The Use of Score Tests for Inference on Variance Components

Geert Molenberghs

Center for Statistics
LUC, Diepenbeek, Belgium

Geert Verbeke

Biostatistical Centre
KUL, Leuven, Belgium

Winter Workshop on Longitudinal Data Analysis, Florida, January 8, 2005

Overview

- *Linear mixed model: marginal* versus *hierarchical view*
- *Implications for likelihood ratio test*
- *Results for the score test case*
- *Results for scalar and vector case*
- *Application:* The rat data

A Simple Linear Mixed Model

$$y_{ij} = \boldsymbol{x_{ij}}'\boldsymbol{\beta} + b_i + \varepsilon_{ij}$$

$$\begin{cases} b_i \sim N(0, \tau^2) \\ \\ \varepsilon_{ij} \sim N(0, \sigma^2) \end{cases} \text{ind.}$$

$\Longrightarrow$

$$\boldsymbol{y_i} = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{in_i} \end{pmatrix} \sim N\left[X_i\boldsymbol{\beta}, \begin{pmatrix} \sigma^2+\tau^2 & \cdots & \tau^2 \\ \vdots & \ddots & \vdots \\ \tau^2 & \cdots & \sigma^2+\tau^2 \end{pmatrix}\right]$$

Hierarchical:

$$\begin{cases} \sigma^2 > 0 \\ \tau^2 \geq 0 \end{cases}$$

$\Longrightarrow$

Marginal: $V_i = \tau^2 J_{n_i} + \sigma^2 I_{n_i}$ PD

The General Linear Mixed Model

$$\boldsymbol{y_i} = X_i\boldsymbol{\beta} + Z_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i$$

$$\left\{\begin{array}{l}\boldsymbol{b}_i \sim N(\mathbf{0}, D) \\ \\ \boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \Sigma_i)\end{array}\right. \quad \text{ind.}$$

$\Longrightarrow$

$$\boldsymbol{y_i} = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{in_i} \end{pmatrix} \sim N\left[X_i\boldsymbol{\beta} \ , \ V_i = Z_iDZ_i' + \Sigma_i \right]$$

Hierarchical:

$$\left\{\begin{array}{ll}\Sigma_i & \text{PD} \\ D & \text{PD}\end{array}\right.$$

$\Longrightarrow$

Marginal: $V_i = Z_iDZ_i' + \Sigma_i$ PD

Marginal $\neq$ Hierarchical

- Balanced data, two measurements per subject ($n_i = 2$)

- **Model 1** can follow from a hierarchical model:

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} = N \left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix} \right]$$

- **Model 2** does not follow from a hierarchical model:

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} = N \left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} 2 & -1 \\ -1 & 3 \end{pmatrix} \right]$$

Marginal $\neq$ Hierarchical

- Balanced data, two measurements per subject ($n_i = 2$)

- **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*

$$V = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} = \begin{pmatrix} d_1 + \sigma^2 & d_1 \\ d_1 & d_1 + d_2 + \sigma^2 \end{pmatrix}$$

Hierarchical versus Marginal: Testing

- **Marginal:**
  - ▷ Typically two-sided
  - ▷ Hypotheses in compound-symmetric case: $\boxed{\tau^2 = 0 \quad \textit{vs.} \quad \tau^2 \neq 0}$
  - ▷ Null distribution for likelihood ratio and score test: $\boxed{\chi^2_1}$

- **Hierarchical:**
  - ▷ Must be one-sided
  - ▷ Hypotheses in compound-symmetric case: $\boxed{\tau^2 = 0 \quad \textit{vs.} \quad \tau^2 > 0}$
  - ▷ Null distribution for likelihood ratio and . . .: $\boxed{\frac{1}{2}\chi^2_0 + \frac{1}{2}\chi^2_1}$

