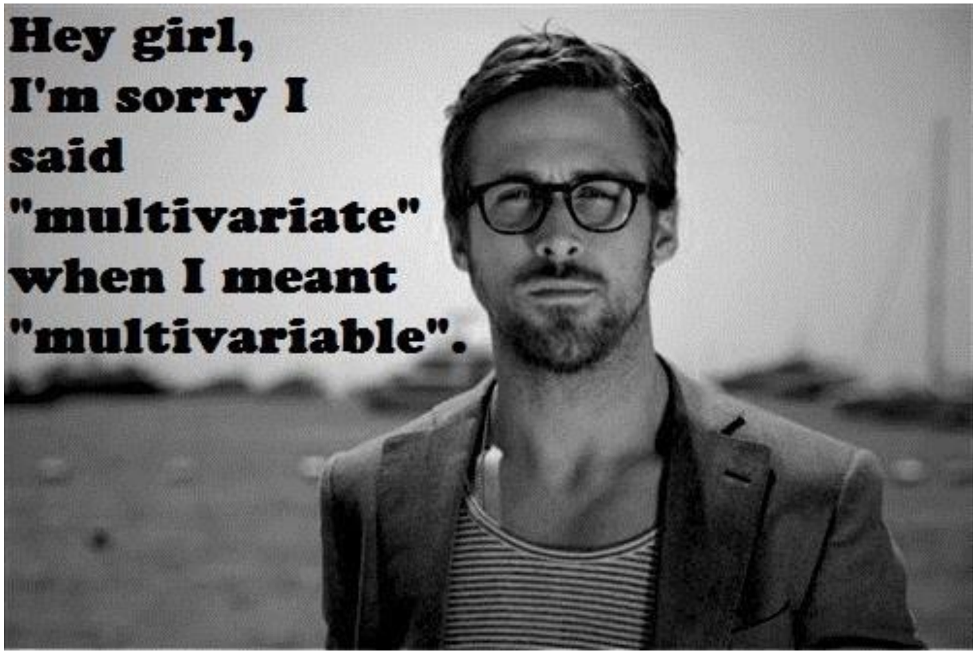
BIOS 6612
Lecture 13

Linear Mixed Models (LMM)
Introduction



Reading: Longitudinal Note Set Section 1 & 2.1
Additional Reading:

- Linear Mixed Models for Longitudinal Data; Verbeke and Molenberghs; Springer; 2000.
- Longitudinal Data Analysis, Hedeker and Gibbons, Wiley, 2006, Chapters 4-7.
- Applied Longitudinal Analysis, Fitzmaurice, Laird and Ware, Wiley, 2011, Chapters 8-9.

Review (Lecture 12)/ Current (Lecture 13)/ Preview (Midterm Review)

- Lecture 12: General Linear Models III
  - Hypothesis tests
    - t-tests
    - F-tests
    - Main effect tests
      - Estimation vs Contrasts

- Lecture 13: Linear Mixed Models
  - 2 time points per persons
    - Paired t-test
    - Linear regression
    - Random intercept model
  - Notation
  - Linear mixed models with a random intercept
  - Linear mixed models specifying the covariance structure

- Midterm Review
  - Potential topics
  - No guarantee on the material
  - SAS output separate from exam

General Linear Models III: Example 2 Way ANOVA Model

- The model: $Y_{ijk} = \mu + \alpha_i + \tau_j + \gamma_{ij} + \varepsilon_{ijk}$ for group i=1,2 time j=1,2,3 replication k=1,2,3,4

$E[\boldsymbol{Y}] = \boldsymbol{X\beta}$ where $\boldsymbol{\beta}^T = (\mu \quad \alpha_1 \quad \alpha_2 \ \tau_1 \quad \tau_2 \quad \tau_3 \ \gamma_{11} \quad \gamma_{12} \quad \gamma_{13} \ \gamma_{21} \quad \gamma_{22} \quad \gamma_{23})$

- For example, say we want to compare means for treatment groups at 24 hours
    - The entire **L** would be (0 1 -1 0 0 0 1 0 0 -1 0 0)

    - In SAS, the estimate statement is:

```
ESTIMATE 'group diffs at 24 hrs' intercept 0 group 1 -1 time 0 0 0
    group*time 1 0 0 -1 0 0;
```

- Note: to find L, easier to consider each group first, and then take the difference:

		Int	Group		Time			Group*Time					
Control group at 24 hours:	**L**=	(1	1	0	1	0	0	1	0	0	0	0	0)
Myostatin group at 24 hours:	**L**=	(1	0	1	1	0	0	0	0	0	1	0	0)
Difference:	**L**=	(0	1	-1	0	0	0	1	0	0	-1	0	0)

