Linear Mixed Models

Objectives:

- Fixed and random effects
- Likelihood inference using MV normal
- Model specification and interpretation

Linear Mixed Model & ML / REML Estimation

○ Linear Mixed Model

- Combines random effects and serial structures.
- Clustered and nested data.
- Mean model and covariance model.

○ Maximum Likelihood

- Assuming multivariate normality.
- $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\Sigma}}$.
- Matrix derivatives.

○ Restricted Maximum Likelihood

- Correction for degrees of freedom in $\widehat{\boldsymbol{\beta}}$.
- Bayesian motivation.

Motivation

Cystic Fibrosis and Pulmonary Function

- Several specific aspects are of interest:

1. What is the rate of decline in FEV1?
2. Is the time course different for males and females?
3. Is the time course different for F508 homozygous subjects ?

- **Reference**: Davis P.B. (1997) *Journal of Pediatrics*

Data

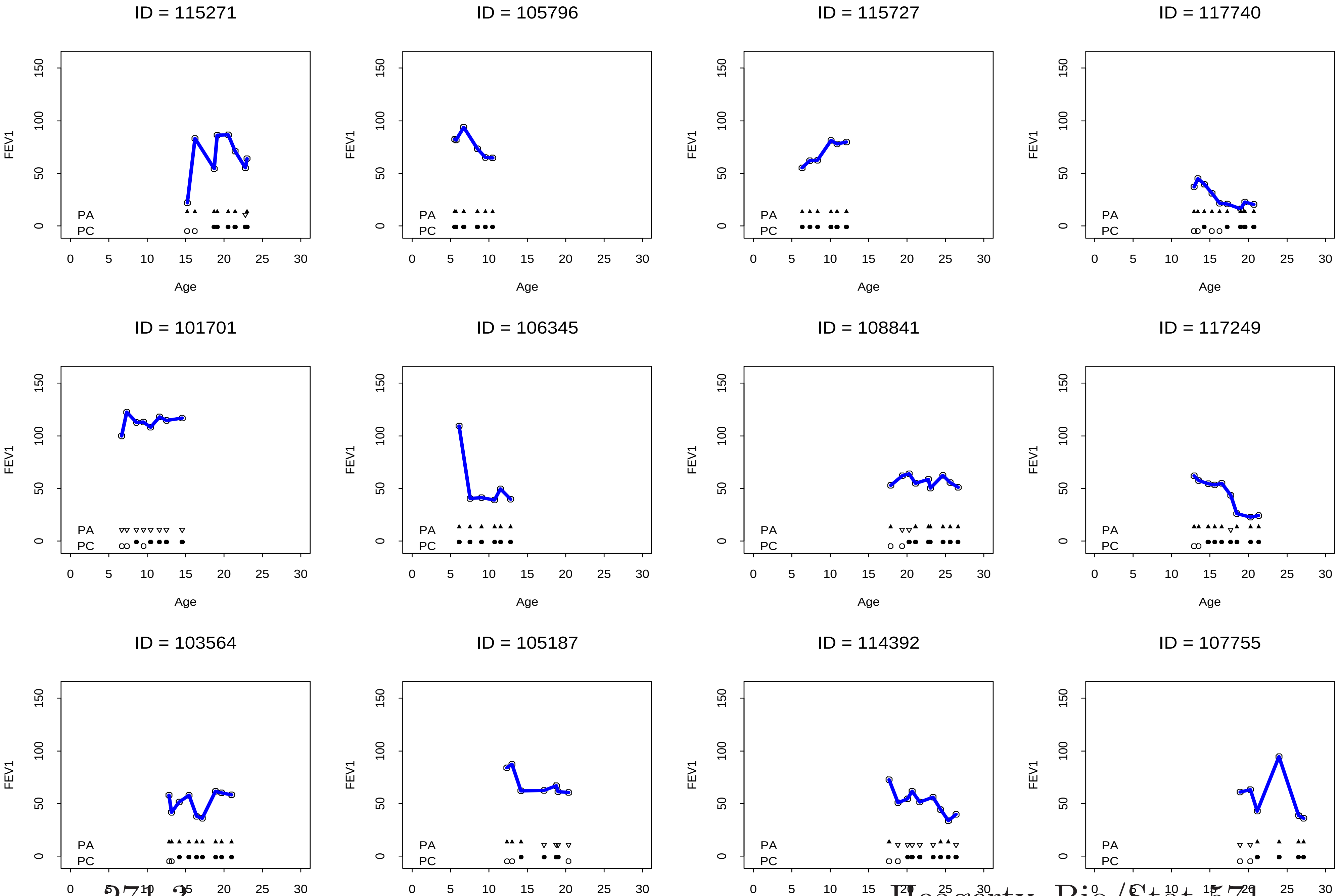
```
  ID        = patient id
  FEV1      = percent-predicted forced expiratory volume in 1 second
  AGE       = age (years)
  GENDER    = sex (1=male, 2=female)
  PSEUDOA   = infection with Pseudomonas Aeruginosa (0=no, 3=yes)
  F508      = genotype (1=homozygous, 2=heterozygous, 3=none)
  PANCREAT  = pancreatic enzyme supplmentation (0,1=no, 2=yes)

100073 113.8   8.452 2 3 1 2
100073  98.18  8.783 2 3 1 2
100073  98.73  9.785 2 3 1 2
100073 101.79 10.538 2 3 1 2
100073  98.04 12.329 2 3 1 2
100073  94.32 13.306 2 3 1 2
100073  95.48 14.418 2 3 1 2
100111  96.85 12.515 2 0 3 1
100111 101.05 13.103 2 0 3 2
100111 100.33 15.105 2 0 3 2
100111  90.92 16.838 2 0 3 2
100111 109.78 17.582 2 0 3 2
100111 107.76 18.847 2 0 3 1
```

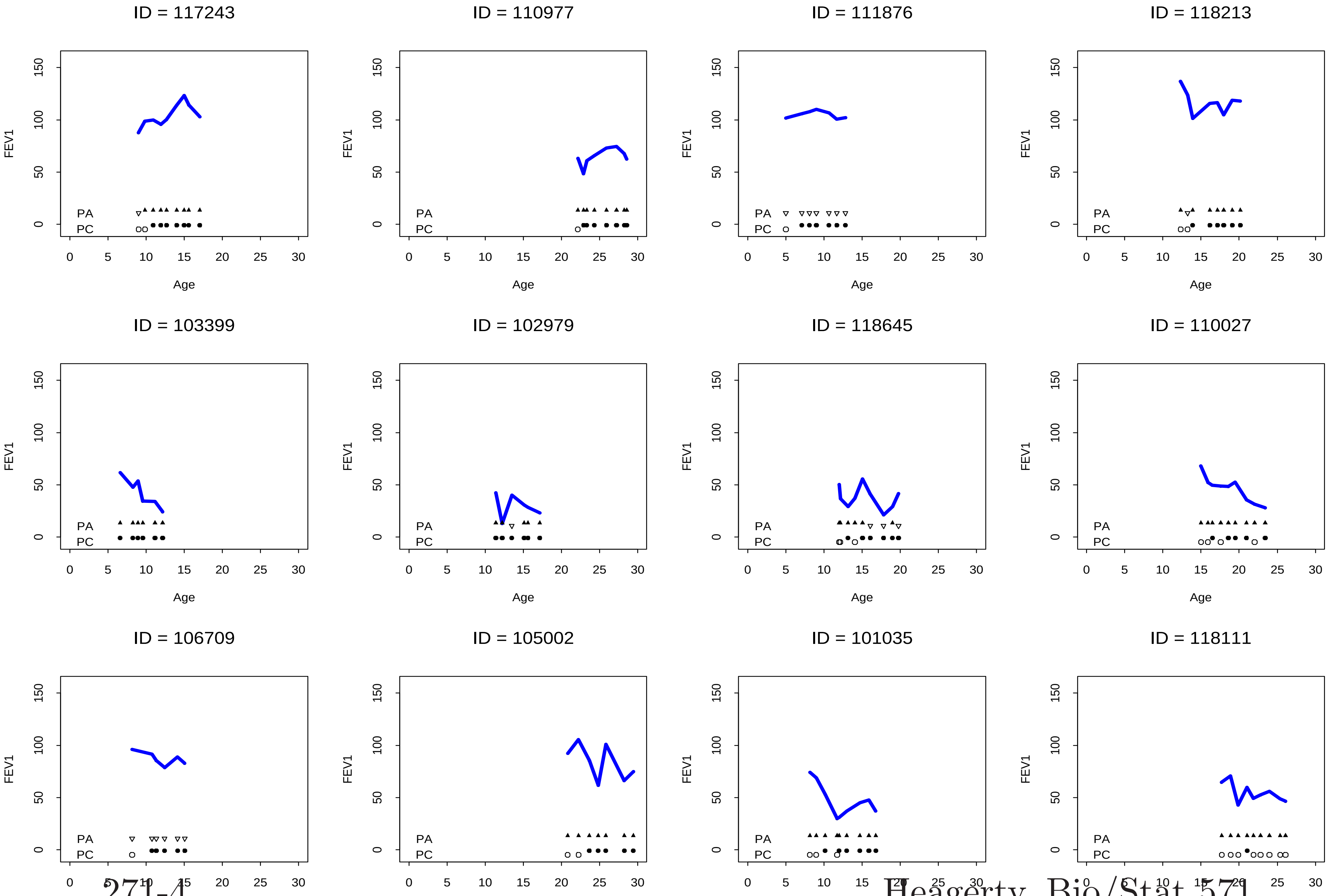
EDA: Numerical Summaries

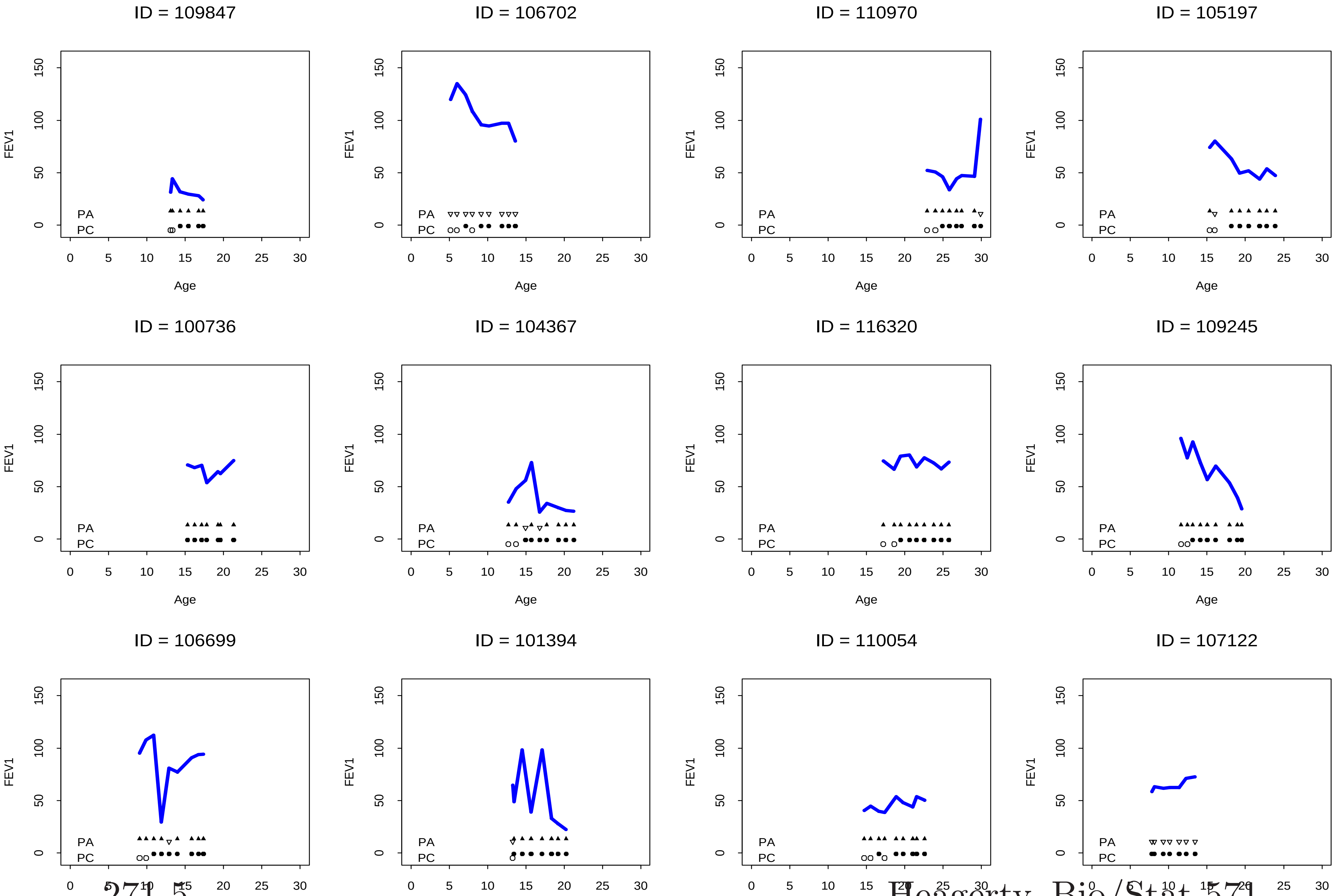
```
Total number of subjects = 200


Number of observations (number of subjects with ni):
  6  7  8  9
 49 52 36 63


Distribution of males / females
 male female
  102     98


Number of mutations of f508
  0  1  2
 23 87 90
```

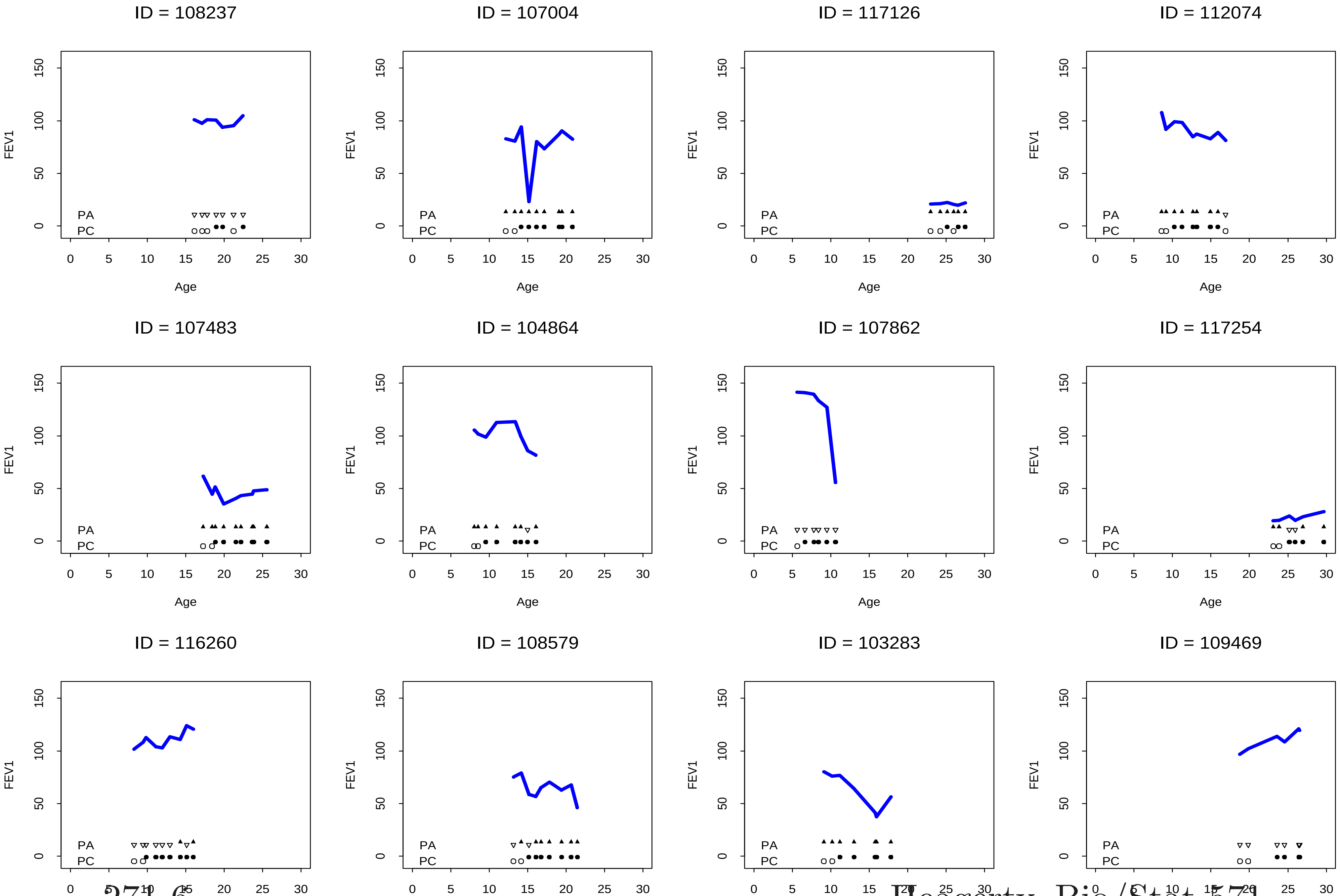
EDA: Numerical Summaries

```
Age at entry

N = 200   Median = 11.9655
Quartiles = 7.758, 15.3235

Decimal point is at the colon

    5 : 002355666788889
    6 : 01112223345555567789999
    7 : 0001234446778889
    8 : 011223345566899
    9 : 00011244788
   10 : 0111113349
   11 : 2223446678
   12 : 0011122233445557788888999
   13 : 01234455
   14 : 111245555779
   15 : 001223357
   16 : 0012347899
   17 : 1223567779
   18 : 4899
   19 : 4
   20 : 0123778
   21 : 15577
   22 : 2459
   23 : 001128
```

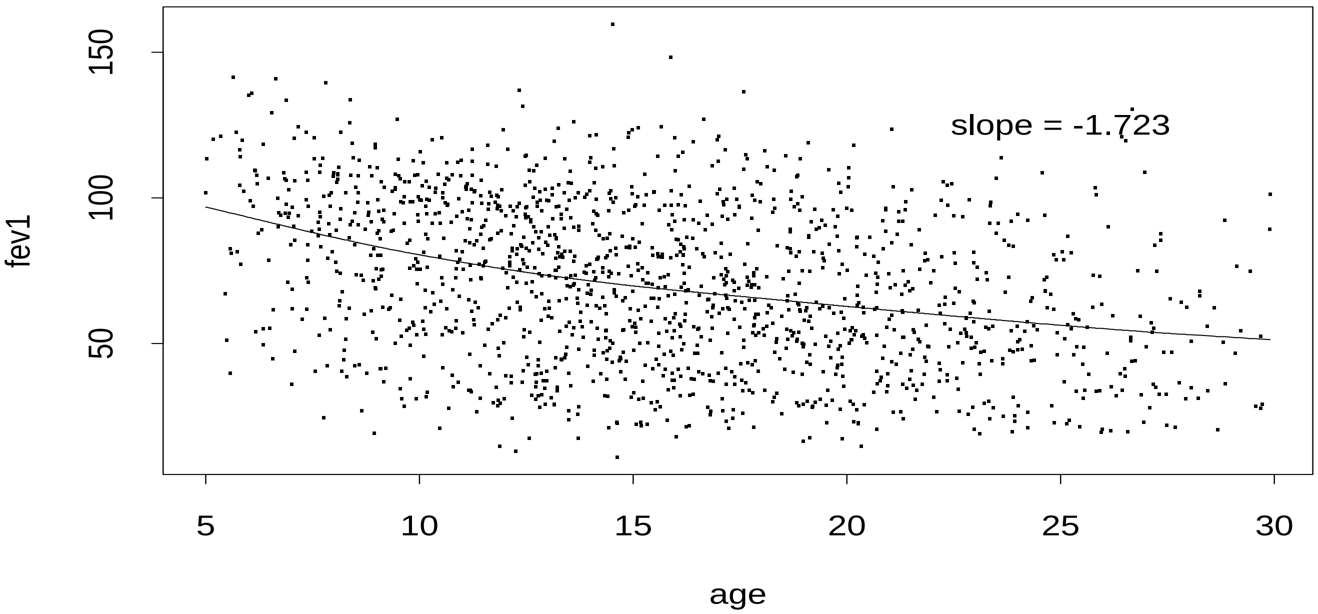

ID = 115271
ID = 105796
ID = 115727
ID = 117740
ID = 101701
ID = 106345
ID = 108841
ID = 117249
ID = 103564
ID = 105187
ID = 114392
ID = 107755
FEV1
Age
PA
PC


ID = 117243
ID = 110977
ID = 111876
ID = 118213
ID = 103399
ID = 102979
ID = 118645
ID = 110027
ID = 106709
ID = 105002
ID = 101035
ID = 118111
FEV1
Age
PA
PC

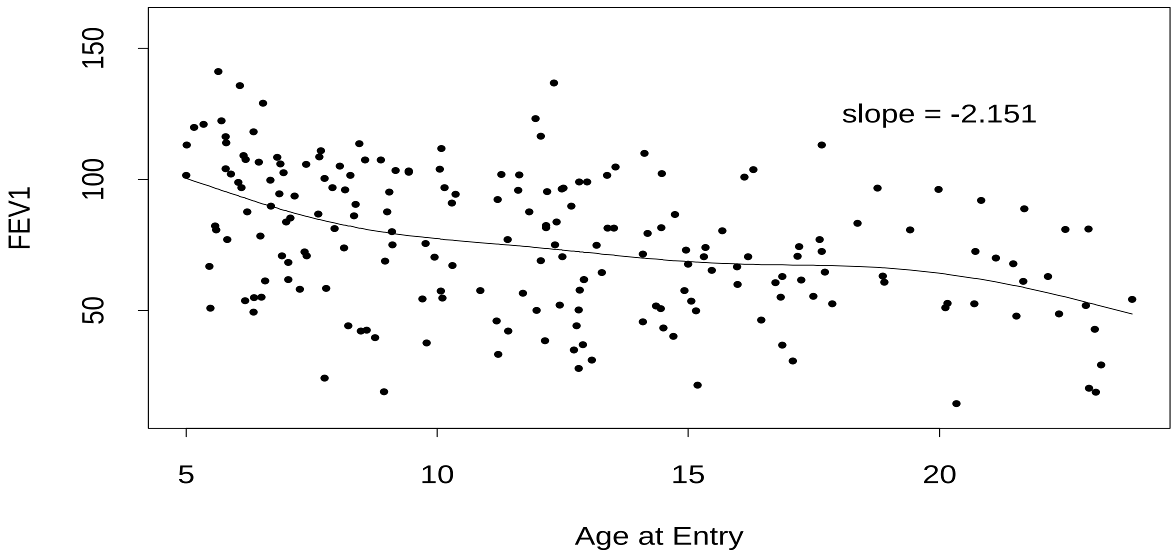

ID = 109847
ID = 106702
ID = 110970
ID = 105197
ID = 100736
ID = 104367
ID = 116320
ID = 109245
ID = 106699
ID = 101394
ID = 110054
ID = 107122
FEV1
Age
PA
PC
0 50 100 150
0 5 10 15 20 25 30


