BIST P8110: Applied Regression II

Lecture 17: Linear Mixed Effects Model

Qixuan Chen

Department of Biostatistics
Columbia University

Spring 2017

This lecture's big ideas

- Linear mixed effects model
- Case study

Linear mixed effects model

- Random coefficients models
  - Random intercept model
  - Random slope model
- A more general model framework: linear mixed effects model
  - Random coefficients models are a special case of it
  - Rich set of models for characterizing repeated measurement data

Random intercept model

The random intercept model: $Y_{ij} = \beta_0 + \beta_1 X_{ij} + b_{0i} + \epsilon_{ij}$

- can be re-written as

$$\boldsymbol{Y}_i = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{Z}_i \boldsymbol{b}_i + \boldsymbol{\epsilon}_i$$

where $\boldsymbol{X}_i = \begin{bmatrix} 1 & x_{i1} \\ \vdots & \vdots \\ 1 & x_{in_i} \end{bmatrix}$, $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$

$$\boldsymbol{Z}_i = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \ \boldsymbol{b}_i = \begin{bmatrix} b_{0i} \end{bmatrix}, \ \boldsymbol{\epsilon}_i = \begin{bmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{in_i} \end{bmatrix}$$

Random slope model

The random intercept and slope model:

$Y_{ij} = \beta_0 + \beta_1 X_{ij} + b_{0i} + b_{1i} X_{ij} + \epsilon_{ij}$

- can be re-written as

$$\boldsymbol{Y}_i = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{Z}_i \boldsymbol{b}_i + \boldsymbol{\epsilon}_i$$

where $\boldsymbol{X}_i = \begin{bmatrix} 1 & x_{i1} \\ \vdots & \vdots \\ 1 & x_{in_i} \end{bmatrix}$, $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$

$$\boldsymbol{Z}_i = \begin{bmatrix} 1 & x_{i1} \\ \vdots & \vdots \\ 1 & x_{in_i} \end{bmatrix}, \boldsymbol{b}_i = \begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix}, \boldsymbol{\epsilon}_i = \begin{bmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{in_i} \end{bmatrix}$$

Linear mixed effects models

- A wide variety of subject-specific models of the form

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\beta + \boldsymbol{Z}_i\boldsymbol{b}_i + \epsilon_i$$

  by suitably defining $\boldsymbol{X}_i$, β, $\boldsymbol{Z}_i$ and $\boldsymbol{b}_i$.
- This model in its general form is known as the linear mixed effects models.

Linear mixed effects models

- With $\boldsymbol{Y}_i$ a $(n_i \times 1)$ vector of responses for the ith unit, $i = 1, \ldots, m$,

$$\boldsymbol{Y}_i = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{Z}_i \boldsymbol{b}_i + \epsilon_i$$

  - $\boldsymbol{X}_i$ is a $(n_i \times p)$ “design matrix" that characterizes the **systematic** part of the response
  - $\boldsymbol{\beta}$ is a $(p \times 1)$ vector of **fixed effects**
  - $\boldsymbol{Z}_i$ is a $(n_i \times k)$ “design matrix" that characterizes random variation in the response attributable to **between-subject** sources
  - $\boldsymbol{b}_i$ is a $(k \times 1)$ vector of **random effects**
  - ϵ_i is a $(n_i \times 1)$ vector of **within-subject deviation** characterizing variation due to sources like within-subject fluctuations and measurement error.

Assumptions on random variation

- The model components $\boldsymbol{b}_i$ and ϵ_i characterize the two sources of variation, between- and within-subjects.
- The usually assumptions
  - $\epsilon_i \sim \boldsymbol{N}_{n_i}(\boldsymbol{0}, \boldsymbol{R}_i)$, $\boldsymbol{R}_i$ is a $(n_i \times n_i)$ covariance matrix, characterizes variance and correlation due to within-subject sources
  - $\boldsymbol{b}_i \sim \boldsymbol{N}_k(\boldsymbol{0}, \boldsymbol{G})$, $\boldsymbol{G}$ is a $(k \times k)$ covariance matrix, characterizes variation due to between-subject sources
  - For $\boldsymbol{R}_i$ and $\boldsymbol{G}$, we must select some covariance structure, e.g.
    - variance components
    - compound symmetry
    - unstructured
    - autoregressive

Model estimation

- $\boldsymbol{R}_i$ and $\boldsymbol{G}$ can be estimated using likelihood-based methods
  - maximum likelihood (ML)
  - restricted/residual maximum likelihood (REML)
- We won't talk about how, but $\boldsymbol{\beta}$ and $\boldsymbol{b}_i$ can be estimated by solving some mixed model equations
  - the resulted $\hat{\boldsymbol{\beta}}$ is called the best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$
  - the resulted $\hat{\boldsymbol{b}}_i$ is called the best linear unbiased predictor (BLUP) of $\boldsymbol{b}_i$