- Or it may be easier to consider:

Control group at 24 hours: $E[Y_{11k}] = \mu + \alpha_1 + \tau_1 + \gamma_{11}$

Myostatin group at 24 hours: $E[Y_{21k}] = \mu + \alpha_2 + \tau_1 + \gamma_{21}$

Difference:

$$E[Y_{11k}] - E[Y_{21k}] = (\mu + \alpha_1 + \tau_1 + \gamma_{11}) - (\mu + \alpha_2 + \tau_1 + \gamma_{21}) = \alpha_1 - \alpha_2 + \gamma_{11} - \gamma_{21}$$

Repeated Measures

- The terms “repeated measurements” or “repeated measures” are sometimes used as rough synonyms for “longitudinal data”
    - However, there are sometimes slight differences in the meaning of these terms
- Repeated measures are also multiple measurements on each of several individuals
    - But they are not necessarily through time
    - For example:
        - Measurements of chemical concentration in the leaves of a plant taken at different locations (low, medium and high on the plant)
    - For Example:
        - In the COPDGene study, an investigator wants to determine if pulmonary function as measured by forced expiratory volume in 1 second (FEV1) and forced vital capacity (FVC), emphysema, and pack-years of cigarettes smoked are jointly associated with a SNP on chromosome 15 in the nicotine receptor CHRNA3/5, adjusting for age, gender, and genetic ancestry via principal components

Repeated Measures

- Repeated measures could also be viewed as multiple measurements on a unit
    - For example:
        - Standardized test scores from students in the same classroom in same school
- In addition, repeated measures may occur across the levels of some controlled factor
    - For example:
        - Crossover studies involving repeated measures
        - In a crossover study, subjects are assigned to multiple treatments (usually 2 or 3) sequentially with a washout period in between
        - i.e. two period crossover experiment involves subjects who each get treatments A and B, some in the order AB, and others in the order BA
- In all cases, however, we are referring to multiple measurements on a given subject or unit of observation

Longitudinal Data

- Longitudinal data consist of observations (i.e., measurements) taken repeatedly through time on a sample of experimental units (i.e., individuals, subjects)

- The experimental units or subjects can be human patients, animals, agricultural plots, etc

- Typically, the terms "longitudinal data" and "longitudinal study" refer to situations in which data are collected through time under controlled or uncontrolled circumstances

    - For example: subjects with torn ACLs in their knees are assigned to one of two methods of surgical repair and then followed through time
        - Examined at 6, 12, ,18, 24 months for knee stability

    - For example: dental measurements at 4 ages
        - Measure the ramus bone in lower jaw in mm
        - On 20 boys at four fixed ages: 8, 8.5, 9 & 9.5
        - Prospective study that has existed for over 40 years
        - Used by dentists to establish a growth curve for the ramus

- Longitudinal data are often contrasted with cross-sectional data
    - Cross-sectional data contains measurements on a sample of subjects at only one point in time

Advantages of Longitudinal Data

- Although time effects can be investigated in cross-sectional studies in which different subjects are examined at different time points
    - Only longitudinal data gives information on individual patterns of change
- Longitudinal studies economize on subjects
    - In investigating time effects in a longitudinal design or treatment effects in a crossover design, each subject can “serve as his or her own control”
    - Comparisons can be made within a subject rather than between subjects
    - This eliminates between-subjects sources of variability from the experimental error
        - This makes inferences more efficient/powerful
- Since the same variables are measured repeatedly on the same subjects, the reliability of those measurements can be assessed, and purely from a measurement standpoint, reliability is higher

Disadvantages of Longitudinal Data

- For longitudinal or clustered data, it is typically reasonable to assume independence across clusters
    - But repeated measures within a cluster are almost always correlated
    - This may complicate the analysis
- Clustered data are often unbalanced or partially incomplete (involve missing data)
    - Loss to follow-up
        - Some subjects move away, die, miss appointments, etc.
    - For other types of clustered data, the cluster size may vary
        - Family data, where family size varies
    - This may complicate the analysis
- As a practical matter, methods and/or software may not exist or may be complex, so obtaining results and interpreting them may be difficult

Simple Longitudinal Data Analysis Example (Treatment Difference)

- 2 roughly normally distributed measurements per person at 2 equally spaced visits
    - Visit 1: Serum cholesterol measurements (mcg/dl) on standard American diet
    - Visit 2: Serum cholesterol measurements on vegetarian diet one month later

