

# Longitudinal Data Analysis with Mixed Models

## A Graphical Overview

Georges Monette, Ph.D., P.Stat.

[georges@yorku.ca](mailto:georges@yorku.ca)

[www.math.yorku.ca/~georges](http://www.math.yorku.ca/~georges)

Workshop on Longitudinal Data Analysis  
for Business and Biostatistics

May 10, 2007

Plan:.....	4
Take 1: The basic theory .....	5
A traditional example .....	6
One wrong analysis .....	8
Fixed effects regression model.....	14
Other approaches .....	21
Multilevel Models .....	22
From Multilevel Model to Mixed Model .....	24
Mixed Model in SAS: .....	27
Mixed Model in Matrices .....	29
Fitting the mixed model .....	31
SAS code and output for Mixed Model .....	32
Estimate statement.....	32
Sas code .....	32
Ouput .....	33
Diagnostics .....	41
Modeling dependencies in time.....	43
G-side vs. R-side .....	44
Simpler Models .....	47
BLUPS: Estimating Within-Subject Effects .....	48
Where the EBLUP comes from : looking at a single subject.....	54
When is a BLUP a BLUPPER?.....	60

Interpreting G .....	61
Differences between PROC GLM and MIXED with balanced data .....	68
Take 2: Learning lessons from unbalanced data .....	73
SAS code and output .....	77
Between, Within and Pooled Models .....	78
The Mixed Model .....	86
A serious a problem? a simulation .....	90
Splitting age into two variables .....	92
Using PROC MIXED with a contextual mean .....	102
Simulation Revisited .....	105
Power .....	107
Some links .....	107
A few books.....	108
Appendix A: Synopsis of SAS commands in PROC MIXED .....	109
Appendix B: PROC MIXED all-dressed.....	114
Appendix C: Reinterpreting weights.....	116

## Plan:

1. Introduce multilevel (hierarchical) models with a 'nice' example
2. Segue to mixed models
3. Consider PROC MIXED code and output
4. Autocorrelation in time
5. G-side and R-side: two jobs but with overlap
6. BLUPs
7. What happens with a not-so-nice example?
8. The relationship between ecological, marginal and conditional regression
9. What does the mixed model do? When is it biased?
10. Splitting unbalanced time-varying variables into two components
11. Synopsis of commands in PROC MIXED

## Take 1: The basic theory

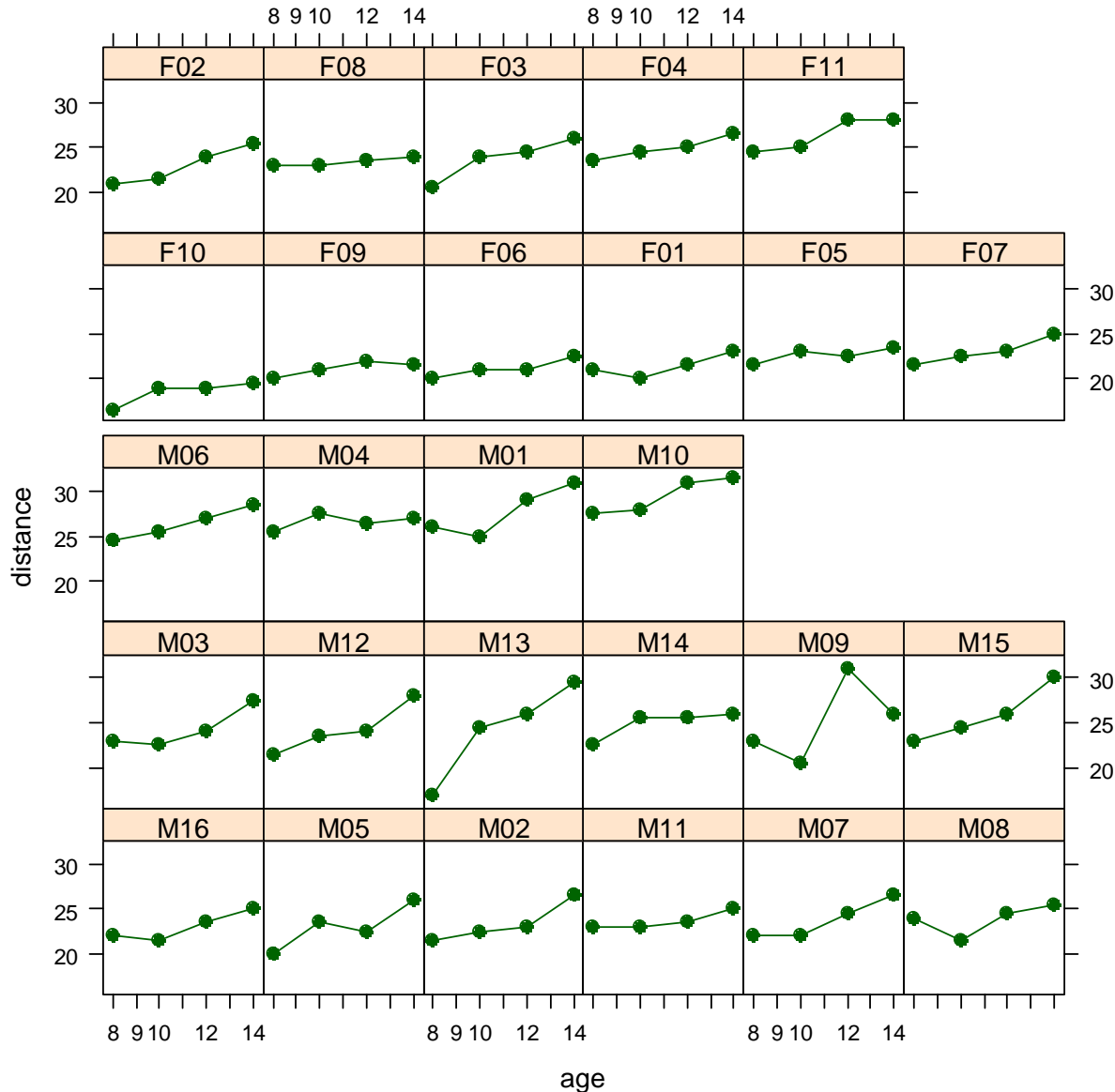
This part of the workshop focuses on models for a normally distributed response with linear predictors.

Formally:

$$\text{General Linear Model: } \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$$

$$\text{Mixed Model: } \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{R}) \perp \boldsymbol{\gamma} \sim N(\mathbf{0}, \mathbf{G})$$

# A traditional example



Balanced data:

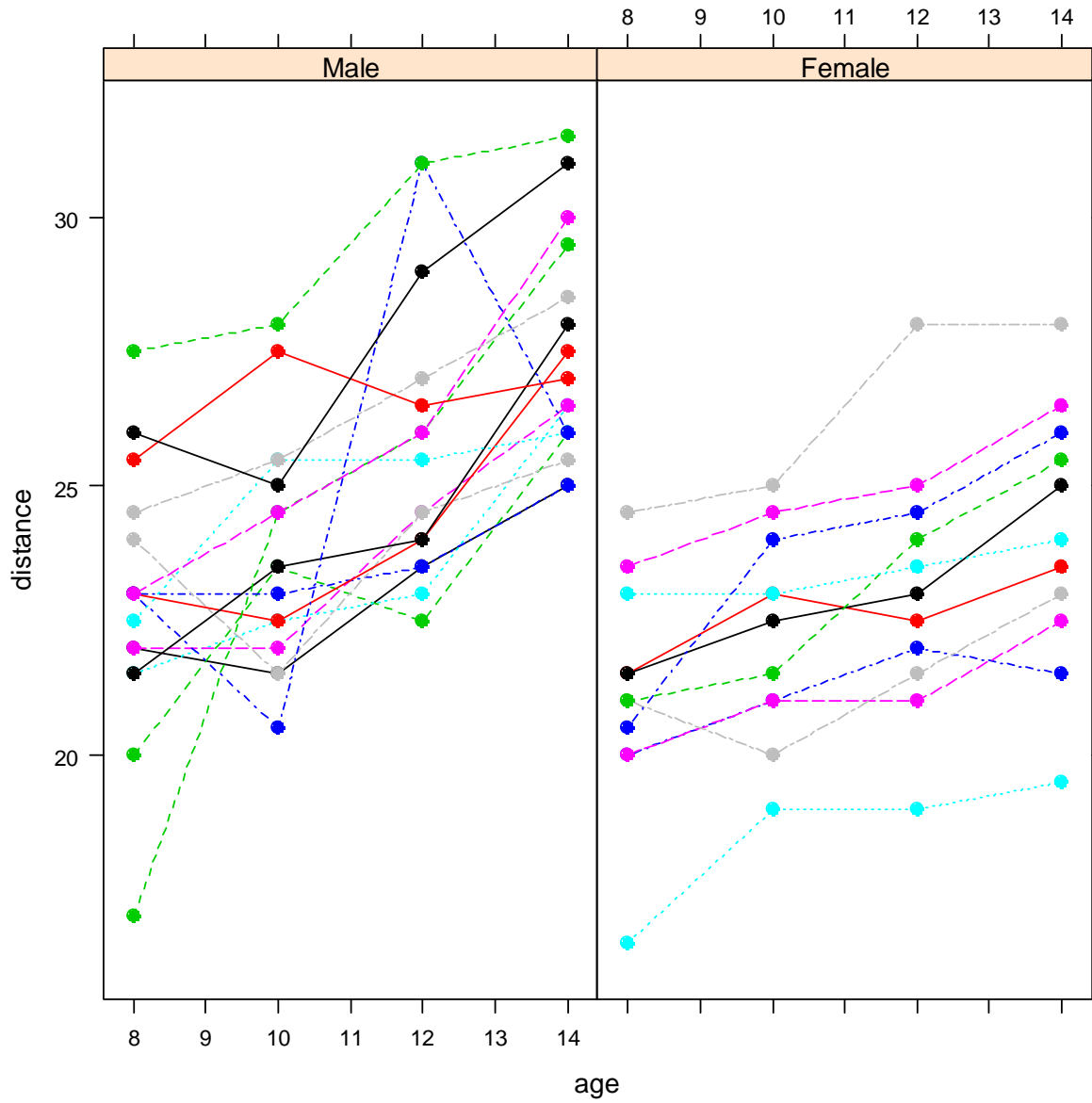
- everyone measured at the same set of ages
- could use a classical repeated measures analysis

Some terminology:

**Cluster:** the set of observations on one subject

**Occasion:** observations at a given time for each subject

Figure 1: Pothoff and Roy dental measurements in boys and girls.



**Figure 2: A different view by sex**

## One wrong analysis

Ordinary least-squares:

$$y_{it} = \beta_0 + \beta_{age} age_{it} + \beta_{sex} sex_i + \beta_{age \times sex} age_{it} sex_i + \varepsilon_{it}$$

$i = 1, \dots, N$  (number of subjects [clusters])

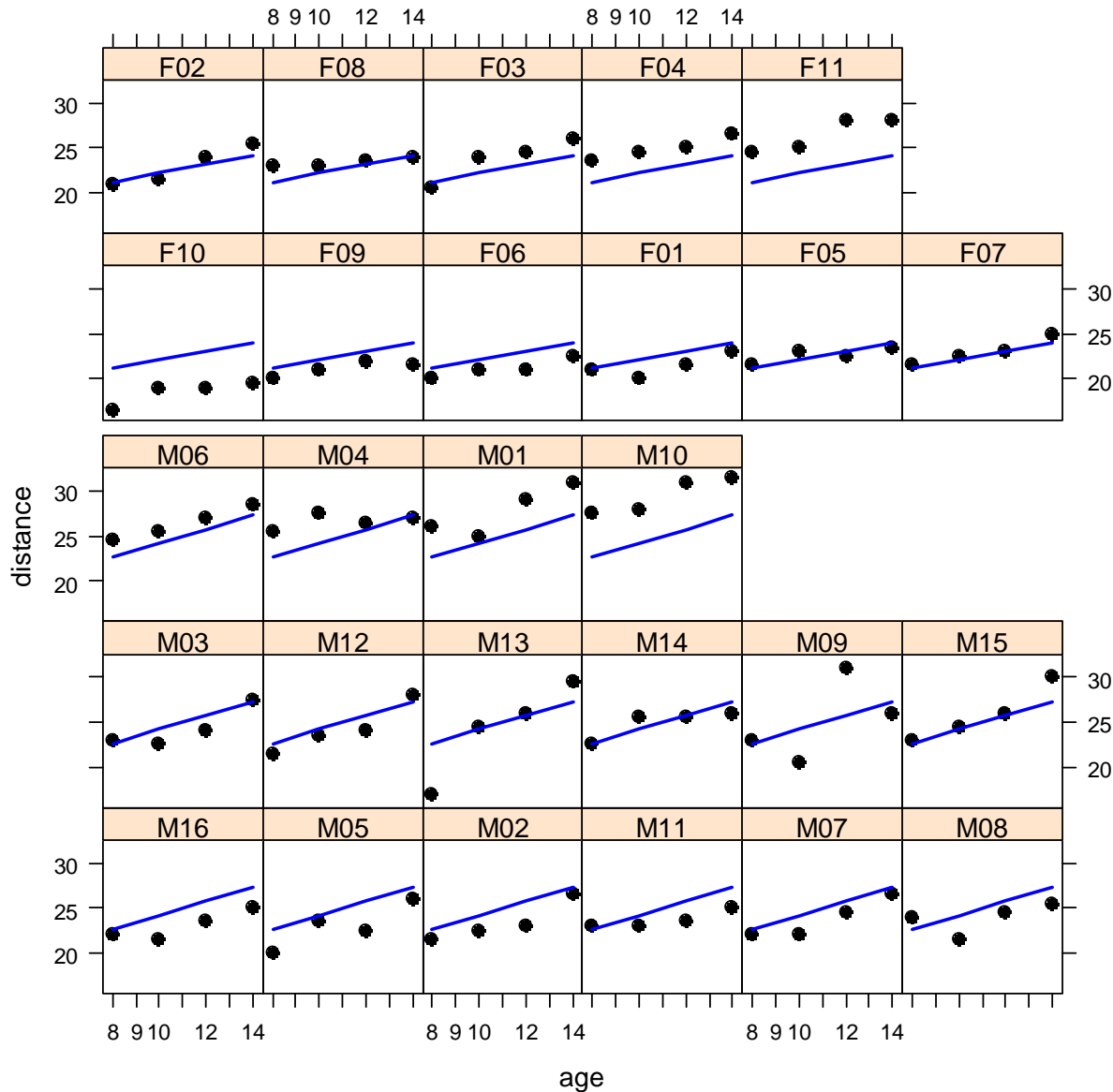
$t = 1, \dots, T_i$  (number of occasions for  $i$ th subject)

SAS:

```
PROC GLM DATA = ORTHO;  
  CLASS SEX;  
  MODEL DISTANCE = SEX AGE AGE*SEX / SOLUTION;
```

Parameter		Estimate		Standard Error	t Value	Pr >  t
Intercept		16.34062500	B	1.41622419	11.54	<.0001
Sex	Female	1.03210227	B	2.21879688	0.47	0.6428
Sex	Male	0.00000000	B	.	.	.
age		0.78437500	B	0.12616728	6.22	<.0001
<b>age*Sex</b>	<b>Female</b>	<b>-0.30482955</b>	<b>B</b>	<b>0.19766614</b>	<b>-1.54</b>	<b>0.1261</b>
age*Sex	Male	0.00000000	B	.	.	.



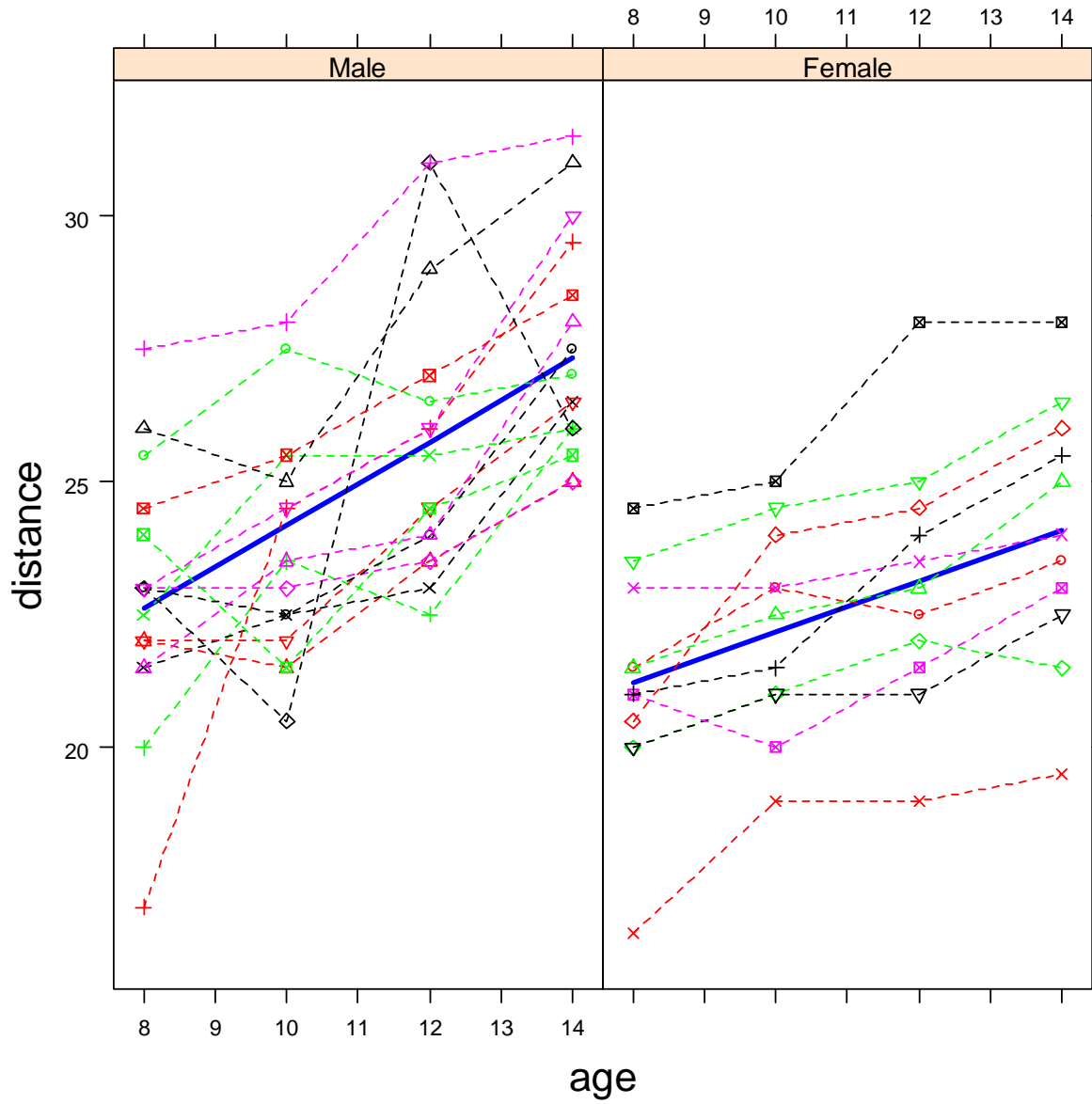


Estimated variance within each subject:

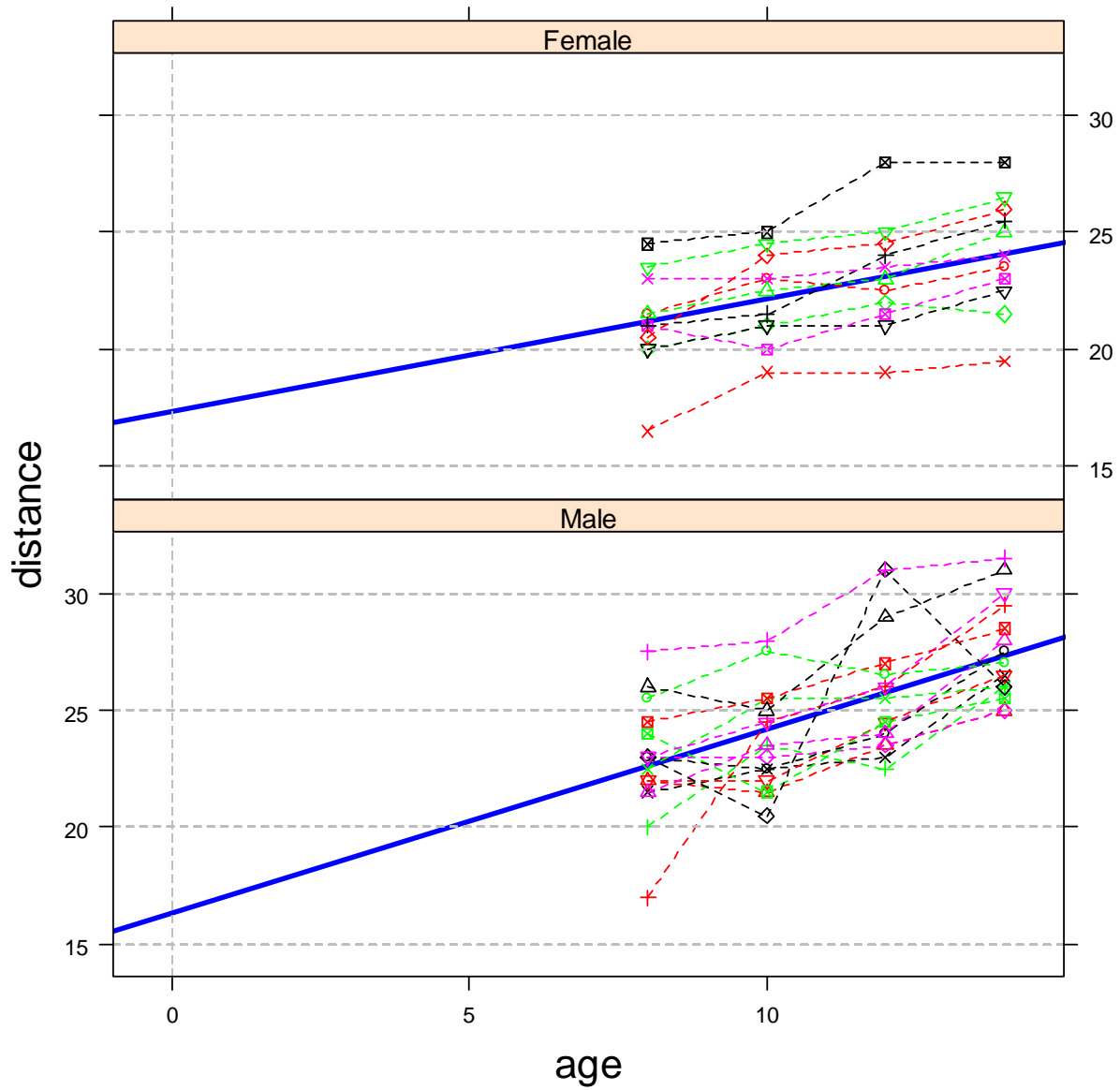
$$\begin{bmatrix} 2.34^2 & \cdot & \cdot & \cdot \\ \cdot & 2.34^2 & \cdot & \cdot \\ \cdot & \cdot & 2.34^2 & \cdot \\ \cdot & \cdot & \cdot & 2.34^2 \end{bmatrix}$$

Why is this wrong?

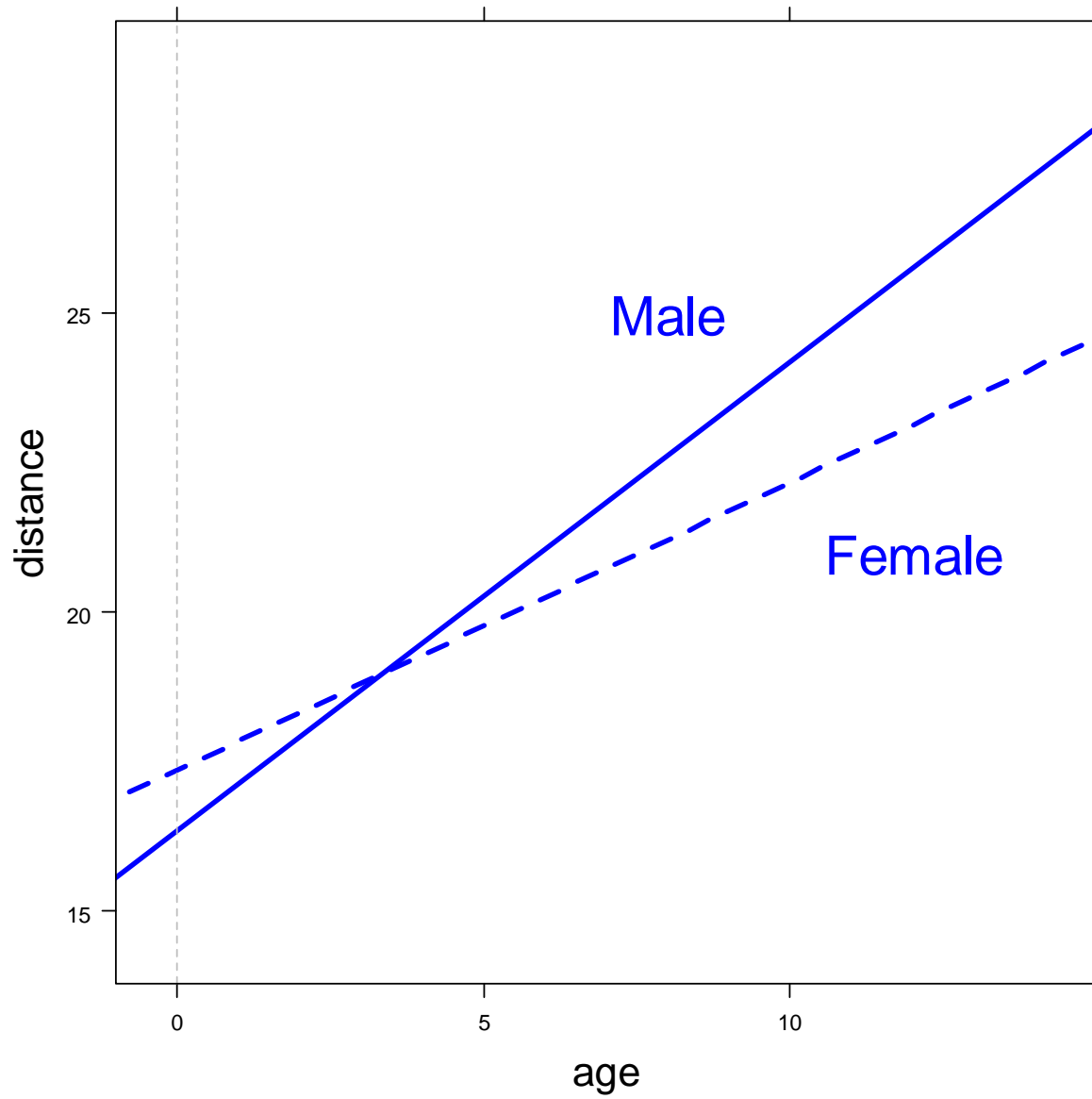
- Residuals within clusters are not independent; they tend to be highly correlated with each other



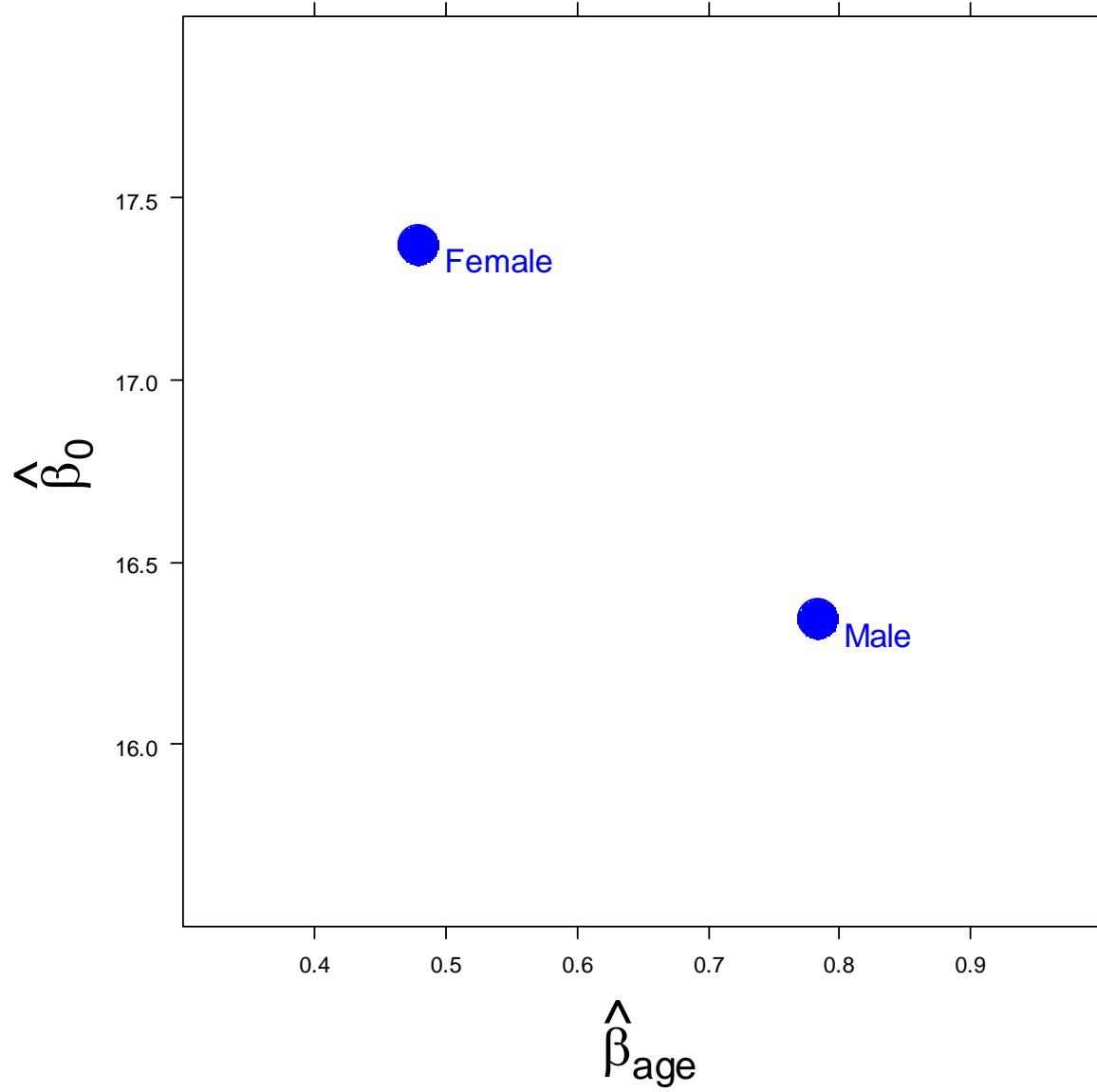
Fitted lines in 'data space'



Determining the intercept and slope of each line



Fitted lines in  
'data' space



Fitted 'lines' in  
'beta' space

## ***Fixed effects regression model***

See

Paul D. Allison (2005) *Fixed Effects Regression Methods for Longitudinal Data Using SAS*. SAS Institute – a great book on basics of **mixed models!**

- Treat Subject as a factor
- Lose Sex unless it is constructed as a Subject contrast
- Fits a separate OLS model to each subject:

$$y_{it} = \psi_{i0} + \psi_{i\text{age}} \text{age}_{it} + \varepsilon_{it}$$

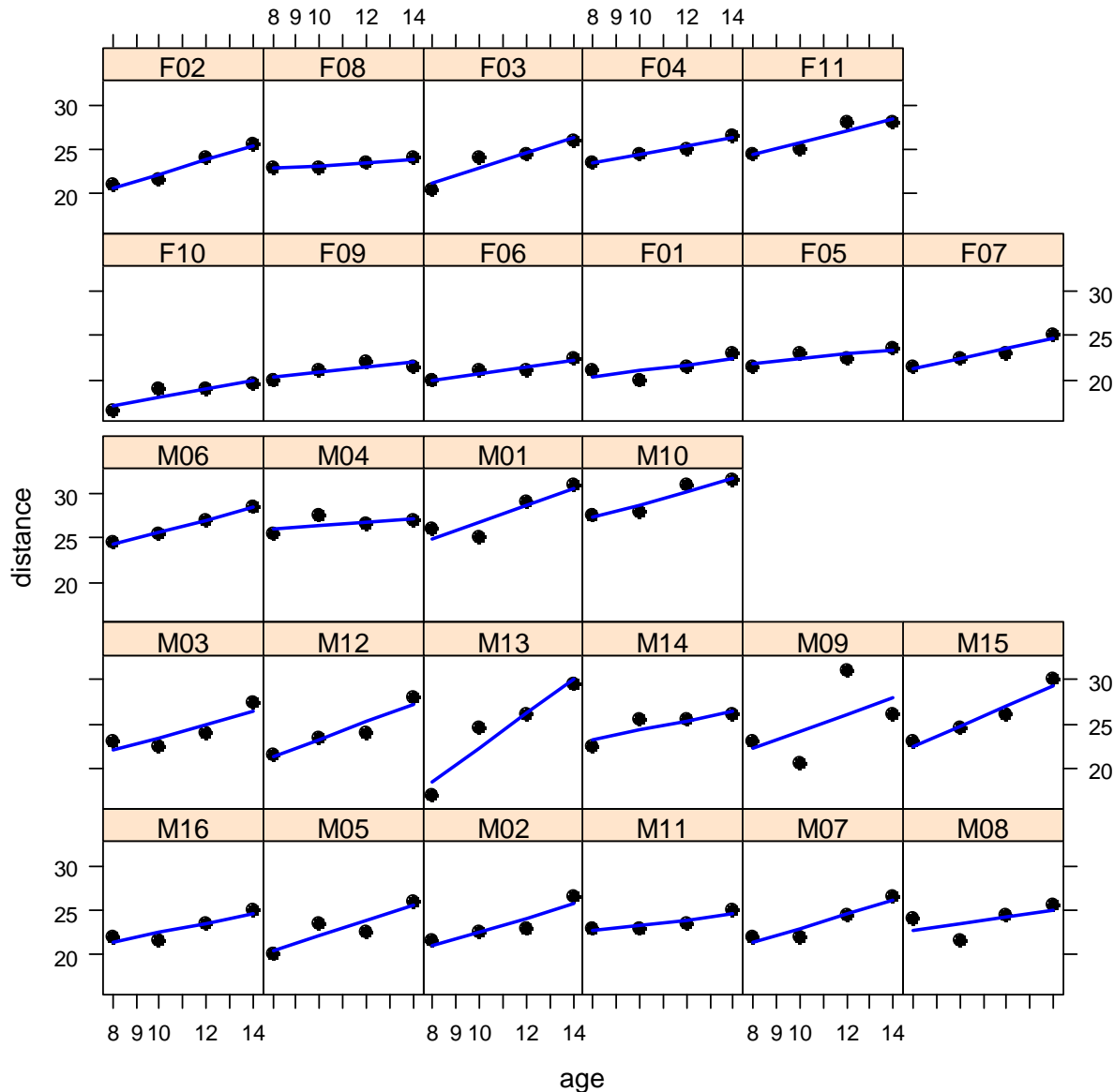
SAS:

```
PROC GLM DATA = ORTHO ;  
  CLASS SUBJECT ;  
  MODEL DISTANCE = SUBJECT AGE SUBJECT*AGE / ESTIMATE ;
```

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Subject	26	518.3796296	19.9376781	11.62	<.0001
age	1	235.3560185	235.3560185	137.14	<.0001
age*Subject	26	71.2814815	2.7415954	1.60	0.0735

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Subject	26	66.9693122	2.5757428	1.50	0.1040
age	1	235.3560185	235.3560185	137.14	<.0001
age*Subject	26	71.2814815	2.7415954	1.60	0.0735

Parameter	Estimate	Standard Error	t Value	Pr >  t
Intercept	16.95000000 B	3.28817325	5.15	<.0001
Subject F01	0.30000000 B	4.65017921	0.06	0.9488
Subject F02	-2.75000000 B	4.65017921	-0.59	0.5567
Subject F03	-2.55000000 B	4.65017921	-0.55	0.5857
Subject F04	2.70000000 B	4.65017921	0.58	0.5639
Subject F05	2.65000000 B	4.65017921	0.57	0.5711
Subject F06	0.05000000 B	4.65017921	0.01	0.9915
Subject F07	0.00000000 B	4.65017921	0.00	1.0000
Subject F08	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
age*Subject M11	-0.22500000 B	0.41427089	-0.54	0.5893
age*Subject M12	0.45000000 B	0.41427089	1.09	0.2822
age*Subject M13	1.40000000 B	0.41427089	3.38	0.0014
age*Subject M14	-0.02500000 B	0.41427089	-0.06	0.9521
age*Subject M15	0.57500000 B	0.41427089	1.39	0.1708
age*Subject M16	0.00000000 B	.	.	.



