

# Applied Statistics with R: Mixed models

Cristina Sartori & Hugo Toledo-Alvarado

20-22 February 2018

cristina.sartori@unipd.it

<https://github.com/Hugo-Toledo/Applied-Statistics-R-UNIPD>



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

**DAFNAE**

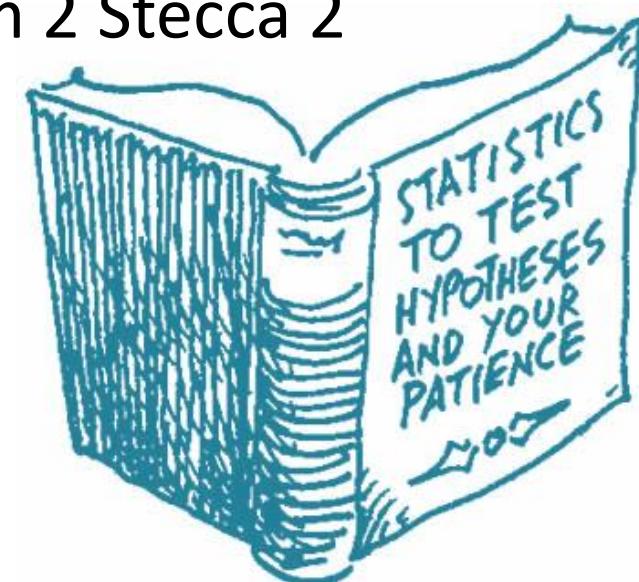
Department of Agronomy, Food, Natural  
resources, Animals and Environment

**Ph.D. ANIMAL  
& FOOD  
COURSE  
SCIENCE**  
UNIVERSITY OF PADOVA

PhD Course in Animal and Food Science

# General information about the course

- Part II of the optional course «Advanced Statistics with SAS» for PhD students
- *When/Where:*
  - February 22<sup>nd</sup>, 2-5 pm / Room 2 Stecca 1
  - February 22<sup>nd</sup>, 11am-1.30 pm / Room 1 Stecca 2
  - February 22<sup>nd</sup>, 2.30-6.30 pm / Room 2 Stecca 2
    - ...sorry for the strange timetable and rooms,
    - we have to deal with the “Scegli il tuo domani” event...
- **9 hours**



# General information about the course

## ■ *The program in brief*

1. Mixed models theory (including some basics of Matrix Algebra)
2. Mixed models analysis with R
3. Mixed models for experimental design
4. Mixed models and longitudinal data: designs with repeated measurements



# General information about the course

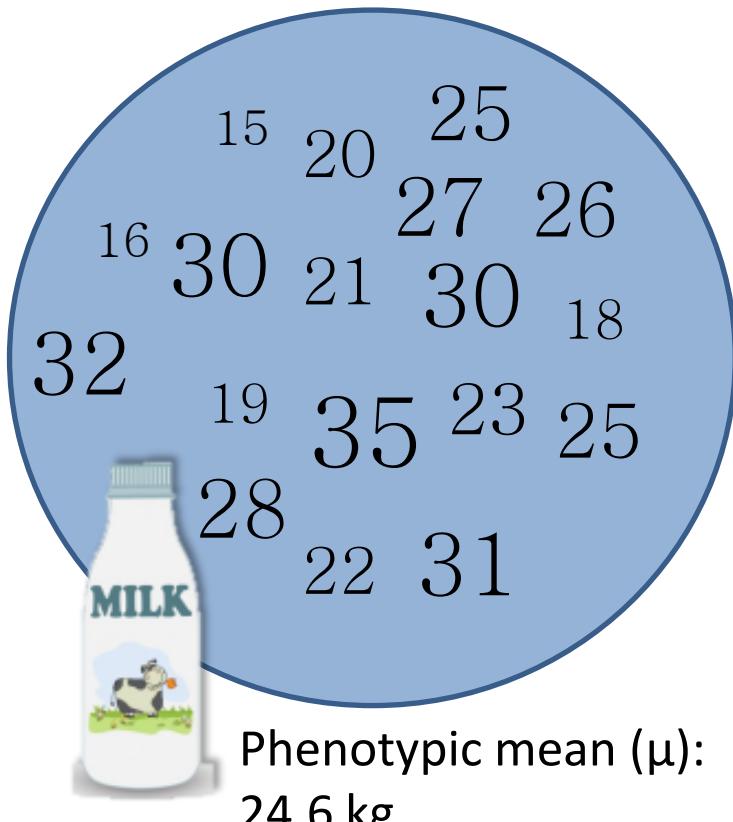
Some issues from the previous parts



# Fixed effects vs. random effects



- ✓ **Fixed effects:** interest in the estimation of the levels of the effect



- ✓ 1<sup>st</sup> parity: -5 kg
- ✓ Lsmean: 19 kg



- ✓ 2<sup>nd</sup> parity:  
+0.5 kg
- Lsmean: 21 kg



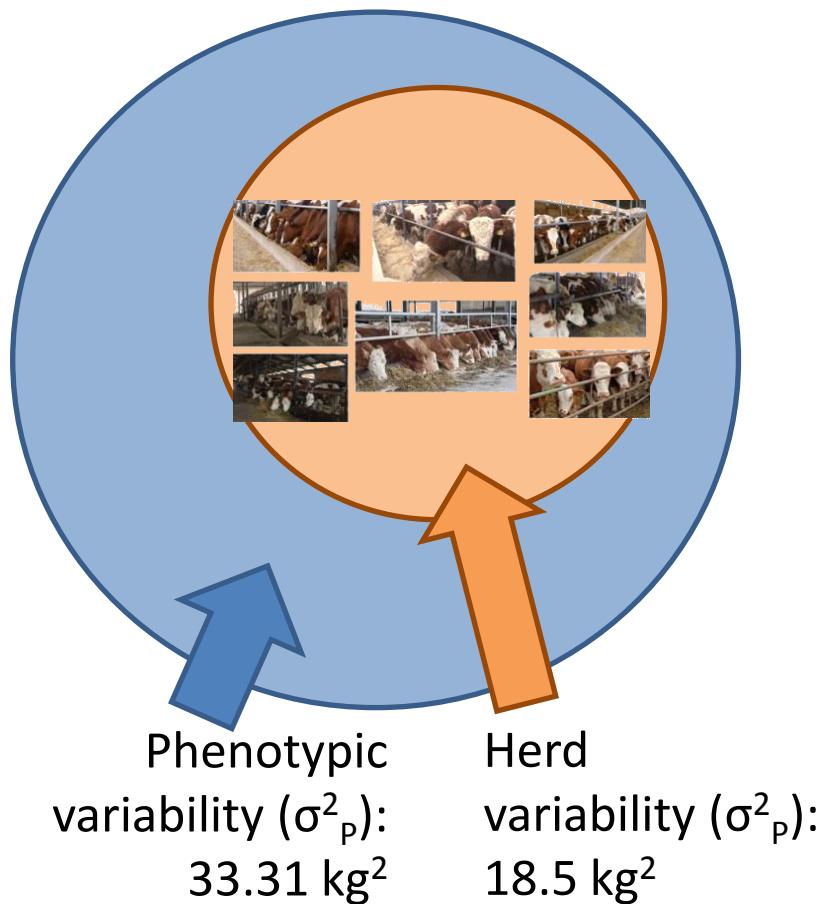
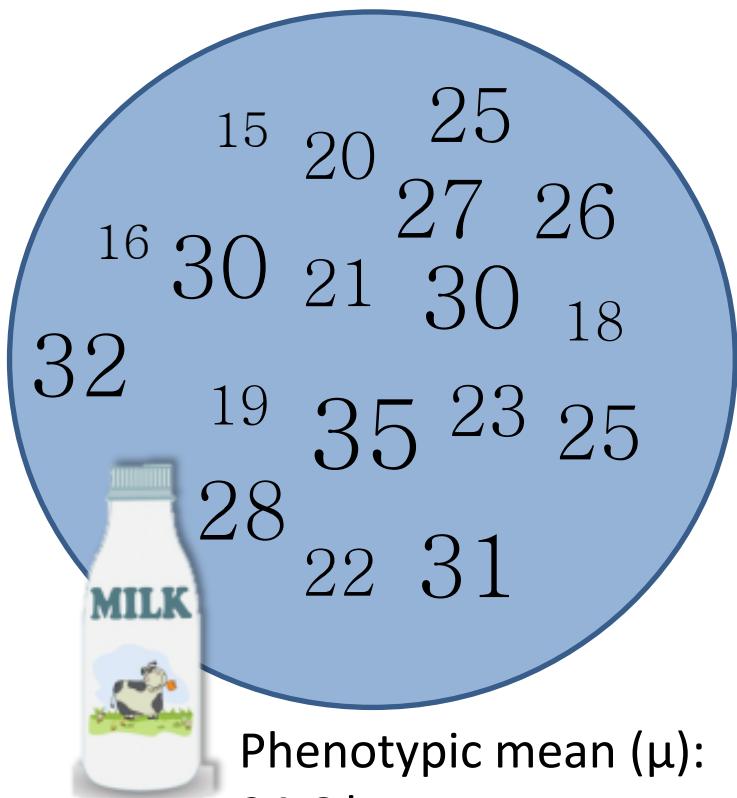
- ✓ 3<sup>rd</sup> parity: +3 kg
- ✓ Ls mean: 28 kg

(Values by chance, not calculated)

# Fixed effects vs. random effects



- ✓ **Random effects:** interest in the estimation of the variability of the effect



# Fixed effects vs. random effects



- In case of **continuous variables**, results are similar



Var.  
Herd



Var.  
Herd



10 cows

Var.  
Herd

Same n° records, same variance:  
**Fixed effect ~ Random effect**

- If an effect is **random** and when **between-levels variance is large**, within levels variance become less important, and large and small levels tend to be weighted equally



Var.  
Herd



Var.  
Herd



100 cows

Var.  
Herd

Different n° records,  
(almost) same variance:  
**Better Random effect**

- If there is **no heterogeneity**, fixed and random effects → similar results



Var.  
Herd



10 cows

Var.  
Herd



10 cows

Var.  
Herd

Same n° records,  
different variance:  
**Better Random effect**

# Mixed models

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

**y**  
Vector of observations

Incidence matrix  
for fixed effects

**X**

Vector of parameters  
associated with the fixed  
factors (e.g.,  $\mu, \alpha_1, \alpha_2, \alpha_3, \beta$ )

**b**

**Z**

Incidence matrix for  
random effects  
(+ random matrices if  
+1 random effects)

**e**

Vector of residual errors

**u**

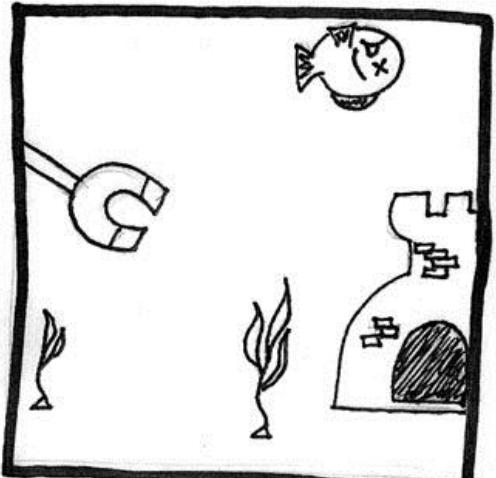
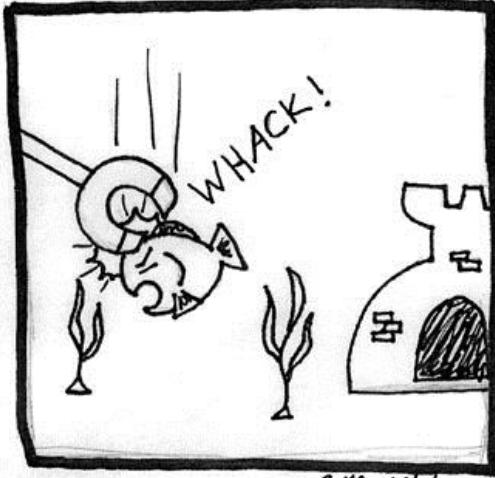
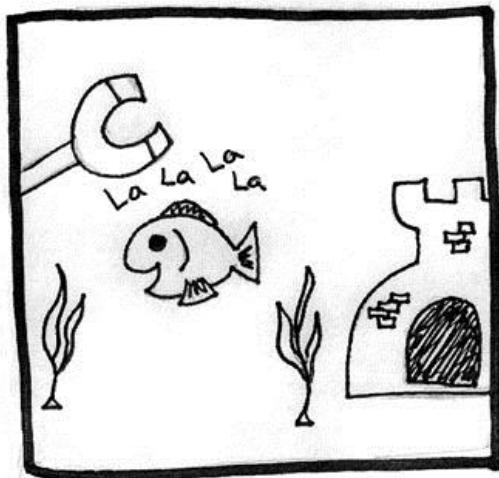
Vector of parameters  
associated with the random  
factors (e.g.,  $u_1, u_2$ )

- ✓ Observe  $y, X, Z$
- ✓ Estimate fixed effects (**b**)
- ✓ Estimate random effects (**u**)



### 3. Mixed models & experimental designs

#### The Importance of Experimental Design



Let's see if the subject  
responds to magnetic  
stimuli... ADMINISTER  
THE MAGNET!

Interesting...there seems  
to be a significant  
decrease in heart rate.  
The fish must sense the  
magnetic field.

# Mixed models & experimental designs



- ✓ Linear models (GLMs) and Mixed Models can be alternatively used to analyse data from different experimental designs, but in some cases one test is more proper than another...

