



# 2AMS10: Longitudinal Data Analysis 2022-2023 Linear Mixed Models

October 6, 2022

Edwin van den Heuvel  
[e.r.v.d.heuvel@tue.nl](mailto:e.r.v.d.heuvel@tue.nl)  
Professor in Statistics

# Linear Mixed Models

## Contents

- Introduction
- General formulation
  - Statistical model
  - SAS procedure MIXED
- Subject-specific models
  - Polynomial time profiles
  - Fractional polynomials
- Marginal models
  - Categorical time points
  - Continuous time profiles
- Model selection
  - General strategy
  - Covariance structures
  - Information criteria

# Introduction

## *History of longitudinal data analysis*

### Data structure:

- The set of outcomes on unit  $i$  is  
$$\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{in_i})^T$$
  - With  $y_{ij}$  measured at time  $t_{ij}$
  - And  $i \in \{1, 2, \dots, m\}$
- Covariates (**continuous or categorical**)  
$$\mathbf{x}_{ri} = (x_{ri1}, x_{ri2}, \dots, x_{rin_i})^T$$
  - With  $r = 1, 2, \dots, p$
  - Covariates do not have to be time-varying
- Data structure is like repeated outcomes on one unit
  - Thus, ANOVA models seemed natural for the analysis of longitudinal data

### First simple analysis model:

- An extension of the regression model  
$$y_{ij} = \sum_{r=1}^p \beta_r x_{ijr} + b_i + e_{ij}$$
  - Adding the random effect  $b_i$  to the regression model introduce correlation
  - Time could be included as a regressor
  - Covariates can enter the model in different ways (transformations and interactions)
- Model has several restrictions:
  - Correlation over time is constant
  - Relations between covariates and outcome is linear in the parameters
  - It does not describe individual trajectories

# Introduction

## *History of longitudinal data analysis*

### Two-step approach:

- In growth curve analysis, individual regression curves were applied
$$y_{ij} = \sum_{r=1}^p \beta_{ir} t_{ij}^{r-1} + e_{ij}$$
- Then the regression coefficients were further analyzed (like features)
- This approach was also applied to other longitudinal data (sometimes referred to as NIH approach)
- This two-step approach is the birth of a more formal analysis where coefficients are treated as random

### Restrictions two-step approach:

- The first step can only include time-varying covariates
  - Thus time-constant variables can only be introduced in the second step
- The two-step approach often imposed restrictive modeling designs
- The two-step approach does not consider the potential imbalances in collected data
  - Some individuals being more precisely modeled than other individuals

# Introduction

## *Case study: Primary school children*

### Study design information:

- 4106 children from primary schools (~ 11 years)
- 216 randomly selected schools (1 class/school)
- Classes can be multi-grade or single grade
- Measured both in grade 7 and in grade 8

### Collected information:

- Language and arithmetic tests in grades 7 and 8
- Verbal and performal IQ (obtained in grade 7)
- Sex of child and minority status (born outside industrialized countries)
- Social economic status of child's family
- Class size (not all children participated)

### Research questions:

- Understanding cognitive development over time
  - How much is it changed?
  - Differences between certain subgroups (sex, minority)?
- How much are language and arithmetic scores affected by intelligence?
- What is the variation between children and schools/classes?

# Introduction

## *Case study: Primary school children*

	CLASS	CHILD	COMBI	SIZE	SSES	GIRL	MINORITY	SITTERS	CSES	IQV	IQP	PRE_LANG	POST_LANG	PRE_ARITH	POST_ARITH	DIFF_A	
1		180	1	0	29	11	1	1	0	10.5	7.3333333333	33	-	10	12	2	
2		180	2	0	29	11	1	1	0	-	14	14.3333333333	44	50	18	30	12
3		180	3	0	29	11	0	0	0	23	15	12.329999924	36	46	14	24	10
4		180	4	0	29	11	0	1	0	10	14.5	10	36	45	12	19	7
5		180	5	0	29	11	0	0	0	15	9.5	11	33	33	10	24	14
6		180	6	0	29	11	0	0	0	23	11	10	29	46	13	26	13
7		180	7	0	29	11	0	0	0	10	8	6.6659998894	19	20	8	9	1
8		180	8	0	29	11	0	1	0	10	9.5	9	22	30	8	13	5
9		180	9	0	29	11	0	1	0	23	9.5	10.329999924	20	30	7	13	6
10		180	10	0	29	11	0	0	0	10	13	14.329999924	44	57	17	30	13
11		180	11	0	29	11	0	1	1	13	9.5	8.6660003662	34	36	10	23	13
12		180	12	0	29	11	0	1	0	15	11	15	31	36	14	22	8
13		180	13	0	29	11	1	1	0	10	5.5	9	18	29	11	19	8
14		180	14	0	29	11	0	0	0	18	14	9.3330001831	36	40	10	23	13
15		180	15	0	29	11	0	1	0	15	9	9	31	41	10	18	8
16		180	16	0	29	11	0	1	0	20	10.5	13	34	47	12	22	10
17		180	17	0	29	11	0	1	0	10	10	8.6660003662	31	33	9	15	6
18		180	18	0	29	11	0	0	0	20	11	12.329999924	32	37	13	21	8
19		180	19	0	29	11	1	0	0	13	6.5	7.3330001831	23	29	9	13	4
20		180	20	0	29	11	1	1	1	10	4	9.3330001831	20	26	9	12	3
21		180	21	0	29	11	1	1	0	10	11	6.6659998894	27	37	7	11	4
22		180	22	0	29	11	1	1	0	15	11	11	31	40	16	23	7
23		180	23	0	29	11	1	0	1	10	11	11	24	27	16	17	1
24		180	24	0	29	11	1	0	0	13	12.5	10	35	43	9	16	7
25		180	25	0	29	11	1	1	0	10	11.5	11	32	39	14	24	10
26		180	26	0	29	11	1	1	0	10	10	6.6659998894	25	21	12	16	4
27		180	27	0	29	11	1	1	0	10	10	9.3330001831	33	42	17	11	-6
28		280	28	1	19	11	1	0	1	15	11.5	7.3330001831	22	21	3	6	3
29		280	29	1	19	11	0	0	0	15	10.5	9.3330001831	33	27	4	8	4
30		280	30	1	19	11	0	0	0	20	8	9.3330001831	22	16	11	9	-2
31		280	31	1	19	11	1	0	1	15	7.5	7.6659998894	25	31	9	9	0
32		280	32	1	19	11	0	1	1	10	7.5	11.329999924	27	21	11	18	7

# Introduction

## *Case study: Longitudinal clinical trial*

### Study design information:

- 24 participants
- 3 treatments to suppress pain: P, G, and R
  - With P being placebo
  - Participants were randomized to treatments
- Four time points (120, 135, 150, 165 minutes)

### Collected information:

- At five vertical lines on the arm pain is induced with more pain going up the arm
- Different clinical outcomes
  - Central line: a distance on the arm where pain is felt
  - AUC: an area on the arm where pain is felt

### Research questions:

- Do treatments G and/or R reduce the pain experience compared to Placebo
  - At a single time point
  - Progressed over time
- Is there a difference in pain experience between treatments G and R
  - At a single time point
  - Progressed over time

# Introduction

## *Case study: Longitudinal clinical trial*

	SUBJECT	TRT	TIME	CLINE	AREA
1		13 R	120	10.5	62.932503526
2		13 R	135	7.8	34.319427625
3		13 R	150	7.5	41.896076785
4		13 R	165	5.3	28.171134162
5		14 R	120	5.5	34.825008973
6		14 R	135	5	31.112698372
7		14 R	150	6	28.54943629
8		14 R	165	4.5	25.19067908
9		15 Placebo	120	10.9	14.379016395
10		15 Placebo	135	7	21.432406538
11		15 Placebo	150	7.2	52.021845892
12		15 Placebo	165	6.2	23.037538931
13		16 G	120	13	43.663843738
14		16 G	135	3.5	10.253048327
15		16 G	150	11	25.986174209
16		16 G	165	2	4.2426406871

# Introduction

## *Case study: Longitudinal clinical trial*

Time	Treatment P	
	Ave	St. Dev
120	9.06	4.04
135	9.74	1.44
150	9.65	1.89
165	8.45	2.55

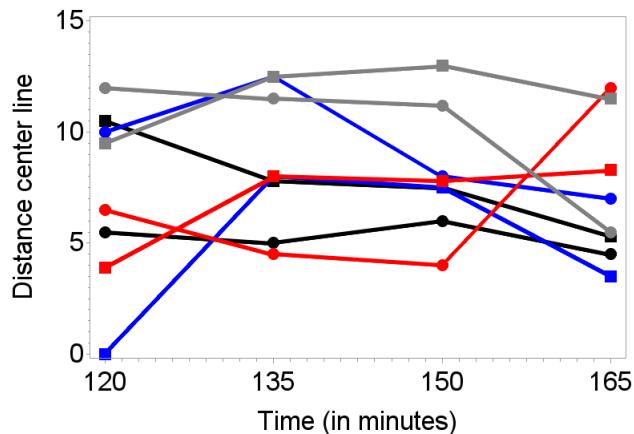
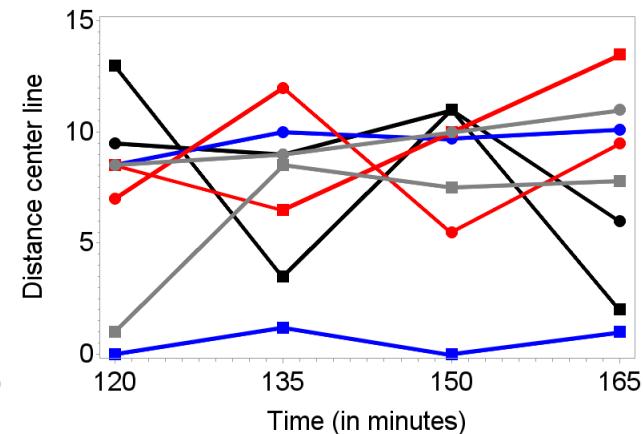
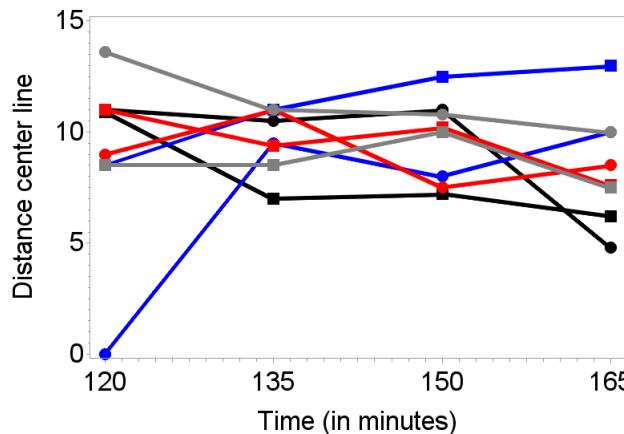
Treatment P

Time	Treatment G	
	Ave	St. Dev
120	7.00	4.38
135	7.46	3.56
150	8.09	3.77
165	7.61	4.37

Treatment G

Time	Treatment R	
	Ave	St. Dev
120	7.24	4.02
135	8.73	3.16
150	8.13	2.82
165	7.20	3.17

Treatment R



# Introduction

## *Case study: Growth data*

### Study design information:

- 714 children from 34 different area's
- 11 observations per child
  - Period: Approximately from 2 month to 13.5 years

### Collected information:

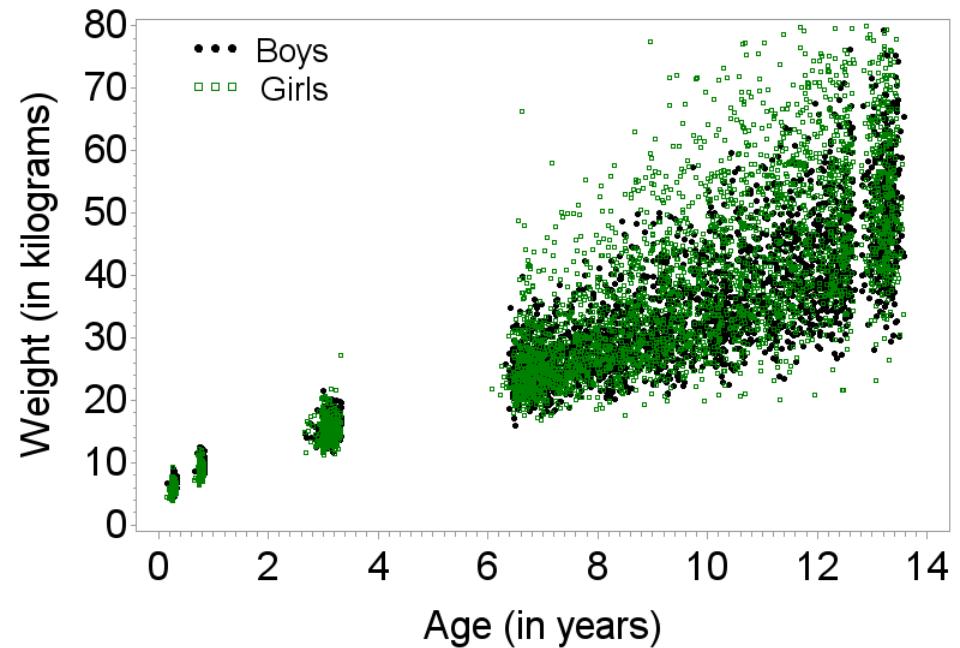
- Response: weight
- Demographic information
  - Age of child in years
  - Sex of child [0 = boys; 1 = girls]
  - Education parents [1 = low; 2 = medium; 3 = high]
  - Smoking [0 = no; 1 = passive; 2 = maternal]
  - BMI mother and father
  - Age mother [1: <25; 2: 25 – 29; 3: 30 – 34; 4: ≥ 35]

### Research questions:

- Can we adequately describe weight as function of age?
- Can we asses normal growth curves?
  - What is a normal range
  - What is a normal shape
- Is growth affected by certain factors
  - Sex
  - Smoking
  - BMI of parents
  - Education

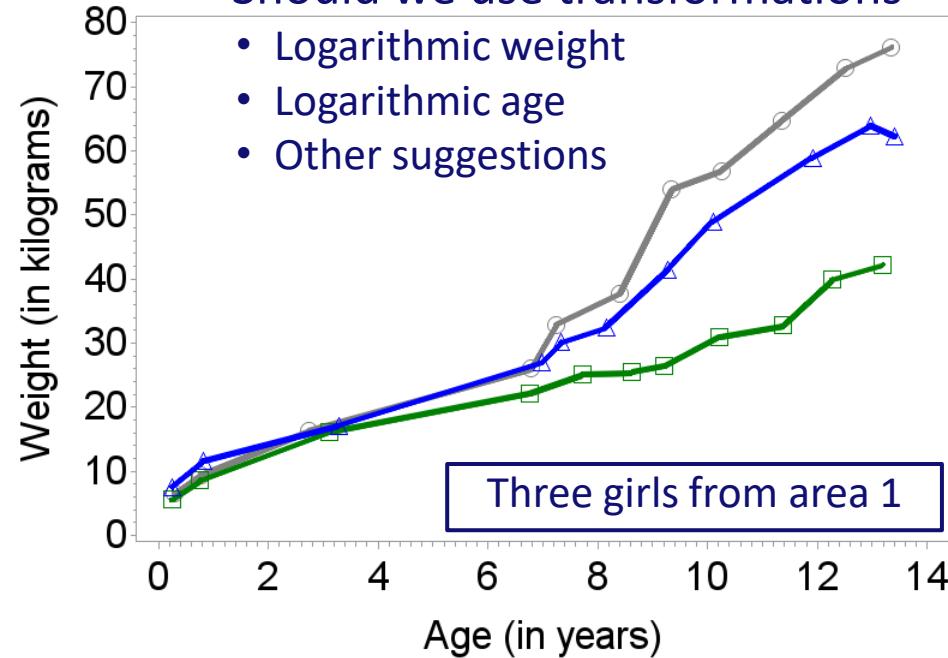
# Introduction

## *Case study: Growth data*



### Analysis considerations:

- Functional form is not obvious
- Should we use transformations



# General formulation

## Statistical model

### Data single unit:

- Let  $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{in_i})^T$  be the set of repeated outcomes on unit  $i$
- The linear mixed model is

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

- With  $\mathbf{X}_i$  an observed matrix of covariates (**continuous** or **categorical**) for unit  $i$
- With  $\boldsymbol{\beta}$  a set of unknown fixed parameters (including an intercept)
- With  $\mathbf{Z}_i$  a known matrix describing the structure of the random effects
- With  $\mathbf{b}_i$  the random effects for unit  $i$
- With  $\mathbf{e}_i$  the residual of unit  $i$

### Distributional assumptions:

- Both random effects and residuals are multivariate normally distributed
  - $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{G}_i)$
  - $\mathbf{e}_i \sim N(\mathbf{0}, \mathbf{R}_i)$
- With  $\mathbf{b}_i$  and  $\mathbf{e}_i$  independent
- With  $\mathbf{G}_i$  and  $\mathbf{R}_i$  any covariance matrix
- Matrices across subjects are only different in size due to sample size not in structure
- For ANOVA models we typically assume independence
  - $\mathbf{G}_i$  is a diagonal matrix
  - $\mathbf{R}_i$  is a diagonal matrix

# General formulation

## Statistical model

### Distribution of $y_i$ :

- First two moments of  $y_i$

- **Mean:**

$$\mathbb{E}(y_i) = \mathbb{E}[X_i\beta + Z_i b_i + e_i] = X_i\beta$$

- Like linear regression

- **Variance:**

$$\begin{aligned}\text{VAR}(y_i) &= \mathbb{E}[Z_i b_i + e_i][Z_i b_i + e_i]^T \\ &= \mathbb{E}[Z_i b_i b_i^T Z_i^T] + \mathbb{E}[e_i e_i^T] \\ &= Z_i G_i Z_i^T + R_i \equiv V_i\end{aligned}$$