ID = 108237
ID = 107004
ID = 117126
ID = 112074
ID = 107483
ID = 104864
ID = 107862
ID = 117254
ID = 116260
ID = 108579
ID = 103283
ID = 109469
FEV1
Age
PA
PC

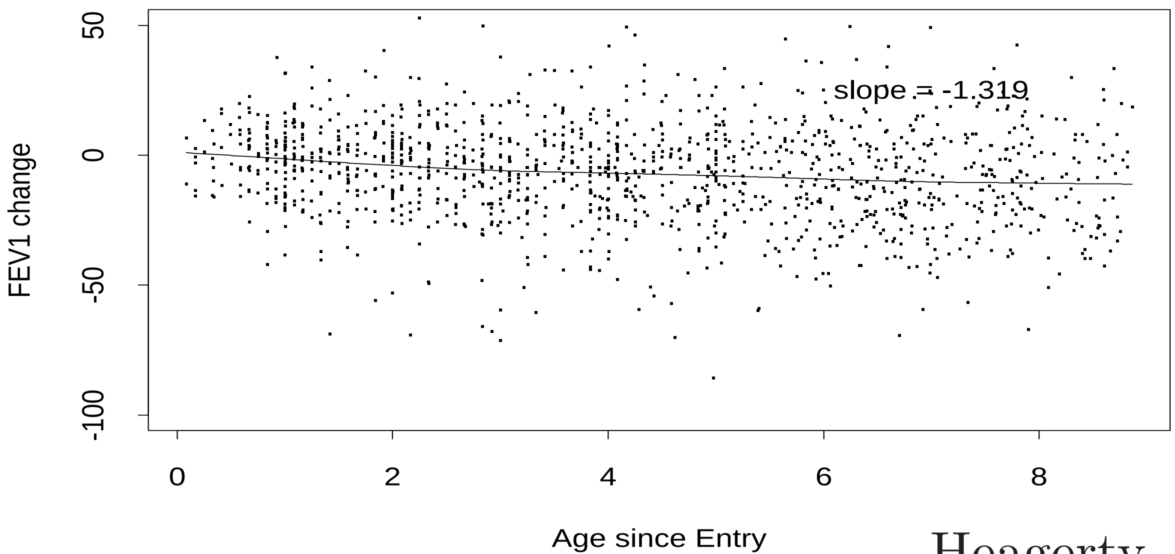
FEV1 versus Age

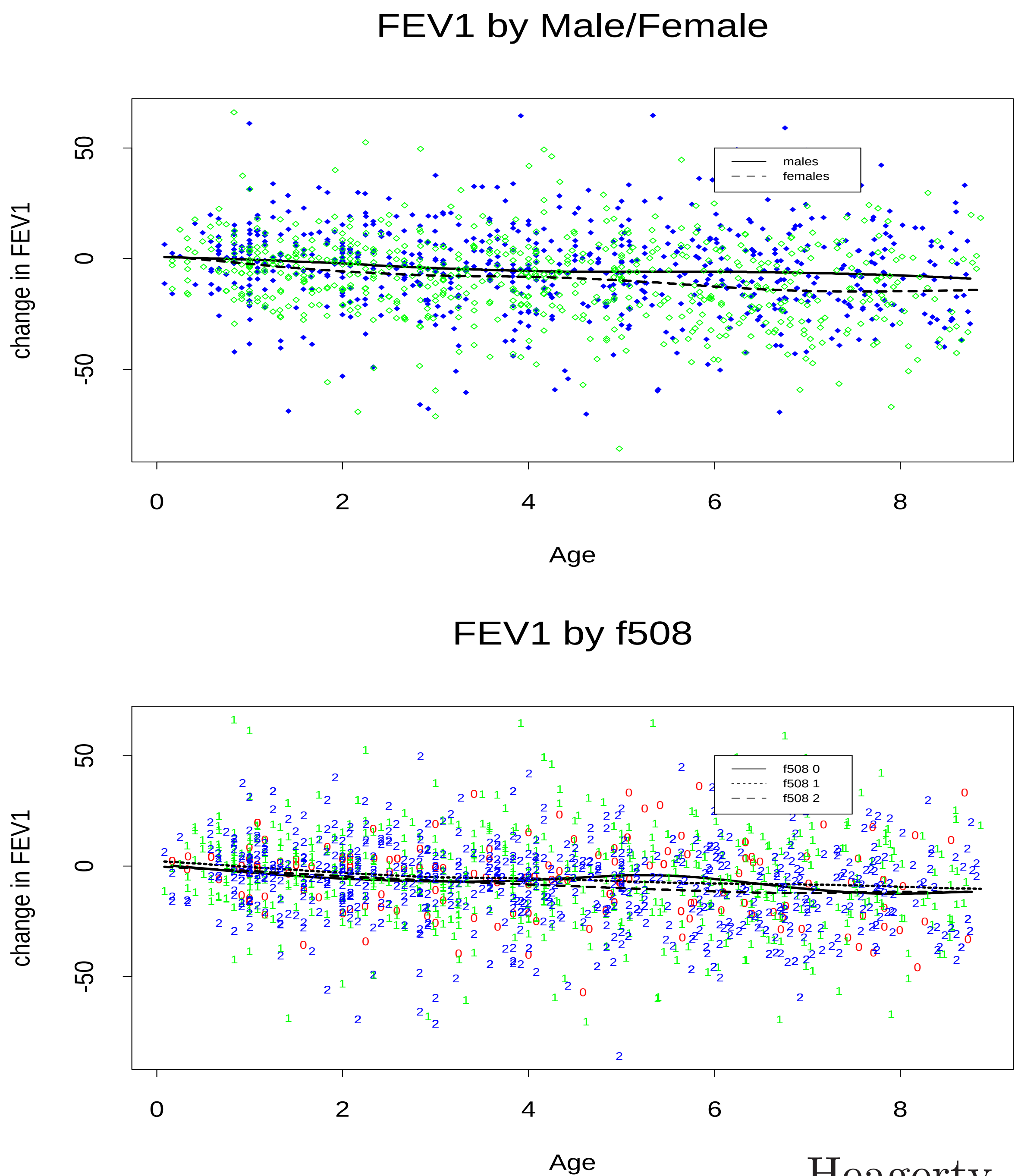


FEV1 versus Age-at-Entry



FEV1 Change versus Age-since-Entry




FEV1 by Male/Female
change in FEV1
Age
males
females
50
0
-50
0
2
4
6
8
FEV1 by f508
change in FEV1
Age
f508 0
f508 1
f508 2
50
0
-50
0
2
4
6
8

EDA Summary

Observations

- Systematic trends: time, gender, F508.
- Random variation: individual, observation.

Questions

- Two time scales?
- Estimation / testing for rates of decline?
- Other?

Linear Mixed Model

- **Regression model:**
  mean response as a function of covariates.
  "systematic variation"

- **Random effects:**
  variation from subject-to-subject in trajectory.
  "random between-subject variation"

- **Within-subject variation:**
  variation of individual observations over time
  "random within-subject variation"

Scientific Questions as Regression

$\boxed{\star}$ Questions concerning the rate of decline refer to the time slope for FEV1:

$$E[\text{ FEV1 } \mid \boldsymbol{X} = \text{age, gender, f508}] \quad = \quad \beta_0(\boldsymbol{X}) + \beta_1(\boldsymbol{X}) \cdot \text{time}$$

CF Regression Model

Model:

$$
\begin{aligned}
E[\texttt{FEV} \mid \boldsymbol{X}_i] &= \beta_0 \\
&\quad + \beta_1 \cdot \texttt{age0} + \beta_2 \cdot \texttt{ageL} \\
&\quad + \beta_3 \cdot \texttt{female} \\
&\quad + \beta_4 \cdot \texttt{f508} = \texttt{1} + \beta_5 \cdot \texttt{f508} = \texttt{2} \\
&\quad + \beta_6 \cdot \texttt{female} \cdot \texttt{ageL} \\
&\quad + \beta_7 \cdot \texttt{f508} = \texttt{1} \cdot \texttt{ageL} + \beta_8 \cdot \texttt{f508} = \texttt{2} \cdot \texttt{ageL} \\
&= \beta_0(\boldsymbol{X}_i) + \beta_1(\boldsymbol{X}_i) \cdot \texttt{ageL}
\end{aligned}
$$

Intercept

	f508=0	f508=1	f508=2
male	$\beta_0 + \beta_1 \cdot \mathsf{age0}$	$\beta_0 + \beta_1 \cdot \mathsf{age0}$ $+\beta_4$	$\beta_0 + \beta_1 \cdot \mathsf{age0}$ $+\beta_5$
female	$\beta_0 + \beta_1 \cdot \mathsf{age0}$ $+\beta_3$	$\beta_0 + \beta_1 \cdot \mathsf{age0}$ $+\beta_3 + \beta_4$	$\beta_0 + \beta_1 \cdot \mathsf{age0}$ $+\beta_3 + \beta_5$

Slope

	f508=0	f508=1	f508=2
male	β_2	$\beta_2 + \beta_7$	$\beta_2 + \beta_8$
female	β_2 $+\beta_6$	$\beta_2 + \beta_7$ $+\beta_6$	$\beta_2 + \beta_8$ $+\beta_6$

Define

$$
\begin{aligned}
Y_{ij} &= \text{FEV1 for subject } i \text{ at time } t_{ij} \\
\boldsymbol{X}_i &= (\boldsymbol{X}_{ij}, \ldots, \boldsymbol{X}_{in_i}) \\
\boldsymbol{X}_{ij} &= (X_{ij,1}, X_{ij,2}, \ldots, X_{ij,p}) \\
&\quad \text{age0, ageL, gender, genotype}
\end{aligned}
$$

Issue: Response variables measured on the same subject are correlated.

$$\text{cov}(Y_{ij}, Y_{ik} \mid \boldsymbol{X}_i) \neq 0$$

Mean and Covariance Models for FEV1

Models:

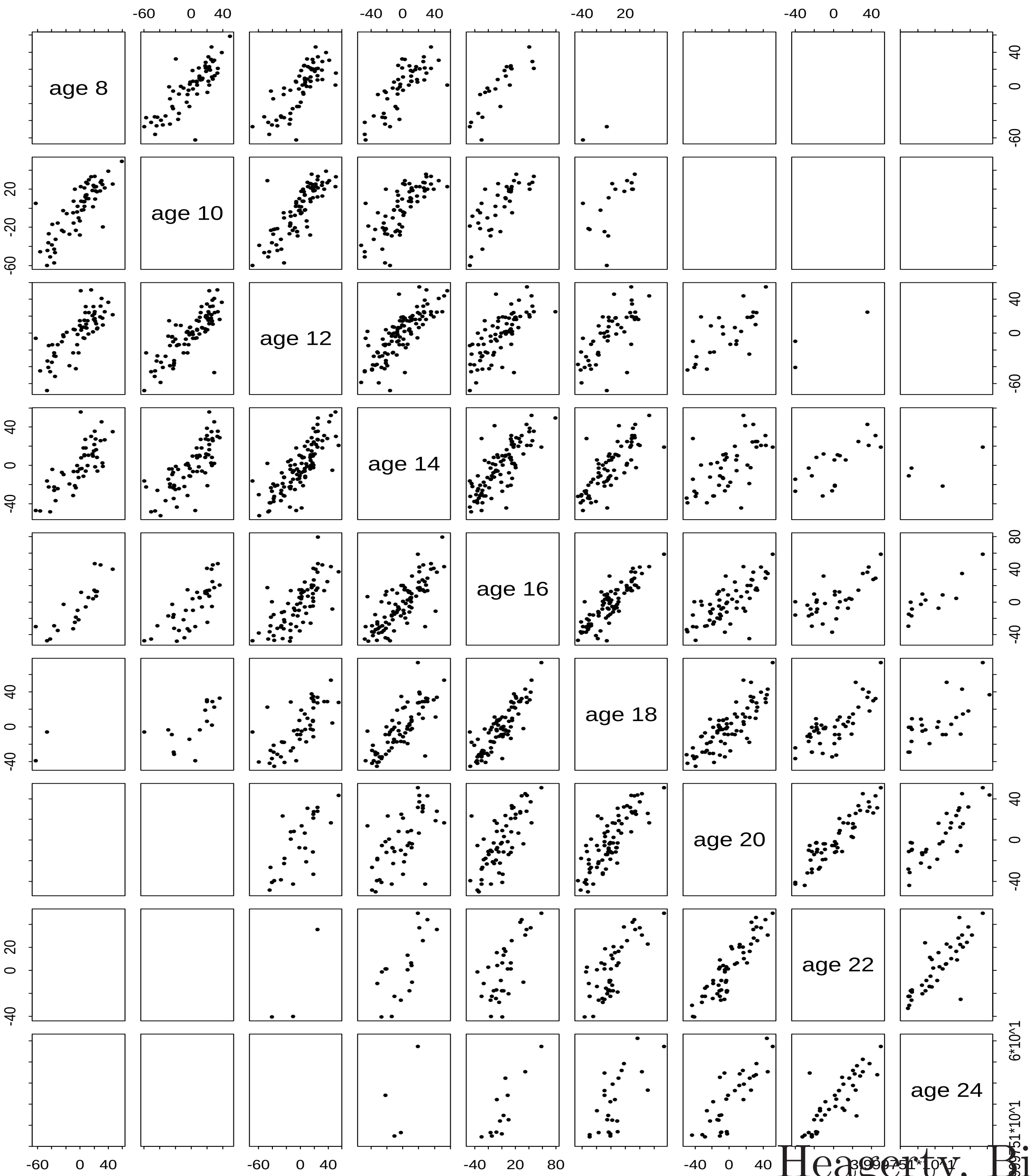
$$E(Y_{ij} \mid \boldsymbol{X}_i) \quad = \quad \mu_{ij} \ \ (\texttt{regression})$$

$$\text{cov}(\boldsymbol{Y}_i \mid \boldsymbol{X}_i) \quad = \quad \boldsymbol{\Sigma}_i = \underbrace{\boldsymbol{Z}_i \boldsymbol{D} \boldsymbol{Z}_i^T}_{\texttt{between-subjects}} \ + \ \underbrace{\boldsymbol{R}_i}_{\texttt{within-subjects}}$$

Q: What are appropriate covariance models for the FEV1 data?

“Wide Data”

```
Data array for the residuals (first 10 rows of ‘‘rmat’’)

      age8  age10 age12  age14  age16  age18  age20  age22 age24
 [1,] 34.64 21.90 25.66  26.75     NA     NA     NA     NA   NA
 [2,]    NA    NA 16.48  21.68  17.94  38.07     NA     NA   NA
 [3,]    NA    NA    NA -34.29 -34.06 -32.15 -49.71     NA   NA
 [4,]    NA    NA    NA     NA     NA     NA     NA -14.01 -6.1
 [5,]    NA    NA    NA  41.59 -11.18  11.36  18.68     NA   NA
 [6,] 20.55 22.71 22.78  21.55  14.61     NA     NA     NA   NA
 [7,]    NA    NA    NA     NA   3.28  -8.39   3.38  19.04   NA
 [8,]    NA    NA    NA     NA  -8.56 -11.92 -19.37 -24.16   NA
 [9,] 39.82 39.16 36.79     NA     NA     NA     NA     NA   NA
[10,]    NA    NA 19.76  24.59  26.90  29.11  27.28     NA   NA
      .
      .
      .
```


age 8
age 10
age 12
age 14
age 16
age 18
age 20
age 22
age 24

```
Empirical Covariance Matrix:

       [,1]   [,2]   [,3]   [,4]   [,5]    [,6]   [,7]   [,8]   [,9]
[1,] 714.16 575.86 515.58 470.58 648.56 1346.10     NA     NA     NA
[2,]   0.00 633.62 523.40 444.44 467.09  383.60     NA     NA     NA
[3,]   0.00   0.00 681.02 504.08 520.35  514.70 492.00 973.33     NA
[4,]   0.00   0.00   0.00 579.61 523.86  493.50 404.72 395.84 315.28
[5,]   0.00   0.00   0.00   0.00 663.82  527.98 440.31 374.98 488.25
[6,]   0.00   0.00   0.00   0.00   0.00  597.54 501.07 406.39 462.75
[7,]   0.00   0.00   0.00   0.00   0.00    0.00 605.90 487.05 511.91
[8,]   0.00   0.00   0.00   0.00   0.00    0.00   0.00 512.90 437.70
[9,]   0.00   0.00   0.00   0.00   0.00    0.00   0.00   0.00 618.17


Empirical Correlation Matrix:

     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
[1,]    1 0.86 0.74 0.73 0.94 2.06   NA   NA   NA
[2,]    0 1.00 0.80 0.73 0.72 0.62   NA   NA   NA
[3,]    0 0.00 1.00 0.80 0.77 0.81 0.77 1.65   NA
[4,]    0 0.00 0.00 1.00 0.84 0.84 0.68 0.73 0.53
[5,]    0 0.00 0.00 0.00 1.00 0.84 0.69 0.64 0.76
[6,]    0 0.00 0.00 0.00 0.00 1.00 0.83 0.73 0.76
[7,]    0 0.00 0.00 0.00 0.00 0.00 1.00 0.87 0.84
[8,]    0 0.00 0.00 0.00 0.00 0.00 0.00 1.00 0.78
[9,]    0 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1.00
```

```
Number of observations (pairs):

      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
[1,]    68   66   64   50   22    2    0    0    0
[2,]     0   87   83   69   39   16    0    0    0
[3,]     0    0  121  105   73   50   27    3    0
[4,]     0    0    0  127   93   70   45   19    4
[5,]     0    0    0    0  111   85   61   33   13
[6,]     0    0    0    0    0  102   77   47   25
[7,]     0    0    0    0    0    0   85   54   33
[8,]     0    0    0    0    0    0    0   66   42
[9,]     0    0    0    0    0    0    0    0   48
```

How to build models for correlation?

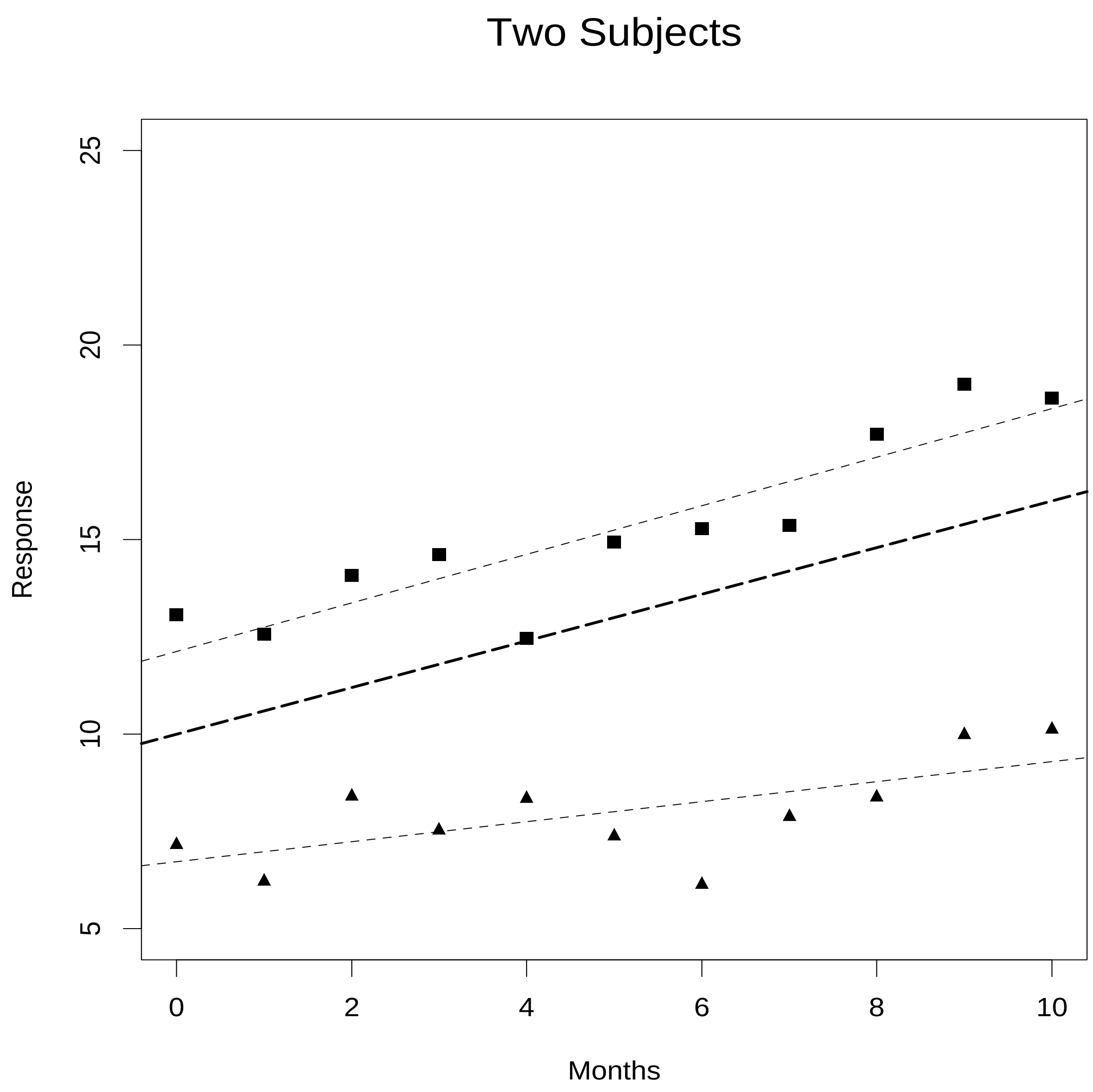
- Mixed models
  - ▷ "random effects"
  - ▷ within-subject similarity due to sharing trajectory
- Serial correlation
  - ▷ close in time implies strong similarity
  - ▷ correlation decreases as time separation increases

Linear Mixed Model

- **Regression model:**
  mean response as a function of covariates.
  "systematic variation"

- **Random effects:**
  variation from subject-to-subject in trajectory.
  "random between-subject variation"

- **Within-subject variation:**
  variation of individual observations over time
  "random within-subject variation"



Two Subjects
Response
25
20
15
10
5
0
2
4
6
8
10
Months

Levels of Analysis

- We first consider the distribution of **measurements** within **subjects**:

$$
\begin{aligned}
Y_{ij} &= \beta_{0,i} + \beta_{1,i} \cdot t_{ij} + e_{ij} \\
e_{ij} &\sim \mathcal{N}(0, \sigma^2) \\
E[\boldsymbol{Y}_i \mid \boldsymbol{X}_i, \boldsymbol{\beta}_i] &= \beta_{0,i} + \beta_{1,i} \cdot t_{ij} \\
&= [1, \texttt{time}_{ij}] \begin{bmatrix} \beta_{0,i} \\ \beta_{1,i} \end{bmatrix} \\
&= \boldsymbol{X}_i \boldsymbol{\beta}_i
\end{aligned}
$$