1. Completely Randomized Design (CRD)
2. Randomized Block Design
3. Latin Square Design Example\*
4. Factorial Design
5. Nested design Example\* \*\*
6. Split-Plot design (analysis of longitudinal data)

\* Apart Nested design & Split-Plot design, experimental designs have been already extensively treated in the previous course, therefore we will consider just some examples

\*\* The analysis of longitudinal data is the last part of this course

# Randomized block design

*From the previous course*

- ✓ **BLOCK:** Experimental units are randomly distributed within similar/homogeneous groups (BLOCKS): e.g., land parcels, cages of animals...
- ✓ **TREATMENT:** All the treatments of the study are assigned to the experimental units (i.e., each treatment to some units) of each block

## BLOCK: Random effect:

E.g.: blocks of similar plot



## TREATMENT: Fixed effect:

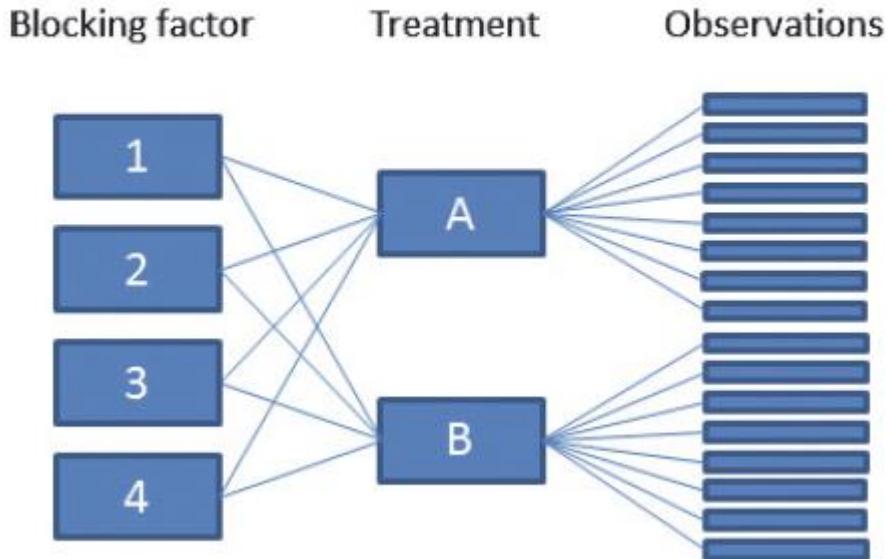
E.g.: fertilizer (2 levels)



Both treated in  
the previous  
course as fixed  
effects

# Randomized block design

## Randomized block design



- ✓ In a randomized block design, the blocking factor is typically conceptually random, even though it might be fitted as a fixed effect
- ✓ The block x treatment interaction variance can be estimated if there are replicate observations for each block-treatment combination

Methods in Ecology and Evolution 2013, 4, 14–24

doi: 10.1111/j.2041-210x.2012.00251.x



Nested by design: model fitting and interpretation in a mixed model era

Holger Schielzeth<sup>1\*</sup> and Shinichi Nakagawa<sup>2</sup>

# Example of Randomized block design

Block	Fertilizer 1	Fertilizer 2	Fertilizer 3
1	7	10	11
2	12	12	15
3	11	11	13
4	8	9	10
5	9	10	12
6	10	12	13

Average yield (t/ha) within block



Differences in yield between fertilizers?



# Example of Randomized block design

## Dataset to be imported

(`"../Dropbox/UNIPD/Biostatistics Curse R Spring 2018/curso STAT phD 2018 Mixed Models/data/yield_rbd.txt"`)

block	fertilizer	yield
1	1	7
1	2	10
1	3	11
2	1	12
2	2	12
2	3	15
3	1	11
3	2	11
3	3	13
4	1	8
4	2	9
4	3	10
5	1	9
5	2	10
5	3	12
6	1	10
6	2	12
6	3	13



# Example of Randomized block design



## R script (RBD linear model - LM vs. mixed model - MM)

```
#rm(list = ls()) # Clean the workspace
#install.packages("lme4") # If necesary install the library
library(lme4) # Call the library to the workspace
library(car) # Call the library to the workspace
library(lsmeans) # Call the library to the workspace
library(multcomp) # Call the library to the workspace

# Read the data
rbd<-read.table(file="C:/Users/toledo/Dropbox/UNIPD/Biostatistics Curse R Spring 2018/curso STAT phD 20
stringsAsFactors = TRUE,header = TRUE,sep = "\t")

rbd$block<-as.factor(rbd$block) # Set block as factor
rbd$fertilizer<-as.factor(rbd$fertilizer) # Set fertilizer as factor
contrasts(rbd$block)<-contr.SAS # Change the reference grid to SAS
contrasts(rbd$fertilizer)<-contr.SAS # Change the reference grid to SAS

# Fit the linear model
model.1<-lm(yield ~ fertilizer + block, data=rbd) # Fit the model
summary(model.1) # See the results
```

block & fertilizer set  
as categorical effects  
(since they are numbers,  
otherwise they are read as  
covariates)

Contrasts written as  
SAS (last level set to zero)

First analysis: **linear model** (no random effects)

Second analysis: **mixed model** → block included as random effect

# RBD: linear model, block as fixed effect



# Fit the linear model

```
model.1<-lm(yield ~ fertilizer + block, data=rbd) # Fit the model  
summary(model.1) # See the results
```

```
##
```

```
## Call:
```

```
## lm(formula = yield ~ fertilizer + block, data = rbd)
```

```
##
```

```
## Residuals:
```

```
##      Min      1Q Median      3Q      Max  
## -1.00000 -0.29167  0.08333  0.33333  0.83333
```

```
##
```

```
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )
## (Intercept)	1.317e+01	4.389e-01	30.002	3.96e-11 ***
## fertilizer1	-2.833e+00	3.801e-01	-7.455	2.17e-05 ***
## fertilizer2	-1.667e+00	3.801e-01	-4.385	0.001366 **
## block1	-2.333e+00	5.375e-01	-4.341	0.001464 **
## block2	1.333e+00	5.375e-01	2.481	0.032504 *
## block3	1.480e-16	5.375e-01	0.000	1.000000
## block4	-2.667e+00	5.375e-01	-4.961	0.000569 ***
## block5	-1.333e+00	5.375e-01	-2.481	0.032504 *
## ---				
## Signif. codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'
##	1			

```
##
```

```
## Residual standard error: 0.6583 on 10 degrees of freedom
```

```
## Multiple R-squared: 0.9328, Adjusted R-squared: 0.8858
```

```
## F-statistic: 19.84 on 7 and 10 DF, p-value: 4.171e-05
```

Linear model analysis (LM):  
block as fixed effect

Block6 set to zero

Standard error of residuals

Coefficients of determination

# RBD: linear model, block as fixed effect



```
Anova(model.1, type=3, test.statistic = "F") # ANOVA table SS type III
```

```
## Anova Table (Type III tests)
##
## Response: yield
##           Sum Sq Df F value    Pr(>F)
## (Intercept) 390.06  1 900.144 3.959e-11 ***
## fertilizer   24.33  2  28.077 7.893e-05 ***
## block       35.83  5  16.538 0.0001489 ***
## Residuals    4.33 10
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



Analysis of Variance: both fertilizer & block effects are significant

```
lsmeans(model.1, "fertilizer") # LSM for factor
```

```
## fertilizer  lsmean      SE df lower.CL upper.CL
## 1          9.50000 0.2687419 10  8.901206 10.09879
## 2         10.66667 0.2687419 10 10.067872 11.26546
## 3         12.33333 0.2687419 10 11.734539 12.93213
##
## Results are averaged over the levels of: block
## Confidence level used: 0.95
```



Least square means for fertilizer effect

# RBD: mixed model, block as random effect



# Fit the mixed model

```
model.2<-lmer(yield ~ fertilizer + (1 | block) ,data = rbd, REML = TRUE)
summary(model.2) # Results of the mixed model
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: yield ~ fertilizer + (1 | block)
##   Data: rbd
##
## REML criterion at convergence: 49.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max 
## -1.65689 -0.39715  0.01943  0.61353  1.12814
##
## Random effects:
##   Groups   Name        Variance Std.Dev.
##   block    (Intercept) 2.2444   1.4981
##   Residual           0.4333   0.6583
##   Number of obs: 18, groups: block, 6
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept) 12.3333   0.6681 18.462
## fertilizer1 -2.8333   0.3801 -7.455
## fertilizer2 -1.6667   0.3801 -4.385
## ...
```

Mixed model analysis (MM):  
block as random effect

Variances for random effects  
(block,  $\sigma^2_u$ ; residuals,  $\sigma^2_e$ )

```
Anova(model.1, type=3,test.statistic = "F") #
## Anova Table (Type III tests)
##
## Response: yield
##              Sum Sq Df F value    Pr(>F)
## (Intercept) 390.06  1 900.144 3.959e-11 ***
## fertilizer   24.33  2  28.077 7.893e-05 ***
## block       35.83  5 16.538 0.0001489 ***
## Residuals   4.33 10
```

Same residual variance using LM or MM

Solutions (BLUE) for fixed effects;  
fertilizer3 was set to zero

# RBD: mixed model, block as random effect



```
AIC(model.2)          # Akaike's Information Criterion (small is better)  
  
## [1] 59.42816  
  
BIC(model.2)         # Bayesian Information Criterion (small is better)  
  
## [1] 63.88002  
  
Anova(model.2, type=3,test.statistic = "F") # ANOVA table SS type III
```

```
## Analysis of Deviance Table (Type III Wald F tests with Kenward-Roger df)  
##  
## Response: yield  
##           F Df Df.res   Pr(>F)  
## (Intercept) 340.830 1 6.2368 1.114e-06 ***  
## fertilizer    28.077 2 10.0000 7.893e-05 ***  
## ---  
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ANOVA for fixed effects (fertilizer)

```
Anova(model.1, type=3,test.statistic = "F") # ANOVA table SS type III  
  
## Anova Table (Type III tests)  
##  
## Response: yield  
##           Sum Sq Df F value    Pr(>F)  
## (Intercept) 390.06  1 900.144 3.959e-11 ***  
## fertilizer    24.33  2  28.077 7.893e-05 ***  
## block        35.83  5 16.538 0.0001489 ***  
## Residuals     4.33 10  
## ---  
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Same ANOVA statistics (F value, P) for the common fixed effects included in the models (fertilizer)

# RBD: mixed model, block as random effect



```
lsmeans(model.2, "fertilizer")      # LSM
```

MM

```
##   fertilizer    lsmean       SE   df lower.CL upper.CL
## 1            9.50000 0.6680541 6.24  7.880260 11.11974
## 2          10.66667 0.6680541 6.24  9.046926 12.28641
## 3          12.33333 0.6680541 6.24 10.713593 13.95307
##
## Degrees-of-freedom method: satterthwaite
## Confidence level used: 0.95
```

Least square means for fertilizer effect

```
lsmeans(model.1, "fertilizer")      # LSM for factor
```

LM

```
##   fertilizer    lsmean       SE df lower.CL upper.CL
## 1            9.50000 0.2687419 10  8.901206 10.09879
## 2          10.66667 0.2687419 10 10.067872 11.26546
## 3          12.33333 0.2687419 10 11.734539 12.93213
##
## Results are averaged over the levels of: block
## Confidence level used: 0.95
```

Same least square means in LM, different standard errors (SE)

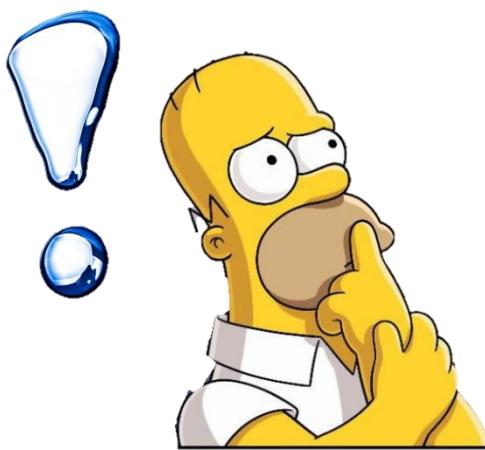
# Focus on: LSMeans standard errors

***Why errors of LSmeans are different in linear & mixed models?***

Due to the matrix constructed to compute them: the approximate standard errors for the LS-mean is computed as the square root of  $(\mathbf{L}\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{L}^T$  ( $\mathbf{L}$  is the vector including the indications about the level to consider for target LSmean).