- Since  $b_i$  and  $e_i$  are both normally distributed,  $y_i$  is normally distributed  
 $y_i \sim N(X_i\beta, V_i)$
- Units are assumed independent

### Data all units:

- Full data:  $y = (y_1, y_2, \dots, y_m)^T$

- The full linear mixed model is

$$y = X\beta + Zb + e$$

- With  $X$  and  $Z$  given by

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_m \end{pmatrix}, \quad Z = \begin{pmatrix} Z_1 \\ \vdots \\ Z_m \end{pmatrix}$$

- With  $b$  and  $e$  given by

$$b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}, \quad e = \begin{pmatrix} e_1 \\ \vdots \\ e_m \end{pmatrix}$$

- With  $\beta$  being the same for all units

# General formulation

## Statistical model

**Subject-specific models:** to describe individual units as good as possible

- **Advantage:** Knowledge on individuals
- **Disadvantage:** Requires several modeling assumptions that are difficult to verify
- Modeling focus:
  - **$G$**  matrix: Using extensive random variables to describe individual units
  - **$X$**  matrix: Using individual characteristics to define subgroups of units
  - **$R$**  matrix: Kept as simple as possible to maintain model identifiability [either identity matrix or autoregressive]

**Marginal models:** to describe the population as good as possible

- **Advantage:** Less sensitive to assumptions and direct interpretation to the population
- **Disadvantage:** No knowledge on units
- Modeling focus:
  - **$G$**  matrix: Typically, not used at all
  - **$X$**  matrix: Using individual characteristics to define subgroups of units in population
  - **$R$**  matrix: Used extensively to describe correlations between units to make sure we obtain proper standard errors

# General formulation

## Statistical model

### Example ANOVA model:

- Crossed two-way mixed effects model  
$$y_{hij} = \mu + \alpha_h + a_i + (\alpha a)_{hi} + e_{hij}$$
- With  $y_{hij}$  outcome of child  $j \in \{1, \dots, J_{hi}\}$  in class  $i \in \{1, \dots, I_h\}$  with sex  $h \in \{1, \dots, H\}$
- With  $\mu$  the overall mean outcome
- With  $\alpha_h$  the effect of sex ( $\alpha_H = 0$  for model identifiability)
- With  $a_i \sim N(0, \sigma_C^2)$  the effect of class
- With  $(\alpha a)_{hi} \sim N(0, \sigma_{SC}^2)$  the interaction effect of sex and class
- With  $e_{hij} \sim N(0, \sigma_R^2)$  the residual
- With all random terms independent

### Correlation structure:

- ANOVA model has only three correlation coefficients
  - Within class and sex  
$$\text{CORR}(y_{hij_1}, y_{hij_2}) = \frac{\sigma_C^2 + \sigma_{SC}^2}{\sigma_C^2 + \sigma_{SC}^2 + \sigma_R^2}$$
  - Within class for different sex  
$$\text{CORR}(y_{1ij_1}, y_{2ij_2}) = \frac{\sigma_C^2}{\sigma_C^2 + \sigma_{SC}^2 + \sigma_R^2}$$
  - Between children from different classes (irrespective of same or different sex)  
$$\text{CORR}(y_{hi_1j_1}, y_{hi_2j_2}) = 0$$

# General formulation

## Statistical model

Example ANOVA model:  $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{in_i})^T$  results for class  $i$

$$\begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{ik_i} \\ y_{ik_i+1} \\ \vdots \\ y_{in_i} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \end{pmatrix} + \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_i \\ (\alpha a)_{i1} \\ (\alpha a)_{i2} \end{pmatrix} + \begin{pmatrix} e_{i1} \\ e_{i2} \\ \vdots \\ e_{ik_i} \\ e_{ik_i+1} \\ \vdots \\ e_{in_i} \end{pmatrix}$$

- With first  $k_i$  observations for boys and second  $n_i - k_i$  observations for girls
- Two fixed effects  $\beta = (\mu, \alpha_1)^T$  and three random effects  $\mathbf{b}_i = (a_i, (\alpha a)_{i1}, (\alpha a)_{i2})^T$
- With  $\mathbf{G} = \text{VAR}(\mathbf{b}_i) = \begin{pmatrix} \sigma_C^2 & 0 & 0 \\ 0 & \sigma_{SC}^2 & 0 \\ 0 & 0 & \sigma_{SC}^2 \end{pmatrix}$  and  $\mathbf{R} = \text{VAR}(\mathbf{e}_i) = \begin{pmatrix} \sigma_R^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_R^2 \end{pmatrix}$

# General formulation

## Statistical model

### Example ANOVA model:

- The mean of  $\mathbf{y}_i$  is

$$\mathbb{E}(\mathbf{y}_i) = \begin{pmatrix} \boldsymbol{\mu} + \boldsymbol{\alpha}_1 \\ \boldsymbol{\mu} \end{pmatrix}$$

- With  $\boldsymbol{\mu} + \boldsymbol{\alpha}_1$  a vector of length  $k_i$
- With  $\boldsymbol{\mu}$  a vector of length  $n_i - k_i$
- Variance of  $\mathbf{y}_i$  is an  $n_i \times n_i$  matrix

$$\mathbf{V}_i = \begin{pmatrix} \mathbf{V}_{i11} & \mathbf{V}_{i12} \\ \mathbf{V}_{i21} & \mathbf{V}_{i22} \end{pmatrix}$$

- With  $\mathbf{V}_{i11}$  a  $k_i \times k_i$  matrix
- With  $\mathbf{V}_{i12} = \mathbf{V}_{i21}^T$  a  $k_i \times (n_i - k_i)$  matrix
- With  $\mathbf{V}_{i22}$  a  $(n_i - k_i) \times (n_i - k_i)$  matrix

- The submatrices are

$$\mathbf{V}_{i11} = \begin{pmatrix} \sigma_{\text{TOT}}^2 & \sigma_C^2 + \sigma_{SC}^2 & \sigma_C^2 + \sigma_{SC}^2 \\ \sigma_C^2 + \sigma_{SC}^2 & \ddots & \sigma_C^2 + \sigma_{SC}^2 \\ \sigma_C^2 + \sigma_{SC}^2 & \sigma_C^2 + \sigma_{SC}^2 & \sigma_{\text{TOT}}^2 \end{pmatrix}$$

$$\mathbf{V}_{i22} = \begin{pmatrix} \sigma_{\text{TOT}}^2 & \sigma_C^2 + \sigma_{SC}^2 & \sigma_C^2 + \sigma_{SC}^2 \\ \sigma_C^2 + \sigma_{SC}^2 & \ddots & \sigma_C^2 + \sigma_{SC}^2 \\ \sigma_C^2 + \sigma_{SC}^2 & \sigma_C^2 + \sigma_{SC}^2 & \sigma_{\text{TOT}}^2 \end{pmatrix}$$

$$\mathbf{V}_{i12} = \begin{pmatrix} \sigma_C^2 & \dots & \sigma_C^2 \\ \vdots & \ddots & \vdots \\ \sigma_C^2 & \dots & \sigma_C^2 \end{pmatrix}$$

- With  $\sigma_{\text{TOT}}^2 = \sigma_C^2 + \sigma_{SC}^2 + \sigma_R^2$

# General formulation

## Statistical model

### Alternative mixed model:

- Model:  $y_{hij} = \mu + \alpha_h + (\alpha a)_{hi} + e_{hij}$ 
  - No random effect of class anymore
  - With  $((\alpha a)_{i1}, (\alpha a)_{i2})$  bivariate normal
$$\begin{pmatrix} (\alpha a)_{i1} \\ (\alpha a)_{i2} \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right)$$
- Interpretation of model parameters:
  - $\sigma_h^2$  is the variability between classes for children with sex  $h$
  - $\rho$  is a correlation coefficient related to children within class of different sex
  - $\sigma_R^2$  is the variability between children within class of the same sex

### Correlation structure:

- Four correlation coefficients
  - Within class and for sex  $h$ 
$$\text{CORR}(y_{hij_1}, y_{hij_2}) = \frac{\sigma_h^2}{\sigma_h^2 + \sigma_R^2}$$
  - Within class between different sexes
$$\text{CORR}(y_{1ij_1}, y_{2ij_2}) = \frac{\rho\sigma_1\sigma_2}{\sqrt{\sigma_1^2 + \sigma_R^2}\sqrt{\sigma_2^2 + \sigma_R^2}}$$
  - Between children from different classes (irrespective of same or different sex)
$$\text{CORR}(y_{hi_1j_1}, y_{hi_2j_2}) = 0$$
- Correlation structure deviates from correlations from ANOVA models

# General formulation

## Statistical model

Alternative mixed model:  $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{in_i})^T$  results for class  $i$

$$\begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{ik_i} \\ y_{ik_i+1} \\ \vdots \\ y_{in_i} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix} \begin{pmatrix} (\alpha a)_{i1} \\ (\alpha a)_{i2} \end{pmatrix} + \begin{pmatrix} e_{i1} \\ e_{i2} \\ \vdots \\ e_{ik_i} \\ e_{ik_i+1} \\ \vdots \\ e_{in_i} \end{pmatrix}$$

- With first  $k_i$  observations for boys and second  $n_i - k_i$  observations for girls
- Two fixed effects  $\beta = (\mu, \alpha_1)^T$  and two random effects  $\mathbf{b}_i = ((\alpha a)_{i1}, (\alpha a)_{i2})^T$
- With  $\mathbf{G} = \text{VAR}(\mathbf{b}_i) = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$  and  $\mathbf{R} = \text{VAR}(\mathbf{e}_i) = \begin{pmatrix} \sigma_R^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_R^2 \end{pmatrix}$

# General formulation

## Statistical model

### Alternative mixed model:

- The mean of  $\mathbf{y}_i$  is

$$\mathbb{E}(\mathbf{y}_i) = \begin{pmatrix} \boldsymbol{\mu} + \boldsymbol{\alpha}_1 \\ \boldsymbol{\mu} \end{pmatrix}$$

- With  $\boldsymbol{\mu} + \boldsymbol{\alpha}_1$  a vector of length  $k_i$
- With  $\boldsymbol{\mu}$  a vector of length  $n_i - k_i$
- Variance of  $\mathbf{y}_i$  is an  $n_i \times n_i$  matrix

$$\mathbf{V}_i = \begin{pmatrix} \mathbf{V}_{i11} & \mathbf{V}_{i12} \\ \mathbf{V}_{i21} & \mathbf{V}_{i22} \end{pmatrix}$$

- With  $\mathbf{V}_{i11}$  a  $k_i \times k_i$  matrix
- With  $\mathbf{V}_{i12} = \mathbf{V}_{i21}^T$  a  $k_i \times (n_i - k_i)$  matrix
- With  $\mathbf{V}_{i22}$  a  $(n_i - k_i) \times (n_i - k_i)$  matrix
- The structure is identical to ANOVA model but the content of  $\mathbf{V}$ 's is different

- The submatrices are now

$$\mathbf{V}_{i11} = \begin{pmatrix} \sigma_1^2 + \sigma_R^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \ddots & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 + \sigma_R^2 \end{pmatrix}$$

$$\mathbf{V}_{i22} = \begin{pmatrix} \sigma_2^2 + \sigma_R^2 & \sigma_2^2 & \sigma_2^2 \\ \sigma_2^2 & \ddots & \sigma_2^2 \\ \sigma_2^2 & \sigma_2^2 & \sigma_2^2 + \sigma_R^2 \end{pmatrix}$$

$$\mathbf{V}_{i12} = \begin{pmatrix} \rho\sigma_1\sigma_2 & \dots & \rho\sigma_1\sigma_2 \\ \vdots & \ddots & \vdots \\ \rho\sigma_1\sigma_2 & \dots & \rho\sigma_1\sigma_2 \end{pmatrix}$$

- Similar to an heteroscedastic ANOVA model
- When  $\sigma_1^2 = \sigma_2^2$  the mixed model returns to the ANOVA model [check this yourself]

# General formulation

## Exercise L1

### Theoretical questions

- Consider the following three-way ANOVA model

$$y_{hijk} = \mu + \alpha_h + \beta_i + (\alpha\beta)_{hi} + a_{j(i)} + (\alpha a)_{hj(i)} + e_{hijk}$$

- With  $\mu$  the overall mean
- With  $\alpha_h$  the effect of sex ( $\alpha_2 = 0$ )
- With  $\beta_i$  the effect of class type ( $\beta_2 = 0$ )
- With  $(\alpha\beta)_{hi}$  the interaction effect of sex and class type ( $(\alpha\beta)_{11} \neq 0$ )
- With  $a_{j(i)} \sim N(0, \sigma_{C(T)}^2)$  the effect of class
- With  $(\alpha a)_{hj(i)} \sim N(0, \sigma_{SC(T)}^2)$  the interaction effect of sex and class with type of class
- With  $e_{hijk} \sim N(0, \sigma_R^2)$  the residual
- Determine the  $X_i$  and  $Z_i$  matrix for unit  $i$  being class
- Determine the  $V_i$  matrix

# General formulation

## SAS procedure MIXED

```
PROC MIXED DATA=name  
  METHOD=REML options;  
  
  CLASS factors;  
  
  MODEL outcome = factors  
    + covariates /SOLUTION CL  
    DDFM=SAT RESIDUAL options;  
  
  RANDOM factors/options;  
  
  REPEATED factors/options;  
  
  LSMEANS factors/options;  
  
  ESTIMATE 'label' fixed  
  effects | random  
  effects/options;  
  
RUN;
```

- Model statement can handle both categorical as numerical outcomes
  - It should not contain the random effects
  - **It describes the  $X_i$  matrix**
- The random statement can handle random coefficients
  - Including multivariate random effects
  - **It describes the  $G$  matrix**
- Repeated statement can handle correlated residuals
  - **It describes the  $R$  matrix**
- LSMEANS and ESTIMATE can make model inferences

# General formulation

## SAS procedure MIXED

### Alternative mixed model:

Model:  $y_{hij} = \mu + \alpha_h + b_{hi} + e_{hij}$

$$\begin{pmatrix} b_{1i} \\ b_{2i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right)$$

```
PROC MIXED DATA=name  
  METHOD=REML ASYCOV CL;  
  CLASS CLASS GIRL;  
  MODEL IQV = GIRL /  
    SOLUTION CL DDFM=SAT;  
  RANDOM GIRL /  
    SUBJECT=CLASS TYPE=UN V;  
 RUN;
```

### Explanation statements:

- RANDOM GIRL: creates a random effect for boys and girls
- SUBJECT=CLASS: random effects for boys and girls vary with class
- TYPE=UN: selects the unstructured  $\mathbf{G}$ :  
$$\mathbf{G} = \text{VAR}(\mathbf{b}_i) = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$
- V: prints  $V_1$
- No repeated statement:  $R_i = \sigma_R^2 I_{n_i \times n_i}$ 
  - With  $I_{k \times k}$  a  $k \times k$  matrix that has 1 on the diagonal and zeros elsewhere

# General formulation

SAS procedure **MIXED**

Alternative mixed model:

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Alpha	Lower	Upper
UN(1,1)	CLASS	0.5869	0.05	0.4316	0.8445
UN(2,1)	CLASS	0.6046	0.05	0.4319	0.7773
UN(2,2)	CLASS	0.6847	0.05	0.4997	0.9959
Residual		3.8055	0.05	3.6371	3.9860

Asymptotic Covariance Matrix of Estimates					
Row	Cov Parm	CovP1	CovP2	CovP3	CovP4
1	UN(1,1)	0.009956	0.006330	0.003835	-0.00087
2	UN(2,1)	0.006330	0.007764	0.007740	-0.00029
3	UN(2,2)	0.003835	0.007740	0.01430	-0.00117
4	Residual	-0.00087	-0.00029	-0.00117	0.007905

Solution for Fixed Effects									
Effect	GIRL	Estimate	Standard Error	DF	t Value	Pr >  t	Alpha	Lower	Upper
Intercept		11.7627	0.07287	170	161.42	<.0001	0.05	11.6189	11.9066
GIRL	0	0.09775	0.06502	195	1.50	0.1344	0.05	-0.03049	0.2260
GIRL	1	0	.	.	.	.	.	.	.

