Using SAS PROC MIXED for the Analysis of Continuous Response of Repeated Measurements

24. November 2006
Table of Contents

- Linear Model
- Estimation
  - Estimators allowing for covariance structure
  - Estimators of the covariance matrix
- Approximation for the degrees of freedom
- Example
- Conclusion
- References
Motivation I

- classical linear model: \( y = X\beta + e \)
  \( e \sim N(0, \Sigma) \) with \( \sigma^2 \) on the main diagonal of \( \Sigma \)
- assumption in linear models is the independence of the measurements
- but with repeated measurements
  - measurements on the same person are not independent
  - measurements on different persons are independent
Motivation II

- with repeated measurements the model is $y = X\beta + e$ and $e \sim N(0, V)$ for $V$ see later
- need methods for
  - estimators for $\beta$ and $V$
  - approximation of the degrees of freedom
Covariance structure

Different structures of the covariance matrix $\mathbf{V}$

- **Compound Symmetry**

  $\begin{pmatrix}
  \sigma^2_b + \sigma^2_w & \sigma^2_b & \sigma^2_b \\
  \sigma^2_b & \sigma^2_b + \sigma^2_w & \sigma^2_b \\
  \sigma^2_b & \sigma^2_b & \sigma^2_b + \sigma^2_w
  \end{pmatrix}$

- **First-Order Autoregressive**

  $\begin{pmatrix}
  \sigma^2 & \sigma^2 \rho & \sigma^2 \rho^2 \\
  \sigma^2 \rho & \sigma^2 & \sigma^2 \rho \\
  \sigma^2 \rho^2 & \sigma^2 \rho & \sigma^2
  \end{pmatrix}$

- **Unstructured**

  $\begin{pmatrix}
  \sigma^2_{11} & \sigma^2_{21} & \sigma^2_{31} \\
  \sigma^2_{21} & \sigma^2_{22} & \sigma^2_{32} \\
  \sigma^2_{31} & \sigma^2_{32} & \sigma^2_{33}
  \end{pmatrix}$

- ... more in SAS
Estimators in the classical linear model

- Least-Squares-Estimator $\hat{\beta}_{LS} = (X'X)^{-1}X'y$
- Maximum-Likelihood-Estimator $\hat{\beta}_{ML} = (X'X)^{-1}X'y$
- the ML-Estimator is identical to the LS-Estimator
- but observations must be independent
Generalized-Least-Squares-Estimator

- minimize $(y - X\beta)'V^{-1}(y - X\beta)$ with respect to $\beta$
- $\hat{\beta}_{\text{GLS}} = (X'V^{-1}X)^{-1}X'V^{-1}y$
- $V$ is assumed to be known, estimators for $V$ are shown in the next section
Maximum-Likelihood-Estimator

\[
\hat{\beta}_{ML} = (X'V^{-1}X)^{-1}X'V^{-1}y
\]

- \( V \) is assumed to be known
- \( \hat{\beta}_{GLS} \) and \( \hat{\beta}_{ML} \) are identical
Using SAS PROC MIXED for the Analysis of Continuous Response of Repeated Measurements

Estimation

Estimators of the covariance matrix

Maximum-Likelihood-Estimator for $V$

- maximize the log-likelihood

$$-\frac{1}{2} \log |V| - \frac{1}{2} (y - X\hat{\beta})' V^{-1} (y - X\hat{\beta}) - \frac{n}{2} \log(2\pi)$$
Restricted-Maximum-Likelihood-Estimator for $\mathbf{V}$

- is defined as a maximum likelihood estimator based on a linearly transformed set of data $\mathbf{Y}^* = \mathbf{A}\mathbf{Y}$ such that the distribution of $\mathbf{Y}^*$ does not depend on $\mathbf{\beta}$

- for example $\mathbf{A}$ is defined as $\mathbf{A} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$

- maximize the log-likelihood

$$-\frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} \log |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| - \frac{1}{2} (\mathbf{y} - \mathbf{X}\hat{\mathbf{\beta}})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{\beta}}) - \frac{n-p}{2} \log(2\pi)$$
Comparison of the ML-Estimator with the REML-Estimator

