Faculty of Health Sciences

Linear mixed models

Analysis of repeated measurements, 10th March 2015

Julie Lyng Forman & Lene Theil Skovgaard

Department of Biostatistics, University of Copenhagen

Program

Topics:

- Random effects & variance components
- Linear mixed models in general.

Read: Fitzmaurice et al. (2011): chapters 8, 21, 22.

Examples:

- Random effects ANOVA
- Multi-level models
- Random regression
- Cross-over trials
- Comparison of measurement methods

Outline

What are repeated measurements?

Repeated measurements refer to data where the same outcome has been measured in different situations (or at different spots) on the same individuals.

- Special case: longitudinal means repeatedly over time.

What is clustered data?

Repeated measurements are termed clustered data when the same outcome is measured on groups of individuals from the same families/workplaces/school classes/villages/etc.

Analysis of repeated measurements

Many applications:

- Longitudinal data
- Treatments applied to multiple limbs, teeth, etc within the same person.
- Cross-over trials.
- Cluster randomized trials / multi-center studies.
- Comparisons / reliability of measurement methods.

ATT: Measurements belonging to the same subject/cluster are correlated. If we fail to take this correlation into account we will experience:

- p-values that are too small or too large.
- confidence intervals that are too wide or too narrow.

Outline

One-way analysis of variance – with **random** variation

Comparison of k groups or clusters, satisfying:

- The groups are of no individual interest and it is of no relevance to test whether they have identical means.
- The groups may be thought of as representatives from a population, that we want to describe.

Measurements belonging to the same subject/cluster tend to be correlated (look alike) due to e.g.

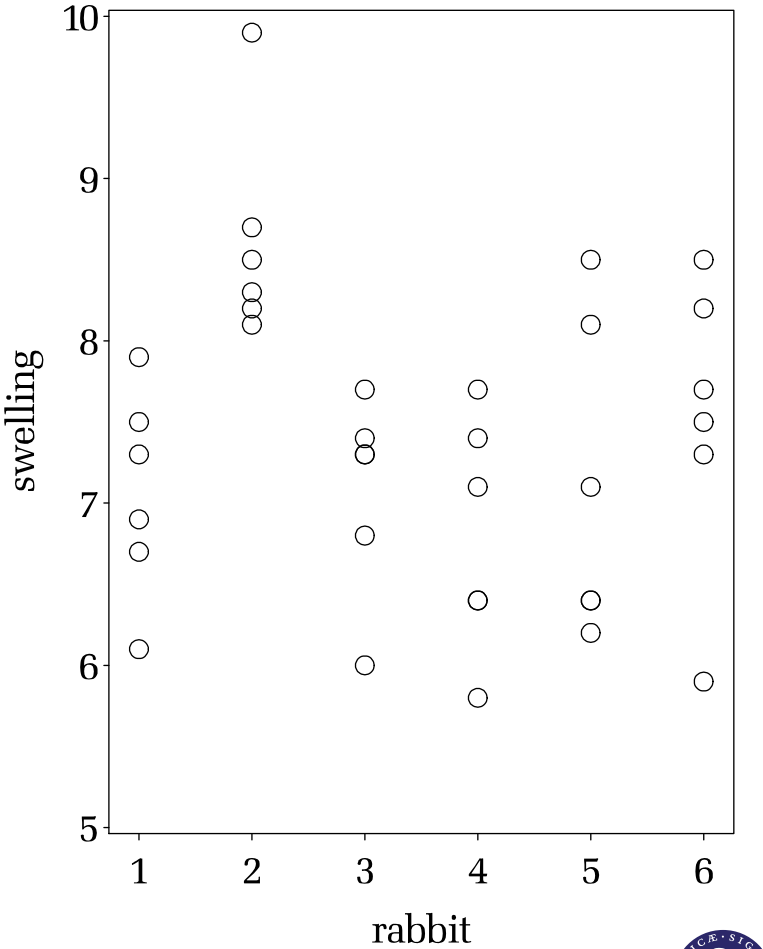
- Environmental variation.
  - Between regions, hospitals or countries.
- Biological variation.
  - Between individuals, families or animals.

Example: Rabbit data

- $R = 6$ rabbits vaccinated.
- In $S = 6$ spots on the back.

Response: swelling in cm^2

Research question:
How much swelling can be expected in reaction to the vaccine?



Random effects anova (the two-level model)

We let each rabbit have its own level of swelling described as

$$Y_{rs} = A_r + \varepsilon_{rs}$$

- We **assume** that these individual levels are randomly sampled from a normally distributed population,

$$A_r \sim \mathcal{N}(\mu, \omega_B^2)$$

- The error terms are considered to be independent normal,

$$\varepsilon_{rs} \sim \mathcal{N}(0, \sigma_W^2)$$

The rabbit levels are so-called random effects and the variances ω_B^2 and σ_W^2 are so-called variance components describing the variance **between rabbits** and **within rabbits**, respectively.

Implications of random effects anova

All observations are considered as randomly sampled measurements from the **same population**. Thus, the model implies that all measurements follow the same normal distribution:

$$Y_{rs} \sim N(\mu, \omega_B^2 + \sigma_W^2)$$

- Population mean μ, the grand mean.
- Population variance $\omega_B^2 + \sigma_W^2$, the total variation.

But: Measurements made on the same rabbit are correlated with the so-called intra-class correlation

$$\text{Corr}(y_{r1}, y_{r2}) = \rho = \frac{\omega_B^2}{\omega_B^2 + \sigma_W^2}$$

Compound symmetry

The implied covariance of the repeated measurements has a compound symmetry-structure:

$$\Sigma = (\omega_B^2 + \sigma_W^2) \cdot \begin{pmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & & \vdots \\ \rho & \rho & \dots & 1 \end{pmatrix}$$

In particular all pairs of spots on the same rabbit are assumed to be equally correlated (with the intra-class correlation).

- We say that the spots are exchangeable.

Note: If this is not the case, an unstructured covariance migth fit the data better. Say, if some spots are expected to respond more similarly than others.

Random effects ANOVA in PROC MIXED

```
PROC MIXED DATA=rabbit;
  CLASS rabbit spot;
  MODEL swelling = / S;
  RANDOM rabbit;
/* or REPEATED spot / TYPE=CS SUBJECT=rabbit; */
RUN;
```

Covariance Parameter Estimates

Cov Parm	Estimate
rabbit	0.3304
Residual	0.5842

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	7.3667	0.2670	5	27.59	<.0001

Estimation of variance components

Level	Variation	Variance component	Estimate	%of variation
1	Between	ω_B^2	0.3304	36%
2	Within	σ_W^2	0.5842	64%
	Total	$\omega_B^2 + \sigma_W^2$	0.9146	100%

Asymptotic standard errors can be obtained with:

```
PROC MIXED COVTEST DATA=rabbit;
```

- 95% CI for **Intra**-rabbit variation σ_W^2 : (0.37,1.04).
- 95% CI for **Inter**-rabbit variation ω_B^2 : (0.06,2.48).

BUT: The coverage may be poor in small samples.

Estimating variance components

In balanced data we have explicit formulae*:

$$\tilde{\sigma}^2_W = \mathrm{MS}_W \quad \text{and} \quad \tilde{\omega}^2_B = \mathrm{MS}_B - \frac{\mathrm{MS}_W}{n}$$

- n is the number of observations in each cluster
- MS_W and MS_B are Mean Squares within and between clusters, defined as in one-way ANOVA.

* This is deduced from

$$\begin{aligned} E(\mathrm{MS}_B) &= \omega^2_B + \frac{\sigma^2_W}{n} \\ E(\mathrm{MS}_W) &= \sigma^2_W \end{aligned}$$

Describing variation

Typical differences between spots on the **same** rabbit:

$$\begin{aligned} y_{rs_1} - y_{rs_2} &= \varepsilon_{rs_1} - \varepsilon_{rs_2} \\ &\sim N(0, 2\omega_W^2) \end{aligned}$$

- Normal region: $\pm 2\sqrt{2\omega_W^2} = \pm 2.16\ cm^2$

Typical differences between spots on **different** rabbits:

$$\begin{aligned} y_{r_1 s_1} - y_{r_2 s_2} &= \alpha_{r_1} - \alpha_{r_2} + \varepsilon_{r_1 s_1} - \varepsilon_{r_2 s_2} \\ &\sim N(0, 2\sigma_B^2 + 2\omega_W^2) \end{aligned}$$

- Normal region: $\pm 2\sqrt{2\sigma_B^2 + 2\omega_W^2} = \pm 2.70\ cm^2$

Why not use traditional one-way anova?

```
PROC GLM DATA=rabbit;
  CLASS rabbit spot;
  MODEL swelling = rabbit / NOINT SOLUTION;
  ESTIMATE 'grand mean' rabbit 0.167 0.167 0.167 0.167 0.167 0.167;
RUN;
```

- **Test of** $H_0: \ \mu_1 = \ldots = \mu_6$: $P = 0.004$.
- **Estimate of grand mean:** 7.367 (0.127)

But: We are not interested in these particular 6 rabbits, only in rabbits in general, as a **species**!