Two Questions

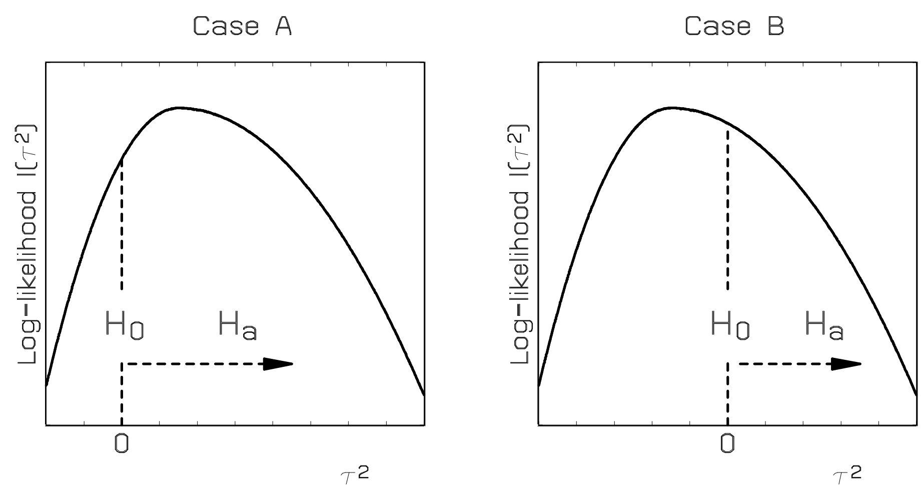
Vector parameter situation ?

and/or

Score test situation ?

One-sided, Scalar, Likelihood Ratio Test

$$T_{LR} = 2 \ln \left[\frac{\max_{H_0} \ell(\tau^2)}{\max_{H_a} \ell(\tau^2)} \right].$$



Case A	Case B
$P(T_{LR} > c \mid H_0, \hat{\tau}^2 \geq 0) P(\hat{\tau}^2 \geq 0 \mid H_0) = 0.5 P(\chi^2_1 > c)$	$P(T_{LR} > c \mid H_0, \hat{\tau}^2 < 0) P(\hat{\tau}^2 < 0 \mid H_0) = 0.5 P(\chi^2_0 > c)$
$P(T_{LR} > c \mid H_0) = \frac{1}{2} P(\chi^2_1 > c) + \frac{1}{2} P(\chi^2_0 > c)$	

More general: Nelder (1954), Chernoff (1954),
Self & Liang (1987), Stram & Lee (1994, 1995),
Shapiro (1988), Raubertas, Lee & Nordheim (1986)

Confusion About the Score Test

"No boundary problem since no fitting under H_a"

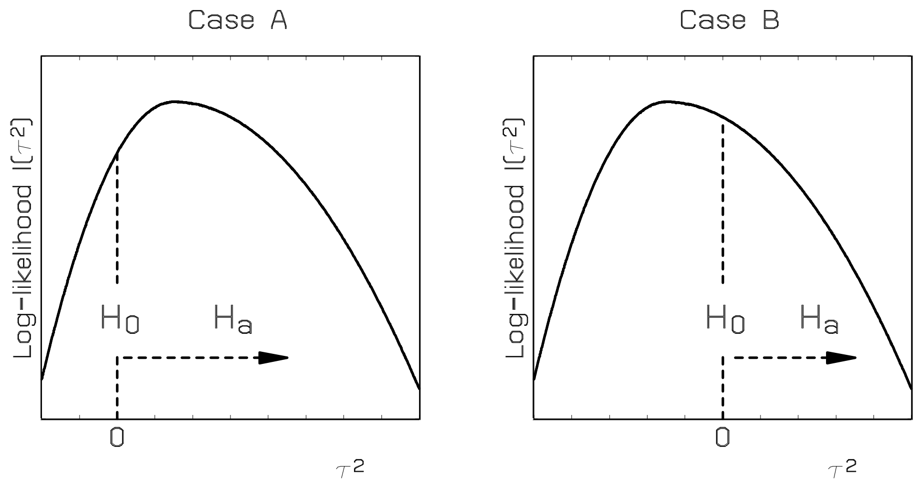
"LR and score not necessarily equivalent under non-standard conditions"