- log-likelihood of the ML-Estimator:
  \[-\frac{1}{2} \log |V| - \frac{1}{2} (y - X\hat{\beta})'V^{-1}(y - X\hat{\beta}) - \frac{n}{2} \log(2\pi)\]

- log-likelihood of the REML-Estimator:
  \[-\frac{1}{2} \log |V| - \frac{1}{2} \log |X'V^{-1}X| - \frac{1}{2} (y - X\hat{\beta})'V^{-1}(y - X\hat{\beta}) - \frac{n-p}{2} \log(2\pi)\]

- the difference is \(\frac{1}{2} \log |X'V^{-1}X|\)
Pass of estimation

- estimate $\beta$ under the assumption that $V$ is known
- estimate with this $\hat{\beta}$ the covariance matrix $V$
- estimate with $\hat{V}$ the parameter vector $\beta$ new
Properties of the estimators

- GLS- and ML-Estimator for $\beta$ are identical
- REML-Estimator for $V$ takes account of the numbers of parameters in the model
- ML-Estimator for $V$ does not take account of the numbers of parameters
- the bias of the REML-Estimator is smaller than the bias of the ML-Estimator
- default in SAS PROC MIXED for $\hat{V}$ is the REML-Estimator
Residual-Method

- $n - \text{rank}(X)$
- Method in the classical model
Satterthwaite-Method I

method for the t-test

- degrees of freedom are calculated as \( \nu = \frac{2 \left( \text{Var} (c\hat{\beta}) \right)^2}{\text{Var} (\text{Var} (c\hat{\beta}))} \)

- \( c \) is a vector defining the contrast which is tested
Satterthwaite-Method II

method for the F-test

- use the spectral decomposition of \( \left( \widehat{\text{Var}}(C\hat{\beta}) \right)^{-1} \) to get
  \[
P' \left( \widehat{\text{Var}}(C\hat{\beta}) \right)^{-1} P = \text{diag}(\lambda_m)
  \]
columns of \( P \) are normalized eigenvectors and \( \lambda_m \) are the eigenvalues

- define \( Q = qF \) with \( q = \text{rank}(C) \) and \( F \) is the Wald F-statistic;
  \( Q \) can be written as \( \sum_{m=1}^{q} t_{\nu_m}^2 \)

- \( \nu_m \) are the degrees of freedom in the \( m \)th t-test
Satterthwaite-Method III

- use the relationship that $\text{E}(F) = \frac{\nu}{\nu - 2}$
- search $\nu$ so that $q^{-1} Q \sim F_{q, \nu}$
- $E(Q) = \sum_{m=1}^{q} \frac{\nu_m}{\nu_m - 2}$
- since $1/q \ E(Q) = \frac{\nu}{\nu - 2}$
  we get $\nu = \frac{2E(Q)}{E(Q)-q}$
Using SAS PROC MIXED for the Analysis of Continuous Response of Repeated Measurements

Approximation for the degrees of freedom

Kenward-Roger-Method

- modification of the estimator of $\text{Var}(\hat{\beta})$
  the bias of the new estimator $\text{Var}^*(\hat{\beta})$ is smaller

- modification of the test statistic
  the test statistic is
  \[ F^* = \delta / q (C\hat{\beta})' (C\text{Var}^*(\hat{\beta}))^{-1} (C\hat{\beta}) \] with $q = \text{rank}(C)$

- $\nu = 4 + \frac{q+2}{q\gamma - 1}$, \( \delta = \frac{\nu}{E(F^*)(\nu - 2)} \), \( \gamma = \frac{\text{Var}(F^*)}{2 E(F^*)} \)

- but the complete derivation of $\text{Var}(F^*)$ and $E(F^*)$ is not described in the literature
Between-Within-Method

- this method divides the residual degrees of freedom into between-subject and within-subject values
- effects that do not change within subjects are assigned the between-subject values (for example sex)
- all others are assigned the within-subject
Discussion

- residual-method and between-within-method ignore the covariance structure
- for small samples the method of Kenward and Roger is better due to correction of the variance estimator of $\hat{\beta}$ and of the test statistic
- but the method of Kenward and Roger is not completely described in the literature
- with more than 400 degrees of freedom it is possible to approximate the degrees of freedom
  - in the t-test with the normal distribution
  - in the F-test for $\nu_1$ an $\nu_2$ with $\frac{\chi^2_{\nu_1}}{\nu_1}$
Data for the example