Marginal and conditional means and residuals

- The marginal mean is $\boldsymbol{X}_i\beta$
- The conditional mean is $\boldsymbol{X}_i\beta + \boldsymbol{Z}_i\boldsymbol{b}_i$
- The marginal residual is $\boldsymbol{Y}_i - \boldsymbol{X}_i\hat{\beta}$
- The conditional residual is $\boldsymbol{Y}_i - \boldsymbol{X}_i\hat{\beta} - \boldsymbol{Z}_i\hat{\boldsymbol{b}}_i$

Case Study: Linear Mixed Effects Models

Project description

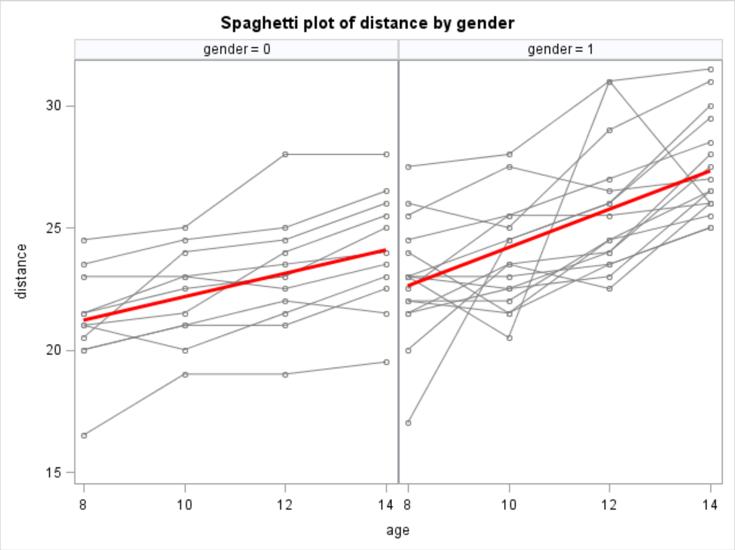
- A dental study dataset of 11 girls and 16 boys
  - obsno: observation number
  - child: child ID number
  - age: age of child at each measurement (8, 10, 12, 14)
  - distance: response
  - gender: gender indicator (0=girl, 1=boy)
- Goal: analysis of the dental study data
  - The model for each child is assumed to be a straight line
  - The intercepts and slopes may have different means depending on gender
  - We use the RANDOM and REPEATED statements in the MIXED procedure to fit models that make several different assumptions about the forms of the matrices $\boldsymbol{R}_i$ and $\boldsymbol{G}$

SAS code: spaghetti plot

```
/*Read in the dataset*/
data dental;
infile 'C:\dental.csv' delimiter = ',';
input obsno child age distance gender;
run;

/*generate spaghetti plots by gender*/
proc sgpanel data=dental NOAUTOLEGEND;
   title 'Spaghetti plot of distance by gender';
   panelby gender;
   series x=age y=distance / group=child BREAK
                             LINEATTRS = (COLOR= gray
                             PATTERN = 1 THICKNESS = 1) ;
   reg x=age y=distance /  LINEATTRS = (COLOR= red
                           PATTERN = 1 THICKNESS = 3)
                           MARKERATTRS = (COLOR= gray);
run;
```

SAS output: spaghetti plot



Model 1

- Model 1: random intercept model

$$Y_{ij} = \beta_0 + \beta_1 \times \text{age}_{ij} + \beta_2 \times \text{gender}_i + \beta_3 \times \text{gender}_i \times \text{age}_{ij} \\ + b_{0i} + \epsilon_{ij}$$

can be re-written as $\boldsymbol{Y}_i = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{Z}_i \boldsymbol{b}_i + \boldsymbol{\epsilon}_i$, where

$\boldsymbol{X}_i =$ $\boldsymbol{\beta} =$

$\boldsymbol{Z}_i =$ $\boldsymbol{b}_i =$

Model 1

- Model 1: random intercept model

$$
\begin{aligned}
Y_{ij} = & \beta_0 + \beta_1 \times \text{age}_{ij} + \beta_2 \times \text{gender}_i + \beta_3 \times \text{gender}_i \times \text{age}_{ij} \\
& + b_{0i} + \epsilon_{ij}
\end{aligned}
$$

$$
\text{where } \boldsymbol{G} = \tau^2, \ \boldsymbol{R}_i = \sigma^2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}
$$

