STAT 574 Linear and Nonlinear Mixed Models

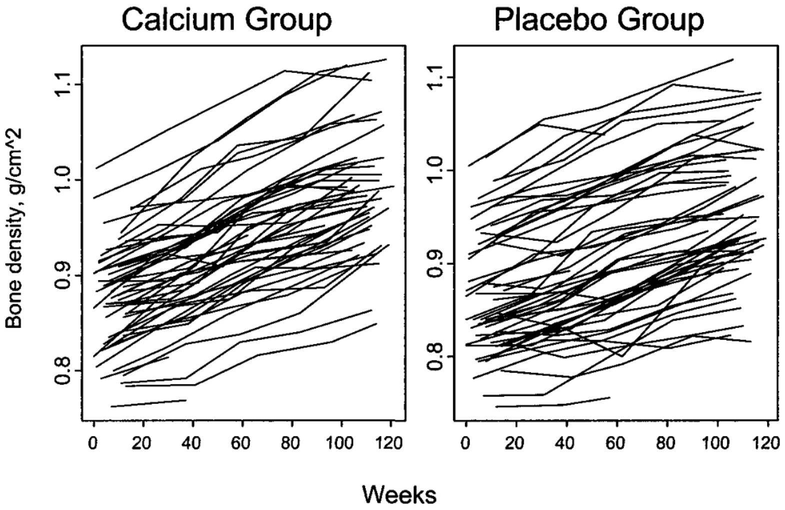
Lecture 4: Growth Curve Models

Chencheng Cai

Washington State University

Linear Growth Curve Model

"Growth Curves"



Linear Growth Curve Model

The dataset on the previous slide:

- Bone density curve over time (2 years)
- 52 persons in the Calcium group and 53 persons in the placebo group

Observations:

- For each person, the bone density curve looks linear in time.
- Intercepts and slopes may differ from person to person.

Linear Growth Curve Model

- To model the linear trend of bone density for each person:

$$y_{ij} = a_{i0} + a_{i1}t_{ij} + \epsilon_{ij}$$

  - t_{ij}: time of j-th measurement for person i.
  - y_{ij}: bone density for person i at time t_{ij}.
  - a_{i0}: intercept for person i.
  - a_{i1}: slope for person i.
  - ϵ_{ij}: random noise.
- Heterogeneity in intercept:

$$a_{i0} = \beta_0 + b_{i0}$$

  - β_0: populational average intercept
  - $b_{i0} \sim \mathcal{N}(0, \sigma^2 d_{00})$: random deviation for person i

Linear Growth Curve Model

- Effect of Calcium on the slope:

$$a_{i1} = \beta_1 + \beta_2 C_i + b_{i1}$$

  - β_1: populational average slope for placebo
  - β_2: effect of Calcium
  - C_i: indicator on person i taking Calcium
  - $b_{i1} \sim \mathcal{N}(0, \sigma^2 d_{11})$: random deviation

Linear Growth Curve Model

Complete model:

$$
\begin{aligned}
y_{ij} &= a_{i0} + a_{i1} t_{ij} + \epsilon_{ij} \\
a_{i0} &= \beta_0 + b_{i0} \\
a_{i1} &= \beta_1 + \beta_2 C_i + b_{i1}
\end{aligned}
$$

- Distributions:

$$
\begin{pmatrix} b_{i0} \\ b_{i1} \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \begin{bmatrix} d_{00} & d_{01} \\ d_{01} & d_{11} \end{bmatrix}\right), \epsilon_{ij} \sim (0, \sigma^2)
$$

- Independence:
  - Persons are independent.
  - Noises at different times are independent.
  - Noise and random effect are independent.