- **Estimate from mixed model:** 7.367 (0.267)

One-way anova with and without random variation

Classical one-way anova

- The rabbit means μ_r are fixed parameters,
  - supposedly of an interest of their own.
- We say that the rabbit factor is a fixed effect.

Random effects one-way anova

- The rabbit levels A_r are considered random and their population mean μ and variance $\omega_B^2 + \sigma_W^2$ is the major interest.
- We say that the rabbit factor is a random effect.
- (If data is from a pilot study used in the planning of some trial, the intra-class correlation will also be of interest).

Fixed or random effect?

How do we decide whether a **factor** should be modeled as fixed or random?

Fixed

- The specific values of the factor have been predetermined when planning the study.
- Allows inference for these particular values only.
- Demands a decent number of observations in each group.

Random

- A representative sample of values of the factor is present.
- Allows inference to be extended beyond the values in the experiment and to the population they were sampled from.

Estimation of individual rabbit means

Sometimes estimates of individual random effects are used for e.g. prediction of future disease status.

How do we estimate them?

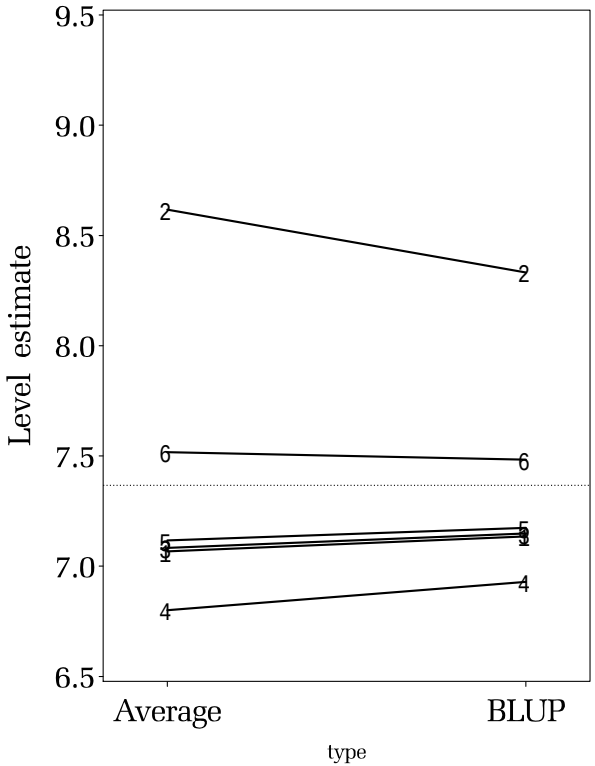
- Simple averages $\bar{y}_{r.}$ of the individual measurements.
- Best unbiased linear predictors (BLUPs) are **weighted averages** of the individual and the population mean:

$$\frac{\tilde{\omega}_B^2}{\tilde{\omega}_B^2 + \frac{\tilde{\sigma}_W^2}{S}} \bar{y}_{r.} + \frac{\frac{\tilde{\sigma}_W^2}{S}}{\tilde{\omega}_B^2 + \frac{\tilde{\sigma}_W^2}{S}} \bar{y}_{..}$$

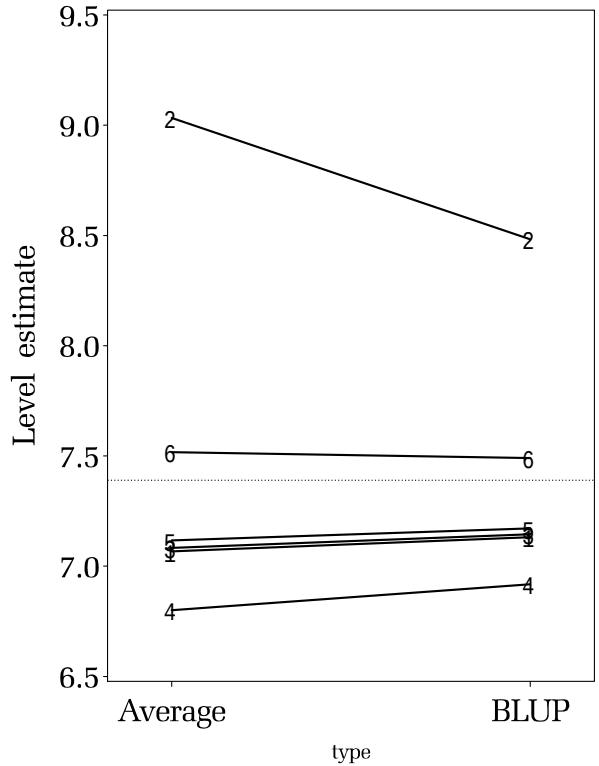
They have been **shrinked** towards the grand mean, $\bar{y}_{..}$;
We are *borrowing strenght from the neighbours*.

BLUPs vs averages

Full data



Reduced data



Note: We see larger shrinkage for rabbit no. 2 when the 3 smallest measurements from this rabbit have been removed.

Outline

General variance component models

Generalisations of ANOVA and GLM models involving several sources of random variation, so-called variance components.

Examples of sources of random variation:

- Environmental variation.
  - Between regions, hospitals or countries.
- Biological variation.
  - Between individuals, families or animals.
- Within-individual variation.
  - Between arms, teeth, days.
- Variation due to uncontrollable circumstances.
  - E.g. time of day, temperature, observer.
- Measurement error.

Multilevel models

Variance component models are also called multilevel models.

- Levels are most often hierarchical.
- We have variation, i.e. a variance component, on each level.
- And possibly systematic effects (covariates) on each level.

individual observation $\rightarrow$ **level 1**	$\rightarrow$	*context/cluster* $\rightarrow$ **level 2**	$\rightarrow$	*context/cluster* **level 3**
students	$\rightarrow$	classes	$\rightarrow$	schools
patient	$\rightarrow$	clinic	$\rightarrow$	regions
time	$\rightarrow$	subject	$\rightarrow$	
spot	$\rightarrow$	rabbit	$\rightarrow$	

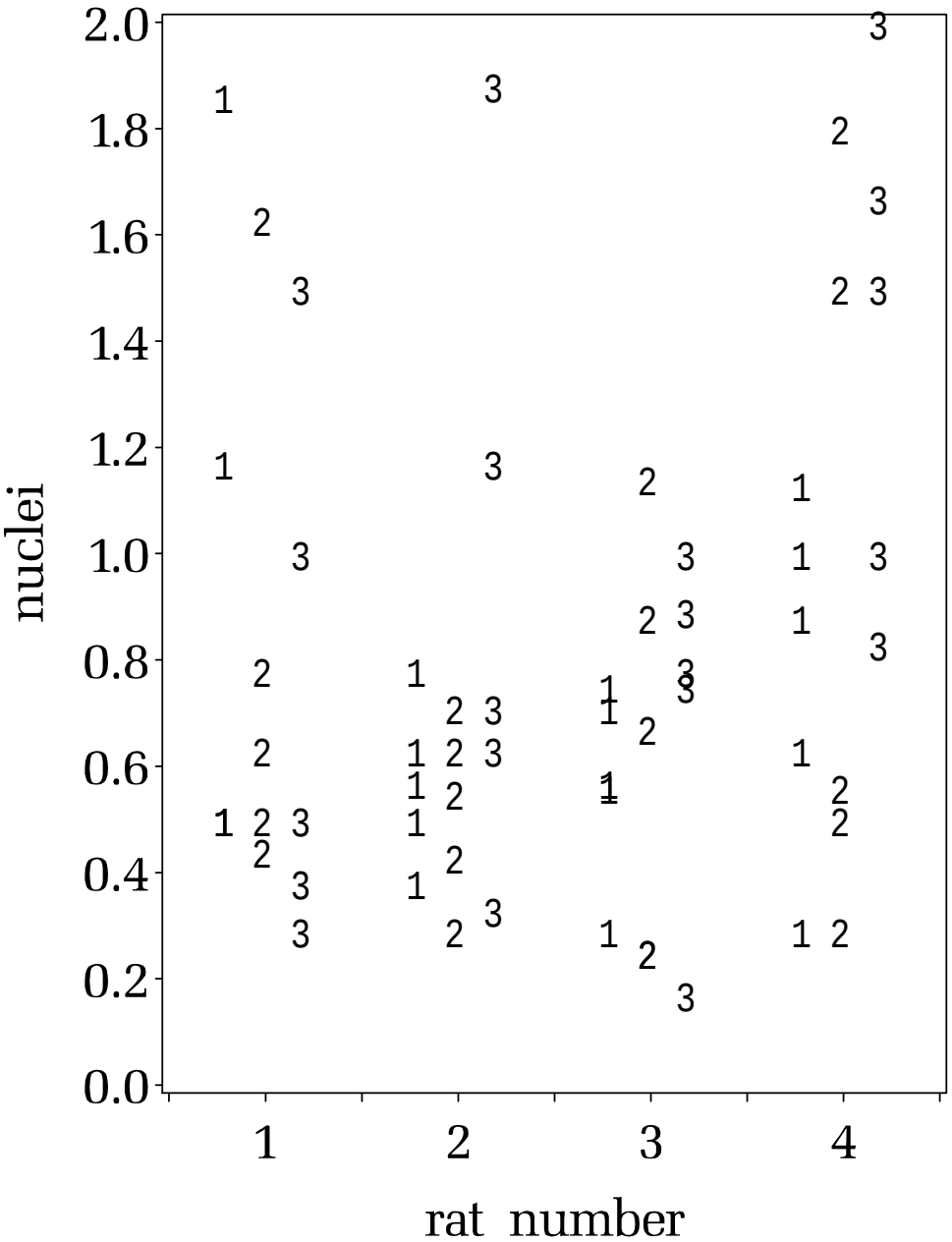
Example: A three-level model

Outcome: Number of nuclei per cell in the rat pancreas (used for the evaluation of cytostatica)