Estimated variance for each subject:

$$\begin{bmatrix} 1.31^2 & . & . & . \\ . & 1.31^2 & . & . \\ . & . & 1.31^2 & . \\ . & . & . & 1.31^2 \end{bmatrix}$$

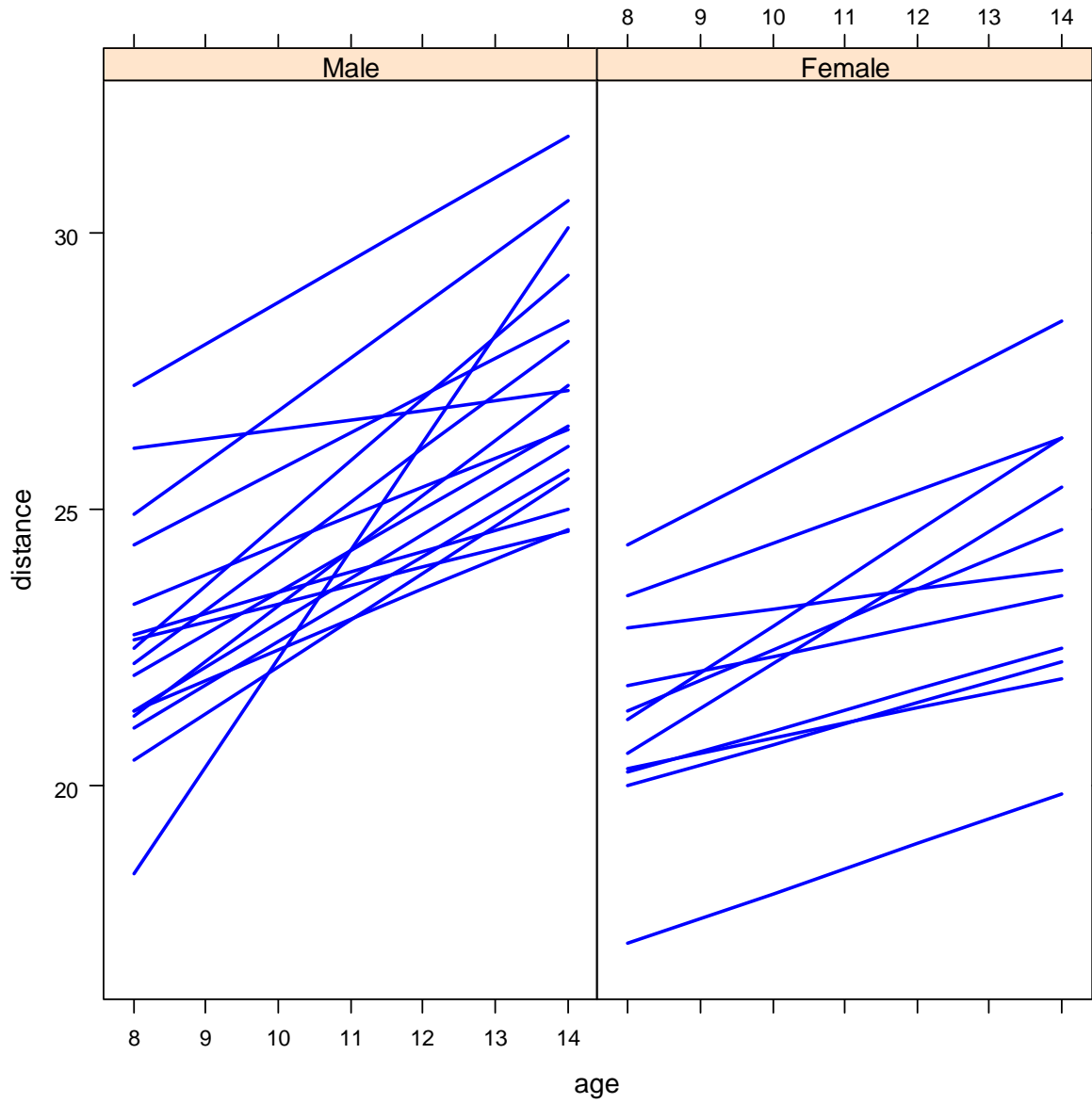
Problems:

No autocorrelation in time

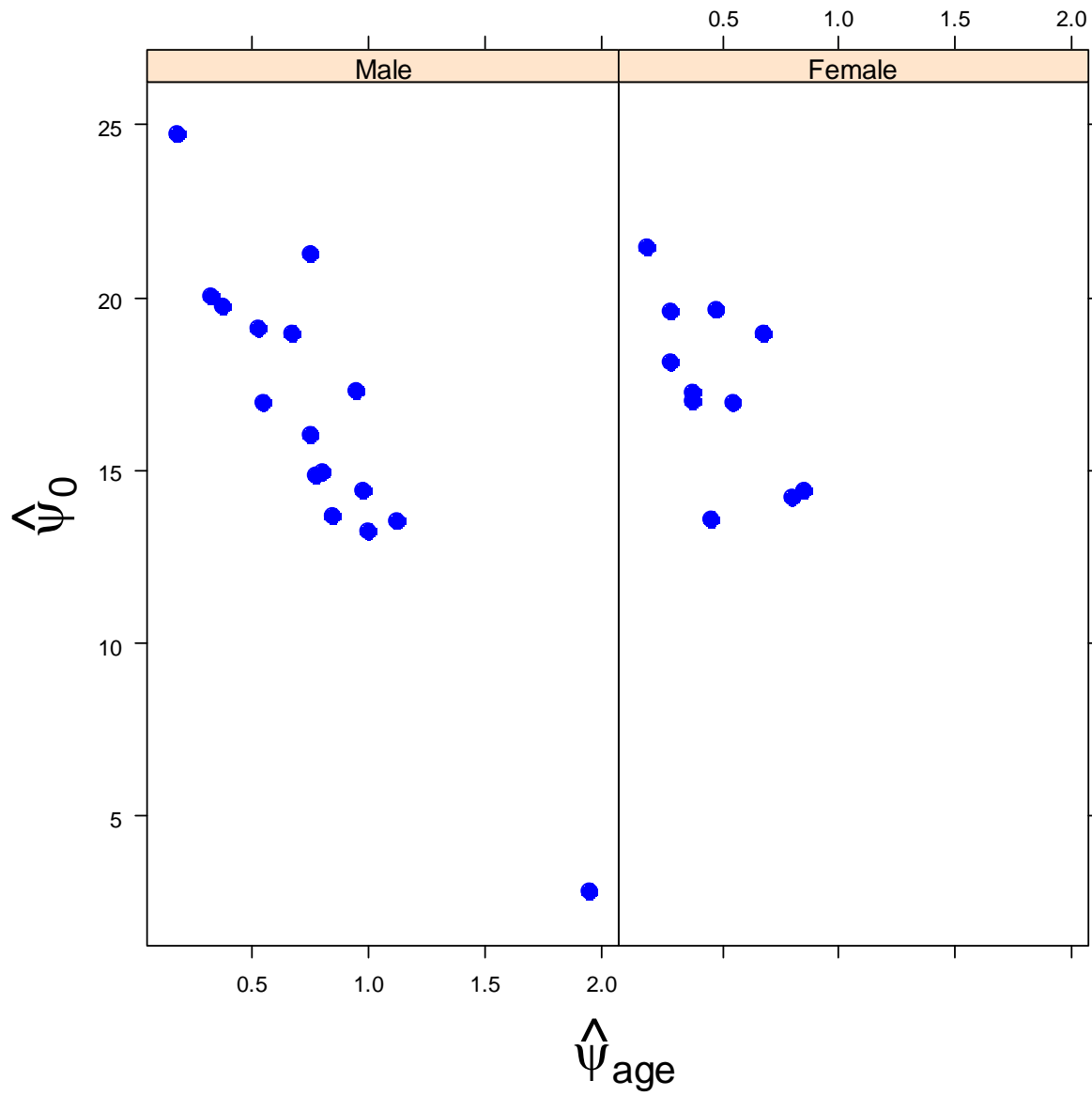
No estimate of sex effect

Can construct sex effect but CI is for difference in this sample, not for the difference in the population

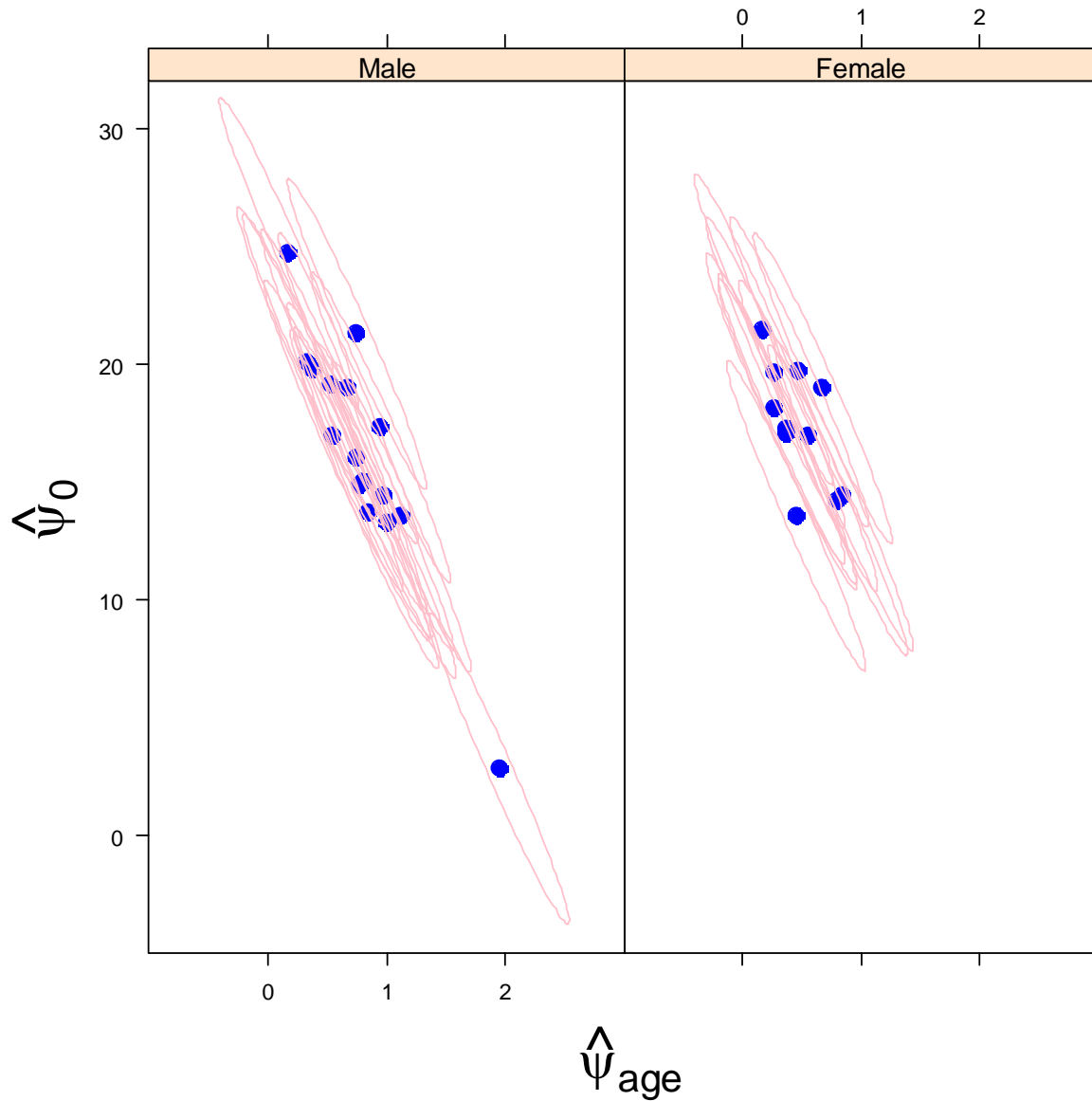




Fitted lines in data space



Fitted lines in beta space



Each within-subject least squares estimate

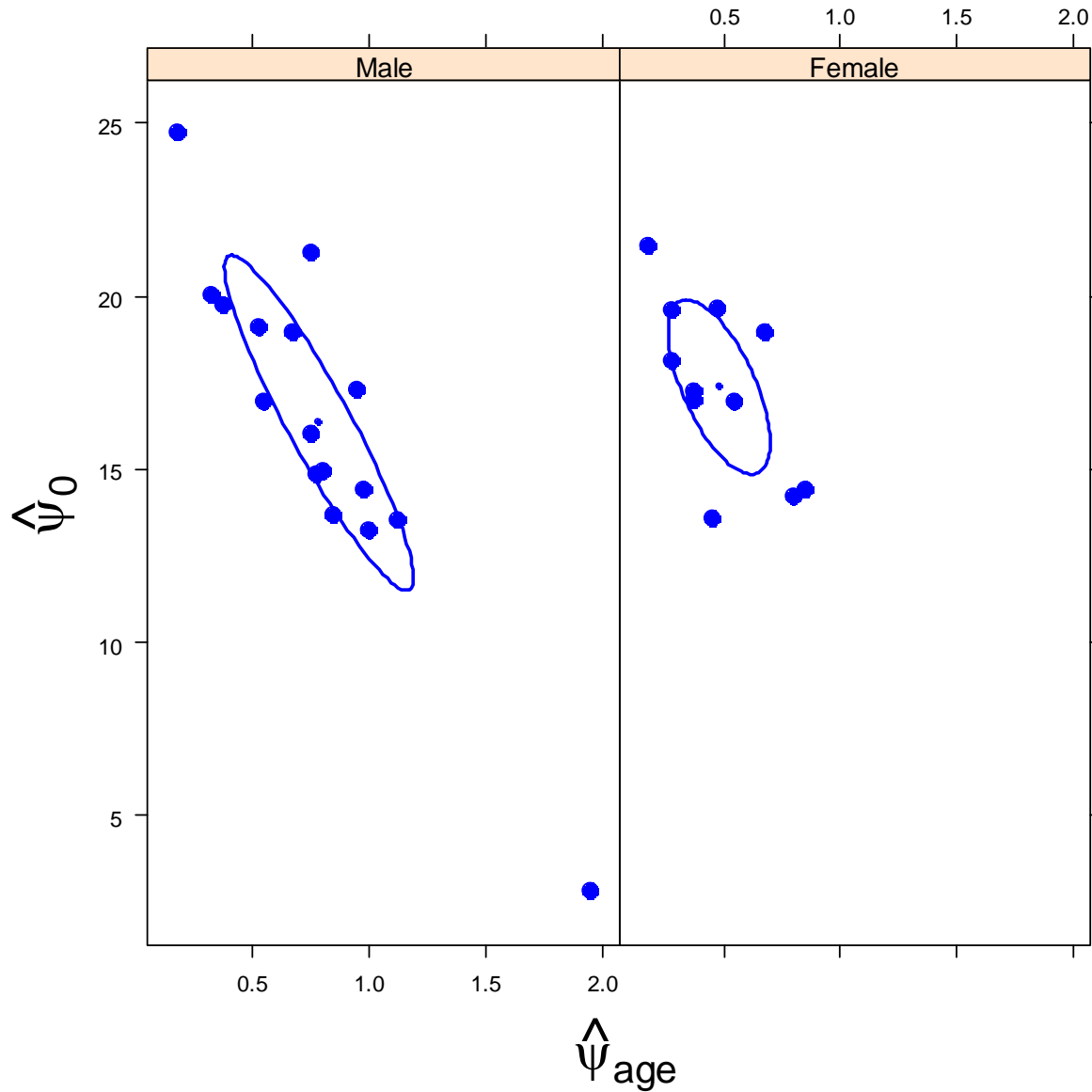
$$\hat{\psi}_i = \begin{bmatrix} \hat{\psi}_{i0} \\ \hat{\psi}_{i\text{age}} \end{bmatrix}$$

has variance  $\sigma^2 (\mathbf{X}_i' \mathbf{X}_i)^{-1}$   
which is used to construct  
a confidence ellipse for the  
'fixed effect'

$$\psi_i = \begin{bmatrix} \psi_{i0} \\ \psi_{i\text{age}} \end{bmatrix}$$

for the  $i$ th subject.

Each CI uses only the  
information from that subject  
(except for the estimate of  $\sigma^2$ )



The dispersion of  $\hat{\psi}_i$ s and the information they provide on the dispersion of  $\psi_i$ s is not used in the model.

The standard error of the estimate of each average Sex line uses the sample distribution of  $\varepsilon_{it}$ s but not the variability in  $\psi_i$ s.

## ***Other approaches***

- Repeated measures (univariate and multivariate)
  - Need same times for each subject, no other time-varying variables
- Two-stage approach: use  $\hat{\psi}_i$ s in second level analysis:
  - If design not balanced, then  $\hat{\psi}_i$ s have different variances, and would need different weights, Using  $\sigma^2(\mathbf{X}_i'\mathbf{X}_i)^{-1}$  does not work because the relevant weight is based on the marginal variance, not the conditional variance given the *i*th subject.

## **Multilevel Models**

Start with the fixed effects model:

**Within-subject model** (same as fixed effects model above):

$$y_{it} = \psi_{i0} + \psi_{i1}X_{it} + \varepsilon_{it} \quad \varepsilon_{i\cdot} \sim N(0, \sigma^2 I) \quad i = 1, \dots, N \quad t = 1, \dots, T_i$$

$\psi_{i0}$  is the ‘true’ intercept and  $\psi_{i1}$  is the ‘true’ slope with respect to  $X$ .

$\sigma^2$  is the within-subject residual variance.

$X$  (*age* in our example) is a time-varying variable. We could have more than one.

Then add:

**Between-subject model (new part):**

We suppose that  $\psi_{i0}$  and  $\psi_{i1}$  vary randomly from subject to subject.

But the distribution might be different for different Sexes (a ‘between-subject’ or ‘time-invariant’ variable). So we assume a multivariate distribution:

$$\begin{aligned}\psi_i = \begin{bmatrix} \psi_{i0} \\ \psi_{i1} \end{bmatrix} &= \begin{bmatrix} \beta_{00} + \beta_{01}W_i \\ \beta_{10} + \beta_{11}W_i \end{bmatrix} + \begin{bmatrix} \gamma_{i0} \\ \gamma_{i1} \end{bmatrix} \quad i = 1, \dots, N \\ &= \begin{bmatrix} \beta_{00} & \beta_{01} \\ \beta_{10} & \beta_{11} \end{bmatrix} \begin{bmatrix} 1 \\ W_i \end{bmatrix} + \begin{bmatrix} \gamma_{i0} \\ \gamma_{i1} \end{bmatrix}\end{aligned}$$

$$\begin{bmatrix} \gamma_{i0} \\ \gamma_{i1} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} g_{00} & g_{01} \\ g_{10} & g_{11} \end{bmatrix}\right) = N(\mathbf{0}, \mathbf{G})$$

where  $W_i$  is a coding variable for Sex, e.g. 0 for Males and 1 for Females.

$$\begin{aligned}E\left(\begin{bmatrix} \psi_{i0} \\ \psi_{i1} \end{bmatrix}\right) &= \begin{bmatrix} \beta_{00} \\ \beta_{10} \end{bmatrix} && \text{among Males} \\ &= \begin{bmatrix} \beta_{00} + \beta_{01} \\ \beta_{10} + \beta_{11} \end{bmatrix} && \text{among Females}\end{aligned}$$

Some software packages use the formulation of the multilevel model, e.g. MLWin. SAS and R use the ‘mixed model’ formulation. It is very useful to know how to go from one formulation to the other.

## ***From Multilevel Model to Mixed Model***

Combine the two levels of the multilevel model by substituting the between subject model into the within-subject model. Then gather together the fixed terms and the random terms:

$$\begin{aligned}y_{it} &= \psi_{i0} + \psi_{i1}X_{it} + \varepsilon_{it} \\&= (\beta_{00} + \beta_{01}W_i + \gamma_{i0}) + (\beta_{10} + \beta_{11}W_i + \gamma_{i1})X_{it} + \varepsilon_{it} \\&= (\beta_{00} + \beta_{01}W_i) + (\beta_{10} + \beta_{11}W_i)X_{it} + \gamma_{i0} + \gamma_{i1}X_{it} + \varepsilon_{it} \\&= \beta_{00} + \beta_{01}W_i + \beta_{10}X_{it} + \beta_{11}W_iX_{it} \quad (\text{fixed part of the model}) \\&\quad + \gamma_{i0} + \gamma_{i1}X_{it} + \varepsilon_{it} \quad (\text{random part of the model})\end{aligned}$$

Anatomy of the fixed part:

$$\begin{aligned}\beta_{00} &\quad (\text{Intercept}) \\+ \beta_{01}W_i &\quad (\text{between-subject, time-invariant variable}) \\+ \beta_{10}X_{it} &\quad (\text{within-subject, time-varying variable}) \\+ \beta_{11}W_iX_{it} &\quad (\text{cross-level interaction})\end{aligned}$$

Interpretation of the fixed part: the parameters reflect population average values.



Anatomy of the random part:

For one occasion:

$$\delta_{it} = \gamma_{i0} + \gamma_{i1} X_{it} + \varepsilon_{it}$$

Putting the observations of one subject together:

$$\begin{bmatrix} \delta_{i1} \\ \delta_{i2} \\ \delta_{i3} \\ \delta_{i4} \end{bmatrix} = \begin{bmatrix} 1 & X_{i1} \\ 1 & X_{i2} \\ 1 & X_{i3} \\ 1 & X_{i4} \end{bmatrix} \begin{bmatrix} \gamma_{i0} \\ \gamma_{i1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \end{bmatrix}$$

$$\delta_{i\cdot} = \mathbf{Z}_i \gamma_{i\cdot} + \varepsilon_{i\cdot}$$

Note: the random-effects design uses only time-varying variables

Distribution assumption:

$$\gamma_{i\cdot} \sim N(0, \mathbf{G}) \quad \text{independent of} \quad \varepsilon_{i\cdot} \sim N(0, \mathbf{R}_i) \quad \text{where, so far, } \mathbf{R}_i = \sigma^2 I$$

## Notes:

- $\mathbf{G}$  (usually) does not vary with  $i$ . It is usually a free positive definite matrix or it may be a structured pos-def matrix. More on  $\mathbf{G}$  later.
- $\mathbf{R}_i$  (usually) does change with  $i$  – as it must if  $T_i$  is not constant.  $\mathbf{R}_i$  is expressed as a function of parameters. The simplest example is  $\mathbf{R}_i = \sigma^2 I_{n_i \times n_i}$ . Later we will use  $\mathbf{R}_i$  to include auto-regressive parameters for longitudinal modeling.
- We can't estimate  $\mathbf{G}$  and  $\mathbf{R}$  directly. We estimate them through:

$$\mathbf{V}_i = \text{Var}(\delta_{i.}) = \mathbf{Z}_i \mathbf{G} \mathbf{Z}_i' + \mathbf{R}_i$$

- Some things can be parametrized either on the G-side or on the R-side. If they're done in both, you lose identifiability. Ill-conditioning due “collinearity” between the G- and R-side models is a common problem.

## ***Mixed Model in SAS:***

```
PROC SORT DATA = ortho;  
  BY subject age;  
PROC MIXED;  
  CLASS subject age;  
  MODEL distance = age sex sex*age;  
  RANDOM INTERCEPT age / TYPE = FA0(2) SUB = subject;  
  REPEATED / TYPE = AR(1) SUB = subject;
```

- SORT: the data set must be sorted on the SUBJECT variable – otherwise PROC MIXED will silently report nonsense. Many longitudinal analyses also require sorting on the time variable.
- MODEL statement:
  - specifies the fixed model
  - includes the INTERCEPT by default
  - contains time-varying, time-invariant and cross-level variables together

- RANDOM statement:
  - Specifies the variables in the random model
  - TYPE: Specifies the  $\mathbf{G}$  matrix. Most people use TYPE = UN (for ‘unstructured’) but then  $\mathbf{G}$  is not constrained to be non-negative definite.  $\text{FA0}(q)$ , where  $q$  is the size of the matrix, generates a free non-negative definite  $\mathbf{G}$  using the Choleski factor for parametrization.
  - SUB: name of the grouping variable

- REPEATED statement:

- Specifies the model for the  $\mathbf{R}_i$  matrices
- Omitted to get the default:  $\mathbf{R}_i = \sigma^2 \mathbf{I}_{n_i \times n_i}$
- Here we illustrate the use of an AR(1) structure producing for example

$$R_i = \sigma^2 \begin{bmatrix} 1 & \rho^1 & \rho^2 & \rho^3 \\ \rho^1 & 1 & \rho^1 & \rho^2 \\ \rho^2 & \rho^1 & 1 & \rho^1 \\ \rho^3 & \rho^2 & \rho^1 & 1 \end{bmatrix} \text{ in a cluster with 4 occasions.}$$

## Mixed Model in Matrices

In the  $i$ th cluster:

$$y_{it} = \beta_{00} + \beta_{01}W_i + \beta_{10}X_{it} + \beta_{11}W_iX_{it} + \gamma_{i0} + \gamma_{i1}X_{it} + \varepsilon_{it}$$

$$\begin{bmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \\ y_{i4} \end{bmatrix} = \begin{bmatrix} 1 & W_i & X_{i1} & W_i \\ 1 & W_i & X_{i2} & W_i \\ 1 & W_i & X_{i3} & W_i \\ 1 & W_i & X_{i4} & W_i \end{bmatrix} \begin{bmatrix} \beta_{00} \\ \beta_{01} \\ \beta_{10} \\ \beta_{11} \end{bmatrix} + \begin{bmatrix} 1 & X_{i1} \\ 1 & X_{i2} \\ 1 & X_{i3} \\ 1 & X_{i4} \end{bmatrix} \begin{bmatrix} \gamma_{i0} \\ \gamma_{i1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \end{bmatrix}$$

$$\mathbf{y}_{i\cdot} = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i.$$

[Could we fit this model in cluster  $i$ ?]

where

$$\boldsymbol{\gamma}_i \sim N(\mathbf{0}, \mathbf{G})$$

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

$$\boldsymbol{\delta}_i = \mathbf{Z}_i\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i.$$

$$\boldsymbol{\delta}_i \sim N(\mathbf{0}, \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i' + \mathbf{R}_i)$$

For the whole sample

$$\begin{bmatrix} \mathbf{y}_{1.} \\ \vdots \\ \mathbf{y}_{N.} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_N \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \mathbf{Z}_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{Z}_N \end{bmatrix} \begin{bmatrix} \boldsymbol{\gamma}_{1.} \\ \vdots \\ \boldsymbol{\gamma}_{N.} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_{1.} \\ \vdots \\ \boldsymbol{\varepsilon}_{N.} \end{bmatrix}$$

Finally making the complex look deceptively simple:

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \\ &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\delta} \end{aligned}$$

$$\text{Var}(\boldsymbol{\varepsilon}) = \mathbf{R} = \begin{bmatrix} \mathbf{R}_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{R}_N \end{bmatrix}$$

$$\text{Var}(\boldsymbol{\gamma}) = \mathbf{G}$$

$$\boldsymbol{\delta} = \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$$

$$\text{Var}(\boldsymbol{\delta}) = \mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}$$

## *Fitting the mixed model*

Use Generalized Least Squares on

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{ZGZ}' + \mathbf{R})$$

$$\hat{\boldsymbol{\beta}}^{GLS} = (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1} \mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{y}$$

We need  $\hat{\boldsymbol{\beta}}^{GLS}$  to get  $\hat{\mathbf{V}}$  and vice versa so one algorithm iterates from one to the other until convergence. **COULD SAY SOMETHING ABOUT ML AND REML**

We used OLS above:

$$\hat{\boldsymbol{\beta}}^{OLS} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$$

- How does OLS differ from GLS?
  - Do they differ only in that GLS produces more accurate standard errors?
  - Or can  $\hat{\boldsymbol{\beta}}^{OLS}$  be very different from  $\hat{\boldsymbol{\beta}}^{GLS}$ ?
  - With balanced data they will be the same. With unbalanced data they can be dramatically different.