- $\text{UN}(1,1) = \sigma_1^2$
- $\text{UN}(2,1) = \rho\sigma_1\sigma_2$
- $\text{UN}(2,2) = \sigma_2^2$
- $\text{UN}(1,1)$  represents boys, since SAS uses alphabetical order (0 = boys; 1 = girls)

# General formulation

## SAS procedure MIXED

$V_{i11}$

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9	Col10	Col11	Col12	Col13	Col14	Col15	Col16	Col17
1	4.3924	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.6046	0.6046
2	0.5869	4.3924	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.6046	0.6046
3	0.5869	0.5869	4.3924	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.6046	0.6046
4	0.5869	0.5869	0.5869	4.3924	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.6046	0.6046
5	0.5869	0.5869	0.5869	0.5869	4.3924	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.6046	0.6046
6	0.5869	0.5869	0.5869	0.5869	0.5869	4.3924	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.6046	0.6046
7	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	4.3924	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.6046	0.6046
8	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	4.3924	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.6046	0.6046
9	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	4.3924	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.6046	0.6046
10	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	4.3924	0.5869	0.5869	0.5869	0.5869	0.5869	0.6046	0.6046
11	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	4.3924	0.5869	0.5869	0.5869	0.5869	0.6046	0.6046
12	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	4.3924	0.5869	0.5869	0.5869	0.6046	0.6046
13	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	4.3924	0.5869	0.5869	0.6046	0.6046
14	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	4.3924	0.5869	0.6046	0.6046
15	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	0.5869	4.3924	0.6046	0.6046
16	0.6046	0.6046	0.6046	0.6046	0.6046	0.6046	0.6046	0.6046	0.6046	0.6046	0.6046	0.6046	0.6046	0.6046	0.6046	4.4902	0.6847
17	0.6046	0.6046	0.6046	0.6046	0.6046	0.6046	0.6046	0.6046	0.6046	0.6046	0.6046	0.6046	0.6046	0.6046	0.6046	0.6847	4.4902

$V_{i22}$

$V_{i12}$

- There are 27 children in the first class
  - 15 boys
  - 12 girls
- Complete matrix  $V_i$  is not shown due to dimension
- If children are not properly sorted, boys and girls will be unordered

# General formulation

## Exercise L2

### Data analytic question:

- Fit the following mixed effects model [report the codes and the output]

$$y_{hijk} = \mu + \alpha_h + \beta_i + (\alpha\beta)_{hi} + b_{hj(i)} + e_{hijk}$$

- With  $\mu$  the overall mean
- With  $\alpha_h$  the effect of sex ( $\alpha_2 = 0$ )
- With  $\beta_i$  the effect of class type ( $\beta_2 = 0$ )
- With  $(\alpha\beta)_{hi}$  the interaction effect of sex and class type ( $(\alpha\beta)_{11} \neq 0$ )
- With  $b_{hj(i)}$  the random effect of sex  $h$  within class  $i$  for type of class  $j$

$$(b_{11(i)}, b_{21(i)}, b_{12(i)}, b_{22(i)})^T \sim N \left( \mathbf{0}, \begin{pmatrix} \mathbf{G}_{11} & \mathbf{G}_{12} \\ \mathbf{G}_{21} & \mathbf{G}_{22} \end{pmatrix} \right),$$

- with  $\mathbf{G}_{jj} = \begin{pmatrix} \sigma_1^2(j) & \rho_j \sigma_1(j) \sigma_2(j) \\ \rho_j \sigma_1(j) \sigma_2(j) & \sigma_2^2(j) \end{pmatrix}$  and  $\mathbf{G}_{12} = \mathbf{G}_{21} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$
- With  $e_{hijk} \sim N(0, \sigma_R^2)$  the residual

# General formulation

## SAS procedure **MIXED**

### School data analyzed:

- Arithmetic score  $\mathbf{y}_i = (y_{i1}, y_{i2})^T$ 
  - Unit  $i$  is child
  - Two time points:  $\mathbf{t}_i = (0,1)^T$  representing grade 7 and grade 8
- Several baseline regression variables:
  - Sex
  - Type of class
  - IQV
  - IQP
  - SES of child
- One random effect:
  - Class within type of class

### SAS code:

```
PROC MIXED DATA=LONGITUDINAL  
METHOD=REML CL ASYCOV;  
CLASS CLASS CHILD COMBI GIRL  
GRADE;  
MODEL ARITH = GRADE GIRL  
COMBI CSES IQV IQP /SOLUTION  
CL DDFM=SAT;  
RANDOM CLASS (COMBI) / GROUP =  
COMBI VCORR;  
REPEATED GRADE / SUBJECT =  
CHILD(GIRL*COMBI) TYPE = UNR;  
RUN;
```

# General formulation

## SAS procedure MIXED

### Linear mixed model:

- Statistical model:

$$y_{ij} = \mu + \alpha_j + \beta_G x_{Gi} + \beta_T x_{Ti} + \beta_S x_{Si} \\ + \beta_V x_{Vi} + \beta_P x_{Pi} \\ + b_{1i} z_{1i} + b_{2i} z_{2i} + e_{ij}$$

- With  $\mu$  an intercept at grade 8
- With  $\alpha_j$  effect of grade ( $\alpha_2 = 0$ )
- With  $x_{Gi}$  a binary indicator for girl and  $\beta_G$  the regression coefficient
- With  $x_{Ti}$  a binary indicator for type of class and  $\beta_T$  the regression coefficient
- With  $x_{Si}$  the numeric variable for SES and  $\beta_S$  the regression coefficient

- With  $x_{Vi}$  the numeric variable for IQV and  $\beta_V$  the regression coefficient
- With  $x_{Pi}$  the numeric variable for IQP and  $\beta_P$  the regression coefficient
- With  $b_{ri} \sim N(0, \sigma_{C,r}^2)$  the random effect of class for type of class  $r$  (heteroscedasticity)
- With  $z_{ri}$  a binary indicator for type of class  $r$  (such that  $z_{1i} + z_{2i} = 1$ )
- With  $e_i = (e_{i1}, e_{i2})^T$  bivariate normal  
$$e_i \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{R,1}^2 & \rho\sigma_{R,1}\sigma_{R,2} \\ \rho\sigma_{R,1}\sigma_{R,2} & \sigma_{R,2}^2 \end{pmatrix}\right)$$
- Here we have a non-diagonal residual covariance matrix

# General formulation

## SAS procedure MIXED

### Correlation structure:

- Correlation between children within class for single grade classes
  - In grade 7:  $ICC_{BC}^{S7} = \frac{\sigma_{C,1}^2}{\sigma_{C,1}^2 + \sigma_{R,1}^2}$
  - In grade 8:  $ICC_{BC}^{S8} = \frac{\sigma_{C,1}^2}{\sigma_{C,1}^2 + \sigma_{R,2}^2}$
- Correlation between children within class for multi-grade classes
  - In grade 7:  $ICC_{BC}^{M7} = \frac{\sigma_{C,2}^2}{\sigma_{C,2}^2 + \sigma_{R,1}^2}$
  - In grade 8:  $ICC_{BC}^{M8} = \frac{\sigma_{C,2}^2}{\sigma_{C,2}^2 + \sigma_{R,2}^2}$

- Correlation between children within class with scores in different grades
  - Single grade:  $ICC_{BC}^{S78} = \frac{\sigma_{C,1}^2}{\sqrt{\sigma_{C,1}^2 + \sigma_{R,1}^2} \sqrt{\sigma_{C,1}^2 + \sigma_{R,2}^2}}$
  - Multi-grade:  $ICC_{BC}^{M78} = \frac{\sigma_{C,2}^2}{\sqrt{\sigma_{C,2}^2 + \sigma_{R,1}^2} \sqrt{\sigma_{C,2}^2 + \sigma_{R,2}^2}}$
  - Confidence intervals for  $ICC_{BC}^{S78}$  and  $ICC_{BC}^{M78}$  are not straightforward
- Correlation between scores on the same child:  $ICC_{WC} = \rho$
- Correlation between children from different classes is equal to zero

# General formulation

SAS procedure **MIXED**

## Variances and covariance:

Solution for Fixed Effects											
Effect	GRADE	GIRL	COMBI	Estimate	Standard Error	DF	t Value	Pr >  t	Alpha	Lower	Upper
Intercept				5.5715	0.3505	926	15.90	<.0001	0.05	4.8837	6.2593
GRADE	7			-7.7579	0.08799	3572	-88.17	<.0001	0.05	-7.9304	-7.5853
GRADE	8			0	.	.	.	.	.	.	.
GIRL		0		0.05457	0.08948	3486	0.61	0.5420	0.05	-0.1209	0.2300
GIRL		1		0	.	.	.	.	.	.	.
COMBI			0	0.3273	0.2143	139	1.53	0.1290	0.05	-0.09644	0.7510
COMBI			1	0	.	.	.	.	.	.	.
CSES				0.04057	0.004917	3574	8.25	<.0001	0.05	0.03093	0.05021
IQV				0.4905	0.02544	3558	19.28	<.0001	0.05	0.4406	0.5404
IQP				0.6250	0.02312	3522	27.03	<.0001	0.05	0.5797	0.6704

### Intercept:

- No real meaning when numerical variables are non-standardized

### Significant effects:

- Grade: 7.76 [7.59; 7.93]
- SES: 0.041 [0.031; 0.050]
- IQV: 0.491 [0.441; 0.540]
- IQP: 0.625 [0.580; 0.670]

### Non-significant effects:

- Sex: 0.055 [-0.121; 0.230]
- Type: 0.327 [-0.096; 0.751]

# General formulation

SAS procedure **MIXED**

## Variances and covariance:

Covariance Parameter Estimates						
Cov Parm	Subject	Group	Estimate	Alpha	Lower	Upper
CLASS(COMBI)		COMBI 0	1.6196	0.05	1.2144	2.2689
CLASS(COMBI)		COMBI 1	1.6424	0.05	1.0968	2.7285
Var(1)	CHILD(GIRL*COMBI)		7.1542	0.05	6.8114	7.5236
Var(2)	CHILD(GIRL*COMBI)		26.1543	0.05	24.8550	27.5587
Corr(2,1)	CHILD(GIRL*COMBI)		0.2542	0.05	0.2176	0.2907

- Variability between children within class is much larger in grade 8 than in grade 7
- Variability between classes is similar for type of class
- Correlation between scores on the same child:  
 $ICC_{WC} = 0.254 [0.218; 0.291]$

## Correlations:

$$ICC_{BC}^{S7} = 0.185 [0.139; 0.233]$$

$$ICC_{BC}^{S8} = 0.058 [0.042; 0.077]$$

$$ICC_{BC}^{M7} = 0.187 [0.121; 0.256]$$

$$ICC_{BC}^{M8} = 0.059 [0.036; 0.086]$$

$$ICC_{BC}^{S78} = 0.104$$

$$ICC_{BC}^{M78} = 0.105$$

- Confidence intervals are calculated with general approach on ICC's

# General formulation

## SAS procedure MIXED

### Large data sets:

- Procedure `MIXED` has many options, but numerical issues may arise for
  - Many fixed effects
  - Many random effects
  - Many observations
- Examples of problems:
  - Thousands of fixed effects (e.g., for genetics analysis)
  - Hierarchical nested models with hundreds of fixed or random effects
- Require high-performance computing

### SAS offers: HPMIXED

- Works the same as procedure `MIXED`
- Is still in an experimental state
- Default (and only) estimation method is REML
- Has less options than `MIXED`
  - No Satterthwaite degrees of freedom
  - No covariance parameters VC's
  - No confidence intervals on VC's
  - No standard residual analysis
- Does not solve all analysis problems, but it is a nice next step

# Subject-specific models

## Polynomial time profiles

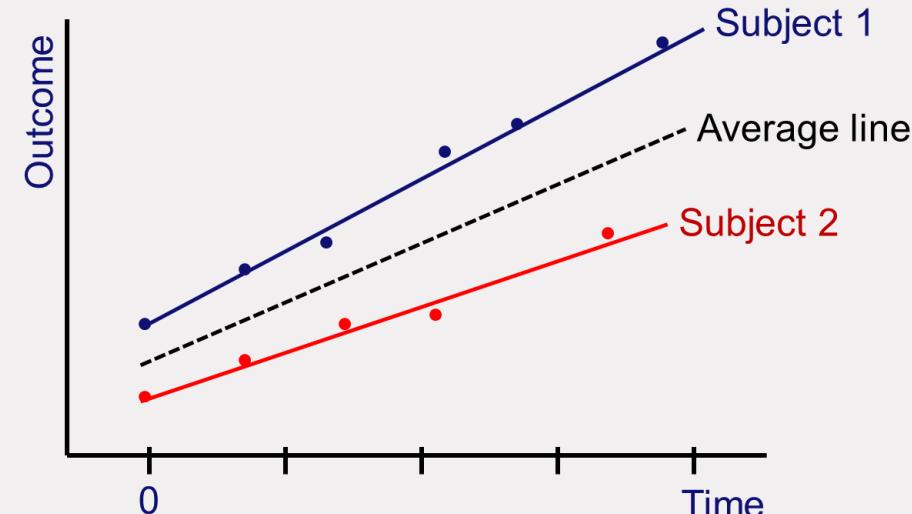
### Without additional covariates:

- Statistical model:

$$y_{ij} = \sum_{r=0}^q b_{ir} t_{ij}^r + e_{ij}$$

- With  $y_{ij}$  the  $j$ th outcome on unit  $i$  observed at time  $t_{ij}$
- With  $\mathbf{b}_i = (b_{i0}, b_{i1}, \dots, b_{iq})^T$  a vector of random coefficients:  $\mathbf{b}_i \sim N(\boldsymbol{\beta}, \mathbf{G}_i)$  and  $\boldsymbol{\beta} = (\beta_0, \dots, \beta_q)^T$  the average coefficients
- With  $\mathbf{e}_i = (e_{i1}, e_{i2}, \dots, e_{in_i})^T$  a vector of residuals:  $\mathbf{e}_i \sim N(\mathbf{0}, \mathbf{R}_i)$
- The model describes for each unit a polynomial time profile

### Linear time profiles [no covariates]:



- If we assume i.i.d.  $e_{ij} \sim N(0, \sigma_e^2)$ :
  - Statistical model has six parameters
  - For  $m$  regression lines

# Subject-specific models

## Polynomial time profiles

### Linear time profiles [No covariates]:

- Statistical model:

$$y_{ij} = b_{i0} + b_{i1}t_{ij} + e_{ij}$$

- Only two coefficients:

$$\begin{pmatrix} b_{i0} \\ b_{i1} \end{pmatrix} \sim N \left( \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \begin{pmatrix} \tau_0^2 & \rho\tau_0\tau_1 \\ \rho\tau_0\tau_1 & \tau_1^2 \end{pmatrix} \right)$$

- First and second moments

$$\mathbb{E}[y_{ij}] = \beta_0 + \beta_1 t_{ij}$$

$$\text{VAR}(y_{ij}) = \tau_0^2 + 2\rho\tau_0\tau_1 t_{ij} + t_{ij}^2\tau_1^2 + \sigma_R^2$$

$$\text{COV}(y_{ij_1}, y_{ij_2}) =$$

$$= \tau_0^2 + \rho\tau_0\tau_1(t_{ij_1} + t_{ij_2}) + t_{ij_1}t_{ij_2}\tau_1^2$$

- All moments depend on the time points

- Minimizing the variance in  $y_{ij}$ 
  - Minimum is equal to  $[1 - \rho^2]\tau_0^2 + \sigma_R^2$
  - Minimum is attained at  $t_0 = -\rho\tau_0/\tau_1$
- Negative correlation ( $\rho < 0$ )
  - Larger intercept decreases growth
  - Minimum variance lies within the data
- No correlation ( $\rho = 0$ )
  - Intercept is not informative for slope
  - Minimum variance is at time zero
- Positive correlation ( $\rho > 0$ )
  - Larger intercept increases growth
  - Minimum variance is before data collection, since time is usually positive

# Subject-specific models

## Polynomial time profiles

### With additional stationary covariates:

- Fixed effects coefficients  $\beta_k$  are being functions of  $\boldsymbol{x}_i = (x_{i1}, \dots, x_{ip})^T$  being time-stationary (continuous and categorical) variables

$$\beta_k(\boldsymbol{x}_i) = \beta_{0k} + \sum_{s=1}^p \beta_{sk} x_{is}$$

- Thus only the mean time profile is changing with characteristics  $\boldsymbol{x}_i$  of unit  $i$
- For example different mean time profiles for men and women
- Modeling variances is more complicated [although factors can affect  $\boldsymbol{G}_i$  and  $\boldsymbol{R}_i$  through the “GROUP =” options in procedure MIXED of SAS]
- General model becomes

$$\begin{aligned} y_{ij} &= \sum_{r=0}^q b_{ir} t_{ij}^r + e_{ij} = \sum_{r=0}^q \beta_r(\boldsymbol{x}_i) t_{ij}^r + \sum_{r=0}^q [b_{ir} - \beta_r(\boldsymbol{x}_i)] t_{ij}^r + e_{ij} \\ &= \sum_{r=0}^q \beta_{0r} t_{ij}^r + \sum_{r=0}^q \sum_{s=1}^p \beta_{sr} x_{is} t_{ij}^r + \sum_{r=0}^q u_{ir} t_{ij}^r + e_{ij} \end{aligned}$$

- With  $\boldsymbol{u}_i = (u_{i1}, u_{i2}, \dots, u_{iq})^T \sim N(\mathbf{0}, \boldsymbol{G}_i)$  and  $\boldsymbol{e}_i = (e_{i1}, e_{i2}, \dots, e_{in_i})^T \sim N(\mathbf{0}, \boldsymbol{R}_i)$

# Subject-specific models

## Polynomial time profiles

### Example: [Linear profile clinical trial]

```
PROC MIXED DATA=DATA_PAIN;  
  CLASS TRT (REF='Placebo')  
  SUBJECT;  
  MODEL RESP = TRT FU TRT*FU  
  /SOLUTION CL DDFM=SAT  
  RESIDUAL;  
  RANDOM INT FU /SUBJECT =  
  SUBJECT TYPE=UNR;
```

**RUN;**

- Variable FU is:  $[t_{ij} - 120]/15$
- Random statement tells that the intercept and slope are random

### Estimated variance components:

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Alpha	Lower	Upper
Var(1)	SUBJECT	8.6257	0.05	4.1666	27.2730
Var(2)	SUBJECT	1.1151	0.05	0.4285	7.0553
Corr(2,1)	SUBJECT	-0.7171	0.05	-1.0587	-0.3755
Residual		5.4935	0.05	3.8203	8.5739

- Based on the confidence intervals
  - Clear subject-to-subject variability
  - Subjects have different slopes
  - A larger intercept reduces the slope:
    - 0.72 [-1.00; -0.38]