- $R = 4$ rats.
- $S = 3$ sections for each rat.
- $F = 5$ randomly chosen fields from each section.

level 1	$\rightarrow$	**level 2**	$\rightarrow$	**level 3**
fields	$\rightarrow$	sections	$\rightarrow$	rats
σ^2		τ^2		ω^2

Reference: Henrik Winther Nielsen, Inst. Med. Anat.


nuclei
rat number
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
1
2
3
4

Estimated variation and correlation

Level	Variation	Estimate
3	Rats (ω^2)	0.0179 (8.2%)
2	Sections (τ^2)	0.0029 (1.3%)
1	Fields (σ^2)	0.1968 (90.4%)
	Total	0.2176 (100%)

Measurements on	Correlation	Typical differences
Different rats	0	$\pm 2\sqrt{2(\omega^2+\tau^2+\sigma^2)} = \pm 1.319$
Different sections of the same rat	$\frac{\omega^2}{\omega^2+\tau^2+\sigma^2} = 0.082$	$\pm 2\sqrt{2(\tau^2+\sigma^2)} = \pm 1.264$
Different fields of the same section	$\frac{\omega^2+\tau^2}{\omega^2+\tau^2+\sigma^2} = 0.096$	$\pm 2\sqrt{2\sigma^2} = \pm 1.255$

Merits of multilevel models

We get a better understanding of the various sources of variation.

Effects *within* may be estimated more precisely (higher power), since some sources of variation are eliminated, e.g. by making comparisons within a family. This is analogous to the **paired design** situation.

When planning investigations, estimates of the variance components are needed in order to compare the power of various designs, and help us decide

- How many replicates do we need at each level?
- Should we randomize entire clusters or randomize *within* the clusters?

Design considerations

(**Note** the analogy with cluster-randomized trials.)

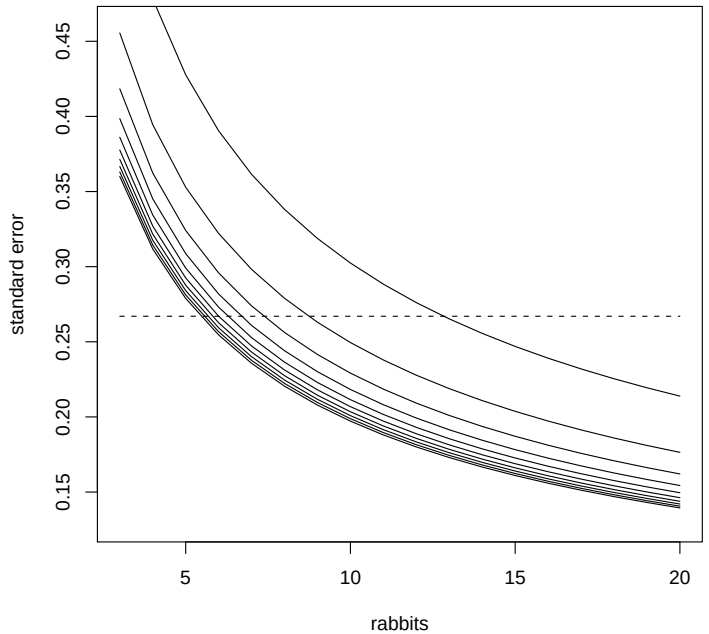
Plan an experiment with:

- R rabbits.
- S spots for each rabbit.
- $R \times S$ measurements.

Std. error of grand mean,

$$\mathrm{Var}(\bar{y}) = \frac{\omega_B^2}{R} + \frac{\sigma_W^2}{RS},$$

decreases with R and S.



The different curves correspond to S varying from 1 to 10.

Effective sample size

How many rabbits would we need to obtain the same precision in estimating the grand mean if we had **only one measurement** on each of R_1 rabbits?

Solve the equation for $\mathrm{Var}(\overline{y})$ to get:

$$R_1 = \frac{R \times S}{1 + \rho(S-1)}$$

where ρ is the within rabbit correlation.

- Estimate: $\rho = \frac{\omega_B^2}{\omega_B^2 + \sigma_W^2} = \frac{0.3304}{0.3304 + 0.5842} = 0.361 \Rightarrow R_1 = 12.8$

I.e. one measurement on each of thirteen rabbits gives the **same precision** as six measurements on each of six rabbits.

Case study: Cortisol

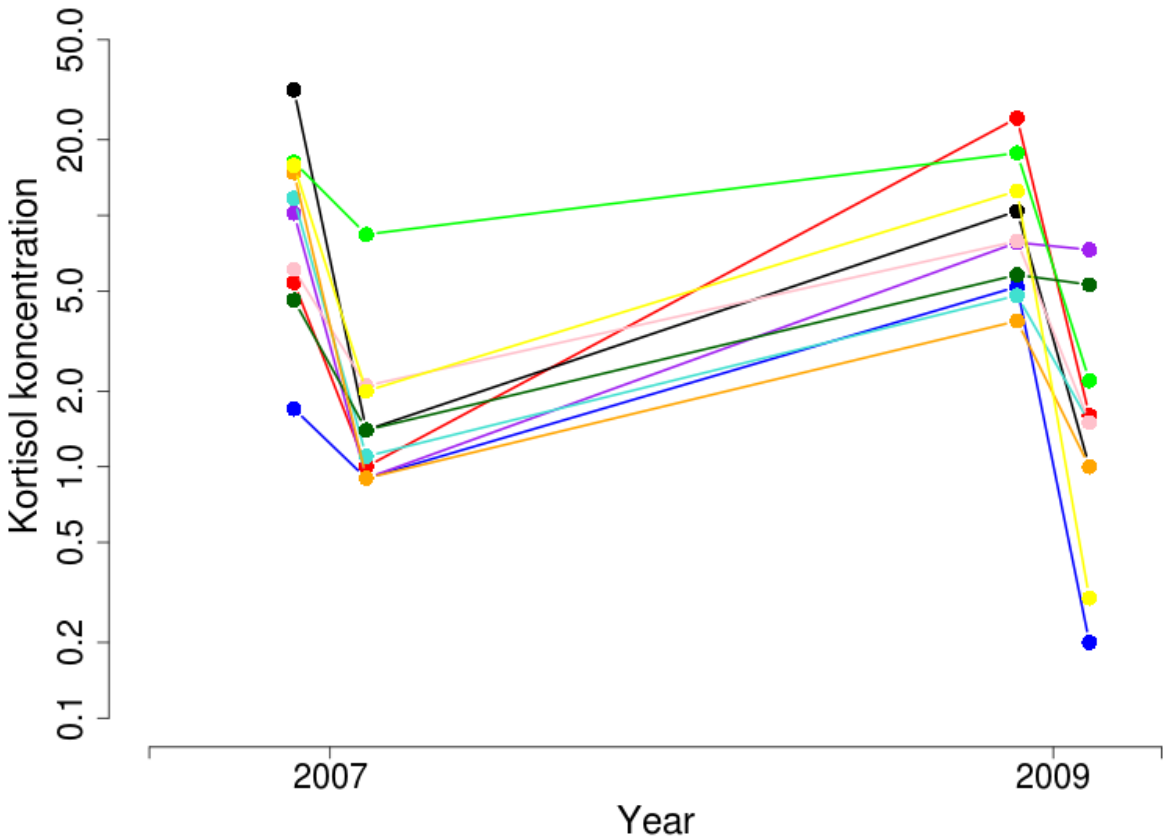
Outcome: Concentration of cortisol in salvia samples taken morning and evening in workers in Aarhus amt and kommune in 2007 (3536 participants) with similar follow-up in 2009 (2408 participants)

Interest: effect of stressors: lifeevents, Effort Reward Index

level	**variation**	**covariates**
3	between persons	`gender, age`
2	within person: between days	`bmi, stressors, year`
1	within person: within days	`time (morning/evening)`

Sample data

From 8 randomly selected men:



NOTE: concentrations on logarithmic scale.

