9.18(2.47)</td>
</tr>
<tr>
<td>Time^2</td>
<td>3</td>
<td>-7.70(2.29)</td>
</tr>
<tr>
<td>Time^2*treat</td>
<td></td>
<td>1.10(0.74)</td>
</tr>
<tr>
<td>Time^2*base</td>
<td></td>
<td>0.07(0.02)</td>
</tr>
<tr>
<td>σ_{11}</td>
<td></td>
<td>173.63(18.01)</td>
</tr>
<tr>
<td>σ_{12}</td>
<td></td>
<td>117.88(17.80)</td>
</tr>
<tr>
<td>σ_{22}</td>
<td></td>
<td>233.86(26.61)</td>
</tr>
<tr>
<td>σ_{13}</td>
<td></td>
<td>89.59(24.56)</td>
</tr>
<tr>
<td>σ_{23}</td>
<td></td>
<td>116.12(34.27)</td>
</tr>
<tr>
<td>σ_{33}</td>
<td></td>
<td>273.98(48.15)</td>
</tr>
</tbody>
</table>
37.4.3 Plot for Three Different Strategies

Strategy 1, ACMV

Strategy 2a

Strategy 2b
37.5 Connection SEM–PMM

Molenberghs, Michiels, Kenward, and Diggle (Stat Neerl 1998) Kenward, Molenberghs, and Thijs (Bka 2002)

Selection Models: $f(D_i|Y_i, \psi) \leftrightarrow f(Y_i|D_i, \theta)$: Pattern-Mixture Models

SeM : MCAR ⊂ MAR ⊂ ¬ future ⊂ MNAR

PMM : MCAR ⊂ ACMV ⊂ ¬ future ⊂ general
## Chapter 38
### Concluding Remarks

<table>
<thead>
<tr>
<th>MCAR/simple</th>
<th>CC</th>
<th>biased</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOCF</td>
<td></td>
<td>inefficient</td>
</tr>
<tr>
<td></td>
<td></td>
<td>not simpler than MAR methods</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MAR</th>
<th>direct likelihood</th>
<th>easy to conduct</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>weighted GEE</td>
<td>Gaussian &amp; non-Gaussian</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MNAR</th>
<th>variety of methods</th>
<th>strong, untestable assumptions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>most useful in sensitivity analysis</td>
</tr>
</tbody>
</table>