Levels of Analysis

- We can equivalently separate the subject-specific regression coefficients into the **average coefficient** and the **specific departure** for subject i:
  - $\beta_{0,i} = \beta_0 + b_{0,i}$
  - $\beta_{1,i} = \beta_1 + b_{1,i}$
- This allows another perspective:

$$
\begin{aligned}
Y_{ij} &= \beta_{0,i} + \beta_{1,i} \cdot t_{ij} + e_{ij} \\
&= (\beta_0 + \beta_1 \cdot t_{ij}) + (b_{0,i} + b_{1,i} \cdot t_{ij}) + e_{ij}
\end{aligned}
$$

$$
E[\boldsymbol{Y}_i \mid \boldsymbol{X}_i, \boldsymbol{\beta}_i] = \underbrace{\boldsymbol{X}_i \boldsymbol{\beta}}_{\textbf{mean model}} + \underbrace{\boldsymbol{X}_i \boldsymbol{b}_i}_{\textbf{between-subject}}
$$

Levels of Analysis

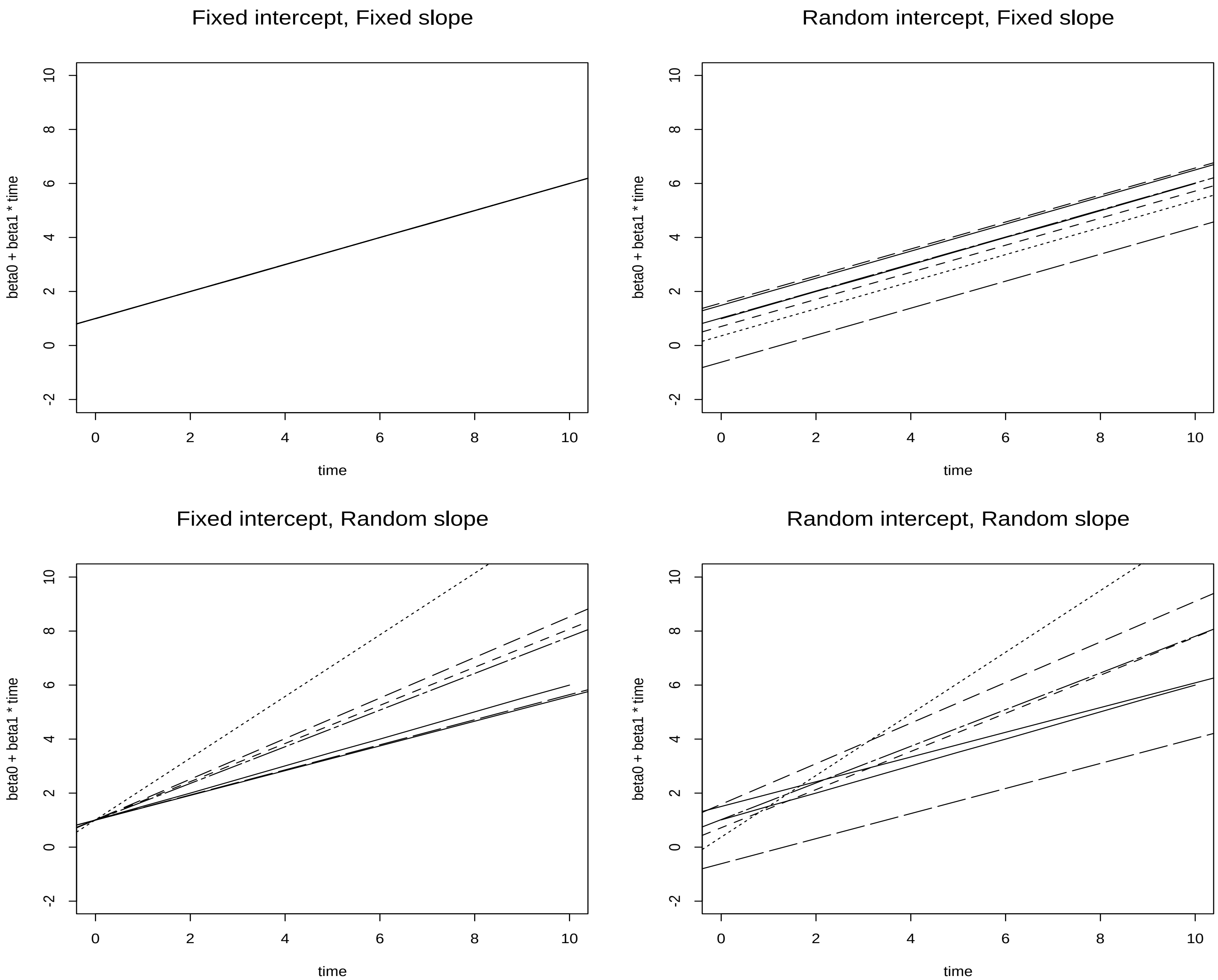
- Next we consider the distribution of **patterns (parameters)** among subjects:

$$\boldsymbol{\beta}_i \sim \mathcal{N}(\boldsymbol{\beta}, \boldsymbol{D})$$

equivalently

$$\boldsymbol{b}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{D})$$

$$\star\star\star \quad \boldsymbol{Y}_i = \underbrace{\boldsymbol{X}_i\boldsymbol{\beta}}_{\textbf{mean model}} + \underbrace{\boldsymbol{X}_i\boldsymbol{b}_i}_{\textbf{between-subject}} + \underbrace{\boldsymbol{e}_i}_{\textbf{within-subject}}$$


Fixed intercept, Fixed slope
Random intercept, Fixed slope
Fixed intercept, Random slope
Random intercept, Random slope
beta0 + beta1 * time
time
-2
0
2
4
6
8
10

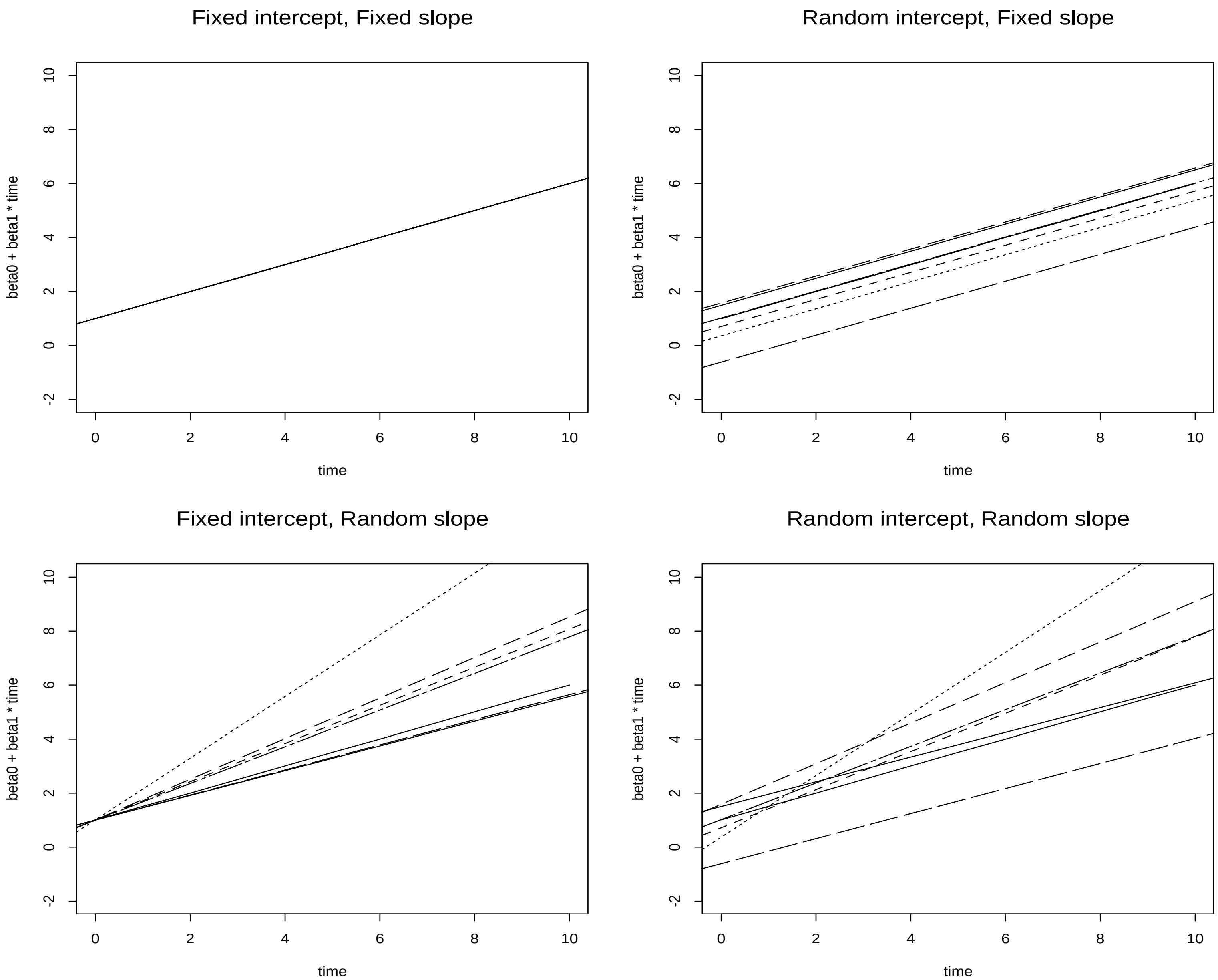
Between-subject Variation

- We can use the idea of random effects to allow different types of between-subject heterogeneity:
- The magnitude of heterogeneity is characterized by $\boldsymbol{D}$:

$$
\begin{aligned}
\boldsymbol{b}_i &= \begin{bmatrix} b_{0,i} \\ b_{1,i} \end{bmatrix} \\
\mathsf{var}(\boldsymbol{b}_i) &= \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}
\end{aligned}
$$

Between-subject Variation

- The components of $\boldsymbol{D}$ can be interpreted as:
  - $\sqrt{D_{11}}$ – the typical subject-to-subject deviation in the overall **level** of the response.
  - $\sqrt{D_{22}}$ – the typical subject-to-subject deviation in the **change** (time slope) of the response.
  - D_{12} – the covariance between individual intercepts and slopes.
    - If positive then subjects with **high levels** also have **high rates** of change.
    - If negative then subjects with **high levels** have **low rates** of change.



Fixed intercept, Fixed slope
Random intercept, Fixed slope
Fixed intercept, Random slope
Random intercept, Random slope
beta0 + beta1 * time
time
-2
0
2
4
6
8
10

Between-subject Variation: Examples

- No random effects:

$$
\begin{aligned}
Y_{ij} &= \beta_0 + \beta_1 \cdot t_{ij} + e_{ij} \\
&= [1, \texttt{time}_{ij}]\boldsymbol{\beta} + e_{ij}
\end{aligned}
$$

- Random intercepts:

$$
\begin{aligned}
Y_{ij} &= (\beta_0 + \beta_1 \cdot t_{ij}) + b_{0,i} + e_{ij} \\
&= [1, \texttt{time}_{ij}]\boldsymbol{\beta} + [\ 1\ ]b_{0,i} + e_{ij}
\end{aligned}
$$

- Random intercepts and slopes:

$$
\begin{aligned}
Y_{ij} &= (\beta_0 + \beta_1 \cdot t_{ij}) + b_{0,i} + b_{1,i} \cdot t_{ij} + e_{ij} \\
&= [1, \texttt{time}_{ij}]\boldsymbol{\beta} + [1, \texttt{time}_{ij}]\boldsymbol{b}_i + e_{ij}
\end{aligned}
$$

Mixed Models and Covariances/Correlation

- **Q**: What is the correlation between outcomes Y_{ij} and Y_{ik} under these random effects models?
- $\boxed{\text{Random Intercept Model}}$

$$
\begin{aligned}
Y_{ij} &= \beta_0 + \beta_1 t_{ij} + b_{0,i} + e_{ij} \\
Y_{ik} &= \beta_0 + \beta_1 t_{ik} + b_{0,i} + e_{ik}
\end{aligned}
$$

$$
\begin{aligned}
\mathsf{var}(Y_{ij}) &= \mathsf{var}(b_{0,i}) + \mathsf{var}(e_{ij}) \\
&= D_{11} + \sigma^2
\end{aligned}
$$

$$
\begin{aligned}
\mathsf{cov}(Y_{ij}, Y_{ik}) &= \mathsf{cov}(b_{0,i} + e_{ij}, b_{0,i} + e_{ik}) \\
&= D_{11}
\end{aligned}
$$

Mixed Models and Covariances/Correlation

- $\boxed{\text{Random Intercept Model}}$

$$
\begin{aligned}
\text{corr}(Y_{ij}, Y_{ik}) &= \frac{D_{11}}{\sqrt{D_{11} + \sigma^2}\sqrt{D_{11} + \sigma^2}} \\
&= \frac{D_{11}}{D_{11} + \sigma^2} = \frac{\textbf{between var}}{\textbf{between var} + \textbf{within var}}
\end{aligned}
$$

- Therefore, any two outcomes have the same correlation. Doesn't depend on the specific times, nor on the distance between the measurements.

- "**Exchangeable**" correlation model.

- Assuming: var$(e_{ij}) = \sigma^2$, and cov$(e_{ij}, e_{ik}) = 0$.

Mixed Models and Covariances/Correlation

- $\boxed{\text{Random Intercept and Slope Model}}$

$$\begin{aligned} Y_{ij} &= (\beta_0 + \beta_1 t_{ij}) + (b_{0,i} + b_{1,i} t_{ij}) + e_{ij} \\ Y_{ik} &= (\beta_0 + \beta_1 t_{ik}) + (b_{0,i} + b_{1,i} t_{ik}) + e_{ik} \end{aligned}$$

$$\begin{aligned} \text{var}(Y_{ij}) &= \text{var}(b_{0,i} + b_{1,i} t_{ij}) + \text{var}(e_{ij}) \\ &= D_{11} + 2 \cdot D_{12} t_{ij} + D_{22} t_{ij}^2 + \sigma^2 \end{aligned}$$

$$\begin{aligned} \text{cov}(Y_{ij}, Y_{ik}) &= \text{cov}[(b_{0,i} + b_{1,i} t_{ij} + e_{ij}), (b_{0,i} + b_{1,i} t_{ik} + e_{ik})] \\ &= D_{11} + D_{12}(t_{ij} + t_{ik}) + D_{22} t_{ij} t_{ik} \end{aligned}$$

Mixed Models and Covariances/Correlation

- $\boxed{\text{Random Intercept and Slope Model}}$

$$\begin{aligned} \rho_{ijk} &= \text{corr}(Y_{ij}, Y_{ik}) \\ &= \frac{D_{11} + D_{12}(t_{ij} + t_{ik}) + D_{22}t_{ij}t_{ik}}{\sqrt{D_{11} + 2 \cdot D_{12}t_{ij} + D_{22}t_{ij}^2 + \sigma^2}\sqrt{D_{11} + 2 \cdot D_{12}t_{ik} + D_{22}t_{ik}^2 + \sigma^2}} \end{aligned}$$

- Therefore, two outcomes may not have the same correlation. Correlation depends on the specific times for the observations, and does not have a simple form.

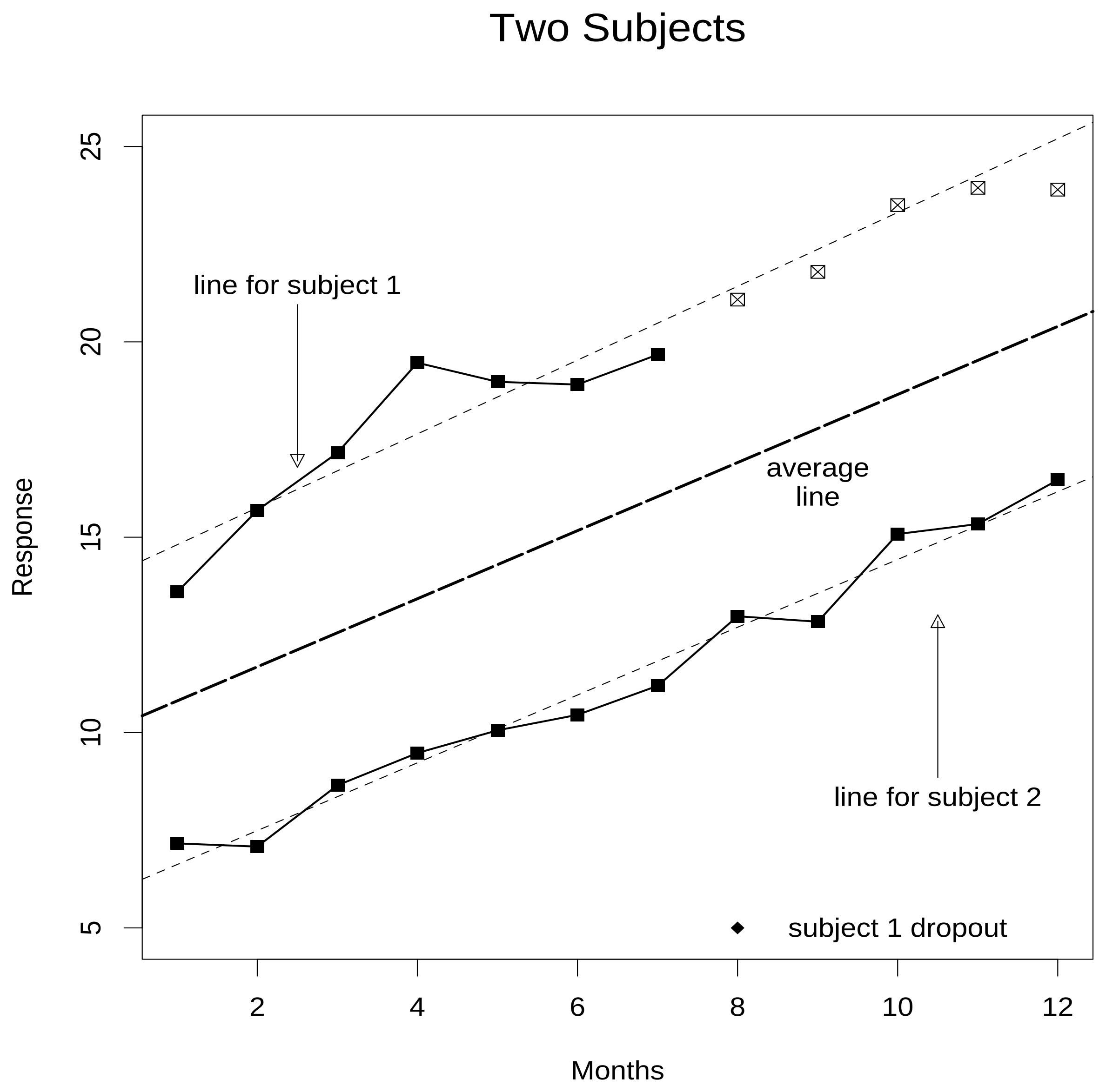
- Assuming: $\text{var}(e_{ij}) = \sigma^2$, and $\text{cov}(e_{ij}, e_{ik}) = 0$.

Linear Mixed Model

- **Regression model:**
  mean response as a function of covariates.
  "systematic variation"

- **Random effects:**
  variation from subject-to-subject in trajectory.
  "random between-subject variation"

- **Within-subject variation:**
  variation of individual observations over time
  "random within-subject variation"


Two Subjects
line for subject 1
average
line
line for subject 2
subject 1 dropout
Response
Months
25
20
15
10
5
2
4
6
8
10
12

LMM and components of variation

Within-Subject: Independence Model :

$$\boldsymbol{R}_i = \sigma^2 \boldsymbol{I}$$

or general diagonal matrix

Then, assuming normal errors we have that $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{i,n_i})$ are conditionally independent given $\boldsymbol{b}_i$.

- This model assumes that the within-subject errors do not have any serial correlation.

More on Covariance Models

Within-Subject: Serial Models

- Linear mixed models assume that each subject follows his/her own line. In some situations the dependence is more **local** meaning that observations close in time are more similar than those far apart in time.

- One model that we introduced is called the **autoregressive** model where:

$$\mathsf{cov}(e_{ij}, e_{ik}) = \sigma^2 \rho^{|j-k|}$$

More on Covariance Models

$\boxed{\text{Autoregressive Correlation}}$ **Assume** $t_{ij} = j, n_i \equiv 4$:

$$\text{corr}(\boldsymbol{e}_i) = \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

More on Covariance Models

Autoregressive Correlation **Assume $t_{ij} = j$:**

$$\text{corr}(\boldsymbol{e}_i) = \begin{bmatrix} 1 & \rho & \rho^2 & \ldots & \rho^{(n-1)} \\ \rho & 1 & \rho & \ldots & \rho^{(n-2)} \\ \rho^2 & \rho & 1 & \ldots & \rho^{(n-3)} \\ \vdots & & & \ddots & \vdots \\ \rho^{(n-1)} & \rho^{(n-2)} & \rho^{(n-3)} & \ldots & 1 \end{bmatrix}$$

More on Covariance Models

Autoregressive Correlation **Assume t_{ij} unique**:

$$\text{corr}(\boldsymbol{e}_i) = \begin{bmatrix} 1 & \rho^{|t_{i1}-t_{i2}|} & \rho^{|t_{i1}-t_{i3}|} & \ldots & \rho^{|t_{i1}-t_{in}|} \\ \rho^{|t_{i2}-t_{i1}|} & 1 & \rho^{|t_{i2}-t_{i3}|} & \ldots & \rho^{|t_{i2}-t_{in}|} \\ \rho^{|t_{i3}-t_{i1}|} & \rho^{|t_{i3}-t_{i2}|} & 1 & \ldots & \rho^{|t_{i3}-t_{in}|} \\ \vdots & & \ddots & & \vdots \\ \rho^{|t_{in}-t_{i1}|} & \rho^{|t_{in}-t_{i2}|} & \rho^{|t_{in}-t_{i3}|} & \ldots & 1 \end{bmatrix}$$

Linear Mixed Model

$\boxed{\text{Mixed Models}}$ Laird & Ware (1982)

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

$\boldsymbol{Y}_i$: $(n_i \times 1)$ response vector

$\boldsymbol{X}_i$: $(n_i \times p)$ design matrix for fixed effects

$\boldsymbol{\beta}$: $(p \times 1)$ regression coefficient for fixed effects

Covariance Models

Mixed Models Laird & Ware (1982)

$$\begin{array}{lcll} \boldsymbol{Z}_i & : & (n_i \times q) & \text{design matrix for random effects} \\ \boldsymbol{b}_i & : & (q \times 1) & \text{vector of random effects} \\ \boldsymbol{e}_i & : & (n_i \times 1) & \text{vector of errors} \end{array}$$

For the random components of the model we typically assume:

$$\boldsymbol{b}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{D})$$

$$\boldsymbol{e}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{R}_i)$$

Laird & Ware

Chair, Dept. Biostatistics HSPH 1990-1999

Associate Dean HSPH

LMM and components of variation

This yields a covariance structure:

$$\text{cov}(\boldsymbol{Y}_i \mid \boldsymbol{X}_i) \quad = \quad \underbrace{\boldsymbol{Z}_i \boldsymbol{D} \boldsymbol{Z}_i^T} \quad + \quad \underbrace{\boldsymbol{R}_i}$$

between-cluster var + **within-cluster var**

- We assume that observations on different subjects are independent.