## ***SAS code and output for Mixed Model***

### **Estimate statement**

One last detail: the ESTIMATE statement: same as GLM for fixed effects. Suppose we want to estimate and test the difference between Sexes at Age 14:

Set out values of SAS X matrix for Male at Age 14 and for Female at Age 14, then subtract:

Variable	INT	Age	Sex=F	Sex =M	Sex =Age*F	Sex =Age*M
Male at Age 14	1	14	0	1	0	14
Female at Age 14	1	14	1	0	14	0
Difference (F-M)	0	0	1	-1	14	-14

### **Sas code**

```
ODS HTML;  
ODS GRAPHICS ON; /* new for diagnostics in Version 9 */  
PROC SORT DATA = ORTHO;  
    BY subject age;  
PROC MIXED ASYCOV ASYCORR;  
    CLASS subject sex;
```



```

MODEL distance = age sex sex*age
           / S CORRB COVB COVBI CL;
RANDOM INTERCEPT age / TYPE = FA0(2) SUB = subject
      G GC GCORR V VCORR
      INFLUENCE
      INFLUENCE ( EFFECT = subject );
REPEATED / TYPE = AR(1) SUB = subject
      R RC RCI RCORR;
ESTIMATE `gap at 14'   sex 1 -1   age*sex 14   -14;
ESTIMATE `gap at 11.5' sex 1 -1   age*sex 11.5 -11.5;

```

## Output

### The Mixed Procedure

#### Model Information

Data Set	WORK.ORTHO
Dependent Variable	distance
Covariance Structures	Factor Analytic, Autoregressive
Subject Effects	Subject, Subject
Estimation Method	REML

Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

### Class Level Information

Class	Levels	Values
Subject	27	F01 F02 F03 F04 F05 F06 F07 F08 F09 F10 F11 M01 M02 M03 M04 M05 M06 M07 M08 M09 M10 M11 M12 M13 M14 M15 M16
Sex	2	Female Male

### Dimensions

Covariance Parameters	5
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

### Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	483.55911746	
1	2	428.86186900	0.00042400
2	1	428.80826312	0.00000516
3	1	428.80764501	0.00000000

Convergence criteria met.

### Estimated R Matrix for Subject F01

Row	Col1	Col2	Col3	Col4
1	1.1924	-0.5644	0.2671	-0.1265
2	-0.5644	1.1924	-0.5644	0.2671
3	0.2671	-0.5644	1.1924	-0.5644
4	-0.1265	0.2671	-0.5644	1.1924

Estimated R Correlation Matrix for Subject F01

Row	Col1	Col2	Col3	Col4
1	1.0000	-0.4733	0.2240	-0.1060
2	-0.4733	1.0000	-0.4733	0.2240
3	0.2240	-0.4733	1.0000	-0.4733
4	-0.1060	0.2240	-0.4733	1.0000

Estimated Chol(R) Matrix for Subject F01

Row	Col1	Col2	Col3	Col4
1	1.0920			
2	-0.5169	0.9619		
3	0.2447	-0.4553	0.9619	
4	-0.1158	0.2155	-0.4553	0.9619

Estimated InvChol(R) Matrix for Subject F01

Row	Col1	Col2	Col3	Col4
1	0.9158			
2	0.4921	1.0396		
3		0.4921	1.0396	
4			0.4921	1.0396

Estimated G Matrix

Row	Effect	Subject	Col1	Col2
1	Intercept	F01	11.3783	-0.8151
2	age	F01	-0.8151	0.08455

Estimated Chol(G) Matrix

Row	Effect	Subject	Col1	Col2
1	Intercept	F01	3.3732	
2	age	F01	-0.2416	0.1618

Estimated G Correlation Matrix

Row	Effect	Subject	Col1	Col2
1	Intercept	F01	1.0000	-0.8310
2	age	F01	-0.8310	1.0000

Estimated V Matrix for Subject F01

Row	Col1	Col2	Col3	Col4
1	4.9411	2.9071	3.4613	2.7905
2	2.9071	4.7248	3.0290	3.9214
3	3.4613	3.0290	5.1849	3.8273
4	2.7905	3.9214	3.8273	6.3214

Estimated V Correlation Matrix for Subject F01

Row	Col1	Col2	Col3	Col4
1	1.0000	0.6017	0.6839	0.4993
2	0.6017	1.0000	0.6120	0.7175
3	0.6839	0.6120	1.0000	0.6685
4	0.4993	0.7175	0.6685	1.0000

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
FA(1,1)	Subject	3.3732
FA(2,1)	Subject	-0.2416
FA(2,2)	Subject	0.1618
AR(1)	Subject	-0.4733
Residual		1.1924

Asymptotic Covariance Matrix of Estimates

Row	Cov Parm	CovP1	CovP2	CovP3	CovP4	CovP5
1	FA(1,1)	0.5338	-0.04666	0.001766	-0.04933	-0.04393
2	FA(2,1)	-0.04666	0.005763	-0.00036	0.006055	0.004865
3	FA(2,2)	0.001766	-0.00036	0.000649	-0.00097	-0.00059
4	AR(1)	-0.04933	0.006055	-0.00097	0.03500	0.009928
5	Residual	-0.04393	0.004865	-0.00059	0.009928	0.05547

Asymptotic Correlation Matrix of Estimates

Row	Cov Parm	CovP1	CovP2	CovP3	CovP4	CovP5
1	FA(1,1)	1.0000	-0.8412	0.09491	-0.3609	-0.2553
2	FA(2,1)	-0.8412	1.0000	-0.1875	0.4264	0.2721

3	FA(2,2)	0.09491	-0.1875	1.0000	-0.2043	-0.09846
4	AR(1)	-0.3609	0.4264	-0.2043	1.0000	0.2253
5	Residual	-0.2553	0.2721	-0.09846	0.2253	1.0000

### Fit Statistics

-2 Res Log Likelihood	428.8
AIC (smaller is better)	438.8
AICC (smaller is better)	439.4
BIC (smaller is better)	445.3

### Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
4	54.75	<.0001

### Solution for Fixed Effects

Effect	Sex	Estimate	Standard Error	DF	t Value	Pr >  t	Alpha	Lower	Upper
Intercept		16.1524	0.9985	25	16.18	<.0001	0.05	14.0960	18.2088
age		0.7980	0.08707	25	9.16	<.0001	0.05	0.6186	0.9773
Sex	Female	1.2647	1.5643	54	0.81	0.4224	0.05	-1.8715	4.4010
Sex	Male	0	.	.	.	.	.	.	.
age*Sex	Female	-0.3222	0.1364	54	-2.36	0.0218	0.05	-0.5957	-0.04876
age*Sex	Male	0	.	.	.	.	.	.	.

### Covariance Matrix for Fixed Effects

Row	Effect	Sex	Col1	Col2	Col3	Col4	Col5	Col6
1	Intercept		0.9969	-0.07620	-0.9969		0.07620	

2	age		-0.07620	0.007581	0.07620	-0.00758
3	Sex	Female	-0.9969	0.07620	2.4470	-0.1870
4	Sex	Male				
5	age*Sex	Female	0.07620	-0.00758	-0.1870	0.01861
6	age*Sex	Male				

Correlation Matrix for Fixed Effects

Row	Effect	Sex	Col1	Col2	Col3	Col4	Col5	Col6
1	Intercept		1.0000	-0.8765	-0.6383		0.5595	
2	age		-0.8765	1.0000	0.5595		-0.6383	
3	Sex	Female	-0.6383	0.5595	1.0000		-0.8765	
4	Sex	Male				1.0000		
5	age*Sex	Female	0.5595	-0.6383	-0.8765		1.0000	
6	age*Sex	Male						1.0000

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
age	1	25	87.18	<.0001
Sex	1	54	0.65	0.4224
age*Sex	1	54	5.58	0.0218

Estimates  
Standard

Label	Estimate	Error	DF	t Value	Pr >  t
gap at 14	-3.2467	0.9258	54	-3.51	0.0009



gap at 11.5      -2.4411      0.7785      54      -3.14      0.0028

## ***Diagnostics***

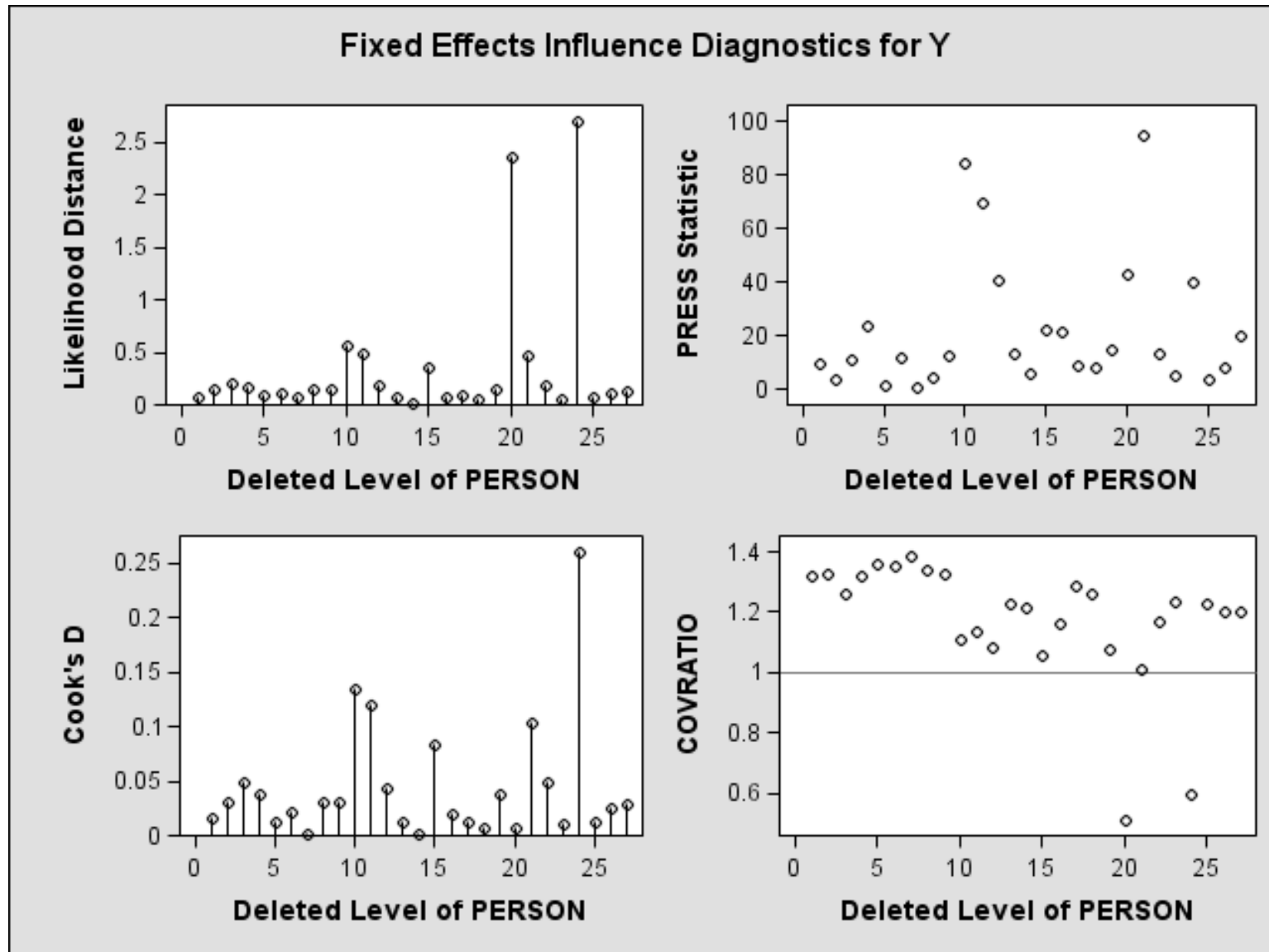
Regression diagnostics are available in SAS version 9. You must specify:

```
ODS HTML ;
```

```
ODS GRAPHICS ON ;
```

before invoking PROC MIXED and you must use the INFLUENCE and/or the RESIDUAL options in the MODEL statement. The INFLUENCE statement can be used to see the effect of dropping an occasion or an entire subject. To drop occasions, use INFLUENCE alone, to drop subjects use INFLUENCE ( EFFECT = subject-variable ). Often it will be desirable. As far as I can see, to get diagnostics for both, you need to run PROC MIXED twice. The following output shows diagnostics for deleting entire subjects. The influential case most clearly identified is subject 24 who has the lowest starting value and the steepest growth. Subject 20 with zigzagging growth is influential but not, generally, as much. Try rerunning the code with the following options for INFLUENCE ( EFFECT = SUBJECT ITER = 3 ). This will also produce output showing influence on the covariance structure.

An example of a diagnostic plot:



## Modeling dependencies in time

The main difference between using mixed models for multilevel modeling as opposed to longitudinal modeling are the assumptions about  $\varepsilon_i$ . For observations observed in time, part of the correlation between  $\varepsilon$ s should be related to their distance in time.

R-side model allows the modeling of temporal and spatial dependence.

REPEATED statement: TYPE option	R
AR(1)	$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$
ARMA(1,1)	$\sigma^2 \begin{bmatrix} 1 & \gamma & \gamma\rho & \gamma\rho^2 \\ \gamma & 1 & \gamma & \gamma\rho \\ \gamma\rho & \gamma & 1 & \gamma \\ \gamma\rho^2 & \gamma\rho & \gamma & 1 \end{bmatrix}$
SP(POW)(time) AR(1) in continuous time e.g. supposing a subject with times 1,2, 5.5 and 10 <sup>1</sup>	$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^{4.5} & \rho^9 \\ \rho & 1 & \rho^{3.5} & \rho^8 \\ \rho^{4.5} & \rho^{3.5} & 1 & \rho^{4.5} \\ \rho^9 & \rho^8 & \rho^{4.5} & 1 \end{bmatrix}$

<sup>1</sup> Note that the times and the number of times – hence the indices – can change from subject to subject but  $\sigma^2$  and  $\rho$  have the same value.

## ***G-side vs. R-side***

- A few things can be done with either side. But don't do it with both in the same model. The redundant parameters will not be identifiable. For example, the G-side random intercept model is 'almost' equivalent to the R-side compound symmetry model.
- With OLS the linear parameters are orthogonal to the variance parameter. Collinearity among the linear parameters is determined by the design,  $\mathbf{X}$ , and does not depend on values of parameters. Computational problems due to collinearity can be addressed by orthogonalizing the  $\mathbf{X}$  matrix.
- With mixed models the variance parameters are generally not orthogonal to each other and, with unbalanced data, the linear parameters are not orthogonal to the variance parameters.
- G-side parameters can be highly collinear even if the  $\mathbf{X}$  matrix is orthogonal. Centering the matrix around the "point of minimal variance" will help but the resulting design matrix may be highly collinear.
- G-side and R-side parameters can be highly collinear. The degree of collinearity may depend on the value of the parameters.

- For example, our model identifies  $\rho$  through:

$$\hat{\mathbf{V}} = \begin{bmatrix} 1 & -3 \\ 1 & -1 \\ 1 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} g_{00} & g_{01} \\ g_{10} & g_{11} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ -3 & -1 & 1 & 3 \end{bmatrix} + \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

For values of  $\rho$  above 0.5, the Hessian is very ill-conditioned. The lesson may be that to use AR, ARMA models effectively, you need some subjects observed on many occasions.

- R-side only: population average models
- G-side only: hierarchical models with conditionally independent observations in each cluster
- Population average longitudinal models can be done on the R-side with AR, ARMA structures, etc.
- The absence of the G-side may be less crucial with balanced data.

- The G-side is not enough to provide control for unmeasured between subject confounders if the time-varying predictors are unbalanced (more on this soon).
- A G-side random effects model DOES NOT provide the equivalent of temporal correlation.

## ***Simpler Models***

The model we've looked at is deliberately complex including examples of the main typical components of a mixed model. We can use mixed models for simpler problems.

Using  $X$  as a generic time-varying (within-subject) predictor and  $W$  as a generic time-invariant (between-subject) predictor we have the following:

	<b><i>MODEL</i></b>	<b><i>RANDOM</i></b>	<b><i>Formula</i></b>
One-way ANOVA with random effects	INT	INT	$y_{it} = \beta_{00} + \gamma_{i0} + \epsilon_{it}$
Means as outcomes	INT W	INT	$y_{it} = \beta_{00} + \beta_{01}W_i + \gamma_{i0} + \epsilon_{it}$
One-way ANCOVA with random effects	INT X	INT	$y_{it} = \beta_{00} + \beta_{10}X_{it} + \gamma_{i0} + \epsilon_{it}$
Random coefficients model	INT X	INT X	$y_{it} = \beta_{00} + \beta_{10}X_{it} + \gamma_{i0} + \gamma_{i1}X_{it} + \epsilon_{it}$
Intercepts and slopes as outcomes	INT X W X*W	INT X	$y_{it} = \beta_{00} + \beta_{01}W_i + \beta_{10}X_{it} + \beta_{11}W_iX_{it} + \gamma_{i0} + \gamma_{i1}X_{it} + \epsilon_{it}$
Non- random slopes	INT X W X*W	INT	$y_{it} = \beta_{00} + \beta_{01}W_i + \beta_{10}X_{it} + \beta_{11}W_iX_{it} + \gamma_{i0} + \epsilon_{it}$

## **BLUPS: Estimating Within-Subject Effects**

We've seen how to estimate  $\beta$ ,  $\mathbf{G}$  and  $\mathbf{R}$ . Now we consider  $\psi_i = \begin{bmatrix} \psi_{i0} \\ \psi_{i1} \end{bmatrix}$ .

We've already estimated  $\psi_i$  using the fixed-effects model with a OLS regression within each subject. Call this estimator:  $\hat{\psi}_i$ . How good is it?

$$\text{Var}(\hat{\psi}_i - \psi_i) = \sigma^2 (\mathbf{X}_i' \mathbf{X}_i)^{-1}$$

Can we do better? We have another 'estimator' of  $\psi_i$ .

Suppose we know  $\beta$ s. We could also **predict**<sup>2</sup>  $\psi_i$  by using the within Sex mean intercepts and slopes, e.g. for Males we could use:  $\begin{bmatrix} \beta_{00} \\ \beta_{10} \end{bmatrix}$  with error variance:

$$\text{Var} \left( \begin{bmatrix} \beta_{00} \\ \beta_{10} \end{bmatrix} - \begin{bmatrix} \psi_{i0} \\ \psi_{i1} \end{bmatrix} \right) = \mathbf{G}$$

---

<sup>2</sup> Non-statisticians are always thrown for a loop when we 'predict' something that happened in the past. We know what we mean.



We could then combine  $\hat{\psi}_i$  and  $\begin{bmatrix} \beta_{00} \\ \beta_{10} \end{bmatrix}$  by weighting them by inverse variance. This yields the BLUP (Best Linear Unbiased Predictor):

$$\left\{ \mathbf{G}^{-1} + \left[ \sigma^2 (\mathbf{X}_i' \mathbf{X}_i)^{-1} \right]^{-1} \right\}^{-1} \left\{ \mathbf{G}^{-1} \begin{bmatrix} \beta_{00} \\ \beta_{10} \end{bmatrix} + \left[ \sigma^2 (\mathbf{X}_i' \mathbf{X}_i)^{-1} \right]^{-1} \hat{\psi}_i \right\}$$

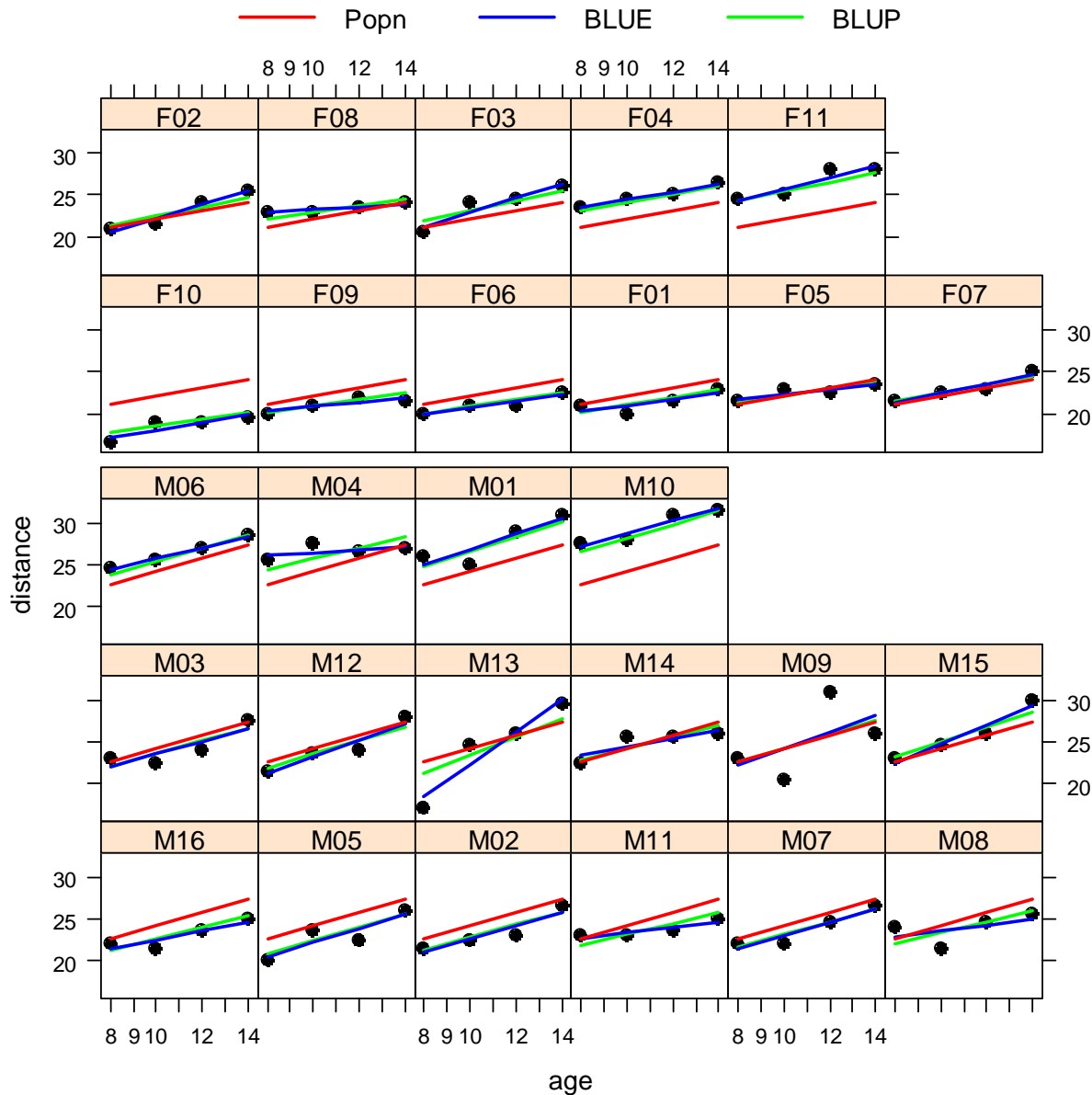
If we replace the unknown parameters with their estimates, we get the EBLUP (Empirical BLUP):

$$\tilde{\psi}_i = \left\{ \hat{\mathbf{G}}^{-1} + \left[ \hat{\sigma}^2 (\mathbf{X}_i' \mathbf{X}_i)^{-1} \right]^{-1} \right\}^{-1} \left\{ \hat{\mathbf{G}}^{-1} \begin{bmatrix} \hat{\beta}_{00} \\ \hat{\beta}_{10} \end{bmatrix} + \left[ \hat{\sigma}^2 (\mathbf{X}_i' \mathbf{X}_i)^{-1} \right]^{-1} \hat{\psi}_i \right\}$$

The EBLUP ‘optimally’ combines the information from the  $i$ th cluster with the information from the other clusters. We **borrow strength** from the other clusters.

The process ‘shrinks’  $\hat{\psi}_i$  towards  $\begin{bmatrix} \hat{\beta}_{00} \\ \hat{\beta}_{10} \end{bmatrix}$  along a path determined by the locus of osculation

of the families of ellipses with shape  $\hat{\mathbf{G}}$  around  $\begin{bmatrix} \hat{\beta}_{00} \\ \hat{\beta}_{10} \end{bmatrix}$  and shape  $\left[ \hat{\sigma}^2 (\mathbf{X}_i' \mathbf{X}_i)^{-1} \right]$  around  $\hat{\psi}_i$ .

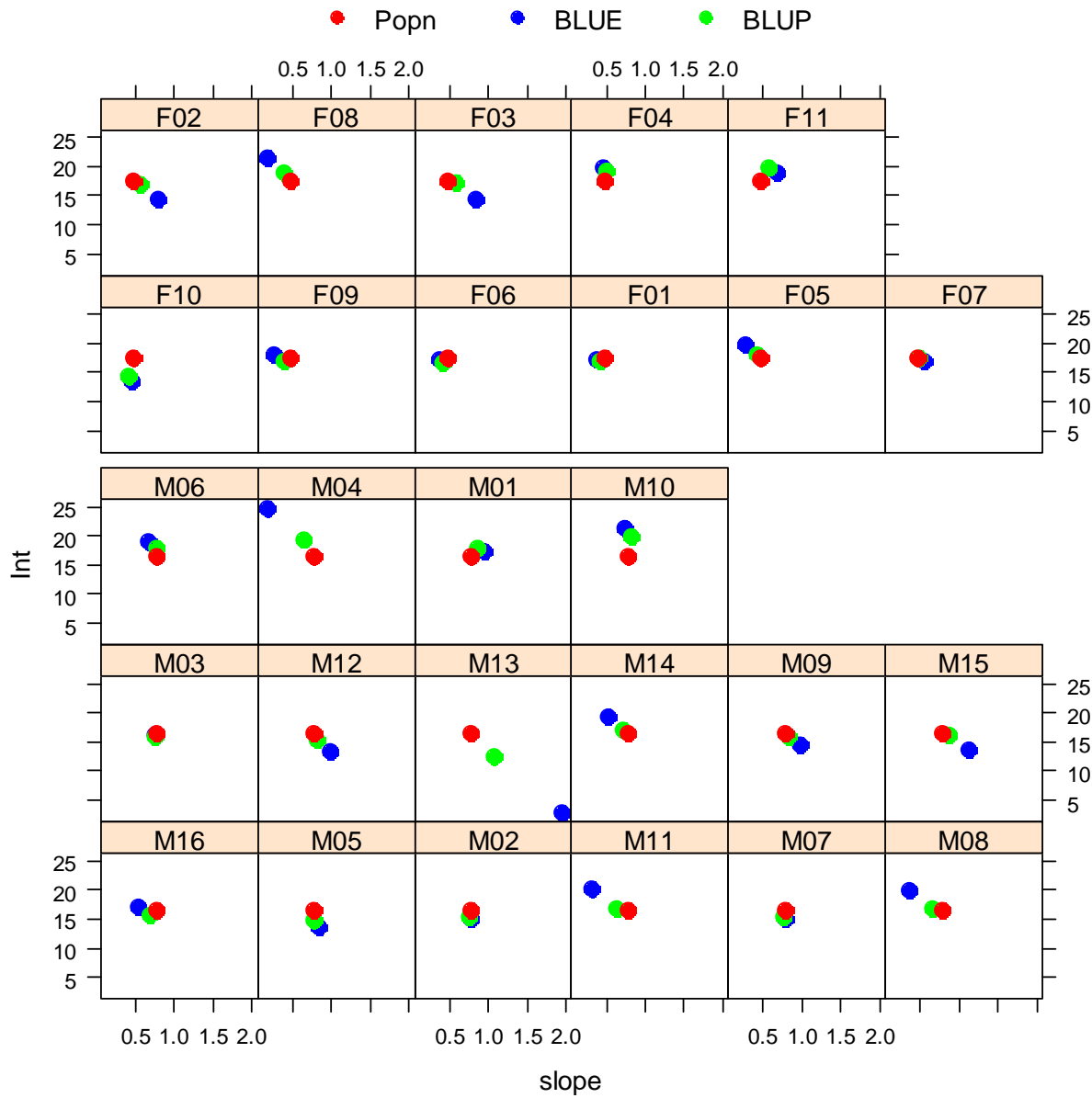


The slope of the BLUP is close to the population slope

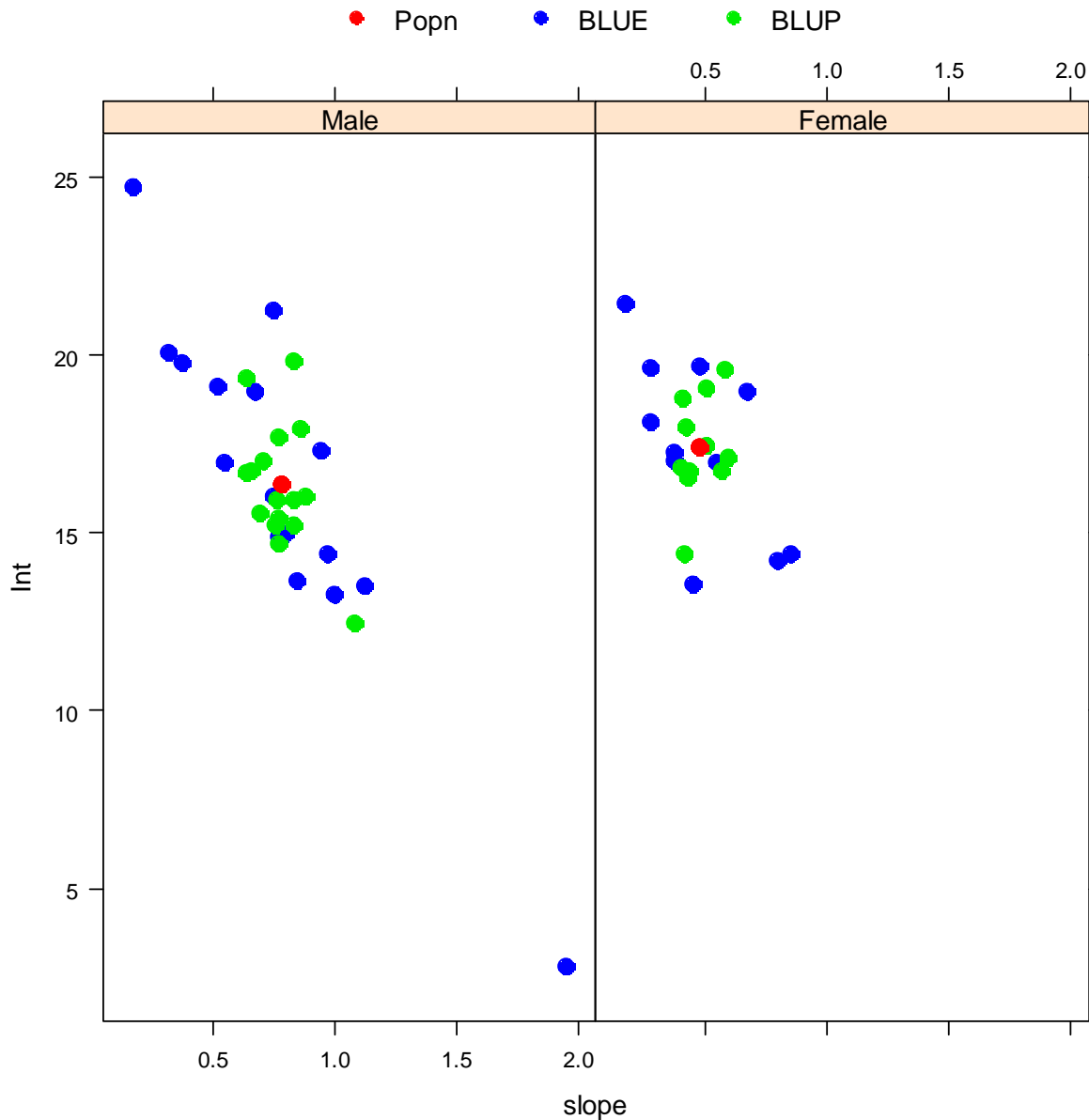
but

the level of the BLUP is close to the level of the BLUE

This suggests that  $\mathbf{G}$  has a large variance for intercepts and a small variance for slopes