- data from a multicentre, open study for patients with rheumatoid arthritis
- the patients were observed for at least 12 weeks, with an optional extension period thereafter
- response: the activity of the rheumatoid arthritis was measured with the Disease Activity Score (DAS)
- explanatory variables: sex, Health Assessment Questionnaire Index (HAQ), C-Reactive Protein (CRP), number of previous therapies with Disease Modifying Antirheumatic Drugs (DMARDs)
- 6430 patients with 29571 observations in at most 36 weeks (visit 8 with 3042 patients)
Methods

- the parameter vector will be estimated with the REML-method
- the Compound Symmetry, the First-Order Autoregressiv and the Unstructured covariance matrices will be used
- the Satterthwaite, the Kenward and Roger, the Between-Within and the Residual Method will be used for the approximation of the degrees of freedom
Programmcode in SAS PROC MIXED

1  PROC MIXED DATA = daten METHOD = REML;
   CLASS patient visit sex haqba_kat
       crpba_kat pdmard_kat ;
   MODEL das28kat = visit sex haqba_kat
       crpba_kat pdmard_kat
       / DDFM = SATTERTH CL INTERCEPT ;
   REPEATED / TYPE = UN SUBJECT = patient
       R RCORR ;
   RUN ;

DDFM = SATTERTH for Satterthwaite, KENWARDROGER for Kenward and Roger, RESIDUAL for Residual, BETWITHIN for Between-Within TYPE= CS for Compound Symmetry, AR(1) for Autoregressive(1), UN for unstructured
Using SAS PROC MIXED for the Analysis of Continuous Response of Repeated Measurements

Example

The estimated covariance matrix

- **Compound Symmetry**
  
  $\begin{pmatrix}
  1,82 & 1,14 & 1,14 & \ldots \\
  1,14 & 1,82 & 1,14 & \ldots \\
  1,14 & 1,14 & 1,82 & \ldots \\
  \vdots & \vdots & \vdots & \ddots
  \end{pmatrix}
  $

- **First-Order Autoregressive**
  
  $\begin{pmatrix}
  1,80 & 1,28 & 0,91 & \ldots \\
  1,28 & 1,80 & 1,28 & \ldots \\
  0,91 & 1,28 & 1,80 & \ldots \\
  \vdots & \vdots & \vdots & \ddots
  \end{pmatrix}
  $

- **Unstructured**
  
  $\begin{pmatrix}
  2,12 & 1,44 & 1,21 & \ldots \\
  1,44 & 1,87 & 1,28 & \ldots \\
  1,21 & 1,28 & 1,74 & \ldots \\
  \vdots & \vdots & \vdots & \ddots
  \end{pmatrix}
  $
### Estimated contrasts

<table>
<thead>
<tr>
<th></th>
<th>CS</th>
<th>AR(1)</th>
<th>Un</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Contrast between</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visit 8 and 3</td>
<td>−1,0</td>
<td>0,020</td>
<td>−1,0</td>
</tr>
<tr>
<td>Visit 7 and 3</td>
<td>−1,0</td>
<td>0,017</td>
<td>−1,0</td>
</tr>
<tr>
<td>Visit 6 and 3</td>
<td>−0,9</td>
<td>0,016</td>
<td>−0,9</td>
</tr>
<tr>
<td>Visit 5 and 3</td>
<td>−0,7</td>
<td>0,015</td>
<td>−0,7</td>
</tr>
<tr>
<td>Visit 4 and 3</td>
<td>−0,4</td>
<td>0,015</td>
<td>−0,4</td>
</tr>
<tr>
<td>male and female</td>
<td>−0,2</td>
<td>0,037</td>
<td>−0,2</td>
</tr>
<tr>
<td>high HAQ-Index and low</td>
<td>0,7</td>
<td>0,030</td>
<td>0,7</td>
</tr>
<tr>
<td>high CRP and low</td>
<td>0,4</td>
<td>0,029</td>
<td>0,4</td>
</tr>
<tr>
<td>min. 5 DMARDs and max. 1</td>
<td>0,2</td>
<td>0,045</td>
<td>0,2</td>
</tr>
<tr>
<td>4 DMARDs and max. 1</td>
<td>0,0</td>
<td>0,048</td>
<td>0,0</td>
</tr>
<tr>
<td>3 DMARDs and max. 1</td>
<td>−0,1</td>
<td>0,044</td>
<td>−0,1</td>
</tr>
<tr>
<td>2 DMARDs and max. 1</td>
<td>−0,2</td>
<td>0,042</td>
<td>−0,2</td>
</tr>
</tbody>
</table>
## Degrees of Freedom with a Compound Symmetry Structure