**V** matrix is different in linear & mixed models:

**V=R** in linear models (only residual variance); **V=G+R** in mixed models (both residual and random effect variance)



# Mixed model of an experimental design:

$$Y = \text{treatment structure} + \text{design structure} + \text{residual}$$

Factorial or **nested structure**;  
generally **fixed effects**

**Restrictions on randomization**;  
experimental unit (plot & split-plot)

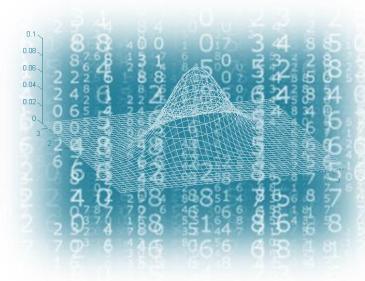


**Random effects** are used when **levels** of effect are considered as **samples** of a population (not all the levels possible of the factor), and when there is an interest in **taking into account** (correcting for) **natural variation** regarding experimental material and conditions (blocking)

# Linear model vs. Mixed Model



*In some experimental design it is incorrect do not include a random effect; e.g., unbalanced designs (different num. obs. in levels of an effect)*



- ✓ Standard errors for estimates and LS-means based on the **fixed-effects** model might be **smaller** than those based on a random-effects model →
- ✓ Some functions that are estimable under a true random-effects model **might not even be estimable under the fixed-effects model**
- ✓ For balanced data designs, the test statistics have similar distribution only when the design is balanced; for unbalanced designs, the values for the **F tests differ**

# Fixed effects vs. random effects



- In case of **continuous variables**, results are similar



10 cows

Var.  
Herd



10 cows

Var.  
Herd



10 cows

Var.  
Herd

Same n° records, same variance:  
**Fixed effect ~ Random effect**

- If an effect is **random** and when **between-levels variance is large**, within levels variance become less important, and large and small levels tend to be weighted equally



2 cows

Var.  
Herd



10 cows

Var.  
Herd



100 cows

Var.  
Herd

Different n° records,  
(almost) same variance:  
**Better Random effect**

- If there is **no heterogeneity**, fixed and random effects → similar results



10 cows

Var.  
Herd



10 cows

Var.  
Herd



10 cows

Var.  
Herd

Same n° records,  
different variance:  
**Better Random effect**

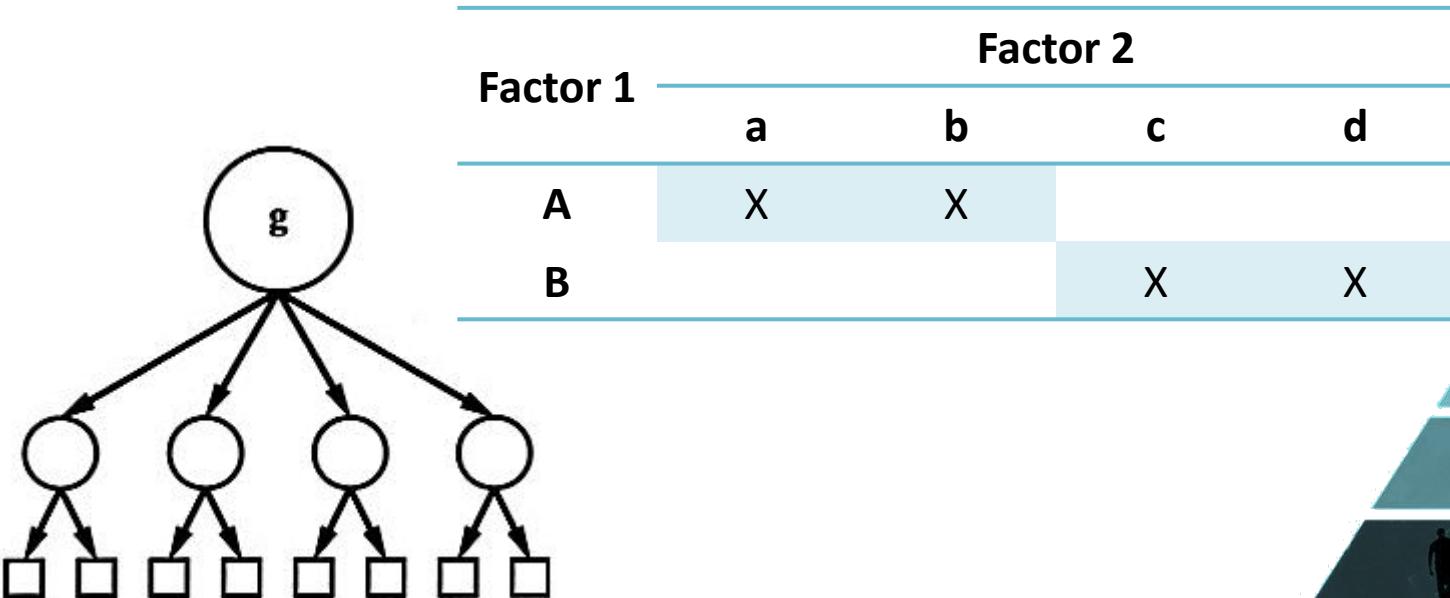
# Nested designs (hierarchical models)



**Hierarchical models** (also called multi-level and, in some cases mixed-effects)

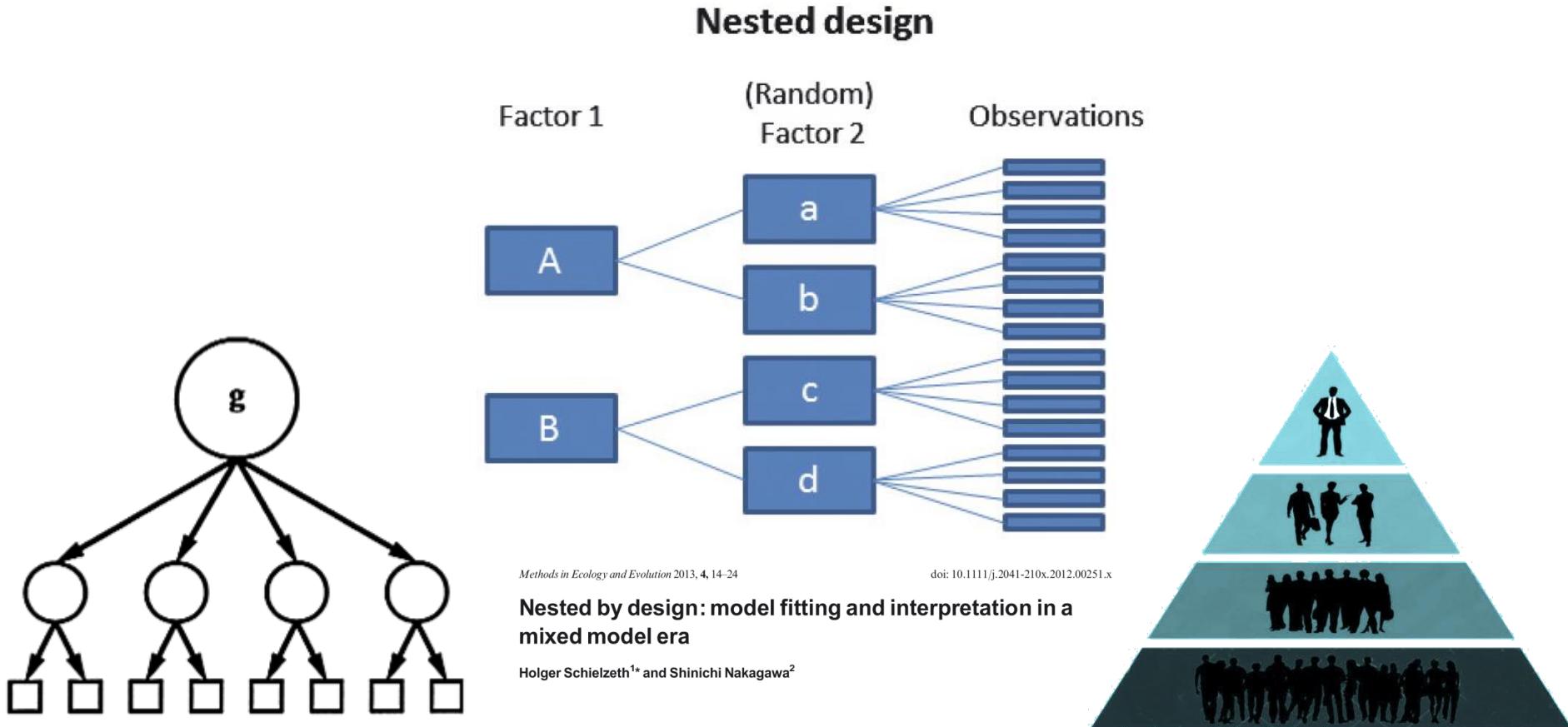
Designed to handle the mutual dependence among data, that can be clustered in groups.

- ✓ Two factors: Factor 1 is hierarchically upper than Factor 2, since **each level of Factor 2 is included just within 1 level of Factor 1**
- ✓ Factor 2 is considered hierarchically nested / subordinated into Factor 1



# Nested designs (hierarchical models)

- In a nested design, the nested factor is typically conceptually random, even though it might be fitted as a fixed effect (Factor 1 is a group-level predictor relative to Factor 2).

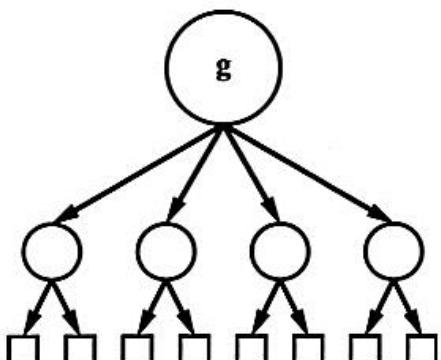


# Example of Hierarchical Model

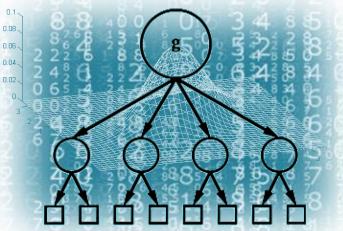
Es: Dataset of 124 cows → 74 Brown Swiss (BS) & 50 Holstein Friesian, divided in 6 single-breed herds



- ✓ The herd is **NESTED** within the breed
- ✓ The **herds** considered are **just sample** of bigger populations



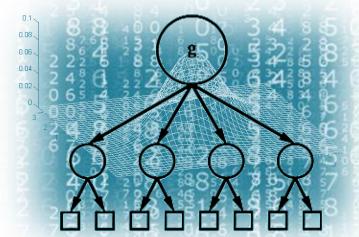
# Example of Hierarchical Model



Example: Dataset of 124 cows → 74 Brown Swiss (BS) & 50 Holstein Friesian, divided in 6 single-breed herds (C:/Users/toledo/Dropbox/UNIPD/Biostatistics Curse R Spring 2018/curso STAT PhD 2018 ANOVA/data/cows.txt)

MATR	breed	herd	parity	milk	dim	CLDIM	fat	protein	cellu	casein	r	sh	ph
IT014TN003B016	BrownSwiss	1805639	7	15.6	266	6	4.08	3.84	1006	2.78	17.4	3.1	6.76
IT014TN003C020	BrownSwiss	1805639	6	26.4	151	4	3.57	3.58	510	2.72	15	2.58	6.8
IT022000032815	BrownSwiss	1805639	4	36.4	51	2	4.84	3.21	186	2.75	10.2	3.26	6.67
IT022000039390	BrownSwiss	1805639	3	32.9	185	5	4.4	3.53	714	2.71	16.2	3.05	6.78
IT022000039394	BrownSwiss	1805639	4	25.2	176	4	4.77	3.99	374	2.98	20.4	3.39	6.72
IT022000039417	BrownSwiss	1805639	4	34.9	35	1	3.42	3.21	44	2.69	12	3.16	6.74
IT022000089451	BrownSwiss	1805639	3	26.2	207	5	3.45	3.74	158	2.88	13.2	3.45	6.66
IT022000089456	BrownSwiss	1805639	4	30	31	1	3.16	3.25	1119	2.43	11	2.72	6.74
IT022000089457	BrownSwiss	1805639	3	25.3	181	5	3.81	3.61	62	2.8	10.4	3.39	6.65
IT022000089510	BrownSwiss	1805639	3	28.5	142	4	6.53	3.4	371	2.75	12.2	2.91	6.73
IT022000093948	BrownSwiss	1805639	2	22.1	240	6	4.71	4.09	134	3.07	15	3.3	6.72
IT022000093951	BrownSwiss	1805639	2	27.2	169	4	4.2	3.69	644	2.84	24	2.94	6.8
IT022000093975	BrownSwiss	1805639	2	21.1	272	7	5.16	4.08	191	3.25	14	2.87	6.75
IT022000093988	BrownSwiss	1805639	2	33	50	2	4.62	3.25	97	2.74	17.2	2.86	6.78
IT022000094004	BrownSwiss	1805639	2	18.6	197	5	4.36	3.83	205	2.93	16.4	3.5	6.7
IT022000094009	BrownSwiss	1805639	2	24.8	184	5	3.53	3.64	34	2.89	19.2	3.46	6.71
IT022000099200	BrownSwiss	1805639	2	33.6	74	2	4.4	3.14	290	2.6	14	2.98	6.73
IT022000099205	BrownSwiss	1805639	1	23.5	131	3	3.26	4.04	108	3.08	18	3.14	6.76
IT022000099218	BrownSwiss	1805639	2	31.2	34	1	3	3.38	1465	2.52	17.4	2.73	6.81
IT022000099225	BrownSwiss	1805639	2	28.2	60	2	2.37	3.11	88	2.43	15.2	3.07	6.76
IT022000099231	BrownSwiss	1805639	1	18.7	346	7	4.57	4.05	843	3.11	22.4	3.04	6.81
IT022000099233	BrownSwiss	1805639	1	26.6	276	7	3.46	3.65	129	2.76	17	3.04	6.78
IT022000099238	BrownSwiss	1805639	1	19.7	73	2	4.16	3.35	115	2.74	18.4	3.13	6.77
IT022000099251	BrownSwiss	1805639	2	36.5	15	1	3.86	3.25	18	2.41	11	3.12	6.68
IT022000099265	BrownSwiss	1805639	2	22.2	182	5	4.29	3.77	154	2.9	14.4	3.48	6.67

# Example of Hierarchical Model



R script:

```
rm(list = ls()) # Clean the workspace
#install.packages("lme4") # If necessary install the library
library(lme4) # Call the library to the workspace
library(car) # Call the library to the workspace
library(lsmeans) # Call the library to the workspace
library(multcomp)

# Read the data
myfile<- "C:/Users/toledo/Dropbox/UNIPD/Biostatistics Curse R Spring 2018/curso STAT PhD 2018"

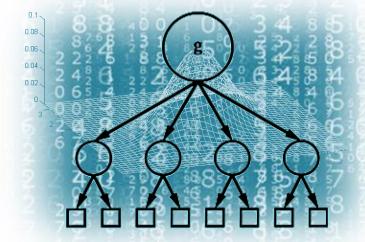
mydata<-read.table(file=myfile,stringsAsFactors = TRUE,header = TRUE,sep = "\t")

mydata$parity<-as.factor(mydata$parity) # Set as factor
mydata$herd<-as.factor(mydata$herd) # Set as factor
contrasts(mydata$parity)<-contr.SAS # Change the reference grid to SAS
contrasts(mydata$breed)<-contr.SAS # Change the reference grid to SAS
contrasts(mydata$herd)<-contr.SAS # Change the reference grid to SAS

# Fit the mixed model
model.1<-lmer(milk ~ parity + dim + breed + (1 |breed:herd), data = mydata, REML = TRUE)
summary(model.1) # Results of the mixed model
```