Multilevel analysis

```
PROC MIXED DATA=prism COVTEST; WHERE sex EQ 'male';
  CLASS id year time;
  MODEL logcortisol = time  / SOLUTION CL DDFM=SATTERTH;
  RANDOM id id*year;
RUN;
```

Covariance Parameter Estimates

Cov Parm	Estimate	Std.Error	Z Value	Pr > Z
id	0.05993	0.01266	4.73	<.0001
id*year	0	.	.	.
Residual	0.5385	0.01794	30.01	<.0001

The *between days*-variance component estimate is a **zero**!

- Level 2 covariates (stressors) can only have **very little impact on individual cortisol koncentrations**!

Negative variance components

In case one of the variance component estimates becomes negative, SAS repports a zero.

What does it mean?

- The zero-estimate may be a chance finding due to statistical uncertainty.
- Or it might be the result of truly negative correlation within clusters - e.g. from competition (plants grown in same pot).

What can we do about it?

- Re-fit the model without the problematic random effect.
- Use a covariance pattern model which allows for negative correlation (e.g. an unstructured covariance).
- Include more covariates at the lower levels.

Estimated time-effect

Solution for Fixed Effects

Effect	time	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
Intercept		0.4106	0.02209	448	18.59	<.0001	0.05	0.3672	0.4540
time	morn	2.0137	0.02872	1305	70.12	<.0001	0.05	1.9573	2.0700
time	even	0	.	.	.	.	.	.	.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
time	1	1305	4916.89	<.0001

Estimates show that median levels of kortisol is about $\exp(2.0137) \simeq 7.49$ times higher in the morning than in the evening.

We should account for **exact time of measurement**!

Outline

Specification of linear mixed models (LMMs)

Mixed refers to *mixed* fixed and random effects.

Systematic variation

- covariates: time, treatment, gender, age, etc.,
  describing population parameters.

Random variation:

- Random effects,
  describing subject specific parameters.
- Serial correlation
- Measurement error

Interactions between systematic and random effects are always random effects.

Technical model description for LMMs

Model repeated outcomes on subject/cluster i as:

$$Y_i = X_i\beta + Z_i b_i + \varepsilon_i$$

- Systematic effects β with designmatrices X_i.
- Random effects b_i with designmatrices Z_i.
- Possibly dependent residual error terms ε_i

We assume that the b_i's and ε_i's are independent normally distributed with mean zero and covariance matrices given by:

- The G-matrix: $\mathrm{Var}(b_i) = G$.
- The R-matrix: $\mathrm{Var}(\varepsilon_i) = R$.

Implied covariance for LMMs

The covariance of the repeated measurements on subject/cluster i is given by the general formula:

$$V_i = Z_i^T G Z_i + R_i$$

Note:

- This is the so-called V-matrix.
- Print with option `vcorr` in `proc mixed`.

SAS: PROC MIXED

`model`: describes the mean value structure (i.e. covariates / fixed effects)

`random`: describes the random effects

`repeated`: describes the residual covariance.

Very flexible modeling framework!

Example: It is possible to model, e.g.

- longitudinal series of measurements (2 levels) ...
- with repeated series on each subject and with different treatments along the way (3 levels) ...
- and subjects belonging to different clusters (4 levels).

Nonidentifiability

Warning: Make sure you understand your model!

- Modeling random effects together with a residual error covariance may result in unidentifiable covariance parameters, i.e. nonconvergence, unless done with some care.

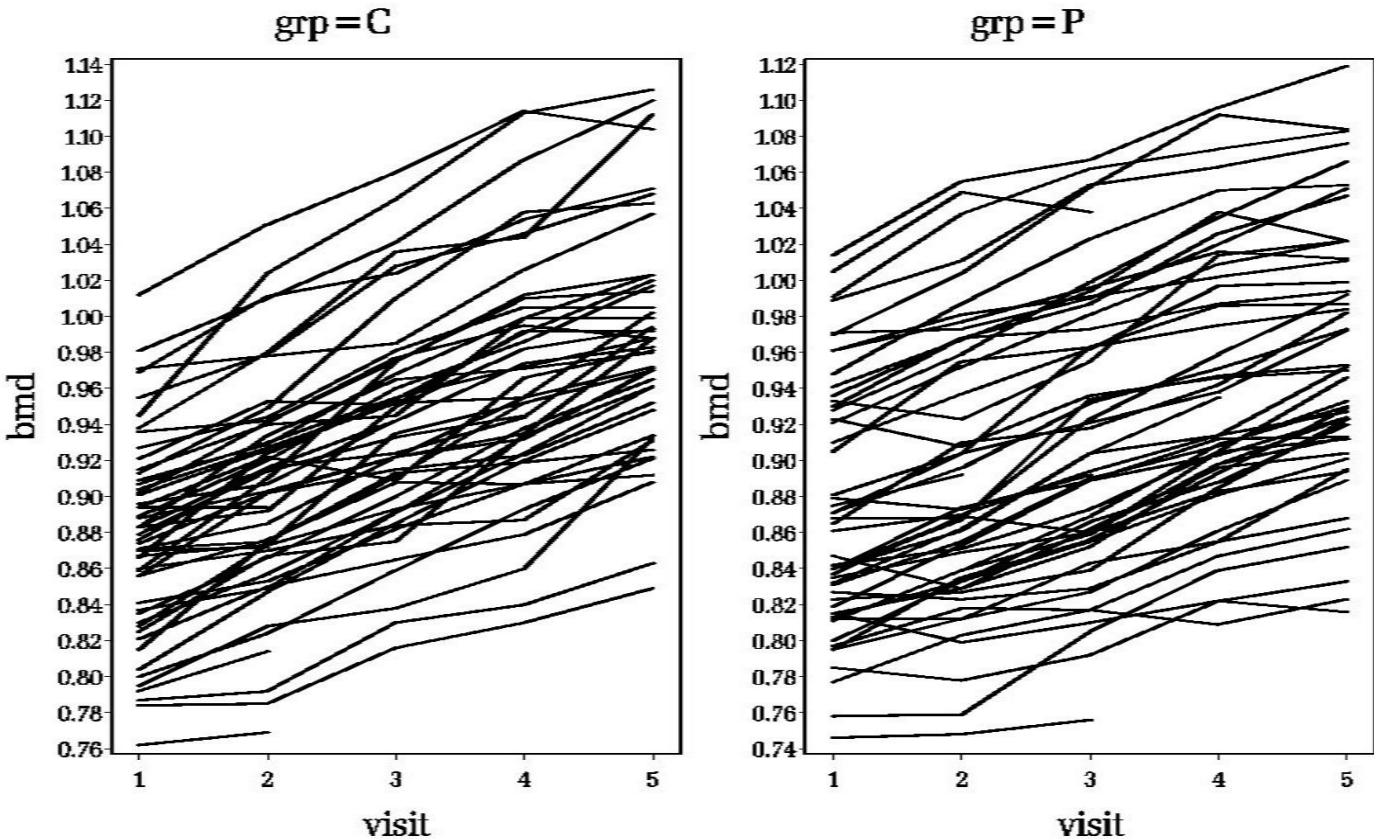
Example: Compound symmetry can be specified as either of:

- `RANDOM id;`
- `RANDOM intercept / SUBJECT=id;`
- `REPEATED time / TYPE=CS SUBJECT=id;`

in case two of these lines are included in the **same program**, it will not converge.

Outline

Calcium data



The time course looks reasonably linear, but maybe the individual girls have different growth rates ...

Random regression

We let each girl have her own level A_i and her own slope B_i

We **assume** these individual 'parameters', A_i and B_i,

- the random effects

follow a bivariate normal distribution in the population

$$\begin{pmatrix} A_i \\ B_i \end{pmatrix} \sim N_2 \left(\begin{pmatrix} \alpha_{g(i)} \\ \beta_{g(i)} \end{pmatrix}, \begin{pmatrix} \tau_a^2 & \omega_{ab} \\ \omega_{ab} & \tau_b^2 \end{pmatrix} \right)$$

The covariance is the so-called G-matrix:

- it describes the **population variance** of the lines, i.e. the inter-individual variation.

PROC MIXED: random regression

```
PROC MIXED DATA=calcium;
CLASS grp girl;
MODEL bmd=visit1 grp*visit1 / SOLUTION DDFM=SATTERTHWAITE;
RANDOM intercept visit1 / TYPE=UN SUBJECT=girl(grp) G;
RUN;
```

Individual intercepts and slopes must be specified in the **`random`-statement**.