Population estimate  
 BLUE and BLUP  
 in beta space



The marginal dispersion of BLUEs comes from:

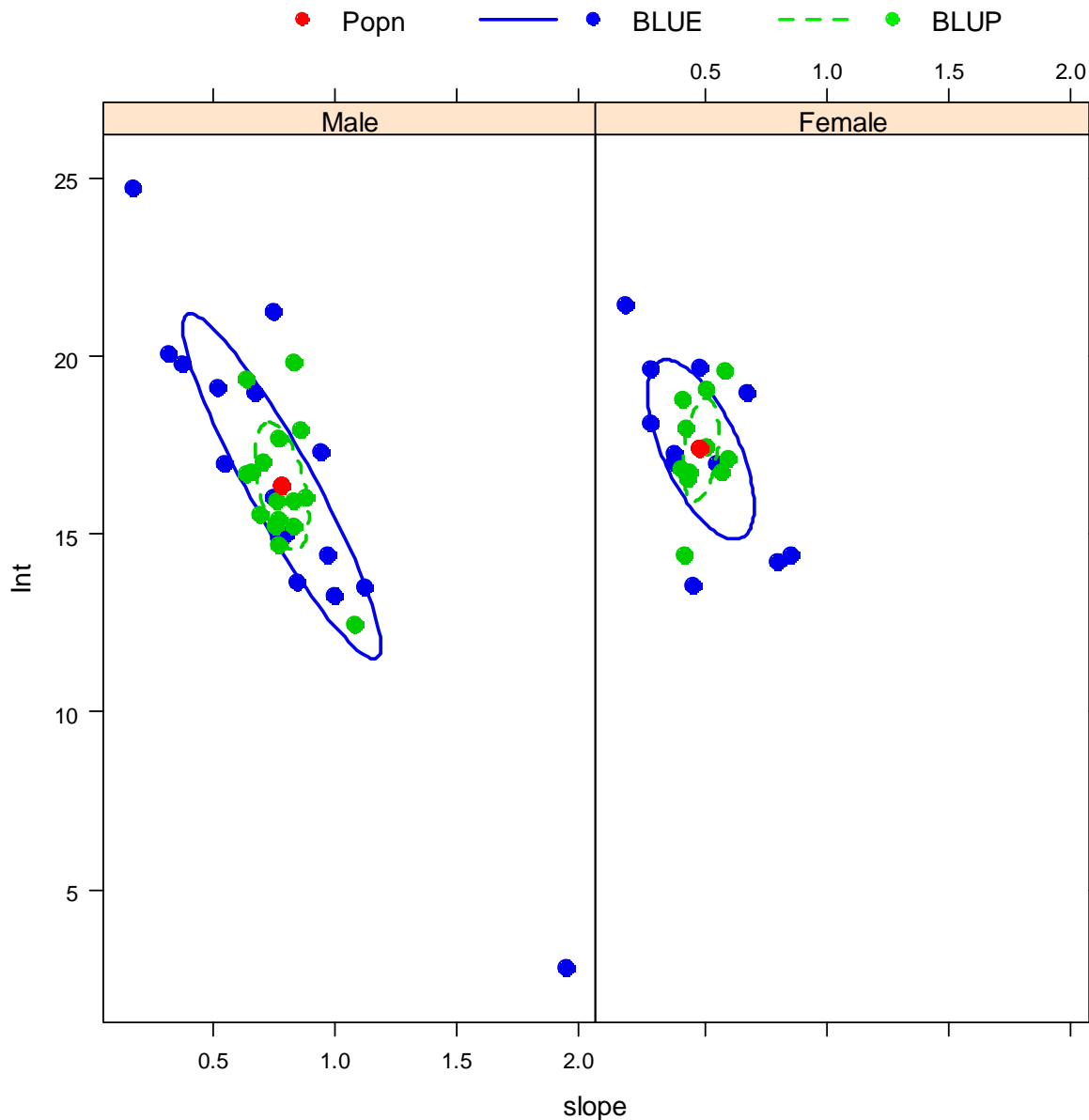
$$\text{Var}(\hat{\psi}_i) = \mathbf{G} + \sigma^2 (\mathbf{X}_i' \mathbf{X}_i)^{-1}$$

$$\text{Var}(\hat{\psi}_{i1}) \approx g_{11} + \frac{\sigma^2}{T_i \times S_{X_i}^2}$$

- $\text{Var}(\psi_i) = \mathbf{G}$  [population var.]
- $\text{Var}(\hat{\psi}_i | \psi_i) = \sigma^2 (\mathbf{X}_i' \mathbf{X}_i)^{-1}$   
[cond'l var. resampling from  $i^{\text{th}}$  subject]
- $\text{E}(\hat{\psi}_i | \psi_i) = \psi_i$  [BLUE]

So:

$$\begin{aligned} \text{Var}(\hat{\psi}_i) &= \text{Var}(\text{E}(\hat{\psi}_i | \psi_i)) + \\ &\quad \text{E}\{\text{Var}(\hat{\psi}_i | \psi_i)\} \\ &= \text{Var}(\psi_i) + \text{E}\{\text{Var}(\hat{\psi}_i | \psi_i)\} \\ &= \mathbf{G} + \sigma^2 (\mathbf{X}_i' \mathbf{X}_i)^{-1} \end{aligned}$$



While the expected variance of the BLUEs is larger than  $\mathbf{G}$

the expected variance of the BLUPs is smaller than  $\mathbf{G}$ .

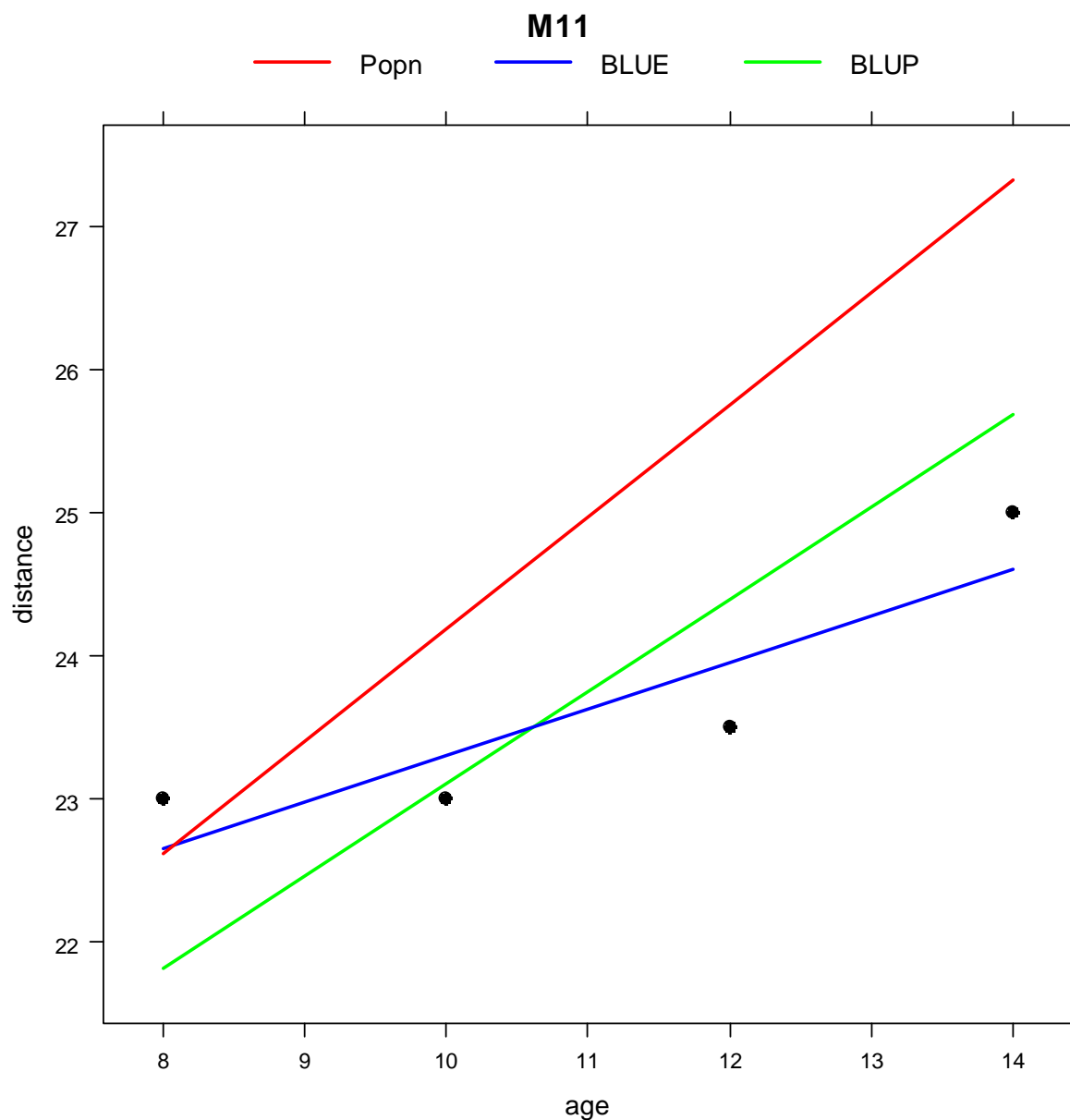
Beware of drawing conclusions about  $\mathbf{G}$  from the dispersion of the BLUPs.

The estimate of  $\mathbf{G}$  can be unstable and often collapses to singularity leading to non-convergence for many methods.

Possible remedies:

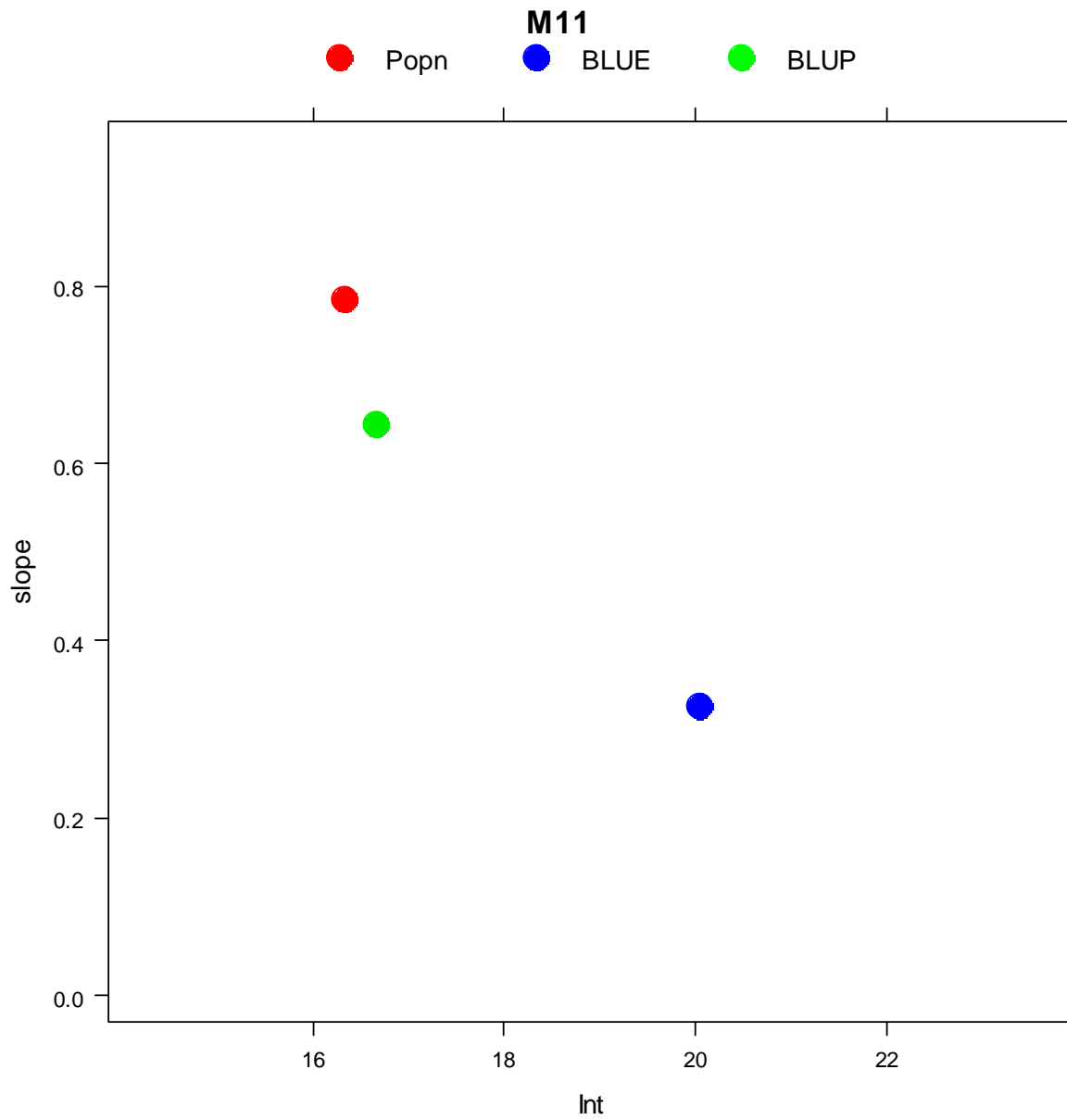
- Recentre  $\mathbf{X}$  near point of minimal variance,
- Use a smaller  $\mathbf{G}$
- Change the model

## Where the EBLUP comes from : looking at a single subject

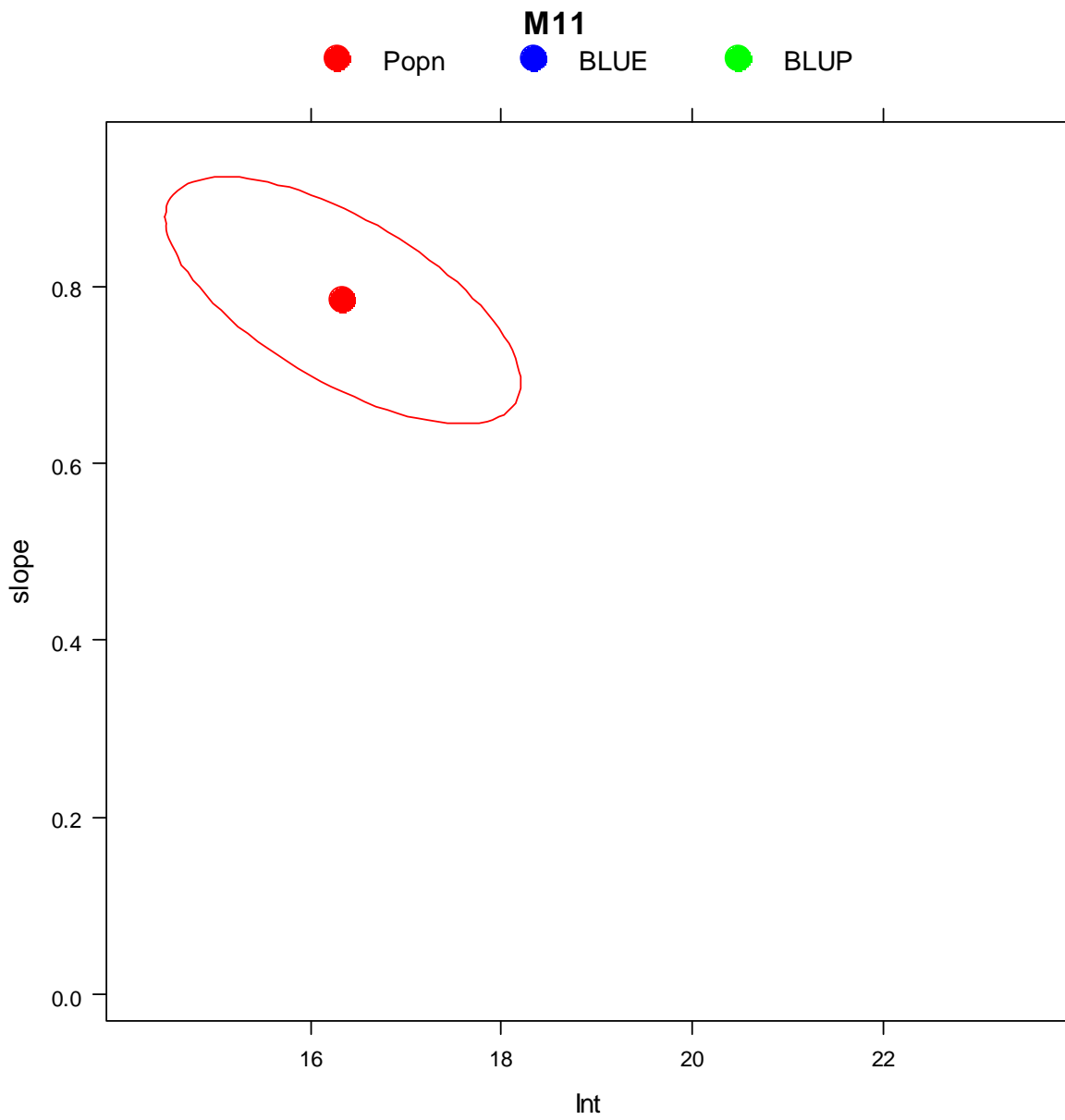


Note that the EBLUP's slope is close to the slope of the population estimate (i.e. the male population conditioning on between-subject predictors) while the level of the line is close to level of the BLUE.

The relative precisions of the BLUE and of the population estimate on slope and level are reflected through the shapes of  $\mathbf{G}$  and  $\sigma^2(\mathbf{X}_i' \mathbf{X}_i)^{-1}$

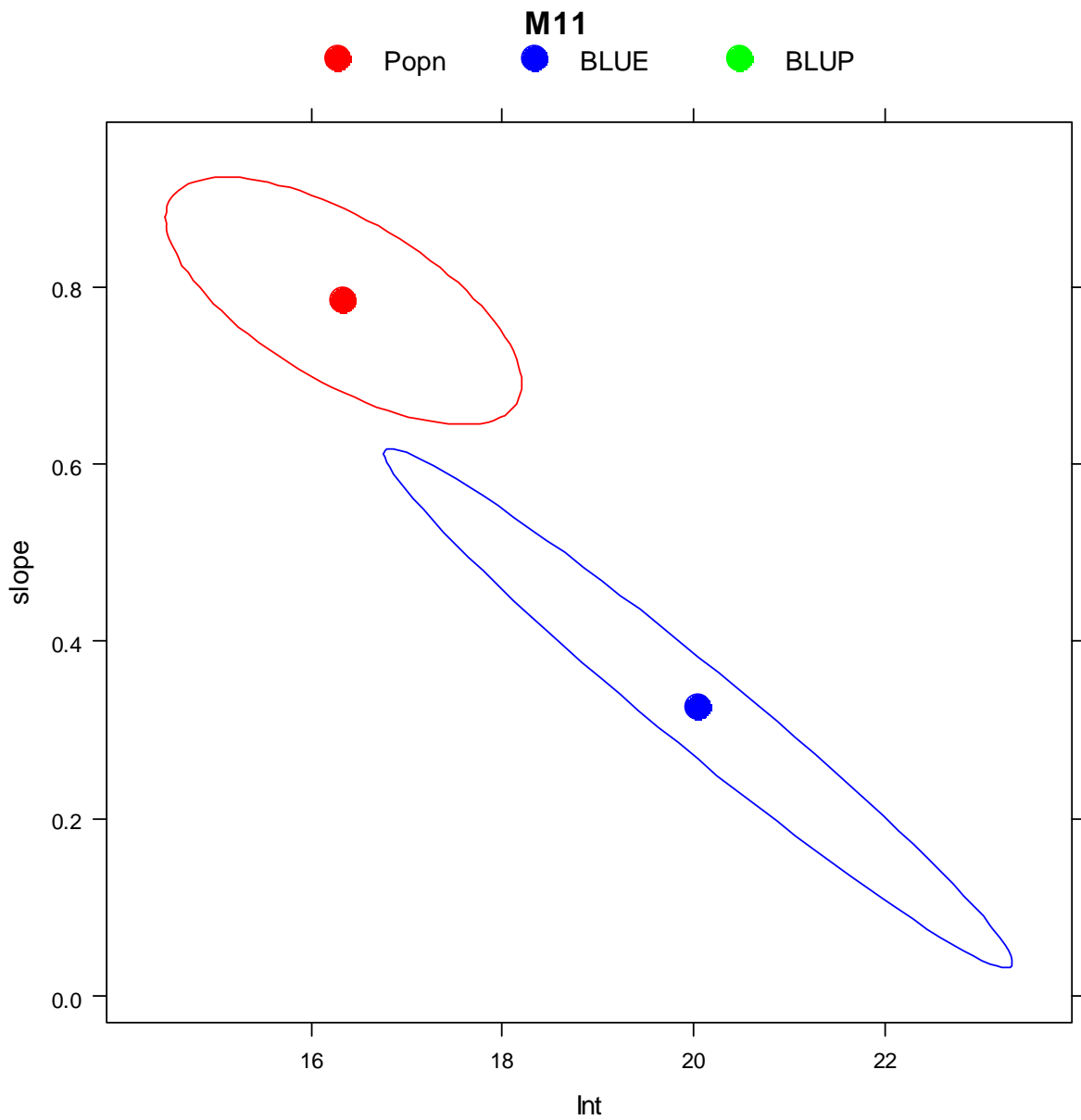


The same picture in  
“beta-space”



The population estimate  
with a SD ellipse.

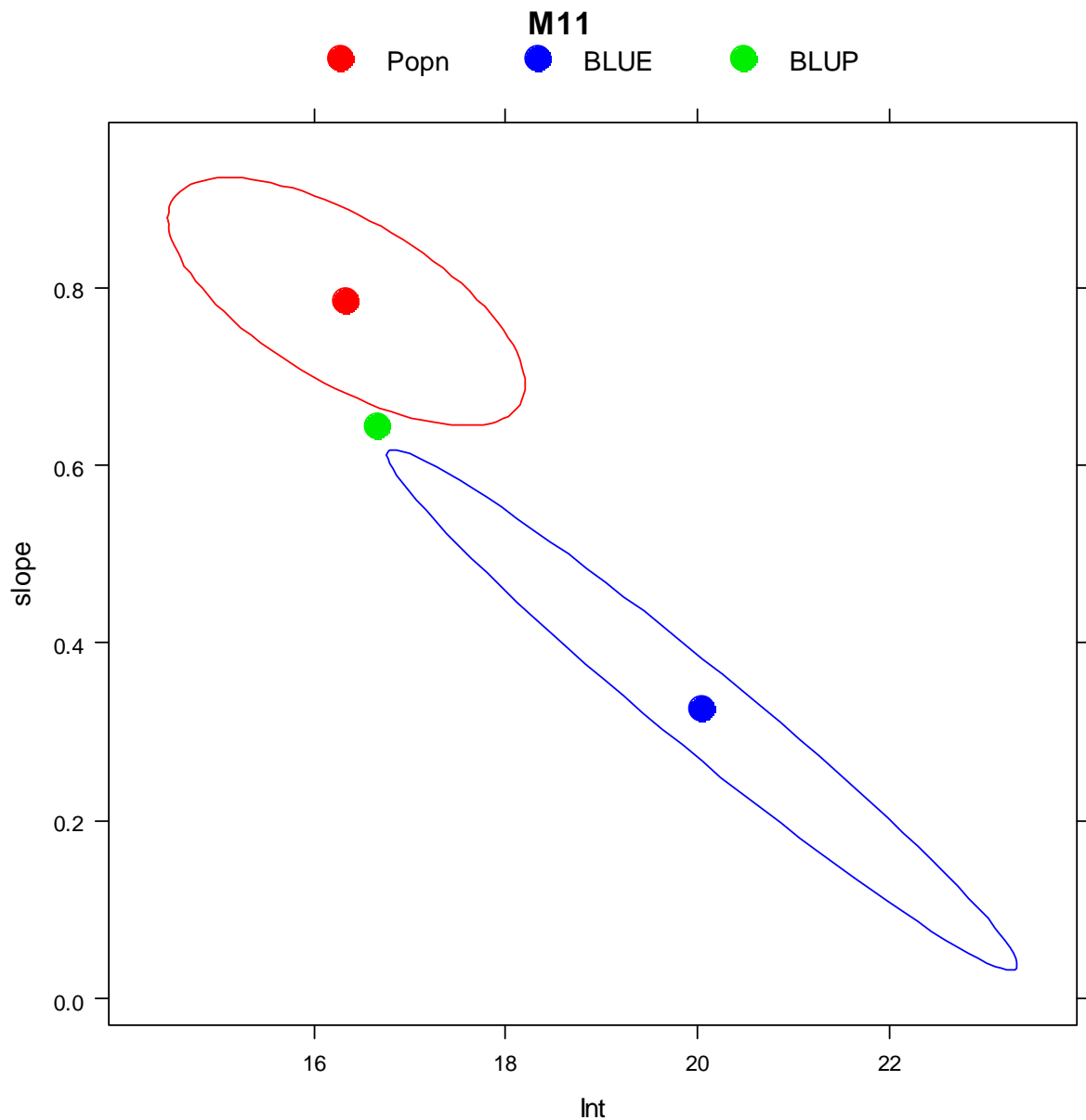




The population estimate  
with a SD ellipse

and

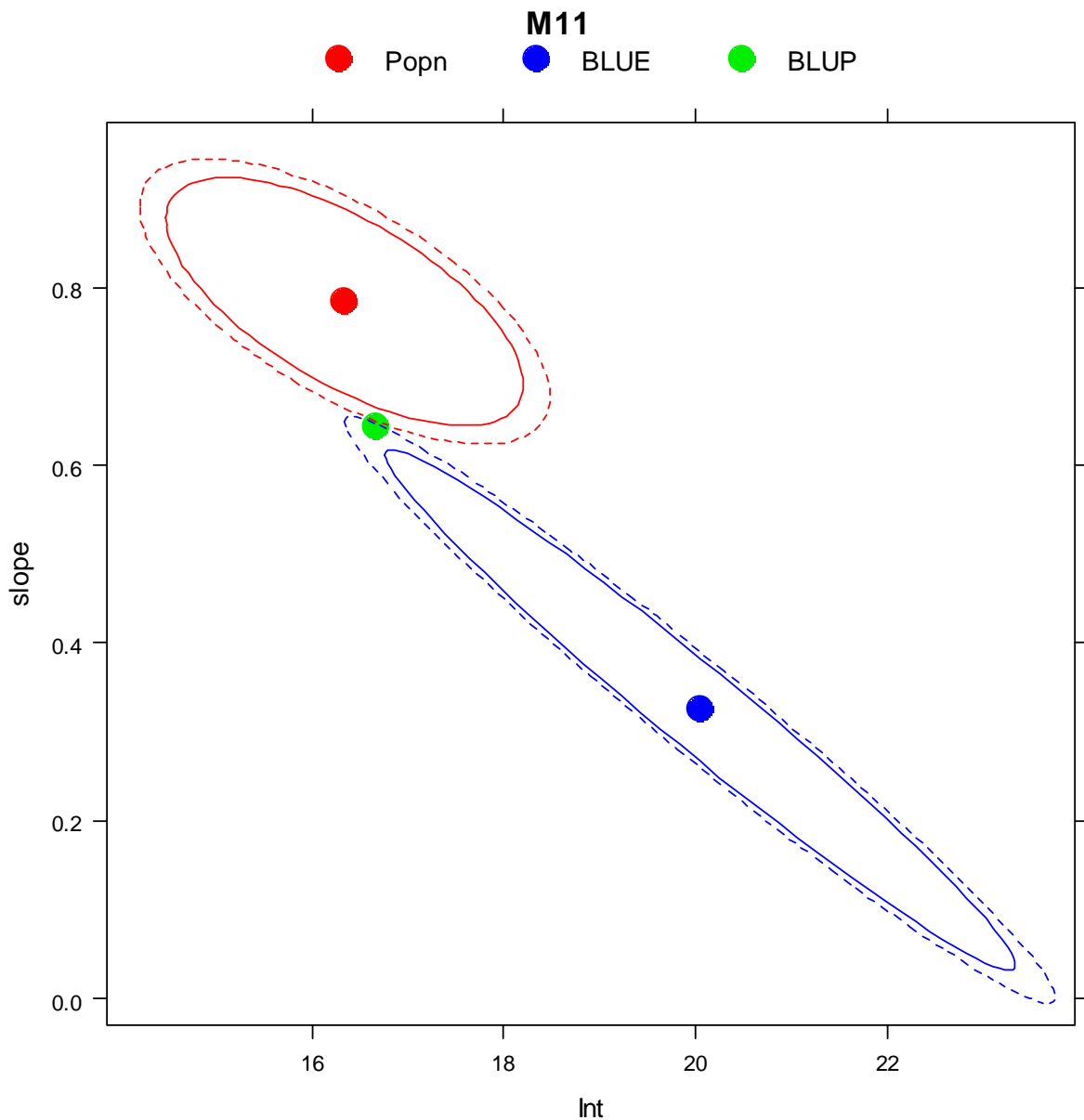
the BLUE with its SE  
ellipse



The EBLUP is an Inverse Variance Weighted mean of the BLUE and of the population estimate.

We can think of taking the BLUE and ‘shrinking’ it towards the population estimate along a path that optimally combines the two components.

The path is formed by the osculation points of the families of ellipses around the BLUE and the population estimate.



The amount and direction of shrinkage depends on the relative shapes and sizes of

**G**

and

$$\text{Var}(\hat{\psi}_i | i) = \sigma^2 (\mathbf{X}_i' \mathbf{X}_i)^{-1} \approx \frac{\sigma^2}{T_i} \mathbf{S}_{\mathbf{X}_i}^{-1}$$

The BLUP is at an osculation point of the families of ellipses generated around the BLUE and population estimate.

Imagine what could happen if **G** were oriented differently:

Paradoxically, both the slope and the intercept could be far outside the population estimate and the BLUE.

## When is a BLUP a BLUPPER?

The rationale behind BLUPs is based on *exchangeability*. No outside information should make this cluster stand out from the others and the mean of the population deserves the same weight in prediction for this cluster as it deserves for any other cluster that doesn't stand out.

If a cluster stands out somehow, then the BLUP might be a BLUPPER.

## ***Interpreting G***

The parameters of  $\mathbf{G}$  give the variance of the intercepts, the variance of the slopes and the covariance between intercepts and the slopes.