More on Covariance Models

Mixed + Serial

- Diggle (1988) proposed the following model

$$Y_{ij} \quad = \quad \boldsymbol{X}_{ij}\boldsymbol{\beta} + b_{i,0} + W_i(t_{ij}) + \epsilon_{ij}$$

Covariance Models

Mixed + Serial

The most general type of covariance model will combine some **random effects** with some additional aspects that characterize within-subject **serial correlation**.

One such model contains three sources of random variation:

random intercept	$b_{i,0}$
serial process	$W_i(t_{ij})$
measurement error	ϵ_{ij}

We assume:

$$
\begin{aligned}
\mathsf{var}(b_{i,0}) &= \nu^2 \\
\mathsf{cov}[W(s), W(t)] &= \sigma^2 \rho^{|s-t|} \\
\mathsf{var}(\epsilon_{ij}) &= \tau^2
\end{aligned}
$$

Then:

$$
\textbf{Total Variance} = \nu^2 + \sigma^2 + \tau^2
$$

$$
\textbf{Covariance}(Y_{ij}, Y_{ik}) = \nu^2 + \sigma^2 \rho^{|t_{ij} - t_{ik}|}
$$

Covariance Models

$\boxed{\text{Mixed + Serial}}$ **Q**: How to biologically interpret these three sources of variation?

- **random intercept**: This represents a "trait" of the subject.
  - ▷ FEV1 – child "size" not captured by age and height.
  - ▷ CD4 – subject's "normal" steady-state level.
- **serial variation**: This represents a "state" for the subject.
  - ▷ FEV1 – child current health status (infected with PseudoA)
  - ▷ CD4 – subject's current immune status (diet? treatment?)
- **measurement error**: This represents the instrumentation or process used to generate the final quantitative measurement.
  - ▷ FEV1 – result of only one "trial" with expiration.
  - ▷ CD4 – blood sample, lab processing.

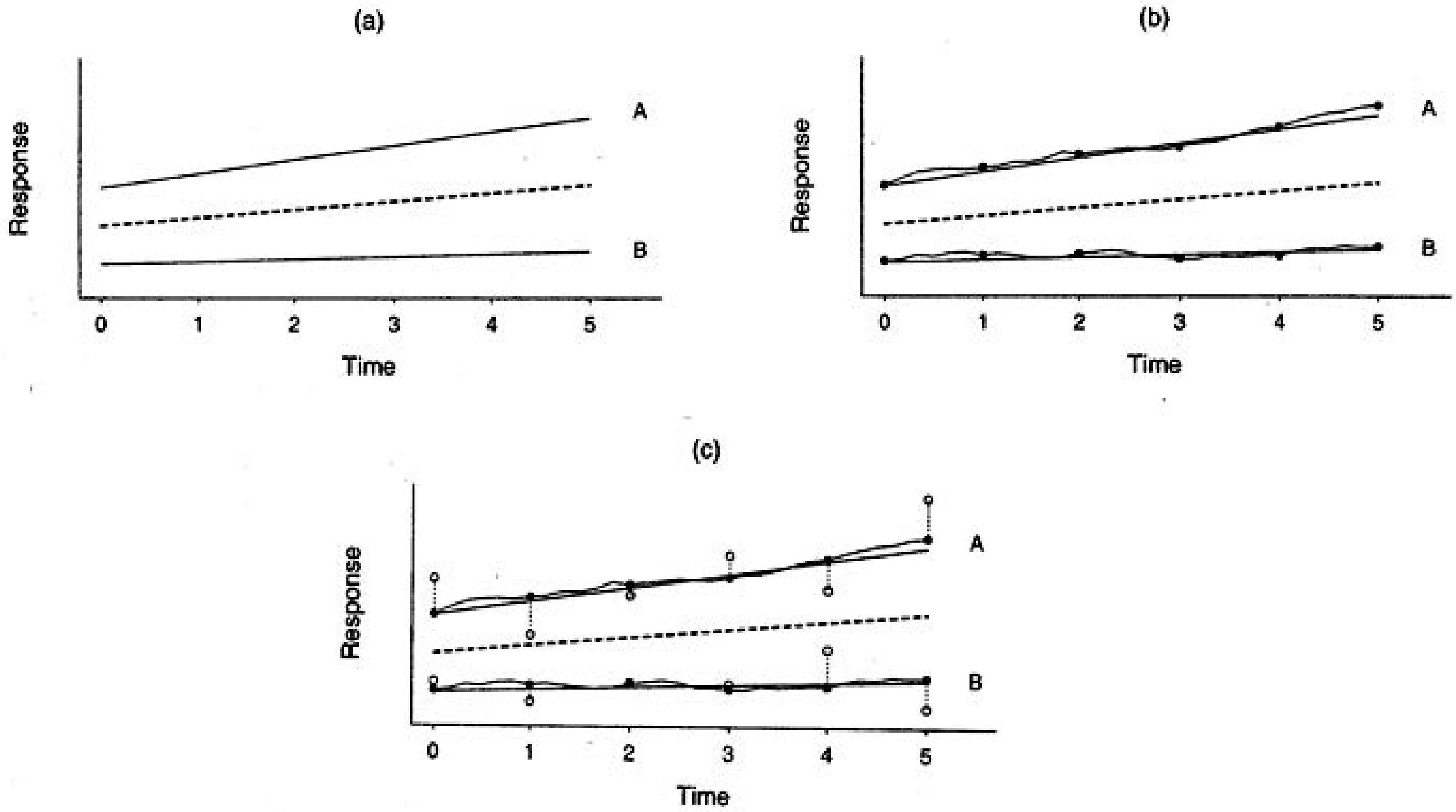


Fig. 2.4 Graphical representation of the cumulative impact of three sources of variability in longitudinal data: (a) between-individual heterogeneity, (b) within-individual biological variation (where • denotes repeated measure free of measurement error), and (c) measurement error (where ∘ denotes observed repeated measure with measurement error).

LMM: Linear Mixed Models and FEV1 Decline

- We can use linear mixed models to assess the evidence for differences in the rate of decline for subgroups defined by covariates.

- **S+** / **R** has a function `lme()`.

- **SAS** has the `MIXED` procedure.

SAS Program:

```
options linesize=80 pagesize=60;

data cfkids;
  infile 'NewCFkids-SAS.data';
  input id fev1 age female pseudoA f508 panc age0 ageL;
run;

data cfkids; set cfkids;
  f508_1 = 0;
  if f508=1 then f508_1 = 1;
  f508_2 = 0;
  if f508=2 then f508_2 = 1;
run;

proc mixed data=cfkids method=reml;
        class id;
        model fev1 = age0 ageL female f508_1 f508_2 female*ageL
                     f508_1*ageL f508_2*ageL / s;
        repeated / type=cs  subject=id;
run;
```

```
proc mixed data=cfkids method=reml;
        class id;
        model fev1 = age0 ageL female f508_1 f508_2 female*ageL
                     f508_1*ageL f508_2*ageL / s;
        random intercept / type=un subject=id g;
run;

proc mixed data=cfkids method=reml;
        class id;
        model fev1 = age0 ageL female f508_1 f508_2 female*ageL
                     f508_1*ageL f508_2*ageL / s;
        random intercept ageL / type=un subject=id g;
run;

proc mixed data=cfkids method=reml;
        class id;
        model fev1 = age0 ageL female f508_1 f508_2 female*ageL
                     f508_1*ageL f508_2*ageL / s;
        random intercept / type=un subject=id g;
        repeated / type=sp(pow)(ageL) subject=id;
run;

proc mixed data=cfkids method=reml;
        class id;
        model fev1 = age0 ageL female f508_1 f508_2 female*ageL
                     f508_1*ageL f508_2*ageL / s;
        random intercept / type=un subject=id g;
        repeated / type=sp(pow)(ageL) subject=id local;
```

```
run;

proc mixed data=cfkids method=reml;
        class id;
        model fev1 = age0 ageL female f508_1 f508_2 female*ageL
                     f508_1*ageL f508_2*ageL / s;
        random intercept ageL / type=un subject=id g;
        repeated / type=sp(pow)(ageL) subject=id;
run;

proc mixed data=cfkids method=reml;
        class id;
        model fev1 = age0 ageL female f508_1 f508_2 female*ageL
                     f508_1*ageL f508_2*ageL / s;
        random intercept ageL / type=un subject=id g;
        repeated / type=sp(pow)(ageL) subject=id local;
run;
```

Comments on Syntax and Model

Model: $Y_{ij} = \mu_{ij} + \overbrace{\boxed{b_{i,0}}}^{\texttt{(a)}} + \overbrace{\boxed{e_{ij}}}^{\texttt{(b)}}$

- SYNTAX: `random intercept / type=un subject=id g;`

- DESCRIPTION: The `random` statement is used to declare random effects. After the forward slash an ID variable must be specified using `subject =` *your-id-variable-name*. The option `type=un` is not necessary here (intercept only). The option g simply asks that the output display the random effects covariance matrix (we've called this $\mathbf{D}$) be written out as a matrix.

- NOTE: The default is to include the errors `(b)` above as independent errors with a constant variance.

Comments on Syntax and Model

Model: $Y_{ij} = \mu_{ij} + \overbrace{\boxed{b_{i,0} + b_{i,1}\texttt{ageL}}}^{\texttt{(a)}} + \overbrace{\boxed{e_{ij}}}^{\texttt{(b)}}$

- SYNTAX: `random intercept ageL / type=un subject=id g;`
- DESCRIPTION: The `random` statement is used to declare random effects. The option `type=un` asks that the variances and the covariance of random effects be an arbitrary (unstructured) matrix we've called this $\mathbf{D}$. One could specify other options such as asking for independent random effects, but for linear mixed models this isn't usually of interest.
- NOTE: The default is to include the errors `(b)` above as independent errors with a constant variance.

Comments on Syntax and Model

$$\boxed{\textbf{Model}:}\; Y_{ij} = \mu_{ij} + \overbrace{\boxed{b_{i,0}}}^{\texttt{(a)}} + \overbrace{\boxed{W_i(t_{ij})}}^{\texttt{(b)}}$$

- SYNTAX for (a): `random intercept / type=un subject=id g;`

- SYNTAX for (b): `repeated / type=sp(pow)(ageL) subject=id;`

- DESCRIPTION: The use of the `repeated` command allows one to relax the assumption that within-subject errors are independent. To include a component of serial correlation (autocorrelated errors) we can use commands like `type = ar(1)` which assume that observations j and k for a subject have within-subject errors with covariance $\sigma^2\rho^{|j-k|}$. When observations are not equally spaced in time the command `type=sp(pow)(ageL)` allows a covariance $\sigma^2\rho^{d_{jk}}$, where the distance is computed as $|t_{ij} - t_{ik}|$, and the argument ageL is specifying the time variable to compute distance.

Comments on Syntax and Model

Model: $Y_{ij} = \mu_{ij} + \overbrace{\boxed{b_{i,0}}}^{(a)} + \overbrace{\boxed{W_i(t_{ij})}}^{(b)} + \overbrace{\boxed{e_{ij}}}^{(c)}$

- SYNTAX for (a): `random intercept / type=un subject=id g;`
- SYNTAX for (b) and (c):
  `repeated / type=sp(pow)(ageL) subject=id local;`
- DESCRIPTION: This is similar to the previous model, but now the option `local` asks for the inclusion of the measurement errors, e_{ij}, which are assumed to be independent. Thus the within-subject errors for this model have both a serial component, and a pure noise component.

SAS Fit 1 **Random Intercepts + Slopes**

```
                    The MIXED Procedure

                  Class Level Information

      Class     Levels  Values

      ID           200  100073 100111 100185 100329
                        100352 100636 100736 100815
                          .      .      .     .
                          .      .      .     .
                          .      .      .     .

              REML Estimation Iteration History

     Iteration  Evaluations     Objective     Criterion

             0            1  11288.083105
             1            2  9625.5053208    0.00009459
             2            1  9625.0138752    0.00000165
             3            1  9625.0058252    0.00000000

                  Convergence criteria met.
```

G Matrix

Effect	ID	Row	COL1	COL2
INTERCEPT	100073	1	512.41416519	-7.62541488
AGEL	100073	2	-7.62541488	4.51229421

Covariance Parameter Estimates (REML)

Cov Parm	Subject	Estimate
UN(1,1)	ID	512.41416519
UN(2,1)	ID	-7.62541488
UN(2,2)	ID	4.51229421
Residual		118.02545375

Model Fitting Information for FEV1

Description	Value
Observations	1513.000
Res Log Likelihood	-6194.59
Akaike's Information Criterion	-6198.59
Schwarz's Bayesian Criterion	-6209.22
-2 Res Log Likelihood	12389.17
Null Model LRT Chi-Square	1663.077
Null Model LRT DF	3.0000
Null Model LRT P-Value	0.0000

Solution for Fixed Effects

Effect	Estimate	Std Error	DF	t	Pr > |t|
INTERCEPT	104.51880927	6.64425792	195	15.73	0.0001
AGE0	-1.91050551	0.33103925	1113	-5.77	0.0001
AGEL	-0.60278138	0.59026446	196	-1.02	0.3084
FEMALE	-1.30051066	3.37009028	1113	-0.39	0.6996
F508_1	-4.23810256	5.56362771	1113	-0.76	0.4464
F508_2	-6.65228283	5.59443908	1113	-1.19	0.2347
AGEL*FEMALE	-0.76242845	0.38119258	1113	-2.00	0.0457
AGEL*F508_1	-0.50010405	0.63571767	1113	-0.79	0.4316
AGEL*F508_2	-0.74589389	0.63445750	1113	-1.18	0.2400

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
AGE0	1	1113	33.31	0.0001
AGEL	1	196	1.04	0.3084
FEMALE	1	1113	0.15	0.6996
F508_1	1	1113	0.58	0.4464
F508_2	1	1113	1.41	0.2347
AGEL*FEMALE	1	1113	4.00	0.0457
AGEL*F508_1	1	1113	0.62	0.4316
AGEL*F508_2	1	1113	1.38	0.2400

SAS Fit 2 **Random Intercepts + AR errors**

```
                    The MIXED Procedure

             REML Estimation Iteration History

Iteration  Evaluations     Objective     Criterion

        0            1  11288.083105
        1            2  9842.5776314  152.37610437
        2            2  9666.7258252    0.01014002
        3            1  9612.7982174    0.00047006
        4            2  9610.8212742    0.00002116
        5            1  9610.7155276    0.00000009
        6            1  9610.7151109    0.00000000

                Convergence criteria met.
```

G Matrix

Effect	ID	Row	COL1
INTERCEPT	100073	1	483.46131336

Covariance Parameter Estimates (REML)

Cov Parm	Subject	Estimate
UN(1,1)	ID	483.46131336
SP(POW)	ID	0.33940770
Residual		172.66675040

Model Fitting Information for FEV1

Description	Value
Observations	1513.000
Res Log Likelihood	-6187.44
Akaike's Information Criterion	-6190.44
Schwarz's Bayesian Criterion	-6198.41
-2 Res Log Likelihood	12374.88
Null Model LRT Chi-Square	1677.368
Null Model LRT DF	2.0000
Null Model LRT P-Value	0.0000

Solution for Fixed Effects

Effect	Estimate	Std Error	DF	t	Pr > |t|
INTERCEPT	104.19222115	6.76123647	195	15.41	0.0001
AGE0	-1.85599115	0.33226611	1309	-5.59	0.0001
AGEL	-0.58899693	0.50099307	1309	-1.18	0.2399
FEMALE	-1.25755036	3.46074067	1309	-0.36	0.7164
F508_1	-4.70904830	5.71244977	1309	-0.82	0.4099
F508_2	-6.76205358	5.74144814	1309	-1.18	0.2391
AGEL*FEMALE	-0.85649279	0.32240863	1309	-2.66	0.0080
AGEL*F508_1	-0.42254215	0.54066747	1309	-0.78	0.4346
AGEL*F508_2	-0.70366210	0.53899040	1309	-1.31	0.1919

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
AGE0	1	1309	31.20	0.0001
AGEL	1	1309	1.38	0.2399
FEMALE	1	1309	0.13	0.7164
F508_1	1	1309	0.68	0.4099
F508_2	1	1309	1.39	0.2391
AGEL*FEMALE	1	1309	7.06	0.0080
AGEL*F508_1	1	1309	0.61	0.4346
AGEL*F508_2	1	1309	1.70	0.1919

SAS Fit 3 Random Intercepts + AR errors + meas. error

```
                    The MIXED Procedure

             REML Estimation Iteration History

     Iteration  Evaluations     Objective     Criterion

             0            1  11288.083105
             1            2  9777.1891608   83.13453204
             2            2  9632.0064884    8.05381107
             .
            19            1  9579.7150151    0.00000001

                 Convergence criteria met.
```

G Matrix

Effect	ID	Row	COL1
INTERCEPT	100073	1	390.13143397

Covariance Parameter Estimates (REML)

Cov Parm	Subject	Estimate
UN(1,1)	ID	390.13143397
Variance	ID	191.19638318
SP(POW)	ID	0.83528160
Residual		72.87412247

Model Fitting Information for FEV1

Description	Value
Observations	1513.000
Res Log Likelihood	-6171.94
Akaike's Information Criterion	-6175.94
Schwarz's Bayesian Criterion	-6186.57
-2 Res Log Likelihood	12343.88
Null Model LRT Chi-Square	1708.368
Null Model LRT DF	3.0000
Null Model LRT P-Value	0.0000

Solution for Fixed Effects

Effect	Estimate	Std Error	DF	t	Pr > |t|
INTERCEPT	104.77553777	6.82547939	195	15.35	0.0001
AGE0	-1.87512100	0.33131471	1309	-5.66	0.0001
AGEL	-0.71223608	0.58209781	1309	-1.22	0.2213
FEMALE	-1.20683124	3.52025659	1309	-0.34	0.7318
F508_1	-5.02332453	5.80876288	1309	-0.86	0.3873
F508_2	-7.10525357	5.83759113	1309	-1.22	0.2238
AGEL*FEMALE	-0.82492646	0.37579975	1309	-2.20	0.0283
AGEL*F508_1	-0.31356545	0.62802526	1309	-0.50	0.6177
AGEL*F508_2	-0.58034315	0.62637981	1309	-0.93	0.3544

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
AGE0	1	1309	32.03	0.0001
AGEL	1	1309	1.50	0.2213
FEMALE	1	1309	0.12	0.7318
F508_1	1	1309	0.75	0.3873
F508_2	1	1309	1.48	0.2238
AGEL*FEMALE	1	1309	4.82	0.0283
AGEL*F508_1	1	1309	0.25	0.6177
AGEL*F508_2	1	1309	0.86	0.3544

Likelihood Summaries

Model	q	$\log L$	AIC	BIC
int + e	2	-6249.0	-6251.0	-6256.4
int+slope + e	4	-6194.6	-6198.6	-6209.2
int + AR	3	-6187.4	-6190.4	-6198.4
int + AR + e	4	-6171.9	-6175.9	-6186.6
int+slope + AR	5	-6174.2	-6179.2	-6192.5
int+slope + AR + e	6	-6170.0	-6176.0	-6192.0

AIC $= \log L - q$

BIC $= \log L - q \cdot \log(\sum n_i)/2$

Likelihood Estimation for Linear Mixed Models

$\boxed{\text{Parameters:}}$

$\boldsymbol{\beta}$: regression parameter, fixed effects coefficient

$\boldsymbol{\alpha}$: variance components

$$\boldsymbol{\alpha} \Rightarrow \quad \boldsymbol{D}(\boldsymbol{\alpha}) \quad \text{and} \quad \boldsymbol{R}(\boldsymbol{\alpha})$$

where $\text{cov}(\boldsymbol{Y}_i) = \boldsymbol{Z}_i \boldsymbol{D} \boldsymbol{Z}_i^T + \boldsymbol{R}_i$

Normality:

$$
\begin{aligned}
E(\boldsymbol{Y}_i) &= \boldsymbol{X}_i\boldsymbol{\beta} \\
\mathsf{cov}(\boldsymbol{Y}_i) &= \boldsymbol{\Sigma}(\boldsymbol{\alpha})
\end{aligned}
$$

$$
\begin{aligned}
f(\boldsymbol{Y}_i;\boldsymbol{\beta},\boldsymbol{\alpha}) &= |\boldsymbol{\Sigma}|^{-1/2}(2\pi)^{-n_i/2} \times \\
& \quad \exp\left[-\frac{1}{2}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})\right]
\end{aligned}
$$

Maximum Likelihood:

Find the values for the regression coefficients, $\boldsymbol{\beta}$, and the variance components that maximizes the likelihood – e.g. put the highest available probability on the observed data.

R.A. Fisher

ML versus REML

- There is a variant of ML estimation known as **REML**.
  - ▷ "Residual" ML
  - ▷ "Restricted" ML
- REML is used to provide slightly less biased estimates of variance components.
- However, be careful using REML when you change the covariates in your model since one can not use changes in REML log likelihoods to test for fixed effects.
- Useful for a single fitted model, or to compare covariance models with a fixed regression model.

Inference in the Linear Mixed Model

Likelihood Ratio Tests – Variance Components

We may want to test whether we have random intercepts and slopes, or just random intercepts.

$$H_0 : \boldsymbol{D} = \begin{bmatrix} D_{11} & 0 \\ 0 & 0 \end{bmatrix} \quad \text{versus} \quad H_1 : \boldsymbol{D} = \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}$$

$\boxed{\mathbf{Q}:}$ What is the distribution of the likelihood ratio statistic

$$\texttt{LRstat} = 2 \cdot \log \frac{\mathcal{L}_{ML}(\widehat{\boldsymbol{\theta}}_{ML,\text{model 1}})}{\mathcal{L}_{ML}(\widehat{\boldsymbol{\theta}}_{ML,\text{model 0}})}$$

LR Testing for Variance Components

$\boxed{\circ}$ $D_{22} = 0$ is on the boundary of the parameter space!!!

$\Rightarrow$ This violates the standard assumption that we use to justify the $\chi^2(p_1 - p_0)$ distribution of the LR statistic.

$\boxed{\star}$ We appeal to results in Stram and Lee (1994) that build upon results in Self & Liang (1987) showing that `LR stat` is a <u>mixture of χ^2</u>.

Note: For a fixed mean strucure we can use the LR based on either ML or REML. (Why?)

See: Verbeke and Molenberghs (1997) pages 108-111.