SAS code: model 1

```
proc mixed method=ml data=dental;
  class gender child;
  model distance = age gender gender*age / solution;
  random intercept / subject=child g;
run;
```

SAS output: model 1

```
                     Model Information

Data Set                        WORK.DENTAL
Dependent Variable              distance
Covariance Structure            Variance Components
Subject Effect                  child
Estimation Method               ML
Residual Variance Method        Profile
Fixed Effects SE Method         Model-Based
Degrees of Freedom Method       Containment

                  Class Level Information

  Class     Levels    Values

  gender         2    0 1
  child         27    1 2 3 4 5 6 7 8 9 10 11 12 13
                      14 15 16 17 18 19 20 21 22 23
                      24 25 26 27

                        Dimensions

          Covariance Parameters             2
          Columns in X                      6
          Columns in Z Per Subject          1
          Subjects                         27
          Max Obs Per Subject               4
```

SAS output: model 1

Estimated G Matrix

Row	Effect	child	Col1
1	Intercept	1	3.0306

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
Intercept	child	3.0306
Residual		1.8746

Fit Statistics

-2 Log Likelihood	428.6
AIC (smaller is better)	440.6
AICC (smaller is better)	441.5
BIC (smaller is better)	448.4

SAS output: model 1

Solution for Fixed Effects

Effect	gender	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		16.3406	0.9631	25	16.97	<.0001
age		0.7844	0.07654	79	10.25	<.0001
gender	0	1.0321	1.5089	79	0.68	0.4960
gender	1	0	.	.	.	.
age*gender	0	-0.3048	0.1199	79	-2.54	0.0130

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
age	1	79	111.10	<.0001
gender	1	79	0.47	0.4960
age*gender	1	79	6.46	0.0130

Model 1: interpretation

- What is th estimated ICC?
- What is the estimated average intercept among boys?
- What is the estimaed average intercept among girls?
- What is the estimated slope among boys?
- What is the estimated slope among girls?

Model 2

- Model 2: random intercept model with different within-subject covariance matrix between boys and girls

$$\begin{aligned} Y_{ij} = & \beta_0 + \beta_1 \times \text{age}_{ij} + \beta_2 \times \text{gender}_i + \beta_3 \times \text{gender}_i \times \text{age}_{ij} \\ & + b_{0i} + \epsilon_{ij} \end{aligned}$$

$$\text{where } \boldsymbol{G} = \tau^2,\ \boldsymbol{R}_{\text{girl}} = \sigma_0^2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\boldsymbol{R}_{\text{boy}} = \sigma_1^2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

SAS code: model 2

```
proc mixed method=ml data=dental;
  class gender child;
  model distance = age gender gender*age / solution;
  repeated / group=gender subject=child;
  random intercept / subject=child g gcorr;
run;
```

SAS output: model 2

```
                    Model Information

Data Set                        WORK.DENTAL
Dependent Variable              distance
Covariance Structure            Variance Components
Subject Effects                 child, child
Group Effect                    gender
Estimation Method               ML
Residual Variance Method        None
Fixed Effects SE Method         Model-Based
Degrees of Freedom Method       Containment

                 Class Level Information

  Class     Levels    Values

  gender         2    0 1
  child         27    1 2 3 4 5 6 7 8 9 10 11 12 13
                      14 15 16 17 18 19 20 21 22 23
                      24 25 26 27

                      Dimensions

          Covariance Parameters               3
          Columns in X                        6
          Columns in Z Per Subject            1
          Subjects                           27
          Max Obs Per Subject                 4
```

SAS output: model 2

Estimated G Matrix

Row	Effect	child	Col1
1	Intercept	1	3.1405

Estimated G Correlation Matrix

Row	Effect	child	Col1
1	Intercept	1	1.0000

Covariance Parameter Estimates

Cov Parm	Subject	Group	Estimate
Intercept	child		3.1405
Residual	child	gender 0	0.5920
Residual	child	gender 1	2.7286

Fit Statistics

-2 Log Likelihood	409.4
AIC (smaller is better)	423.4
AICC (smaller is better)	424.5
BIC (smaller is better)	432.4

SAS output: model 2

Solution for Fixed Effects

Effect	gender	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		16.3406	1.1272	25	14.50	<.0001
age		0.7844	0.09234	79	8.49	<.0001
gender	0	1.0321	1.3767	79	0.75	0.4557
gender	1	0	.	.	.	.
age*gender	0	-0.3048	0.1059	79	-2.88	0.0051
age*gender	1	0	.	.	.	.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
age	1	79	142.41	<.0001
gender	1	79	0.56	0.4557
age*gender	1	79	8.28	0.0051

Model 2: Interpretation

- How to calculate ICC in model 2?

