Lab 6
Overview of the Linear Mixed Effects Model

Linear mixed effects model

Recall that in ordinary linear regression, we observe n independent random variables Y_i ($i = 1, ..., n$) such that $Y_i = X_i'\beta + \epsilon_i$ with $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ (Lecture 2, Slide 7). In the setting of longitudinal data, we have not one, but multiple observations for the same subject:

$$Y_i = X_i'\beta + \epsilon_i \longrightarrow Y_{ij} = X_{ij}'\beta + \epsilon_{ij}$$

Repeated measures on the same individuals are likely to be correlated and must be accounted for to obtain valid inferences (Lecture 1, Slide 6). Linear mixed effects models extend traditional linear models by allowing a subset of the regression parameters to vary randomly from one individual to another, thereby accounting for sources of natural heterogeneity in the population that arise from both between- and within- subject variability (Lecture 10, Slide 17):

$$Y_{ij} = X_{ij}'\beta + \epsilon_{ij} \longrightarrow Y_{ij} = X_{ij}'\beta + Z_{ij}'b_i{}' + \epsilon_{ij}$$

Specifically, by assuming that $b_i \overset{iid}{\sim} N(0, G)$ and $\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$, we allow the mean model for each subject (aka conditional mean)

$$E(Y_{ij}|X_{ij}, b_i) = X_{ij}'\beta + Z_{ij}'b_i$$

to differ from the population mean model (aka marginal mean)

$$E(Y_{ij}|X_{ij}) = X_{ij}'\beta$$

by $Z_{ij}'b_i$ (Lecture 10, Slides 33). Furthermore, each observation Y_{ij} is allowed to vary above or below its subject-specific mean model by ϵ_{ij}. You can therefore think of G as effectively capturing the between-subject variability and σ^2 as the within-subject error variance.

General interpretation of the parameters

- β_1: mean response for baseline group (averaged over subjects)
- β_k: change in mean response for every unit change in X_{ijk}, conditional on all $X_{ijk'}$, $k' \neq k$ (and averaged over subjects)
- b_{1i}, b_{2i}: 1st and 2nd random effect for individual i
- $Var(b_{1i}) = G_{11}$, $Var(b_{2i}) = G_{22}$: between-subject variance of the 1st and 2nd random effect
- $Cov(b_{1i}, b_{2i}) = G_{12} = G_{21}$: between-subject covariance of the 1st and 2nd random effect
- σ^2: within-subject error variance
- $Cov(Y_i) = Z_i G Z_i' + \sigma^2 I_{n_i}$: marginal covariance of the observations

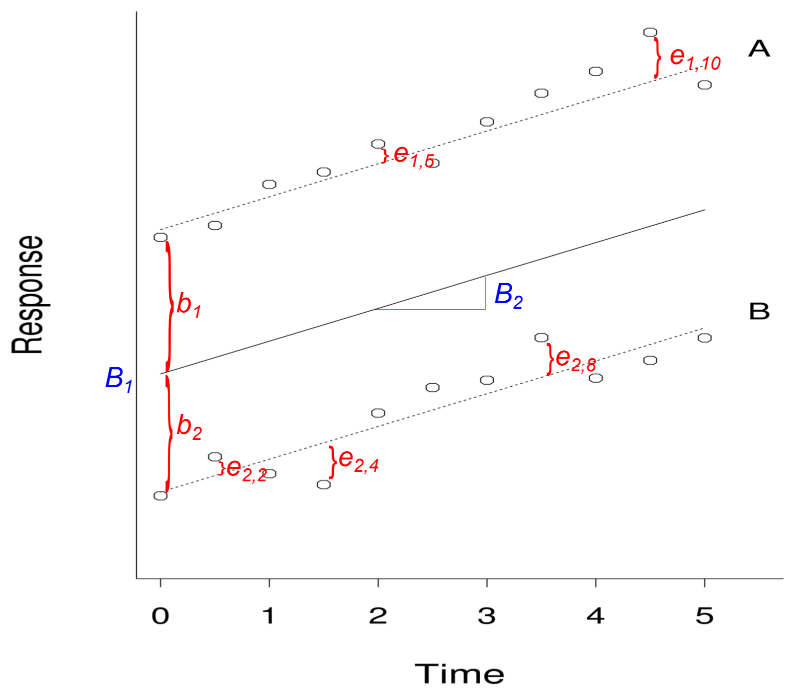
Example : Random intercept model

Consider the linear trend model from Lecture 10, Slide 24:

$$
\begin{aligned}
Y_{ij} &= \begin{pmatrix} 1 & t_{ij} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} 1 \end{pmatrix} \begin{pmatrix} b_i \end{pmatrix} + \epsilon_{ij} \\
&= (\beta_1 + b_i) + \beta_2 t_{ij} + \epsilon_{ij}
\end{aligned}
$$

Under this model,

- subject-specific mean responses for individuals A and B (broken lines) deviate from the population trend (solid line) by b_1 and b_2.
- inclusion of measurement errors, ϵ_{ij}, allows responses at any occasion to vary randomly above or below subject-specific trajectories.