- Solutions for Simple Longitudinal Data:
    - 1. Change-score model: $\Delta = Y_{post} - Y_{pre}$ as outcome
        - Fit the linear regression model with just intercept: $E[\Delta_i] = E\left[Y_{post_i} - Y_{pre_i}\right] = \beta_0$
        - $H_0 : \beta_0 = 0$ (mean cholesterol difference=0)
            - (1) Fit with linear regression or (2) paired t-test

    - 2. Baseline-as-covariate model: outcome= Y_{post} and covariate= Y_{pre}
        - Fit the linear regression model: $E\left[Y_{post_i}\right] = \alpha_0 + \alpha_b Y_{pre_i}$
        - $H_0 : \alpha_b = 0$ (mean post cholesterol is not associated with pre cholesterol levels)

    - 3. Hybrid model: $\Delta = Y_{post} - Y_{pre}$ as outcome and covariate= Y_{pre}
        - Fit the linear regression model $E[\Delta_i] = E\left[Y_{post_i} - Y_{pre_i}\right] = \gamma_0 + \gamma_b Y_{pre_i}$
            - Slope γ_b indicates whether change in cholesterol related to base line value

Article: BaselineChange.pdf article discusses when not to adjust for baseline value

- 4. Linear Mixed Model
    - Random Intercept
        - Allow for time varying covariates

Simple Longitundinal Data Analysis Example (Time Difference)

- 2 roughly normally distributed measurements per person
  - Measurements taken at 2 visits (2nd visit is 5 years later)
    - Not a difference in treatment, but a difference of time
    - i.e. FEV_1, 6 minute walk

- Solutions:
  - 1. Model the average of Y_{visit2} and Y_{visit1} as the outcome
    - May NOT be appropriate given the question of interest
  - 2. Model Y_{visit2} as outcome, with Y_{visit1} as a covariate using linear regression
  - 3. Consider $\Delta = Y_{visit2} - Y_{visit1}$ as outcome
    - Y_{visit1} may be included as a covariate in this model
      - Slope indicates whether changes over time are related to base line value
  - 4. Longitudinal model for the data (i.e. general linear mixed models)
    - Outcome= both base line and follow-up measurements
    - Able to model time sensitive covariates
    - Example: Random intercept model
    - Model time as class variable or continuous

Solutions 1-4: can include covariates for precision or to account for confounders
Solutions 1-3: Not able to estimate variances at each time or correlation between time points

You need to consider the question of interest first, before deciding on the methods!!!

Linear Mixed Models

- Linear mixed model methods are methods for analyzing clustered data
  - When the outcome variable is continuous and approximately normally distributed
- Useful for analyzing repeated measures and longitudinal data
- Both SAS and R are used to fit linear mixed models in these notes
- If the outcome is not 'approximately normal', one of the following might be considered:
  - Transform the outcome
  - Mixture distribution model
  - Non-normal outcome model
- A mixed model may have random as well as fixed effects
- Mixed models allow for more complicated error covariance structures
  - Unlike simpler general linear models

Methods for Linear Mixed Models

- Standard methods to conduct inference in a linear mixed model (LMM)
  - Parameters usually estimated using Maximum likelihood (ML) or restricted maximum likelihood (REML) methods
  - Random effects are estimated using empirical Bayes methods
  - Tests for fixed-effect parameters in the model are usually conducted using functions of estimated parameters that have exact or approximate t or F statistics

- Other methods can be used to fit LMMs than "standard" LMM methodology:
  - A simple mixed model with a random intercept and other fixed effects could be fit using repeated measures ANOVA
  - A model without random effects but with a non-independent error covariance structure could be fit using generalized least squares
  - 'Multivariate GLM' methods such as MANOVA could be used with an LMM with no random effects but unstructured error covariance structure

- Differences in results obtained between using these alternative approaches and standard LMM methods are often minor, and in some cases they will be the same
  - However, these alternative approaches have their limitations in that they can only be applied to specific types of mixed models

Linear regression in matrix notation

$$\mathbf{Y} = \mathbf{X\beta}+\boldsymbol{\varepsilon} \ \textit{where} \ \varepsilon_i \sim N(0, \sigma^2 \mathrm{I}_{nxn})$$

$$\mathbf{Y}_{n\times 1} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \boldsymbol{\beta}_{(p+1)\times 1} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \boldsymbol{\varepsilon}_{n\times 1} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$$\mathbf{X}_{n\times(p+1)} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$$

Note that the first column of the **X** matrix of independent variables contains only 1's. This is the general convention used for any regression model containing an intercept (i.e., a constant term β_0).