# Subject-specific models

## Polynomial time profiles

### Example: [Linear profile for clinical trial]

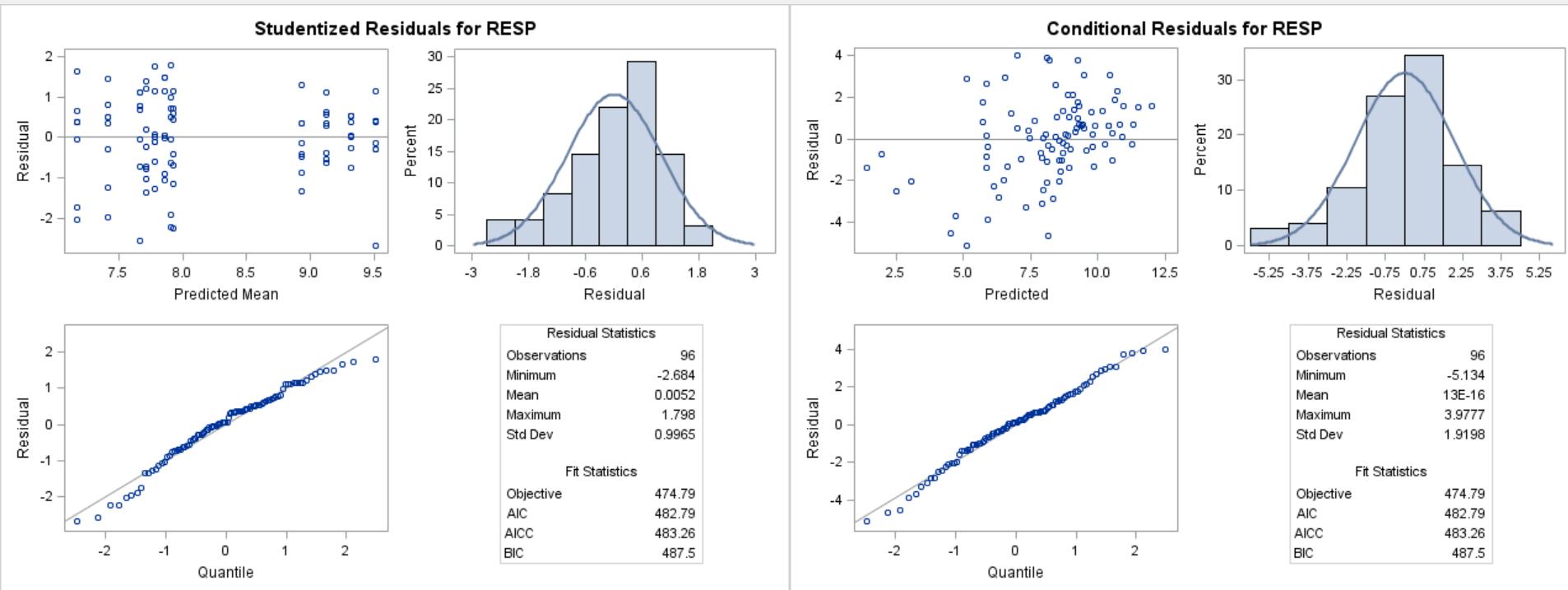
Solution for Fixed Effects									
Effect	TRT	Estimate	Standard Error	DF	t Value	Pr >  t	Alpha	Lower	Upper
Intercept		9.5138	1.2486	21	7.62	<.0001	0.05	6.9173	12.1102
TRT	G	-2.3425	1.7657	21	-1.33	0.1989	0.05	-6.0145	1.3295
TRT	R	-1.5850	1.7657	21	-0.90	0.3795	0.05	-5.2570	2.0870
TRT	Placebo	0	.	.	.	.	.	.	.
FU		-0.1925	0.5260	21	-0.37	0.7181	0.05	-1.2865	0.9015
FU*TRT	G	0.4388	0.7439	21	0.59	0.5616	0.05	-1.1084	1.9859
FU*TRT	R	0.1213	0.7439	21	0.16	0.8721	0.05	-1.4259	1.6684
FU*TRT	Placebo	0	.	.	.	.	.	.	.

### Results:

- Pain score at 120 minutes is largest for Placebo, but effect of G and R are not significant
- Progression of pain score is decreasing for Placebo, but less so for treatment G and R [although not significant]

# Subject-specific models

## Polynomial time profiles



# Subject-specific models

## Polynomial time profiles

**Reference limits:** provide a range of values common to the population

- Often quantified by  $100\% \times (1 - 2\alpha)$  prediction intervals
- They are typically age or time dependent
- Might be constructed for different subgroups of units (e.g., sex)
- Consider linear profiles:

$$y_{ij} = b_{i0} + b_{i1}t_{ij} + e_{ij}$$

- With the two random coefficients:

$$\begin{pmatrix} b_{i1} \\ b_{i2} \end{pmatrix} \sim N \left( \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \begin{pmatrix} \tau_0^2 & \rho\tau_0\tau_1 \\ \rho\tau_0\tau_1 & \tau_1^2 \end{pmatrix} \right)$$

- Assume we collect an arbitrary unit at time  $t$ , then what is  $U(t)$  such that  $P(y(t) \leq U(t)) = 1 - \alpha$ 
  - With  $y(t)$  the observed outcome at  $t$
- We have learned that  $y(t) \sim N(\mu(t), \sigma^2(t))$ 
  - With  $\mu(t) = \beta_0 + \beta_1 t$
  - With  $\sigma^2(t) = \tau_0^2 + 2\rho\tau_0\tau_1 t + t^2\tau_1^2 + \sigma_R^2$
- Thus theoretically we have  $U(t) = \mu(t) + z_{1-\alpha}\sigma(t)$ 
  - With  $z_q$  the  $q$ th quantile of the normal distribution [QUANTILE('Normal',  $q$ , 0, 1)]
  - But we have to estimate the parameters

# Subject-specific models

## Polynomial time profiles

### Asymptotic reference limits:

- Substituting the (RE)ML estimators
- $100\% \times (1 - 2\alpha)$  upper and lower reference limits are

$$U(t) = \hat{\mu}(t) + z_{1-\alpha} \hat{\sigma}(t)$$

$$L(t) = \hat{\mu}(t) - z_{1-\alpha} \hat{\sigma}(t)$$

- They only work for
  - Large sample sizes, since they do not take into account the uncertainty in estimation of  $\mu(t)$  and  $\sigma(t)$
  - The existing data set, since the data and reference limits share the same potential bias or selection issues

### Incorporating estimation uncertainty:

- Estimator for mean  $\mu(t)$ :
  - $\hat{\mu}(t) = \hat{\beta}_0 + \hat{\beta}_1 t$  is approximately normal when data is not too small
  - The mean is approximately:
$$\mathbb{E}(\hat{\mu}(t)) = \mu(t)$$
  - The variance is approximately:
$$\eta^2(t) = \text{VAR}(\hat{\mu}(t)) = \text{VAR}(\hat{\beta}_0) + 2\text{COV}(\hat{\beta}_0, \hat{\beta}_1)t + \text{VAR}(\hat{\beta}_1)t^2$$
- Thus  $y(t) - \hat{\mu}(t)$  is approximately normal with mean zero and variance  $\sigma^2(t) + \text{VAR}(\hat{\mu}(t))$
- **Thus this variance must be estimated**

# Subject-specific models

## Polynomial time profiles

### Finite data reference limits:

- Estimation of  $\eta^2(t) = \text{VAR}(\hat{\mu}(t))$  can be done with SAS output
  - Use COVB as option in model statement
  - The output reports the estimates for the covariance matrix
$$\begin{pmatrix} \text{VAR}(\hat{\beta}_0) & \text{COV}(\hat{\beta}_0, \hat{\beta}_1) \\ \text{COV}(\hat{\beta}_0, \hat{\beta}_1) & \text{VAR}(\hat{\beta}_1) \end{pmatrix}$$
- Estimator of variance of  $y(t) - \hat{\mu}(t)$ :  
 $\hat{\sigma}^2(t) + \hat{\eta}^2(t)$ 
  - Distribution of this estimator is unknown
  - Whether approximation with chi-square is appropriate is unknown

- The quantile  $z_{1-\alpha}$  does not account for uncertainty in  $\hat{\sigma}^2(t) + \hat{\eta}^2(t)$ 
  - Possibly better to use a quantile ( $t_d^{-1}(1 - \alpha)$ ) of the  $t$ -distribution, but determination of degrees of freedom  $d$  is non-trivial
  - $\text{VAR}(\hat{\sigma}^2(t))$  is unreliable for Satterthwaite degrees of freedom due to the term of  $t^4$
  - One option is to use number of units  $m$  or  $m - p$  corrected for  $p$  fixed effects
- Proposed new reference limits:

$$U(t) = \hat{\mu}(t) + t_m^{-1}(1 - \alpha) \sqrt{\hat{\sigma}^2(t) + \hat{\eta}^2(t)}$$

$$L(t) = \hat{\mu}(t) - t_m^{-1}(1 - \alpha) \sqrt{\hat{\sigma}^2(t) + \hat{\eta}^2(t)}$$

# Subject-specific models

## Polynomial time profiles

### Confidence limits time profile:

- Provide a range of time profiles that are also likely to have occurred with the proposed confidence
- They follow the reasoning of the reference limits
  - But variability of the individual is excluded
- $100\% \times (1 - 2\alpha)$  confidence interval
$$U(t) = \hat{\mu}(t) + t_d^{-1}(1 - \alpha)\hat{\eta}(t)$$
$$L(t) = \hat{\mu}(t) - t_d^{-1}(1 - \alpha)\hat{\eta}(t)$$
- Degrees of freedom  $d$  is taken from the degrees of freedom of fixed effects

### Reference limits: [pain data]

```
ODS OUTPUT SOLUTIONF=PARMS_FE;  
ODS OUTPUT COVPARMS=PARMS_RE;  
ODS OUTPUT COVB=COVS_FE;  
PROC MIXED DATA=DATA_LONG CL  
ASYCOV;  
CLASS TRT SUBJECT;  
MODEL RESP = TRT FU TRT*FU  
/SOLUTION CL COVB  
DDFM=SAT;  
RANDOM INT FU /SUBJECT =  
SUBJECT TYPE=UNR;  
RUN;
```

# Subject-specific models

## Polynomial time profiles

### Fixed effects estimates:

Solution for Fixed Effects									
Effect	TRT	Estimate	Standard Error	DF	t Value	Pr >  t	Alpha	Lower	Upper
Intercept		9.5137	1.2485	21	7.62	<.0001	0.05	6.9174	12.1101
TRT	G	-2.3425	1.7656	21	-1.33	0.1989	0.05	-6.0144	1.3294
TRT	R	-1.5850	1.7656	21	-0.90	0.3795	0.05	-5.2569	2.0869
TRT	Placebo	0	.	.	.	.	.	.	.
FU		-0.1925	0.5260	21	-0.37	0.7181	0.05	-1.2864	0.9014
FU*TRT	G	0.4387	0.7439	21	0.59	0.5616	0.05	-1.1083	1.9858
FU*TRT	R	0.1212	0.7439	21	0.16	0.8721	0.05	-1.4258	1.6683
FU*TRT	Placebo	0	.	.	.	.	.	.	.

- Estimated mean time profile:  
 $\hat{\mu}(t) = 9.51 - 0.19t$
- Remember:  $t = [t_{ij} - 120]/15$

### Covariances fixed effects:

Covariance Matrix for Fixed Effects									
Row	Effect	TRT	Col1	Col2	Col3	Col4	Col5	Col6	Col7
1	Intercept		1.5588	1.5588	-1.5588		-0.4839	0.4839	0.4839
2	TRT	G	-1.5588	3.1175	1.5588		0.4839	-0.9679	-0.4839
3	TRT	R	-1.5588	1.5588	3.1175		0.4839	-0.4839	-0.9679
4	TRT	Placebo							
5	FU		-0.4839	0.4839	0.4839		0.2767	0.2767	-0.2767
6	FU*TRT	G	0.4839	-0.9679	-0.4839		-0.2767	0.5534	0.2767
7	FU*TRT	R	0.4839	-0.4839	-0.9679		-0.2767	0.2767	0.5534
8	FU*TRT	Placebo							

- Estimated variance mean time profile:  
 $\hat{\eta}(t) = 1.559 - 0.968t + 0.277t^2$

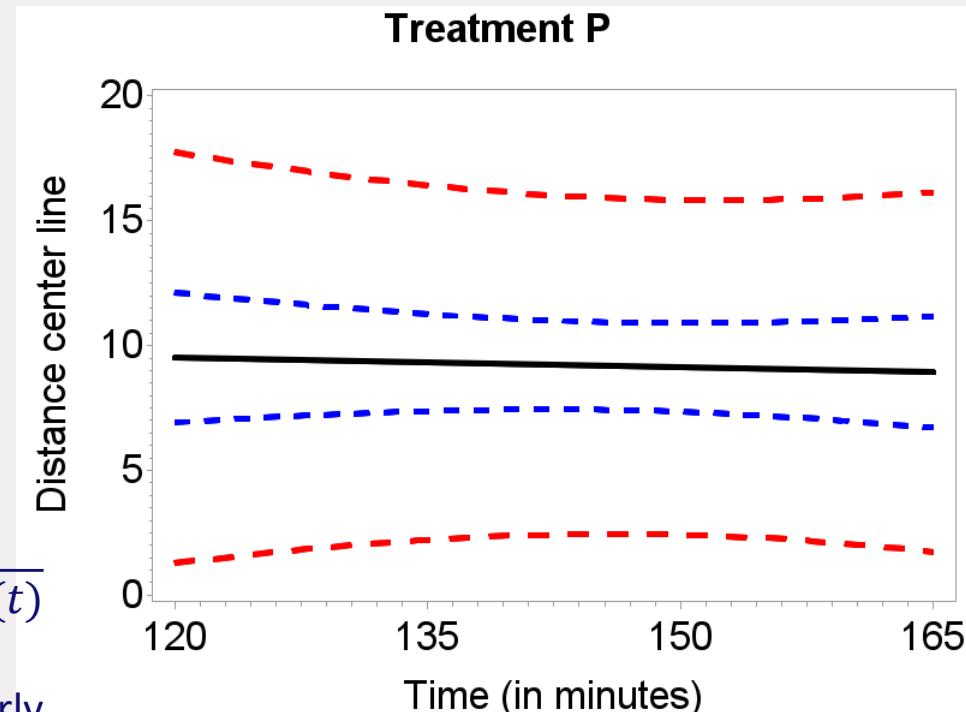
# Subject-specific models

## Polynomial time profiles

### Variance components:

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Alpha	Lower	Upper
UN(1,1)	SUBJECT	8.6246	0.05	4.1658	27.2735
UN(2,1)	SUBJECT	-2.2234	0.05	-5.0915	0.6447
UN(2,2)	SUBJECT	1.1150	0.05	0.4284	7.0556
Residual		5.4935	0.05	3.8203	8.5739

- Variance  $\sigma^2(t)$  is estimated  
$$\hat{\sigma}^2(t) = 14.1 - 4.44t + 1.12t^2$$
- Limits [degrees of freedom 21]:  
$$LPL(t) = \hat{\mu}(t) - 2.080\sqrt{\hat{\sigma}^2(t) + \hat{\eta}^2(t)}$$
$$LCL(t) = \hat{\mu}(t) - 2.080\hat{\eta}(t)$$
  - $UCL(t)$  &  $UPL(t)$  are calculated similarly



# Subject-specific models

## Exercise L3

### Theoretical questions:

- Assume the subject-specific model with linear time profiles:
  - Formulate the correlation for two repeated observations at time  $t$  and  $(1 + c)t$
  - Formulate 95% reference (prediction) limits for the linear time profiles

### Data analytic questions:

- Use the outcome AUC of the pain data and fit the subject-specific model with linear time profile and treatment
  - Report the model parameters with their 95% confidence intervals
  - Do you believe that treatment has an effect on the outcome?
  - Do you believe that time has an effect on the outcome?
  - Estimate the correlation between observations at 135 and 150
  - Estimate the 95% confidence and reference limits for the three treatments

# Subject-specific models

## Fractional polynomials (FPs)

### Parsimonious parametric modeling:

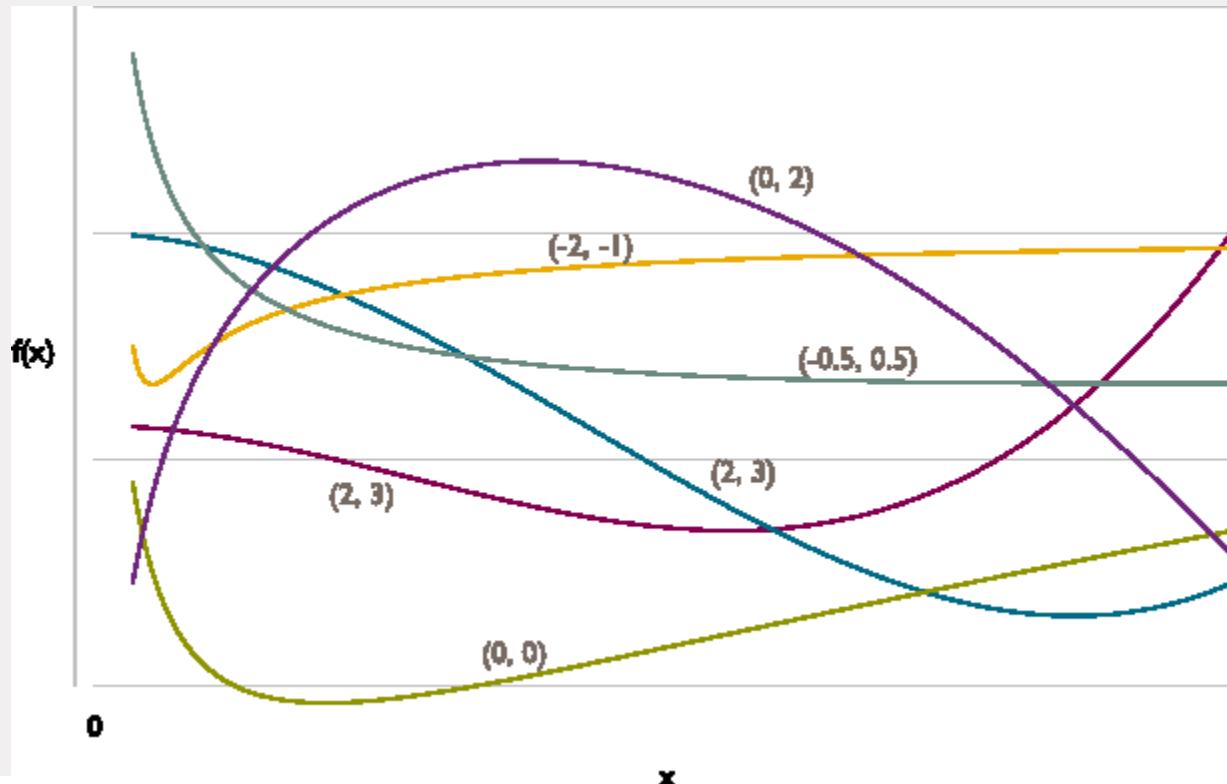
- More flexible in the shape than regular polynomials time profiles
- FP's of degree  $m$   
$$\beta_0 + \sum_{r=1}^m \beta_r t^{p_r}$$
  - With  $\beta_r$  unknown regression parameters
  - With  $p_r \in \{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$  and with  $p_1 \leq p_2 \leq \dots \leq p_m$
  - With  $t^0 \equiv \log t$
  - When  $p_r = p_s$ , we take  $\beta_r t^{p_r} + \beta_s t^{p_s} = \beta_r t^{p_r} + \beta_s t^{p_s} \log t$
- Common degree is equal to  $m = 2$ 
  - Number of possible FP's is 36