- Here `visit` is used as a continuous covariate, with the intercept moved to `visit=1`. Due to randomization at baseline the main effect of `grp` omitted so that intercepts are the same in both groups.
- Note that `type=un` refers to a unstructured specification of the G-matrix. If it is omitted, we may experience convergence problems and sometimes totally incomprehensible results.

Output from random regression

Estimated G Matrix

Row	Effect	grp	girl	Col1	Col2
1	Intercept	C	101	0.004155	0.000051
2	visit1	C	101	0.000051	0.000048

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
Residual		0.000125

Fit Statistics

-2 Res Log Likelihood	-2347.7
AIC (smaller is better)	-2339.7

Solution for Fixed Effects

Effect	grp	Estimate	StdError	DF	t Value	Pr > |t|
Intercept		0.8752	0.006149	111	142.32	<.0001
visit1		0.02245	0.001097	96	20.46	<.0001
visit1*grp	C	0.004429	0.001570	96.5	2.82	0.0058
visit1*grp	P	0	.	.	.	.

We find an extra increase in BMD of **0.0044 (0.0016) g/cm^3** per half year, when giving calcium supplement.

Implied covariance

The random regression model implies a particular covariance-structure:

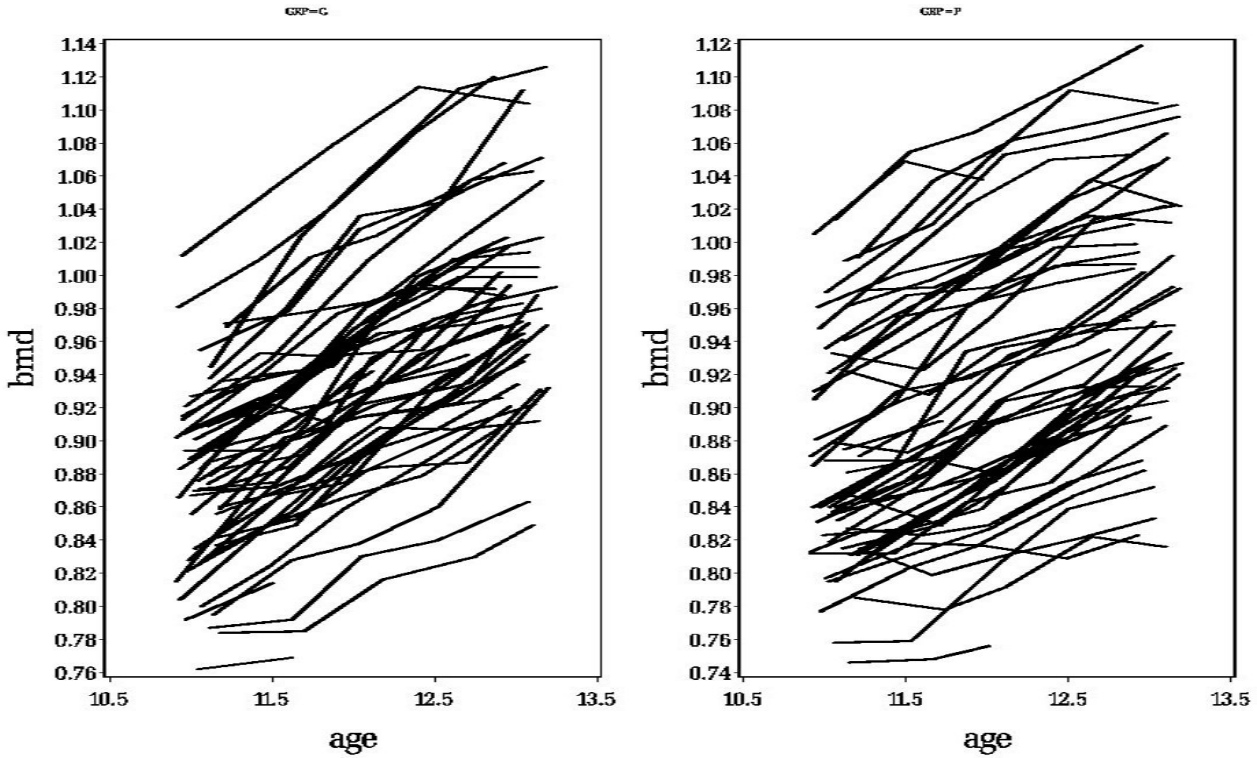
$$\begin{aligned}\mathrm{Cov}(Y_{ij}, Y_{ik}) &= \mathrm{Cov}(A_i + B_i t_j + \varepsilon_{ij}, A_i + B_i t_k + \varepsilon_{ik}) \\ &= \mathrm{Var}(A_i) + (t_j + t_k)\mathrm{Cov}(B_i, A_i) + t_j t_k \mathrm{Var}(B_i) \\ &= \tau_a^2 + (t_j + t_k)\omega_{ab} + t_j t_k \tau_b^2\end{aligned}$$

- Option `v` and `vcorr` makes SAS print the V-matrix and the associated correlation matrix.

Estimated V Matrix for girl(grp) 101 C

Row	Col1	Col2	Col3	Col4	Col5
1	0.004280	0.004207	0.004258	0.004309	0.004360
2	0.004207	0.004430	0.004405	0.004503	0.004602
3	0.004258	0.004405	0.004676	0.004698	0.004844
4	0.004309	0.004503	0.004698	0.005017	0.005086
5	0.004360	0.004602	0.004844	0.005086	0.005453

Nonequidistant time points



- The girls are only seen **approximately twice a year**.
- Perhaps we get better estimates of the slopes when replacing `visit` with the actual age of the girl.

Random regression, using actual age

Estimated G Matrix

Row	Effect	grp	girl	Col1	Col2
1	Intercept	C	101	0.004208	0.000095
2	age11	C	101	0.000095	0.000179

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
Residual		0.000124

Fit Statistics

-2 Res Log Likelihood	-2356.3
AIC (smaller is better)	-2348.3

Solution for Fixed Effects

Effect	grp	Estimate	StdError	DF	t Value	Pr > |t|
Intercept		0.8721	0.006193	111	140.84	<.0001
age11		0.04534	0.002151	96.2	21.08	<.0001
age11*grp	C	0.008803	0.003074	96.8	2.86	0.0051
age11*grp	P	0	.	.	.	.

In this model, we quantify the effect of a calcium supplement to **0.0088 (0.0031) g/cm^3** per **year**.

Results from random regression

Time variable	Difference in Slopes	P-value
`visit1`	0.0089 (0.0031)	0.0051
`age11`	0.0044 (0.0016)	0.0058
P	0.37	0.0048

Seemingly steeper slopes than when `visit` was used as the time-variable.

- Due to quantificantion (per year vs per 1/2 year)!

Note: In some cases replacing proxy age with exact age would result in steeper slopes due to bias reduction (recall measurement error in the independent variable causes bias towards the null).

Modeling the covariance

Random regression implies a particular covariance pattern.

- Does this fit the data well?

No benchmark for model comparisons:

- An **unstructured covariance** cannot be esimated from non-equidistant data!

Instead, non-nested models can be compared using Akaikes information criterion (AIC) which balances goodness of fit against model complexity.

- Smaller values of AIC indicates a better model fit.

Non-equidistant covariance patterns

In case subject are measured at individual or otherwise non-equally spaced time points only a limited number of stationary covariance pattern models are available:

- The variance is **constant over time**.
- The correlation **depend only on the time-distance** between the observations.

`proc mixed` `type=`	$\mathrm{Cov}(Y_{ij}, Y_{ik})$	no. param
`CS`	$\sigma^2[I\{j=k\} + \rho \cdot I\{j \neq k\}]$	2
`SP(POW)(ctime)`	$\sigma^2 \rho^{\lvert t_{ij} - t_{ik} \rvert}$	2
`SP(GAU)(ctime)`	$\sigma^2 e^{-\lvert t_{ij} - t_{ik} \rvert^2 / \gamma^2}$	2
`SP(LIN)(ctime)`	$\sigma^2(1 - \rho \lvert t_{ik} - t_{ij} \rvert) \cdot I\{\rho \lvert t_{ik} - t_{ij} \rvert \leq 1\}]$	2

The `ctime`-variable must be a **numerical variable** in `SAS`.

Tests of treatment effect

Comparison of slopes for different covariance structures:

Covariance structure	AIC	Cov.par.	Difference in slopes	P
Independence	-1251.3	1	0.0094 (0.0086)	0.27
Compound symmetry	-2253.9	2	0.0091 (0.0020)	$<$0.0001
Power (Autoregressive)	-2374.3	2	0.0099 (0.0030)	0.0014
Random Regression	-2348.3	4	0.0088 (0.0031)	0.0051

- Confidence intervals and tests depend on the covariance!