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \Rightarrow Var(Y) = \begin{pmatrix} \sigma^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \sigma^2 \end{pmatrix} = \sigma^2 \mathrm{I}_{nxn}$$

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \varepsilon_i \Rightarrow E[Y_i \mid \mathrm{X}_i] = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}$$

$$Var(\mathrm{Y_i}) = \mathrm{cov}(\mathrm{Y_i}, \mathrm{Y}_i) = \sigma^2 \ \textit{and} \ \ \mathrm{cov}(\mathrm{Y_i}, \mathrm{Y_j}) = 0 \ \ \forall \ i \neq j$$

Linear Mixed Model

- The general linear mixed model can be defined as:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Zb} + \boldsymbol{\varepsilon}$$

$\boldsymbol{Y}$ is the vector that contains the responses

$\boldsymbol{X}$ is a known matrix (design matrix)

$\boldsymbol{\beta}$ is the vector that contains the overall mean and all the fixed effects parameters

$\boldsymbol{Z}$ is a known matrix (the design matrix for the random effects)

$\boldsymbol{b}$ is the vector that contains all the random-effects variables

$\boldsymbol{\varepsilon}$ is the vector that contains the random errors

and

$$\begin{pmatrix} \mathbf{b} \\ \boldsymbol{\varepsilon} \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{G} & 0 \\ 0 & \mathbf{R} \end{pmatrix} \right)$$

We can specify G and/or R to account for the correlation between measurements.

So that,

$$V = Var(\mathbf{Y}) = \mathbf{ZGZ'} + \mathbf{R}$$

Specified by SAS RANDOM statement.

Specified by SAS REPEATED statement.

Linear Mixed Models Notation

- 3 basic ways that linear mixed models can be expressed:
    - subject-time level
    - subject level
    - complete or full data level

- Subject-time level: useful when the particular experiment and variables are defined
    - For the myostatin mean model example for group i time j replicate k,

$$Y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$$

- Subject level data:

$$\underset{r_i\times 1}{\mathbf{Y}_i} = \underset{r_i\times p}{\mathbf{X}_i}\ \underset{p\times 1}{\boldsymbol{\beta}} + \underset{r_i\times q}{\mathbf{Z}_i}\ \underset{q\times 1}{\mathbf{b}_i} + \underset{r_i\times 1}{\boldsymbol{\varepsilon}_i}, \qquad \text{for subjects } i=1,\ldots,n.$$

- $\mathbf{Y}_i$ are the $r_i\times 1$ responses for subject i
- $\mathbf{X}_i$ is the matrix of known covariates associated with fixed effects
- $\boldsymbol{\beta}$ are the $p\times 1$ fixed effects
- $\mathbf{Z}_i$ is the matrix of known covariates associated with the random effects
- $\boldsymbol{\varepsilon}_i$ is the residual error vector

- Complete or full level data:

$$\mathbf{Y} = \mathbf{X}\,\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon}$$

- No indices to denote full-data version

Linear Mixed Models Notation

- The subject models can be combined into one ‘complete-data’ model by essentially stacking the n subject-specific models:

$$\underbrace{\begin{pmatrix} \underset{r_1\times 1}{\mathbf{Y}_1} \\ \underset{r_2\times 1}{\mathbf{Y}_2} \\ \vdots \\ \underset{r_n\times 1}{\mathbf{Y}_n} \end{pmatrix}}_{r_{tot}\times 1} = \underbrace{\begin{pmatrix} \underset{r_1\times p}{\mathbf{X}_1} \\ \underset{r_2\times p}{\mathbf{X}_2} \\ \vdots \\ \underset{r_n\times p}{\mathbf{X}_n} \end{pmatrix}}_{r_{tot}\times p} \underset{p\times 1}{\boldsymbol{\beta}} + \underbrace{\begin{pmatrix} \underset{r_1\times q}{\mathbf{Z}_1} & \underset{r_1\times q}{\mathbf{0}} & \cdots & \underset{r_1\times q}{\mathbf{0}} \\ \underset{r_2\times q}{\mathbf{0}} & \underset{r_2\times q}{\mathbf{Z}_2} & & \underset{r_2\times q}{\mathbf{0}} \\ \vdots & & \ddots & \\ \underset{r_n\times q}{\mathbf{0}} & \underset{r_n\times q}{\mathbf{0}} & & \underset{r_n\times q}{\mathbf{Z}_n} \end{pmatrix}}_{r_{tot}\times q_{tot}} \underbrace{\begin{pmatrix} \underset{q\times 1}{\mathbf{b}_1} \\ \underset{q\times 1}{\mathbf{b}_2} \\ \vdots \\ \underset{q\times 1}{\mathbf{b}_n} \end{pmatrix}}_{q_{tot}\times 1} + \underbrace{\begin{pmatrix} \underset{r_1\times 1}{\boldsymbol{\varepsilon}_1} \\ \underset{r_2\times 1}{\boldsymbol{\varepsilon}_2} \\ \vdots \\ \underset{r_n\times 1}{\boldsymbol{\varepsilon}_n} \end{pmatrix}}_{r_{tot}\times 1},$$