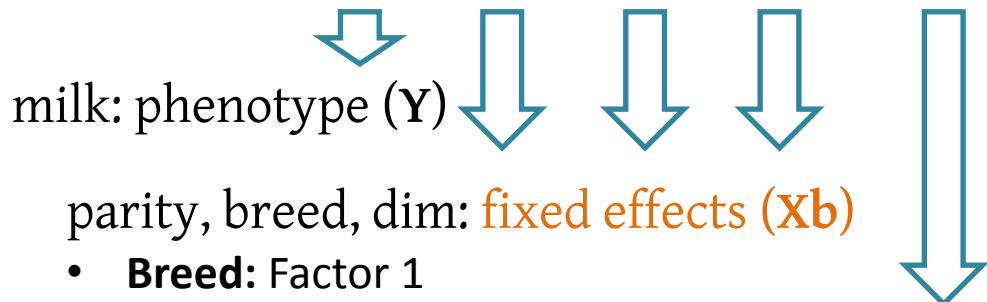
Mixed model including the **random effect (herd, Factor 2)** as nested within the hierarchically upper **fixed effect (breed, Factor 1)**

# Example of Hierarchical Model



```
# Fit the mixed model  
model.1<-lmer(milk ~ parity + dim + breed + (1 |breed:herd), data = mydata, REML = TRUE)  
summary(model.1) # Results of the mixed model
```

model.1<-lmer(milk ~ parity + dim + breed + (1 |breed:herd), data = mydata, REML = TRUE)



(1|breed:herd): **random effect (Zu)** of the herd:

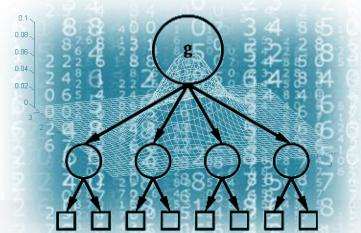
- **Herd:** Factor 2, nested within breed

The colon (:) indicates that the 2<sup>nd</sup> effect is nested within the 1<sup>st</sup>  
(in a different analysis, if you want to include also breed as random effect, you can write (1|breed/herd) instead of (1|breed) + (1|breed:herd) )

$Milk = \mu + \text{parity} + \text{dim} + \text{breed} + \text{herd:breed} + e \rightarrow$  Model in linear notation

$\mathbf{Y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e} \rightarrow$  Model in matrix notation

# Example of Hierarchical Model



```
# Fit the mixed model
```

```
model.1<-lmer(milk ~ parity + dim + breed + (1 |breed:herd), data = mydata, REML = TRUE)  
summary(model.1) # Results of the mixed model
```

Linear mixed model fit by REML t-tests use Satterthwaite approximations to degrees of freedom [  
lmerMod]  
Formula: milk ~ parity + dim + breed + (1 | breed:herd)  
Data: mydata

REML criterion at convergence: 731.2

Scaled residuals:

Min	1Q	Median	3Q	Max
-2.60214	-0.58489	0.03607	0.69114	2.02577

Random effects:

Groups	Name	Variance	Std.Dev.
breed:herd	(Intercept)	13.34	3.653
Residual		24.18	4.917

Number of obs: 124, groups: breed:herd, 6

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(> t )						
(Intercept)	39.339899	5.786365	80.400000	6.799	1.69e-09	***					
parity1	1.563274	5.101115	111.340000	0.306	0.760						
parity2	3.954127	5.098720	111.330000	0.776	0.440						
parity3	6.092434	5.166230	111.680000	1.179	0.241						
parity4	5.999030	5.412165	111.210000	1.108	0.270						
parity5	10.365640	7.137449	111.560000	1.452	0.149						
parity6	4.777401	6.986695	110.930000	0.684	0.496						
dim	-0.052370	0.005915	113.760000	-8.855	1.29e-14	***					
breed1	-6.890854	3.162900	3.880000	-2.179	0.097	.					
---											
Signif. codes:	0	'***'	0.001	'**'	0.01	'*'	0.05	'.'	0.1	' '	1

Value of -2log Likelihood when  
REML reached the convergence

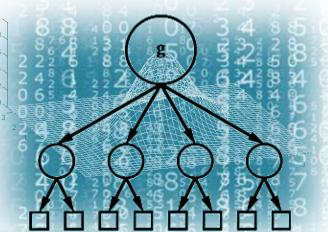
Variances for random effects  
(herd:breed,  $\sigma^2_u$ , residuals,  $\sigma^2_e$ )

6 herds, 2 breeds, 3 herds within breed

Solutions (BLUE) for fixed  
effects:

7 parities considered for this  
breed (parity7 set to zero);  
dim as covariate;  
2 breeds (breed2 set to zero)

# Example of Hierarchical Model



```
Anova(model.1, type=3,test.statistic = "F") # ANOVA table SS type III  
lsmeans(model.1,"parity") # LSM  
lsmeans(model.1,"breed") # LSM
```

```
> AIC(model.1) # Akaike's Information Criterion (small is better)  
[1] 753.1553  
> BIC(model.1) # Bayesian Information Criterion (small is better)  
[1] 784.1784  
> Anova(model.1, type=3,test.statistic = "F") # ANOVA table SS type III  
Analysis of Deviance Table (Type III Wald F tests with Kenward-Roger df)
```

Response: milk

	F	Df	Df.res	Pr(>F)
(Intercept)	46.0339	1	81.073	1.744e-09 ***
parity	2.2011	6	112.369	0.04797 *
dim	77.1992	1	113.791	1.855e-14 ***
breed	4.7359	1	3.987	0.09537 .

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

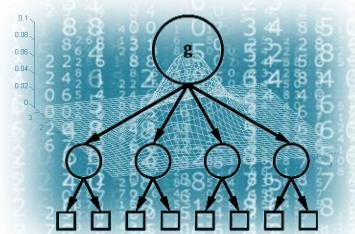
>

- ✓ In this analysis, the breed effect, hierarchically superior than the herd, is not statistically significant (just a nominal significance, P<0.1)
- ✓ Maybe the analysis of other traits than milk (e.g., fat, protein, casein...) could provide different results

Model fitting statistics

ANOVA on fixed effects:  
parity, dim & breed

# Example of Hierarchical Model



```
lsmeans(model.1, "parity")      # LSM  
lsmeans(model.1, "breed")       # LSM
```

```
> lsmeans(model.1, "parity")      # LSM  
parity   lsmean      SE    df lower.CL upper.CL  
1        29.40115  1.751183  5.69 25.05979 33.74251  
2        31.79200  1.681236  4.95 27.62404 35.95996  
3        33.93031  1.805198  6.56 29.45504 38.40558  
4        33.83690  2.596578 25.44 27.39972 40.27409  
5        38.20351  5.228135 103.82 25.24243 51.16460  
6        32.61527  5.219682 103.97 19.67514 45.55540  
7        27.83787  5.278241 104.48 14.75257 40.92317
```

Results are averaged over the levels of: breed

Degrees-of-freedom method: satterthwaite

Confidence level used: 0.95

```
> lsmeans(model.1, "breed")      # LSM  
breed   lsmean      SE    df lower.CL upper.CL  
BRUNA   29.07129  2.482912  5.89 22.96936 35.17322  
FRISONA 35.96214  2.620832  7.25 29.52127 42.40302
```

Results are averaged over the levels of: parity

Degrees-of-freedom method: satterthwaite

Confidence level used: 0.95

```
> |
```

Least square means for parity, as it were the only fixed effect in the model  
(Results are averaged over the levels of: breed)

Least square means for breed, as it were the only fixed effect in the model  
(Results are averaged over the levels of: parity)

# Hierarchical model with just fixed effects

```
# Model including just fixed effects  
model.2<-lm(milk ~ parity + dim + breed + herd%in%breed, data = mydata)  
summary(model.2) # Results of the mixed model
```

```
Call:  
lm(formula = milk ~ parity + dim + breed %in% breed, data = mydata)
```

Residuals:

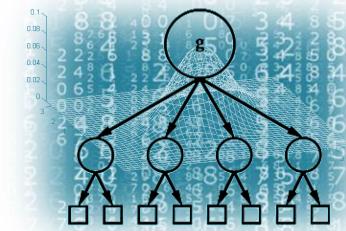
Min	1Q	Median	3Q	Max
-12.3093	-3.0542	0.1615	2.9618	10.1044

Coefficients: (6 not defined because of singularities)

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	44.240943	5.703326	7.757	4.55e-12	***
parity1	1.429784	5.104848	0.280	0.779935	
parity2	3.626959	5.102358	0.711	0.478675	
parity3	5.531636	5.174263	1.069	0.287359	
parity4	5.595854	5.414718	1.033	0.303639	
parity5	10.171414	7.146107	1.423	0.157439	
parity6	4.657332	6.985627	0.667	0.506346	
dim	-0.053415	0.005962	-8.959	8.86e-15	***
breed1	-12.649202	1.982793	-6.379	4.22e-09	***
breedBRUNA:herd1	-1.783482	1.393996	-1.279	0.203423	
breedFRISONA:herd1	NA	NA	NA	NA	
breedBRUNA:herd2	5.961697	1.674231	3.561	0.000546	***
breedFRISONA:herd2	NA	NA	NA	NA	
breedBRUNA:herd3	NA	NA	NA	NA	
breedFRISONA:herd3	NA	NA	NA	NA	
breedBRUNA:herd4	NA	NA	NA	NA	
breedFRISONA:herd4	-7.206510	1.962285	-3.673	0.000371	***
breedBRUNA:herd5	NA	NA	NA	NA	
breedFRISONA:herd5	-5.656240	2.307552	-2.451	0.015799	*

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.916 on 111 degrees of freedom  
Multiple R-squared: 0.6239, Adjusted R-squared: 0.5833  
F-statistic: 15.35 on 12 and 111 DF, p-value: < 2.2e-16



Why it is better to include the nested effect as random: Alternative model including herd as a fixed effect

Some solutions of herd effect (breed:herd, or herd%in%breed) are not defined because the herd effect is nested within breed  
↓  
there are singularities:  
It is better to treat herd as a random effect

# Hierarchical model with just fixed effects

```
Anova(model.2) # ANOVA table SS type III  
lsmeans(model.2,"parity")      # LSM  
lsmeans(model.2,"breed")       # LSM
```

```
> Anova(model.2) # ANOVA table SS type III
```

```
Anova Table (Type II tests)
```

```
Response: milk  
          Sum Sq Df F value    Pr(>F)  
parity     276.19  6 1.9049  0.08618 .  
dim        1939.31  1 80.2550 8.864e-15 ***  
breed      1113.45  1 46.0781 5.845e-10 ***  
breed:herd  863.46  4  8.9332 2.723e-06 ***  
Residuals  2682.25 111  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'  
> lsmeans(model.2,"parity")      # LSM  
   parity  lsmean        SE  df lower.CL upper.CL  
1 29.68148 0.9365801 111 27.82559 31.53738  
2 31.87866 0.7811267 111 30.33080 33.42651  
3 33.78334 1.0210600 111 31.76004 35.80663  
4 33.84755 2.1296472 111 29.62752 38.06759  
5 38.42311 5.0167446 111 28.48210 48.36413  
6 32.90903 5.0058113 111 22.98968 42.82838  
7 28.25170 5.0683260 111 18.20847 38.29493
```

Results are averaged over the levels of: herd, dim, breed

Confidence level used: 0.95

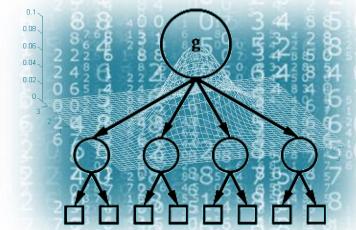
```
> lsmeans(model.2,"breed")      # LSM
```

NOTE: Results may be misleading due to involvement in interactions

```
breed  lsmean        SE  df lower.CL upper.CL  
BRUNA 29.19769 1.312593 111 26.59670 31.79868  
FRISONA 36.16657 1.563315 111 33.06875 39.26438
```

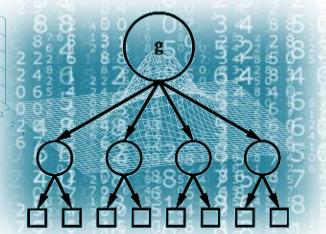
Results are averaged over the levels of: parity, herd, dim

Confidence level used: 0.95



All fixed effects apart from parity are highly significant, but the mixed model with herd as a random factor should be preferred due to the data structure

# Another example of Hierarchical Model



This example uses the same data of the Nested model example considered in the previous course with Prof. Roberto Mantovani, but analyses are performed using a Mixed Model instead of a Linear Model

GitHub, Inc. [US] | <https://github.com/Hugo-Toledo/Applied-Statistics-R-UNIPD>

## R scripts

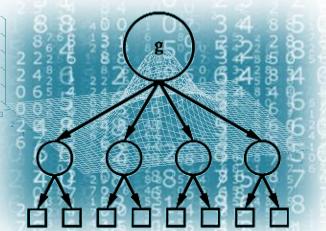
If you want to use the scripts, please select, copy and paste the script in R. Otherwise you can download the PDF with the solutions.

1. [Descriptive Statistics and Normality](#)
  - [PDF with solutions](#)
2. [One-Way ANOVA](#)
  - [PDF with solutions](#)
3. [Complete Randomized Block Design](#)
  - [PDF with solutions](#)
4. [Two-Way ANOVA](#)
  - [PDF with solutions](#)
5. [Contrasts](#)
  - [PDF with solutions](#)
6. [Completely Randomized Design with a Covariate](#)
  - [PDF with solutions](#)

7. [Random Block Design](#)
8. [Latin Square](#)
9. [Nested](#)
10. [Split Plot](#)

Here there is the script for the nested design, linear model analysis (not mixed model)

# Another example of Hierarchical Model



This example uses the same data of the Nested model example considered in the previous course with Prof. Roberto Mantovani, but analyses are performed using a Mixed Model instead of a Linear Model

Boar	Sow									
	A	B	C	D	E	F	G	H	I	L
1	X, X	X, X								
2			X, X	X, X						
3					X, X	X, X				
4							X, X	X, X		
5									X, X	X, X

Each level of «Boar» effect (hierarchically upper effect) there are different and unique levels of «sow» (nested effect).

In each combination 2 repeats (X) are recorded (2 piglets, for example).