Would it make sense to assume that the covariance is 0 to reduce the number of parameters in the model? To address this, consider that the variance of the heights of individual regression lines a fixed value of  $X$  is:

$$\begin{aligned}\text{Var}(\psi_{i0} + \psi_{i1}X) &= \text{Var}\left(\begin{bmatrix} 1 & X \end{bmatrix} \begin{bmatrix} \psi_{i0} \\ \psi_{i1} \end{bmatrix}\right) \\ &= \begin{bmatrix} 1 & X \end{bmatrix} \begin{bmatrix} g_{00} & g_{01} \\ g_{10} & g_{11} \end{bmatrix} \begin{bmatrix} 1 \\ X \end{bmatrix} \\ &= g_{00} + 2g_{01}X + g_{11}X^2\end{aligned}$$

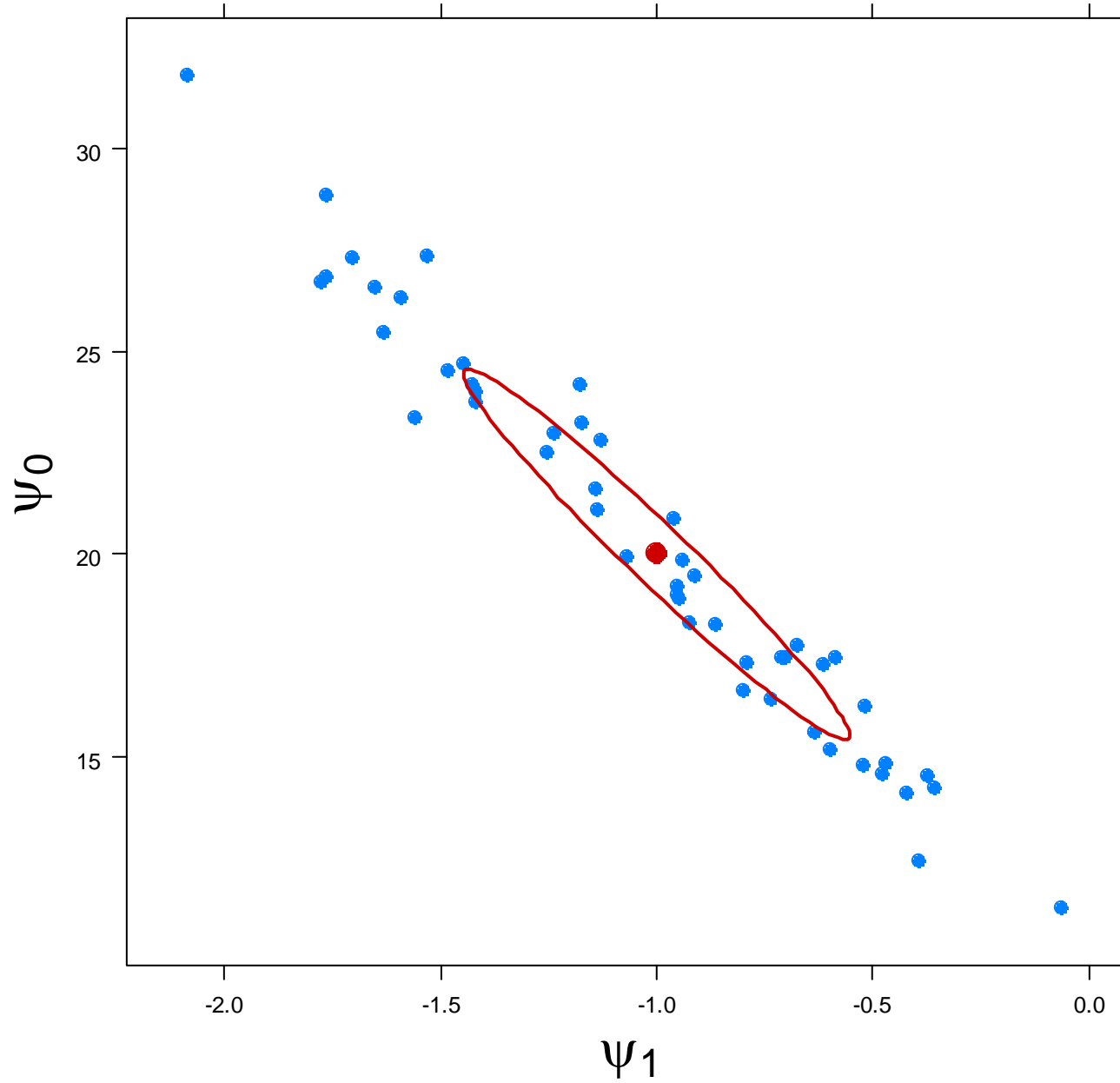
So  $\text{Var}(\psi_{i0} + \psi_{i1}X)$  has a minimum at  $-\frac{g_{01}}{g_{11}}$

and the minimum variance is  $g_{00} - \frac{g_{01}^2}{g_{11}}$

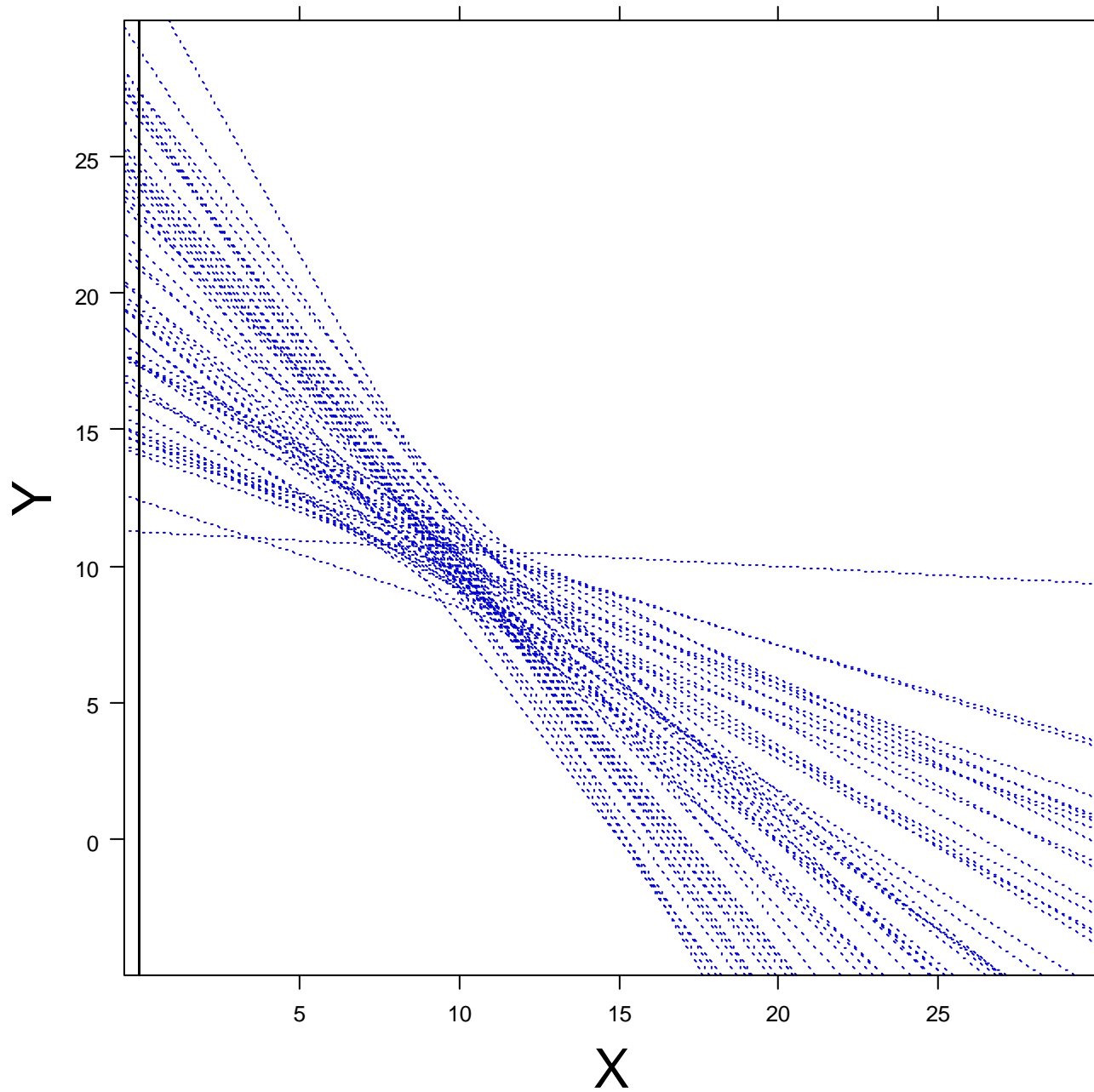
Thus assuming the covariance is 0 is equivalent to assuming that the minimum variance occurs when  $X = 0$ . This is an assumption that is not invariant with location transformations of  $X$ . It is similar to removing a main effect that is marginal to an interaction in a model, something that should not be done without a thorough understanding of its consequences.

Example: Let  $\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 20 \\ -1 \end{bmatrix}$  and  $\mathbf{G} = \begin{bmatrix} 10.5 & -1 \\ -1 & 0.1 \end{bmatrix}$

concentration ellipse

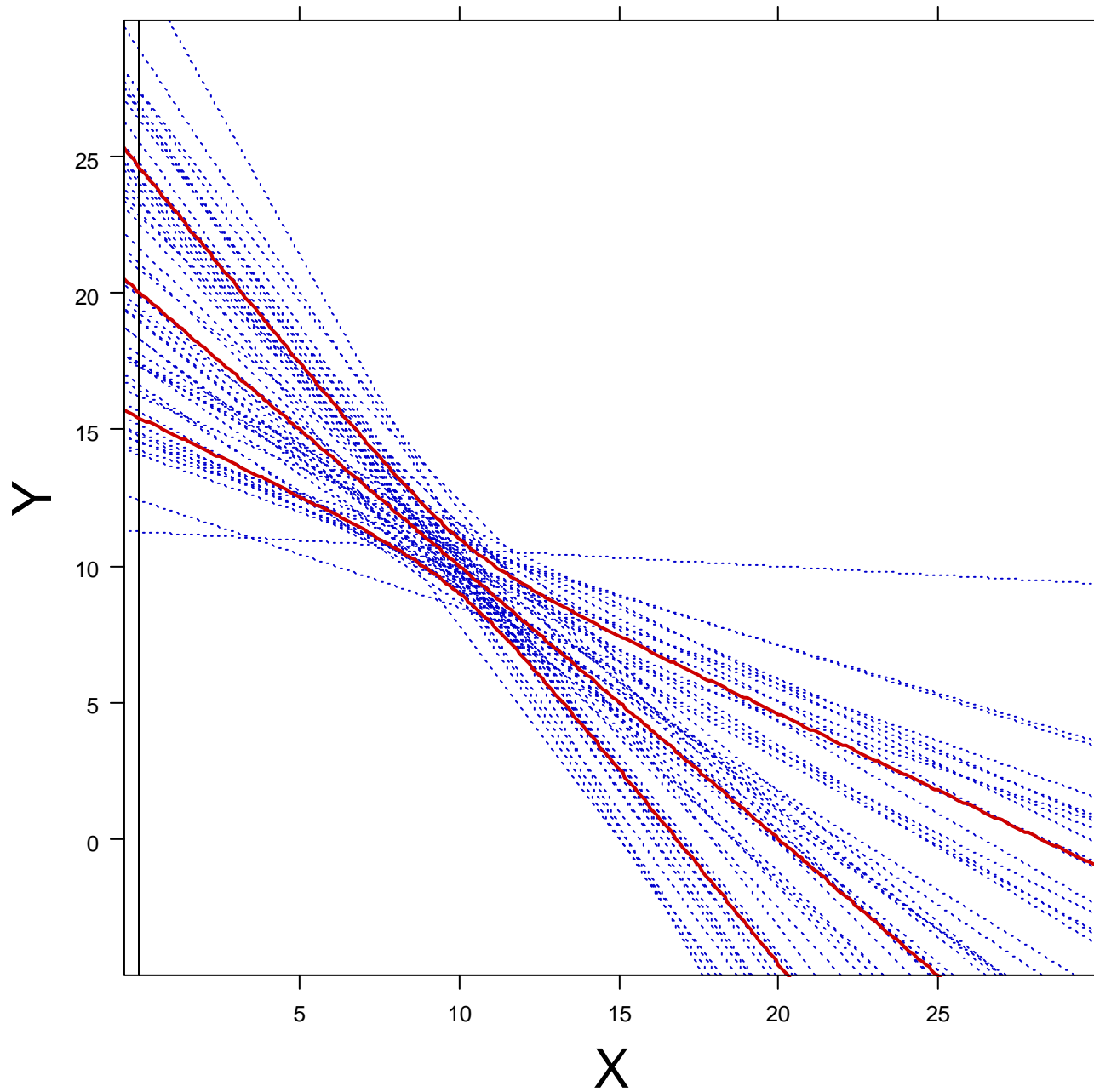


A sample of lines in beta space

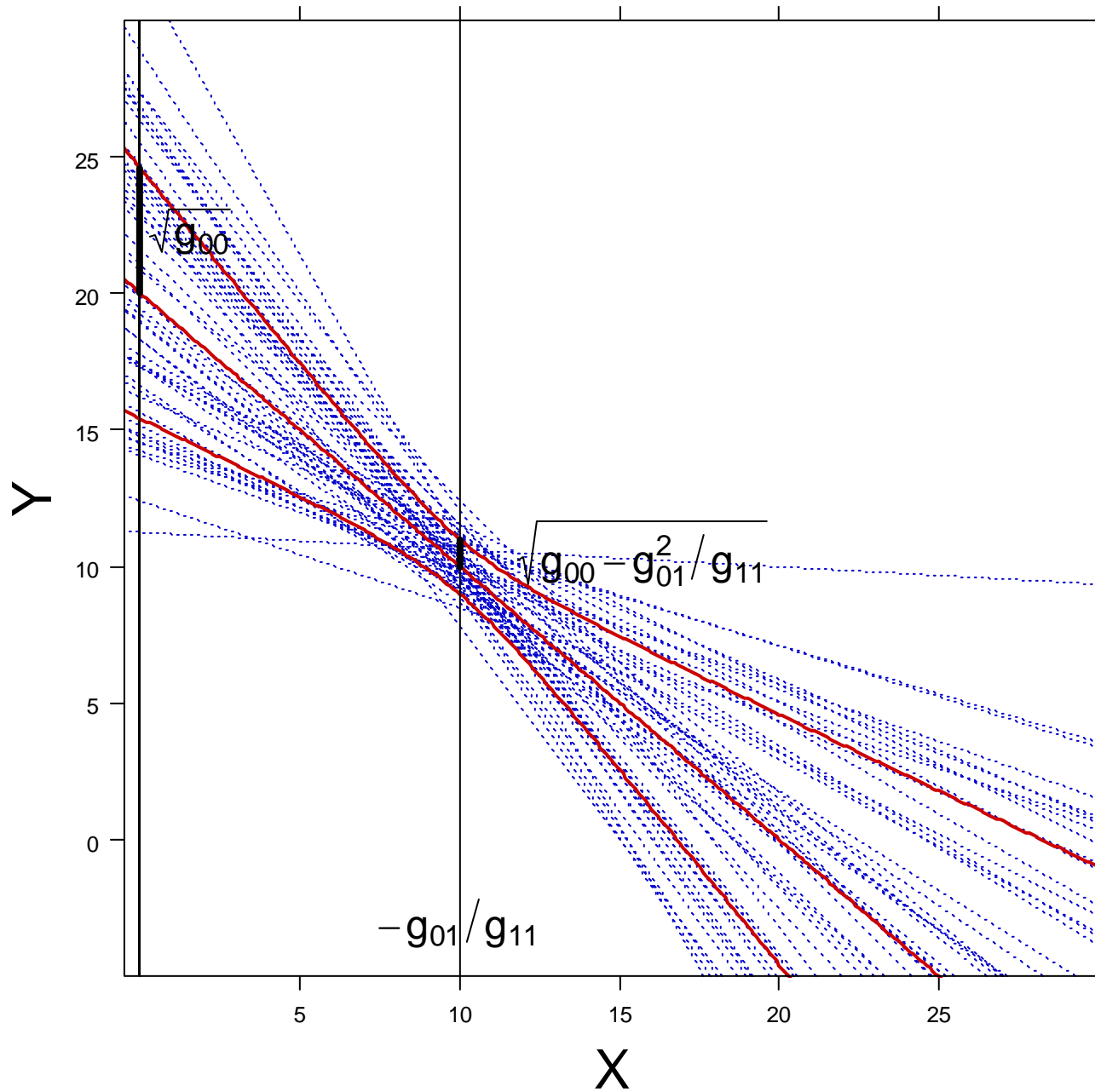


The same lines in data space.





The same lines in data space with the population mean line and lines at one SD above and below the population mean line



The parameters of  $\mathbf{G}$  determine the location and value of the minimum standard deviation of lines

With two time-varying variables with random effects, the  $\mathbf{G}$  matrix would look like:

$$\text{Var} \left( \begin{bmatrix} \psi_{i0} \\ \psi_{i1} \\ \psi_{i2} \end{bmatrix} \right) = \begin{bmatrix} g_{00} & g_{01} & g_{02} \\ g_{10} & g_{11} & g_{12} \\ g_{20} & g_{21} & g_{22} \end{bmatrix}$$

$g_{01}$  and  $g_{02}$  are related to the point of minimum variance in  $X_1, X_2$  space but  $g_{12}$  an interpretation that is origin-invariant and could be candidate for an assumption that it is 0. PROC MIXED code to do this is shown in an appendix.

## Differences between PROC GLM and MIXED with balanced data

Just looking at regression coefficients:

PROC GLM;

MODEL RESPONSE = SEX AGE AGE\*SEX;

Parameter	Estimate	Standard Error	t Value	Pr >  t
Intercept	16.34062500 B	1.41622419	11.54	<.0001
age	0.78437500 B	0.12616728	6.22	<.0001
Sex Female	1.03210227 B	2.21879688	0.47	0.6428
Sex Male	0.00000000 B	.	.	.
age*Sex Female	-0.30482955 B	0.19766614	-1.54	0.1261
age*Sex Male	0.00000000 B	.	.	.

```

PROC MIXED;
  MODEL RESPONSE = SEX AGE AGE*SEX;
  RANDOM INT / SUB = SUBJECT TYPE = FA0(1);

```

Solution for Fixed Effects

Standard

Effect	Sex	Estimate	Error	DF	t Value	Pr >  t
Intercept		16.3406	0.9813	104	16.65	<.0001
age		0.7844	0.07750	79	10.12	<.0001
Sex	Female	1.0321	1.5374	104	0.67	0.5035
Sex	Male	0	.	.	.	.
age*Sex	Female	-0.3048	0.1214	79	-2.51	0.0141
age*Sex	Male	0	.	.	.	.

PROC MIXED;

MODEL RESPONSE = SEX AGE AGE\*SEX;

RANDOM INT AGE / SUB = SUBJECT TYPE = FA0(2);

Effect	Sex	Estimate	Standard Error	DF	t Value	Pr >  t
Intercept		16.3406	1.0185	25	16.04	<.0001
age		0.7844	0.08600	25	9.12	<.0001
Sex	Female	1.0321	1.5957	25	0.65	0.5237
Sex	Male	0	.	.	.	.
age*Sex	Female	-0.3048	0.1347	25	-2.26	0.0326
age*Sex	Male	0	.	.	.	.

PROC MIXED also gives estimates of  $\mathbf{G}$  and  $\mathbf{V}_1 = \mathbf{Z}_1 \mathbf{G} \mathbf{Z}_1' + \mathbf{R}_1$

### Estimated G Matrix

Row	Effect	Subject	Col1	Col2
1	Intercept	F01	5.7864	-0.2896
2	age	F01	-0.2896	0.03252

### Estimated Chol(G) Matrix

Row	Effect	Subject	Col1	Col2
1	Intercept	F01	2.4055	
2	age	F01	-0.1204	0.1343

### Estimated G Correlation Matrix

Row	Effect	Subject	Col1	Col2
1	Intercept	F01	1.0000	-0.6676
2	age	F01	-0.6676	1.0000

### Estimated V Matrix for Subject F01

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

Estimated V Correlation Matrix for Subject F01

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

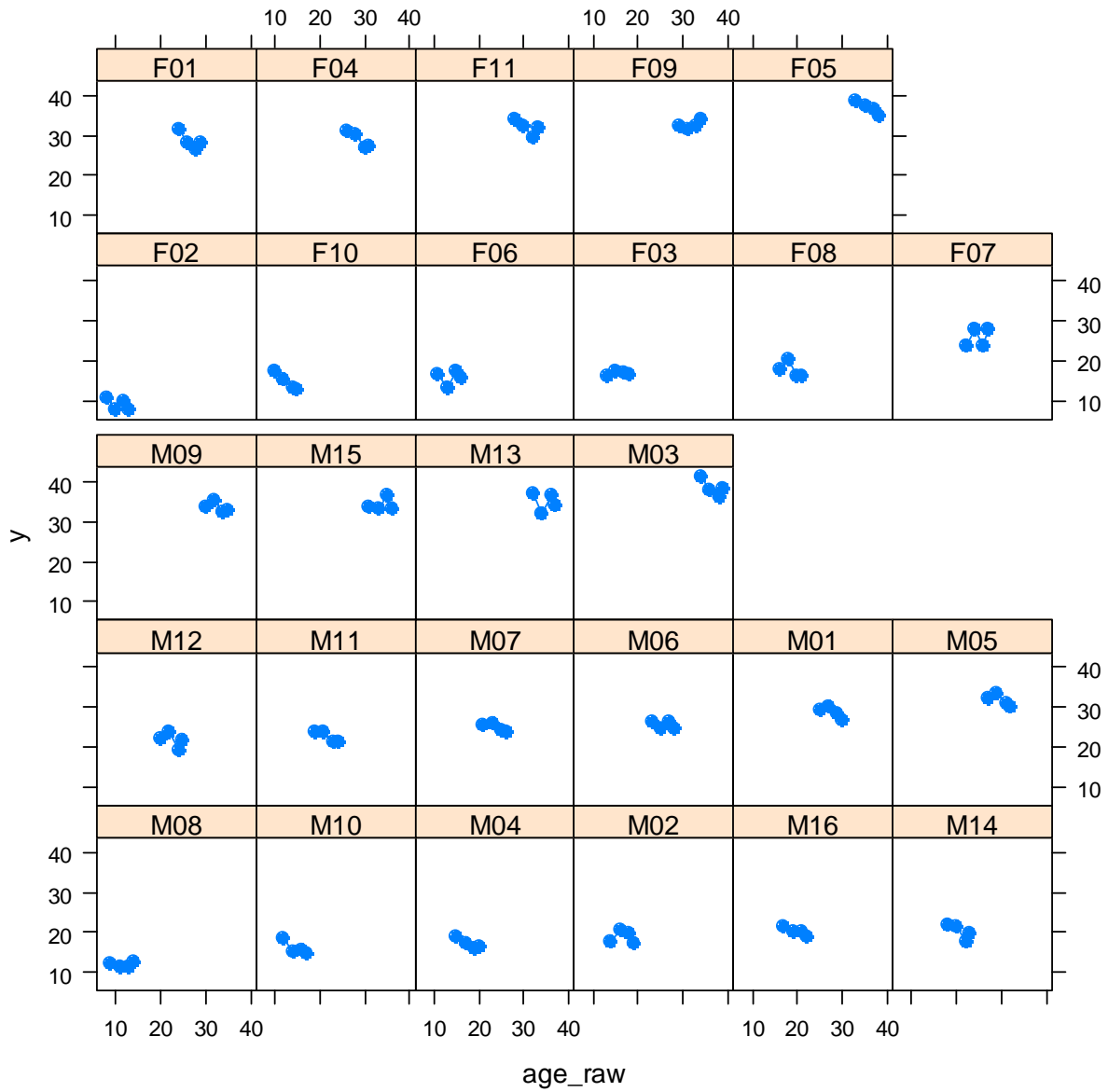
When data are balanced, PROC GLM and PROC MIXED produce the same  $\hat{\beta}$ s with different SEs.

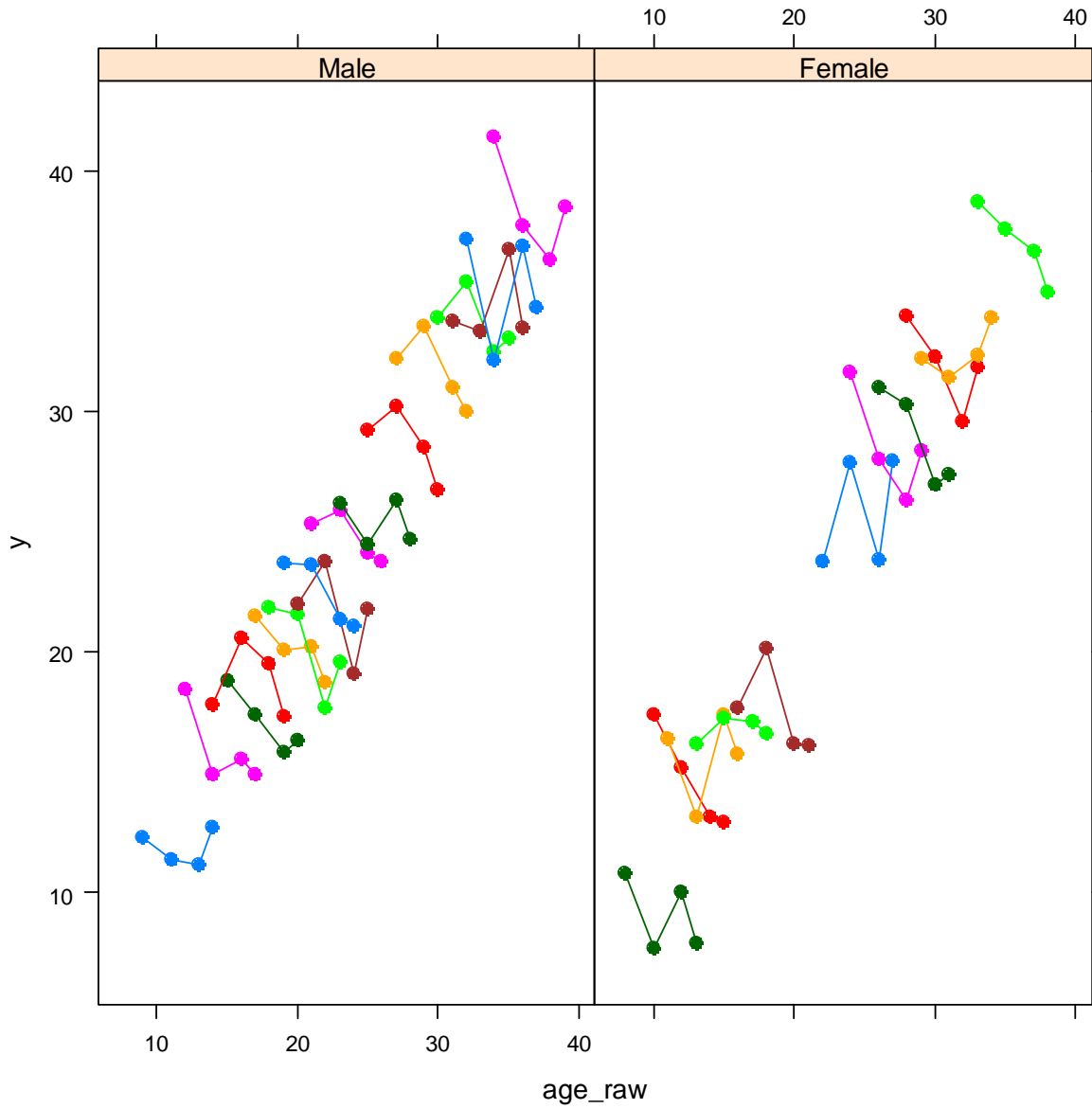


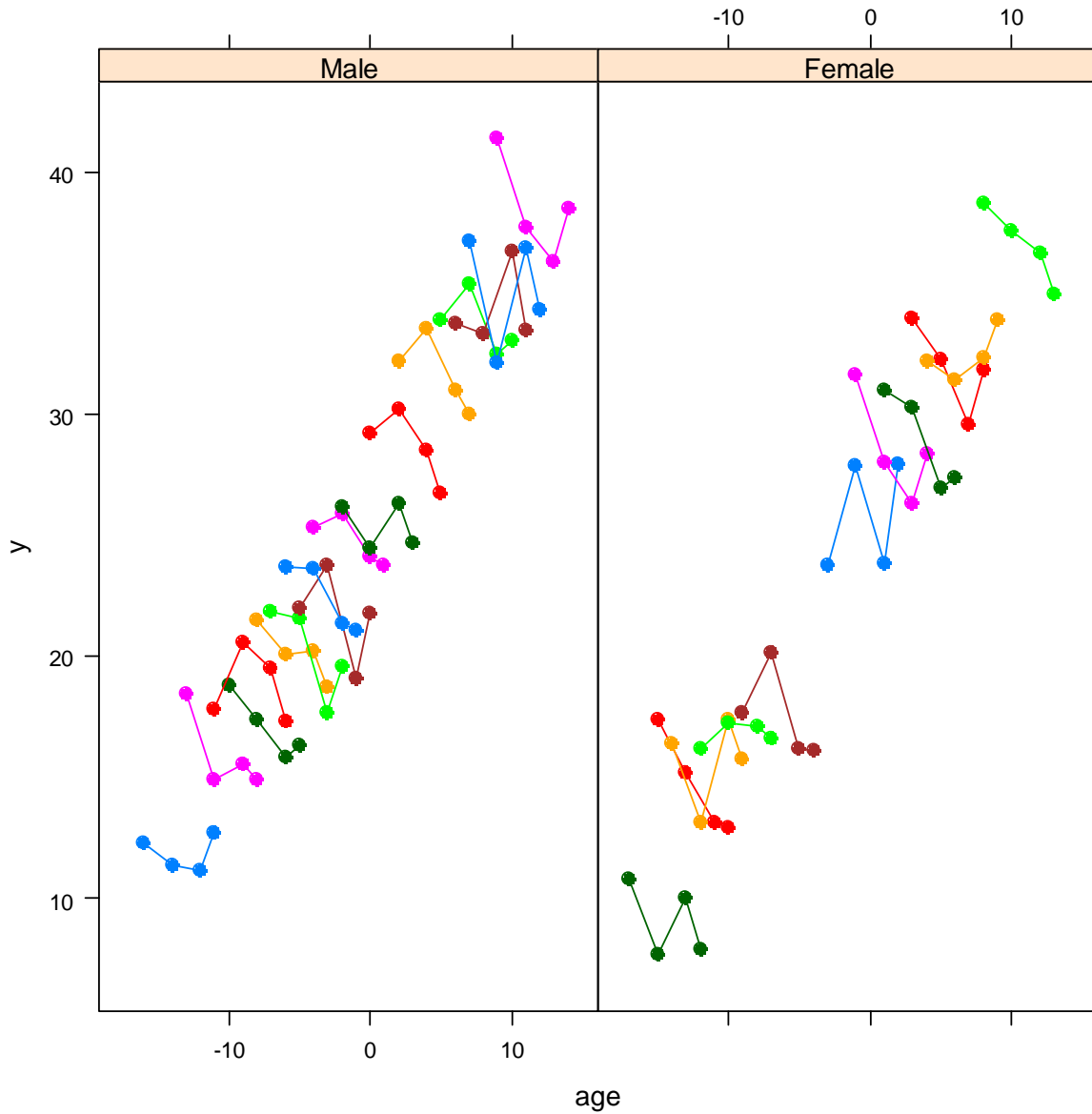
## Take 2: Learning lessons from unbalanced data

What can happen with unbalanced data? Here is some data that is similar to the Pothoff and Roy data but with:

- different age ranges for different subjects
- a between-subject effect of age that is different from the within-subject effect of age







Using age centered at 25.

Why? Like the ordinary regression model, the mixed model is **equivariant** under **global centering** but convergence may be improved because the  $\mathbf{G}$  matrix is less eccentric.