Model 3

- Model 3: random slope model

$$
\begin{aligned}
Y_{ij} = & \beta_0 + \beta_1 \times \text{age}_{ij} + \beta_2 \times \text{gender}_i + \beta_3 \times \text{gender}_i \times \text{age}_{ij} \\
& + b_{0i} + b_{1i} \times \text{age}_{ij} + \epsilon_{ij}
\end{aligned}
$$

can be re-written as $\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i$, where

$\boldsymbol{X}_i =$ $\qquad$ $\boldsymbol{\beta} =$

$\boldsymbol{Z}_i =$ $\qquad$ $\boldsymbol{b}_i =$

Model 3

- Model 3: random slope model

$$
\begin{aligned}
Y_{ij} = &\beta_0 + \beta_1 \times \text{age}_{ij} + \beta_2 \times \text{gender}_i + \beta_3 \times \text{gender}_i \times \text{age}_{ij} \\
&+ b_{0i} + b_{1i} \times \text{age}_{ij} + \epsilon_{ij}
\end{aligned}
$$

$$
\text{where } \boldsymbol{G} = \begin{bmatrix} \tau_1^2 & \tau_{12} \\ \tau_{12} & \tau_2^2 \end{bmatrix}, \ \boldsymbol{R}_i = \sigma^2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}
$$

SAS code: model 3

```
proc mixed method=ml data=dental;
  class gender child;
  model distance = age gender gender*age / solution;
  random intercept age / type=un subject=child g gcorr;
run;
```

SAS output: model 3

```
                     Model Information

Data Set                     WORK.DENTAL
Dependent Variable           distance
Covariance Structure         Unstructured
Subject Effect               child
Estimation Method            ML
Residual Variance Method     Profile
Fixed Effects SE Method      Model-Based
Degrees of Freedom Method    Containment

                  Class Level Information

  Class     Levels    Values

  gender         2    0 1
  child         27    1 2 3 4 5 6 7 8 9 10 11 12 13
                      14 15 16 17 18 19 20 21 22 23
                      24 25 26 27

                      Dimensions

          Covariance Parameters             4
          Columns in X                      6
          Columns in Z Per Subject          2
          Subjects                         27
          Max Obs Per Subject               4
```

SAS output: model 3

Estimated G Matrix

Row	Effect	child	Col1	Col2
1	Intercept	1	4.5569	-0.1983
2	age	1	-0.1983	0.02376

Estimated G Correlation Matrix

Row	Effect	child	Col1	Col2
1	Intercept	1	1.0000	-0.6025
2	age	1	-0.6025	1.0000

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
UN(1,1)	child	4.5569
UN(2,1)	child	-0.1983
UN(2,2)	child	0.02376
Residual		1.7162

Fit Statistics

-2 Log Likelihood	427.8
AIC (smaller is better)	443.8
AICC (smaller is better)	445.3
BIC (smaller is better)	454.2

SAS output: model 3

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
3	50.44	<.0001

Solution for Fixed Effects

Effect	gender	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		16.3406	0.9801	25	16.67	<.0001
age		0.7844	0.08275	25	9.48	<.0001
gender	0	1.0321	1.5355	54	0.67	0.5043
gender	1	0	.	.	.	.
age*gender	0	-0.3048	0.1296	54	-2.35	0.0224
age*gender	1	0	.	.	.	.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
age	1	25	95.04	<.0001
gender	1	54	0.45	0.5043
age*gender	1	54	5.53	0.0224

Model 4

- Model 3: random slope model with an AR(1) within-subject covariance matrix

$$
\begin{aligned}
Y_{ij} = &\beta_0 + \beta_1 \times \text{age}_{ij} + \beta_2 \times \text{gender}_i + \beta_3 \times \text{gender}_i \times \text{age}_{ij} \\
&+ b_{0i} + b_{1i} \times \text{age}_{ij} + \epsilon_{ij}
\end{aligned}
$$

where $\boldsymbol{G} = \begin{bmatrix} \tau_1^2 & \tau_{12} \\ \tau_{12} & \tau_2^2 \end{bmatrix}$, $\boldsymbol{R}_i = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$

SAS code: model 4

```
proc mixed method=ml data=dental;
  class gender child ;
  model distance = age gender gender*age / solution ;
  repeated / type=ar(1) subject=child rcorr;
  random intercept age / type=un subject=child g gcorr;
run;
```