# Another example of Hierarchical Model

This example uses the same data of the Nested model example considered in the previous course with Prof. Roberto Mantovani, but analyses are performed using a Mixed Model instead of a Linear Model

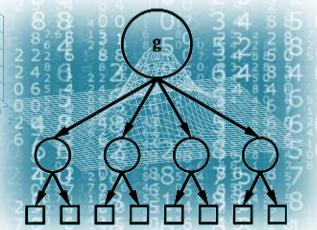
R script:

```
#rm(list = ls()) # Clean the workspace
library(lsmeans) # Call the library
library(lme4) # Call the library
library(car) # Call the library
# Read Data
mydata<-read.table(file = "C:/Users/toledo/Dropbox/UNIPD/Biostatistics Curse R Spring 2018/curso STAT Pl
                         sep = "\t",header = TRUE,stringsAsFactors = TRUE)
mydata$Boar<-as.factor(mydata$Boar) # Set the variable as factor
str(mydata) # See the structure of my data

## 'data.frame': 20 obs. of 4 variables:
## $ Boar      : Factor w/ 5 levels "1","2","3","4",...: 1 1 1 1 2 2 2 2 3 3 ...
## $ Sow       : Factor w/ 10 levels "A","B","C","D",...: 1 1 2 2 3 3 4 4 5 5 ...
## $ Replicate: int 1 2 1 2 1 2 1 2 1 2 ...
## $ ADG       : num 2.77 2.38 2.58 2.94 2.28 2.22 3.01 2.61 2.36 2.71 ...
contrasts(mydata$Boar)<-contr.SAS # Set the contrast as SAS
contrasts(mydata$Sow)<-contr.SAS # Set the contrast as SAS
table(mydata$Sow,mydata$Boar) # Frequencies for factors

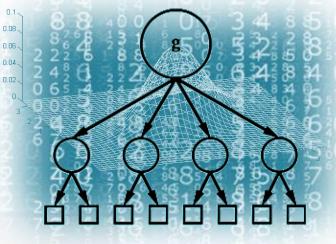
#### Random Model

mymodel.3<-lmer(ADG ~ Boar + (1|Sow), data = mydata ) # Fit a model with Random effect
summary(mymodel.3) # See the results
```



Let's go to run the random model!

# Another example of Hierarchical Model



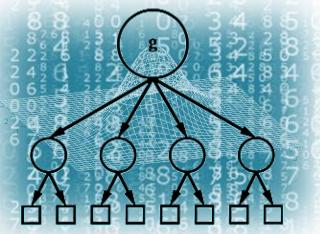
```
#### Random Model
```

```
mymodel.3<-lmer(ADG ~ Boar + (1|Sow), data = mydata ) # Fit a model with Random effect  
summary(mymodel.3) # See the results
```

```
## Linear mixed model fit by REML ['lmerMod']  
## Formula: ADG ~ Boar + (1 | Sow)  
##   Data: mydata  
##  
## REML criterion at convergence: 6.1  
##  
## Scaled residuals:  
##     Min      1Q  Median      3Q     Max  
## -1.1527 -0.5563 -0.1764  0.6277  1.5054  
##  
## Random effects:  
##   Groups   Name        Variance Std.Dev.  
##   Sow       (Intercept) 0.03701  0.1924  
##   Residual            0.03870  0.1967  
## Number of obs: 20, groups:  Sow, 10  
##  
## Fixed effects:  
##             Estimate Std. Error t value  
## (Intercept)  2.5700    0.1679 15.310  
## Boar1        0.0975    0.2374  0.411  
## Boar2       -0.0400    0.2374 -0.168  
## Boar3        0.0625    0.2374  0.263  
## Boar4       -0.1000    0.2374 -0.421  
## ...
```

You can write both  $(1|Sow)$  or  $(1|Boar:Breed)$ , the results are the same (since it is a random factor)

# Another example of Hierarchical Model



```
Anova(mymodel.3,type = 3,test.statistic = "F") # Anova table SS Type III
```

```
## Analysis of Deviance Table (Type III Wald F tests with Kenward-Roger df)
```

```
##
```

```
## Response: ADG
```

```
##          F Df Df.res   Pr(>F)
```

```
## (Intercept) 234.4033  1      5 2.156e-05 ***
```

```
## Boar        0.2212  4      5   0.9155
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
lsmeans(mymodel.3,"Boar")           # LSM for factor
```

```
##    Boar lsmean      SE df lower.CL upper.CL
```

```
## 1     2.6675 0.1678616 5 2.235998 3.099002
```

```
## 2     2.5300 0.1678616 5 2.098498 2.961502
```

```
## 3     2.6325 0.1678616 5 2.200998 3.064002
```

```
## 4     2.4700 0.1678616 5 2.038498 2.901502
```

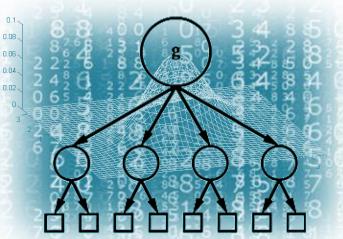
```
## 5     2.5700 0.1678616 5 2.138498 3.001502
```

```
##
```

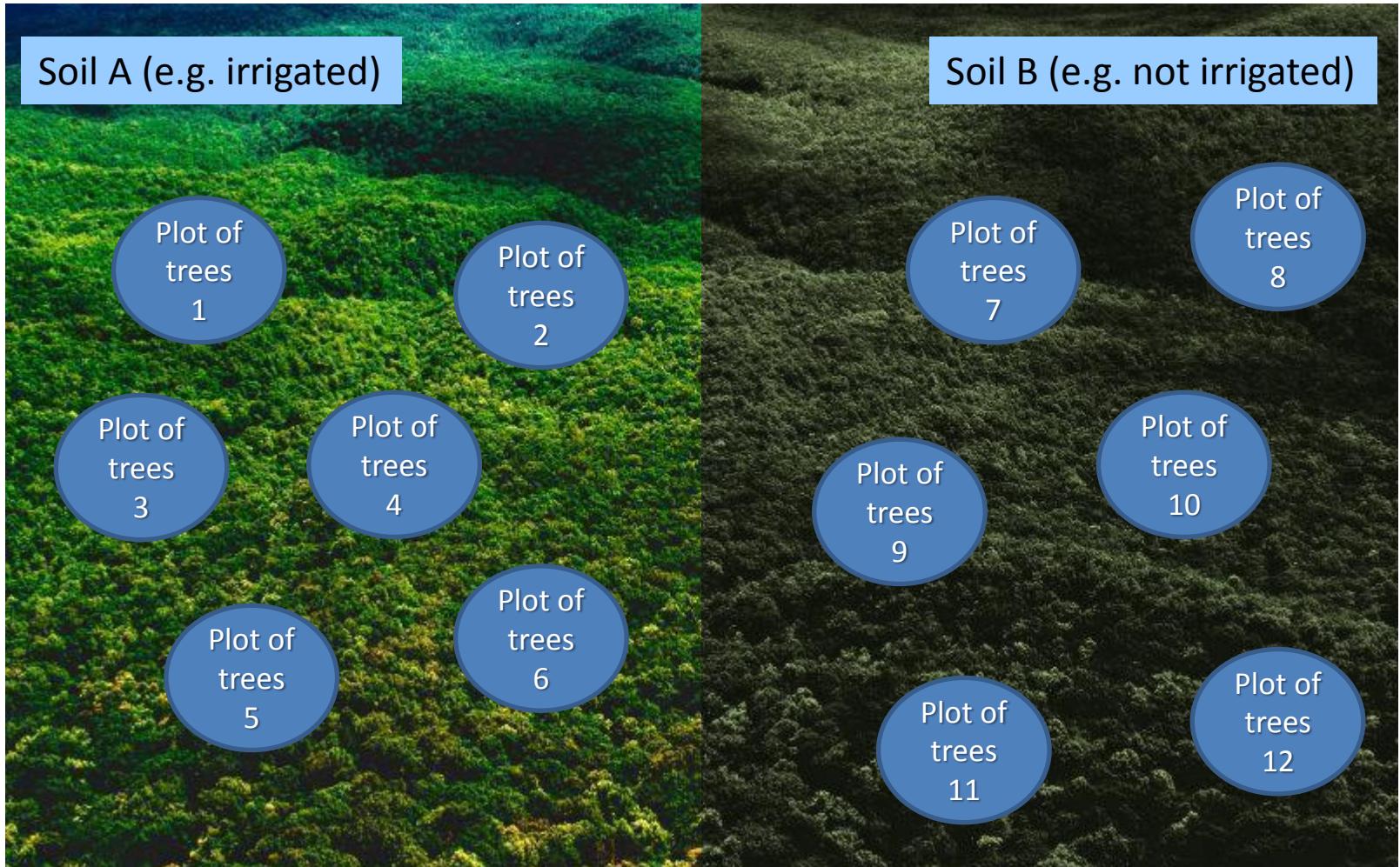
```
## Degrees-of-freedom method: satterthwaite
```

```
## Confidence level used: 0.95
```

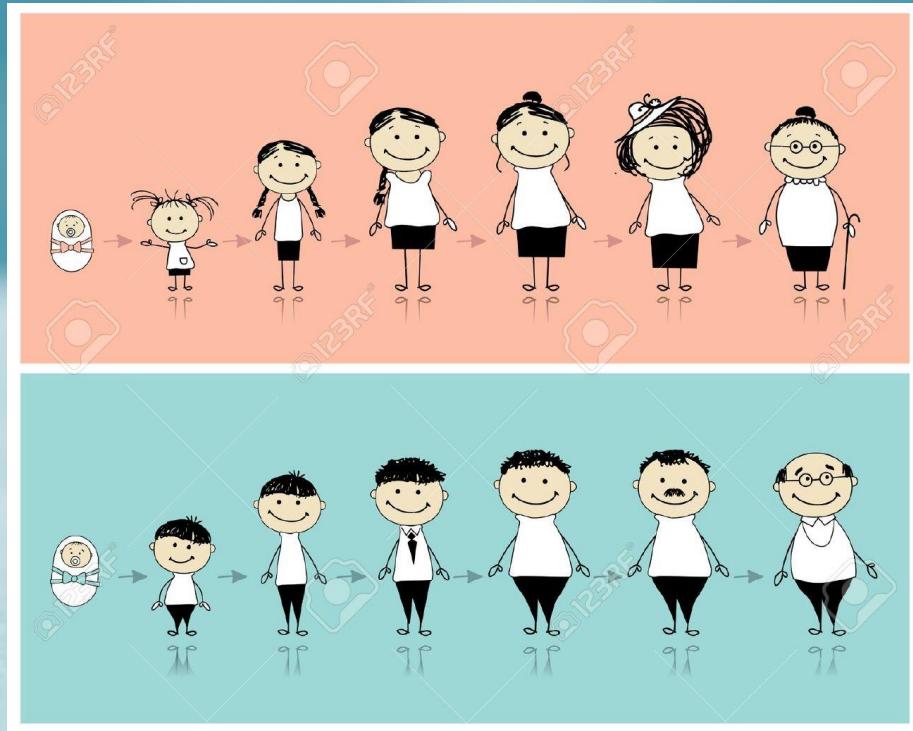
# Another example of Hierarchical Model



You can find examples of hierarchical models also in Crop Science & in Forestry Science (e.g.: soil as fixed effect, plot as random effect):



# 4. Analysis of longitudinal data



# Analysis of longitudinal data



- ✓ **Repeated measurements over time**, or on another specific dimension (drug dosage, space, etc.)
- ✓ Chance to **compare trends of response over time** (within or between treatments)
- ✓ **Dependency (covariance) between observations** performed on the **same units** (e.g. repeated measurements over time in target individuals)
- ✓ **Variance** of measures **may vary** over time; correlation between measures may differ depending on distance
- ✓ The following factors are included:
  - **Treatment** (between-subjects effect)
  - **Time** (within-subjects effect)
  - **Interaction treatment\*time**



# Longitudinal data: SPLIT-PLOT

Modified from prev. course



	t1	t2	t3	t4
Sub. 1				
Sub. 2				
Sub. 3				
Sub. 4				
Sub. 5				
Sub. 6				
Sub. 7				
Sub. 8				

Treat. 1

	t1	t2	t3	t4
Sub. 9				
Sub. 10				
Sub. 11				
Sub. 12				
Sub. 13				
Sub. 14				
Sub. 15				
Sub. 16				

Treat. 2

- ✓ Two treatments on 16 subjects (e.g., cows, pigs, seedling, plants); different measurements (e.g. weight, height...) at subsequent time intervals (t1, t2, t3, t4)

# Longitudinal data: SPLIT-PLOT

- ✓ Split-plot with one Subject effect

$$y_{ijk} = \mu + \rho_i + \alpha_{ij} + \beta_k + (\rho\beta)_{jk} + \varepsilon_{ijk}$$

$$i = 1, \dots, I \quad j = 1, \dots, J \quad k = 1, \dots, K \quad N = I * J * K$$

$\rho_i$ : Treatment effect

$\alpha_{ij}$ : Interaction Subject within treatment  $\sim N(0, \sigma^2_v)$ :  
RANDOM EFFECT

$\beta_k$ : Time effect

$\rho\beta_{ik}$ : Interaction Treatment x Time

$\varepsilon_{ijk}$ : residual or error of sub-plot  $\sim N(0, \sigma^2_e)$



# Analysis of longitudinal data



- ✓ **Repeated measurements** on the same subject → tend to be **more similar** than measures taken on different subjects
- ✓ Measures made **close in time** on the same subjects tend to be **more highly correlated** than measures taken at far temporal moments
- ✓ The analysis of repeated measures data should account for the presence of **correlation between observations** on the same subject and for possible **non-constant variance**
- ✓ Data analysis alternatives: univariate **(split-plot on time)** or **multi-trait** (each time as a different variable (trait); or intermediate → more proper **(co)variance structures** (e.g. auto-regressive)



# Analysis of longitudinal data

- ✓ Different strategies for analysing longitudinal data have been developed over time; nowadays **mixed models accounting for different (co)variance structures for residuals** are preferred