### Natural extensions:

- Including random effects  
$$b_0 + \sum_{r=1}^m b_r t^{p_r}$$
  - With  $\boldsymbol{b}_i = (b_{i0}, b_{i1}, \dots, b_{iq})^T$  a vector of random coefficients:  $\boldsymbol{b}_i \sim N(\boldsymbol{\beta}, \boldsymbol{G})$
  - For  $m = 2$  we have three random effects
- Including covariates in the fixed effects parameters:  
$$\beta_k(x_i) = \beta_{0k} + \sum_{s=1}^p \beta_{sk} x_{is}$$
  - For categorical variables, we create subject-specific profiles for each level
  - Need an approach to select best FP

# Subject-specific models

## Fractional polynomials (FPs)



### Benefits FP's:

- FP's can take many different shapes
- FP's are continuous and smooth functions
- FP's can have an asymptote at the origin
- FP's are parametric and thus easy to fit with standard software

# Subject-specific models

## Fractional polynomials (FPs)

### Example: [Growth curve data]

- Model choices:

- We have selected  $p_1 = 0$  and  $p_2 = 2$
- We have included sex:  $x_i \in \{0,1\}$

- Statistical model:

$$y_{ij} = b_{0i} + b_{1i} \log(t_{ij}) + b_{2i} t_{ij}^2 + e_{ij}$$

- With  $b_i = (b_{0i}, b_{1i}, b_{2i})^T$  multivariate normal with mean and variance

$$\beta_r = \beta_{0r} + \beta_{1r} x_i$$

$$\boldsymbol{G}_i = \begin{pmatrix} \tau_0^2 & \rho_{01}\tau_0\tau_1 & \rho_{02}\tau_0\tau_2 \\ \rho_{01}\tau_0\tau_1 & \tau_1^2 & \rho_{12}\tau_1\tau_2 \\ \rho_{02}\tau_0\tau_2 & \rho_{12}\tau_1\tau_2 & \tau_2^2 \end{pmatrix}$$

- With  $e_{ij}$  i.i.d. normal  $N(0, \sigma_R^2)$

### SAS programming codes:

```
PROC MIXED DATA=ANALYSIS CL;  
  CLASS SEX ID;  
  MODEL RESP = SEX AGE0 AGE2  
    SEX*AGE0 SEX*AGE2 / SOLUTION  
    CL DDFM=SAT;  
  RANDOM INT AGE0 AGE2  
    / SUBJECT=ID TYPE=UNR;  
RUN;
```

# Subject-specific models

## Fractional polynomials (FPs)

### Variance-covariance estimates:

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Alpha	Lower	Upper
Var(1)	ID	1.4580	0.05	1.2047	1.8010
Var(2)	ID	1.9313	0.05	1.6582	2.2783
Var(3)	ID	0.004394	0.05	0.003837	0.005083
Corr(2,1)	ID	1.0000	.	.	.
Corr(3,1)	ID	0.1123	0.05	-0.09456	0.3193
Corr(3,2)	ID	-0.1185	0.05	-0.2511	0.01415
Residual		7.0308	0.05	6.7939	7.2804

- Correlation intercept and slope for  $\log t_{ij}$  are perfectly correlated
- Residual variance is relatively small

### Warning message:

NOTE: Convergence criteria met.

NOTE: Estimated G matrix is not positive definite.

NOTE: Asymptotic variance matrix of covariance parameter estimates has been found to be singular and a generalized inverse was used. Covariance parameters with zero variance do not contribute to degrees of freedom computed by DDFM=SATTERTH.

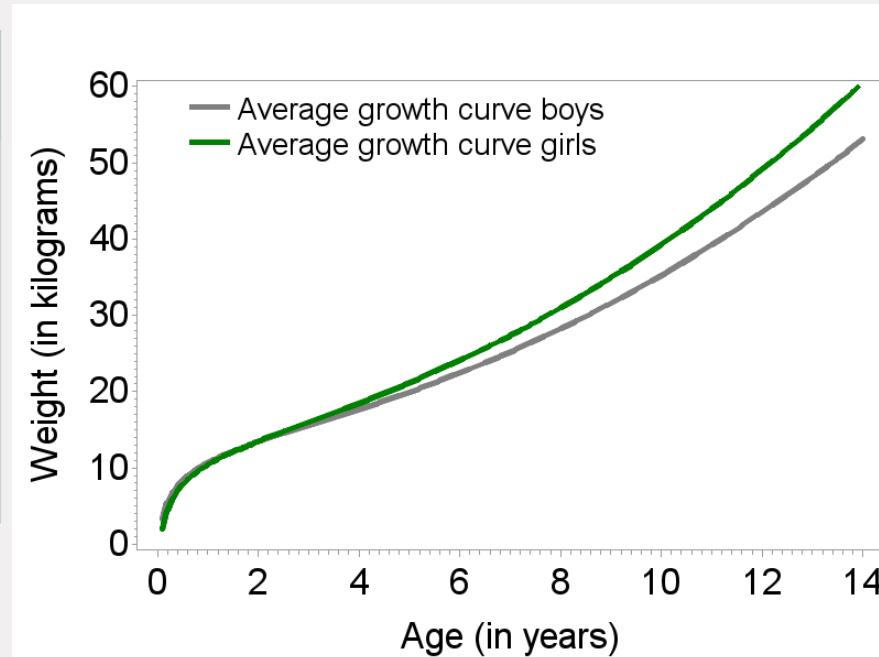
# Subject-specific models

## Fractional polynomials (FPs)

### Average growth curves:

Solution for Fixed Effects									
Effect	SEX	Estimate	Standard Error	DF	t Value	Pr >  t	Alpha	Lower	Upper
Intercept		10.2145	0.09332	970	109.45	<.0001	0.05	10.0314	10.3976
SEX	0	0.3292	0.1368	969	2.41	0.0163	0.05	0.06066	0.5977
SEX	1	0	.	.	.	.	.	.	.
AGE0		3.5825	0.09012	788	39.75	<.0001	0.05	3.4056	3.7594
AGE2		0.2085	0.003581	484	58.23	<.0001	0.05	0.2015	0.2156
AGE0*SEX	0	-0.4388	0.1321	786	-3.32	0.0009	0.05	-0.6980	-0.1795
AGE0*SEX	1	0	.	.	.	.	.	.	.
AGE2*SEX	0	-0.03348	0.005252	484	-6.38	<.0001	0.05	-0.04380	-0.02317
AGE2*SEX	1	0	.	.	.	.	.	.	.

- Boys:  $y_{ij} = 10.5 + 3.14 \log(t_{ij}) + 0.175 t_{ij}^2$
- Girls:  $y_{ij} = 10.2 + 3.58 \log(t_{ij}) + 0.209 t_{ij}^2$



# Subject-specific models

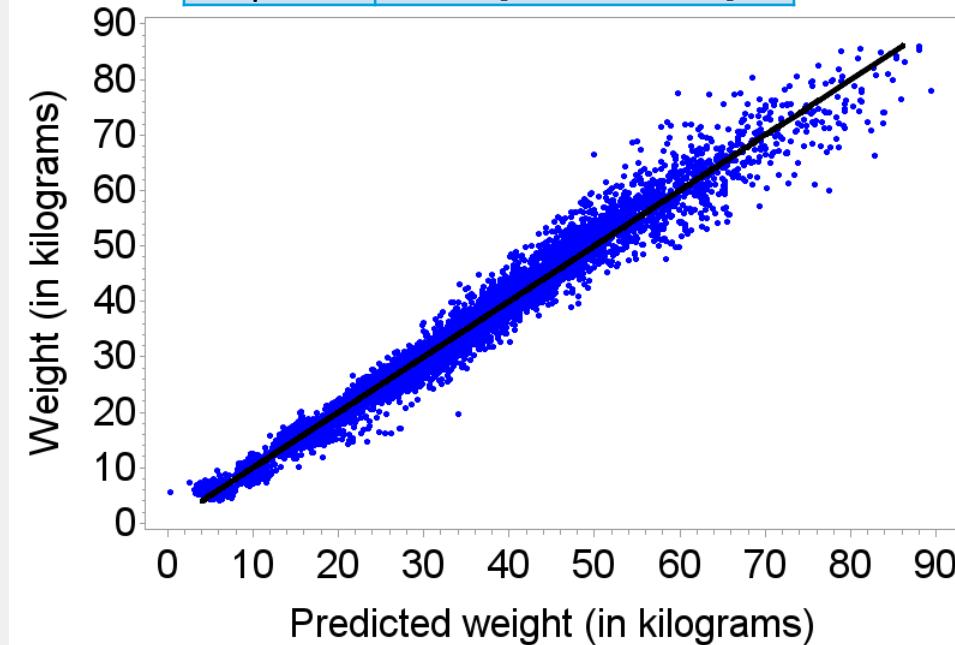
## Fractional polynomials (FPs)

### Example: [Growth curve data]

- Using the BLUP's we can predict each individual weight value
- Use option OUTP = in the MODEL statement
- Run a standard regression analysis
  - Outcome is the response weight
  - Independent variable is the prediction

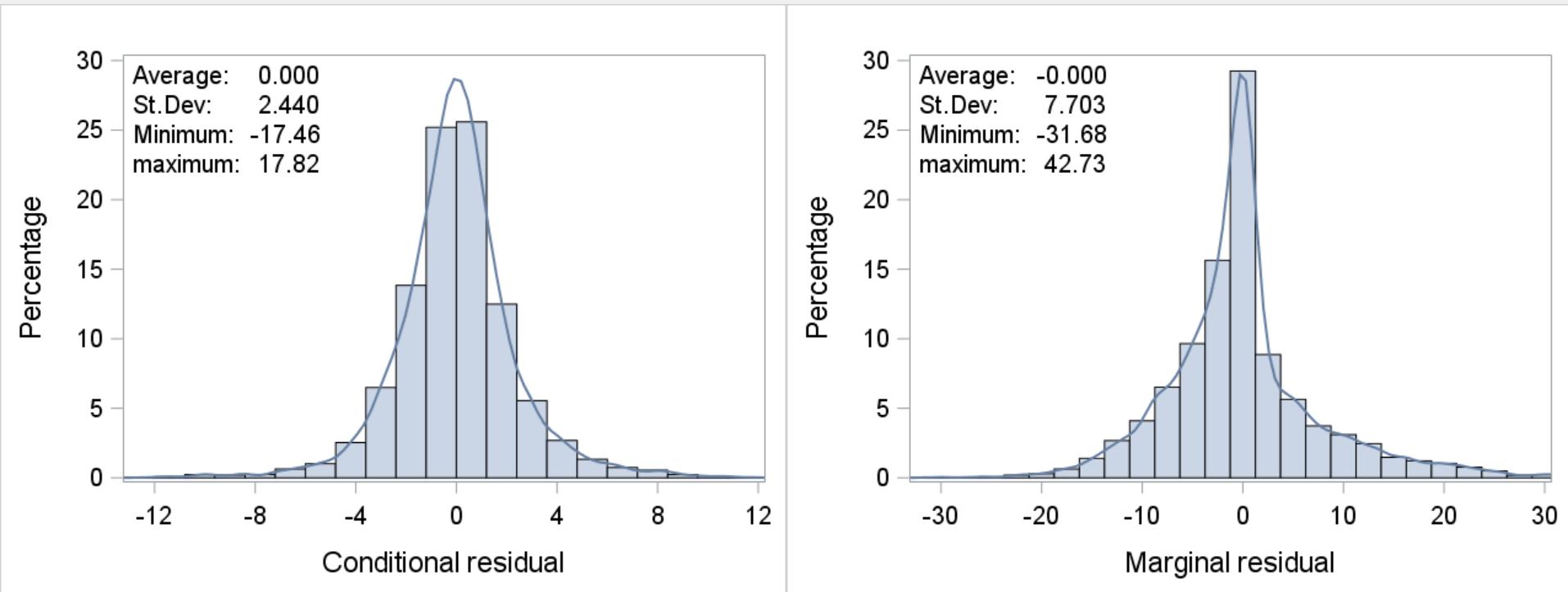
```
PROC GLM DATA=PRED  
    PLOTS (MAXPOINTS=8000);  
  
    MODEL RESP=PRED/CLPARM;  
  
RUN;
```

R-Square	Coeff Var	Root MSE	RESP Mean
0.979002	7.966345	2.439210	30.61893
Intercept	-0.119 [-0.233; -0.006]		
Slope	1.004 [1.001; 1.007]		



# Subject-specific models

## Fractional polynomials (FPs)



# Subject-specific models

## Exercise L4

### Data analytic questions:

- Consider the growth data on child weight
  - Use the logarithmic transformed weight as outcome
- Fit a fractional polynomial with  $p_1 = 1$  and  $p_2 = -0.5$  assuming
  - A separate average FP for sexes
  - Assume that the three coefficients have a multivariate normal distribution
  - Assume that the residual is i.i.d. normal
- Report on the following aspects:
  - Are the average FPs for sexes different?
  - Does the model predicts the observations
  - Do the residuals show better or worse patterns than the residuals in the course slides
  - Calculate reference limits for the weight growth of children for both sexes separately

# Marginal models

## Categorical time points

- Let  $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{in_i})^T$  be the set of repeated outcomes on unit  $i$
- Ignoring random effects leads to the linear mixed model:

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{e}_i$$

- With  $\mathbf{X}_i$  an observed matrix of covariates (**continuous** or **categorical**) for unit  $i$
- With  $\boldsymbol{\beta}$  a set of unknown fixed parameters (including an intercept)
- With  $\mathbf{e}_i$  the residual of unit  $i$  having zero mean and covariance matrix  $\mathbf{R}_i$
- This model works (best) when time points  $t_{ij} = t_j$  are aligned across units

## Covariance structures:

- Unstructured (TYPE=UN):
  - Only four time points ( $n_i = 4$ )
$$\begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} \end{pmatrix}$$
  - With  $\sigma_{rr} = \sigma_r^2 = \text{VAR}(e_{ir})$  the variance of the residual at time  $t_r$
  - With  $\sigma_{rs} = \sigma_{sr} = \text{COV}(e_{ir}, e_{is})$  and  $r \neq s$ , the covariance between residuals
- Other representation (TYPE=UNR)
  - $\sigma_{rs} = \rho_{rs} \sigma_r \sigma_s$
  - Correlations are reported

# Marginal models

## Categorical time points

### Example: [Pain data]

- The statistical model:

$$y_{ij} = \mu + \alpha_j + \sum_{h=1}^3 (\beta_h + \gamma_{hj}) x_{hi} + e_{ij}$$

- With  $i \in \{1, 2, \dots, m\}$  and  $j \in \{1, 2, \dots, n\}$
- With  $\mu$  the average at time point  $n$  for one of the treatments
- With  $\alpha_j$  effect of time point  $j$  ( $\alpha_n = 0$ )
- With  $x_{hi}$  an indicator variable for treatment  $h \in \{1, 2, 3\}$
- With  $\beta_h$  the effect of treatment  $h$  ( $\beta_3 = 0$ )
- With  $\gamma_{hj}$  the interaction effect of time  $j$  and treatment  $h$  ( $\gamma_{hn} = 0, \gamma_{3j} = 0$ )
- With  $e_i \sim N(0, R_i)$ ,  $R_i$  unstructured

### SAS programming codes:

```
PROC MIXED DATA=DATA_LONG CL;
  CLASS TRT (REF='Placebo')
  SUBJECT TIME;
  MODEL RESP = TRT TIME
  TRT*TIME/SOLUTION CL
  DDFM=SAT;
  REPEATED TIME
  / SUBJECT=SUBJECT TYPE=UNR R;
  RUN;
```

- Only repeated statement is used
- Estimation procedure is REML
- Time can be continuous as well

# Marginal models

## Categorical time points

### Example: [Pain data]: Variance structure $R_i$

- Pattern in variances over time may fit well with a subject-specific model
- Alternative reasons for this pattern
  - At time 120, medication is not yet fully active
  - At time 165, medication is worn out
  - These assumptions may suggest that variance is treatment dependent
- Unexpected orders in correlations:
  - $\rho_{13}$  is estimated larger than  $\rho_{12}$
  - $\rho_{24}$  is estimated larger than  $\rho_{34}$