<table>
<thead>
<tr>
<th></th>
<th>Residual</th>
<th>Between</th>
<th>Satterth</th>
<th>Kenward</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td></td>
<td></td>
<td>7092</td>
<td>7092</td>
</tr>
<tr>
<td>Contrast between</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visit 8 and 3</td>
<td>30000</td>
<td>23000</td>
<td>24000</td>
<td>24000</td>
</tr>
<tr>
<td>Visit 7 and 3</td>
<td>30000</td>
<td>23000</td>
<td>24000</td>
<td>24000</td>
</tr>
<tr>
<td>Visit 6 and 3</td>
<td>30000</td>
<td>23000</td>
<td>24000</td>
<td>24000</td>
</tr>
<tr>
<td>Visit 5 and 3</td>
<td>30000</td>
<td>23000</td>
<td>23000</td>
<td>23000</td>
</tr>
<tr>
<td>Visit 4 and 3</td>
<td>30000</td>
<td>23000</td>
<td>23000</td>
<td>23000</td>
</tr>
<tr>
<td>Male and female</td>
<td>30000</td>
<td>6396</td>
<td>6306</td>
<td>6306</td>
</tr>
<tr>
<td>High HAQ-Index and low</td>
<td>30000</td>
<td>6396</td>
<td>6294</td>
<td>6294</td>
</tr>
<tr>
<td>High CRP and low</td>
<td>30000</td>
<td>6396</td>
<td>6318</td>
<td>6318</td>
</tr>
<tr>
<td>Min. 5 DMARDs and max. 1</td>
<td>30000</td>
<td>6396</td>
<td>6328</td>
<td>6328</td>
</tr>
<tr>
<td>4 DMARDs and max. 1</td>
<td>30000</td>
<td>6396</td>
<td>6299</td>
<td>6299</td>
</tr>
<tr>
<td>3 DMARDs and max. 1</td>
<td>30000</td>
<td>6396</td>
<td>6284</td>
<td>6284</td>
</tr>
<tr>
<td>2 DMARDs and max. 1</td>
<td>30000</td>
<td>6396</td>
<td>6295</td>
<td>6295</td>
</tr>
</tbody>
</table>
### Degrees of Freedom with a AR(1) Structure

<table>
<thead>
<tr>
<th></th>
<th>Residual</th>
<th>Between</th>
<th>Satterth</th>
<th>Kenward</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>30000</td>
<td>6396</td>
<td>8172</td>
<td>8172</td>
</tr>
<tr>
<td>Contrast between</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visit 8 and 3</td>
<td>30000</td>
<td>23000</td>
<td>28000</td>
<td>28000</td>
</tr>
<tr>
<td>Visit 7 and 3</td>
<td>30000</td>
<td>23000</td>
<td>29000</td>
<td>29000</td>
</tr>
<tr>
<td>Visit 6 and 3</td>
<td>30000</td>
<td>23000</td>
<td>29000</td>
<td>29000</td>
</tr>
<tr>
<td>Visit 5 and 3</td>
<td>30000</td>
<td>23000</td>
<td>27000</td>
<td>27000</td>
</tr>
<tr>
<td>Visit 4 and 3</td>
<td>30000</td>
<td>23000</td>
<td>22000</td>
<td>22000</td>
</tr>
<tr>
<td>male and female</td>
<td>30000</td>
<td>6396</td>
<td>7099</td>
<td>7099</td>
</tr>
<tr>
<td>high HAQ-Index and low</td>
<td>30000</td>
<td>6396</td>
<td>7074</td>
<td>7074</td>
</tr>
<tr>
<td>high CRP and low</td>
<td>30000</td>
<td>6396</td>
<td>7110</td>
<td>7110</td>
</tr>
<tr>
<td>min. 5 DMARDs and max. 1</td>
<td>30000</td>
<td>6396</td>
<td>7133</td>
<td>7133</td>
</tr>
<tr>
<td>4 DMARDs and max. 1</td>
<td>30000</td>
<td>6396</td>
<td>7081</td>
<td>7081</td>
</tr>
<tr>
<td>3 DMARDs and max. 1</td>
<td>30000</td>
<td>6396</td>
<td>7052</td>
<td>7052</td>
</tr>
<tr>
<td>2 DMARDs and max. 1</td>
<td>30000</td>
<td>6396</td>
<td>7051</td>
<td>7051</td>
</tr>
</tbody>
</table>
### Degrees of Freedom with an unstructured covariance matrix