Or $\underset{r_{tot}\times 1}{\mathbf{Y}} = \underset{r_{tot}\times p}{\mathbf{X}} \underset{p\times 1}{\boldsymbol{\beta}} + \underset{r_{tot}\times q_{tot}}{\mathbf{Z}} \underset{q_{tot}\times 1}{\mathbf{b}} + \underset{r_{tot}\times 1}{\boldsymbol{\varepsilon}}$

where $\begin{pmatrix} \underset{q_{tot}\times 1}{\mathbf{b}} \\ \underset{r_{tot}\times 1}{\boldsymbol{\varepsilon}} \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} \underset{q_{tot}\times 1}{\mathbf{0}} \\ \underset{r_{tot}\times 1}{\mathbf{0}} \end{pmatrix}, \begin{pmatrix} \underset{q_{tot}\times q_{tot}}{\mathbf{G}} & \underset{q_{tot}\times r_{tot}}{\mathbf{0}} \\ \underset{r_{tot}\times q_{tot}}{\mathbf{0}} & \underset{r_{tot}\times r_{tot}}{\mathbf{R}} \end{pmatrix}\right]$

$q_{tot} = nq$, $r_{tot} = \Sigma\, r_i$, $\underset{q_{tot}\times q_{tot}}{\mathbf{G}} = \underset{i=1}{\overset{n}{\mathrm{diag}}}\left\{\underset{q\times q}{\mathbf{G}_i}\right\}$ and $\underset{r_{tot}\times r_{tot}}{\mathbf{R}} = \underset{i=1}{\overset{n}{\mathrm{diag}}}\left\{\underset{r_i\times r_i}{\mathbf{R}_i}\right\}$

Linear Mixed Models Notation

- Even when $\mathbf{R}_i$ or $\mathbf{G}_i$ are the same across subjects (this is usually the case for $\mathbf{G}_i$), we keep the subscript i since $\mathbf{R}$ and $\mathbf{G}$ are used for complete data form
  - When $\mathbf{R}_i$ does differ between subjects due to missing data, you can keep dimensions of $\mathbf{R}_i$ the same across subjects by partitioning the matrix into 'observed' and 'missing' pieces

- For the model above, we assume

$$\underset{q\times 1}{\mathbf{b}_i} \sim iid\ \mathrm{N}\left[\underset{q\times 1}{\mathbf{0}}, \underset{q\times q}{\mathbf{G}_i}\right] \text{ and } \underset{r_i\times 1}{\boldsymbol{\varepsilon}_i} \sim iid\ \mathrm{N}\left[\underset{r_i\times 1}{\mathbf{0}}, \underset{r_i\times r_i}{\mathbf{R}_i}\right]$$

  - Often assume that these random vectors are independent
  - Often assume that subjects are independent of each other

- Generally speaking,
  - $\mathbf{G}_i$ will be used to account for variability between subjects
  - $\mathbf{R}_i$ will be used to account for covariances between repeated measures within subjects
  - However, it will also be demonstrated that there are many ways to model correlated data that combine $\mathbf{G}_i$ and $\mathbf{R}_i$

Linear Mixed Models with a Random Intercept

- Special case of linear mixed models: random intercept model
    - A general linear model with an additional random effect called a random intercept
    - This random intercept can be defined for any cluster unit, but here we consider it for subjects.
    - This model offers one simple way to account for longitudinal data
- Considering longitudinal studies, a random intercept term for subjects will
    - Account for between-subject variability
    - Induce a correlation structure for the responses
    - Often over simplistic for longitudinal data
        - But generally an improvement over no correlation structure at all

Linear Mixed Models with a Random Intercept

- The basic model (subject-time level) is

$$Y_{ij} = \beta_0 + \beta_1 x_{1ij} + \ldots + \beta_{p-1} x_{p-1,ij} + b_i + \varepsilon_{ij}$$
$$= \mathbf{X}_{ij}\boldsymbol{\beta} + b_i + \varepsilon_{ij}$$