Linear Growth Curve Model

Vectorized version:

$$\boldsymbol{y}_i = a_{i0}\mathbf{1} + a_{i1}\boldsymbol{t}_i + \boldsymbol{\epsilon}_i$$
$$\begin{pmatrix} a_{i0} \\ a_{i1} \end{pmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & C_i \end{bmatrix} \boldsymbol{\beta} + \boldsymbol{b}_i$$

- $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I})$
- $\boldsymbol{b}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{D})$

Linear Growth Curve Model

Linear Growth Curve (LGC) Model

$$\boldsymbol{y}_i = \boldsymbol{Z}_i \boldsymbol{a}_i + \boldsymbol{\epsilon}_i$$
$$\boldsymbol{a}_i = \boldsymbol{A}_i \boldsymbol{\beta} + \boldsymbol{b}_i$$

- $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I})$
- $\boldsymbol{b}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{D})$

Is a LGC model a LME model?

Linear Growth Curve Model

A LGC model is a LME model:

$$\boldsymbol{y}_i = \boldsymbol{Z}_i \boldsymbol{A}_i \boldsymbol{\beta} + \boldsymbol{Z}_i \boldsymbol{b}_i + \boldsymbol{\epsilon}_i$$

- Design matrix for fixed effects: $\boldsymbol{X}_i = \boldsymbol{Z}_i \boldsymbol{A}_i$
- Fixed effect coefficients: $\boldsymbol{\beta}$
- Design matrix for random effects: $\boldsymbol{Z}_i$
- Random effect coefficients: $\boldsymbol{b}_i$

Linear Growth Curve Model — Special Cases

- When $\boldsymbol{Z}_I = \boldsymbol{I}$, LGC is a random-coefficient model
- $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B}\boldsymbol{Z} + \boldsymbol{\epsilon}$ is another LGC model.
- Growth curve models are sometimes called "latent growth curve" models.
- Growth curve models are related to structure equation modeling (SEM) in econometrics.

Linear Growth Curve Model — Practice

- Fama-French Three-Factor Model:

$$r_i = R_f + \beta_i(R_m - R_f) + b_{i1}\mathrm{SMB} + b_{i2}\mathrm{HML} + \epsilon_i$$

  - r_i: return of stock i .
  - R_f: risk-free rate
  - R_m: market return
  - SMB: Small minus big in capitalization
  - HML: High minus low in book-to-market ratio

Extend it to a LGC model!

LGC Model — Known matrix $\boldsymbol{D}$

If $\boldsymbol{D}$ is known or estimated, we have

$$\hat{\boldsymbol{\beta}}_{GLS} = \left(\sum_{i=1}^{N} \boldsymbol{A}_i^T \boldsymbol{Z}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{Z}_i \boldsymbol{A}_i \right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{A}_i^T \boldsymbol{Z}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{y}_i \right)$$

with covariance

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}_{GLS}) = \sigma^2 \left(\sum_{i=1}^{N} \boldsymbol{A}_i^T \boldsymbol{Z}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{Z}_i \boldsymbol{A}_i \right)^{-1}$$

LGC Model — Known matrix $\boldsymbol{D}$

Let $\boldsymbol{W}_i = \boldsymbol{D} + (\boldsymbol{Z}_i^T\boldsymbol{Z}_i)^{-1}$, we have

$$\begin{aligned}
\boldsymbol{Z}_i^T\boldsymbol{V}_i^{-1}\boldsymbol{Z}_i &= \boldsymbol{Z}_i^T(\boldsymbol{I} - \boldsymbol{Z}_i(\boldsymbol{D}^{-1} + \boldsymbol{Z}_i^T\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i^T)\boldsymbol{Z}_i \\
&= \boldsymbol{Z}_i^T\boldsymbol{Z}_i - \boldsymbol{Z}_i^T\boldsymbol{Z}_i(\boldsymbol{D}^{-1} + \boldsymbol{Z}_i^T\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i^T\boldsymbol{Z}_i \\
&= \boldsymbol{Z}_i^T\boldsymbol{Z}_i(\boldsymbol{D}^{-1} + \boldsymbol{Z}_i^T\boldsymbol{Z}_i)^{-1}\boldsymbol{D}^{-1} \\
&= \boldsymbol{Z}_i^T\boldsymbol{Z}_i(\boldsymbol{I} + \boldsymbol{Z}_i^T\boldsymbol{Z}_i\boldsymbol{D})^{-1} \\
&= \left((\boldsymbol{Z}_i^T\boldsymbol{Z}_i)^{-1} + \boldsymbol{D}\right)^{-1} = \boldsymbol{W}_i^{-1}
\end{aligned}$$