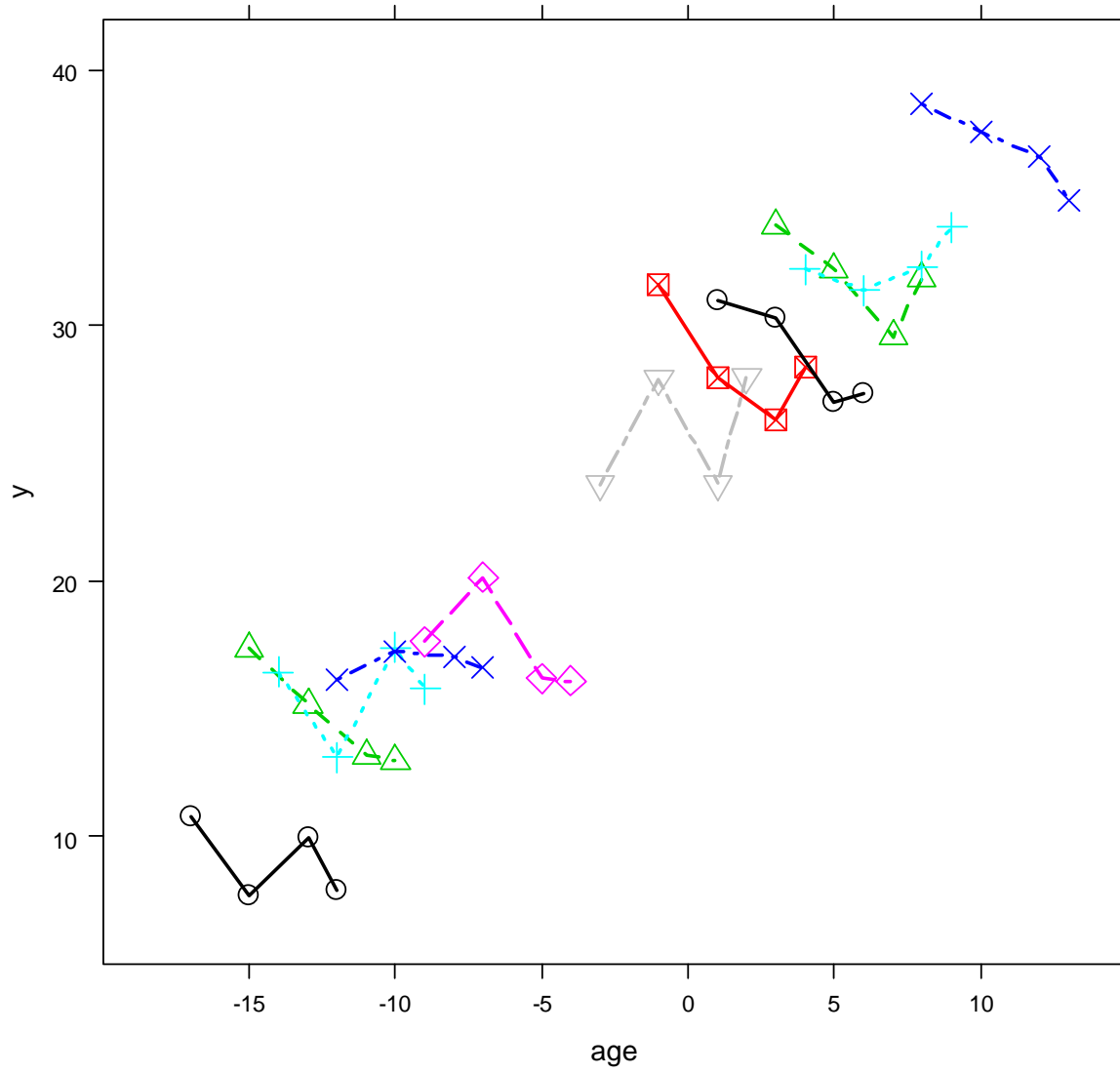
## SAS code and output

```
PROC MIXED ;  
  CLASS subject SEX;  
  MODEL y = age sex sex*age  
    / SOLUTION DDFM = SATTERTH;  
  RANDOM age INTERCEPT / TYPE = FA0(2) SUB = subject ;  
RUN;
```

### Solution for Fixed Effects

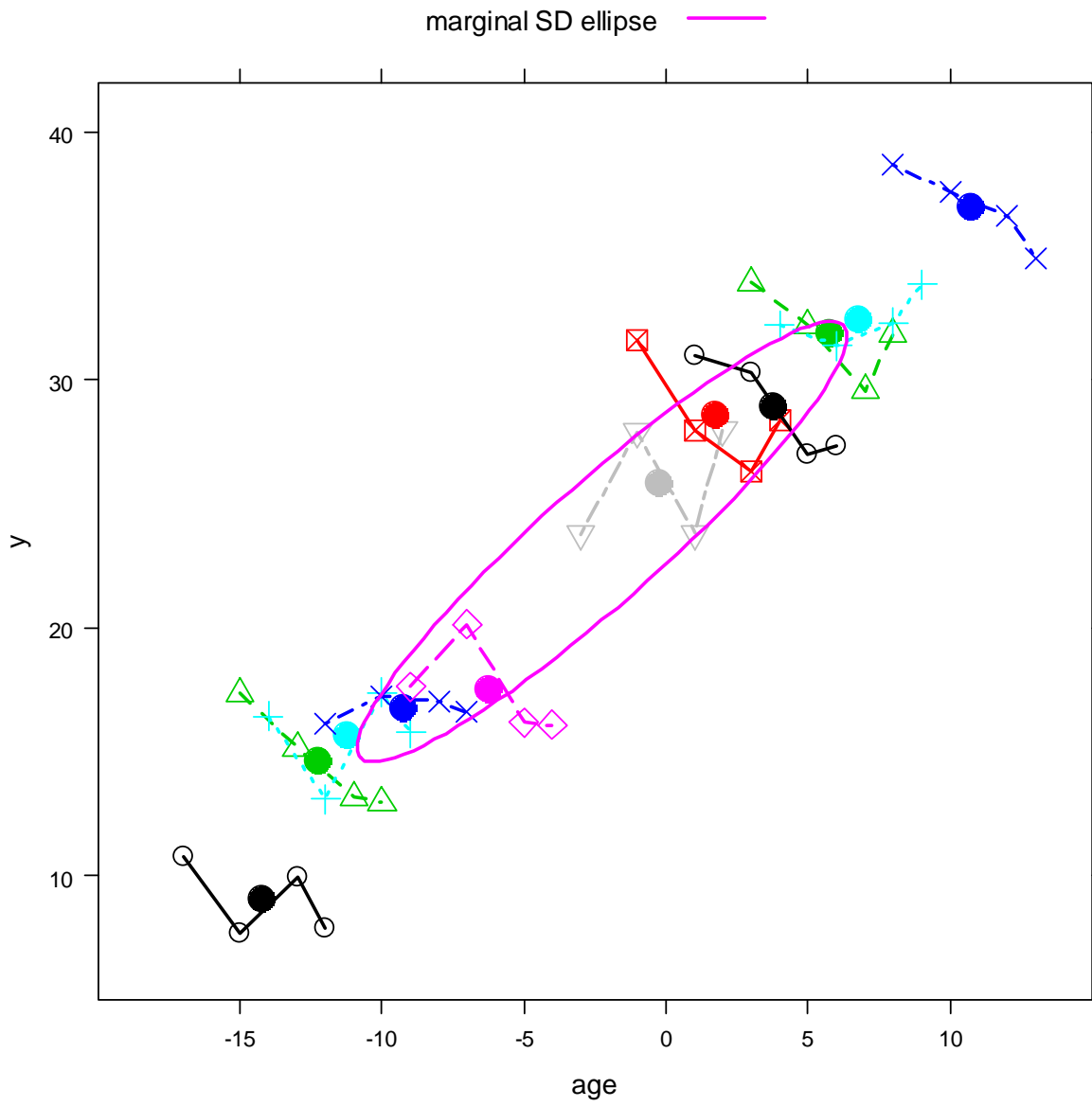
Effect	Sex	Estimate	Standard Error	DF	t Value	Pr >  t
Intercept		23.3275	2.8521	20	8.18	<.0001
age		-0.3778	0.1083	87.3	-3.49	0.0008
Sex	Female	-0.1476	4.4687	20	-0.03	0.9740
Sex	Male	0	.	.	.	.
age*Sex	Female	-0.1268	0.1724	83	-0.74	0.4642
age*Sex	Male	0	.	.	.	.

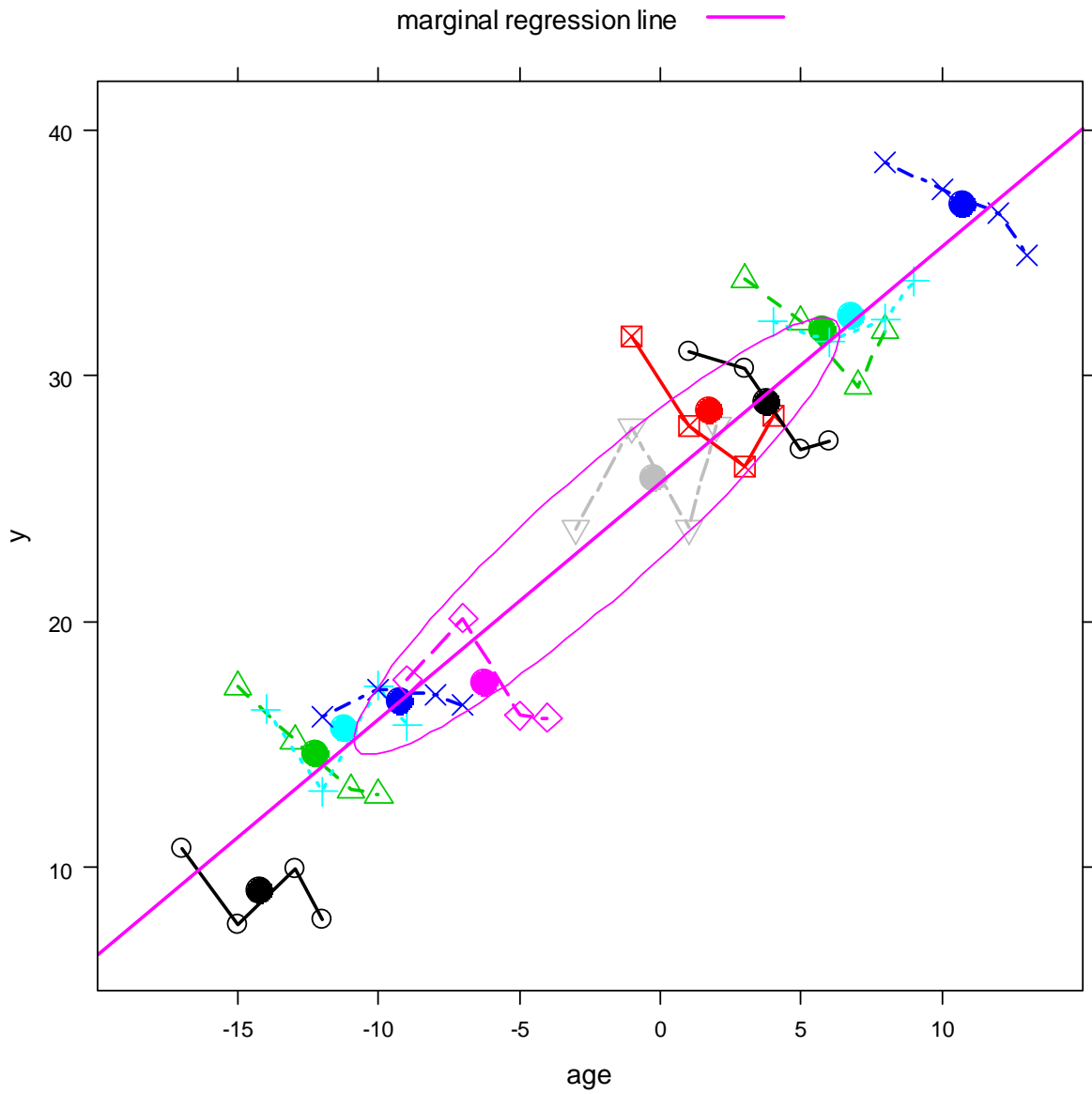
## *Between, Within and Pooled Models*



We first focus on one group, the female data:

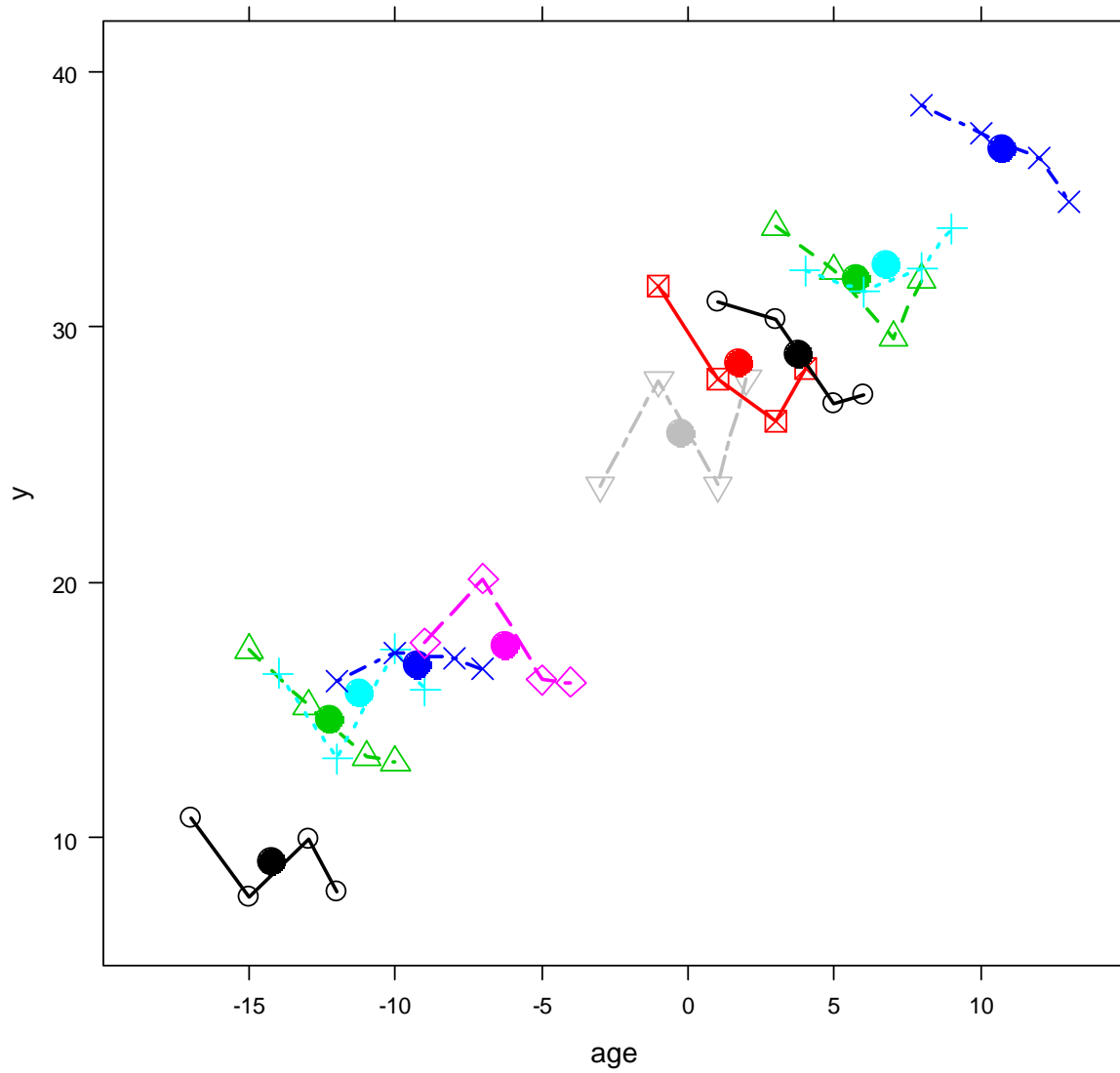
What models could we fit to this data?



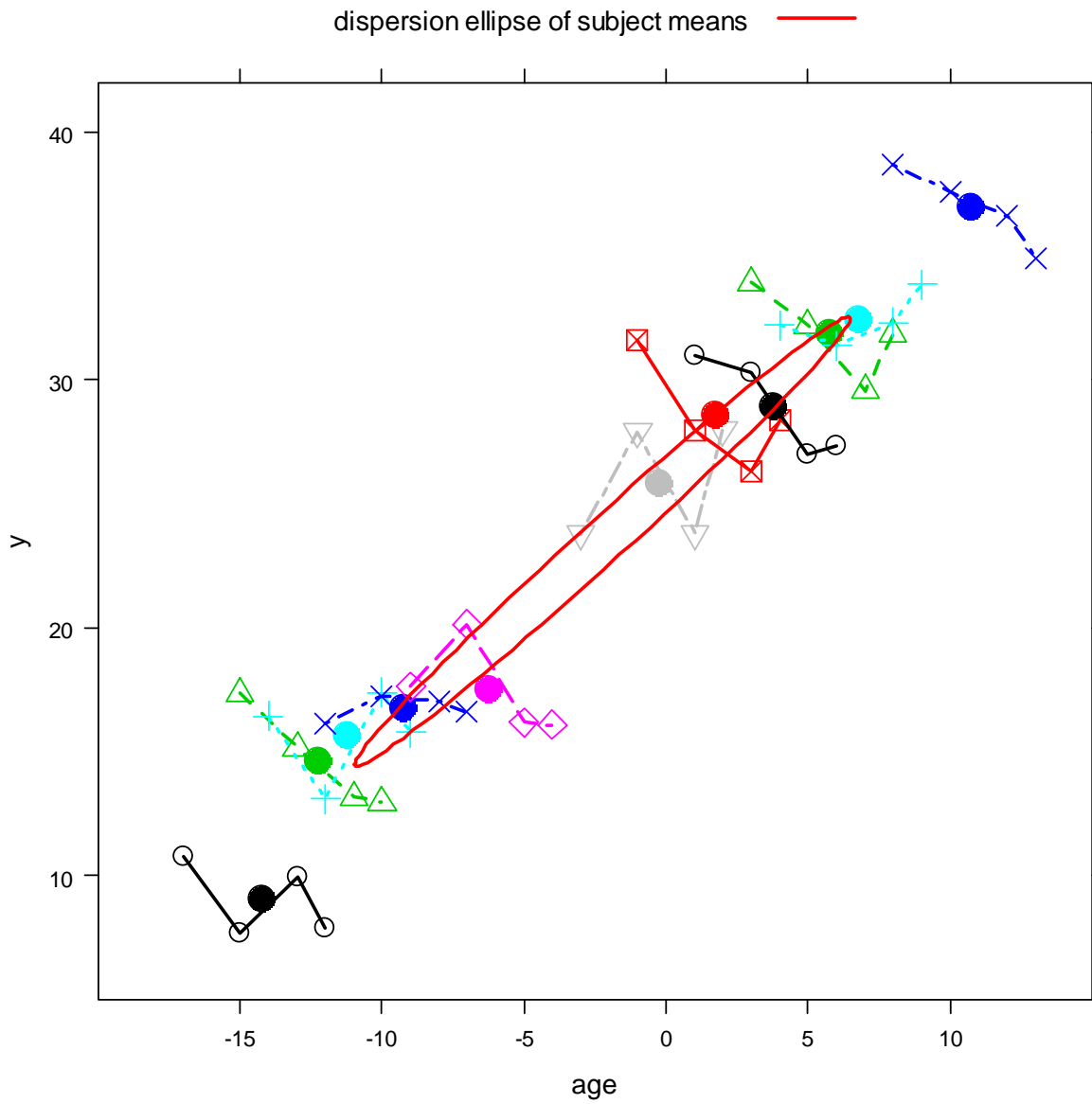


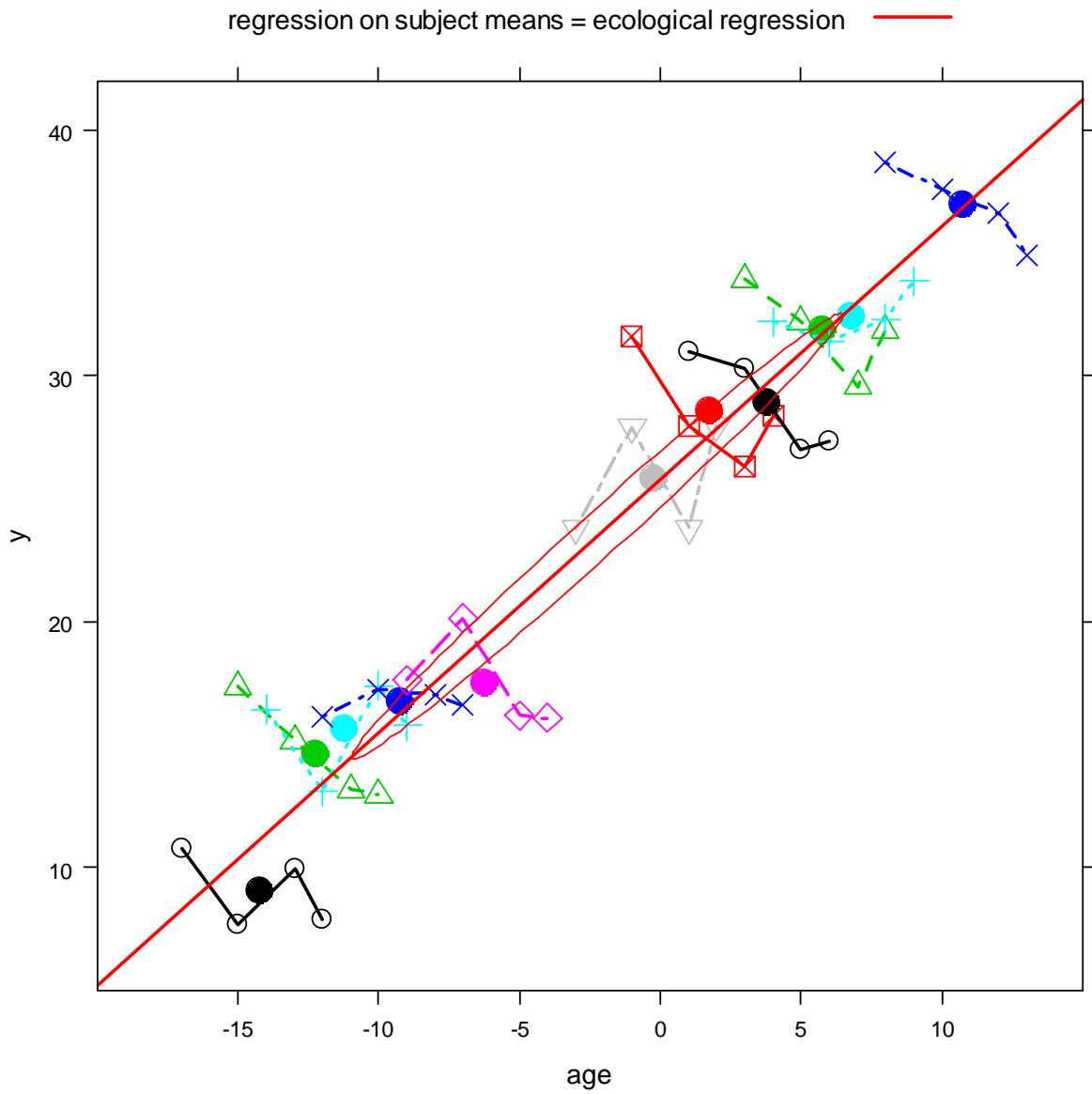
Regressing with the pooled data – ignoring Subject – yields the marginal (unconditional) estimate of the slope:  $\hat{\beta}_P$



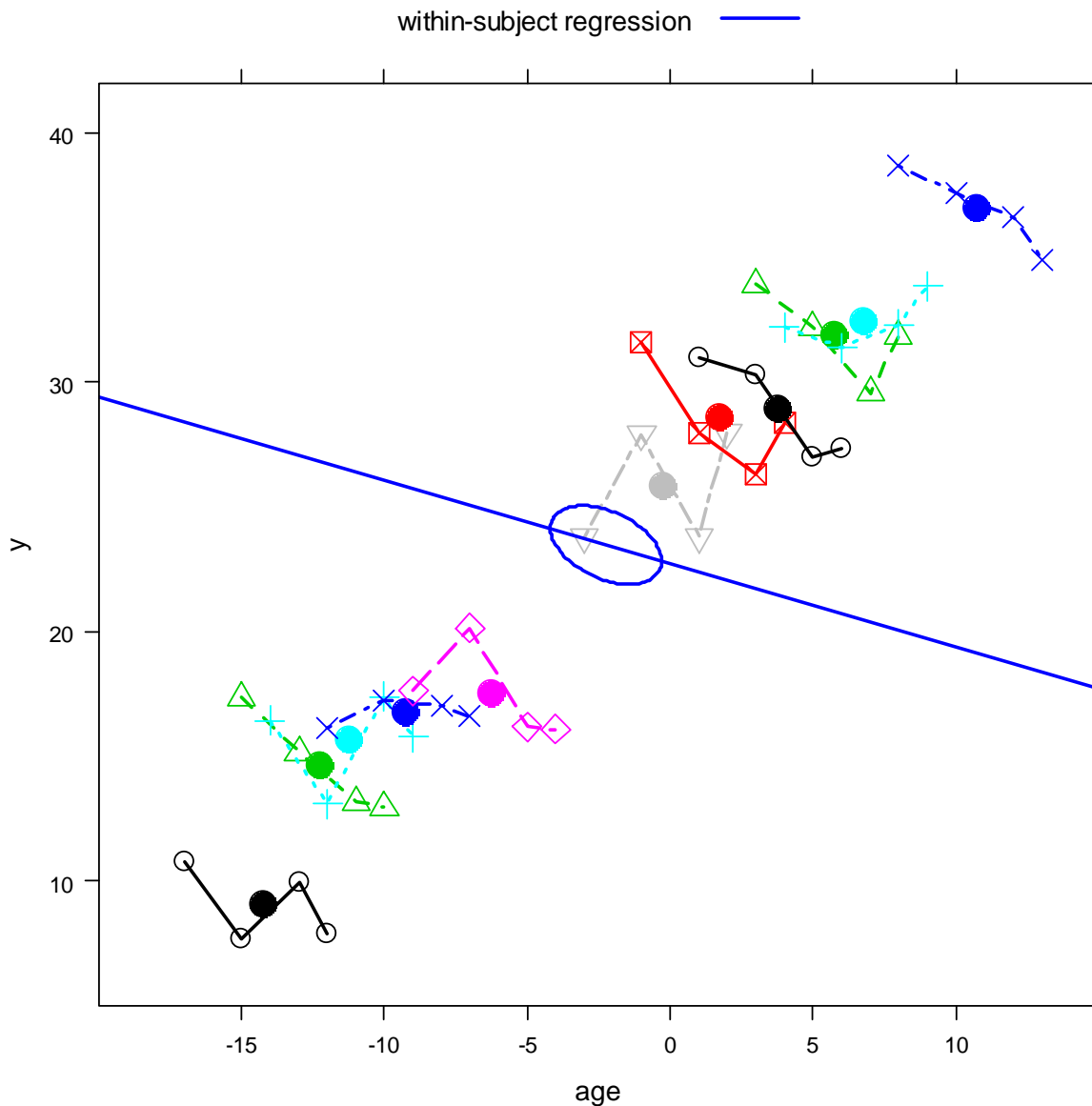


We could replace each Subject by its means for x and y and use the resulting aggregated data with one point for each Subject.





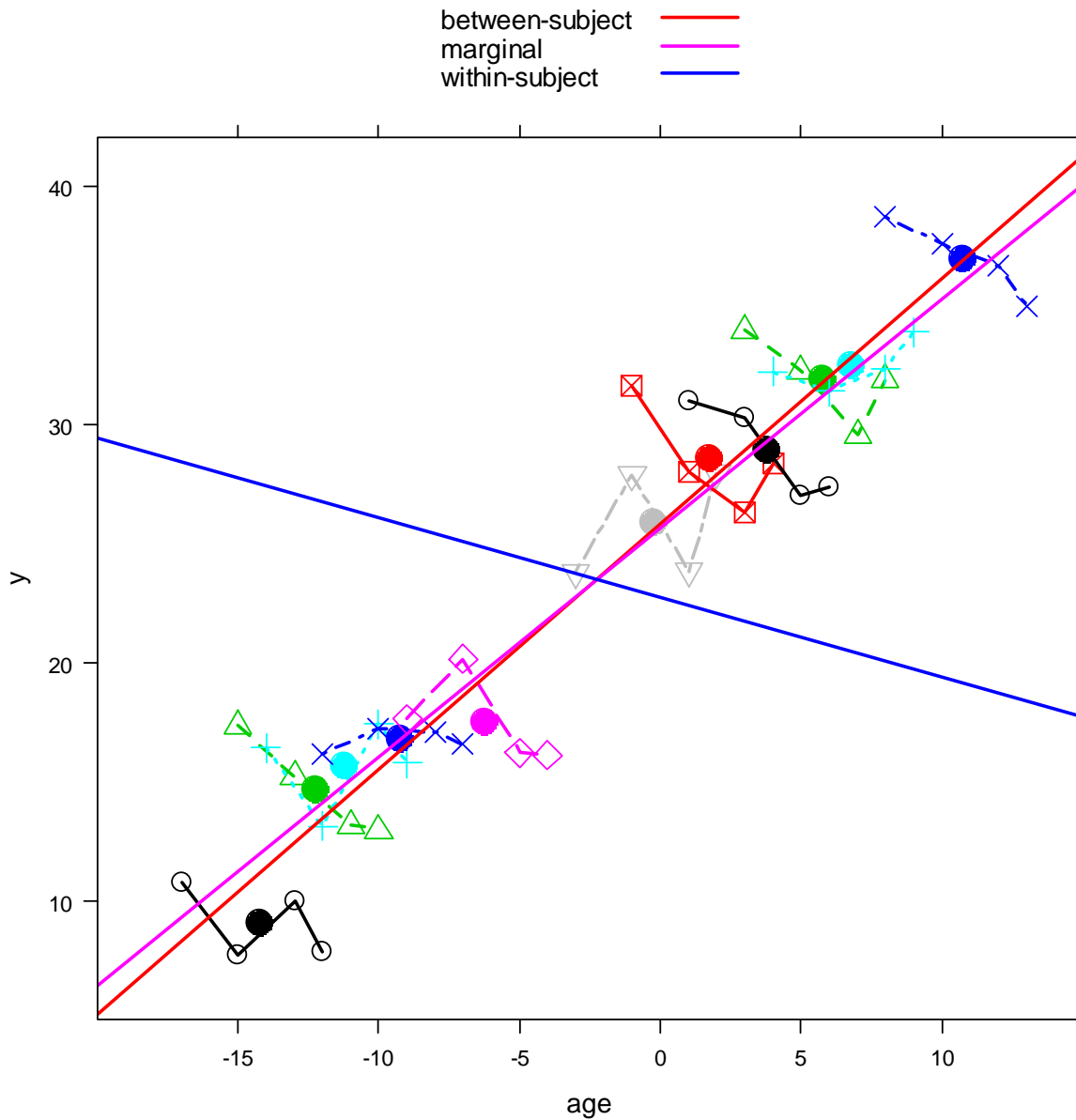
Performing a regression on the aggregated data yields the 'between-subject' regression, in some contexts called an 'ecological regression'



We can combine all within-subject regressions to get a combined estimate of the within-subject slope. This is the estimate obtained with a fixed-effects model using age and Subject additively. Equivalently, we can perform a regression using (the within-subject residuals of  $y$  minus mean  $y$ ) on (age minus mean age).

Q: Which is better:  $\hat{\beta}_B$ ,  $\hat{\beta}_W$  or  $\hat{\beta}_P$ ?

A: None. They answer different questions. Typically,  $\hat{\beta}_P$  would be used for prediction across the population;  $\hat{\beta}_W$  for ‘causal’ inference controlling for between-subject confounders.



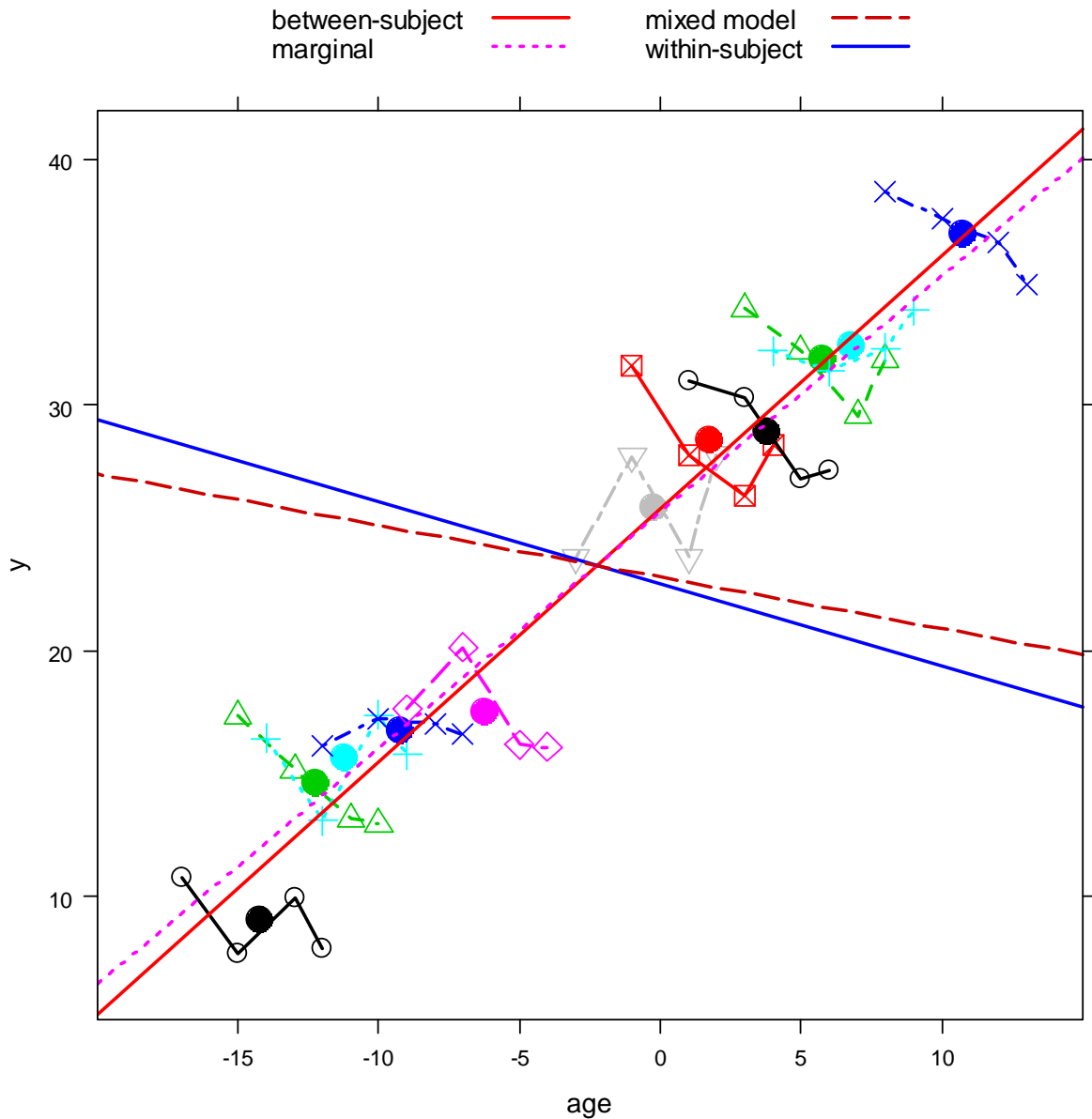
The relationship among estimators:

$\hat{\beta}_P$  combines  $\hat{\beta}_B$  and  $\hat{\beta}_W$ :

$$\hat{\beta}_P = (W_B + W_W)^{-1} (W_B \hat{\beta}_B + W_W \hat{\beta}_W)$$

The weights depend only on the design, not of estimated variances of the response.