Outline

Example: Cross-over study of headache

Patients with chronic headache are randomized into two groups:

- Both groups receive LNMMA and placebo, on two different days, with a suitable wash-out period in-between
- Group G1 was treated first with placebo (period 1), and then with LNMMA (period 2)
- Group G2 was treated first with LNMMA (period 1), and then with placebo (period 2)

Pain was measured subjectively on a VAS-scale (small is good), at baseline and at 30, 60, 90 and 120 minutes after treatment.

Ashina, Lassen, Bendtsen, Jensen og Olesen (1999), Lancet, pp.287-289

Picture ignoring period effect and pairing

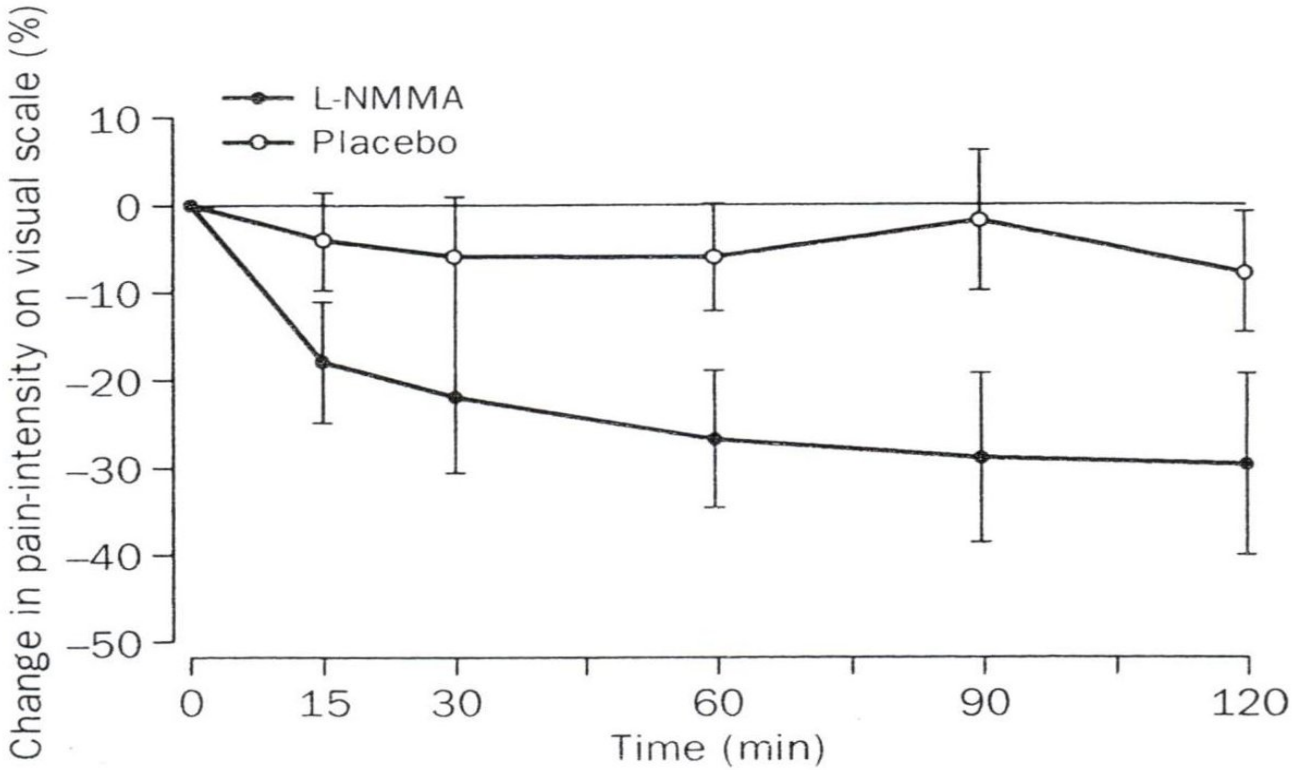


Figure 2: **Mean percentage change from baseline in pain intensity on 100 mm visual analogue scale**
Bars=SE.

Model building for cross over study

Fixed effect:

- `time`, `treat` `treat*time`, `period`
- possibly a carry-over effect: `treat*period(*time)`?

Covariance structure:

- We expect that observations from the same period (and same patient) are more strongly correlated when they are close in time, e.g.

```
RANDOM patient;
REPEATED time / TYPE=SP(POW)(time) LOCAL
                SUBJECT=patient*period;
```

where `LOCAL` adds an additional measurement error.

Extract data

Unfortunately, we do not have access to the full data with repeated measurements over time.

New outcome: Difference between average follow-up measurements and baseline,

$$Y_{30} + Y_{60} + Y_{120} - 3\,Y_0$$

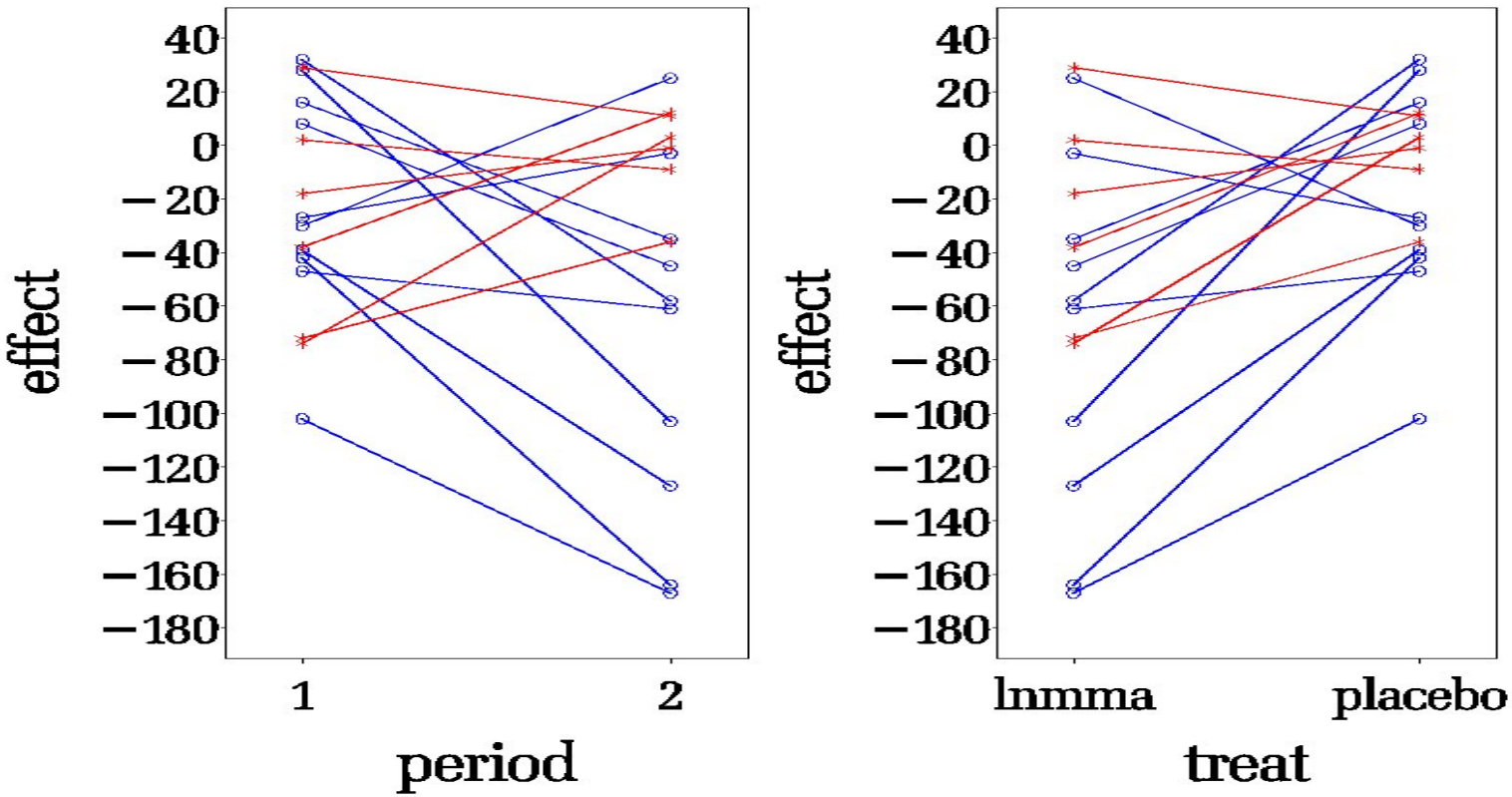
(recall, for this to be efficient the correlation must be strong).