and similarly, $\boldsymbol{Z}_i^T\boldsymbol{V}_i^{-1}\boldsymbol{y}_i = \boldsymbol{W}_i^{-1}\boldsymbol{a}_i^0$ with $\boldsymbol{a}_i^0 = (\boldsymbol{Z}_i^T\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i^T\boldsymbol{y}$.
Then

$$\hat{\boldsymbol{\beta}}_{GLS} = \left(\sum_{i=1}^N \boldsymbol{A}_i^T\boldsymbol{W}_i^{-1}\boldsymbol{A}_i\right)^{-1}\left(\sum_{i=1}^N \boldsymbol{A}_i^T\boldsymbol{W}_i^{-1}\boldsymbol{a}_i^0\right) \quad \mathrm{Cov}(\hat{\boldsymbol{\beta}}_{GLS}) = \sigma^2\left(\sum_{i=1}^N \boldsymbol{A}_i^T\boldsymbol{W}_i^{-1}\boldsymbol{A}_i\right)^{-1}$$

LGC Model — Known matrix $\boldsymbol{D}$

Recall the LGC Model:

$$\boldsymbol{y}_i = \boldsymbol{Z}_i \boldsymbol{a}_i + \boldsymbol{\epsilon}_i$$
$$\boldsymbol{a}_i = \boldsymbol{A}_i \boldsymbol{\beta} + \boldsymbol{b}_i$$

We defined $\boldsymbol{a}_i^0 = (\boldsymbol{Z}_i^T \boldsymbol{Z}_i)^{-1} \boldsymbol{Z}_i^T \boldsymbol{y}$ — the OLS solution for the first equation!!
$\hat{\boldsymbol{\beta}}_{GLS}$ formula is the GLS solution for the second equation with covariance $\boldsymbol{W}_i$!!
To see it:

$$\boldsymbol{a}_i^0 = (\boldsymbol{Z}_i^T \boldsymbol{Z}_i)^{-1} \boldsymbol{Z}_i^T \boldsymbol{y} = \boldsymbol{a}_i + (\boldsymbol{Z}_i^T \boldsymbol{Z}_i)^{-1} \boldsymbol{Z}_i^T \boldsymbol{\epsilon}_i$$

and

$$\mathrm{Cov}(\boldsymbol{a}_i^0) = \sigma^2 (\boldsymbol{D} + (\boldsymbol{Z}_i^T \boldsymbol{Z}_i)^{-1}) = \sigma^2 \boldsymbol{W}_i$$

LGC Model — Known matrix $\boldsymbol{D}$

$$\boldsymbol{y}_i = \boldsymbol{Z}_i \boldsymbol{a}_i + \boldsymbol{\epsilon}_i$$
$$\boldsymbol{a}_i = \boldsymbol{A}_i \boldsymbol{\beta} + \boldsymbol{b}_i$$

Two-step estimation procedure:

- Estimate $\boldsymbol{a}_i^0$ for each i using OLS for the first equation.
- Estimate $\boldsymbol{\beta}$ using GLS for the second equation with covariance $\boldsymbol{W}_i$.

LGC Model — MLE

The MLE maximizes

$$\ell(\boldsymbol{\beta}, \sigma^2, \boldsymbol{D}) = -\frac{1}{2}\left\{N_T \log \sigma^2 + \sum_{i=1}^{N} \log|\boldsymbol{V}_i| + \sigma^{-2} \sum_{i=1}^{N} (\boldsymbol{y}_i - \boldsymbol{Z}_i \boldsymbol{A}_i \boldsymbol{\beta})^T \boldsymbol{V}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{Z}_i \boldsymbol{A}_i \boldsymbol{\beta})\right\}$$

or using the previous trick:

$$\ell(\boldsymbol{\beta}, \sigma^2, \boldsymbol{D}) = -\frac{1}{2}\left\{N_T \log \sigma^2 + \sum_{i=1}^{N} \log|\boldsymbol{W}_i| + \sigma^{-2}\left[S_0 + \sum_{i=1}^{N} (\boldsymbol{a}_i^0 - \boldsymbol{A}_i \boldsymbol{\beta})^T \boldsymbol{W}_i^{-1} (\boldsymbol{a}_i^0 - \boldsymbol{A}_i \boldsymbol{\beta})\right]\right\}$$

where $S_0 = \sum_{i=1}^{N} \|\boldsymbol{y}_i - \boldsymbol{Z}_i \boldsymbol{a}_i^0\|^2$.