Various Views on Score Test Case

Unconstrained, two-sided view	Implicit	Jacqmin-Gadda & Commenges (1995) Lin (1997) le Cessie & van Houwelingen (1995)
	Explicit	Paul & Islam (1995)
Constrained, one-sided view		Silvapulle & Silvapulle (1995) Hall & Præstgaard (2001) Verbeke & Molenberghs (2003)

One-sided, Scalar, Score Test

$$T_S = \left[\left. \frac{\partial \ell(\tau^2)}{\partial \tau^2} \right|_{\tau^2=0} \right]^2 \left[- \left. \frac{\partial^2 \ell(\tau^2)}{\partial \tau^2 \partial \tau^2} \right|_{\tau^2=0} \right]^{-1}$$



Case A	**Case B**
$T_S = \left[\left. \frac{\partial \ell(\tau^2)}{\partial \tau^2} \right\|_{\tau^2=0} \right]^2 \left[- \left. \frac{\partial^2 \ell(\tau^2)}{\partial \tau^2 \partial \tau^2} \right\|_{\tau^2=0} \right]^{-1}$	$T_S = 0$
$P(T_S > c \| H_0, \hat{\tau}^2 \geq 0) P(\hat{\tau}^2 \geq 0 \| H_0) = 0.5 P(\chi^2_1 > c)$	$P(T_S > c \| H_0, \hat{\tau}^2 < 0) P(\hat{\tau}^2 < 0 \| H_0) = 0.5 P(\chi^2_0 > c)$
$P(T_S > c \| H_0) = \frac{1}{2} P(\chi^2_1 > c) + \frac{1}{2} P(\chi^2_0 > c)$	

Likelihood Ratio Test: General Results

- Historical: Nelder (1954)

  Chernoff (1954)

- Fundamental result: Self and Liang (1987)

- Variance components in mixed models: Stram and Lee (1994, 1995)

- Reparametrization (e.g., Cholesky decomposition) does not help

- Fairly general hypotheses: D_k *versus* D_{k+1}

- Null distribution for likelihood ratio test: $\frac{1}{2}\chi^2_k + \frac{1}{2}\chi^2_{k+1}$

- More general: Shapiro (1988)

  Raubertas, Lee, and Nordheim (1986)

Score Test: General Results

$$H_0 : \boldsymbol{\psi} = \mathbf{0} \qquad \textit{versus} \qquad H_a : \boldsymbol{\psi} \in \mathcal{C}$$

- $\mathcal{C}$: closed, convex cone in Euclidean space; vertex at origin

- Silvapulle & Silvapulle (1995)

- **A one-sided score test statistic:**

$$T_S := \boldsymbol{Z}'_N H^{-1}_{\psi\psi}(\widehat{\boldsymbol{\theta}}_H)\boldsymbol{Z}_N \;-\; \inf\big\{(\boldsymbol{Z}_N - \boldsymbol{b})' H^{-1}_{\psi\psi}(\widehat{\boldsymbol{\theta}}_H)(\boldsymbol{Z}_N - \boldsymbol{b}) | \boldsymbol{b} \in \mathcal{C}\big\}$$

One-sided Score Test

- Simplified notation:

$$T_S = \boldsymbol{Z}' H^{-1} \boldsymbol{Z} \ - \ \inf \left\{ (\boldsymbol{Z} - \boldsymbol{b})' H^{-1} (\boldsymbol{Z} - \boldsymbol{b}) | \boldsymbol{b} \in \mathcal{C} \right\}$$

- Special case: $\mathcal{C} = \mathbb{R}_+$

  Form of statistic:

$$\begin{cases} Z \geq 0 \ : \ T_S = \boldsymbol{Z} H^{-1} \boldsymbol{Z} \\ Z < 0 \ : \ T_S = 0 \end{cases}$$