Analysis Variable : effect

treat	period	N Obs	N	Mean	Std Dev
lnmma	1	6	6	-28.5000000	40.9865832
	2	10	10	-73.8000000	65.0022222
placebo	1	10	10	-20.3000000	41.5452899
	2	6	6	-3.3333333	17.8063659

Observations vs. period and treatment

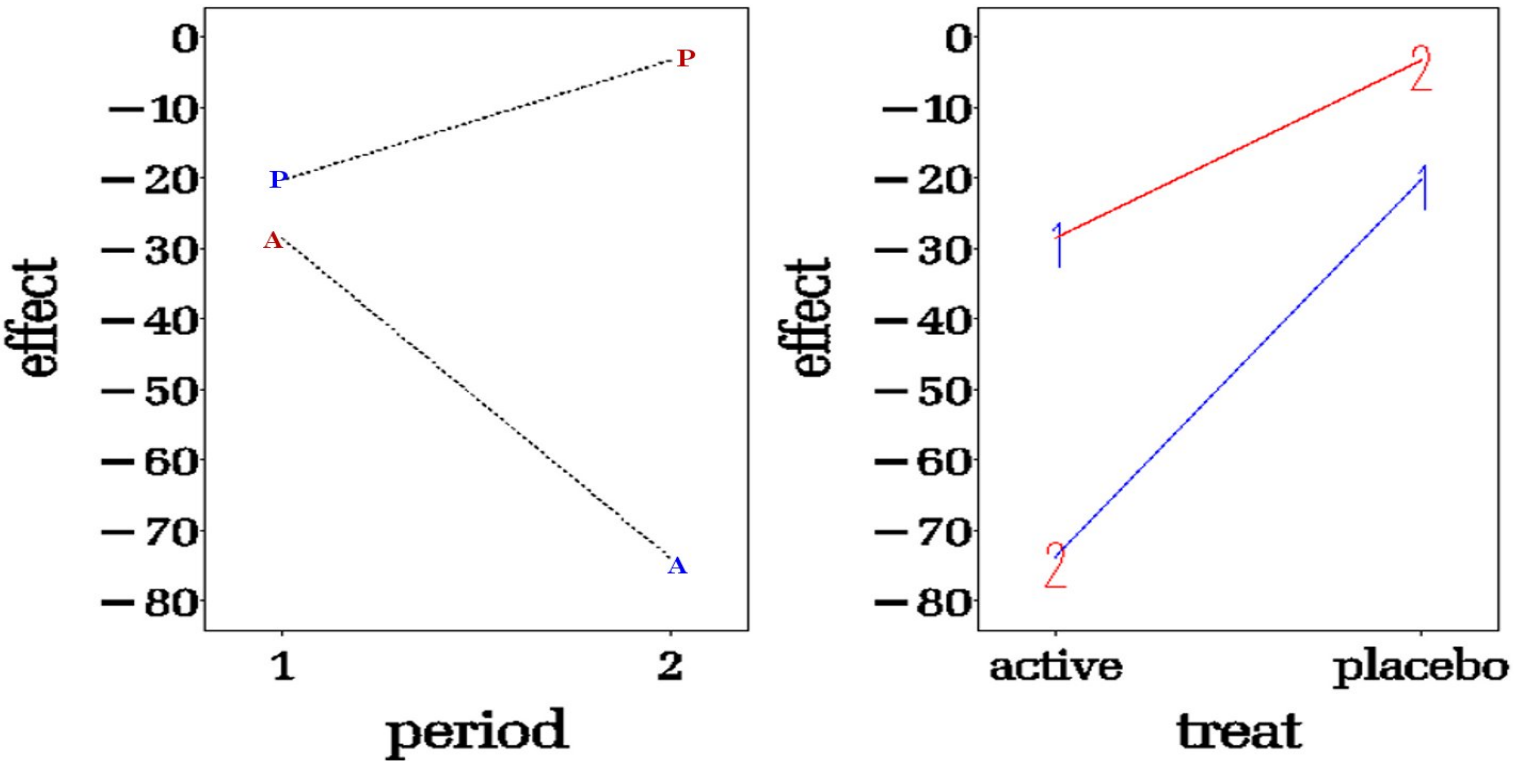
Legend: Group G1 (P+A), Group G2 (A+P)



Correlation looks reasonably strong.

Average over patients

A, P denote the treatments, 1 and 2 denote the periods



Seemingly much larger treatment effect in period 2.

Model for cross-over study

For subject i, treatment t and period p:

$$Y_{tpi} = \alpha + \beta_t + \gamma_p + \delta_{tp} + b_i + \varepsilon_{tpi}$$

- $b_i \sim N(0, \omega_B^2)$ are the random subject effect
- $\varepsilon_{tpi} \sim N(0, \sigma_W^2)$ are the residuals
- δ_{tp} is the carry-over effect.

Parameter of interest: Treatment effect in period 1.

Coded as a mixed effects model

```
PROC MIXED DATA=ashina;
  CLASS patient group treat period;
  MODEL effect=treat period treat*period / S CL DDFM=SATTERTH;
  RANDOM intercept / SUBJECT=patient(group);
RUN;
```

Solution for Fixed Effects

Effect	treat	period	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			-3.3333	19.4487	14	-0.17	0.8664
treat	lmmma		-70.4667	24.6009	14	-2.86	0.0125
treat	placebo		0	.	.	.	.
period		1	-16.9667	24.6009	14	-0.69	0.5017
period		2	0	.	.	.	.
treat*period	lmmma	1	62.2667	40.8798	14	1.52	0.1500
treat*period	lmmma	2	0	.	.	.	.
treat*period	placebo	1	0	.	.	.	.
treat*period	placebo	2	0	.	.	.	.

Interpretation of the carry-over effect

The carry-over effect is usually interpreted as an additional effect of placebo when given after the active treatment.

Estimate 62.3, with 95% CI $(-25.4, 149.9)$, i.e. nonsignificant.

The carry-over effect (placebo following active) has a positive value, corresponding to a worsening of the headache.

This could be explained as a psychological effect, in the sense that subjects expect something better (namely what they experienced in the previous period).

Traditional approach

First test the hypothesis $H_0 : \delta = 0$ (*no carry-over effect*):

- Unpaired T-test (G1 vs G2) with the sum of the two effects as outcome, since the group means are:
  - G1: $2\alpha + \beta + \gamma$
  - G2: $2\alpha + \beta + \gamma + \delta$

If this is accepted, test $H_1 : \beta = 0$ (*no treatment effect*):

- Unpaired T-test (G1 vs G2) with the difference between the two effects (P1-P2) as outcome, since the group means are:
  - G1 (P+A): $(\alpha + \beta + \gamma) - \alpha = \beta + \gamma$
  - G2 (A+P): $(\alpha + \gamma) - (\alpha + \beta) = \gamma - \beta$
- And report the estimated treatment effect.

But what if there is a carry-over effect?

Conclusion on treatment effect

Depends on your protocol!

Method	Effect	Confidence Interval	P-value
Period 1	-8.20	(-53.99,37.59)	0.71
Period 2	-70.47	(-129.40,-11.55)	0.022
Joint*	-39.33	(-68.70,-9.97)	0.012