Example: Random intercept and slope model

Consider the mixed effects analysis from Homework 5, Part 2 (also 2014 Midterm, Question 3):

$$Y_{ij} = \begin{pmatrix} 1 & gender_i & age_{ij} & gender_i * age_{ij} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} + \begin{pmatrix} 1 & age_{ij} \end{pmatrix} \begin{pmatrix} b_{1i} \\ b_{2i} \end{pmatrix} + \epsilon_{ij}$$
$$= (\beta_1 + b_{1i}) + \beta_2 gender_i + (\beta_3 + b_{2i}) age_{ij} + \beta_4 gender_i * age_{ij} + \epsilon_{ij}$$

Based on the SAS output below,

- Within-subject error variance = $\hat{\sigma}^2 = 1.7162$
- Between-subject variance of the intercepts = $\widehat{G}_{11} = 5.7864$
- Difference in mean intercept by gender = $\hat{\beta}_2 = 1.0321$
- Difference in mean slope by gender = $\hat{\beta}_4 =$ -0.3048
- Standard deviation of the slopes among subjects, conditional upon gender = $\sqrt{\widehat{G}_{22}} = 0.1803$
- Linear rate of change over time in the mean response for males = $\hat{\beta}_3 = 0.7844$
- Linear rate of change over time in the mean response for females = $\hat{\beta}_3 + \hat{\beta}_4 = 0.7844 +$ (-0.3048) = 0.4796
- Estimated mean response at age 10 for males = $\hat{\beta}_1 + 10\hat{\beta}_3 = 16.3406 + 10(0.7844) = 24.1846$
- Estimated mean response at age 10 for females = $(\hat{\beta}_1 + \hat{\beta}_2) + 10(\hat{\beta}_3 + \hat{\beta}_4) = (16.3406 + 1.0321) + 10(0.4796) = 22.1687$
- Range in which 95% of the linear rates of change in males fall = $\hat{\beta}_3 \pm 1.96 * \sqrt{\widehat{G}_{22}} = 0.7844 \pm 1.96 * 0.1803$
- 95% confidence interval for the linear rate of change in the mean response for males = $\hat{\beta}_3 \pm 1.96 * SE(\hat{\beta}_3) = 0.7844 \pm 1.96 * 0.0860$
- The predicted linear change over time for child 1, who happens to be female = $\hat{\beta}_3 + \hat{\beta}_4 + \hat{b}_{21} = 0.7844 + (-0.3048) + (-0.04475) = 0.4349$

The SAS System

The Mixed Procedure

Model Information

Data Set	WORK.DENTAL2
Dependent Variable	y
Covariance Structure	Unstructured
Subject Effect	id
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

Dimensions

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

Number of Observations

Number of Observations Read	108
Number of Observations Used	108
Number of Observations Not Used	0

Estimated G Matrix (Btwn subject var-cov matrix for random effects)

Row	Effect	Subject	Col1	Col2
1	Intercept	1	5.7864	-0.2896
2	age	1	-0.2896	0.03252

Estimated G Correlation Matrix

Row	Effect	Subject	Col1	Col2
1	Intercept	1	1.0000	-0.6676
2	age	1	-0.6676	1.0000

Estimated V Matrix for Subject 1

Row	Col1	Col2	Col3	Col4
1	4.9502	3.1751	3.1162	3.0574
2	3.1751	4.9625	3.3176	3.3888
3	3.1162	3.3176	5.2351	3.7202
4	3.0574	3.3888	3.7202	5.7679

The SAS System

The Mixed Procedure

Estimated V Correlation Matrix for Subject 1

Row	Col1	Col2	Col3	Col4
1	1.0000	0.6406	0.6122	0.5722
2	0.6406	1.0000	0.6509	0.6334
3	0.6122	0.6509	1.0000	0.6770
4	0.5722	0.6334	0.6770	1.0000

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
UN(1,1)	id	5.7864
UN(2,1)	id	-0.2896

UN(2,2)	id	0.03252
Residual		1.7162 (Within-subject error variance)

Fit Statistics

-2 Res Log Likelihood	432.6
AIC (smaller is better)	440.6
AICC (smaller is better)	441.0
BIC (smaller is better)	445.8

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
3	50.98	<.0001

Solution for Fixed Effects

Effect	gender	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		16.3406	1.0185	25	16.04	<.0001
gender	F	1.0321	1.5957	54	0.65	0.5205
gender	M	0	.	.	.	.
age		0.7844	0.08600	25	9.12	<.0001
age*gender	F	-0.3048	0.1347	54	-2.26	0.0277
age*gender	M	0	.	.	.	.

Solution for Random Effects

Effect	Subject	Estimate	Std Err Pred	DF	t Value	Pr > |t|
Intercept	1	-0.6413	1.8112	54	-0.35	0.7247
age	1	-0.04475	0.1543	54	-0.29	0.7729
Intercept	2	-0.6602	1.8112	54	-0.36	0.7169
age	2	0.09029	0.1543	54	0.59	0.5608
Intercept	3	-0.2489	1.8112	54	-0.14	0.8912
age	3	0.1136	0.1543	54	0.74	0.4649
Intercept	4	1.6611	1.8112	54	0.92	0.3632
age	4	0.02821	0.1543	54	0.18	0.8556
Intercept	5	0.5710	1.8112	54	0.32	0.7538
age	5	-0.05496	0.1543	54	-0.36	0.7230

.
.
.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
gender	1	54	0.42	0.5205
age	1	25	88.00	<.0001
age*gender	1	54	5.12	0.0277