- Equivalence: $T_{LR} = T_S + o_p(1)$

p values

from weighted sums of χ^2 variables

Structure of D	**Hypotheses**	**Null distribution**
Unstructured	D_k *vs.* D_{k+1}	$\frac{1}{2}\chi^2_k + \frac{1}{2}\chi^2_{k+1}$
Variance components	D_k diagonal *vs.* $D_{k+k'}$ diagonal	$\sum_{m=0}^{k'} 2^{-k'} \binom{k'}{m} \chi^2_m$

Estimation versus Testing in Mixed-Models Setting

	Estimation	**Testing**
Hierarchical	In line with intuition	*Extended theory*
Marginal	*Confusion about negative variance*	Classical theory

Summary of Result for Score Test

▷ General:

$$H_0 : \boldsymbol{\psi} = \mathbf{0} \qquad \text{vs.} \qquad H_a : \boldsymbol{\psi} \in \mathcal{C}$$

$\mathcal{C}$: closed, convex cone in Euclidean space; vertex at origin

▷ One-sided score test:

$$T_S = \boldsymbol{Z}' H^{-1} \boldsymbol{Z} - \inf\left\{(\boldsymbol{Z} - \boldsymbol{b})' H^{-1} (\boldsymbol{Z} - \boldsymbol{b}) | \boldsymbol{b} \in \mathcal{C}\right\}$$

▷ Equivalence:

$$T_{LR} = T_S + o_p(1)$$

▷ **But:** Calculation of one-sided T_S less straightforward

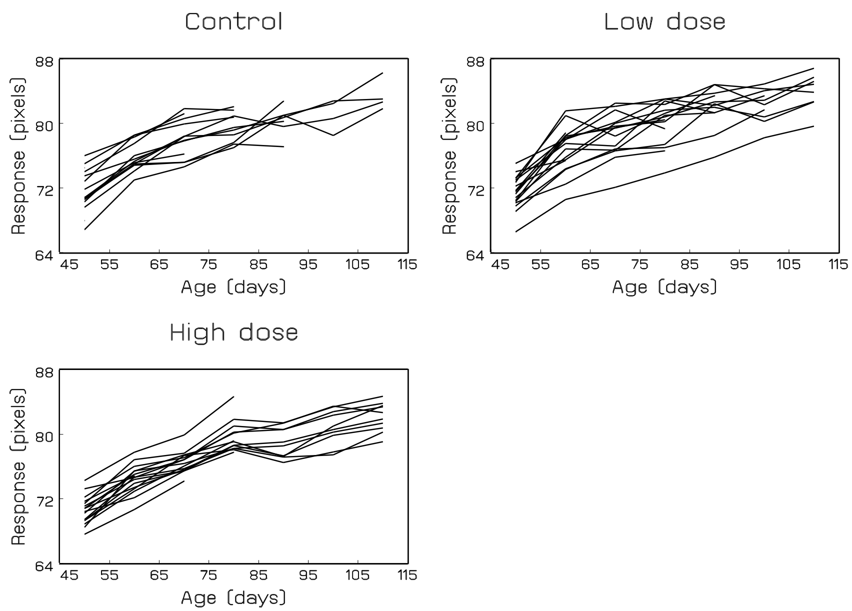
The Rat Data

- Department of 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)

The Rat Data



- 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, reflecting height of skull.

Model for Rat Data

$$Y_{ij} = \begin{cases} \beta_0 + b_{1i} + (\beta_1 + b_{2i})t_{ij} + \varepsilon_{ij}, & \text{if low dose,} \\ \beta_0 + b_{1i} + (\beta_2 + b_{2i})t_{ij} + \varepsilon_{ij}, & \text{if high dose,} \\ \beta_0 + b_{1i} + (\beta_3 + b_{2i})t_{ij} + \varepsilon_{ij}, & \text{if control.} \end{cases}$$

- $t = \ln(1 + (\text{Age} - 45)/10)$

- *Four sub-models:*
  - ▷ **Model 1:** no random effects
  - ▷ **Model 2:** random intercepts only ($\Rightarrow$ marginal: compound symmetry)
  - ▷ **Model 3:** random intercepts and random slopes; uncorrelated
  - ▷ **Model 4:** random intercepts and random slopes; correlated