Univariate analysis of single temporal lags (linear model)

Multivariate analysis – each time is a different trait (linear model)

Mixed model analysis including the experimental unit as random factor

Mixed model analysis accounting for different (co)variance structures for residual variance

Random regression analysis (mixed model analysis)



# Mixed model analysis

- ✓ Mixed model analysis enables a **flexible approach** for modelling (co)variance structures
- ✓ Mixed model analysis handles unbalanced data with unequally spaced time points within and across subjects
- ✓ Different (co)variance structures may be investigated to find the most appropriate one for fitting the model
- ✓ No problems for missing data



# Mixed models: repeated measures



First example  
of RM

- ✓ 28 young bulls divided in 4 pens including 7 bulls each
- ✓ age: age (days) at arrival at performance testing centre
- ✓ bw1-bw8: 8 monthly body weight measures
- ✓ Difference in growth in different pens?



animal	group	age	bw1	bw2	bw3	bw4	bw5	bw6	bw7	bw8
1	A	18	59	80	110	144	155	194	210	257
2	A	16	57	66	98	131	153	197	221	276
3	A	40	79	95	126	161	182	221	245	296
4	A	34	67	85	116	157	193	226	245	296
5	A	39	62	107	118	135	158	185	210	248
6	A	42	82	88	118	145	172	199	220	270
7	A	32	74	94	126	163	193	227	251	296
8	B	19	49	78	104	159	198	243	281	294
9	B	24	52	77	100	134	156	179	207	226
10	B	15	53	67	95	140	177	209	246	271
11	B	13	53	65	85	116	151	186	215	239
12	B	38	60	80	108	146	173	210	248	277
13	B	26	44	64	88	124	156	194	224	252
14	B	31	57	75	102	144	190	228	265	285
15	B	12	50	71	81	82	102	144	170	200
16	C	18	44	91	130	152	202	261	293	320
17	C	38	62	118	164	196	250	296	341	375
18	C	15	52	98	148	179	233	286	325	360
19	C	35	58	96	130	162	209	256	290	317
20	C	42	58	117	155	191	242	293	325	351
21	C	27	71	88	102	155	191	238	273	303
22	C	14	52	88	134	163	216	269	303	328
23	D	45	74	105	133	140	192	224	258	297
24	D	47	78	105	117	142	176	208	231	281
25	D	64	84	117	130	148	179	203	226	260
26	D	46	85	132	146	170	210	237	272	322
27	D	46	73	103	113	156	183	211	244	288
28	D	29	64	83	119	165	202	245	273	326

# Mixed models: repeated measures



Second

example of RM

- ✓ 67 maple seedling, equally divided in 4 growth solutions with
  - a) 0  $\mu\text{M}$ ; b) 100  $\mu\text{M}$ ; c) 300  $\mu\text{M}$ ; d) 600  $\mu\text{M}$  of alluminium

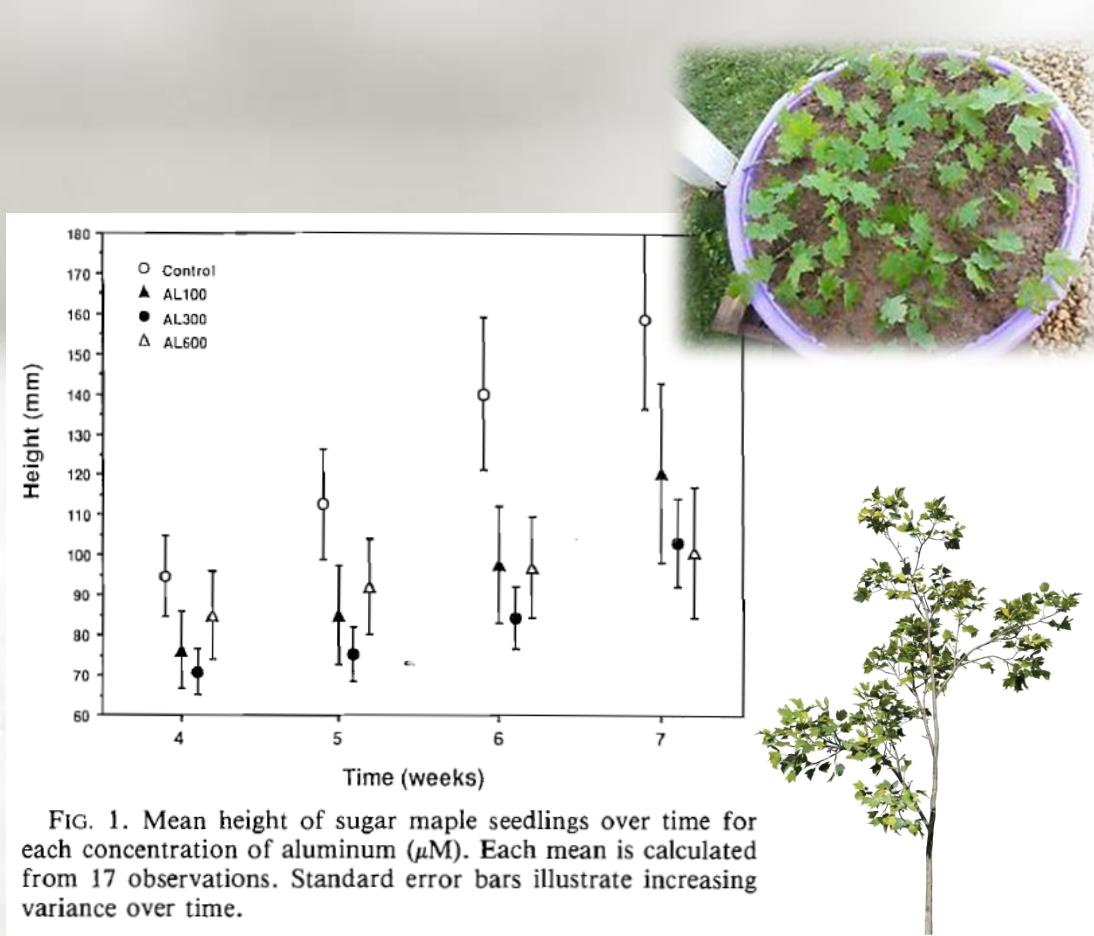


FIG. 1. Mean height of sugar maple seedlings over time for each concentration of aluminum ( $\mu\text{M}$ ). Each mean is calculated from 17 observations. Standard error bars illustrate increasing variance over time.

seedling	alluminium	t1	t2	t3	t4
1	0	60	62	78	104
2	0	41	50	60	60
3	0	85	97	115	120
4	0	88	87	90	80
5	0	66	65	80	95
6	0	106	100	133	172
7	0	61	65	65	54
8	0	52	52	65	50
9	0	194	210	250	300
10	0	86	90	105	163
11	0	140	183	283	310
12	0	41	45	83	135
13	0	99	190	280	262
14	0	144	208	220	221
15	0	117	155	208	323
16	0	102	100	105	100
17	100	70	86	100	90
18	100	167	190	230	320
19	100	116	120	157	291
20	100	85	100	125	180
21	100	90	85	90	150
22	100	44	45	70	82
23	100	71	68	70	65
24	100	40	35	37	34
25	100	29	30	32	32

# Mixed models: repeated measures

First example  
of RM

- ✓ Data have to be imported as usual;
- ✓ different groups (treatments) & subgroups (repeated effect of age) in column
- ✓ Dataset “rep.txt”



animal	group	age	bw
1	A	1	59
2	A	1	57
3	A	1	79
4	A	1	67
5	A	1	62
6	A	1	82
7	A	1	74
8	B	1	49
9	B	1	52
10	B	1	53
11	B	1	53
12	B	1	60
13	B	1	44
14	B	1	57
15	B	1	50
16	C	1	44
17	C	1	62
18	C	1	52
19	C	1	58
20	C	1	58
21	C	1	71
22	C	1	52
23	D	1	74
24	D	1	78
25	D	1	84
26	D	1	85
27	D	1	73
28	D	1	64
1	A	2	80
2	A	2	66
3	A	2	95
...			

# Mixed models: repeated measures



## ✓ 2 steps:

1. Investigation of a **proper (co)variance structure** for modelling individual phenotypic variation over time
  - *Fixed effect models with possibility to display different (co)variance structures*
  
2. Run a number of **mixed models** including the individual within group as random effect, and using different covariance structures **for modelling phenotypic variation over time**
  - *Random effect models, possibility to account but not to display different (co)variance structures*

- ✓ Phenotype: body weight, **bw**
- ✓ Time: 8 **ages** of measurement
- ✓ Individual: **animal** (young bull)
- ✓ **Group**: pens with 7 animals each



# Focus on: R with block-diagonal structure

- ✓ Step 1: in our example, a random effect is not specified, but a random matrix R for residuals is considered
- \* Since blocks refer to repeated measures of a subject, then within-block residual (co)variances are not independent

Since blocks refer to repeated measures of a subject, then within-block residual (co)variances are not independent

Matrix of residual (co)variance:  $\mathbf{R}$   
 $\text{Var}(\mathbf{e}) \rightarrow \text{not more } \mathbf{R} = \sigma_e^2 \cdot \mathbf{I}$ , but more complex structure

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e} *$$

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \\ \mathbf{y}_4 \\ \mathbf{y}_5 \\ \mathbf{y}_6 \\ \mathbf{y}_7 \\ \mathbf{y}_8 \\ \mathbf{y}_9 \\ \mathbf{y}_{10} \\ \mathbf{y}_{11} \\ \mathbf{y}_{12} \end{bmatrix} = 
 \begin{bmatrix} 1 & 1 & 0 & 0 & \mathbf{x}_1 \\ 1 & 1 & 0 & 0 & \mathbf{x}_2 \\ 1 & 1 & 0 & 0 & \mathbf{x}_3 \\ 1 & 1 & 0 & 0 & \mathbf{x}_4 \\ 1 & 0 & 1 & 0 & \mathbf{x}_5 \\ 1 & 0 & 1 & 0 & \mathbf{x}_6 \\ 1 & 0 & 1 & 0 & \mathbf{x}_7 \\ 1 & 0 & 1 & 0 & \mathbf{x}_8 \\ 1 & 0 & 0 & 1 & \mathbf{x}_9 \\ 1 & 0 & 0 & 1 & \mathbf{x}_{10} \\ 1 & 0 & 0 & 1 & \mathbf{x}_{11} \\ 1 & 0 & 0 & 1 & \mathbf{x}_{12} \end{bmatrix} \cdot 
 \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta \end{bmatrix} + 
 \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \cdot 
 \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} + 
 \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \mathbf{e}_3 \\ \mathbf{e}_4 \\ \mathbf{e}_5 \\ \mathbf{e}_6 \\ \mathbf{e}_7 \\ \mathbf{e}_8 \\ \mathbf{e}_9 \\ \mathbf{e}_{10} \\ \mathbf{e}_{11} \\ \mathbf{e}_{12} \end{bmatrix}$$

Vector of residual term

# Focus on: R with block-diagonal structure

Up to now:

Matrix of residual (co)variance  
with uncorrelated residuals:

$$R = \sigma_e^2 \cdot I$$

$$R = \sigma_e^2 \cdot \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$



All off-diagonal elements are = 0

$$R = \begin{bmatrix} \sigma_e^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_e^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_e^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_e^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_e^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_e^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_e^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_e^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_e^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_e^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_e^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_e^2 \end{bmatrix}$$



# Focus on: R with block-diagonal structure



Correlated measures within block, uncorrelated elsewhere



Subject 1, Time 1		0 0 0 0	0 0 0 0	0 0 0 0
Subject 1, Time 2		0 0 0 0	0 0 0 0	0 0 0 0
Subject 1, Time 3		0 0 0 0	0 0 0 0	0 0 0 0
Subject 1, Time 4		0 0 0 0	0 0 0 0	0 0 0 0
Subject 2, Time 1	Block 1 (subject 1)	0 0 0 0		0 0 0 0
Subject 2, Time 2		0 0 0 0		0 0 0 0
Subject 2, Time 3		0 0 0 0	Block 2 (subject 2)	0 0 0 0
Subject 2, Time 4		0 0 0 0		0 0 0 0
Subject 3, Time 1		0 0 0 0	0 0 0 0	Block 3 (subject 3)
Subject 3, Time 2		0 0 0 0	0 0 0 0	
Subject 3, Time 3		0 0 0 0	0 0 0 0	
Subject 3, Time 4		0 0 0 0	0 0 0 0	

# Focus on: R with block-diagonal structure

Matrix of residual (co)variance:  $\mathbf{R}$

**Var(e)** → not more  $\mathbf{R} = \sigma_e^2 \cdot \mathbf{I}$ , but  
more complex structure

- ✓ Each block includes the residual (co)variance structure of one subject
- ✓ All subjects have the same residual (co)variance structure
- ✓ Block 1 = Block 2 = Block 3...