LGC Model — MLE

If we profile out σ^2 with

$$\hat{\sigma}^2 = \frac{1}{N_T} \sum_{i=1}^{N} (\boldsymbol{a}_i^0 - \boldsymbol{A}_i \boldsymbol{\beta})^T \boldsymbol{W}_i^{-1} (\boldsymbol{a}_i^0 - \boldsymbol{A}_i \boldsymbol{\beta})$$

we have

$$\ell_p(\boldsymbol{\beta}, \boldsymbol{D}) = -\frac{1}{2} \left\{ \sum_{i=1}^{N} \log |\boldsymbol{W}_i| + N_T \log \left[S_0 + \sum_{i=1}^{N} (\boldsymbol{a}_i^0 - \boldsymbol{A}_i \boldsymbol{\beta})^T \boldsymbol{W}_i^{-1} (\boldsymbol{a}_i^0 - \boldsymbol{A}_i \boldsymbol{\beta}) \right] \right\}$$

LGC Model — MLE

Two-step procedure without knowing $\boldsymbol{D}$:

- Fit the first equation with OLS to get $\boldsymbol{a}_i^0$ for all i.
- Maximize $\ell_p(\boldsymbol{\beta}, \boldsymbol{D})$ with given $\boldsymbol{a}_i^0$'s.

LGC Model — MLE

Is $\hat{\boldsymbol{\beta}}_{GLS}$ unbiased?

$$\hat{\boldsymbol{\beta}}_{GLS} = \left(\sum_{i=1}^{N} \boldsymbol{A}_i^T \hat{\boldsymbol{W}}_i^{-1} \boldsymbol{A}_i \right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{A}_i^T \hat{\boldsymbol{W}}_i^{-1} \boldsymbol{a}_i^0 \right)$$

LGC Model — MLE

Is $\hat{\boldsymbol{\beta}}_{GLS}$ unbiased?

$$\hat{\boldsymbol{\beta}}_{GLS} = \left(\sum_{i=1}^{N} \boldsymbol{A}_i^T \hat{\boldsymbol{W}}_i^{-1} \boldsymbol{A}_i \right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{A}_i^T \hat{\boldsymbol{W}}_i^{-1} \boldsymbol{a}_i^0 \right)$$

- True, because $\hat{\boldsymbol{D}}$ is an even function of $\boldsymbol{b}_i$'s.

LGC Model — MLE

CLT for MLE:

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{H}),$$

where $\boldsymbol{H} = \lim_{N\to\infty} N^{-1} \left(\sum_{i=1}^{N} \boldsymbol{A}_i^T \boldsymbol{W}_i^{-1} \boldsymbol{A}_i \right)^{-1}$. (Deterministic Scheme)

LGC Model — MLE

Sketch of proof for CLT:

Theorem (Slutsky Theorem)

Let $\{g_n(\theta), n = 1, 2, \dots\}$ *be a sequence of uniformly differentiable functions on* $\mathbb{R}^k$ *with* $\lim_{n\to\infty} g_n = g$. *If* $\theta_n \xrightarrow{P} \theta$ *in* $\mathbb{R}^k$, *then* $g_n(\theta_n) \xrightarrow{P} g(\theta)$.

- With Slutsky Theorem, we can show $\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{D}}) \xrightarrow{P} \hat{\boldsymbol{\beta}}(\boldsymbol{D})$ for $\hat{\boldsymbol{D}} \xrightarrow{P} \boldsymbol{D}$.
- On the other hand, we can show

$$\sqrt{N}(\hat{\boldsymbol{\beta}}(\boldsymbol{D}) - \boldsymbol{\beta}) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{H})$$

- The CLT is approved.