SAS output: model 4

```
                 Model Information

Data Set                     WORK.DENTAL
Dependent Variable           distance
Covariance Structures        Unstructured,
                             Autoregressive
Subject Effects              child, child
Estimation Method            ML
Residual Variance Method     Profile
Fixed Effects SE Method      Model-Based
Degrees of Freedom Method    Containment

              Class Level Information

  Class     Levels    Values

  gender         2    0 1
  child         27    1 2 3 4 5 6 7 8 9 10 11 12 13
                      14 15 16 17 18 19 20 21 22 23
                      24 25 26 27

                      Dimensions

          Covariance Parameters             5
          Columns in X                      6
          Columns in Z Per Subject          2
          Subjects                         27
          Max Obs Per Subject               4
```

SAS output: model 4

Estimated R Correlation Matrix for child 1

Row	Col1	Col2	Col3	Col4
1	1.0000	-0.4680	0.2190	-0.1025
2	-0.4680	1.0000	-0.4680	0.2190
3	0.2190	-0.4680	1.0000	-0.4680
4	-0.1025	0.2190	-0.4680	1.0000

Estimated G Matrix

Row	Effect	child	Col1	Col2
1	Intercept	1	10.1459	-0.7198
2	age	1	-0.7198	0.07508

Estimated G Correlation Matrix

Row	Effect	child	Col1	Col2
1	Intercept	1	1.0000	-0.8248
2	age	1	-0.8248	1.0000

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
UN(1,1)	child	10.1459
UN(2,1)	child	-0.7198
UN(2,2)	child	0.07508
AR(1)	child	-0.4680
Residual		1.1940

SAS output: model 4

Fit Statistics

-2 Log Likelihood	424.1
AIC (smaller is better)	442.1
AICC (smaller is better)	443.9
BIC (smaller is better)	453.7

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
4	54.19	<.0001

Solution for Fixed Effects

Effect	gender	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		16.1544	0.9607	25	16.82	<.0001
age		0.7978	0.08374	25	9.53	<.0001
gender	0	1.2622	1.5051	54	0.84	0.4054
gender	1	0	.	.	.	.
age*gender	0	-0.3220	0.1312	54	-2.45	0.0174
age*gender	1	0	.	.	.	.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
age	1	25	94.23	<.0001
gender	1	54	0.70	0.4054
age*gender	1	54	6.03	0.0174

Model 5

- Model 5: random slope model with different within-subject covariance matrix for boys and girls

$$
\begin{aligned}
Y_{ij} = &\beta_0 + \beta_1 \times \text{age}_{ij} + \beta_2 \times \text{gender}_i + \beta_3 \times \text{gender}_i \times \text{age}_{ij} \\
&+ b_{0i} + b_{1i} \times \text{age}_{ij} + \epsilon_{ij}
\end{aligned}
$$

where $\boldsymbol{G} = \begin{bmatrix} \tau_1^2 & \tau_{12} \\ \tau_{12} & \tau_2^2 \end{bmatrix}$, $\boldsymbol{R}_{\text{girl}} = \sigma_0^2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

$\boldsymbol{R}_{\text{boy}} = \sigma_1^2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

SAS code: model 5

```
proc mixed  method=ml data=dental;
  class  child gender;
  model distance = age gender gender*age / solution;
  repeated / group=gender subject=child;
  random intercept age / type=un subject=child g gcorr;
run;
```

SAS output: model 5

```
                    Model Information

Data Set                         WORK.DENTAL
Dependent Variable               distance
Covariance Structures            Unstructured, Variance
                                 Components
Subject Effects                  child, child
Group Effect                     gender
Estimation Method                ML
Residual Variance Method         None
Fixed Effects SE Method          Model-Based
Degrees of Freedom Method        Containment

                 Class Level Information

  Class      Levels    Values

  child          27    1 2 3 4 5 6 7 8 9 10 11 12 13
                       14 15 16 17 18 19 20 21 22 23
                       24 25 26 27
  gender          2    0 1

                       Dimensions

         Covariance Parameters             5
         Columns in X                      6
         Columns in Z Per Subject          2
         Subjects                         27
         Max Obs Per Subject               4
```

SAS output: model 5

Estimated G Matrix

Row	Effect	child	Col1	Col2
1	Intercept	1	3.1978	-0.1103
2	age	1	-0.1103	0.01976

Estimated G Correlation Matrix

Row	Effect	child	Col1	Col2
1	Intercept	1	1.0000	-0.4388
2	age	1	-0.4388	1.0000

Covariance Parameter Estimates

Cov Parm	Subject	Group	Estimate
UN(1,1)	child		3.1978
UN(2,1)	child		-0.1103
UN(2,2)	child		0.01976
Residual	child	gender 0	0.4449
Residual	child	gender 1	2.6294

Fit Statistics

-2 Log Likelihood	406.0
AIC (smaller is better)	424.0
AICC (smaller is better)	425.9
BIC (smaller is better)	435.7