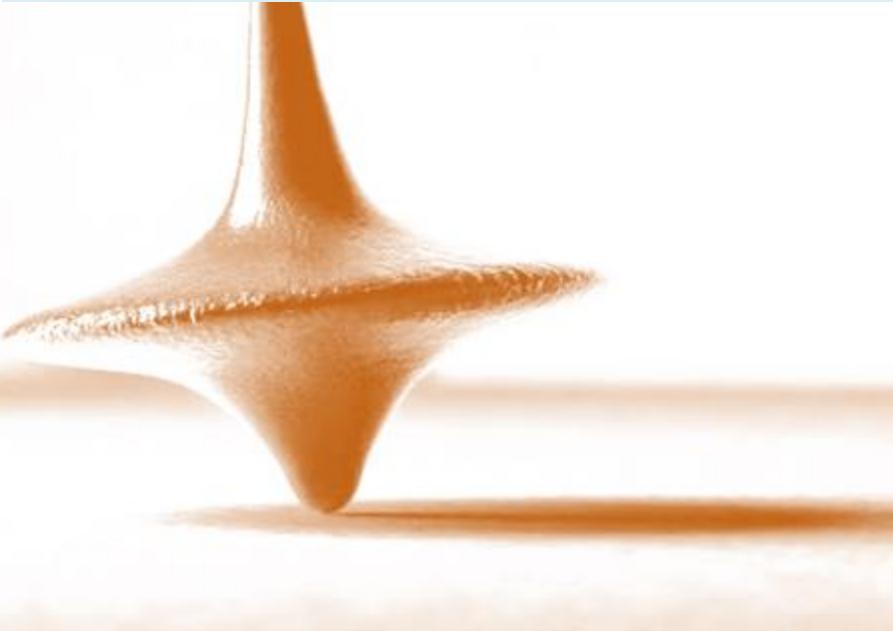


		0	0	0	0	0	0	0	0
	Block 1 (subject 1)	0	0	0	0	0	0	0	0
		0	0	0	0	0	0	0	0
		0	0	0	0	0	0	0	0
R =		0	0	0	0				
		0	0	0	0	Block 2 (subject 2)			
		0	0	0	0				
		0	0	0	0				
		0	0	0	0				
		0	0	0	0	0	0	0	0
		0	0	0	0	0	0	0	0
		0	0	0	0	0	0	0	0
		0	0	0	0	0	0	0	0
		0	0	0	0	0	0	0	0
	Block 3 (subject 3)								



# Focus on: types of covariances structures

- ✓ Overview of (co)variance structures that can be requested to model residual variance within blocks.
- ✓ As a **default**, the **Variance component** structure is used (that is  $V(e) = R = \sigma_e^2 \cdot I$ )



**Table 56.15 Covariance Structure Examples**

Description	Structure	Example
Variance Components	VC (default)	$\begin{bmatrix} \sigma_B^2 & 0 & 0 & 0 \\ 0 & \sigma_B^2 & 0 & 0 \\ 0 & 0 & \sigma_{AB}^2 & 0 \\ 0 & 0 & 0 & \sigma_{AB}^2 \end{bmatrix}$
Compound Symmetry	CS	$\begin{bmatrix} \sigma^2 + \sigma_1 & \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma^2 + \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 \end{bmatrix}$
Unstructured	UN	$\begin{bmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$
Banded Main Diagonal	UN(1)	$\begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{bmatrix}$
First-Order Autoregressive	AR(1)	$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$
Toeplitz	TOEP	$\begin{bmatrix} \sigma^2 & \sigma_1 & \sigma_2 & \sigma_3 \\ \sigma_1 & \sigma^2 & \sigma_1 & \sigma_2 \\ \sigma_2 & \sigma_1 & \sigma^2 & \sigma_1 \\ \sigma_3 & \sigma_2 & \sigma_1 & \sigma^2 \end{bmatrix}$

# Step 1: Examples of models with different (co)variance structure

## 1<sup>st</sup> covariance structure investigated: Unstructured



```
#rm(list = ls()) # Clean the workspace
#install.packages("nlme") # If necesary install the library
library(lsmeans)           # Call the library to the workspace
library(nlme)              # Call the library to the workspace

# Read the data
myfile<-"C:/Users/toledo/Dropbox/UNIPD/Biostatistics Curse R Spring 2018/curso STAT phD 2018 Mixed Model"
mydata<-read.table(file=myfile,stringsAsFactors = TRUE,header = TRUE,sep = "\t")
options(contrasts = c("contr.SAS","contr.poly")) # Change contrast options
mydata$animal<-as.factor(mydata$animal)      # Set as factor
mydata$age<-as.factor(mydata$age)             # Set as factor

# Different covariance Structures
# Model Unstructured
model.un<-gls(bw ~ group + age + group*age, data = mydata,
               correlation = corSymm(form = ~ 1|group/animal),
               weights = varIdent(form = ~ 1|age))

model.un                                # Model fitted
                                         # Model
                                         # General correlation matrix, unstructured
                                         # Constant variance(s), used to allow different
                                         # variances according to the levels of a classification factor
```

**Unstructured:** Constant variance(s), used to allow different variances according to the levels of a classification factor

# Unstructured (UN) covariance matrix



```
# Different covariance Structures  
# Model Unstructured  
model.un<-gls(bw ~ group + age + group*age, data = mydata, # Model  
correlation = corSymm(form = ~ 1|group/animal),  
weights = varIdent(form = ~ 1|age))  
# General correlation matrix, unstructure.  
# Constant variance(s), used to allow different  
# variances according to the levels of a classification factor.
```

- ✓ **UN** specifies a **completely general (unstructured) covariance matrix** parameterized directly in terms of variances and covariances. The variances are constrained to be nonnegative, and the covariances are unconstrained.
- ✓ A number of  $k+k(k-1)/2$  parameters (diagonal + off-diagonal) are estimated
  - Not parsimonious model; **Computationally demanding**
  - Best choice when no assumptions about covariance structure are possible

Unstructured UN

$$\begin{bmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$$

# Unstructured (UN) covariance matrix

```
# Different covariance Structures
# Model Unstructured
model.un<-gls(bw ~ group + age + group*age, data = mydata,           # Model
               correlation = corSymm(form = ~ 1|group/animal),
               weights = varIdent(form = ~ 1|age))                                # General correlation matrix, unstructure.
# Constant variance(s), used to allow different
# variances according to the levels of a classification factor.
```

```
## Generalized least squares fit by REML
##   Model: bw ~ group + age + group * age
##   Data: mydata
##   Log-restricted-likelihood: -670.3923
##
```

```
## Coefficients:
```

	groupA	groupB	groupC	age1
## (Intercept)	295.6666667	-18.6666667	-40.1666667	40.6190476
## age2		age3	age4	age5
## -188.1666667		-169.3333333	-142.1666667	-105.3333333
## age7	groupA:age1	groupB:age1	groupC:age1	groupA:age2
## -45.0000000	10.9047619	16.0833333	-60.2380952	-0.9761905
## groupB:age2	groupC:age2	groupA:age3	groupB:age3	groupC:age3
## 4.7916667	-48.6904762	8.3333333	9.2083333	-29.3809524
## groupA:age4	groupB:age4	groupC:age4	groupA:age5	groupB:age5
## 13.1666667	17.2916667	-22.9761905	0.6190476	12.7083333
## groupC:age5	groupA:age6	groupB:age6	groupC:age6	groupA:age7
## -10.5238095	4.3333333	17.9583333	9.3333333	-3.1428571
## groupB:age7	groupC:age7			
## 21.5000000	15.8571429			
##				
## Correlation Structure: General				
## Formula: ~1   group/animal				

} Animal within group



Solutions (BLUE) for  
fixed effects (groupD  
& age8 set to zero)

# Unstructured (UN) covariance matrix

```
## Correlation Structure: General
## Formula: ~1 | group/animal
## Parameter estimate(s):
## Correlation:
##   1   2   3   4   5   6   7
## 2 0.497
## 3 0.244 0.755
## 4 0.254 0.425 0.709
## 5 0.197 0.389 0.744 0.944
## 6 0.045 0.239 0.669 0.895 0.963
## 7 0.075 0.271 0.678 0.892 0.962 0.989
## 8 0.069 0.197 0.632 0.858 0.919 0.964 0.977
## Variance function:
## Structure: Different standard deviations per stratum
## Formula: ~1 | age
## Parameter estimates:
##   1   2   3   4   5   6   7   8
## 1.000000 1.546902 1.730297 2.234001 2.801870 2.891596 3.275352 3.353758
## Degrees of freedom: 224 total; 192 residual
## Residual standard error: 7.916225
```

	Block 1 (subject 1)	0 0 0 0 0 0 0 0 0
		0 0 0 0 0 0 0 0 0
		0 0 0 0 0 0 0 0 0
		0 0 0 0 0 0 0 0 0
		0 0 0 0 0 0 0 0 0
	Block 2 (subject 2)	0 0 0 0 0 0 0 0 0
		0 0 0 0 0 0 0 0 0
		0 0 0 0 0 0 0 0 0
		0 0 0 0 0 0 0 0 0
	Block 3 (subject 3)	0 0 0 0 0 0 0 0 0
		0 0 0 0 0 0 0 0 0
		0 0 0 0 0 0 0 0 0
		0 0 0 0 0 0 0 0 0

Correlations between each time point within subject

Standard deviations for each stratum (by age)

# Unstructured (UN) covariance matrix

```
anova(model.un) # ANOVA
```

```
## Denom. DF: 192
##                  numDF  F-value p-value
## (Intercept)      1 5283.275 <.0001
## group            3   21.487 <.0001
## age              7   549.657 <.0001
## group:age        21   14.102 <.0001
```

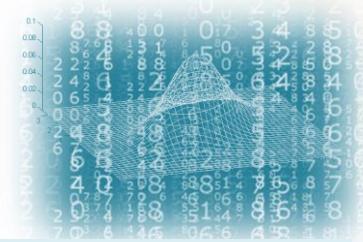
```
lsmeans(model.un, "group") # LSM
```

```
##    group    lsmean       SE  df lower.CL upper.CL
##    A     163.1964 6.097614 192 151.1695 175.2233
##    B     149.9844 5.703796 192 138.7342 161.2345
##    C     200.0000 6.097614 192 187.9731 212.0269
##    D     177.7083 6.586176 192 164.7178 190.6989
##
## Results are averaged over the levels of: age
## Confidence level used: 0.95
```

ANOVA on fixed effects (including the interaction between group & age)

Least square means for group effect

# Unstructured (UN) covariance matrix



**Unstructured matrix of residuals** for each block, **expressed as correlation** (correlation of each element within the block, i.e., the repeated measures of an individual)

```
cors<-corMatrix(model.un$modelStruct$corStruct)[[1]] # Extract the Correlation Matrix
cors
# See the Correlation Matrix

##          [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]      [,8]
## [1,] 1.00000000 0.4969533 0.2435151 0.2544810 0.1972696 0.0447571 0.0752157 0.06923708
## [2,] 0.49695332 1.0000000 0.7546379 0.4254784 0.3885893 0.2393159 0.2714338 0.19731798
## [3,] 0.24351505 0.7546379 1.0000000 0.7092706 0.7444147 0.6688010 0.6776423 0.63246018
## [4,] 0.25448097 0.4254784 0.7092706 1.0000000 0.9439046 0.8945347 0.8916583 0.85806538
## [5,] 0.19726957 0.3885893 0.7444147 0.9439046 1.0000000 0.9633293 0.9615218 0.91879140
## [6,] 0.04475710 0.2393159 0.6688010 0.8945347 0.9633293 1.0000000 0.9892438 0.96421204
## [7,] 0.07521570 0.2714338 0.6776423 0.8916583 0.9615218 0.9892438 1.0000000 0.97650071
## [8,] 0.06923708 0.1973180 0.6324602 0.8580654 0.9187914 0.9642120 0.9765007 1.00000000
..."
```

Unstructured

UN

$$\begin{bmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$$

Closer are the time points (e.g., age 7 & age 8), higher is the correlation between these time points

# Unstructured (UN) covariance matrix

## Unstructured (co)variance matrix of residuals for each block

```
stdev.st<-c(1.000000, 1.546902, 1.730297, 2.234001,  
           2.801870, 2.891596, 3.275352, 3.353758) # Get the stratum SD  
vars<-stdev.st*model.un$sigma^2 # Multiply the stratum SD by the Error Variance  
covs<-outer(vars,vars,function(x,y)sqrt(x)*sqrt(y)) # Get the Variance & Covariance Matrix  
round(cors*covs,3) # Get the R matrix: Covariance estimate for subjects
```

##	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]
## [1,]	62.667	38.733	20.073	23.836	20.693	4.769	8.530	7.946
## [2,]	38.733	96.939	77.369	49.566	50.697	31.718	38.288	28.164
## [3,]	20.073	77.369	108.432	87.388	102.715	93.748	101.094	95.476
## [4,]	23.836	49.566	87.388	139.997	147.989	142.477	151.149	147.185
## [5,]	20.693	50.697	102.715	147.989	175.584	171.832	182.536	176.499
## [6,]	4.769	31.718	93.748	142.477	171.832	181.207	190.782	188.167
## [7,]	8.530	38.288	101.094	151.149	182.536	190.782	205.255	202.817
## [8,]	7.946	28.164	95.476	147.185	176.499	188.167	202.817	210.169



By multiplying the standard deviations of one age stratum (**stdev.st**) by the error variance of the model (**model.un\$sigma^2**), the unstructured covariance matrix is obtained: each parameter is different (the same parameters above & below diagonal)

Unstructured

UN

$$\begin{bmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$$

# Unstructured (UN) covariance matrix

## Focus on: correlations

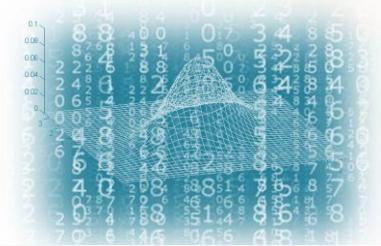
- \* ✓ Moving from diagonal, correlations decrease → close time measures have a greater covariance (they are more related) than far measures
- ✓ The program has estimated 36 different covariance parameters → very computationally demanding

```
cors<-corMatrix(model.un$modelStruct$corStruct)[[1]] # Extract the Correlation Matrix  
cors # See the Correlation Matrix
```

```
## [,1] [,2] [,3] [,4] [,5] [,6] [,7] * [,8]  
## [1,] 1.0000000 0.4969533 0.2435151 0.2544810 0.1972696 0.0447571 0.0752157 0.06923708  
## [2,] 0.49695332 1.0000000 0.7546379 0.4254784 0.3885893 0.2393159 0.2714338 0.19731798  
## [3,] 0.24351505 0.7546379 1.0000000 0.7092706 0.7444147 0.6688010 0.6776423 0.63246018  
## [4,] 0.25448097 0.4254784 0.7092706 1.0000000 0.9439046 0.8945347 0.8916583 0.85806538  
## [5,] 0.19726957 0.3885893 0.7444147 0.9439046 1.0000000 0.9633293 0.9615218 0.91879140  
## [6,] 0.04475710 0.2393159 0.6688010 0.8945347 0.9633293 1.0000000 0.9892438 0.96421204  
## [7,] 0.07521570 0.2714338 0.6776423 0.8916583 0.9615218 0.9892438 1.0000000 0.97650071  
## [8,] 0.06923708 0.1973180 0.6324602 0.8580654 0.9187914 0.9642120 0.9765007 1.00000000  
...
```