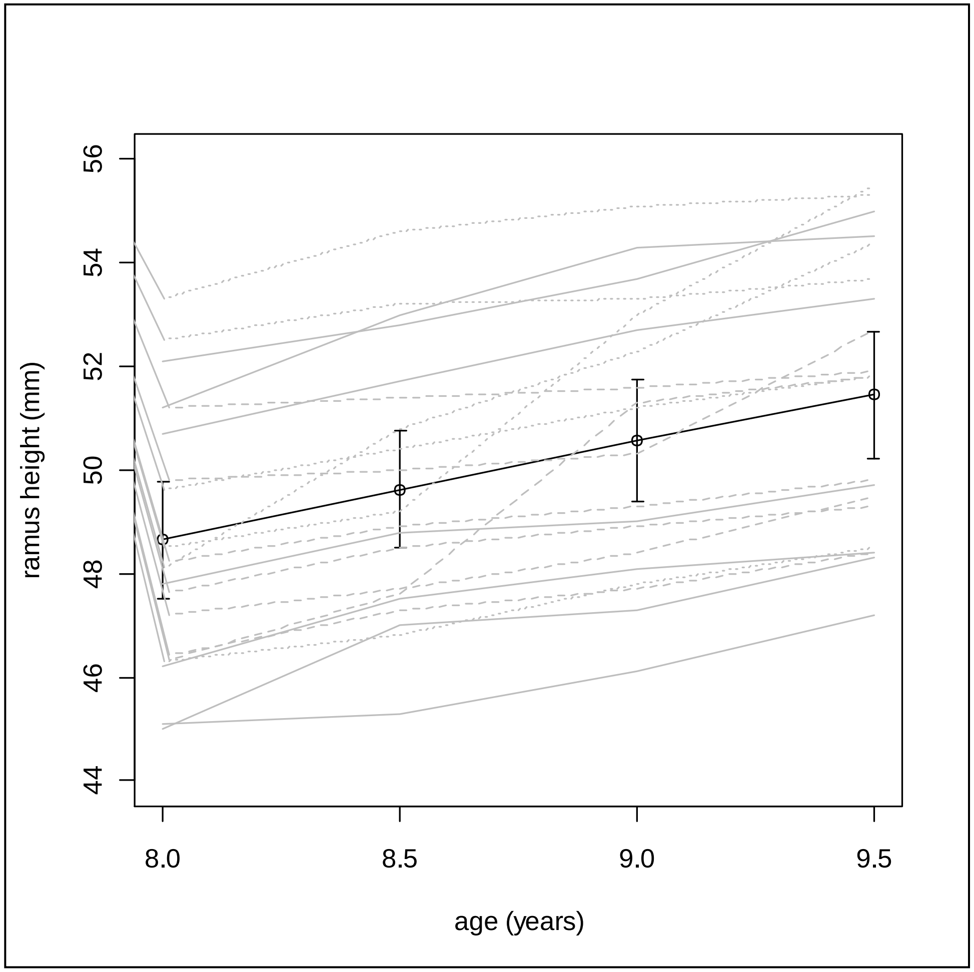
  - i indexes the subject 1…n
  - j indexes the time
  - p= number of covariates
  - *Y* is the outcome
  - $\mathbf{X}_{ij} = (x_1,\ldots,x_{p-1})$ is a row vector of predictors, both for subject *i* at time *j*
  - $\varepsilon_{ij} \sim N(0,\sigma_\varepsilon^2)$ and $b_i \sim N(0,\sigma_b^2)$
    - These random terms are assumed to be independent of each other

- The main element that distinguishes this from a general linear model is the addition of the random term b_i

- Ramus data example for random intercept model
  - Look at the 3 model forms: subject-time level, subject level, and full level
  - Look at the form of the variance covariance matrix

Ramus Data Example

- Ramus bone in lower jaw was measured on 20 boys at four fixed ages: 8, 8½, 9 & 9½
  - Prospective study that has existed for over 40 years
  - Used by dentists to establish a growth curve for the ramus
  - Ages 8 (h1), 8½ (h2), 9 (h3) and 9½ (h4) in mm



Consider a Random Intercept model for the Dental Measurements Example

- 20 boys at four fixed ages: 8, 8½, 9 & 9½
- Subject-time level:

$$Y_{ij} = \beta_0 + \beta_{age} age_{ij} + b_i + \varepsilon_{ij}$$

Where $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ and $b_i \sim N(0, \sigma_b^2)$