SAS output: model 5

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
4	72.20	<.0001

Solution for Fixed Effects

Effect	gender	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		16.3406	1.1114	25	14.70	<.0001
age		0.7844	0.09722	25	8.07	<.0001
gender	0	1.0321	1.3344	54	0.77	0.4426
gender	1	0	.	.	.	.
age*gender	0	-0.3048	0.1152	54	-2.65	0.0106
age*gender	1	0	.	.	.	.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
age	1	25	120.38	<.0001
gender	1	54	0.60	0.4426
age*gender	1	54	7.00	0.0106

Model 6

- Model 6: random slope model with different AR(1) within-subject covariance matrix for boys and girls

$$
\begin{aligned}
Y_{ij} = &\beta_0 + \beta_1 \times \text{age}_{ij} + \beta_2 \times \text{gender}_i + \beta_3 \times \text{gender}_i \times \text{age}_{ij} \\
&+ b_{0i} + b_{1i} \times \text{age}_{ij} + \epsilon_{ij}
\end{aligned}
$$

where $\boldsymbol{G} = \begin{bmatrix} \tau_1^2 & \tau_{12} \\ \tau_{12} & \tau_2^2 \end{bmatrix}$, $\boldsymbol{R}_{\text{girl}} = \sigma_0^2 \begin{bmatrix} 1 & \rho_0 & \rho_0^2 & \rho_0^3 \\ \rho_0 & 1 & \rho_0 & \rho_0^2 \\ \rho_0^2 & \rho_0 & 1 & \rho_0 \\ \rho_0^3 & \rho_0^2 & \rho_0 & 1 \end{bmatrix}$

$\boldsymbol{R}_{\text{boy}} = \sigma_1^2 \begin{bmatrix} 1 & \rho_1 & \rho_1^2 & \rho_1^3 \\ \rho_1 & 1 & \rho_1 & \rho_1^2 \\ \rho_1^2 & \rho_1 & 1 & \rho_1 \\ \rho_1^3 & \rho_1^2 & \rho_1 & 1 \end{bmatrix}$

SAS code: model 6

```
proc mixed  method=ml data=dental;
  class  child gender;
  model distance = age gender gender*age / solution;
  repeated / type=ar(1) group=gender subject=child;
  random intercept age / type=un subject=child g gcorr;
run;
```

SAS output: model 6

```
                    Model Information

Data Set                     WORK.DENTAL
Dependent Variable           distance
Covariance Structures        Unstructured,
                             Autoregressive
Subject Effects              child, child
Group Effect                 gender
Estimation Method            ML
Residual Variance Method     None
Fixed Effects SE Method      Model-Based
Degrees of Freedom Method    Containment

                 Class Level Information

  Class     Levels    Values

  child         27    1 2 3 4 5 6 7 8 9 10 11 12 13
                      14 15 16 17 18 19 20 21 22 23
                      24 25 26 27
  gender         2    0 1

                      Dimensions

          Covariance Parameters              7
          Columns in X                       6
          Columns in Z Per Subject           2
          Subjects                          27
          Max Obs Per Subject                4
```

SAS output: model 6

Estimated G Matrix

Row	Effect	child	Col1	Col2
1	Intercept	1	4.4794	-0.2206
2	age	1	-0.2206	0.03035

Estimated G Correlation Matrix

Row	Effect	child	Col1	Col2
1	Intercept	1	1.0000	-0.5982
2	age	1	-0.5982	1.0000

Covariance Parameter Estimates

Cov Parm	Subject	Group	Estimate
UN(1,1)	child		4.4794
UN(2,1)	child		-0.2206
UN(2,2)	child		0.03035
Variance	child	gender 0	0.3866
AR(1)	child	gender 0	-0.1325
Variance	child	gender 1	2.2542
AR(1)	child	gender 1	-0.2467

SAS output: model 6

Fit Statistics

-2 Log Likelihood	404.9
AIC (smaller is better)	426.9
AICC (smaller is better)	429.7
BIC (smaller is better)	441.2

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
6	73.29	<.0001

Solution for Fixed Effects

Effect	gender	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		16.2407	1.0061	25	16.14	<.0001
age		0.7914	0.08806	25	8.99	<.0001
gender	0	1.1444	1.2737	54	0.90	0.3729
gender	1	0	.	.	.	.
age*gender	0	-0.3129	0.1101	54	-2.84	0.0063
age*gender	1	0	.	.	.	.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
age	1	25	132.95	<.0001
gender	1	54	0.81	0.3729
age*gender	1	54	8.07	0.0063