Covariance Parameter Estimates						
Cov Parm	Subject	Estimate	Alpha	Lower	Upper	
Var(1)	SUBJECT	17.2094	0.05	10.1863	35.1455	
Var(2)	SUBJECT	8.2415	0.05	4.8782	16.8311	
Var(3)	SUBJECT	8.5726	0.05	5.0741	17.5071	
Var(4)	SUBJECT	11.8823	0.05	7.0332	24.2664	
Estimated R Matrix for SUBJECT 13					Corr(2,1)	SUBJECT
Row	Col1	Col2	Col3	Col4	Corr(3,1)	SUBJECT
1	17.2094	3.4454	6.9699	1.0640	Corr(3,2)	SUBJECT
2	3.4454	8.2415	4.7308	4.3904	Corr(4,1)	SUBJECT
3	6.9699	4.7308	8.5726	3.1382	Corr(4,2)	SUBJECT
4	1.0640	4.3904	3.1382	11.8823	Corr(4,3)	SUBJECT

# Marginal models

## Categorical time points

### Example: [Pain data]

- Intercept 8.45 (1.22) represents the average for treatment P at time 165
- Effects  $-0.84$  (1.72) for G and  $-1.25$  (1.72) for R indicate difference with treatment P at time 165 minutes
- Effects 0.61 (1.84), 1.29 (1.19), and 1.20 (1.33) for times 120, 135, and 150 indicate difference with time 165 for treatment P
- Effects  $-1.23$  (2.60),  $-1.44$  (1.68), and  $-0.73$  (1.88) for times 120, 135, and 150 and treatment G indicate corrections on time and treatment G effects
- Similar for interaction effects of treatment R

Effect	TRT	TIME	Estimate	Solution for Fixed Effects							
				Standard Error	DF	t Value	Pr >  t	Alpha	Lower	Upper	
Intercept			8.4500	1.2187	21	6.93	<.0001	0.05	5.9155	10.9845	
TRT	G		-0.8375	1.7235	21	-0.49	0.6321	0.05	-4.4218	2.7468	
TRT	R		-1.2500	1.7235	21	-0.73	0.4763	0.05	-4.8343	2.3343	
TRT	Placebo		0	.	.	.	.	.	.	.	
TIME		120	0.6125	1.8359	21	0.33	0.7420	0.05	-3.2054	4.4304	
TIME		135	1.2875	1.1907	21	1.08	0.2918	0.05	-1.1888	3.7638	
TIME		150	1.2000	1.3313	21	0.90	0.3776	0.05	-1.5686	3.9686	
TIME		165	0	.	.	.	.	.	.	.	
TRT*TIME	G	120	-1.2250	2.5963	21	-0.47	0.6419	0.05	-6.6244	4.1744	
TRT*TIME	G	135	-1.4375	1.6840	21	-0.85	0.4029	0.05	-4.9395	2.0645	
TRT*TIME	G	150	-0.7250	1.8827	21	-0.39	0.7041	0.05	-4.6403	3.1903	
TRT*TIME	G	165	0	.	.	.	.	.	.	.	
TRT*TIME	R	120	-0.5750	2.5963	21	-0.22	0.8269	0.05	-5.9744	4.8244	
TRT*TIME	R	135	0.2375	1.6840	21	0.14	0.8892	0.05	-3.2645	3.7395	
TRT*TIME	R	150	-0.2750	1.8827	21	-0.15	0.8853	0.05	-4.1903	3.6403	
TRT*TIME	R	165	0	.	.	.	.	.	.	.	
TRT*TIME	Placebo	120	0	.	.	.	.	.	.	.	
TRT*TIME	Placebo	135	0	.	.	.	.	.	.	.	
TRT*TIME	Placebo	150	0	.	.	.	.	.	.	.	
TRT*TIME	Placebo	165	0	.	.	.	.	.	.	.	

# Marginal models

## Categorical time points

### Multiplicity:

- The output for fixed effects tests all individual parameters equal to zero

$$H_0: \alpha_j = 0$$

$$H_0: \beta_h = 0$$

$$H_0: \gamma_{hj} = 0$$

- At least 11 different tests in case study
- The type 1 error for testing all  $H_0$ 's is inflated if all  $H_0$ 's are true
  - The **family wise type 1 error** is approximately  $1 - (1 - \alpha)^K$
  - With  $K$  the number of tests
  - With  $\alpha$  the significance level per test

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
TRT	2	21	1.12	0.3463
TIME	3	21	1.15	0.3526
TRT*TIME	6	21	0.28	0.9401

### Overall test statistics:

- Combining the parameters per term:

$$H_0: \alpha_1 = 0, \alpha_2 = 0, \dots, \alpha_J = 0$$

$$H_0: \beta_1 = 0, \beta_2 = 0, \dots, \beta_H = 0$$

$$H_0: \gamma_{11} = 0, \gamma_{12} = 0, \dots, \gamma_{1J} = 0$$

$$\gamma_{21} = 0, \gamma_{22} = 0, \dots, \gamma_{2J} = 0$$

$$\vdots \quad \vdots \quad \vdots$$

$$\gamma_{H1} = 0, \gamma_{H2} = 0, \dots, \gamma_{HJ} = 0$$

- These combined hypothesis are estimated with an  $F$ -test
- When interactions are present, other effects are difficult to interpret

# Marginal models

## Categorical time points

### Example: [Pain data]

- To obtain the averages per treatment and time point can be obtained with statement:  
LSMEANS TRT\*TIME / CL;
  - CL option provides approximate confidence intervals
- P-values test the hypothesis:  
 $H_0: \mu_{hj} = 0$ 
  - With  $\mu_{hj}$  the mean of treatment  $h$  at time point  $j$
  - This is mostly an uninteresting hypothesis to test for

Least Squares Means											
Effect	TRT	TIME	Estimate	Standard Error	DF	t Value	Pr >  t	Alpha	Lower	Upper	
TRT*TIME	G	120	7.0000	1.4667	21	4.77	0.0001	0.05	3.9499	10.0501	
TRT*TIME	G	135	7.4625	1.0150	21	7.35	<.0001	0.05	5.3517	9.5733	
TRT*TIME	G	150	8.0875	1.0352	21	7.81	<.0001	0.05	5.9348	10.2402	
TRT*TIME	G	165	7.6125	1.2187	21	6.25	<.0001	0.05	5.0780	10.1470	
TRT*TIME	R	120	7.2375	1.4667	21	4.93	<.0001	0.05	4.1874	10.2876	
TRT*TIME	R	135	8.7250	1.0150	21	8.60	<.0001	0.05	6.6142	10.8358	
TRT*TIME	R	150	8.1250	1.0352	21	7.85	<.0001	0.05	5.9723	10.2777	
TRT*TIME	R	165	7.2000	1.2187	21	5.91	<.0001	0.05	4.6655	9.7345	
TRT*TIME	Placebo	120	9.0625	1.4667	21	6.18	<.0001	0.05	6.0124	12.1126	
TRT*TIME	Placebo	135	9.7375	1.0150	21	9.59	<.0001	0.05	7.6267	11.8483	
TRT*TIME	Placebo	150	9.6500	1.0352	21	9.32	<.0001	0.05	7.4973	11.8027	
TRT*TIME	Placebo	165	8.4500	1.2187	21	6.93	<.0001	0.05	5.9155	10.9845	

# Marginal models

## Categorical time points

### Example: [Pain data]: Estimating specific contrast

```
PROC MIXED DATA=DATA_LONG CL;
  CLASS TRT(REF='Placebo') SUBJECT TIME;
  MODEL RESP = TRT*TIME/NOINT SOLUTION CL DDFM=SAT;
  REPEATED TIME /SUBJECT=SUBJECT TYPE=UNR R;
  LSMEANS TRT*TIME/DIFF=ALL CL;
  ESTIMATE "TRT G AT TIME 135"
    TRT*TIME 0 1 0 0 0 0 0 0 0 0 0 0/CL;
  ESTIMATE "TRT P MINUS G AT TIME 135"
    TRT*TIME 0 -1 0 0 0 0 0 0 0 1 0 0/CL;
  ESTIMATE "TIME 165 MINUS 120 AT TRT G"
    TRT*TIME 0 0 0 0 0 0 0 -1 0 0 0 1/CL;
RUN;
```

# Marginal models

## Categorical time points

### Example: [Pain data]: Estimating specific contrast

Estimates								
Label	Estimate	Standard Error	DF	t Value	Pr >  t	Alpha	Lower	Upper
TRT G AT TIME 135	7.4625	1.0150	21	7.35	<.0001	0.05	5.3517	9.5733
TRT P MINUS G AT TIME 135	2.2750	1.4354	21	1.58	0.1279	0.05	-0.7101	5.2601
TIME 165 MINUS 120 AT TRT G	-0.6125	1.8359	21	-0.33	0.7420	0.05	-4.4304	3.2054

- Combinations of comparisons are also possible

```
ESTIMATE "AVERAGE DIFF TRT G MINUS P AT TIMES 135 AND 150"  
TRT*TIME 0 0.5 0.5 0 0 0 0 0 -0.5 -0.5 0/CL;
```

AVERAGE DIFF TRT G MINUS P AT TIMES 135 AND 150	-1.9188	1.2815	21	-1.50	0.1492	0.05	-4.5838	0.7463
-------------------------------------------------	---------	--------	----	-------	--------	------	---------	--------

# Marginal models

## Categorical time points

### Example: [Pain data]: Estimating specific contrast

- DIFF=ALL comes with several options for multiple testing:
- ADJUST=
  - **Bonferroni:** Divides significance level by number of tests  $\alpha_B = \alpha/K$
  - **Dunnett:** Same as Bonferroni, but now with respect to a control group
  - **Sidak:** Corrects significance level using powers  $\alpha_S = 1 - (1 - \alpha)^{1/k}$
  - **Simulation:** simulates from the joint  $t$ -distribution of all tests

Effect	TRT	TIME	_TRT	_TIME	Estimate	Differences of Least Squares Means						
						Standard Error	DF	t Value	Pr >  t	Alpha	Lower	Upper
TRT*TIME	G	120	G	135	-0.4625	1.5232	21	-0.30	0.7644	0.05	-3.6301	2.7051
TRT*TIME	G	120	G	150	-1.0875	1.2167	21	-0.89	0.3815	0.05	-3.6177	1.4427
TRT*TIME	G	120	G	165	-0.6125	1.8359	21	-0.33	0.7420	0.05	-4.4304	3.2054
TRT*TIME	G	120	R	120	-0.2375	2.0742	21	-0.11	0.9099	0.05	-4.5511	4.0761
TRT*TIME	G	120	R	135	-1.7250	1.7836	35.1	-0.97	0.3401	0.05	-5.3457	1.8957
TRT*TIME	G	120	R	150	-1.1250	1.7952	29.9	-0.63	0.5356	0.05	-4.7918	2.5418
TRT*TIME	G	120	R	165	-0.2000	1.9070	40.4	-0.10	0.9170	0.05	-4.0528	3.6528
TRT*TIME	G	120	Placebo	120	-2.0625	2.0742	21	-0.99	0.3314	0.05	-6.3761	2.2511
TRT*TIME	G	120	Placebo	135	-2.7375	1.7836	35.1	-1.53	0.1338	0.05	-6.3582	0.8832
TRT*TIME	G	120	Placebo	150	-2.6500	1.7952	29.9	-1.48	0.1504	0.05	-6.3168	1.0168
TRT*TIME	G	120	Placebo	165	-1.4500	1.9070	40.4	-0.76	0.4514	0.05	-5.3028	2.4028
TRT*TIME	G	135	G	150	-0.6250	0.9587	21	-0.65	0.5215	0.05	-2.6187	1.3687
TRT*TIME	G	135	G	165	-0.1500	1.1907	21	-0.13	0.9010	0.05	-2.6263	2.3263
TRT*TIME	G	135	R	120	0.2250	1.7836	35.1	0.13	0.9003	0.05	-3.3957	3.8457
TRT*TIME	G	135	R	135	-1.2625	1.4354	21	-0.88	0.3891	0.05	-4.2476	1.7226
TRT*TIME	G	135	R	150	-0.6625	1.4497	31.9	-0.46	0.6608	0.05	-3.6159	2.2909
TRT*TIME	G	135	R	165	0.2625	1.5860	34.3	0.17	0.8695	0.05	-2.9595	3.4845
TRT*TIME	G	135	Placebo	120	-1.6000	1.7836	35.1	-0.90	0.3758	0.05	-5.2207	2.0207
TRT*TIME	G	135	Placebo	135	-2.2750	1.4354	21	-1.58	0.1279	0.05	-5.2601	0.7101
TRT*TIME	G	135	Placebo	150	-2.1875	1.4497	31.9	-1.51	0.1412	0.05	-5.1409	0.7669
TRT*TIME	G	135	Placebo	165	-0.9875	1.5860	34.3	-0.62	0.5376	0.05	-4.2095	2.2345
TRT*TIME	G	150	G	165	0.4750	1.3313	21	0.36	0.7248	0.05	-2.2936	3.2436
TRT*TIME	G	150	R	120	0.8500	1.7952	29.9	0.47	0.6393	0.05	-2.8168	4.5168
TRT*TIME	G	150	R	135	-0.6375	1.4497	31.9	-0.44	0.6631	0.05	-3.5909	2.3159
TRT*TIME	G	150	R	150	-0.03750	1.4639	21	-0.03	0.9798	0.05	-3.0819	3.0069

# Marginal models

## Categorical time points

### Strategy multiplicity:

#### [Only specific effects are of interest]

- Examples:
  - Difference treatment G and P at time 165
  - Difference treatment R and P at time 165
- Test all the effects and use an appropriate correction method
  - For small number of effects without a specific control use Bonferroni or Sidak
  - For small number of effects with control use Dunnett
  - For larger number of effects use the simulation option

#### [No specific effects are of interest]

- Examples:
  - Is there a treatment and/or time effect
- Step 1: Investigate interaction terms
  - Not significant: Eliminate interaction terms
  - Significant: Time & treatment effect exists
  - All interesting effects can be calculated without correction of significance
- Step 2: Model with no interactions
  - Use the overall significance tests
  - Not significant: No effect of these factors
  - Significant: Calculate all interesting effects without correction of significance

# Marginal models

## Exercise L5

### Data analytic questions:

- Consider the AUC of the pain data
- Fit a marginal model with categorical time points
  - Include the effects time, treatment, and the interaction time and treatment
  - Use an unstructured residual variance
- Report on the following aspects:
  - The p-values of the overall test statistics for the fixed effects
  - Describe the correlation structure between time points
  - Report the difference (and its 95% CI) between treatment G and P and R and P at time point 165. Are they significant?
  - Report the difference (and its 95% CI) between the results at time 120 and the average of the three time points 135, 150, and 165 for the three treatments. Are they significant?

# Marginal models

## Continuous time profiles

### Polynomial time profiles:

- Only mean profile  $X_i\beta$ :

$$\sum_{r=0}^q \beta_r t_{ij}^r$$

- With coefficients  $\beta_k$  given by

$$\beta_k(x_i) = \beta_{0k} + \sum_{s=1}^p \beta_{sk} x_{is}$$

- With  $x_i$  the set of covariates

### Fractional polynomials:

- Only mean profile  $X_i\beta$ :

$$\beta_0 + \sum_{r=1}^m \beta_r t^{pr}$$

- With coefficients  $\beta_k$  given by

$$\beta_k(x_i) = \beta_{0k} + \sum_{s=1}^p \beta_{sk} x_{is}$$

- With  $x_i$  the set of covariates

### Covariance structure:

- No random effects:  $G$  does not exists
- Modeling the residual variance  $R_i$ 
  - Most general structure is TYPE=UN:

$$\begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} \end{pmatrix}$$

- Marginal model is different from subject-specific model:
  - The variance of  $y_{ij}$  is not a smooth function of time
  - We can't obtain profiles for individual units

# Marginal models

## Continuous time profiles

### Example: [Pain data]

```
PROC MIXED DATA=DATA_LONG CL;  
  CLASS TRT(REF='Placebo')  
  SUBJECT TIME;  
  
  MODEL RESP = TRT FU  
  TRT*FU/SOLUTION CL  
  DDFM=SAT;  
  
  REPEATED TIME/  
  SUBJECT=SUBJECT TYPE=UN R;  
  
RUN;
```

- Time is treated as continuous variable
  - Requires two time variable in data set: one for fixed effects and one for residuals

Solution for Fixed Effects										
Effect	TRT	Estimate	Standard Error	DF	t Value	Pr >  t	Alpha	Lower	Upper	
Intercept		9.5311	1.0961	21	8.70	<.0001	0.05	7.2516	11.8106	
TRT	G	-2.5142	1.5501	21	-1.62	0.1197	0.05	-5.7379	0.7095	
TRT	R	-1.2484	1.5501	21	-0.81	0.4297	0.05	-4.4720	1.9753	
TRT	Placebo	0	.	.	.	.	.	.	.	
FU		-0.1537	0.4242	21	-0.36	0.7207	0.05	-1.0360	0.7285	
FU*TRT	G	0.5064	0.5999	21	0.84	0.4081	0.05	-0.7412	1.7541	
FU*TRT	R	-0.04482	0.5999	21	-0.07	0.9412	0.05	-1.2925	1.2028	
FU*TRT	Placebo	0	.	.	.	.	.	.	.	
Estimated R Matrix for SUBJECT 13										
Row	Col1	Col2	Col3	Col4						
1	16.6971	3.6365	6.5488	1.5028						
2	3.6365	8.1766	4.9156	4.1918						
3	6.5488	4.9156	8.3480	3.3458						
4	1.5028	4.1918	3.3458	11.6990						
Type 3 Tests of Fixed Effects										
Effect	Num DF	Den DF	F Value	Pr > F						
TRT		2	21	1.32	0.2896					
FU		1	21	0.00	0.9996					
FU*TRT		2	21	0.52	0.6015					