- Subject level:

$$\mathbf{Y}_1 = \mathbf{X}_1\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{b}_1 + \boldsymbol{\varepsilon}_1$$

$$\mathbf{Y}_1 = \begin{bmatrix} Y_{(1)1} \\ Y_{(1)2} \\ Y_{(1)3} \\ Y_{(1)4} \end{bmatrix} = \begin{bmatrix} 47.8 \\ 48.8 \\ 49 \\ 49.7 \end{bmatrix}$$

$$\mathbf{X}_1\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{b}_1 + \boldsymbol{\varepsilon}_1 = \begin{bmatrix} 1 & age1 \\ 1 & age2 \\ 1 & age3 \\ 1 & age4 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_{age} \end{bmatrix} + \mathbf{Z}_1\mathbf{b}_1 + \boldsymbol{\varepsilon}_1 = \begin{bmatrix} \beta_0 + 8\beta_{age} \\ \beta_0 + 8.5\beta_{age} \\ \beta_0 + 9\beta_{age} \\ \beta_0 + 9.5\beta_{age} \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \mathrm{b}_1 + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{14} \end{bmatrix}$$

Consider a Random Intercept model for the Dental Measurements Example

- Then for full or complete data level

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

$$V = Var(\mathbf{Y}) = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}$$

$\mathbf{Y} = \mathbf{X}\beta$+Zb+$\varepsilon$

$$\mathbf{Y}_{4(20)\times 1} = \begin{pmatrix} \begin{bmatrix} Y_{(1)1} \\ Y_{(1)2} \\ Y_{(1)3} \\ Y_{(1)4} \end{bmatrix} \\ \vdots \\ \begin{bmatrix} Y_{(n)1} \\ Y_{(n)2} \\ Y_{(n)3} \\ Y_{(n)4} \end{bmatrix} \end{pmatrix} =, \mathbf{X}_{4(20)\,\mathrm{x}(1+1)} = \begin{pmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} & \begin{bmatrix} age1 \\ age2 \\ age3 \\ age4 \end{bmatrix} \\ \vdots & \vdots \\ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} & \begin{bmatrix} age1 \\ age2 \\ age3 \\ age4 \end{bmatrix} \end{pmatrix}, \boldsymbol{\beta}_{(1+1)\times 1} = \begin{pmatrix} \beta_0 \\ \beta_{age} \end{pmatrix}, (\mathbf{X}\boldsymbol{\beta})_{4(20)\times 1} = \begin{pmatrix} \beta_0 + age1 * \beta_{age} \\ \beta_0 + age2 * \beta_{age} \\ \beta_0 + age3 * \beta_{age} \\ \beta_0 + age4 * \beta_{age} \\ \vdots \\ \beta_0 + age1 * \beta_{age} \\ \beta_0 + age2 * \beta_{age} \\ \beta_0 + age3 * \beta_{age} \\ \beta_0 + age4 * \beta_{age} \end{pmatrix}$$

$$\mathbf{Z}_{4(20)\times 1(20)}\mathbf{G}_{20\times 20}\mathbf{Z}^{T}{}_{20\times 4(20)} = \begin{pmatrix} 1 & 0 & & 0 \\ 1 & 0 & \cdots & 0 \\ 1 & 0 & & 0 \\ 1 & 0 & & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & 1 & & 0 \\ 0 & 1 & & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & & 1 \\ 0 & 0 & & 1 \\ 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \sigma_b^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_b^2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & & & & & & & \vdots & \ddots & \vdots & & & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 1 & 1 & 1 \end{pmatrix}$$

$$\mathbf{Z}_{4(20)\times 1(20)}\mathbf{G}_{20\times 20}\mathbf{Z}^{T}{}_{20\times 4(20)} = \sigma_b^2 \begin{pmatrix} 1 & \cdots & 0 \\ 1 & & 0 \\ 1 & & 0 \\ 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \\ 0 & & 1 \\ 0 & & 1 \\ 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & & & \vdots & \ddots & \vdots & & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 1 & 1 & 1 \end{pmatrix} = \sigma_b^2 \begin{pmatrix} 1 & 1 & 1 & 1 & & & & & 0 \\ 1 & 1 & 1 & 1 & & & & & \\ 1 & 1 & 1 & 1 & & & & & \\ 1 & 1 & 1 & 1 & & & & & \\ & & & & \ddots & & & & \\ & & & & & 1 & 1 & 1 & 1 \\ & & & & & 1 & 1 & 1 & 1 \\ & & & & & 1 & 1 & 1 & 1 \\ 0 & & & & & 1 & 1 & 1 & 1 \end{pmatrix}_{4(20)\times 4(20)}$$

$$\begin{pmatrix} \mathbf{b} \\ \boldsymbol{\varepsilon} \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{G} & 0 \\ 0 & \mathbf{R} \end{pmatrix} \right)$$