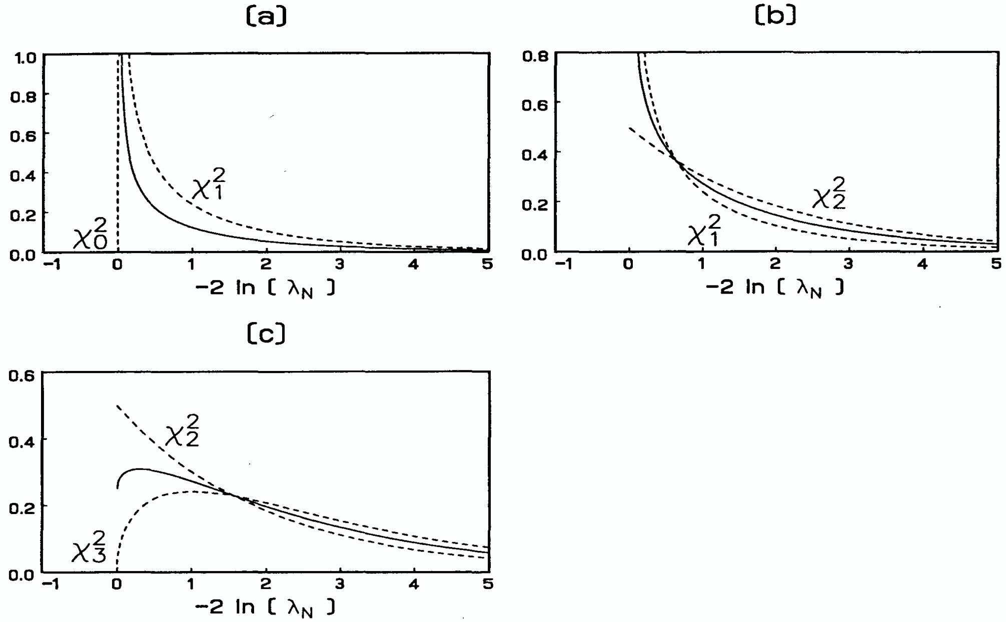


FIGURE 3.4. *Graphical representation of the asymptotic null distribution of the likelihood ratio statistic for testing the significance of random effects in a linear mixed model, for three different types of hypotheses. For each case, the distribution (solid line) is a mixture of two chi-squared distributions (dashed lines), with both weights equal to 0.5:*
(a) Case 1: no random effects versus one random effect.
(b) Case 2: one random effect versus two random effects.
(c) Case 3: two random effects versus three random effects.

Case 1: No Random Effects versus One Random Effect

For testing $H_0 : D = 0$ versus $H_1 : D = d_{11}$, where d_{11} is a non-negative scalar, we have that the asymptotic null distribution of $-2\ln\lambda_N$ is a mixture of χ^2_1 and χ^2_0 with equal weights 0.5. The χ^2_0 distribution is the distribution which gives probability mass one to the value 0. The mixture is shown in panel (a) of Figure 3.4. Note that if the classical null distribution would be used, all p-values would be overestimated. Therefore, the null hypothesis would be accepted too often, resulting in incorrectly simplifying the covariance structure of the model, which may seriously invalidate inferences, as shown by Altham (1984).

Case 2: One versus Two Random Effects

In the case one wishes to test

$$H_0 : D = \begin{pmatrix} d_{11} & 0 \\ 0 & 0 \end{pmatrix},$$

for some strictly positive d_{11}, versus H_1 that D is a (2×2) positive semi-definite matrix, we have that the asymptotic null distribution of $-2 \ln \lambda_N$ is a mixture with equal weights 0.5 for χ^2_2 and χ^2_1, shown in Figure 3.4(b). Similar to case 1, we have that ignoring the boundary problems may result in too parsimonious covariance structures.

Case 3: q versus $q + 1$ Random Effects

For testing the hypothesis

$$H_0 : D = \begin{pmatrix} D_{11} & \mathbf{0} \\ \mathbf{0}' & 0 \end{pmatrix}, \tag{3.26}$$

in which D_{11} is a $(q \times q)$ positive definite matrix, versus H_1 that D is a general $((q + 1) \times (q + 1))$ positive semi-definite matrix, the large-sample behavior of the null distribution of $-2 \ln \lambda_N$ is a mixture of χ^2_{q+1} and χ^2_q, again with equal weights 0.5. A graphical representation for the case of testing two random effects ($q = 2$) versus three random effects is given in the third panel of Figure 3.4. Again, we have that the correction due to the boundary problems reduces the p-values in order to protect against the use of oversimplified covariance structures.

Inference using ML/REML

- Inference on Regression Coefficients
  - $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$
  - $\boxed{H_0 : \;\; \boldsymbol{\beta}_2 = \mathbf{0}}$

- **Likelihood Ratio Tests**
  - Testing $\boldsymbol{\beta}_2 = 0$ can be done using standard likelihood ratio tests if using ML.
  - Testing $\boldsymbol{\beta}_2 = 0$ can **NOT** be done using likelihood ratio tests if using **REML**.

Inference using ML/REML

- **Multivariate Wald Tests**
  - Testing $\boldsymbol{\beta}_2 = 0$ can be done using standard multivariate Wald test.
  - Testing can be based on **empirical standard errors** since ML/REML use WLS. (see SAS MIXED and `empirical` option!)

S+ LMM Program:

```
#
# cfkids-CDA-NewLMM.q
#
# ------------------------------------------------------------
#
# PURPOSE:  Use linear mixed models to characterize longitudinal
#            change by gender and genotype.
#
# AUTHOR:  P. Heagerty
#
# DATE:  00/07/10  Revised 14Feb2002
#
# ------------------------------------------------------------
#
#
#####
##### Read data
#####
#
source("cfkids-read.q")
#
#
#####
##### Trellis plots of individuals and groups
#####
```

```
#
# Create Grouped Data Set
#
ntotal <-  cumsum( unlist( lapply( split( cfkids$id, cfkids$id) , length ) ) )
cf.subset <- groupedData(
    fev1 ~ age | id, outer = ~ factor(f508)*female,
    data = cfkids[ 1:ntotal[(8*4*1)], ] )
#
cfkids <- groupedData(
    fev1 ~ age | id, outer = ~ factor(f508)*female,
    data = cfkids )
#
# trellis plot, by id, first 1 pages, 8x4
#
postscript( file="cfkids-trellis.1.ps", horiz=F )
plot( cf.subset, layout = c(4,8) )
graphics.off()
postscript( file="cfkids-trellis.2.ps", horiz=T )
par( pch="." )
plot( cfkids, outer = ~ factor(f508)*factor(female), layout=c(3,2),
     aspect=1 )
graphics.off()
#
#####
##### Linear Mixed Models
#####
#
options( contrasts=c("contr.treatment","contr.helmert") )
```

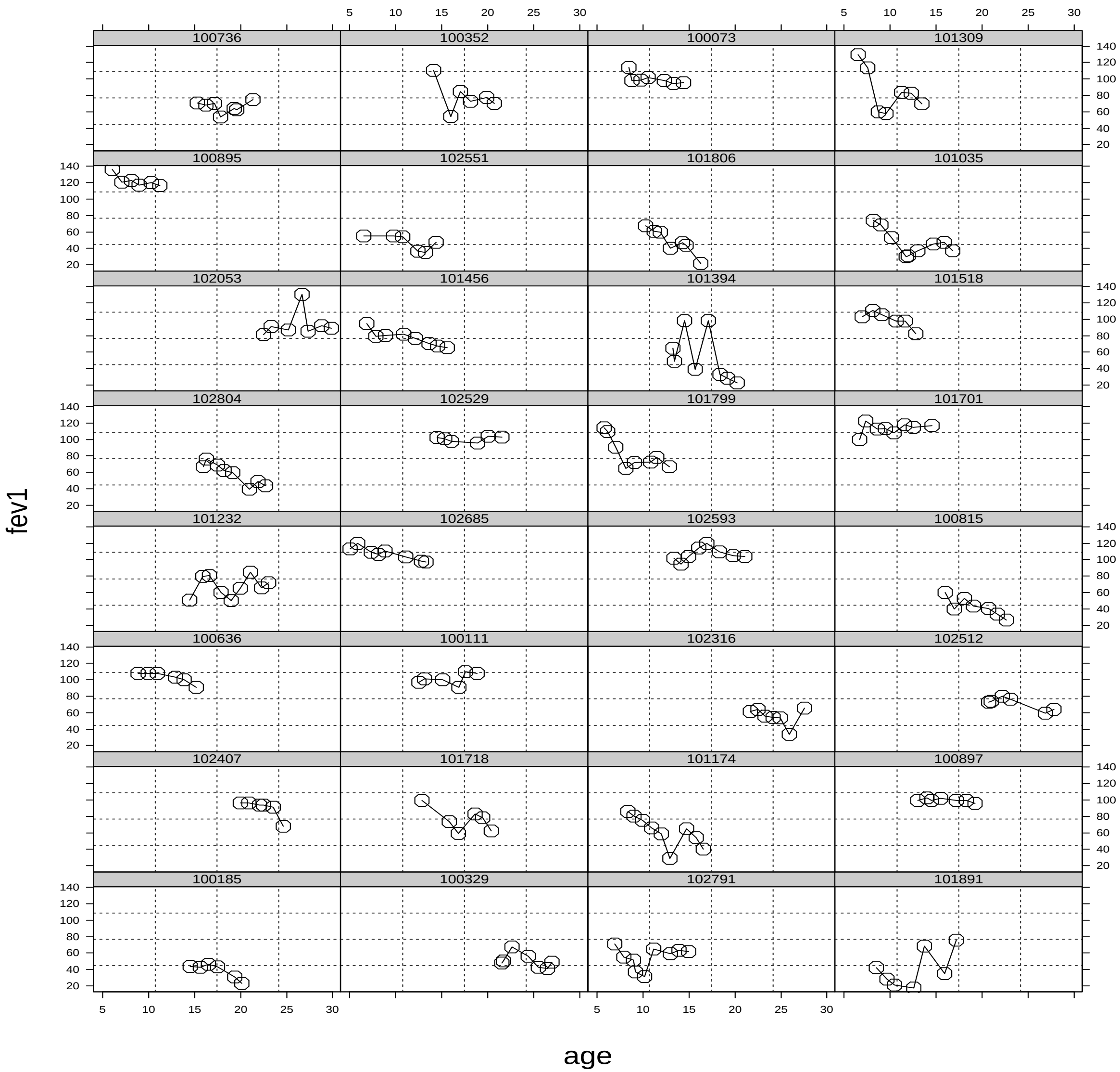
```
#

### Intercept only

fit0 <- lme( fev1 ~ age0 + ageL + female*ageL + factor(f508)*ageL,
            method = "ML",
            random = reStruct( ~ 1 | id, pdClass="pdSymm", REML=F),
            data = cfkids )
summary( fit0 )

### Intercept plus Slope

fit1 <- lme( fev1 ~ age0 + ageL + female*ageL + factor(f508)*ageL,
            method = "ML",
            random = reStruct( ~ 1 + ageL | id, pdClass="pdSymm", REML=F),
            data = cfkids )
summary( fit1 )

### EDA for serial correlation

postscript( file="cfkids-Variogram.ps", horiz=T )
plot( Variogram(  fit0, form = ~ age | id , resType="response" ) )
graphics.off()


### Intercept plus AR(1)

fit2a <- lme( fev1 ~ age0 + ageL + female*ageL + factor(f508)*ageL,
```

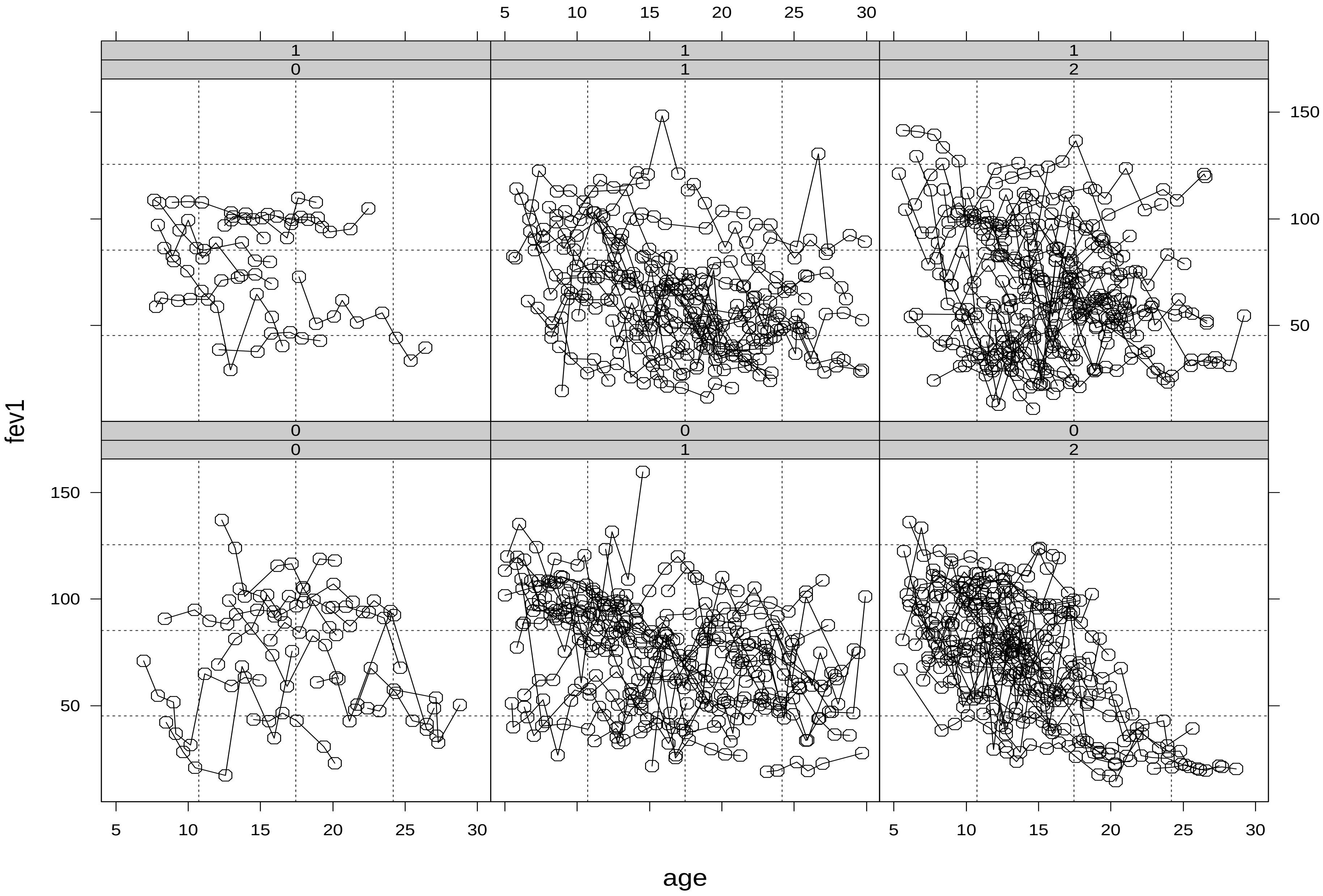
```
            method = "ML",
            random = reStruct( ~ 1 | id, pdClass="pdSymm", REML=F),
            correlation = corAR1( form = ~ 1 | id ),
            data = cfkids )
summary( fit2a )

### another way

fit2b <- lme( fev1 ~ age0 + ageL + female*ageL + factor(f508)*ageL,
            method = "ML",
            random = reStruct( ~ 1 | id, pdClass="pdSymm", REML=F),
            correlation = corExp( form = ~ ageL | id, nugget=F),
            data = cfkids )
summary( fit2b )

### another way

fit2c <- lme( fev1 ~ age0 + ageL + female*ageL + factor(f508)*ageL,
            method = "ML",
            random = reStruct( ~ 1 | id, pdClass="pdSymm", REML=F),
            correlation = corCAR1( form = ~ ageL | id ),
            data = cfkids )
summary( fit2c )

fit2 <- fit2b

### Intercept plus AR(1) plus measurement error
```

```
fit3 <- lme( fev1 ~ age0 + ageL + female*ageL + factor(f508)*ageL,
            method = "ML",
            random = reStruct( ~ 1 | id, pdClass="pdSymm", REML=F),
            correlation = corExp( form = ~ ageL | id, nugget=T),
            data = cfkids )
summary( fit3 )

#
##### compare these models
#
anova( fit0, fit1, fit2, fit3 )

#
#####
##### Residual Analysis -- using fit3
#####
#
pop.res <- resid( fit3, level=0 )
cluster.res <- resid( fit3, level=1 )
print( var( pop.res ) )
print( var( cluster.res ) )
#
postscript( file="cfkids-NewResiduals.ps", horiz=F )
par( mfrow=c(2,1) )
plot( cfkids$age0, pop.res, pch="." )
lines( smooth.spline( cfkids$age0, pop.res, df=5 ) )
title("Residuals (pop) vs Age0")
abline( h=0, lty=2 )
```

```
plot( cfkids$ageL, pop.res, pch="." )
lines( smooth.spline( cfkids$ageL, pop.res, df=5 ) )
title("Residuals (pop) vs AgeL")
abline( h=0, lty=2 )
graphics.off()
#
postscript( file="cfkids-NewResiduals2.ps", horiz=F )
par( mfrow=c(2,1) )
plot( cfkids$age0, cluster.res, pch="." )
lines( smooth.spline( cfkids$age0, cluster.res, df=5 ) )
abline( h=0, lty=2 )
title("Residuals (cluster) vs Age0")
b0 <- unlist( fit2$coefficients$random )
age0 <- unlist( lapply( split( cfkids$age0, cfkids$id ), min ) )
plot( age0, b0 )
lines( smooth.spline( age0, b0, df=5 ) )
abline( h=0, lty=2 )
title("EB b0 versus Age0")
graphics.off()

#
##### Do we need a quadratic age0???
#
fit4 <- lme( fev1 ~ age0 + age0^2 + ageL + female*ageL + factor(f508)*ageL,
            method = "ML",
            random = reStruct( ~ 1 | id, pdClass="pdSymm", REML=F),
            correlation = corExp( form = ~ ageL | id, nugget=T),
            data = cfkids )
```

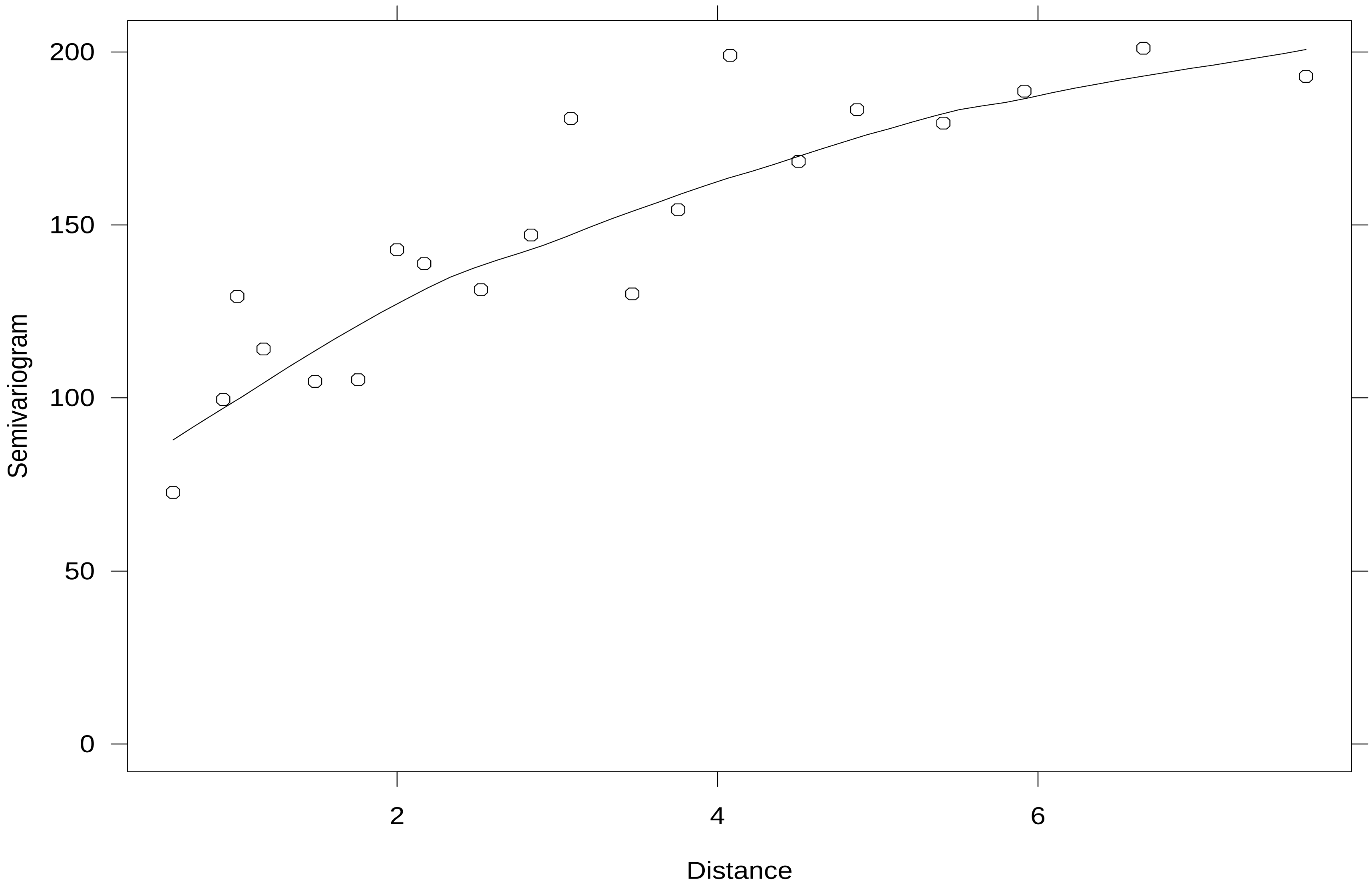
```
summary( fit4 )
#
anova( fit3, fit4 )
#
# end-of-file...
```


100736
100352
100073
101309
100895
102551
101806
101035
102053
101456
101394
101518
102804
102529
101799
101701
101232
102685
102593
100815
100636
100111
102316
102512
102407
101718
101174
100897
100185
100329
102791
101891
fev1
age


1
0
1
1
1
2
0
0
0
1
0
2
fev1
age
5
10
15
20
25
30
50
100
150

$\boxed{\text{Fit 0}}$ Random Intercepts

```
Linear mixed-effects model fit by maximum likelihood

 Data: cfkids
       AIC      BIC    logLik
  12532.01 12590.55 -6255.005

Random effects:
 Formula:  ~ 1 | id
        (Intercept) Residual
StdDev:    22.30435 12.17881

Fixed effects: fev1 ~ age0 + ageL + female * ageL + factor(f508) * ageL
                      Value Std.Error   DF   t-value p-value
      (Intercept)  103.8074  6.644683 1309  15.62262  <.0001
             age0   -1.8553  0.331087  195  -5.60372  <.0001
             ageL   -0.5881  0.393625 1309  -1.49417  0.1354
           female   -1.1620  3.367634  195  -0.34505  0.7304
    factor(f508)1   -4.2821  5.561343  195  -0.76998  0.4422
    factor(f508)2   -6.7427  5.592591  195  -1.20564  0.2294
      female:ageL   -0.8257  0.249786 1309  -3.30577  0.0010
ageLfactor(f508)1   -0.4873  0.423076 1309  -1.15191  0.2496
ageLfactor(f508)2   -0.6568  0.421964 1309  -1.55655  0.1198

Number of Observations: 1513
Number of Groups: 200
```