# Marginal models

## Continuous time profiles

### Reference limits:

- The smooth profile provides an average result at each time point
- The residual variance only provides a variance estimate at discrete times
- Thus we have no variance estimate in between time points unless the
  - Variance over time is independent of time  
 $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2$
  - Variance is also modeled as function of time [outside scope], for example,  
 $\sigma_k^2 = \sigma_0^2 [\mu(t_k)]^p$

### Linear profile: [pain data]

- Statistical model

$$y_{ij} = \beta_{i0} + \beta_{i1} t_j + e_{ij}$$

- With  $\beta_{ik} = \beta_{1k}x_{1i} + \beta_{2k}x_{2i} + \beta_{3k}x_{3i}$
- With  $x_{hi} \in \{0,1\}$  indicator for treatment  $h$  such that  $x_{1i} + x_{2i} + x_{3i} = 1$
- Equal Toeplitz covariance matrix:

$$\sigma^2 \begin{pmatrix} 1 & \rho_{12} & \rho_{13} & \rho_{14} \\ \rho_{12} & 1 & \rho_{12} & \rho_{13} \\ \rho_{13} & \rho_{12} & 1 & \rho_{12} \\ \rho_{14} & \rho_{13} & \rho_{12} & 1 \end{pmatrix}$$

- Variance & correlation between sequential measurements independent of time

# Marginal models

## Continuous time profiles

### Linear profile: [pain data]

- Estimator for mean  $\mu(t)$ :
  - $\hat{\mu}_h(t) = \hat{\beta}_{h0} + \hat{\beta}_{h1}t$  is approximately normal when data is not too small
  - The mean is approximately:
$$\mathbb{E}(\hat{\mu}_h(t)) = \mu_h(t)$$
  - The variance is approximately:
$$\eta_h^2(t) = \text{VAR}(\hat{\mu}_h(t)) = \text{VAR}(\hat{\beta}_{h0}) + 2\text{COV}(\hat{\beta}_{h0}, \hat{\beta}_{h1})t + \text{VAR}(\hat{\beta}_{h1})t^2$$
- $y(t) - \hat{\mu}_h(t)$  approximately normal:
  - With  $y(t) - \hat{\mu}_h(t) \sim N(0, \sigma^2 + \eta_h^2(t))$
  - With  $\sigma^2 = \text{VAR}(y_{ij})$  independent of time [may depend on treatment if needed]

### Proposed reference limits:

$$U(t) = \hat{\mu}_h(t) + t_m^{-1}(1 - \alpha)(\hat{\sigma}^2 + \hat{\eta}_h^2(t))^{1/2}$$

$$L(t) = \hat{\mu}_h(t) - t_m^{-1}(1 - \alpha)(\hat{\sigma}^2 + \hat{\eta}_h^2(t))^{1/2}$$

- With  $\hat{\mu}_h(t)$  the (RE)ML estimate of  $\mu_h(t)$
- With  $\hat{\sigma}^2$  the (RE)ML estimate of  $\sigma^2$
- With  $\hat{\eta}_h^2(t)$  the (RE)ML estimate of  $\eta_h^2(t)$ , making use of COVB: the estimates of

$$\begin{pmatrix} \text{VAR}(\hat{\beta}_0) & \text{COV}(\hat{\beta}_0, \hat{\beta}_1) \\ \text{COV}(\hat{\beta}_0, \hat{\beta}_1) & \text{VAR}(\hat{\beta}_1) \end{pmatrix}$$

- With  $t_d^{-1}(q)$  the  $q$ th quantile of the  $t$ -distribution with  $d$  degrees of freedom
- With  $m$  the degrees of freedom: obtained from fixed effects or the number of units

# Marginal models

## Continuous time profiles

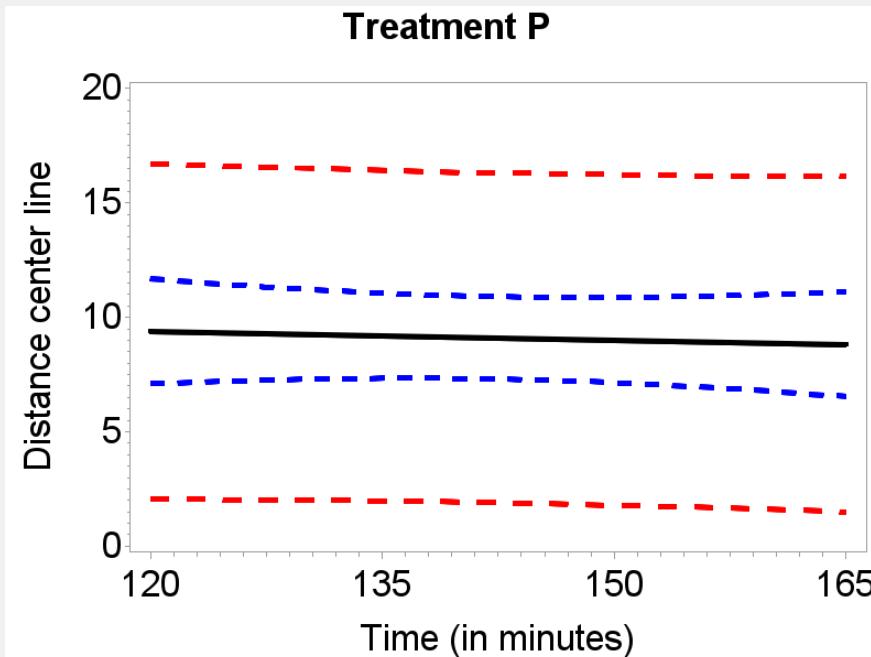
### Linear profile: [pain data]

- $100\% \times (1 - 2\alpha)$  confidence interval

$$U(t) = \hat{\mu}_h(t) + t_m^{-1}(1 - \alpha)\hat{\eta}_h(t)$$

$$L(t) = \hat{\mu}_h(t) - t_m^{-1}(1 - \alpha)\hat{\eta}_h(t)$$

```
PROC MIXED DATA=DATA_LONG CL;  
  
    CLASS TRT (REF='Placebo')  
    SUBJECT TIME;  
  
    MODEL RESP = TRT FU TRT*FU  
    /SOLUTION CL COVB DDFM=SAT;  
    REPEATED TIME /  
    SUBJECT=SUBJECT TYPE=TOEP;  
  
RUN;
```



- REML estimates ( $h = P$ ):

$$\hat{\mu}_P(t) = 9.387 - 0.1865t$$

$$\hat{\sigma}^2 = 11.376$$

$$\hat{\eta}_P^2(t) = 1.23 - 0.63t + 0.209t^2$$

# Marginal models

## Exercise L6

### Data analytic questions:

- Consider the growth data on child weight
  - Use the logarithmic transformed weight as outcome
- Fit a marginal model with fractional polynomial (with  $p_1 = 1$  and  $p_2 = -0.5$ )
  - Use a separate average FP for sexes
  - Assume that the structure of residual variance matrix is unstructured
- Report on the following aspects:
  - Are the average FPs for sexes different?
  - Calculate reference limits for the weight growth of children for both sexes separately
  - Compare these reference limits with the limits of the subject-specific model

# Model selection

## General strategy

### Modeling issues:

- Longitudinal data has 3 dimensions
  - Number of units ( $m$ )
  - Number of repeats ( $n$ )
  - Number of covariates ( $p$ )
- Difficult to formulate model selection for all settings:
  - $p \gg n$ : impossible to fit a model with all covariates included in model
  - $n \gg m$ : impossible to fit a multivariate normal (marginal) model
- Goal is to select parsimonious model to prevent overfitting

### Non-sparse setting: $[m \gg n, m \gg p]$

- General strategy
  - First obtain optimal covariance structure
  - Then obtain the fixed effects model under the optimal covariance structure
  - Proper covariance structure is needed to obtain correct fixed effects standard errors
- Step 1: Overspecify fixed effects model
  - Include any ***potential*** fixed effect term
  - In doubt: include, but make sure it makes sense (subject-matter knowledge)
  - The goal is to obtain left-over variability that is unexplained by any of the covariates

# Model selection

## General strategy

- Step 2: Select appropriate covariance structure using REML estimation
    - Use subject matter knowledge for expected correlation structures
    - Use graphical tools to examine correlation patterns in the data
    - Use **likelihood ratio tests** for nested covariance structures with the REML estimation method
    - Use **information criteria** (AIC, AICC, BIC) for comparing different (non-hierarchical) covariance structures
  - Step 3: Fix best covariance structure
  - Step 4: Select the best fixed effects model with ML estimation
    - Use **likelihood ratio tests** for nested fixed effects models
    - Use **information criteria** (AIC, AICC, BIC) for (non-hierarchical) fixed effects models
  - Step 5: Fix the fixed effects model
  - Step 6: Refit final model with REML estimates and collect residuals
  - Step 7: Investigate goodness-of-fit
- Iterations may be needed if the residuals and goodness-of-fit dictate this*

# Model selection

## Covariance structures

- Covariance structures can be obtained with `TYPE=`
- Covariance structures can be used for random and repeated statements
- Common options for  $G$ 
  - UN (or UNR)
  - VC for independent effects
- All structures would make sense for  $R_i$
- Structures with @ are for multivariate outcomes

Structure	Description	Parms	$(i, j)$ element
ANTE(1)	Antedependence	$2t - 1$	$\sigma_i \sigma_j \prod_{k=i}^{j-1} \rho_k$
AR(1)	Autoregressive(1)	2	$\sigma^2 \rho^{ i-j }$
ARH(1)	Heterogeneous AR(1)	$t + 1$	$\sigma_i \sigma_j \rho^{ i-j }$
ARMA(1,1)	ARMA(1,1)	3	$\sigma^2 [\gamma \rho^{ i-j -1} \mathbf{1}(i \neq j) + \mathbf{1}(i = j)]$
CS	Compound symmetry	2	$\sigma_1 + \sigma^2 \mathbf{1}(i = j)$
CSH	Heterogeneous CS	$t + 1$	$\sigma_i \sigma_j [\rho \mathbf{1}(i \neq j) + \mathbf{1}(i = j)]$
FA( $q$ )	Factor analytic	$\frac{q}{2}(2t - q + 1) + t$	$\sum_{k=1}^{\min(i,j,q)} \lambda_{ik} \lambda_{jk} + \sigma_i^2 \mathbf{1}(i = j)$
FA0( $q$ )	No diagonal FA	$\frac{q}{2}(2t - q + 1)$	$\sum_{k=1}^{\min(i,j,q)} \lambda_{ik} \lambda_{jk}$
FA1( $q$ )	Equal diagonal FA	$\frac{q}{2}(2t - q + 1) + 1$	$\sum_{k=1}^{\min(i,j,q)} \lambda_{ik} \lambda_{jk} + \sigma^2 \mathbf{1}(i = j)$
HF	Huynh-Feldt	$t + 1$	$(\sigma_i^2 + \sigma_j^2)/2 + \lambda \mathbf{1}(i \neq j)$
LIN( $q$ )	General linear	$q$	$\sum_{k=1}^q \theta_k \mathbf{A}_{ij}$
TOEP	Toeplitz	$t$	$\sigma_{ i-j +1}$
TOEP( $q$ )	Banded Toeplitz	$q$	$\sigma_{ i-j +1} \mathbf{1}( i-j  < q)$
TOEPH	Heterogeneous TOEP	$2t - 1$	$\sigma_i \sigma_j \rho_{ i-j }$
TOEPH( $q$ )	Banded hetero TOEP	$t + q - 1$	$\sigma_i \sigma_j \rho_{ i-j } \mathbf{1}( i-j  < q)$
UN	Unstructured	$t(t + 1)/2$	$\sigma_{ij}$
UN( $q$ )	Banded	$\frac{q}{2}(2t - q + 1)$	$\sigma_{ij} \mathbf{1}( i-j  < q)$
UNR	Unstructured corrs	$t(t + 1)/2$	$\sigma_i \sigma_j \rho_{\max(i,j) \min(i,j)}$
UNR( $q$ )	Banded correlations	$\frac{q}{2}(2t - q + 1)$	$\sigma_i \sigma_j \rho_{\max(i,j) \min(i,j)}$
UN@AR(1)	Direct product AR(1)	$t_1(t_1 + 1)/2 + 1$	$\sigma_{i_1 j_1} \rho^{ i_2 - j_2 }$
UN@CS	Direct product CS	$t_1(t_1 + 1)/2 + 1$	$\begin{cases} \sigma_{i_1 j_1} & i_2 = j_2 \\ \sigma^2 \sigma_{i_1 j_1} & i_2 \neq j_2 \\ 0 \leq \sigma^2 \leq 1 & \end{cases}$
UN@UN	Direct product UN	$t_1(t_1 + 1)/2 + t_2(t_2 + 1)/2 - 1$	$\sigma_{1,j_1 j_1} \sigma_{2,j_2 j_2}$
VC	Variance components	$q$	$\sigma_k^2 \mathbf{1}(i = j)$ and $i$ corresponds to $k$ th effect

# Model selection

## Covariance structures

Structure	Description	Parms	$(i, j)$ element
SP(EXP)(c-list )	Exponential	2	$\sigma^2 \exp\{-d_{ij}/\theta\}$
SP(EXPA)(c-list )	Anisotropic exponential	$2c + 1$	$\sigma^2 \prod_{k=1}^c \exp\{-\theta_k d(i, j, k)^{p_k}\}$
SP(EXPGA)(c <sub>1</sub> c <sub>2</sub> )	2D exponential, geometrically anisotropic	4	$\sigma^2 \exp\{-d_{ij}(\theta, \lambda)/\rho\}$
SP(GAU)(c-list )	Gaussian	2	$\sigma^2 \exp\{-d_{ij}^2/\rho^2\}$
SP(GAUGA)(c <sub>1</sub> c <sub>2</sub> )	2D Gaussian, geometrically anisotropic	4	$\sigma^2 \exp\{-d_{ij}(\theta, \lambda)^2/\rho^2\}$
SP(LIN)(c-list )	Linear	2	$\sigma^2 (1 - \rho d_{ij}) \mathbf{1}(\rho d_{ij} \leq 1)$
SP(LINL)(c-list )	Linear log	2	$\sigma^2 (1 - \rho \log(d_{ij}))$ $\times \mathbf{1}(\rho \log(d_{ij}) \leq 1, d_{ij} > 0)$
SP(MATERN)(c-list )	Matérn	3	$\sigma^2 \frac{1}{\Gamma(v)} \left( \frac{d_{ij}}{2\rho} \right)^v 2K_v(d_{ij}/\rho)$
SP(MATHSW)(c-list )	Matérn (Handcock-Stein-Wallis)	3	$\sigma^2 \frac{1}{\Gamma(v)} \left( \frac{d_{ij}\sqrt{v}}{\rho} \right)^v 2K_v \left( \frac{2d_{ij}\sqrt{v}}{\rho} \right)$
SP(POW)(c-list )	Power	2	$\sigma^2 \rho^{d_{ij}}$
SP(POWA)(c-list )	Anisotropic power	$c + 1$	$\sigma^2 \rho_1^{d(i,j,1)} \rho_2^{d(i,j,2)} \dots \rho_c^{d(i,j,c)}$
SP(SPH)(c-list )	Spherical	2	$\sigma^2 [1 - (\frac{3d_{ij}}{2\rho}) + (\frac{d_{ij}^3}{2\rho^3})] \mathbf{1}(d_{ij} \leq \rho)$
SP(SPHGA)(c <sub>1</sub> c <sub>2</sub> )	2D Spherical, geometrically anisotropic	4	$\sigma^2 [1 - (\frac{3d_{ij}(\theta, \lambda)}{2\rho}) + (\frac{d_{ij}(\theta, \lambda)^3}{2\rho^3})]$ $\times \mathbf{1}(d_{ij}(\theta, \lambda) \leq \rho)$

### Spatial-Temporal:

- For temporal correlations the *list* typically includes the time variable
  - Distance between time points are used
- For spatial correlations  $(x, y)$ -coordinates are typically used in  $(c_1, c_2)$
- The structures can be combined with GROUP= to stratify correlations

# Model selection

## Covariance structures

### Example: [pain data]

- Marginal model
  - Smooth linear time profile
  - Log linear autocorrelation

```
PROC MIXED DATA=DATA_LONG CL;  
  CLASS TRT(REF='Placebo')  
        SUBJECT TIME;  
  
  MODEL RESP = TRT FU TRT*FU  
    /SOLUTION CL COVB DDFM=SAT;  
  
  REPEATED TIME/SUBJECT=SUBJECT  
    TYPE=SP(LINL)(TIME) R;  
  
RUN;
```

Estimated R Matrix for SUBJECT 13									
Row	Col1	Col2	Col3	Col4					
1	11.5906	5.2038	3.5691	2.6129					
2	5.2038	11.5906	5.2038	3.5691					
3	3.5691	5.2038	11.5906	5.2038					
4	2.6129	3.5691	5.2038	11.5906					
Covariance Parameter Estimates									
Cov Parm	Subject	Estimate	Alpha	Lower	Upper				
SP(LINL)	SUBJECT	0.2035	0.05	0.1466	0.3016				
Residual		11.5906	0.05	8.3882	17.0633				
Solution for Fixed Effects									
Effect	TRT	Estimate	Standard Error	DF	t Value	Pr >  t	Alpha	Lower	Upper
Intercept		9.4284	1.1373	53.8	8.29	<.0001	0.05	7.1480	11.7088
TRT	G	-2.2845	1.6084	53.8	-1.42	0.1613	0.05	-5.5095	0.9404
TRT	R	-1.6406	1.6084	53.8	-1.02	0.3123	0.05	-4.8655	1.5844
TRT	Placebo	0							
FU		-0.1958	0.4912	69	-0.40	0.6914	0.05	-1.1756	0.7841
FU*TRT	G	0.4302	0.6946	69	0.62	0.5377	0.05	-0.9555	1.8159
FU*TRT	R	0.1410	0.6946	69	0.20	0.8397	0.05	-1.2447	1.5267
FU*TRT	Placebo	0							