# The Mixed Model



The mixed model estimate<sup>3</sup> also combines

$\hat{\beta}_B$  and  $\hat{\beta}_W$ :

$$\hat{\beta}_{MM} = (W_B^{MM} + W_W)^{-1} (W_B^{MM} \hat{\beta}_B + W_W \hat{\beta}_W)$$

but with a lower weight on  $\hat{\beta}_B$ :

$$W_B^{MM} = \frac{\sigma^2 / T}{\sigma^2 / T + g_{00}} W_B = \frac{\frac{\sigma^2 / T}{g_{00}}}{1 + \frac{\sigma^2 / T}{g_{00}}} W_B$$

Note that

$$W_B^{MM} \leq W_B$$

<sup>3</sup> Using a random intercept model

- The mixed model estimator is a variance optimal combination of  $\hat{\beta}_B$  and  $\hat{\beta}_W$ .
- It makes perfect sense if  $\hat{\beta}_B$  and  $\hat{\beta}_W$  estimate the same thing, i.e. if  $\beta_B = \beta_W$ !
- Otherwise, it's an arbitrary combination of estimates that estimate different things. The weights in the combination have no substantive interpretation.
- i.e. it's an optimal answer to a meaningless question.

Summary of the relationships among 4 models:

Model	Estimate of slope	Precision
Between Subjects	$\hat{\beta}_B$	$W_B$
Marginal (pooled data)	$\hat{\beta}_P$	
Mixed Model	$\hat{\beta}_{MM}$	
Within Subjects	$\hat{\beta}_W$	$W_W$

The pooled estimate combines  $\hat{\beta}_B$  and  $\hat{\beta}_W$ :

$$\hat{\beta}_P = (W_B + W_W)^{-1} (W_B \hat{\beta}_B + W_W \hat{\beta}_W)$$

Mixed model:

With a random intercept model:

$$y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_{i0} + \varepsilon_{it}, \quad \gamma_{i0} \sim N(0, g_{00}), \quad \varepsilon_{it} \sim N(0, \sigma^2)$$

with  $g_{00}, \sigma^2$  known  $\hat{\beta}_{\text{MM}}$  is also a weighted combination of  $\hat{\beta}_{\text{B}}$  and  $\hat{\beta}_{\text{W}}$  but with less weight on  $\hat{\beta}_{\text{B}}$ :

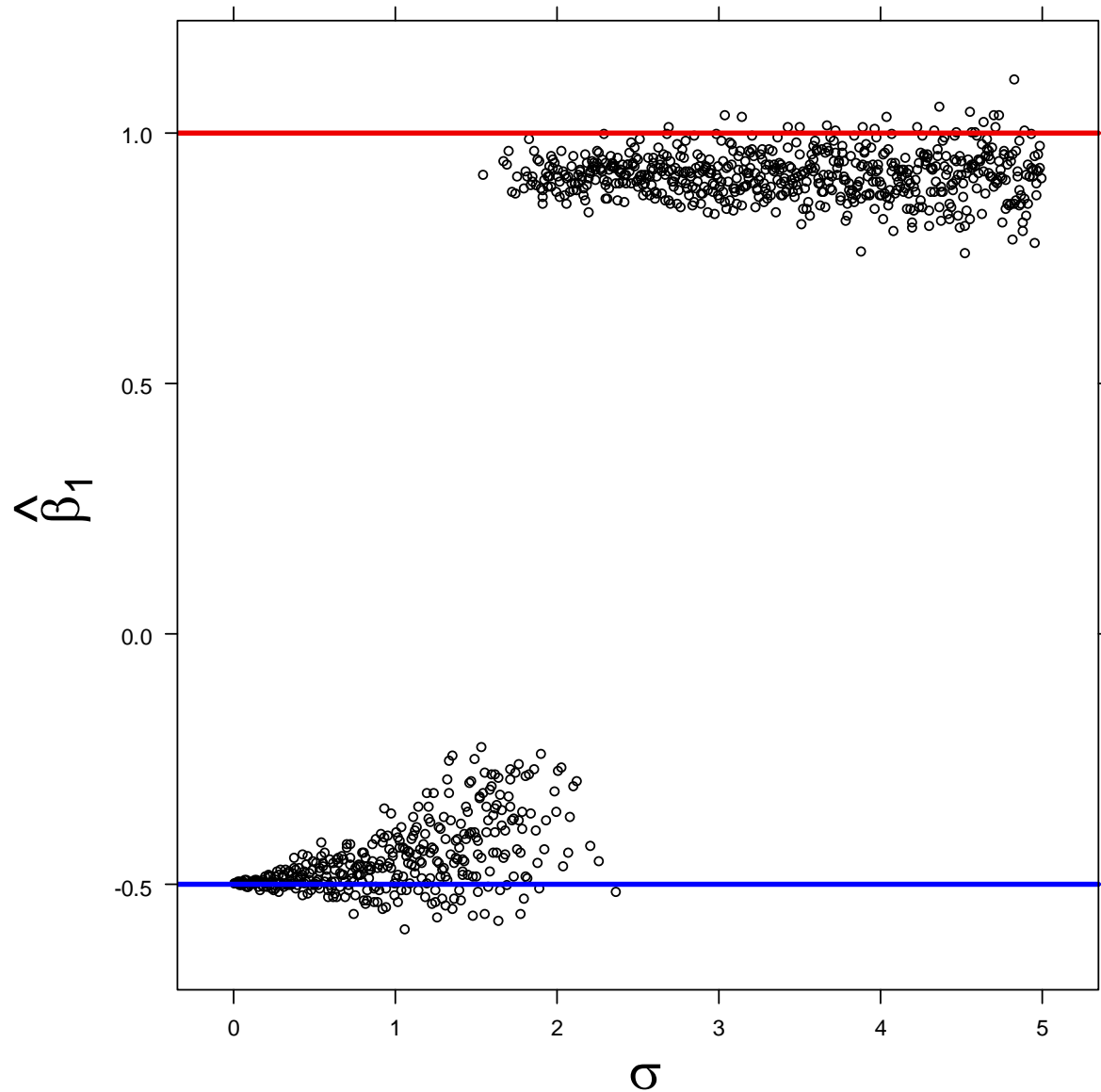
$$\begin{aligned} W_{\text{B}}^{\text{MM}} &= \frac{\sigma^2 / T}{\sigma^2 / T + g_{00}} W_{\text{B}} \\ &= f_{\text{monotone}} \left( \frac{\text{Between-Subject Information}}{\text{Within-Subject Information}} \right) \times W_{\text{B}} \end{aligned}$$

- $\hat{\beta}_{\text{MM}}$  is between  $\hat{\beta}_{\text{W}}$  and  $\hat{\beta}_{\text{P}}$ , i.e. it does better than  $\hat{\beta}_{\text{P}}$  in the sense of being closer to  $\hat{\beta}_{\text{W}}$  but is not equivalent to  $\hat{\beta}_{\text{W}}$ .
- With balanced data  $\hat{\beta}_{\text{W}} = \hat{\beta}_{\text{MM}} = \hat{\beta}_{\text{P}} = \hat{\beta}_{\text{P}}$
- As  $\frac{\sigma^2}{T} \times \frac{1}{g_{00}} \rightarrow 0$ ,  $\hat{\beta}_{\text{MM}} \rightarrow \hat{\beta}_{\text{W}}$ , so a mixed model estimates the within effect asymptotically in  $T$ .

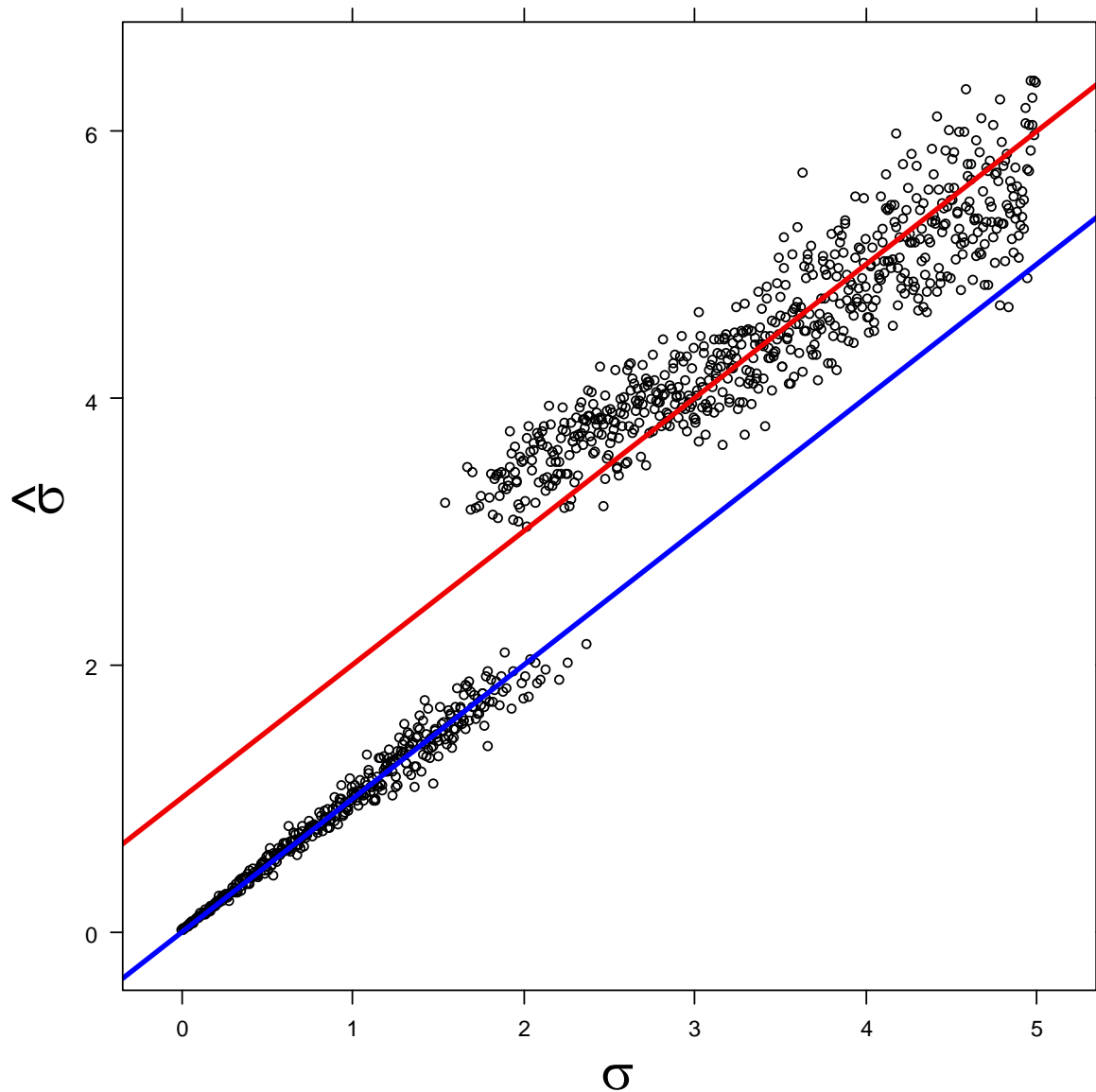


- As  $\frac{\sigma^2}{T} \times \frac{1}{g_{00}} \rightarrow \infty$ ,  $\hat{\beta}_{\text{MM}} \rightarrow \hat{\beta}_{\text{B}}$ . Thus the mixed model estimate fails to control for between-subject confounding factors.

## *A serious a problem? a simulation*



1,000 simulations showing mixed model estimates of slope using the same configuration of Xs with  $\beta_W = -1/2$  and  $\beta_B = 1$ , keeping  $g_{00} = 1/2$  and allowing  $\sigma^2$  to vary from 0.005 to 5



What happened?

As  $\sigma$  gets larger,  
the relatively small value of  $g_{00}$   
is harder to identify  
and  
both sources of variability  
(within-subject and between-  
subject)  
are attributed to  $\sigma$ .

The blue line is the diagonal  
 $\hat{\sigma} = \sigma$  and the equation of the  
red line is  $\hat{\sigma} = \sigma + 1$ .

When  $\hat{g}_{00} \approx 0$ , the between-  
subject relationship is treated as  
if it has very high precision and  
it dominates in forming the  
mixed model estimate.

## ***Splitting age into two variables***

Since age has a within-subject effect that is inconsistent with its between-subject effect we can split it into two variables:

1. Between-subject ‘contextual predictor’: e.g. *age.mean* of each subject (or the starting age), and
2. within-subject predictor:
  - a. *age* itself or
  - b. within-subject residual:  $age.resid = age - age.mean$

So we model:

$$E(y_{it}) = \beta_0 + \beta_{age.mean} age.mean_i + \beta_{age} age_{it} + \delta_{it}$$

or

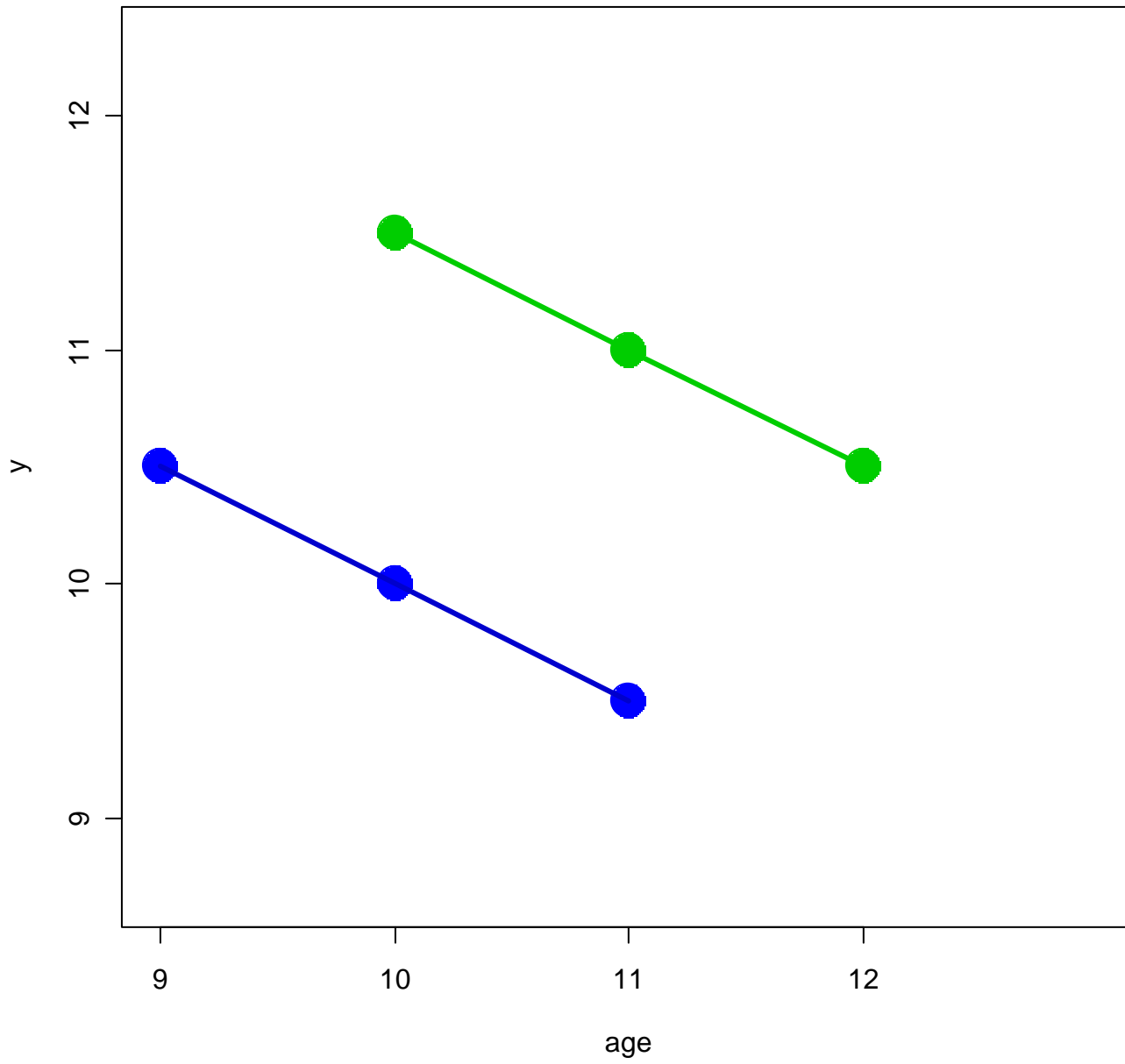
$$E(y_{it}) = \beta_0^* + \beta_{age.mean}^* age.mean_i + \beta_{age.diff}^* age.diff_{it} + \delta_{it}$$

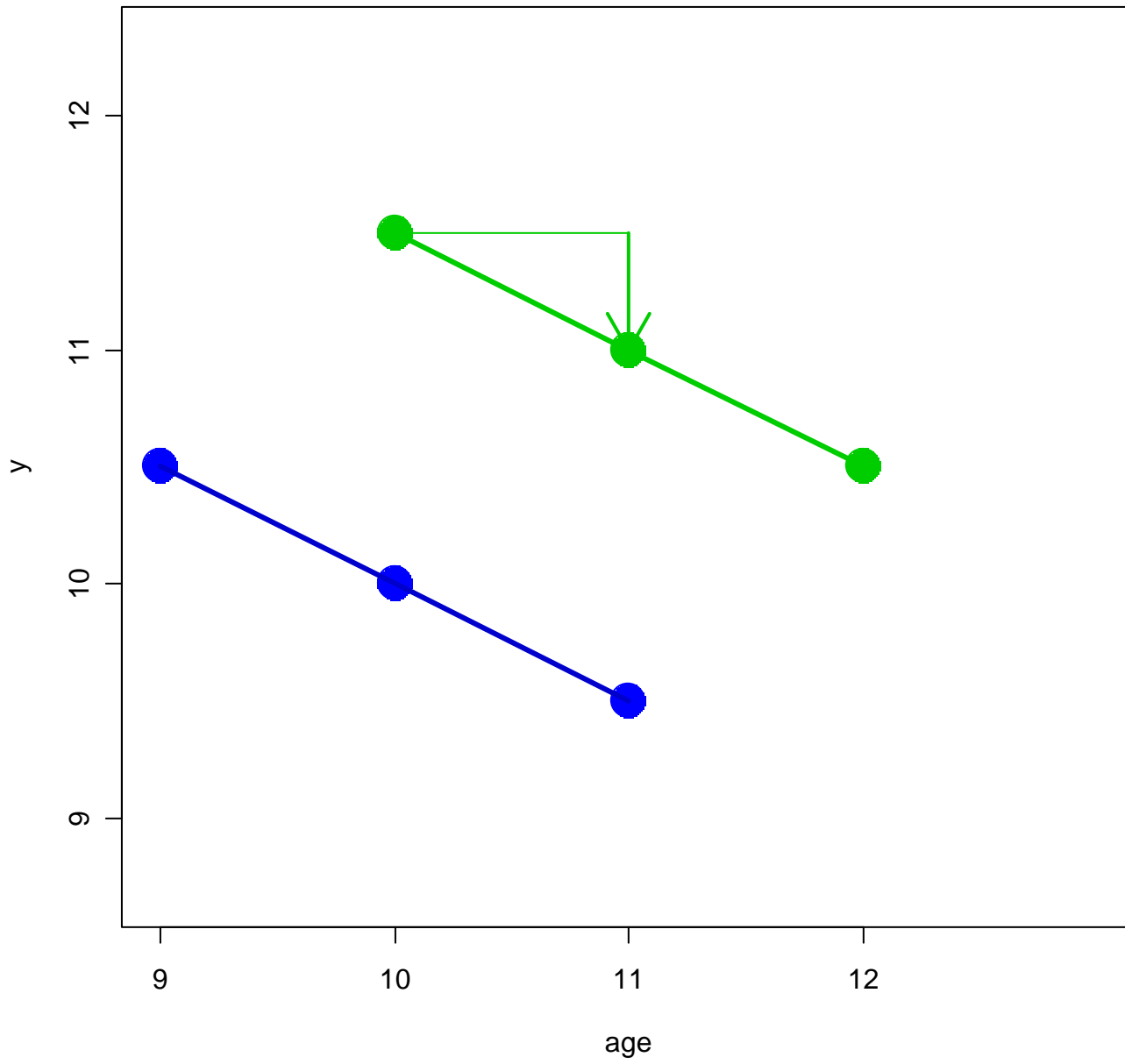
Surprisingly

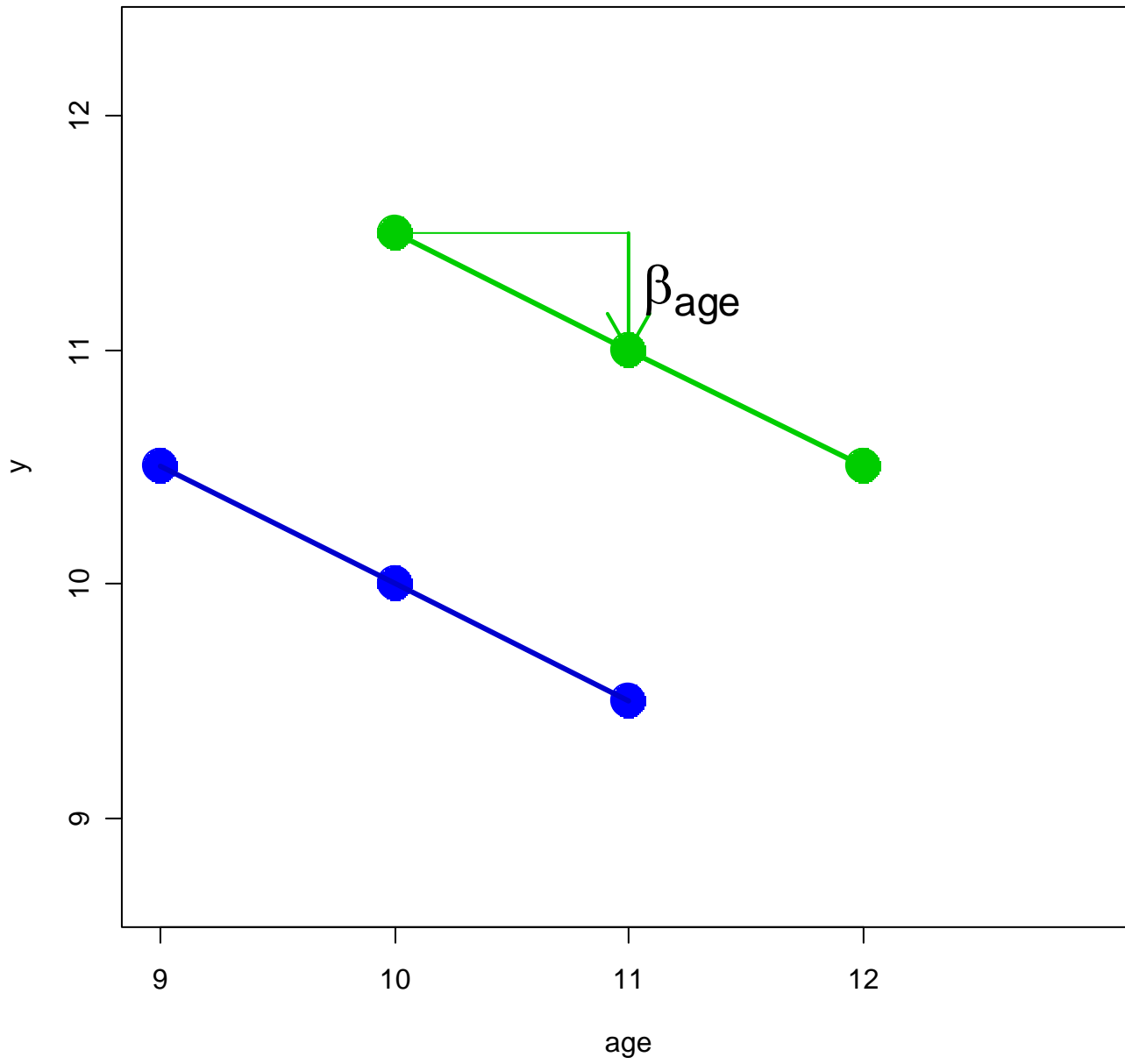
$$\beta_{age.diff}^* = \beta_{age}$$

but

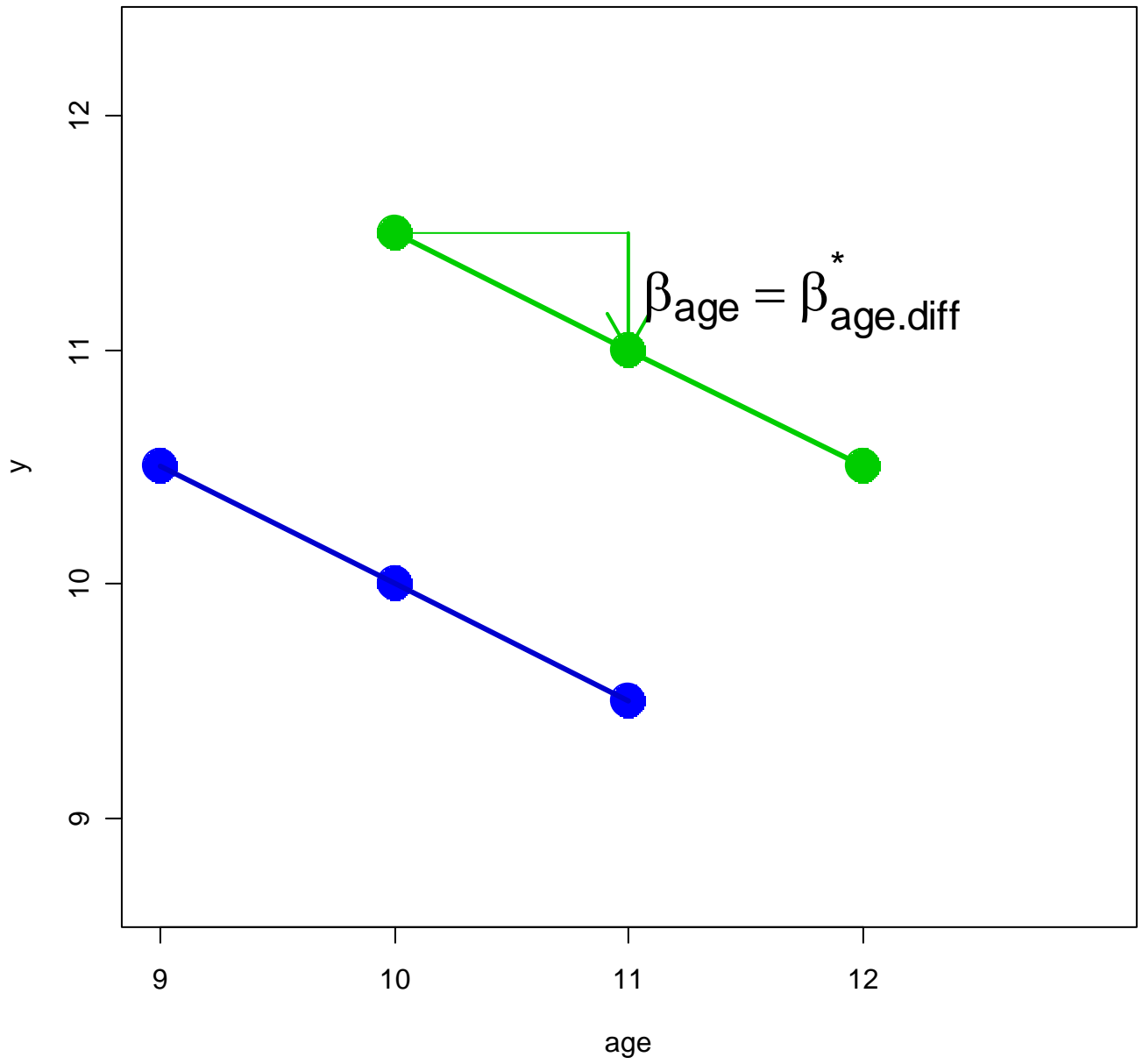
$$\begin{aligned}\beta_{age.mean}^* &= \beta_{age.mean} + \beta_{age.diff}^* \\ &= \beta_{age.mean} + \beta_{age}\end{aligned}$$