*assuming no carry-over effect

Outline

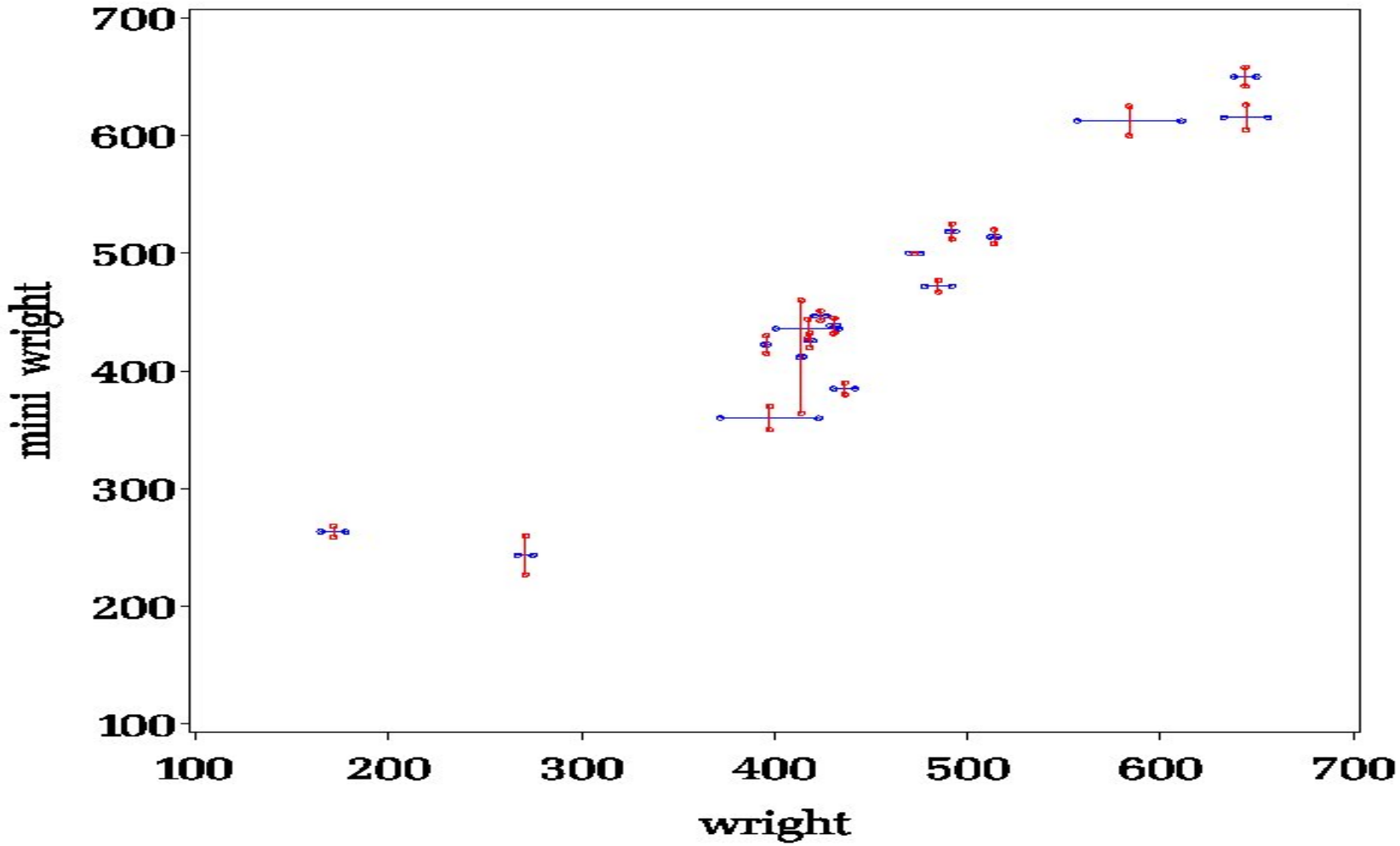
Comparing measurement devices

Example: Peak expiratory flow rate, l/min:

- 17 subjects, 2 measurement devices, two replicates with **each method**.

subject	*Wright*		*mini Wright*	
id	Y_{1p1}	Y_{1p2}	Y_{2p1}	Y_{2p2}
1	494	490	512	525
2	395	397	430	415
3	516	512	520	508
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
15	178	165	259	268
16	423	372	350	370
17	427	421	451	443
Average	450.35	445.41	452.47	455.35
SD	116.31	119.61	113.12	111.32

Reference: Bland and Altman, Lancet (1986).


700
600
500
400
300
200
100
mini wright
100
200
300
400
500
600
700
wright

Aim of investigation

Quantify the **precision** of each measuring device

- Variability / reproducibility.

Quantify the **agreement** between the two devices

- Bias of one method compared to the other.
- Variance of one method compared to the other.

Can the devices be used interchangably in clinic?

Simple approaches

For reliability

- Compare the replicate measuements in Bland-Altman plots* with limits of agreement, i.e.
  - Plot of difference in measurements vs average of measurements.
  - 95% normal range for typical differences.
- for **each method separately**.

For method comparison

- Compare **averages** in a Bland-Altman plot?
- Not good - unless you also do averages in clinic!

$\star$ See: Bland & Altman, Lancet (1986).

Variance component model?

level	variation	covariates
3	between subjects (σ^2)	
2	between methods (τ^2)	method
1	within methods (ω^2)	

Specified as:

$$Y_{ijk} = \mu_j + A_i + B_{ij} + \varepsilon_{ijk}$$

- $A_i \sim \mathcal{N}(0, \sigma^2)$ for subjects $i = 1, \ldots, 17$,
- $B_{ij} \sim \mathcal{N}(0, \tau^2)$ for methods $j = 1, 2$,
- $\varepsilon_{ijk} \sim \mathcal{N}(0, \omega^2)$ for replicate $k = 1, 2$.

Implied covariance structure

- We have 4 measurements on each subject

Covariance matrix with ordering (wright1, wright2, mini1, mini2):

$$\begin{pmatrix} \sigma^2+\tau^2+\omega^2 & \sigma^2+\tau^2 & \sigma^2 & \sigma^2 \\ \sigma^2+\tau^2 & \sigma^2+\tau^2+\omega^2 & \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2 & \sigma^2+\tau^2+\omega^2 & \sigma^2+\tau^2 \\ \sigma^2 & \sigma^2 & \sigma^2+\tau^2 & \sigma^2+\tau^2+\omega^2 \end{pmatrix}$$

- We have stronger correlation between measurements made with the same method than with different methods.
- And same variance for both methods.

Analysis

```
PROC MIXED DATA=wright;
  CLASS method id;
  MODEL flow=method / SOLUTION CL;
  RANDOM intercept method / SUBJECT=id;
RUN;
```

Solution for Fixed Effects

Effect	method	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		447.88	27.7519	16	16.14	<.0001
method	mini	6.0294	8.0532	16	0.75	0.4649
method	wright	0	.	.	.	.

No evidence of **systematic** differences between the measurement methods.

Estimated variance components

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
Intercept	id	12542
method	id	393.57
Residual		315.37

Fit Statistics

-2 Res Log Likelihood	676.0
AIC (smaller is better)	681.6

What does this tell us about the precision of the measurements?

Typical differences

Between replicate measurements using the same method:

$$\begin{aligned} Y_{ijk_1} - Y_{ijk_2} &= \varepsilon_{ijk_1} - \varepsilon_{ijk_2} \\ &\sim \mathcal{N}(0, 2\omega^2) \end{aligned}$$

Limits-of-agreement: $\pm 2\sqrt{2\omega^2} \simeq \pm 50.23.$

Between measurements using the different methods:

$$\begin{aligned} Y_{ij_1k_1} - Y_{ij_2k_1} &= \mu_{j_1} - \mu_{j_2} + B_{ij_1} - B_{ij_2} + \varepsilon_{ij_1k_1} - \varepsilon_{ij_2k_1} \\ &\sim \mathcal{N}(\mu_{j_1} - \mu_{j_2}, 2\tau^2 + 2\omega^2) \end{aligned}$$

Limits-of-agreement: $\mu_1 - \mu_2 \pm 2\sqrt{2\tau^2 + 2\omega^2} \simeq 6.03 \pm 75.31.$

(where we include the non-significant systematic difference).

Comparing precisions

We need a more general model:

$$Y_{ijk} = \mu_j + A_{ij} + \varepsilon_{ijk}$$

- $A_i \sim \mathcal{N}(0, \Sigma)$ for subjects $i = 1, \ldots, 17$,
- $\varepsilon_{ijk} \sim \mathcal{N}(0, \omega_j^2)$ for replicate $k = 1, 2$.

- bivariate random effect.
- method-dependent residual variance.

Analysis

```
PROC MIXED DATA=wright;
CLASS method id;
MODEL flow=method / SOLUTION CL;
RANDOM method / TYPE=UN SUBJECT=id G;
REPEATED / TYPE=simple GROUP=method SUBJECT=id*method;
RUN;
```

Covariance Parameter Estimates

Cov Parm	Subject	Group	Estimate
UN(1,1)	id		12188
UN(2,1)	id		12542
UN(2,2)	id		13683
Residual	method*id	method mini	396.44
Residual	method*id	method wright	234.29

Fit Statistics

-2 Res Log Likelihood	673.8
AIC (smaller is better)	683.8

Comparing precisions

Reproducibility (typical differences):

$$\text{Wright: } \hat{\omega}_1^2 = 234.29 \ \rightarrow \ \pm 2\sqrt{2\omega_1^2} \simeq \pm 43.3$$

$$\text{Mini: } \hat{\omega}_2^2 = 396.44 \ \rightarrow \ \pm 2\sqrt{2\omega_2^2} \simeq \pm 56.3$$

Seemingly Wright is more precise, but is the difference significant?

$$F = \frac{396.44}{234.29} = 1.69 \sim F(17, 17) \rightarrow P = 0.14$$

Don't form too firm a conclusion with too small data.

Overall comparison

Solution for Fixed Effects

Effect	method	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		447.88	28.4914	32	15.72	<.0001
method	mini	6.0294	8.0532	32	0.75	0.4595
method	wright	0	.	.	.	.

No evidence of **systematic** differences between the two methods.

Typical differnces between the two methods:

$$\begin{aligned} Y_{ij_1k_1} - Y_{ij_2k_1} &= \mu_{j_1} - \mu_{j_2} + A_{ij_1} - A_{ij_2} + \varepsilon_{ij_1k_1} - \varepsilon_{ij_2k_1} \\ &\sim \mathcal{N}(\mu_{j_1} - \mu_{j_2}, \sigma_1^2 + \sigma_2^2 - 2\sigma_{12} + \omega_1^2 + \omega_2^2) \end{aligned}$$

Limits-of-agreement: $6.0 \pm 75.3 = (-69.3, 81.3)$.

The end

I hope you have enjoyed the course!

Suggestions for improvements are warmly welcomed.