200
150
100
50
0
Semivariogram
2
4
6
Distance

Fit 1 Random Intercepts and Slopes

```
Linear mixed-effects model fit by maximum likelihood

 Data: cfkids
       AIC      BIC    logLik
  12430.34 12499.52 -6202.168

Random effects:                                              StdDev   Corr
 Formula:  ~ 1 + ageL | id                        (Intercept) 22.330642 (Inter
 Structure: General positive-definite                    ageL  2.087337 -0.156
                                                     Residual 10.865196

Fixed effects: fev1 ~ age0 + ageL + female * ageL + factor(f508) * ageL
                      Value Std.Error   DF   t-value p-value
      (Intercept)  104.5162  6.581633 1309  15.87997  <.0001
             age0   -1.9104  0.327870  195  -5.82672  <.0001
             ageL   -0.6019  0.585611 1309  -1.02777  0.3042
           female   -1.2969  3.338418  195  -0.38847  0.6981
    factor(f508)1   -4.2382  5.511575  195  -0.76896  0.4428
    factor(f508)2   -6.6550  5.542114  195  -1.20081  0.2313
      female:ageL   -0.7636  0.378110 1309  -2.01965  0.0436
ageLfactor(f508)1   -0.5003  0.630694 1309  -0.79323  0.4278
ageLfactor(f508)2   -0.7451  0.629441 1309  -1.18370  0.2367

Number of Observations: 1513
Number of Groups: 200
```

Fit 2a Random Intercepts + AR(1) errors

```
Linear mixed-effects model fit by maximum likelihood

 Data: cfkids
       AIC      BIC    logLik
  12425.01 12488.87 -6200.503

Random effects:                           Correlation Structure: AR(1)
 Formula:  ~ 1 | id                        Formula:  ~ 1 | id
        (Intercept) Residual               Parameter estimate(s):
StdDev:    21.57121 13.25308                      Phi
                                            0.3760331

Fixed effects: fev1 ~ age0 + ageL + female * ageL + factor(f508) * ageL
                      Value Std.Error   DF   t-value p-value
      (Intercept)  104.2928  6.704807 1309  15.55493  <.0001
             age0   -1.8599  0.328983  195  -5.65355  <.0001
             ageL   -0.6754  0.507081 1309  -1.33188  0.1831
           female   -1.2954  3.428658  195  -0.37781  0.7060
    factor(f508)1   -4.6714  5.664678  195  -0.82465  0.4106
    factor(f508)2   -6.8660  5.695078  195  -1.20560  0.2294
      female:ageL   -0.8151  0.325773 1309  -2.50206  0.0125
ageLfactor(f508)1   -0.3865  0.546674 1309  -0.70696  0.4797
ageLfactor(f508)2   -0.6224  0.545237 1309  -1.14152  0.2539

Number of Observations: 1513         Number of Groups: 200
```

Fit 2b Random Intercepts + corExp errors

```
Linear mixed-effects model fit by maximum likelihood

 Data: cfkids
       AIC      BIC    logLik
  12412.76 12476.62 -6194.378

Random effects:                Correlation Structure: Exponential spatial corr
 Formula:  ~ 1 | id             Formula:  ~ ageL | id
        (Intercept) Residual    Parameter estimate(s):
StdDev:    21.70338 13.08926        range
                                0.9136573

Fixed effects: fev1 ~ age0 + ageL + female * ageL + factor(f508) * ageL
                      Value Std.Error   DF   t-value p-value
      (Intercept)  104.1874  6.700086 1309  15.55015  <.0001
             age0   -1.8560  0.329094  195  -5.63975  <.0001
             ageL   -0.5882  0.499669 1309  -1.17720  0.2393
           female   -1.2584  3.430314  195  -0.36684  0.7141
    factor(f508)1   -4.7019  5.662454  195  -0.83036  0.4073
    factor(f508)2   -6.7593  5.691155  195  -1.18769  0.2364
      female:ageL   -0.8559  0.321526 1309  -2.66206  0.0079
ageLfactor(f508)1   -0.4242  0.539225 1309  -0.78677  0.4316
ageLfactor(f508)2   -0.7042  0.537553 1309  -1.30999  0.1904

Number of Observations: 1513         Number of Groups: 200
```

Fit 2c **Random Intercepts + CAR(1) errors**

```
Linear mixed-effects model fit by maximum likelihood

 Data: cfkids
       AIC      BIC    logLik
  12412.76 12476.62 -6194.378

Random effects:                     Correlation Structure: Continuous AR(1)
 Formula:  ~ 1 | id                  Formula:  ~ ageL | id
        (Intercept) Residual         Parameter estimate(s):
StdDev:    21.69983 13.09153               Phi
                                     0.3350188

Fixed effects: fev1 ~ age0 + ageL + female * ageL + factor(f508) * ageL
                      Value Std.Error   DF   t-value p-value
      (Intercept)  104.1877  6.699496 1309  15.55158  <.0001
             age0   -1.8560  0.329056  195  -5.64039  <.0001
             ageL   -0.5883  0.499812 1309  -1.17704  0.2394
           female   -1.2584  3.430073  195  -0.36686  0.7141
    factor(f508)1   -4.7025  5.662058  195  -0.83053  0.4073
    factor(f508)2   -6.7597  5.690753  195  -1.18784  0.2363
      female:ageL   -0.8559  0.321620 1309  -2.66136  0.0079
ageLfactor(f508)1   -0.4241  0.539381 1309  -0.78625  0.4319
ageLfactor(f508)2   -0.7041  0.537708 1309  -1.30942  0.1906

Number of Observations: 1513         Number of Groups: 200
```

Fit 3 **Random Intercepts + corExp + meas. error**

```
Linear mixed-effects model fit by maximum likelihood
 Data: cfkids
       AIC      BIC    logLik
  12384.92 12454.11 -6179.461

Random effects:                Correlation Structure: Exponential spatial corr
 Formula:  ~ 1 | id             Formula:  ~ ageL | id
        (Intercept) Residual    Parameter estimate(s):
StdDev:    19.69916 15.86913       range    nugget
                                5.116653 0.2878059

Fixed effects: fev1 ~ age0 + ageL + female * ageL + factor(f508) * ageL
                      Value Std.Error   DF   t-value p-value
      (Intercept)  104.7466  6.760173 1309  15.49466  <.0001
             age0   -1.8739  0.328085  195  -5.71153  <.0001
             ageL   -0.7095  0.577371 1309  -1.22886  0.2193
           female   -1.2089  3.486725  195  -0.34671  0.7292
    factor(f508)1   -5.0121  5.753669  195  -0.87112  0.3848
    factor(f508)2   -7.0993  5.782222  195  -1.22778  0.2210
      female:ageL   -0.8259  0.372683 1309  -2.21612  0.0269
ageLfactor(f508)1   -0.3159  0.622931 1309  -0.50710  0.6122
ageLfactor(f508)2   -0.5817  0.621290 1309  -0.93627  0.3493

Number of Observations: 1513
Number of Groups: 200
```

ANOVA

```
     Model df      AIC      BIC    logLik   Test  L.Ratio p-value
fit0     1 11 12532.01 12590.55 -6255.005
fit1     2 13 12430.34 12499.52 -6202.168 1 vs 2 105.6739  <.0001
fit2     3 12 12412.76 12476.62 -6194.378 2 vs 3  15.5802   1e-04
fit3     4 13 12384.92 12454.11 -6179.461 3 vs 4  29.8354  <.0001
```

LMM Summary

- Observe $\boldsymbol{Y}_i$, $i = 1, 2, \ldots, m$ independent clusters.

- **Model**: (Laird & Ware, 1982)

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

- $\boldsymbol{\beta}$ is the coefficient that is common to all clusters (fixed across clusters).

- $\boldsymbol{b}_i$ is the deviation of the coefficient that varies from cluster to cluster (random across clusters).

- $(\beta_j + b_{j,i})$ is the coefficient of $X_{i,j}$ for cluster i.

$$
\begin{aligned}
\boldsymbol{b}_i &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{D}) \text{ between-cluster} \\
\boldsymbol{e}_i &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{R}_i) \text{ within-cluster}
\end{aligned}
$$

- **Estimation/Inference**: WLS, ML
  - ▷ Covariance model choice leads to WLS – but estimated regression coefficient is **unbiased** for any choice of weight (covariance).
  - ▷ Covariance model choice determines the **standard error** estimates for the regression coefficients – **correct** covariance model is needed for **correct** standard errors.

Likelihood Estimation for Linear Mixed Models

Parameters:

$\boldsymbol{\beta}$: regression parameter, fixed effects coefficient

$\boldsymbol{\alpha}$: variance components

$$\boldsymbol{\alpha} \Rightarrow \quad \boldsymbol{D}(\boldsymbol{\alpha}) \quad \text{and} \quad \boldsymbol{R}(\boldsymbol{\alpha})$$

where $\text{cov}(\boldsymbol{Y}_i) = \boldsymbol{Z}_i \boldsymbol{D} \boldsymbol{Z}_i^T + \boldsymbol{R}_i$

Normality:

$$\begin{aligned} E(\boldsymbol{Y}_i) &= \boldsymbol{X}_i \boldsymbol{\beta} \\ \mathsf{cov}(\boldsymbol{Y}_i) &= \boldsymbol{\Sigma}_i(\boldsymbol{\alpha}) \end{aligned}$$

$$\begin{aligned} f(\boldsymbol{Y}_i; \boldsymbol{\beta}, \boldsymbol{\alpha}) &= |\boldsymbol{\Sigma}_i|^{-1/2} (2\pi)^{-n_i/2} \times \\ & \exp\left[-\frac{1}{2} (\boldsymbol{Y}_i - \boldsymbol{X}_i \boldsymbol{\beta})^T \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{X}_i \boldsymbol{\beta}) \right] \end{aligned}$$

LMM Score Equations

$$
\begin{aligned}
\boldsymbol{U}_1(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= \sum_i \frac{\partial}{\partial \boldsymbol{\beta}} \log f_i \\
\boldsymbol{U}_2(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= \sum_i \frac{\partial}{\partial \boldsymbol{\alpha}} \log f_i
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{U}_1(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= -\sum_i \boldsymbol{X}_i^T \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{X}_i \boldsymbol{\beta}) \\
\boldsymbol{U}_{2,j}(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= \sum_i \frac{1}{2} \left[(\boldsymbol{Y}_i - \boldsymbol{X}_i \boldsymbol{\beta})^T \boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{\Sigma}_i}{\partial \alpha_j} \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{X}_i \boldsymbol{\beta}) \right. \\
&\qquad \left. - \mathsf{trace} \left(\boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{\Sigma}_i}{\partial \alpha_j} \right) \right]
\end{aligned}
$$

$\boxed{\star}$ Details: see Lindstrom & Bates (1988).

Note: The parameter $\boldsymbol{\alpha}$ is constrained to yield a covariance matrix that is positive definite. There are several computational methods for parameterizing $\boldsymbol{\Sigma}$ to satisfy the constraints (ie. use Cholesky decomposition $\boldsymbol{C}^T\boldsymbol{C} = \boldsymbol{\Sigma}$).

Restricted Maximum Likelihood (REML)

In the balanced and complete case we obtain:

$$
\begin{aligned}
\widehat{\boldsymbol{\Sigma}}_{ML} &= \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{Y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}})(\boldsymbol{Y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}})^T \\
E(\widehat{\boldsymbol{\Sigma}}_{ML}) &= \frac{1}{N}\sum_{i=1}^{N} E\left[(\boldsymbol{Y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}})(\boldsymbol{Y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}})^T\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} E\left[(\boldsymbol{Y}_i - \boldsymbol{\mu}_i + \boldsymbol{\mu}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}})\times\right. \\
&\qquad \left.(\boldsymbol{Y}_i - \boldsymbol{\mu}_i + \boldsymbol{\mu}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}})^T\right] \\
&= \frac{1}{N}\sum_{i=1}^{N}\left[\boldsymbol{\Sigma} - 2\boldsymbol{X}_i\text{cov}(Y_i, \widehat{\boldsymbol{\beta}}) + \boldsymbol{X}_i\text{cov}(\widehat{\boldsymbol{\beta}})\boldsymbol{X}_i^T\right]
\end{aligned}
$$

Restricted Maximum Likelihood (REML)

$$\text{(continued)} \quad E(\widehat{\boldsymbol{\Sigma}}_{ML}) = \frac{1}{N}\sum_{i=1}^{N}\left[\boldsymbol{\Sigma} - \boldsymbol{X}_i\left(\sum_j \boldsymbol{X}_j^T\boldsymbol{\Sigma}_j^{-1}\boldsymbol{X}_j\right)^{-1}\boldsymbol{X}_i^T\right]$$

$$= \boldsymbol{\Sigma} - \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{X}_i\left(\sum_j \boldsymbol{X}_j^T\boldsymbol{\Sigma}_j^{-1}\boldsymbol{X}_j\right)^{-1}\boldsymbol{X}_i^T$$

- Therefore, there is some bias in $\widehat{\boldsymbol{\Sigma}}_{ML}$ due to the estimation of $\boldsymbol{\beta}$.

Q: Can we correct for this bias?

Answer: Yes using REML! (Searle, Cassella & McCulloch, section 6.6)

Restricted Maximum Likelihood (REML)

"restricted maximum likelihood"

"residual maximum likelihood"

$\boxed{\star}$ There are many perspectives that justify REML:

1. Bias correction.
2. Likelihood for $\mathbf{\Sigma}$ based on error contrasts.
3. Conditional likelihood function.
4. Bayesian posterior for $\mathbf{\Sigma}$ with a flat prior on $\boldsymbol{\beta}$.

REML

$$\underset{(n \times 1)}{\boldsymbol{Y}} \sim \mathcal{N}(\underset{(n \times p)\,(p \times 1)}{\boldsymbol{X}\boldsymbol{\beta}}, \underset{(n \times n)}{\boldsymbol{\Sigma}})$$

$\boxed{\text{Define:}}$ "error contrast"

$$\begin{aligned} \boldsymbol{C} \quad &: \quad (n-p) \times n \quad \text{matrix} \\ E(\boldsymbol{C}\boldsymbol{Y}) \quad &= \quad 0 \;\; \forall \, \boldsymbol{\beta} \\ &\Leftrightarrow \quad \boldsymbol{C}\boldsymbol{X} = 0 \end{aligned}$$

$\boxed{\text{Define:}}$ error contrast based on OLS

$$\boldsymbol{A} = \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$$

$\boxed{\text{Define:}}$ Orthonormal basis for column span of $\boldsymbol{A}$

$$\begin{aligned} \boldsymbol{H} &: \quad n \times (n-p) \quad \text{matrix} \\ \boldsymbol{A} &= \boldsymbol{H}\boldsymbol{H}^T \\ \boldsymbol{I}_{(n-p)} &= \boldsymbol{H}^T\boldsymbol{H} \end{aligned}$$

$\boxed{\text{Define:}}$

$$\begin{aligned} \boldsymbol{G} &: \quad n \times p \quad \text{matrix} \\ \boldsymbol{G}^T &= (\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1} \end{aligned}$$

Define:

$$\boldsymbol{T} = \underset{n-p,\;\; p}{[\, \boldsymbol{H} \,,\, \boldsymbol{G} \,]}$$

Note:

$$\begin{aligned}
|\boldsymbol{T}| &= |\boldsymbol{T}^T\boldsymbol{T}|^{1/2} \\
&= \left|[\boldsymbol{H},\boldsymbol{G}]^T\,[\boldsymbol{H},\boldsymbol{G}]\right|^{1/2} \\
&= \left|\begin{array}{cc} \boldsymbol{I} & \boldsymbol{H}^T\boldsymbol{G} \\ \boldsymbol{G}^T\boldsymbol{H} & \boldsymbol{G}^T\boldsymbol{G} \end{array}\right|^{1/2} \\
&= \left|\boldsymbol{G}^T\boldsymbol{G} - \boldsymbol{G}^T\boldsymbol{H}\boldsymbol{H}^T\boldsymbol{G}\right|^{1/2} \\
&= \left|\boldsymbol{G}^T(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{G}\right|^{1/2} \\
&= |(\boldsymbol{X}^T\boldsymbol{X})^{-1}|^{1/2}
\end{aligned}$$

- $\boldsymbol{T}$ is full-rank if $\boldsymbol{X}$ is.
- $\boldsymbol{T}$ is the jacobian of $\boldsymbol{Y} \longrightarrow \boldsymbol{T}^T\boldsymbol{Y}$.

Note:

$$\mathcal{L}(\boldsymbol{Y};\boldsymbol{\beta},\boldsymbol{\Sigma}) \propto \mathcal{L}(\boldsymbol{T}^T\boldsymbol{Y};\boldsymbol{\beta},\boldsymbol{\Sigma})$$

Note:

$$\begin{aligned}
\boldsymbol{T}^T\boldsymbol{Y} &= [\boldsymbol{H},\boldsymbol{G}]^T\,\boldsymbol{Y} \\
&= \left[\begin{array}{c} \boldsymbol{H}^T\boldsymbol{Y} \\ \boldsymbol{G}^T\boldsymbol{Y} \end{array}\right] = \left[\begin{array}{c} \boldsymbol{H}^T\boldsymbol{Y} \\ \widehat{\boldsymbol{\beta}}(\boldsymbol{\Sigma}) \end{array}\right]
\end{aligned}$$

Note:

$$
\begin{aligned}
\mathsf{cov}(\boldsymbol{H}^T\boldsymbol{Y}, \boldsymbol{G}^T\boldsymbol{Y}) &= E[\boldsymbol{H}^T\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{G}] - \underbrace{E[\boldsymbol{H}^T\boldsymbol{Y}]}_{0} E[\boldsymbol{G}^T\boldsymbol{Y}]^T \\
&= \boldsymbol{H}^T \left[\boldsymbol{\Sigma} + \boldsymbol{X}\boldsymbol{\beta}\boldsymbol{\beta}^T\boldsymbol{X}\right] \boldsymbol{G} \\
&= \boldsymbol{H}^T\boldsymbol{\Sigma}\boldsymbol{G} + \underbrace{\boldsymbol{H}\boldsymbol{X}}_{0}\boldsymbol{\beta}\boldsymbol{\beta}^T\boldsymbol{X}\boldsymbol{G} \\
&= \underbrace{\boldsymbol{H}^T\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}}_{0}\left(\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1} \\
&= 0
\end{aligned}
$$

REML

Therefore,

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{Y};\boldsymbol{\beta},\boldsymbol{\Sigma}) &= |\boldsymbol{T}|\mathcal{L}(\boldsymbol{T}^T\boldsymbol{Y};\boldsymbol{\beta},\boldsymbol{\Sigma}) \\
&= |\boldsymbol{X}^T\boldsymbol{X}|^{-1/2}\mathcal{L}(\boldsymbol{H}^T\boldsymbol{Y};\boldsymbol{\Sigma})\ \mathcal{L}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\Sigma});\boldsymbol{\beta},\boldsymbol{\Sigma}) \\
\Rightarrow & \\
\mathcal{L}(\boldsymbol{H}^T\boldsymbol{Y};\boldsymbol{\Sigma}) &= |\boldsymbol{X}^T\boldsymbol{X}|^{1/2}\frac{\mathcal{L}(\boldsymbol{Y};\boldsymbol{\beta},\boldsymbol{\Sigma})}{\mathcal{L}(\widehat{\boldsymbol{\beta}};\boldsymbol{\beta},\boldsymbol{\Sigma})}
\end{aligned}
$$

$$
\begin{aligned}
&= c \cdot |\boldsymbol{X}^T\boldsymbol{X}|^{1/2} \frac{|\boldsymbol{\Sigma}|^{-1/2}\exp[-\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})]}{|\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X}|^{1/2}\exp[-\frac{1}{2}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})^T\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})]} \\
&= c \cdot |\boldsymbol{X}^T\boldsymbol{X}|^{1/2}|\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X}|^{-1/2}|\boldsymbol{\Sigma}|^{-1/2} \times \\
&\qquad \exp[-\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})]
\end{aligned}
$$

$$
\ell(\boldsymbol{H}^T\boldsymbol{Y};\boldsymbol{\Sigma}) = \ell(\boldsymbol{Y};\widehat{\boldsymbol{\beta}}(\boldsymbol{\Sigma}),\boldsymbol{\Sigma}) - \frac{1}{2}\log|\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X}|
$$

REML Comments

1. ℓ_{REML} is free of $\boldsymbol{\beta}$.

2. $\mathcal{L}_{REML}$ is likelihood function for $\boldsymbol{H}^T\boldsymbol{Y}$.

3. $\mathcal{L}_{REML}$ is conditional likelihood function: $\mathcal{L}(\boldsymbol{Y} \mid \widehat{\boldsymbol{\beta}}(\boldsymbol{\Sigma})\ )$.

4. There is a Bayesian interpretation for $\widehat{\boldsymbol{\Sigma}}_{REML}$. See Laird and Ware (1982), Harville (1974, 1976, 1977).

REML and Bayes

- Likelihood function factors:

$$f(\boldsymbol{Y} \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}) = f(\boldsymbol{H}^T\boldsymbol{Y} \mid \boldsymbol{\Sigma}) \cdot f(\widehat{\boldsymbol{\beta}} \mid \boldsymbol{\beta}, \boldsymbol{\Sigma})$$

- posterior $\pi(\boldsymbol{\Sigma} \mid \boldsymbol{Y})$:

$$\begin{aligned}
\pi(\boldsymbol{\Sigma} \mid \boldsymbol{Y}) &= \int_{\boldsymbol{\beta}} \pi(\boldsymbol{\beta}, \boldsymbol{\Sigma} \mid \boldsymbol{Y}) \\
&= \frac{\int_{\boldsymbol{\beta}} f(\boldsymbol{H}^T\boldsymbol{Y} \mid \boldsymbol{\Sigma}) \cdot f(\widehat{\boldsymbol{\beta}} \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}) \cdot \pi(\boldsymbol{\beta}, \boldsymbol{\Sigma})}{\int_{\boldsymbol{\beta}} \int_{\boldsymbol{\Sigma}} f(\boldsymbol{Y} \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}) \cdot \pi(\boldsymbol{\beta}, \boldsymbol{\Sigma})} \\
&= f(\boldsymbol{H}^T\boldsymbol{Y} \mid \boldsymbol{\Sigma}) \frac{\int_{\boldsymbol{\beta}} f(\widehat{\boldsymbol{\beta}} \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}) \cdot \pi(\boldsymbol{\beta}, \boldsymbol{\Sigma})}{\int_{\boldsymbol{\beta}} \int_{\boldsymbol{\Sigma}} f(\boldsymbol{Y} \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}) \cdot \pi(\boldsymbol{\beta}, \boldsymbol{\Sigma})}
\end{aligned}$$

REML and Bayes

$\boxed{\textbf{Q}:}$ What if $\pi(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \pi(\boldsymbol{\Sigma}) \cdot \mathcal{N}(0, \boldsymbol{\Gamma})$ and we let $\boldsymbol{\Gamma}^{-1} \rightarrow \mathbf{0}$?