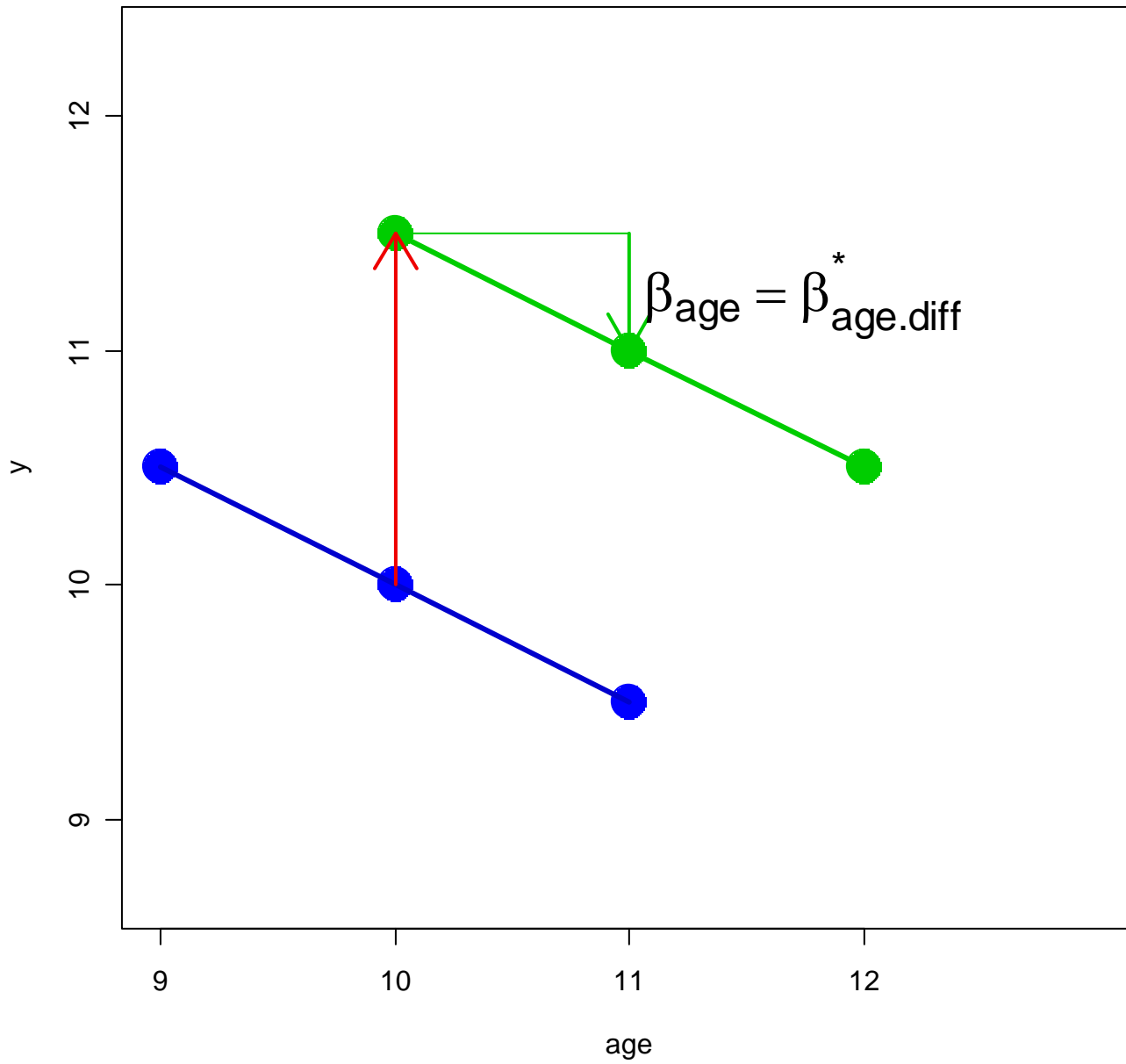


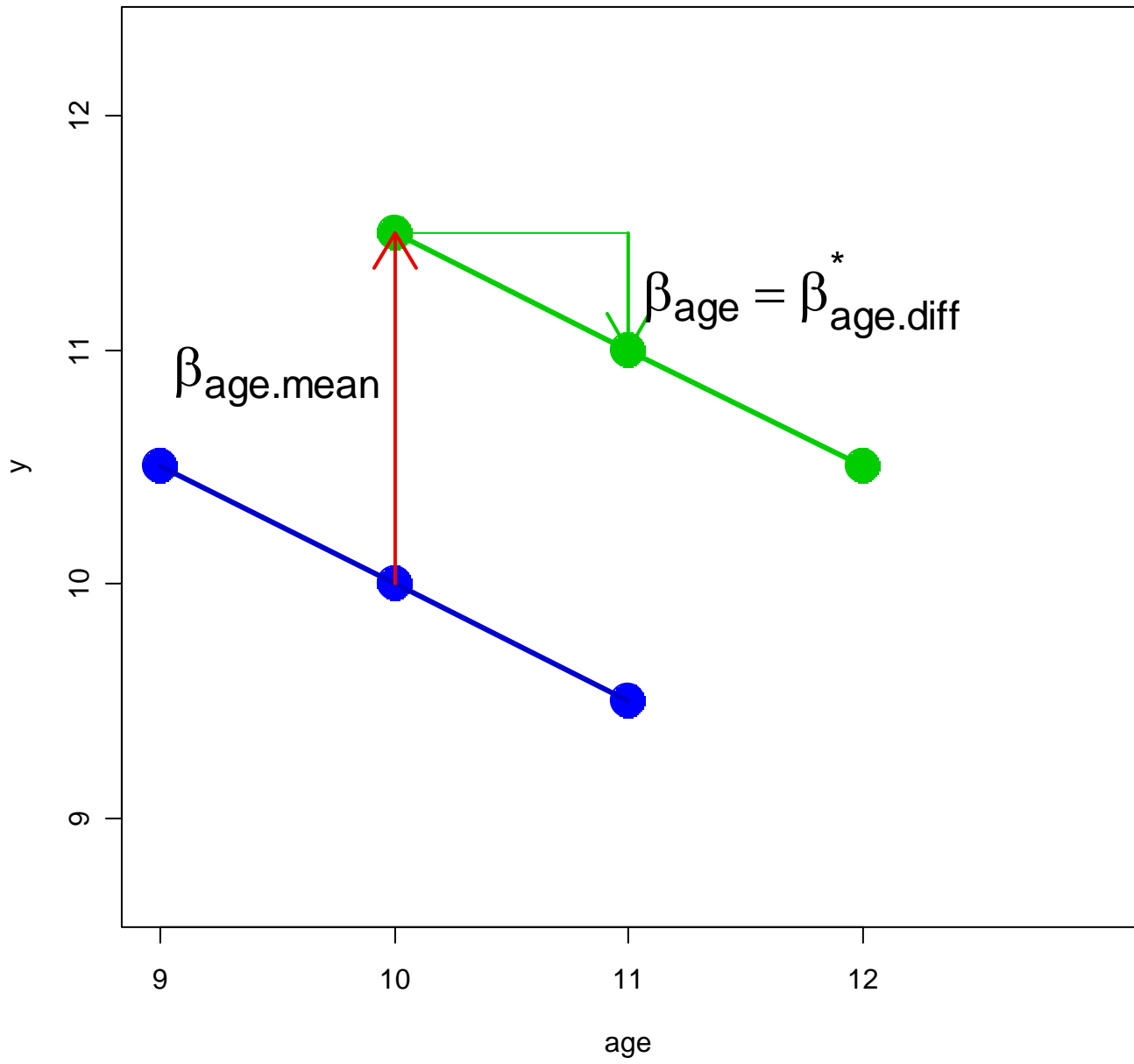


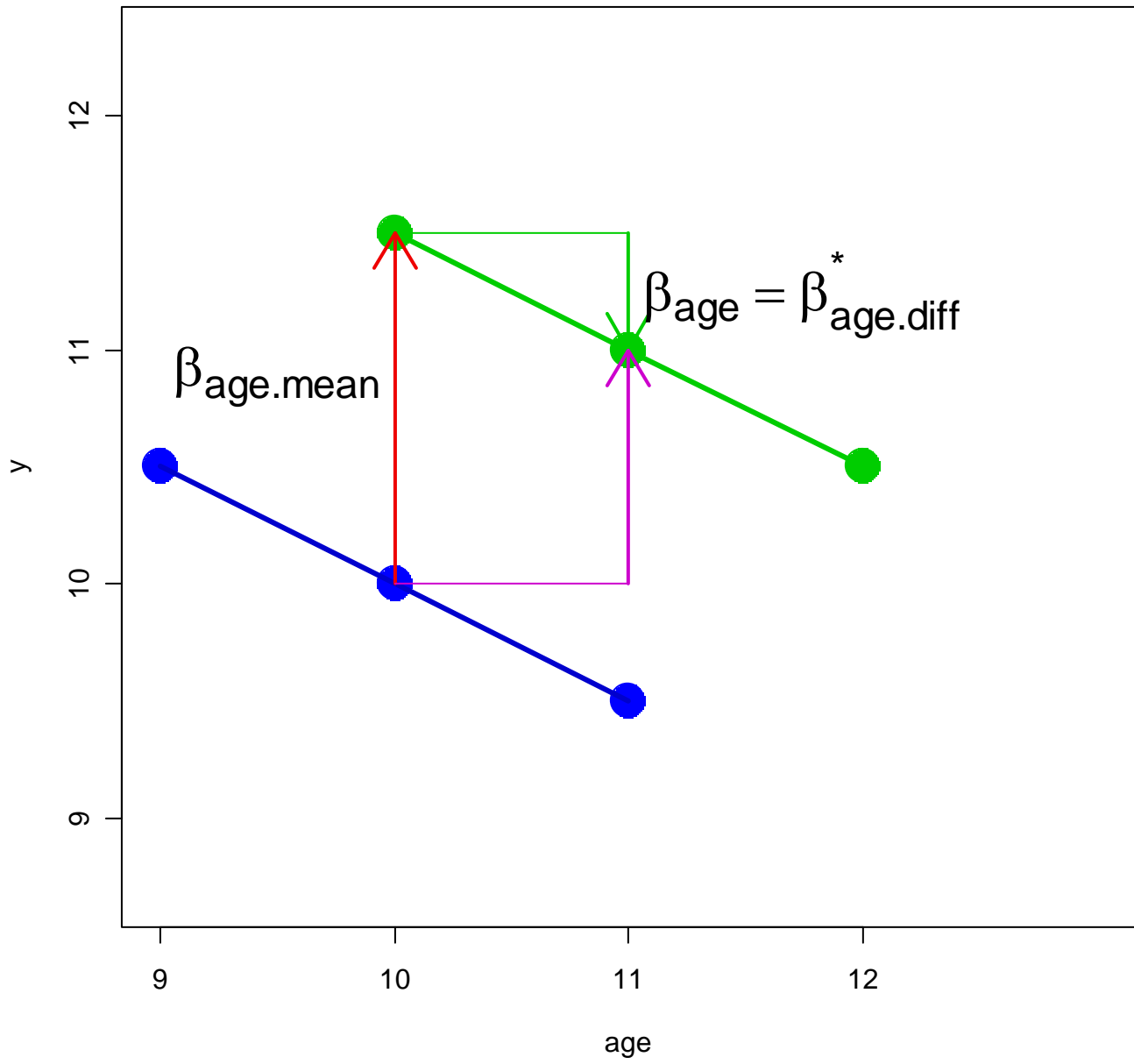


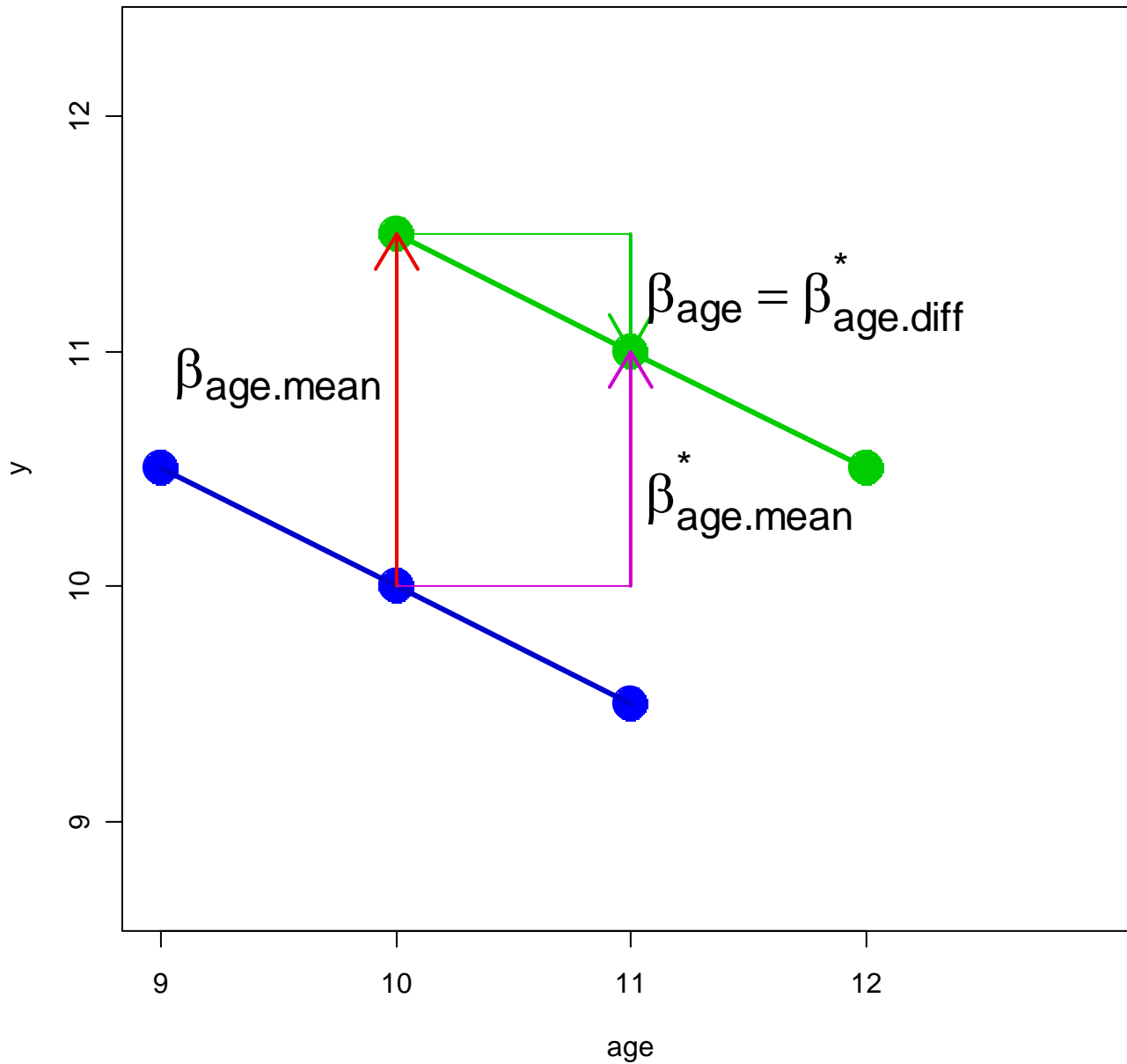












$$\beta_{age.mean}^* = \beta_{age.mean} + \beta_{age}$$

$\beta_{age.mean}^*$  keeps age.diff constant

$\beta_{age.mean}$  keeps age constant

Compositional effect  
 = Contextual effect  
 + Within-subject effect

## ***Using PROC MIXED with a contextual mean***

```
/* Create a data set with subject means */
```

```
PROC MEANS DATA = unbal;  
  VAR age;  
  BY subject;  
  OUTPUT OUT = new MEAN = age_mean;  
RUN;
```

```
PROC SORT DATA = new;  
  BY subject;  
RUN;
```

```
/*
```

```
We now merge the school averages back into the  
main data set and create a variable with each  
subject's deviation, age_diff, in age from their  
average age.
```

```
*/
```

```

DATA unbalm;
  MERGE unbal new;
  BY subject;
  age_diff = age - age_mean;
RUN;

/*
  We are ready to run PROC MIXED on the new variables.
  We add the new inner variable age_diff to the RANDOM
  model.
*/
PROC MIXED DATA = unbalm; /* using age and age_mean */
  CLASS subject SEX;
  MODEL y = age age_mean sex sex*age sex*age_mean
    / SOLUTION DDFM = SATTERTH;
  RANDOM age INTERCEPT / TYPE = FA0(2) SUB = subject ;
RUN;

```

Solution for Fixed Effects

Effect	Sex	Estimate	Standard Error	DF	t Value	Pr >  t
Intercept		25.3832	0.2105	102	120.61	<.0001
age		-0.5468	0.1086	102	-5.04	<.0001
age_mean		1.5386	0.1112	102	13.83	<.0001

Sex	Female	-0.02565	0.3323	102	-0.08	0.9386
Sex	Male	0	.	.	.	.
age*Sex	Female	-0.04850	0.1701	102	-0.29	0.7761
age*Sex	Male	0	.	.	.	.
age_mean*Sex	Female	-0.00188	0.1764	102	-0.01	0.9915
age_mean*Sex	Male	0	.	.	.	.

```

PROC MIXED DATA = unbalm; /* using age_diff and age_mean */
CLASS subject SEX;
MODEL y = age_diff age_mean sex sex*age_diff sex*age_mean
      / SOLUTION DDFM = SATTERTH;
RANDOM age INTERCEPT / TYPE = FA0(2) SUB = subject ;
RUN;

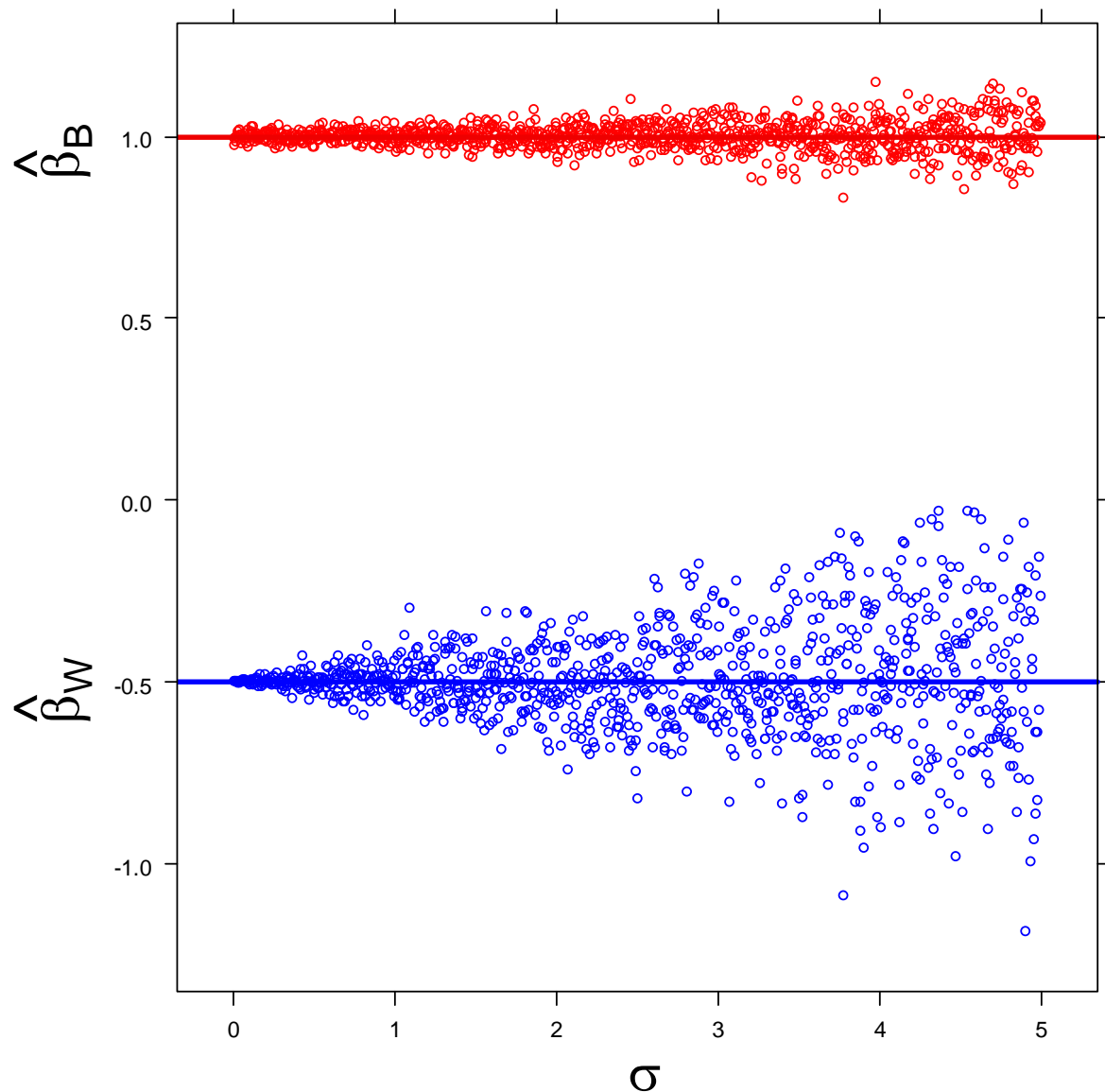
```

Solution for Fixed Effects  
Standard

Effect	Sex	Estimate	Error	DF	t Value	Pr >  t
Intercept		25.3832	0.2105	102	120.61	<.0001
age_diff		-0.5468	0.1086	102	-5.04	<.0001
age_mean		0.9918	0.02407	102	41.21	<.0001
Sex	Female	-0.02565	0.3323	102	-0.08	0.9386
Sex	Male	0	.	.	.	.
age_diff*Sex	Female	-0.04850	0.1701	102	-0.29	0.7761
age_diff*Sex	Male	0	.	.	.	.
age_mean*Sex	Female	-0.05038	0.04660	102	-1.08	0.2822
age_mean*Sex	Male	0	.	.	.	.

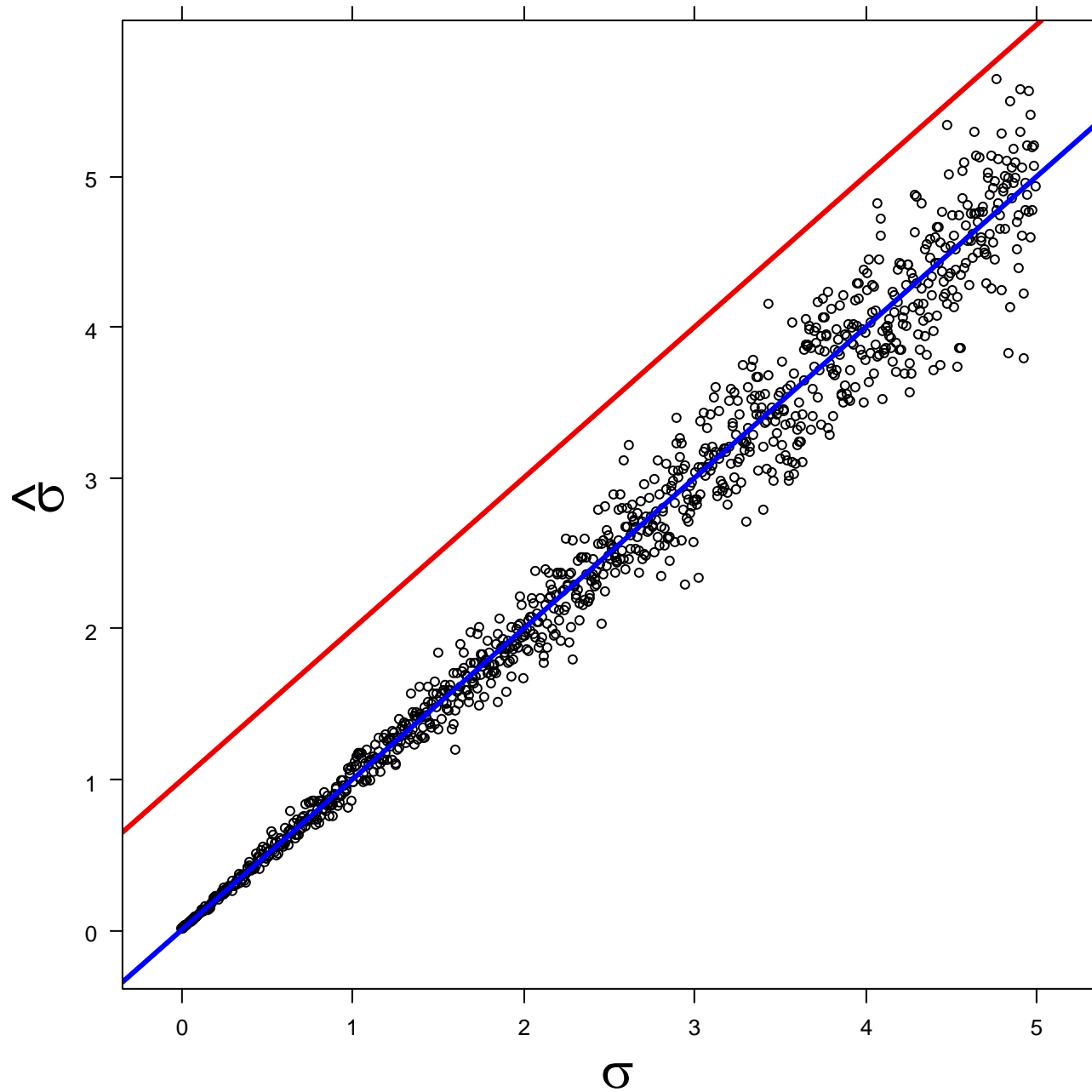


## Simulation Revisited



1,000 simulations using the same models as the earlier simulation, i.e. the same configuration of  $X$ s with  $\beta_W = -1/2$  and  $\beta_B = 1$ , keeping  $g_{00} = 1/2$  and allowing  $\sigma^2$  to vary from 0.005 to 5

Here a mixed model is used with mean age by subject and the within-subject residual of age from mean age.



Including the contextual variable gives better estimates of variance components. The estimate of  $\sigma$  does not eventually include  $g_{00}$

## Power

The best way to carry out power calculations is to simulate. You end up learning about a lot more than power.

Nevertheless, Stephen Raudenbush and colleagues have a nice graphical package available at [Optimal Design Software](#).

## Some links

- There is a very good current bibliography as well as many other resources at the UCLA Academic Technology Services site. Start your visit at [http://www.ats.ucla.edu/stat/sas/topics/repeated\\_measures.htm](http://www.ats.ucla.edu/stat/sas/topics/repeated_measures.htm)
- Another important site is the Centre for Multilevel Modeling, currently at the University of Bristol:  
<http://www.cmm.bristol.ac.uk/learning-training/multilevel-m-support/news.shtml>

## A few books

- Fitzmaurice, Garrett M., Laird, Nan M., Ware, James H. (2004) *Applied Longitudinal Analysis*, Wiley.
- Allison, Paul D. (2005) *Fixed Effects Regression Methods for Longitudinal Data Using SAS*, SAS Institute.
- Littell, Ramon C. et al. (2006) *SAS for Mixed Models* (2<sup>nd</sup> ed.), SAS Institute.
- Singer, Judith D. and Willett, John B. (2003) *Applied Longitudinal Data Analysis : Modeling Change and Event Occurrence*. Oxford University Press.

## Appendix A: Synopsis of SAS commands in PROC MIXED

For a complete list of SAS commands for PROC MIXED, see SAS on-line documentation. The following is a list of the key commands and options. The concepts behind some of these commands have not been covered.

**PROC MIXED:** The following are some of the commonly used options on the PROC statement<sup>4</sup>:

**Input:** **DATA = dsname**

### Output:

**ASYCOV ASYCORR :** asymptotic covariance/correlation of the variance/covariance (**G** and **R**) parameter estimates.

**CL :** confidence limits for above.

**COVTEST :** tests for above.

**RATIO :** produces  $\mathbf{G}/\sigma^2$ .

**LOGNOTE :** for very long runs — so SAS will show signs of life.

**ITDETAILS :** for when things seem to go wrong.

**MMEQ :** matrices of coefficients of “mixed model equations” which are extensions of the OLS “normal” equations.

## Method:

**METHOD = [REML] | ML** : restricted or full likelihood (choose ML if you will compare log-likelihoods between models with different fixed models [i.e. different **MODEL** statements] but **ML** sometimes has trouble converging with near singular **G** matrices).

**EMPIRICAL** : uses robust “sandwich” estimator to estimate variance-covariance of fixed effect estimates. The “sandwich” estimate uses the observed subject-to-subject variance instead of using the inverse Fisher information of the normal model.

## Algorithm:

**CONVF | CONVG | [CONVH]** : choose convergence criterion.

**MAXFUNC=[150] MAXITER=[50]** : maximum number of function evaluations at each iteration and maximum number of iterations.

**NOPROFILE** : treats  $\sigma^2$  like other parameters. e.g. so it can be “held” with the **HOLD** option of the **PARMS** statement.

**CLASS vars** : names categorical variables used in analysis.

**MODEL  $y = x_1 x_2 x_1*x_2$**  : specify fixed model (intercept is included by default)

**Output:**

**S** : show “solution,” i.e. estimated values of parameters.

**CORRB COVB COVBI** : variances/covariances/correlations and inverse covariances of estimates.

**CL** : confidence intervals.

**OUTPREDM = ds1 OUTPRED = ds2** : output datasets with ‘population’ predicted values (no EBLUPS) and ‘cluster-wise’ predicted values (including contribution of EBLUPS).

**[NEW IN VERSION 9] INFLUENCE : optionally INFLUENCE (EFFECT = var);**

**[NEW IN VERSION 9] RESIDUAL :**

**Method:**

**DDFM = SATTERTH** : uses Satterthwaite approximation for denominator degrees of freedom. A newer method **KENWARDROGER** might be better.

**NOINT** : no intercept.

**RANDOM INT x1 x2** : specify random model.

### Output:

**G GC GCORR GI** : print the T matrix, its Cholesky factor (useful to determine rank), its covariances as correlations, the inverse of **G**. Note that the Cholesky factor can be useful to detect rank.

**s** : EBLUPS. Beware if many clusters.

**V VCORR** : within-cluster variance.

### Method:

**SUB = var** : variable identifying cluster. If omitted analysis is done as if for a single cluster. Note that the data set should be sorted on this variable.

**GRP = var** : variable identifying groups with possibly different **G** matrices.

**TYPE = [VC] | UN | FA0(q)** : There are many more variance-covariance structures but these three are generally suitable for random effects. **VC** is appropriate for homoscedastic categorical variables. **UN** and **FA0(q)** where  $q$  is the order of **G** both generate variance matrices with no constraints except that matrices fit with **UN** need not be non-negative definite. **FA0(q)** uses a Cholesky parametrization that might be numerically more stable. Some parameters in both forms can be constrained with the **PARMS** statement.



**REPEATED** : specifies within cluster variance structure. It is especially useful to model autocorrelation for longitudinal data.

**Output:**

**R RC RCI RCORR RI** : various reports on within cluster variance.

**Method:**

**SUB = var** : same as **RANDOM** statement.

**GRP = var** : variable identifying groups with potentially different values of **R**.

**LOCAL = POM(data set) | EXP( var )** : allows variance to vary as a power of  $E(Y)$  or as a function of predictors.

**TYPE = ANTE(1) | AR(1) | ARH(1) | ARMA(1,1) | UN |**

**SP(POW)(var)** : default is  $\sigma^2\mathbf{I}$ . Other options are discussed elsewhere.

They are used primarily for longitudinal data or to fit multilevel multivariate models.

**ESTIMATE CONTRAST LSMEANS** : Used to estimate  $E(Y | \dots)$  or differences. See SAS documentation and notes above.

**PARMS** : sets initial values for **G** and **R** parameters.

**Method:**

**HOLD = i, j** : keeps parameters in position  $i, j$  fixed at initial value.

## Appendix B: PROC MIXED all-dressed

This is an example calling PROC MIXED with many options. The model used has random effects for two inner variables: SES and FEMALE. A constrained Cholesky parametrization is used to fit a covariance of 0 between the random effects for SES and FEMALE. A self-extracting data set for this example can be found at

<http://www.math.yorku.ca/~georges/Courses/Repeated/hs.exe>

```
ODS HTML;      /* new for diagnostics in Version 9 */
ODS GRAPHICS ON; /* new for diagnostics in Version 9 */
PROC MIXED DATA = hs ASYCOV ASYCORR CL COVTEST;
CLASS SCHOOL;
MODEL mathach = SES SECTOR FEMALE SES*SECTOR /
    S CORRB COVB COVBI CL
    DDFM = SATTERTH
    INFLUENCE
    INFLUENCE ( EFFECT = SCHOOL )
    RESIDUAL
    OUTPREDM = hsm OUTPRED = hsc ;
RANDOM SES FEMALE INT / /* note: INT last */
SUB = SCHOOL
```

```
TYPE = FA0(3)
S
G GC GCORR GI
V VCORR; /* big and not very interesting here */
ESTIMATE 'Cath-Pub | low SES'
    SECTOR 1 SES*SECTOR -2 / CL E ;
ESTIMATE 'Cath-Pub | high SES'
    SECTOR 1 SES*SECTOR +2 / CL E ;
PARMS (1) (0) (1) (1) (1) (1) (1)/ HOLD = 2;
RUN;
```

## Appendix C: Reinterpreting weights

The mixed model estimate using a random intercept model can be seen either as a weighted combination of  $\hat{\beta}_B$  and  $\hat{\beta}_W$  or of  $\hat{\beta}_P$  and  $\hat{\beta}_W$

$$\begin{aligned}
 \hat{\beta}_{MM} &= \left( \frac{\sigma^2 / T}{\sigma^2 / T + g_{00}} W_B + W_W \right)^{-1} \left( \frac{\sigma^2 / T}{\sigma^2 / T + g_{00}} W_B \hat{\beta}_B + W_W \hat{\beta}_W \right) \\
 &= \left( \left( \frac{\sigma^2}{T} + g_{00} \right)^{-1} W_B + \left( \frac{\sigma^2}{T} \right)^{-1} W_W \right)^{-1} \left( \left( \frac{\sigma^2}{T} + g_{00} \right)^{-1} W_B \hat{\beta}_B + \left( \frac{\sigma^2}{T} \right)^{-1} W_W \hat{\beta}_W \right) \\
 &= \left( (g_{00})^{-1} (W_B + W_W) + \left( \frac{\sigma^2}{T} \right)^{-1} W_W \right)^{-1} \left( (g_{00})^{-1} (W_B + W_W) \hat{\beta}_P + \left( \frac{\sigma^2}{T} \right)^{-1} W_W \hat{\beta}_W \right)
 \end{aligned}$$