$$Var(\mathbf{Y})_{4(20)\times 4(20)} = \mathbf{Z}_{4(20)\times 20}\mathbf{G}_{20\times 20}\mathbf{Z}^T{}_{20\times 4(20)} + \mathbf{R}_{80\times 80} = \mathbf{Z}_{4(20)\times 20}\mathbf{G}_{20\times 20}\mathbf{Z}^T{}_{20\times 4(20)} + \sigma_e^2\mathbf{I}_{80\times 80}$$

$$= \sigma_b^2 \begin{pmatrix} 1 & 1 & 1 & 1 & & & & & & 0 \\ 1 & 1 & 1 & 1 & & & & & & \\ 1 & 1 & 1 & 1 & & & & & & \\ 1 & 1 & 1 & 1 & & & & & & \\ & & & & \ddots & & & & & \\ & & & & & & 1 & 1 & 1 & 1 \\ & & & & & & 1 & 1 & 1 & 1 \\ & & & & & & 1 & 1 & 1 & 1 \\ 0 & & & & & & 1 & 1 & 1 & 1 \end{pmatrix}_{4(20)\times 4(20)} + \sigma_e^2 \begin{pmatrix} 1 & 0 & 0 & 0 & & & & & & 0 \\ 0 & 1 & 0 & 0 & & & & & & \\ 0 & 0 & 1 & 0 & & & & & & \\ 0 & 0 & 0 & 1 & & & & & & \\ & & & & \ddots & & & & & \\ & & & & & & 1 & 0 & 0 & 0 \\ & & & & & & 0 & 1 & 0 & 0 \\ & & & & & & 0 & 0 & 1 & 0 \\ 0 & & & & & & 0 & 0 & 0 & 1 \end{pmatrix}_{4(20)\times 4(20)}$$

$$= \begin{pmatrix} \sigma_b^2+\sigma_e^2 & \sigma_b^2 & \sigma_b^2 & \sigma_b^2 & & & & & & 0 \\ \sigma_b^2 & \sigma_b^2+\sigma_e^2 & \sigma_b^2 & \sigma_b^2 & & & & & & \\ \sigma_b^2 & \sigma_b^2 & \sigma_b^2+\sigma_e^2 & \sigma_b^2 & & & & & & \\ \sigma_b^2 & \sigma_b^2 & \sigma_b^2 & \sigma_b^2+\sigma_e^2 & & & & & & \\ & & & & \ddots & & & & & \\ & & & & & \sigma_b^2+\sigma_e^2 & \sigma_b^2 & \sigma_b^2 & \sigma_b^2 & \\ & & & & & \sigma_b^2 & \sigma_b^2+\sigma_e^2 & \sigma_b^2 & \sigma_b^2 & \\ & & & & & \sigma_b^2 & \sigma_b^2 & \sigma_b^2+\sigma_e^2 & \sigma_b^2 & \\ 0 & & & & & \sigma_b^2 & \sigma_b^2 & \sigma_b^2 & \sigma_b^2+\sigma_e^2 & \end{pmatrix}_{4(20)\times 4(20)}$$

Different forms of the Cov(Y)

$$Var(\mathbf{Y}) = \mathbf{ZGZ}^T + \mathbf{R} = \begin{pmatrix} \sigma_b^2+\sigma_e^2 & \sigma_b^2 & \sigma_b^2 & \sigma_b^2 & & & & & 0 \\ \sigma_b^2 & \sigma_b^2+\sigma_e^2 & \sigma_b^2 & \sigma_b^2 & & & & & \\ \sigma_b^2 & \sigma_b^2 & \sigma_b^2+\sigma_e^2 & \sigma_b^2 & & & & & \\ \sigma_b^2 & \sigma_b^2 & \sigma_b^2 & \sigma_b^2+\sigma_e^2 & & & & & \\ & & & & \ddots & & & & \\ & & & & & \sigma_b^2+\sigma_e^2 & \sigma_b^2 & \sigma_b^2 & \sigma_b^2 \\ & & & & & \sigma_b^2 & \sigma_b^2+\sigma_e^2 & \sigma_b^2 & \sigma_b^2 \\ & & & & & \sigma_b^2 & \sigma_b^2 & \sigma_b^2+\sigma_e^2 & \sigma_b^2 \\ 0 & & & & & \sigma_b^2 & \sigma_b^2 & \sigma_b^2 & \sigma_b^2+\sigma_e^2 \end{pmatrix}_{4(20)\times 4(20)}$$

- Different structures for the variance covariance matrix
  - Allow for random intercept and random slope
  - Allow for different covariance structures
  - RMANOVA (not different)