$\boxed{\textbf{A}:}$ $$\pi(\boldsymbol{\Sigma} \mid \boldsymbol{Y}) \propto f(\boldsymbol{H}^T \boldsymbol{Y} \mid \boldsymbol{\Sigma}) \cdot \pi(\boldsymbol{\Sigma}).$$

Therefore, there is a Bayesian justification for the use of REML.

Empirical Bayes Estimates

- In certain situations we want to know about the random effects, $\boldsymbol{b}_i$:
  - Individual trajectories.
  - Hospital profiling.

- We can use **posterior** distributions to estimate these unknown "parameters"

$$[\boldsymbol{b}_i \mid \boldsymbol{Y}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}]$$

- If we simply use the MLEs for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ then we can estimate the random effects using

$$E[\boldsymbol{b}_i \mid \boldsymbol{Y}_i, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}}]$$

- These estimates are typically different from those estimates obtained using only the individual subject's data. They are usually "shrunk" toward the values of parameters that other subjects have.

EB Estimates: Motivation

- **Viral Samples in Villages**

$$
\begin{array}{ccccc}
i=1 & i=2 & i=3 & \ldots & i=N \\
V_{11} & V_{21} & V_{31} & & V_{N1} \\
V_{12} & V_{22} & V_{32} & & V_{N2} \\
 & V_{23} & V_{33} & & \\
 & V_{24} & V_{34} & & \\
 & V_{24} & & &
\end{array}
$$

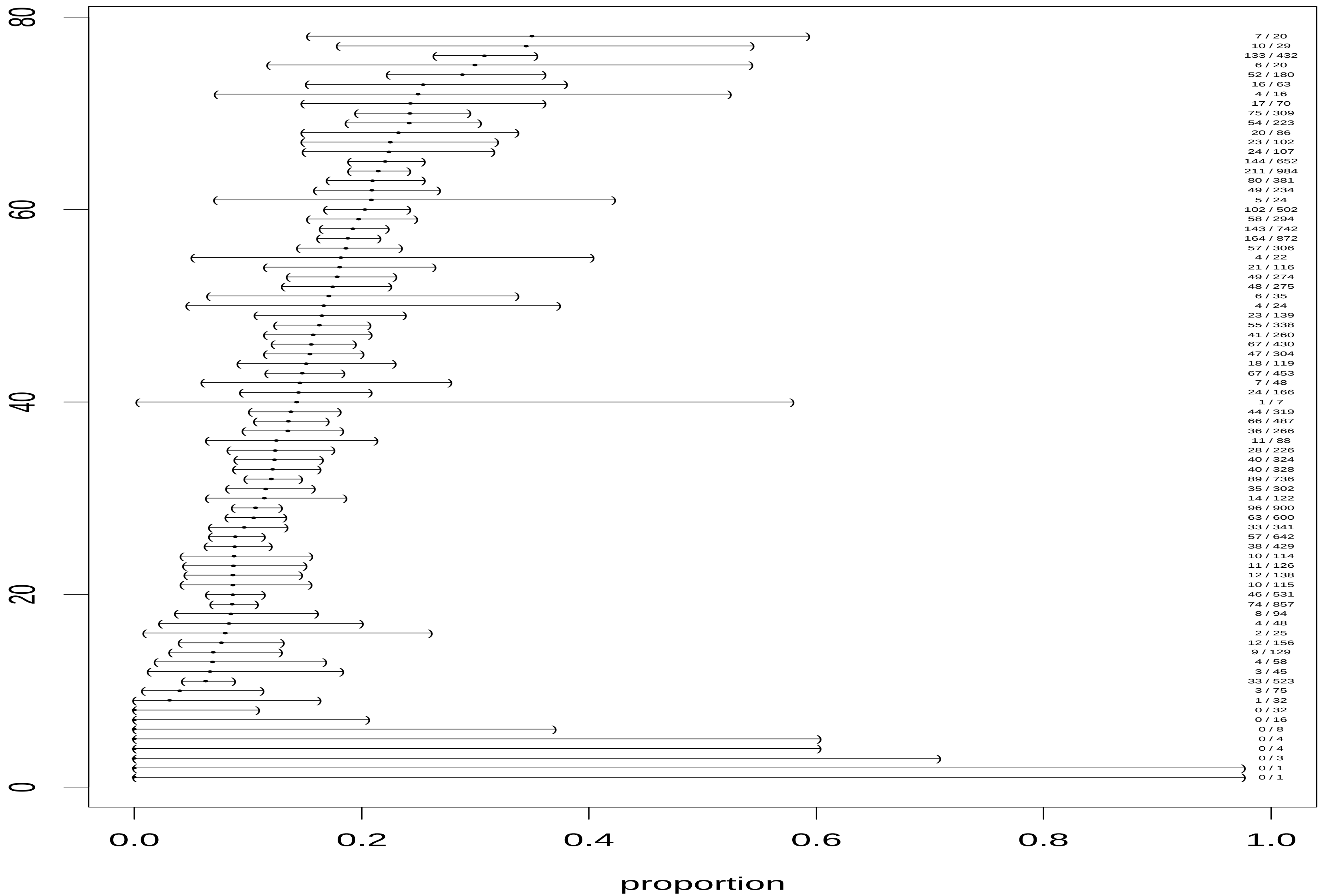
▷ Each measurement V_{ij} is a measure of pathogenicity of the virus.

▷ Different number of samples available

▷ **Q**: which village has the worst mean pathogenicity?

EB Estimates: Motivation

- **Hospital Ratings**
  - ▷ WA State data on GI surgery and adverse events: (S_i, E_i).
  - ▷ E_i/S_i is a crude rate of AE.
  - ▷ **Q**: Estimate the hospital-specific risk?
  - ▷ **ISSUE:** the number of recorded surgeries varies.
  - ▷ **ISSUE:** the "case-mix" varies.
- **Regression Context**
  - ▷ $Y_{i1}, Y_{i2}, \ldots, Y_{in_i}$ $\boxed{\text{group } i}$, μ_i.
  - ▷ **Goal**: Estimate $E(Y_{ij}) = \mu_i$.


80
60
40
20
0
7 / 20
10 / 29
133 / 432
6 / 20
52 / 180
16 / 63
4 / 16
17 / 70
75 / 309
54 / 223
20 / 86
23 / 102
24 / 107
144 / 652
211 / 984
80 / 381
49 / 234
5 / 24
102 / 502
58 / 294
143 / 742
164 / 872
57 / 306
4 / 22
21 / 116
49 / 274
48 / 275
6 / 35
4 / 24
23 / 139
55 / 338
41 / 260
67 / 430
47 / 304
18 / 119
67 / 453
7 / 48
24 / 166
1 / 7
44 / 319
66 / 487
36 / 266
11 / 88
28 / 226
40 / 324
40 / 328
89 / 736
35 / 302
14 / 122
96 / 900
63 / 600
33 / 341
57 / 642
38 / 429
10 / 114
11 / 126
12 / 138
10 / 115
46 / 531
74 / 857
8 / 94
4 / 48
2 / 25
12 / 156
9 / 129
4 / 58
3 / 45
33 / 523
3 / 75
1 / 32
0 / 32
0 / 16
0 / 8
0 / 4
0 / 4
0 / 3
0 / 1
0 / 1
0.0
0.2
0.4
0.6
0.8
1.0
proportion

Empirical Bayes: Simplest Example

- Suppose we have "performance" measures, Y_{ij} and are interested in the "skill" level, $\theta_i = E(Y_{ij})$.
- $\boxed{1}$ **Individual observations**

$$\begin{aligned} Y_{ij} \mid \theta_i &\sim \mathcal{N}(\theta_i, \sigma^2) \\ \overline{Y}_i \mid \theta_i &\sim \mathcal{N}(\theta_i, \frac{\sigma^2}{n_i}) \end{aligned}$$

- $\boxed{2}$ **Population model**

$$\theta_i \sim \mathcal{N}(\mu, \tau^2)$$

- $Y_{ij} = \mu + \underbrace{(\theta_i - \mu)}_{b_i} + e_{ij}$

EB: Example

- Estimation of μ, the population mean, can be based on WLS (Gauss-Markov):

$$\widehat{\mu} = \frac{\sum_i w_i \overline{Y}_i}{\sum_i w_i}$$

where

$$w_i = \left(\frac{\sigma^2}{n_i} + \tau^2 \right)^{-1}$$

- **Q**: but what about **individual** skill estimates?

EB: Example

$$Y_{ij} \mid \theta_i \sim \mathcal{N}(\theta_i, \sigma^2) \qquad \textbf{Tests}$$
$$\theta_i \sim \mathcal{N}(\mu, \tau^2) \qquad \textbf{Population}$$

We can then consider the joint distribution of performance means and the skills:

$\overline{\boldsymbol{Y}} = \text{stack}(\overline{Y}_i) \qquad \boldsymbol{\theta} = \text{stack}(\theta_i)$

$$\begin{bmatrix} \overline{\boldsymbol{Y}} \\ \boldsymbol{\theta} \end{bmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu} \end{pmatrix}, \begin{pmatrix} \text{diag}(\tau^2 + \sigma^2/n_i) & \text{diag}(\tau^2) \\ \text{diag}(\tau^2) & \text{diag}(\tau^2) \end{pmatrix} \right]$$

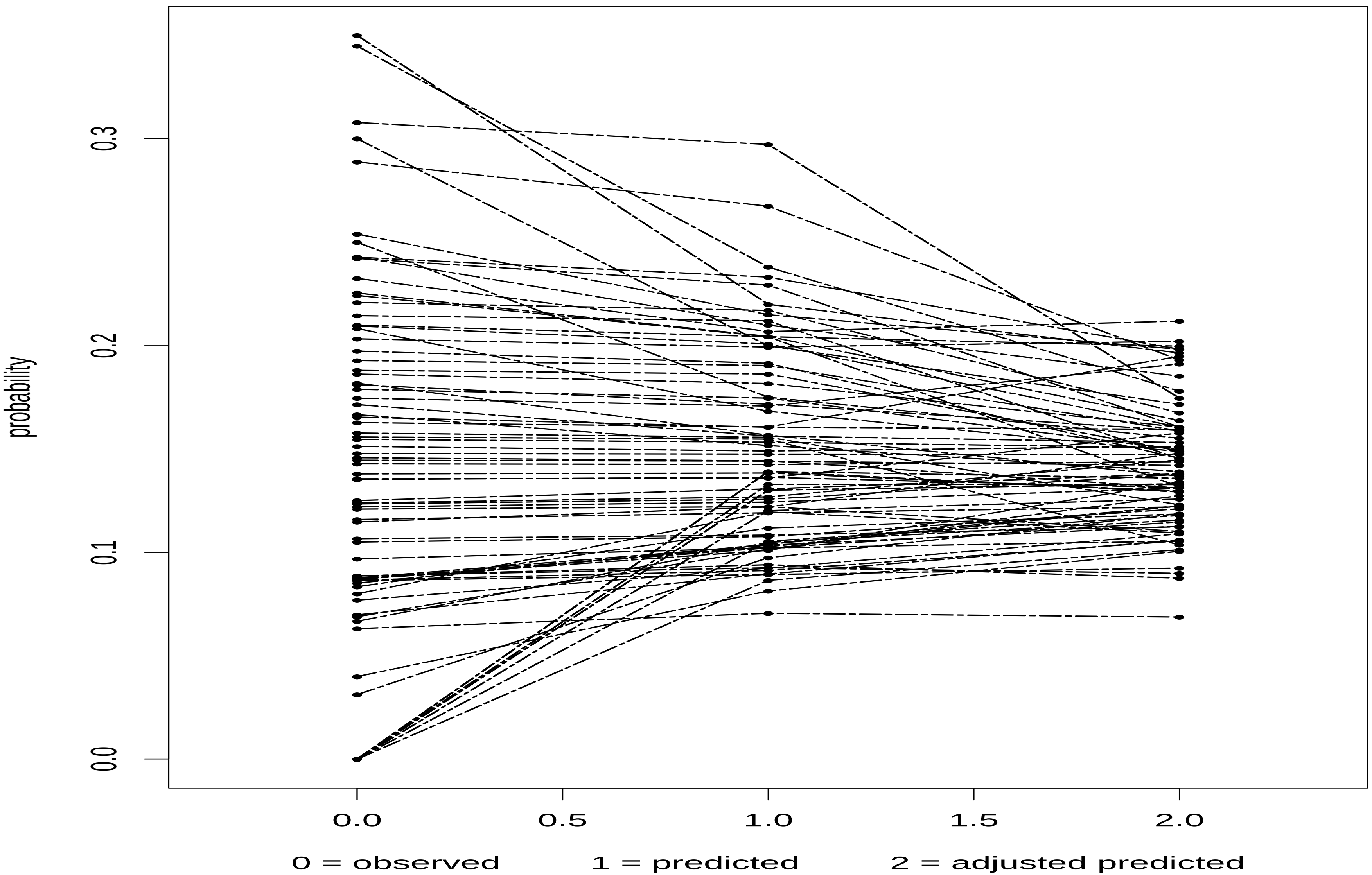
EB: Example

This joint distribution implies the conditional:

$$\boldsymbol{\theta} \mid \overline{\boldsymbol{Y}} \quad \sim \quad \mathcal{N}\left[\boldsymbol{\mu} + \text{diag}(\frac{\tau^2}{\tau^2 + \sigma^2/n_i})(\overline{\boldsymbol{Y}} - \boldsymbol{\mu}), \boldsymbol{\Sigma}\right]$$

So each individual skill can be estimated by it's conditional mean

$$\begin{aligned}
E(\theta_i \mid \overline{\boldsymbol{Y}}) &= \mu + \frac{\tau^2}{\tau^2 + \sigma^2/n_i}(\overline{Y}_i - \mu) \\
&= \frac{\sigma^2/n_i}{\tau^2 + \sigma^2/n_i} \cdot \mu + \frac{\tau^2}{\tau^2 + \sigma^2/n_i} \cdot \overline{Y}_i \\
&= \lambda_i \cdot \mu \; + \; (1 - \lambda_i) \cdot \overline{Y}_i
\end{aligned}$$


probability
0.0
0.1
0.2
0.3
0.0
0.5
1.0
1.5
2.0
0 = observed
1 = predicted
2 = adjusted predicted

EB: Example

- $\boxed{\text{Lindley \& Smith (1972)}}$

$$
\begin{aligned}
\boldsymbol{Y}_i &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{X}\boldsymbol{b}_i + \boldsymbol{e}_i \\
\boldsymbol{b}_i &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{D}) \\
\boldsymbol{e}_i &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{R} = \sigma^2 \boldsymbol{I})
\end{aligned}
$$

Then in this case $E(\boldsymbol{Y}_i \mid \boldsymbol{b}_i) = \boldsymbol{X}(\boldsymbol{\beta} + \boldsymbol{b}_i) = \boldsymbol{X}\boldsymbol{\beta}_i$.

- $\boxed{\textbf{Empirical Bayes Estimates}}$

$$
\begin{aligned}
\tilde{\boldsymbol{\beta}}_i &= \left(\sigma^{-2}\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{D}^{-1}\right)^{-1}\left(\sigma^{-2}\boldsymbol{X}^T\boldsymbol{Y}_i + \boldsymbol{D}^{-1}\widehat{\boldsymbol{\beta}}\right) \\
&= W_1 \cdot \widehat{\boldsymbol{\beta}}_i + W_2 \cdot \widehat{\boldsymbol{\beta}}
\end{aligned}
$$

EB: Example

- $\boxed{\text{Lindley \& Smith (1972)}}$ continued...

$$W_1 = \left(\sigma^{-2}\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{D}^{-1}\right)^{-1}\sigma^{-2}\boldsymbol{X}^T\boldsymbol{X}$$

$$W_2 = \left(\sigma^{-2}\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{D}^{-1}\right)^{-1}\boldsymbol{D}^{-1}$$

- Individual estimate "borrows strength" from other data depending on $\boldsymbol{D}$.

EB: Linear Mixed Model

- **Model**

$$\begin{aligned} \boldsymbol{Y}_i &= \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i \\ \boldsymbol{b}_i &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{D}) \\ \boldsymbol{e}_i &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{R}) \end{aligned}$$

- **Estimation**

  ▷ $\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{D}}, \widehat{\boldsymbol{R}}$ via ML or REML.

  ▷ **Q**: but what about estimates for $\boldsymbol{b}_i$?

EB: Linear Mixed Model

$$\begin{pmatrix} \boldsymbol{Y}_i \\ \boldsymbol{b}_i \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \boldsymbol{X}_i\boldsymbol{\beta} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i^T + \boldsymbol{R}_i & \boldsymbol{Z}_i\boldsymbol{D} \\ \boldsymbol{D}\boldsymbol{Z}_i^T & \boldsymbol{D} \end{pmatrix}\right]$$

This leads to the conditional distribution

$$[\boldsymbol{b}_i \mid \boldsymbol{Y}_i] \sim \mathcal{N}\left(\boldsymbol{D}\boldsymbol{Z}_i^T(\boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i^T + \boldsymbol{R}_i)^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}),\ \boldsymbol{\Sigma}_i^*\right)$$

$$\begin{aligned} \boldsymbol{\Sigma}_i^* &= \boldsymbol{D} - \boldsymbol{D}\boldsymbol{Z}_i^T(\boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i^T + \boldsymbol{R}_i)^{-1}\boldsymbol{Z}_i^T\boldsymbol{D} \\ &= \left(\boldsymbol{Z}_i^T\boldsymbol{R}_i^{-1}\boldsymbol{Z}_i + \boldsymbol{D}^{-1}\right)^{-1} \end{aligned}$$

Therefore, the **empirical Bayes** estimate is

$$\tilde{\boldsymbol{b}}_i = \boldsymbol{D}\boldsymbol{Z}_i^T(\boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i^T + \boldsymbol{R}_i)^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})$$

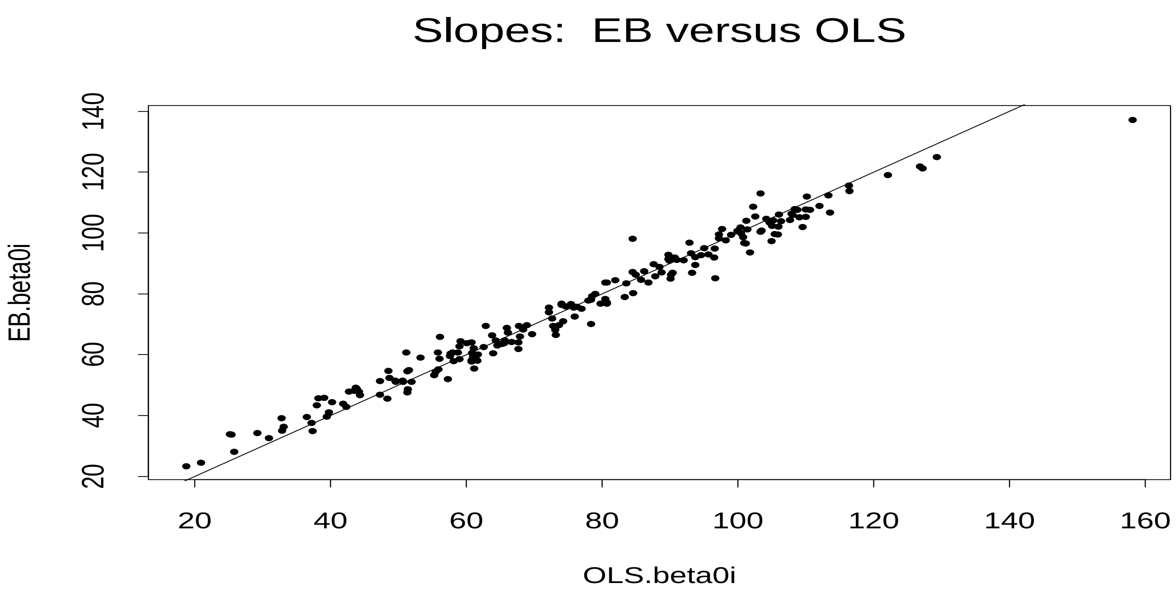
R/S+ Program:

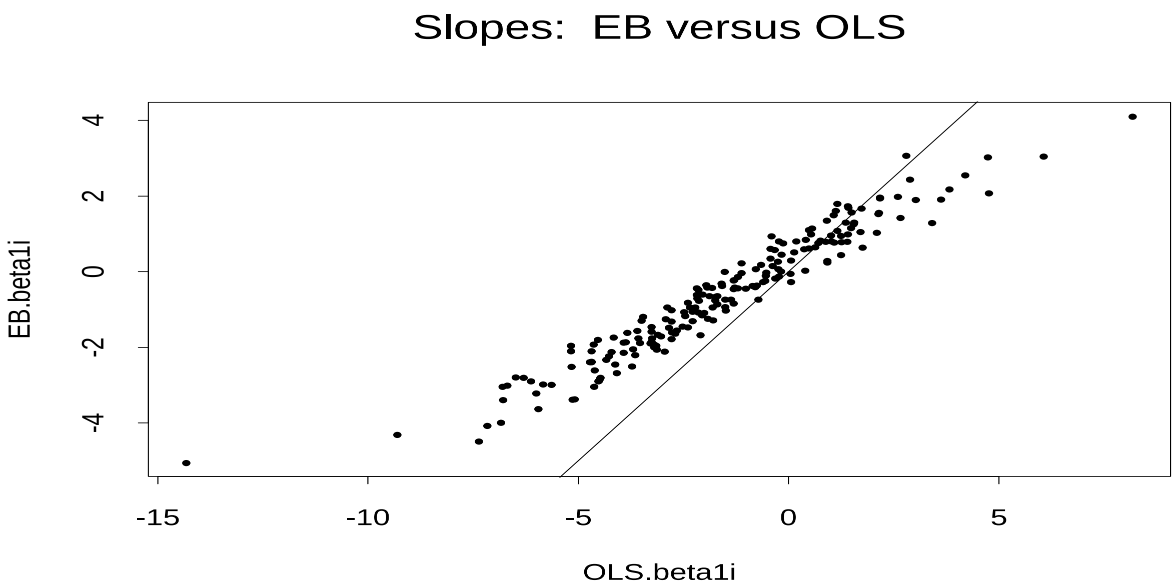
```
#
#####
##### Linear mixed models
#####
#
options( contrasts="contr.treatment" )
#
fit1 <- lme(
            fixed = fev1 ~ age0 + ageL + female*ageL + factor(f508)*ageL,
            random = ~ 1 + ageL,
            cluster = ~id,
            est.method = "ML",
            re.structure = "unstructured",
            data = cfkids )
#
#####
##### Empirical Bayes
#####
#
X <- model.matrix( ~ age0 + ageL + female*ageL + factor(f508)*ageL,
                     data=cfkids )
X[,c(3,7,8,9)] <- 0
beta <- fit1$coefficients$fixed
f0 <- as.vector( X%*%beta )
#
```

```
##### be careful about ID order -- build data sets and merge
#
f0 <- unlist( lapply( split( f0, cfkids$id ), min ) )
id0 <- unlist( lapply( split( cfkids$id, cfkids$id ), min ) )
data0 <- data.frame( id=id0, f0=f0 )
#
b0 <- fit1$coefficients$random[,1]
b1 <- fit1$coefficients$random[,2]
id1 <- as.numeric( dimnames( fit1$coefficients$random )[[1]] )
data1 <- data.frame( id=id1, b0=b0, b1=b1 )
#
combo <- merge( data0, data1, by="id", all=T )
ooo <- order( combo$id )
combo <- combo[ ooo, ]
EB.beta0i <- combo$f0 + combo$b0
EB.beta1i <- beta[3] + combo$b1
#
##### OLS estimates
#
uid <- unique( cfkids$id )
m <- length( uid )
OLS.beta0i <- rep( NA, m )
OLS.beta1i <- rep( NA, m )
id2 <- rep( NA, m )
for( j in 1:m ){
     id2[j] <- uid[j]
     keep <- cfkids$id == uid[j]
     yj <- cfkids$fev1[ keep ]
```