# Model selection

## Information criteria

- Likelihood function increases when the model complicates
  - Like the  $R^2$  in linear regression analysis
- Need a measure that “punishes” for the complexity of the model
  - Like the adjusted  $R^2$  in regression analysis
- In 1970, Hirotugu Akaike developed a likelihood based measure
  - Called **Akaike's information criterion (AIC)**
  - Many alternative information criteria were developed thereafter
  - We will discuss **AIC, AICC, BIC**
  - Is a measure for closeness to true model

### For mixed models:

- Akaike's information criterion
$$AIC = -2 \log(L_M) + 2\dim(M)$$
  - With  $M$  an indicator for the fitted model
  - With  $L_M$  the likelihood value
- Corrected AIC:
$$AIC = -2 \log(L_M) + \frac{2n}{n-\dim(M)-1} \dim(M)$$
  - AIC corrected for the total sample size  $n$
- Bayesian information criteria
$$BIC = -2 \log(L_M) + \log(m) \dim(M)$$
  - With  $m$  the number of units
  - The BIC punishes more than AICC, which punishes more than AIC

# Model selection

## Information criteria

### Marginal model: [pain data]

- Include fixed effects: TRT, TIME, TRT\*TIME [Step 1 model strategy]
- Fit different covariance structures: UN, TOEP, CS, LINL, AR(1)

Criteria	UN	TOEP	CS	LINL	AR(1)	$BIC = -2 \log(L_M) + \log(m) \dim(M) = 459.5 + \log(24) \times 2$
$-2 \log(L_M)$	443.0	451.7	456.6	456.7	459.5	
AIC	463.0	459.7	460.6	460.7	463.5	
AICC	466.0	460.2	460.7	460.9	463.6	$AICC = -2 \log(L_M) + \frac{2n \dim(M)}{n-\dim(M)-1} = 456.6 + \frac{2 \times 96 \times 2}{96-2-1}$
BIC	474.7	464.2	463.0	463.1	465.8	

- Best covariance structure:
  - Likelihood chooses Unstructured
  - AIC and AICC chooses Toeplitz
  - BIC chooses Compound Symmetry

$$AIC = -2 \log(L_M) + 2\dim(M) = 443.0 + 2 \times 10$$

Only covariance parameters are counted with REML

# Model selection

## Information criteria

### Marginal model: [pain data]

- Hierarchical models can be tested with likelihood ratio test
- CS is hierarchically part of Toeplitz:
  - Constraints:  $\rho_{12} = \rho_{13} = \rho_{14}$
  - Thus null hypothesis  $H_0: \rho_{12} = \rho_{13} = \rho_{14}$
- The LRT is defined by
$$LRT = -2 \log(M_0) - (-2 \log(M))$$
  - With  $M_0$  the model under  $H_0$
  - With  $M$  the full model
  - Distribution of  $LRT$  is approximately  $\chi^2$  with degrees of freedom equal to  $\dim(M) - \dim(M_0)$

- Toeplitz structure:

$$\sigma^2 \begin{pmatrix} 1 & \rho_{12} & \rho_{13} & \rho_{14} \\ \rho_{12} & 1 & \rho_{12} & \rho_{13} \\ \rho_{13} & \rho_{12} & 1 & \rho_{12} \\ \rho_{14} & \rho_{13} & \rho_{12} & 1 \end{pmatrix}$$

- Compound symmetry:

$$\sigma^2 \begin{pmatrix} 1 & \rho_{12} & \rho_{12} & \rho_{12} \\ \rho_{12} & 1 & \rho_{12} & \rho_{12} \\ \rho_{12} & \rho_{12} & 1 & \rho_{12} \\ \rho_{12} & \rho_{12} & \rho_{12} & 1 \end{pmatrix}$$

- Likelihood ratio test

$$\begin{aligned} LRT &= 456.6 - 451.7 = 4.9 \\ df &= 4 - 2 = 2 \\ P &= 0.087 [H_0 \text{ not rejected}] \end{aligned}$$

# Model selection

## Information criteria

### Marginal model: [pain data]

- Covariance structure is compound symmetry [Step 2 model strategy]
- For fixed effects we need to change to ML estimation
- Comparing models with: [INT TR TI TR\*TI], [INT TR TI], [INT TI], [INT]

Criteria	[INT TR TI TR*TI]	[INT TR TI]	[INT TI]	[INT]
$-2 \log(L_M)$	480.5	481.7	484.2	486.8
AIC	508.5	497.7	496.2	492.8
AICC	513.7	499.4	497.1	493.1
BIC	525.0	507.2	503.2	496.4

$$AIC = -2 \log(L_M) + 2\dim(M)$$
$$= 480.5 + 2 \times 14 = 508.5$$

All parameters are used for ML

$$AICC = -2 \log(L_M) + \frac{2n \dim(M)}{n - \dim(M) - 1}$$
$$= 481.7 + \frac{2 \times 96 \times 8}{96 - 8 - 1} = 499.4$$

$$BIC = -2 \log(L_M) + \log(m) \dim(M)$$
$$= 486.8 + \log(24) \times 6$$

# Model selection

## Information criteria

### Marginal model: [pain data]

- All information criteria chooses a model without any fixed effects
  - Treatment seems not relevant
  - Time seems not relevant
- The LRT between models
  - TR\*TI:  $481.7 - 480.5 = 1.2$  ( $df = 6$ )
  - TR:  $484.2 - 481.7 = 2.5$  ( $df = 2$ )
  - TI:  $486.8 - 484.2 = 2.4$  ( $df = 3$ )
  - Critical value at 1 degrees of freedom: 3.8415
- Testing can lead to irrelevant significant effects in big data
  - Information criteria are less affected

- The final marginal model is a simple one way ANOVA model:

$$y_{ij} = \mu + a_i + e_{ij}$$

- With  $\mu$  the overall mean
- With  $a_i \sim N(0, \sigma_G^2)$  a random effect for individual
- With  $e_{ij} \sim N(0, \sigma_E^2)$  i.i.d. normal
- Note that the one-way ANOVA can be written as

$$y_{ij} = \mu + \tilde{e}_{ij}$$

- With  $\mu$  the overall mean
- With  $\tilde{e}_i = (\tilde{e}_{i1}, \dots, \tilde{e}_{in})^T \sim N(\mathbf{0}, \mathbf{R}_i)$  and  $\mathbf{R}_i$  having a compound symmetry structure

# Model selection

## Information criteria

### Marginal model: [pain data]

```
PROC MIXED DATA=DATA_LONG CL;  
  CLASS TRT(REF='Placebo')  
  SUBJECT TIME;  
  MODEL RESP = /SOLUTION CL  
  DDFM=SAT;  
  REPEATED TIME/SUBJECT=  
  SUBJECT TYPE=CS R;  
RUN;
```

- Alternative model formulation  
 RANDOM SUBJECT;
- And no repeated statement

Estimated R Matrix for SUBJECT 13				
Row	Col1	Col2	Col3	Col4
1	11.1060	4.1584	4.1584	4.1584
2	4.1584	11.1060	4.1584	4.1584
3	4.1584	4.1584	11.1060	4.1584
4	4.1584	4.1584	4.1584	11.1060

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Alpha	Lower	Upper
CS	SUBJECT	4.1584	0.05	0.7042	7.6126
Residual		6.9476	0.05	5.1383	9.9197

Solution for Fixed Effects								
Effect	Estimate	Standard Error	DF	t Value	Pr >  t	Alpha	Lower	Upper
Intercept	8.1958	0.4956	23	16.54	<.0001	0.05	7.1706	9.2211

# Model selection

## Information criteria

### Subject-specific model: [pain data]

- Overspecify fixed effects [Step 1]
  - Use a quadratic time profile for individuals
  - Coefficients are 3-dimensionally normal
  - One average profile per treatment
  - Make sure that time is standardized
- When we use random effects, we should not overspecify  $R_i$ 
  - Covariance of  $y_i$  is  $V_i = Z_i G Z_i^T + R_i$  and  $G$  already provides structure to  $V_i$
  - Complex  $R_i$  leads to non-identifiability
  - Usually we restrict  $R_i$  to at most autoregressive correlation structures

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
UN(1,1)	SUBJECT	9.7083
UN(2,1)	SUBJECT	-3.9176
UN(2,2)	SUBJECT	1.9400
UN(3,1)	SUBJECT	0.5111
UN(3,2)	SUBJECT	-0.1346
UN(3,3)	SUBJECT	0
Residual		5.4057

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
TRT	2	21.3	0.68	0.5181
FU	1	34.4	2.67	0.1112
FU*FU	1	45	3.37	0.0731
FU*TRT	2	34.4	0.08	0.9188
FU*FU*TRT	2	45	0.21	0.8145

### Suggested model gives issues

```
PROC MIXED DATA=DATA_LONG;
  CLASS TRT (REF='Placebo')
  SUBJECT;
  MODEL RESP = TRT FU FU*FU
  TRT*FU TRT*FU*FU/SOLUTION
  CL DDFM=SAT;
  RANDOM INT FU FU*FU/SUBJECT
  =SUBJECT TYPE=UN;
  RUN;
```

# Model selection

## Information criteria

### Subject-specific model: [pain data]

- Thus no evidence for quadratic form:
  - Not on an individual level
  - Not on an average level
- So, we assume linear profiles

```
PROC MIXED DATA=DATA_LONG;  
  CLASS RT (REF='Placebo')  
  SUBJECT;  
  
  MODEL RESP = TRT FU TRT*FU  
  /SOLUTION CL DDFM=SAT;  
  
  RANDOM INT FU/SUBJECT  
  =SUBJECT TYPE=UN;  
  
RUN;
```

### Covariance structures:

- We can play with both  $R_i$  and  $G$ 
  - $G$ : Intercept and slope are correlated or not: `TYPE=UN` and `TYPE=VC`
  - $R_i$ : Residuals are i.i.d., linear spatial or autoregressive one: `TYPE=none`, `TYPE=SP(LIN)`, `TYPE=AR(1)`
- Thus we must use `TYPE=` in random and repeated statements
- All combinations can be studied: notation  $[G:R_i]$
- Use estimation technique REML

# Model selection

## Information criteria

### Subject-specific model: [pain data]

- Included fixed effects: TRT, FU, TRT\*FU [Step 1 model strategy]
- Investigating different covariance structures:

	[UN:LIN]	[UN:AR(1)]	[UN:NONE]	[VC:LIN]	[VC:AR(1)]	[VC:NONE]
$-2 \log(L_M)$	474.8	472.8	474.8	478.8	478.8	478.8
AIC	484.8	482.8	482.8	486.8	486.8	484.8
AICC	485.5	483.5	483.3	487.3	487.2	485.1
BIC	490.7	488.7	487.5	491.5	491.5	488.3

- Covariance structure [UN:LIN] showed Hessian problems
- Conclusions on covariance structure:
  - AR(1) residual correlation fits better than or is similar to linear correlation in residuals
  - Correlation between random effects fits better than uncorrelated random effects
  - Best correlation structure is [UN:NONE] for both AICC and BIC

# Model selection

## Information criteria

### Subject-specific model: [pain data]

- Fitting the best model so far results into the following results

Covariance Parameter Estimates						
Cov Parm	Subject	Estimate	Alpha	Lower	Upper	
UN(1,1)	SUBJECT	8.6246	0.05	4.1658	27.2735	
UN(2,1)	SUBJECT	-2.2234	0.05	-5.0915	0.6447	
UN(2,2)	SUBJECT	1.1150	0.05	0.4284	7.0556	
Residual		5.4935	0.05	3.8203	8.5739	

- Variances are different from zero
- Covariance is not significantly different from zero
  - Remember:  $FU = [t_{ij} - 120]/15$

Solution for Fixed Effects									
Effect	TRT	Estimate	Standard Error	DF	t Value	Pr >  t	Alpha	Lower	Upper
Intercept		9.5137	1.2485	21	7.62	<.0001	0.05	6.9174	12.1101
TRT	G	-2.3425	1.7656	21	-1.33	0.1989	0.05	-6.0144	1.3294
TRT	R	-1.5850	1.7656	21	-0.90	0.3795	0.05	-5.2569	2.0869
TRT	Placebo	0	.	.	.	.	.	.	.
FU		-0.1925	0.5260	21	-0.37	0.7181	0.05	-1.2864	0.9014
FU*TRT	G	0.4387	0.7439	21	0.59	0.5616	0.05	-1.1083	1.9858
FU*TRT	R	0.1212	0.7439	21	0.16	0.8721	0.05	-1.4258	1.6683
FU*TRT	Placebo	0	.	.	.	.	.	.	.

Fit Statistics		Type 3 Tests of Fixed Effects					
-2 Res Log Likelihood	474.8	Effect	Num DF	Den DF	F Value	Pr > F	
AIC (Smaller is Better)	482.8	TRT		2	21	0.92	0.4153
AICC (Smaller is Better)	483.3	FU		1	21	0.00	0.9849
BIC (Smaller is Better)	487.5	FU*TRT		2	21	0.19	0.8320

# Model selection

## Information criteria

### Subject-specific model: [pain data]

- Assuming independence on random effects depends on shifting time
- Assume the following model

$$(1) \quad y_{ij} = b_{i0} + b_{i1}t_{ij} + e_{ij}$$

- With the random effects given by

$$\begin{pmatrix} b_{i0} \\ b_{i1} \end{pmatrix} \sim N \left( \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \begin{pmatrix} \tau_0^2 & \tau_{01} \\ \tau_{01} & \tau_1^2 \end{pmatrix} \right)$$

- If we shift the time scale with  $c$  we may assume the following model:

$$(2) \quad y_{ij} = \tilde{b}_{i0} + \tilde{b}_{i1}(t_{ij} - c) + e_{ij}$$

- With the random effects given by

$$\begin{pmatrix} \tilde{b}_{i0} \\ \tilde{b}_{i1} \end{pmatrix} \sim N \left( \begin{pmatrix} \tilde{\beta}_0 \\ \tilde{\beta}_1 \end{pmatrix}, \begin{pmatrix} \tilde{\tau}_0^2 & \tilde{\tau}_{01} \\ \tilde{\tau}_{01} & \tilde{\tau}_1^2 \end{pmatrix} \right)$$

- Rewriting the second model we get
  - $\tilde{b}_{i1} = b_{i1}$  and  $\tilde{b}_{i0} = b_{i0} + cb_{i1}$
  - $\tilde{\tau}_1^2 = \tau_1^2$ ,  $\tilde{\tau}_0^2 = \tau_0^2 + 2c\rho\tau_0\tau_1 + c^2\tau_1^2$ , and
  - $\tilde{\tau}_{01} = \tau_{01} + c\tau_1^2$
- If we choose (accidentally)  $c = -\tau_{01}/\tau_1^2$ 
  - Model (2) would prefer independence
  - Model (1) does not prefer independence
- Constant:  $c = \frac{2.2234}{1.115} = 1.994 \approx 2$ 
  - Thus  $FU = [t_{ij} - 120 - 30]/15$  would provide independent random effects

# Model selection

## Information criteria

### Subject-specific model: [pain data]

- Thus the shift parameter in the time variable affect the decision
  - We can always find a shift that forces us to use the unstructured  $\mathbf{G}$  matrix
  - We can use the ideal  $c$  such that we only require  $\tau_0^2$  and  $\tau_1^2$  in matrix  $\mathbf{G}$
  - We might even scale time such that  $\tau_0^2$  and  $\tau_1^2$  are equal
- Information criteria when we fit independent (dependent) random effects

	$[t_{ij} - 150]/15$	$[t_{ij} - 140]/15$	$[t_{ij} - 130]/15$	$[t_{ij} - 120]/15$
$-2 \log(L_M)$	474.8 (474.8)	475.7 (474.8)	477.6 (474.8)	478.8 (474.8)
AIC	480.8 (482.8)	481.7 (482.8)	483.6 (482.8)	484.8 (482.8)
AICC	481.1 (483.3)	481.9 (483.3)	483.9 (483.3)	485.1 (483.3)
BIC	484.3 (487.5)	485.2 (487.5)	487.1 (487.5)	488.3 (487.5)

- Thus shifts in the interval [130,150] may all prefer independence over dependence

# Model selection

## Information criteria

### Subject-specific model: [pain data]

- For investigating different fixed effects models, we
  - Consider  $FU = [t_{ij} - 120]/15$
  - Implement unstructured  $G$  matrix
  - Use ML estimation
  - Compare the models: [TR FU TR\*FU], [TR FU], and [FU]
  - Do not consider the model without time, since we have different time profiles for individuals and this requires a average time profile for all individuals
  - Use the information criteria

	[TR FU TR*FU]	[TR FU]	[FU]
$-2 \log(L_M)$	480.5	481.0	483.2
AIC	500.5	497.0	495.2
AICC	503.1	498.6	496.1
BIC	512.3	506.4	502.2

- All criteria suggest that treatment does not affect the linear time profiles
- Now this model can be compared with the best marginal model
  - AICC: 493.1
  - BIC: 496.2
- Thus the marginal model is best one

# Marginal models

## Exercise L7

### Data analytic questions:

- Consider the growth data on child weight
  - Use the logarithmic transformed weight as outcome
- Fit a marginal model with fractional polynomial (with  $p_1 = 1$  and  $p_2 = -0.5$ )
  - Use a separate average FP for sexes
  - Assume that the structure of residual variance matrix is unstructured
- Investigate different models which considers
  - Covariance structures for the residuals
  - The importance of sex in the model
- Report which model seems most reasonable