Model comparison

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
β_0	13.35 (0.96)	16.34 (1.13)	16.34 (0.98)	16.15 (0.96)	16.34 (1.11)	16.24 (1.01)
β_1	0.78 (0.08)	0.78 (0.09)	0.78 (0.08)	0.80 (0.08)	0.78 (0.10)	0.79 (0.09)
β_2	1.03 (1.51)	1.03 (1.38)	1.03 (1.54)	1.26 (1.51)	1.03 (1.33)	1.14 (1.27)
β_3	-0.30 (0.12)	-0.30 (0.11)	-0.30 (0.13)	-0.32 (0.13)	-0.30 (0.12)	-0.31 (0.11)
σ_0^2	1.87	0.59	1.72	1.19	0.44	0.39
σ_1^2		2.73			2.63	2.25
ρ_0				-0.47		-0.13
ρ_1						-0.25
τ_1^2	3.03	3.14	4.56	10.15	3.20	4.48
τ_{12}			-0.20	-0.72	-0.11	-0.22
τ_2^2			0.02	0.08	0.02	0.03
τ_3^2						
τ_{34}						
τ_4^2						
$-2\log L$	428.6	409.4	427.8	424.1	406.0	404.9
AIC	440.6	423.4	443.8	442.1	424.0	426.9
AICC	441.5	424.5	445.3	443.9	425.9	429.7
BIC	448.4	432.4	454.2	453.7	435.7	441.1

Model 2 is the best! Why? We can use LR tests to compare nested models:

- Models 2 vs. 1 with p-value = $Pr(\chi_1^2 \geq 428.6 - 409.4) < .001$
- Models 5 vs. 2 with p-value = $Pr(\chi_2^2 \geq 409.4 - 406.0) = 0.183$
- Models 6 vs. 2 with p-value = $Pr(\chi_4^2 \geq 409.4 - 404.9) = 0.343$

Revisit Model 2

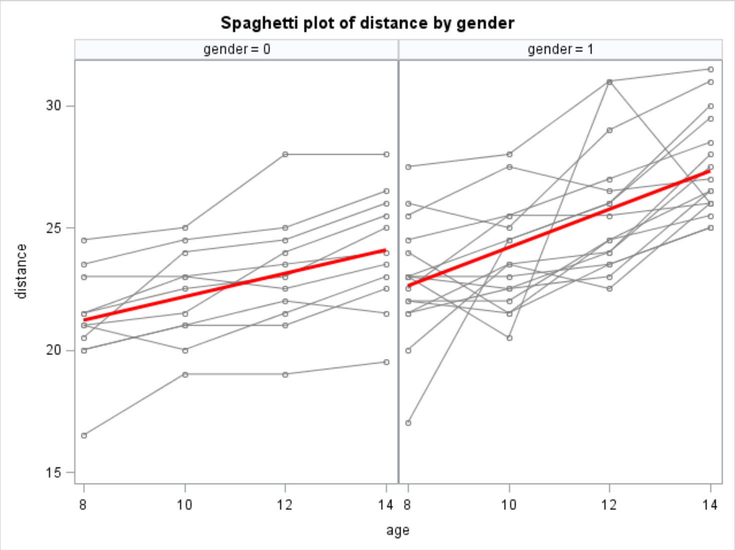
- Model 2: random intercept model with different within-subject covariance matrix between boys and girls

$$\begin{aligned} Y_{ij} = &\beta_0 + \beta_1 \times \text{age}_{ij} + \beta_2 \times \text{gender}_i + \beta_3 \times \text{gender}_i \times \text{age}_{ij} \\ &+ b_{0i} + \epsilon_{ij} \end{aligned}$$

where

$$\boldsymbol{G} = \tau^2, \ \boldsymbol{R}_{\text{girl}} = \sigma_0^2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \boldsymbol{R}_{\text{boy}} = \sigma_1^2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Revisit the data



SAS code: model 2

```
proc mixed method=ml data=dental;
  class gender child;
  model distance = age gender gender*age / solution;
  repeated / group=gender subject=child;
  random intercept / subject=child g gcorr;
run;
```

Revisit model 2

Covariance Parameter Estimates

Cov Parm	Subject	Group	Estimate
Intercept	child		3.1405
Residual	child	gender 0	0.5920
Residual	child	gender 1	2.7286

Solution for Fixed Effects

Effect	gender	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		16.3406	1.1272	25	14.50	<.0001
age		0.7844	0.09234	79	8.49	<.0001
gender	0	1.0321	1.3767	79	0.75	0.4557
gender	1	0	.	.	.	.
age*gender	0	-0.3048	0.1059	79	-2.88	0.0051
age*gender	1	0	.	.	.	.