<table>
<thead>
<tr>
<th></th>
<th>Residual</th>
<th>Between</th>
<th>Satterth</th>
<th>Kenward</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>30000</td>
<td>6396</td>
<td>6510</td>
<td>6510</td>
</tr>
<tr>
<td>Contrast between</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visit 8 and 3</td>
<td>30000</td>
<td>6396</td>
<td>4718</td>
<td>4718</td>
</tr>
<tr>
<td>Visit 7 and 3</td>
<td>30000</td>
<td>6396</td>
<td>6861</td>
<td>6861</td>
</tr>
<tr>
<td>Visit 6 and 3</td>
<td>30000</td>
<td>6396</td>
<td>8722</td>
<td>8722</td>
</tr>
<tr>
<td>Visit 5 and 3</td>
<td>30000</td>
<td>6396</td>
<td>9960</td>
<td>9960</td>
</tr>
<tr>
<td>Visit 4 and 3</td>
<td>30000</td>
<td>6396</td>
<td>9491</td>
<td>9491</td>
</tr>
<tr>
<td>male and female</td>
<td>30000</td>
<td>6396</td>
<td>6246</td>
<td>6246</td>
</tr>
<tr>
<td>high HAQ-Index and low</td>
<td>30000</td>
<td>6396</td>
<td>6183</td>
<td>6183</td>
</tr>
<tr>
<td>high CRP and low</td>
<td>30000</td>
<td>6396</td>
<td>6238</td>
<td>6238</td>
</tr>
<tr>
<td>min. 5 DMARDs and max. 1</td>
<td>30000</td>
<td>6396</td>
<td>6275</td>
<td>6275</td>
</tr>
<tr>
<td>4 DMARDs and max. 1</td>
<td>30000</td>
<td>6396</td>
<td>6192</td>
<td>6192</td>
</tr>
<tr>
<td>3 DMARDs and max. 1</td>
<td>30000</td>
<td>6396</td>
<td>6146</td>
<td>6146</td>
</tr>
<tr>
<td>2 DMARDs and max. 1</td>
<td>30000</td>
<td>6396</td>
<td>6119</td>
<td>6119</td>
</tr>
</tbody>
</table>
Comparison of the AIC and $R^2_{\text{adj.}}$.

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>$R^2_{\text{adj.}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compound Symmetry</td>
<td>86024,4</td>
<td>0,16</td>
</tr>
<tr>
<td>AR(1)</td>
<td>85067,4</td>
<td>0,16</td>
</tr>
<tr>
<td>Unstructured</td>
<td>83975,2</td>
<td>0,16</td>
</tr>
</tbody>
</table>
Conclusions

- GLS- and ML-Estimator for $\beta$ are identical
- the REML-Estimator for the covariance matrix has a smaller bias than the ML-Estimator
- because of simulation studies the method of Kenward and Roger should be used with small samples but this method is not completely described in the literature
- in big samples the degrees of freedom can be approximated with a normal distribution or the quotient of a $\chi^2$-distribution and its degrees of freedom
References I

Statistical Method for the Analysis of Repeated Measurements
Springer-Verlag, New York

Giesbrecht, F. and Burns, J. (1985):
Two-Stage Analysis Based on a Mixed Model: Large-Sample
Asymptotic Theory and Small-Sample Simulation Results
Biometrics, 41, 477–486

Small Sample Inference for Fixed Effects from Restricted
Maximum Likelihood
Biometrics, 53, 983–997
References

SAS Institute Inc., Cary

In *Proceedings of the Twenty-Sixth Annuals SAS Users Group International Conference*, paper 262