Four Models

Model	Structure	One-sided	Two-sided
1	$D = \begin{pmatrix} 0 \end{pmatrix}$	$\widehat{D} = \begin{pmatrix} 0 \end{pmatrix}$	$\widehat{D} = \begin{pmatrix} 0 \end{pmatrix}$
2	$D = \begin{pmatrix} d_{11} \end{pmatrix}$	$\widehat{D} = \begin{pmatrix} 3.44 \end{pmatrix}$	$\widehat{D} = \begin{pmatrix} 3.44 \end{pmatrix}$
3	$D = \begin{pmatrix} d_{11} & 0 \\ 0 & d_{22} \end{pmatrix}$	$\widehat{D} = \begin{pmatrix} 3.44 & 0 \\ 0 & 0.00 \end{pmatrix}$	$\widehat{D} = \begin{pmatrix} 3.77 & 0 \\ 0 & -0.17 \end{pmatrix}$
4	$D = \begin{pmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{pmatrix}$	$\widehat{D} = \begin{pmatrix} 3.41 & 0.01 \\ 0.01 & 0.00 \end{pmatrix}$	$\widehat{D} = \begin{pmatrix} 2.83 & 0.48 \\ 0.48 & -0.33 \end{pmatrix}$

Results

Mod		Ref	LR test		Score test	
			One-sided	Two-sided	One-sided	Two-sided
1	$D = 0$					
2	$D = \left(d_{11} \right)$ $\widehat{D} = \left(3.44 \right)$	1	$1102.7 - 928.7 = 174$ $\frac{1}{2}\chi^2_0 + \frac{1}{2}\chi^2_1$ $p < 0.0001$	$1102.7 - 928.7 = 174$ χ^2_1 $p < 0.0001$	$27.15 - 0.0 = 27.15$ $\frac{1}{2}\chi^2_0 + \frac{1}{2}\chi^2_1$ $p < 0.0001$	27.15 χ^2_1 $p < 0.0001$
3	$D = \begin{pmatrix} d_{11} & 0 \\ 0 & d_{22} \end{pmatrix}$ $\widehat{D} = \begin{pmatrix} 3.77 & 0 \\ 0 & -0.17 \end{pmatrix}$	2	$928.7 - 928.7 = 0.0$ $\frac{1}{2}\chi^2_0 + \frac{1}{2}\chi^2_1$ $p = 1.0000$	$928.7 - 927.4 = 1.3$ χ^2_1 $p = 0.2542$	$1.67 - 1.67 = 0$ $\frac{1}{2}\chi^2_0 + \frac{1}{2}\chi^2_1$ $p = 1.0000$	1.67 χ^2_1 $p = 0.1963$
4	$D = \begin{pmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{pmatrix}$ $\widehat{D} = \begin{pmatrix} 2.83 & 0.48 \\ 0.48 & -0.33 \end{pmatrix}$	2	$928.7 - 928.6 = 0.1$ $\frac{1}{2}\chi^2_1 + \frac{1}{2}\chi^2_2$ $p = 0.8515$	$928.7 - 925.8 = 2.9$ χ^2_2 $p = 0.2346$	$2.03 - 1.93 = 0.1$ $\frac{1}{2}\chi^2_1 + \frac{1}{2}\chi^2_2$ $p = 0.8515$	2.03 χ^2_2 $p = 0.3624$

General Testing Result

- One-sided testing holds generally for vector (random-effects) case:

$$\boldsymbol{\psi} = \mathbf{0} \qquad \textit{versus} \qquad \boldsymbol{\psi} \in \mathcal{C}$$

- Covers important constrained cases in hierarchical models

- Likelihood ratio and score test statistic have same null distribution

- Critical level determined from mixtures of χ^2 variables

Unified Theory

- Same null distribution for:
  - ▷ Likelihood ratio test
  - ▷ Score test
  - ▷ *Wald test*
- Critical level determined from mixtures of χ^2 variables