Interpretation

- For each year increase in age, on average, the distance is increased by
  - _____ (95% CI: [,]) units among boys
  - _____ (95% CI: [,]) units among girls
- The average increase in the distance for each year increase in age is _____ (95% CI: [,]) units faster among boys than girls.

SAS code: 95% CI for β

```
proc mixed method=ml data=dental;
  class gender child;
  model distance = age gender gender*age / solution alpha=0.05;
  repeated / group=gender subject=child;
  random intercept / subject=child g gcorr;
run;
```

SAS output: 95% CI for β

Solution for Fixed Effects

Effect	gender	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
Intercept		16.3406	1.1272	25	14.50	<.0001	0.05	14.0190	18.6622
age		0.7844	0.09234	79	8.49	<.0001	0.05	0.6006	0.9682
gender	0	1.0321	1.3767	79	0.75	0.4557	0.05	-1.7081	3.7723
gender	1	0	.	.	.	.	.	.	.
age*gender	0	-0.3048	0.1059	79	-2.88	0.0051	0.05	-0.5156	-0.09401
age*gender	1	0	.	.	.	.	.	.	.

- For each year increase in age, on average, the distance is increased by
  - 0.78 (95% CI: 0.60-0.97) units among boys
  - 0.48 (95% CI: ??-??) units among girls
- The average increase in the distance for each year increase in age is 0.30 (95% CI: 0.09-0.52) units faster among boys than girls.

SAS output: 95% CI for β

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
Intercept	17.3727	0.7903	25	21.98	<.0001	0.05	15.7451	19.0004
age	0.4795	0.05187	79	9.24	<.0001	0.05	0.3763	0.5828
gender	-1.0321	1.3767	79	-0.75	0.4557	0.05	-3.7723	1.7081
age*gender	0.3048	0.1059	79	2.88	0.0051	0.05	0.09401	0.5156

- Code “girls” as the reference group
- For each year increase in age, on average, the distance is increased by
  - 0.48 (95% CI: 0.38-0.58) units among girls

Other model: categorical age

- If age is a categorical variable and we fit a model

$$
\begin{aligned}
y_{ij} =& \beta_0 + \beta_1 \times I(age_{ij} = 10) + \beta_2 \times I(age_{ij} = 12) \\
& + \beta_3 \times I(age_{ij} = 14) + \beta_4 \times gender_i \\
& + \beta_1 \times I(age_{ij} = 10) \times gender_i \\
& + \beta_2 \times I(age_{ij} = 12) \times gender_i \\
& + \beta_3 \times I(age_{ij} = 14) \times gender_i + \epsilon_{ij}
\end{aligned}
$$

- SAS code:

```
proc mixed method=ml data=dental;
  class gender age child;
  model distance = age gender gender*age / solution alpha=0.05;
  repeated / group=gender subject=child;
  random intercept / subject=child g gcorr;
  lsmeans gender*age /cl diff;
run;
```

SAS output: categorical age

Fit Statistics

-2 Log Likelihood	407.9
AIC (smaller is better)	429.9
AICC (smaller is better)	432.6
BIC (smaller is better)	444.1

Least Squares Means

Effect	gender	age	Estimate	Alpha	Lower	Upper
gender*age	0	8	21.1818	0.05	20.0207	22.3429
gender*age	0	10	22.2273	0.05	21.0662	23.3884
gender*age	0	12	23.0909	0.05	21.9298	24.2520
gender*age	0	14	24.0909	0.05	22.9298	25.2520
gender*age	1	8	22.8750	0.05	21.6755	24.0745
gender*age	1	10	23.8125	0.05	22.6130	25.0120
gender*age	1	12	25.7187	0.05	24.5192	26.9183
gender*age	1	14	27.4688	0.05	26.2692	28.6683

Differences of Least Squares Means

Effect	gender	age	_gender	_age	Estimate	Pr > |t|	Alpha	Lower	Upper
gender*age	0	8	1	8	-1.6932	0.0469	0.05	-3.3626	-0.02374
gender*age	0	10	1	10	-1.5852	0.0624	0.05	-3.2547	0.08422
gender*age	0	12	1	12	-2.6278	0.0024	0.05	-4.2973	-0.9584
gender*age	0	14	1	14	-3.3778	0.0001	0.05	-5.0473	-1.7084

Continuous or categorical age?

- Question: Would you model age as continuous or categorical based on the SAS outputs of the two models?

Summary of key points

- What is linear mixed effects model?
- What are the *R* and *G* matrices in mixed models?
- How to specify *R* and *G* matrices in SAS?
- Which *R* and *G* matrices to choose?
- How to create spaghetti plot using SAS?
- Should the “time" variable modeled as continuous or categorical?

Suggested Reading

- Chapter 10 (Davidian)