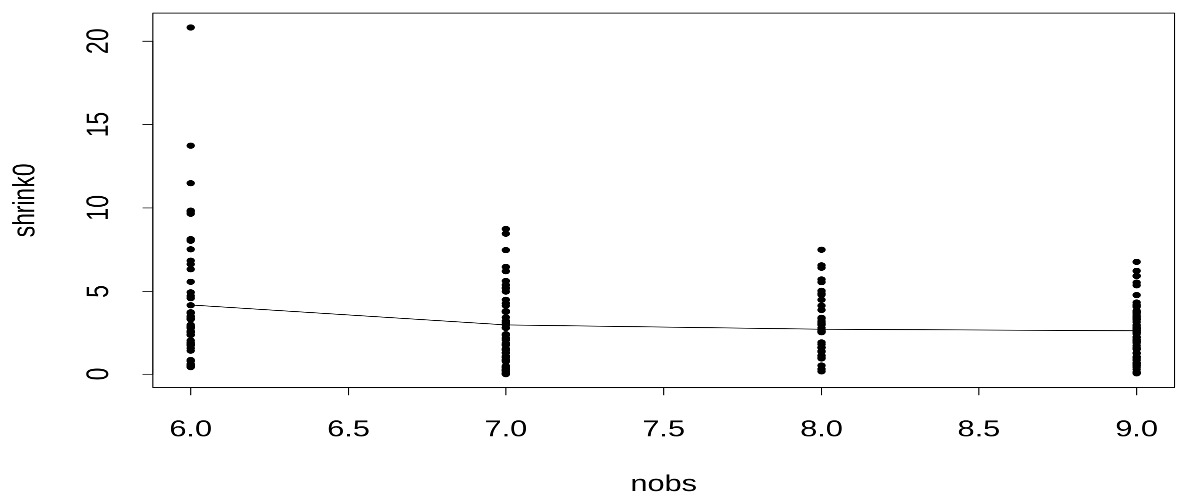
```
     xj <- cfkids$ageL[ keep ]
     fit <- lm( yj ~ xj )
     OLS.beta0i[j] <- fit$coefficient[1]
     OLS.beta1i[j] <- fit$coefficient[2]
}
ooo <- order( id2 )
OLS.beta0i <- OLS.beta0i[ ooo ]
OLS.beta1i <- OLS.beta1i[ ooo ]
#
#####
##### compare the OLS and the EB estimates
#####
#
postscript( file="cfkids-EBplot.ps", horiz=F )
par( mfrow=c(2,1) )
plot( OLS.beta0i, EB.beta0i )
abline( 0, 1 )
title("Slopes:  EB versus OLS")
plot( OLS.beta1i, EB.beta1i )
abline( 0, 1 )
title("Slopes:  EB versus OLS")
graphics.off()
#
##### shrinkage
#
shrink0 <- abs( OLS.beta0i - EB.beta0i )
shrink1 <- abs( OLS.beta1i - EB.beta1i )
#
```

```
nobs <- unlist( lapply( split( cfkids$age, cfkids$id ), length ) )
VarAge <- unlist( lapply( split( cfkids$age, cfkids$id ), var ) )
postscript( file="cfkids-EBshrink.ps", horiz=F )
par( mfrow=c(2,1) )
plot( nobs, shrink0 )
lines( smooth.spline( nobs, shrink0 , df=3 ) )
title("Intercepts:  shrinkage verses N_i")
plot( VarAge, shrink1 )
lines( smooth.spline( VarAge, shrink1, df=3 ) )
title("Slopes:  shrinkage verses var(Age_i)")
graphics.off()
#
# end-of-file...
```

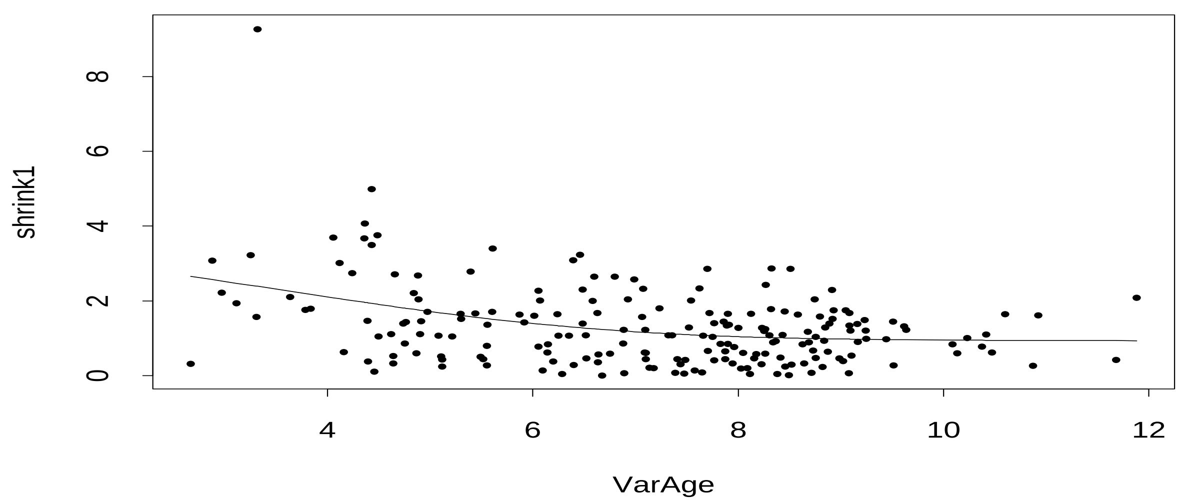

Slopes: EB versus OLS
EB.beta0i
OLS.beta0i
20
40
60
80
100
120
140
160


Slopes: EB versus OLS
EB.beta1i
OLS.beta1i
-4
-2
0
2
4
-15
-10
-5
5

Intercepts: shrinkage verses N_i



Slopes: shrinkage verses var(Age_i)



Residuals for LMM

Define:

$$\text{population residual} \quad : \quad \boldsymbol{r}_i^P = \boldsymbol{Y}_i - \boldsymbol{X}_i \widehat{\boldsymbol{\beta}}$$

$$\text{cluster residual} \quad : \quad \boldsymbol{r}_i^C = \boldsymbol{Y}_i - \boldsymbol{X}_i \widehat{\boldsymbol{\beta}} - \boldsymbol{Z}_i \widehat{\boldsymbol{b}_i}$$

Note: For the CF data using model 2 we have

```
> var( pop.res )
[1] 665.479
> var( cluster.res )
[1] 134.4172
```

- Residuals can be used to assess systematic model.

  Plot r_i^P versus covariate(s) (between- or within-)

  Plot r_i^C versus covariate(s) (within-)

- Residuals can be used to assess variance as a function of covariates.

R/S+ Program:

```
#
# cfkids-CDA-NewLMM.q
#
# ------------------------------------------------------------
#
# PURPOSE:  Use linear mixed models to characterize longitudinal
#            change by gender and genotype.
#
# AUTHOR:  P. Heagerty
#
# DATE:  00/07/10  Revised 14Feb2002
#
# ------------------------------------------------------------
#
#
#####
##### Linear Mixed Models
#####
#
options( contrasts=c("contr.treatment","contr.helmert") )
#

### Intercept plus AR(1) plus measurement error

fit3 <- lme( fev1 ~ age0 + ageL + female*ageL + factor(f508)*ageL,
```

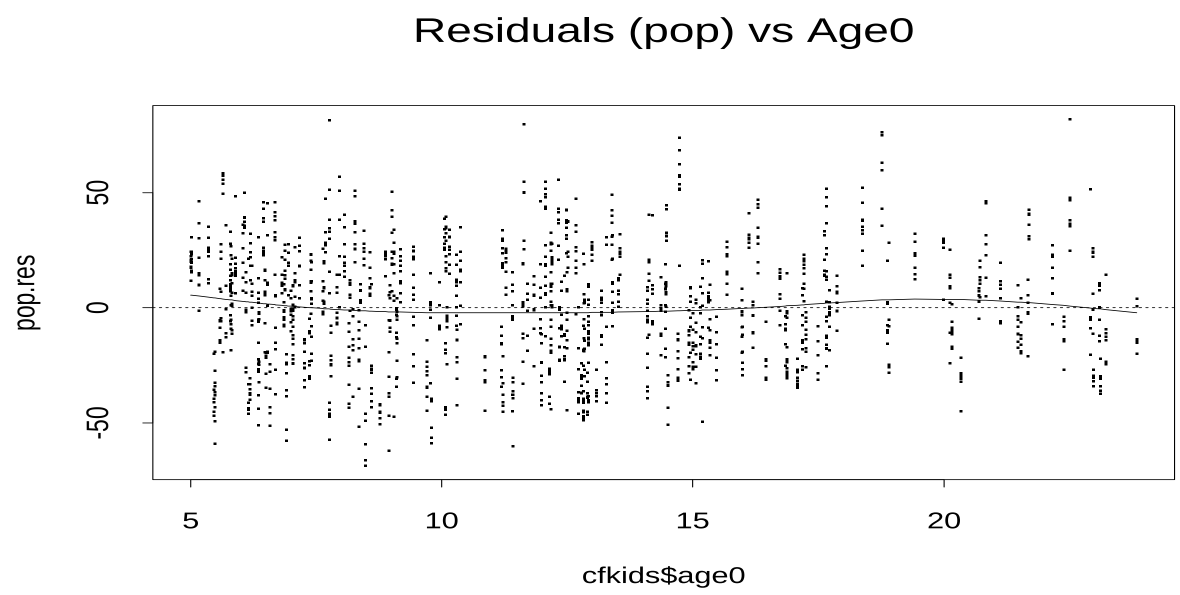
```
             method = "ML",
             random = reStruct( ~ 1 | id, pdClass="pdSymm", REML=F),
             correlation = corExp( form = ~ ageL | id, nugget=T),
             data = cfkids )
summary( fit3 )

#
#####
##### Residual Analysis -- using fit3
#####
#
pop.res <- resid( fit3, level=0 )
cluster.res <- resid( fit3, level=1 )
print( var( pop.res ) )
print( var( cluster.res ) )
#
postscript( file="cfkids-NewResiduals.ps", horiz=F )
par( mfrow=c(2,1) )
plot( cfkids$age0, pop.res, pch="." )
lines( smooth.spline( cfkids$age0, pop.res, df=5 ) )
title("Residuals (pop) vs Age0")
abline( h=0, lty=2 )
plot( cfkids$ageL, pop.res, pch="." )
lines( smooth.spline( cfkids$ageL, pop.res, df=5 ) )
title("Residuals (pop) vs AgeL")
abline( h=0, lty=2 )
graphics.off()
#
```

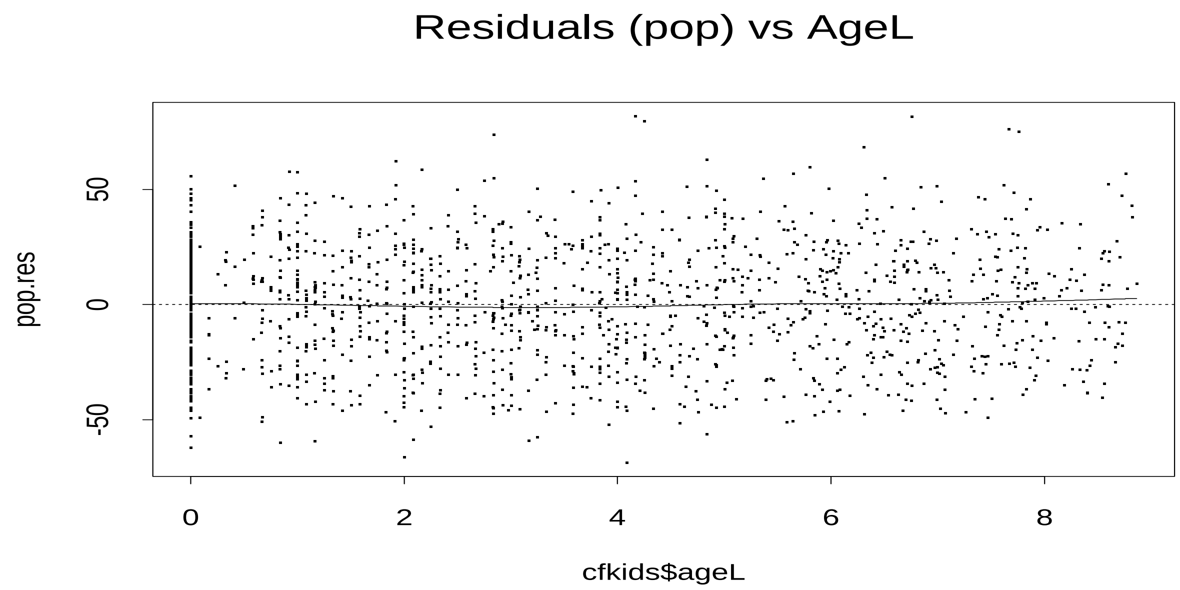
```
postscript( file="cfkids-NewResiduals2.ps", horiz=F )
par( mfrow=c(2,1) )
plot( cfkids$age0, cluster.res, pch="." )
lines( smooth.spline( cfkids$age0, cluster.res, df=5 ) )
abline( h=0, lty=2 )
title("Residuals (cluster) vs Age0")
b0 <- unlist( fit2$coefficients$random )
age0 <- unlist( lapply( split( cfkids$age0, cfkids$id ), min ) )
plot( age0, b0 )
lines( smooth.spline( age0, b0, df=5 ) )
abline( h=0, lty=2 )
title("EB b0 versus Age0")
graphics.off()

#
##### Do we need a quadratic age0???
#
fit4 <- lme( fev1 ~ age0 + age0^2 + ageL + female*ageL + factor(f508)*ageL,
            method = "ML",
            random = reStruct( ~ 1 | id, pdClass="pdSymm", REML=F),
            correlation = corExp( form = ~ ageL | id, nugget=T),
            data = cfkids )
summary( fit4 )
#
anova( fit3, fit4 )
#
# end-of-file...
```

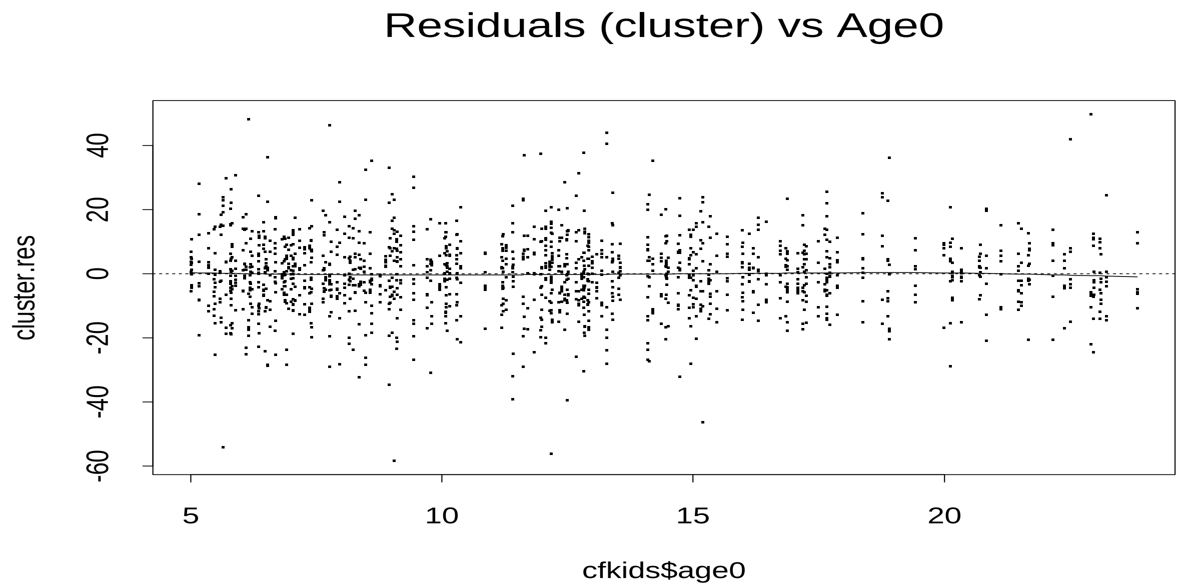
Fit 3 Random Intercepts + corExp + meas. error

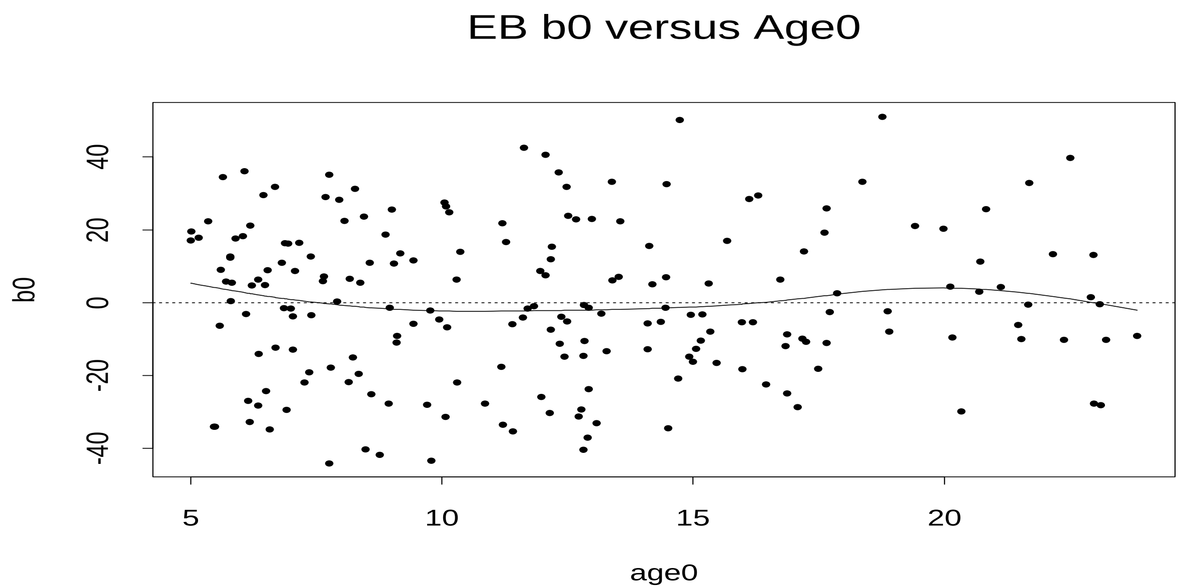
```
Linear mixed-effects model fit by maximum likelihood
 Data: cfkids
       AIC      BIC    logLik
  12384.92 12454.11 -6179.461

Random effects:
 Formula:  ~ 1 | id
        (Intercept) Residual
StdDev:    19.69916 15.86913

Correlation Structure: Exponential spatial correlation
 Formula:  ~ ageL | id
 Parameter estimate(s):
    range    nugget
 5.116653 0.2878059
Fixed effects: fev1 ~ age0 + ageL + female * ageL + factor(f508) * ageL
                      Value Std.Error   DF   t-value p-value
      (Intercept)  104.7466  6.760173 1309  15.49466  <.0001
             age0   -1.8739  0.328085  195  -5.71153  <.0001
             ageL   -0.7095  0.577371 1309  -1.22886  0.2193
           female   -1.2089  3.486725  195  -0.34671  0.7292
    factor(f508)1   -5.0121  5.753669  195  -0.87112  0.3848
    factor(f508)2   -7.0993  5.782222  195  -1.22778  0.2210
```

```
      female:ageL   -0.8259  0.372683 1309  -2.21612  0.0269
ageLfactor(f508)1   -0.3159  0.622931 1309  -0.50710  0.6122
ageLfactor(f508)2   -0.5817  0.621290 1309  -0.93627  0.3493

Standardized Within-Group Residuals:
       Min         Q1          Med         Q3       Max
 -3.734659 -0.5250052 5.392674e-06 0.4646477 3.215006

Number of Observations: 1513
Number of Groups: 200
```


Residuals (pop) vs Age0
pop.res
50
0
-50
5
10
15
20
cfkids$age0


Residuals (pop) vs AgeL
pop.res
50
0
-50
0
2
4
6
8
cfkids$ageL


Residuals (cluster) vs Age0
cluster.res
cfkids$age0
-60
-40
-20
0
20
40
5
10
15
20


EB b0 versus Age0
b0
age0
-40
-20
0
20
40
5
10
15
20

LMM Summary

- Observe $\boldsymbol{Y}_i$, $i = 1, 2, \ldots, m$ independent clusters.

- **Model**: (Laird & Ware, 1982)

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

- $\boldsymbol{\beta}$ is the coefficient(s) that are common to all clusters (fixed across clusters).

- $\boldsymbol{b}_i$ is the deviation of the coefficient(s) that vary from cluster to cluster (random across clusters).

- $(\beta_{i,j} + b_{i,j})$ is the coefficient of $X_{i,j}$ for cluster i.

$$\begin{aligned} \boldsymbol{b}_i &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{D}) \ \text{between-cluster} \\ \boldsymbol{e}_i &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{R}_i) \ \text{within-cluster} \end{aligned}$$

- **Estimation/Inference**: WLS, ML

LMM: Additional Topics

- **Influence Diagnostics**
  - Lesaffre and Verbeke (1998)
- **Alternative Random Effects Distributions**
  - Butler and Louis (1992)
- **Non- and Semi-parametric Regression Structure**
  - Lin and Zhang (1999)