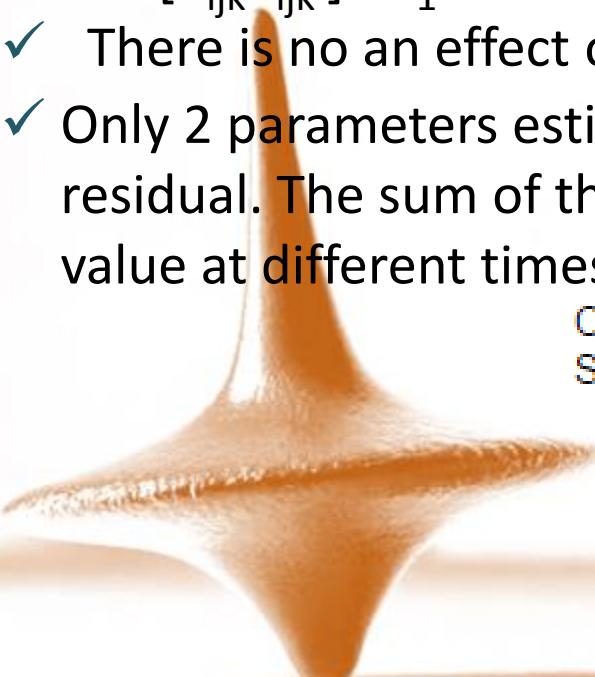


# Compound symmetry (CS) covariance matrix



```
# Model Compound Symmetry Correlation Structure  
model.cs<-gls(bw ~ group + age + group*age, data = mydata,      # Model  
                 correlation = corCompSymm(form = ~ 1|group/animal)) # Compound Symmetry Correlation Structure  
model.cs                         # Model fitted
```

- ✓ **constant variance**  $\text{Var}[Y_{ijk}] = \sigma^2 + \sigma_1$  and **constant covariance** over time  $\text{Cov}[Y_{ijk} Y_{ijk'}] = \sigma_1$
- ✓ There is no an effect of temporal lag (time distance between 2 measures)
- ✓ Only 2 parameters estimated: the covariance at different times ( $\sigma_1$ ) and residual. The sum of this two parameters provides the constant variance value at different times



Compound Symmetry      CS

$$\begin{bmatrix} \sigma^2 + \sigma_1 & \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma^2 + \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 \end{bmatrix}$$

# Compound symmetry (CS) covariance matrix

```
# Model Compound Symmetry Correlation Structure
model.cs<-gls(bw ~ group + age + group*age, data = mydata,      # Model
               correlation = corCompSymm(form = ~ 1|group/animal)) # Compound Symmetry Correlation Structure
model.cs                                         # Model fitted

## Generalized least squares fit by REML
##   Model: bw ~ group + age + group * age
##   Data: mydata
##   Log-restricted-likelihood: -813.9656

## Coefficients:
## (Intercept)      groupA      groupB      groupC      age1
## 295.6666667 -18.6666667 -40.1666667  40.6190476 -219.3333333
##       age2      age3      age4      age5      age6
## -188.1666667 -169.3333333 -142.1666667 -105.3333333 -74.3333333
##       age7 groupA:age1 groupB:age1 groupC:age1 groupA:age2
## -45.0000000 10.9047619 16.0833333 -60.2380952 -0.9761905
## groupB:age2 groupC:age2 groupA:age3 groupB:age3 groupC:age3
## 4.7916667 -48.6904762  8.3333333  9.2083333 -29.3809524
## groupA:age4 groupB:age4 groupC:age4 groupA:age5 groupB:age5
## 13.1666667 17.2916667 -22.9761905  0.6190476 12.7083333
## groupC:age5 groupA:age6 groupB:age6 groupC:age6 groupA:age7
## -10.5238095  4.3333333 17.9583333  9.3333333 -3.1428571
## groupB:age7 groupC:age7
## 21.5000000 15.8571429
##                                         Compound          CS
## Correlation Structure: Compound symmetry
## Formula: ~1 | group/animal
## Parameter estimate(s):
##   Rho
## 0.6232494
## Degrees of freedom: 224 total; 192 residual
## Residual standard error: 19.70444
```

$$\begin{bmatrix} \sigma^2 + \sigma_1 & \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma^2 + \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 \end{bmatrix}$$

Rho ( $\sigma$ ) is the parameter of the matrix

# Compound symmetry (CS) covariance matrix



**Compound symmetry matrix of residuals** for each block, **expressed as correlation** (correlation of each element within the block, i.e., the repeated measures of an individual)

```
cors<-corMatrix(model.cs$modelStruct$corStruct)[[1]] # Extract the Correlation Matrix
cors
# See the Correlation Matrix

##          [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]      [,8]
## [1,] 1.0000000 0.6232494 0.6232494 0.6232494 0.6232494 0.6232494 0.6232494 0.6232494
## [2,] 0.6232494 1.0000000 0.6232494 0.6232494 0.6232494 0.6232494 0.6232494 0.6232494
## [3,] 0.6232494 0.6232494 1.0000000 0.6232494 0.6232494 0.6232494 0.6232494 0.6232494
## [4,] 0.6232494 0.6232494 0.6232494 1.0000000 0.6232494 0.6232494 0.6232494 0.6232494
## [5,] 0.6232494 0.6232494 0.6232494 0.6232494 1.0000000 0.6232494 0.6232494 0.6232494
## [6,] 0.6232494 0.6232494 0.6232494 0.6232494 0.6232494 1.0000000 0.6232494 0.6232494
## [7,] 0.6232494 0.6232494 0.6232494 0.6232494 0.6232494 0.6232494 1.0000000 0.6232494
## [8,] 0.6232494 0.6232494 0.6232494 0.6232494 0.6232494 0.6232494 0.6232494 1.0000000
```

1 in diagonal, the same correlation elsewhere

Compound  
Symmetry

CS

$$\begin{bmatrix} \sigma^2 + \sigma_I & \sigma_I & \sigma_I & \sigma_I \\ \sigma_I & \sigma^2 + \sigma_I & \sigma_I & \sigma_I \\ \sigma_I & \sigma_I & \sigma^2 + \sigma_I & \sigma_I \\ \sigma_I & \sigma_I & \sigma_I & \sigma^2 + \sigma_I \end{bmatrix}$$

# Compound symmetry (CS) covariance matrix



## Compound symmetry

```
anova(model.cs)      # ANOVA

## Denom. DF: 192
##          numDF  F-value p-value
## (Intercept)     1 3172.750 <.0001
## group          3   12.937 <.0001
## age            7 1243.995 <.0001
## group:age      21   10.043 <.0001

lsmeans(model.cs,"group") # LSM
```

```
## group  lsmean      SE  df lower.CL upper.CL
## A      163.1964 6.097667 192 151.1694 175.2234
## B      149.9844 5.703845 192 138.7341 161.2346
## C      200.0000 6.097667 192 187.9730 212.0270
## D      177.7083 6.586233 192 164.7177 190.6990
##
## Results are averaged over the levels of: age
## Confidence level used: 0.95
```

## Unstructured

```
anova(model.un)      # ANOVA

## Denom. DF: 192
##          numDF  F-value p-value
## (Intercept)     1 5283.275 <.0001
## group          3   21.487 <.0001
## age            7   549.657 <.0001
## group:age      21   14.102 <.0001

lsmeans(model.un,"group") # LSM

## group  lsmean      SE  df lower.CL upper.CL
## A      163.1964 6.097614 192 151.1695 175.2233
## B      149.9844 5.703796 192 138.7342 161.2345
## C      200.0000 6.097614 192 187.9731 212.0269
## D      177.7083 6.586176 192 164.7178 190.6989
##
## Results are averaged over the levels of: age
## Confidence level used: 0.95
```

Similar ANOVA and lsmeans (that differ only by SE) using different (co)variance matrices

# Compound symmetry (CS) covariance matrix



Multiplying the error variance by the correlation matrix (2 slides ago), the Compound Symmetry matrix is obtained

```
err.var<-model.cs$sigma^2          # Err. Variance  
round(cors*err.var,3)               # Get the R matrix: Covariance estimate for subjects  
  
## [,1]    [,2]    [,3]    [,4]    [,5]    [,6]    [,7]    [,8]  
## [1,] 388.265 241.986 241.986 241.986 241.986 241.986 241.986 241.986  
## [2,] 241.986 388.265 241.986 241.986 241.986 241.986 241.986 241.986  
## [3,] 241.986 241.986 388.265 241.986 241.986 241.986 241.986 241.986  
## [4,] 241.986 241.986 241.986 388.265 241.986 241.986 241.986 241.986  
## [5,] 241.986 241.986 241.986 241.986 388.265 241.986 241.986 241.986  
## [6,] 241.986 241.986 241.986 241.986 241.986 388.265 241.986 241.986  
## [7,] 241.986 241.986 241.986 241.986 241.986 241.986 388.265 241.986  
## [8,] 241.986 241.986 241.986 241.986 241.986 241.986 241.986 388.265
```

Estimated parameter for covariance  $\sigma_1$  (all the off-diagonal part of the matrix)

Estimated parameter for diagonal variance  $\sigma^2 + \sigma_1$   
( $241.99 + 146.28 = 388.26$ )

Compound Symmetry

CS

$$\begin{bmatrix} \sigma^2 + \sigma_1 & \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma^2 + \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 \end{bmatrix}$$

# Variance components (VC) covariance matrix



A parameter in diagonal (the error term of the model), and zeros elsewhere. This matrix specifies standard variance components and is the **default structure** for linear models & mixed models.

Variance  
Components

VC (default)

$$\begin{bmatrix} \sigma_B^2 & 0 & 0 & 0 \\ 0 & \sigma_B^2 & 0 & 0 \\ 0 & 0 & \sigma_{AB}^2 & 0 \\ 0 & 0 & 0 & \sigma_{AB}^2 \end{bmatrix}$$

1 variance estimated for the diagonal,  
off-diagonal = 0: classical structure  $\mathbf{R}=\sigma_e^2 \cdot \mathbf{I}$

# Variance components (VC) covariance matrix

# A model with No Correlation looks like:

```
diag(1,8,8)*err.var
```

```
##      [,1]   [,2]   [,3]   [,4]   [,5]   [,6]   [,7]   [,8]
## [1,] 388.265 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## [2,] 0.000 388.265 0.000 0.000 0.000 0.000 0.000 0.000
## [3,] 0.000 0.000 388.265 0.000 0.000 0.000 0.000 0.000
## [4,] 0.000 0.000 0.000 388.265 0.000 0.000 0.000 0.000
## [5,] 0.000 0.000 0.000 0.000 388.265 0.000 0.000 0.000
## [6,] 0.000 0.000 0.000 0.000 0.000 388.265 0.000 0.000
## [7,] 0.000 0.000 0.000 0.000 0.000 0.000 388.265 0.000
## [8,] 0.000 0.000 0.000 0.000 0.000 0.000 0.000 388.265
```



Multiplying the error variance by an Identity matrix, the variance component matrix is obtained. Off-diagonal values are set to zero.

Variance  
Components

VC (default)

$$\begin{bmatrix} \sigma_B^2 & 0 & 0 & 0 \\ 0 & \sigma_B^2 & 0 & 0 \\ 0 & 0 & \sigma_{AB}^2 & 0 \\ 0 & 0 & 0 & \sigma_{AB}^2 \end{bmatrix}$$

## Step 2: mixed models with different covariance structures evaluated by their model fitting

```
# Fit the same model with different types of variance and covariance
```

```
model.a<-lme(bw ~ group + age + group*age, data = mydata, # Model  
               random = ~ 1|group/animal, # Random Effect  
               weights = varIdent(form = ~ 1|age), # Constant variance(s)  
               correlation = NULL) # No correlation
```

```
model.c<-lme(bw ~ group + age + group*age, data = mydata, # Model  
               random = ~ 1|group/animal, # Random Effect  
               weights = varIdent(form = ~ 1|age), # Constant variance(s)  
               correlation = corAR1()) # autocorrelation structure of order 1
```

```
model.d<-lme(bw ~ group + age + group*age, data = mydata, # Model  
               random = ~ 1|group/animal, # Random Effect  
               correlation = corCompSymm()) # Compound Symmetry Correlation Structure
```

```
anova(model.a,model.c,model.d) # Model Comparison
```



# First order autoregressive (AR) covariance matrix

```
model1.c<-lme(bw ~ group + age + group*age, data = mydata, # Model  
                 random = ~ 1|group/animal, # Random Effect  
                 weights = varIdent(form = ~ 1|age), # Constant variance(s)  
                 correlation = corAR1()) # autocorrelation structure of order 1
```



- ✓ The correlations in the same individuals and at different times is **function of the distance between the times ( $\rho, \rho^2, \rho^3, \dots$ )**
- ✓ Observations at close time points (e.g. T1 & T2) is  $= \rho$  disregard the time point (T1 & T2, T2 & T3, T3 & T4..)
- ✓ Observations at 2-time-points-distance (T1 & T3) have a correlation of  $\rho^2$
- ✓ 2 variance parameters to estimate ( $\rho$  and residual  $\sigma_e^2$ ), as for CS

First-Order  
Autoregressive

AR(1)

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

# Step 2: mixed models with different covariance structures: Variance Components



# Fit the same model with different types of variance and covariance

```
model.a<-lme(bw ~ group + age + group*age, data = mydata, # Model  
               random = ~ 1|group/animal, # Random Effect  
               weights = varIdent(form = ~ 1|age), # Constant variance(s)  
               correlation = NULL) # No correlation
```

model.a

Random effects:  
Formula: ~1 | group  
 (Intercept)  
StdDev: 8.299868

Formula: ~1 | animal %in% group  
 (Intercept) Residual  
StdDev: 22.5659 23.29137

Variance function:  
Structure: Different standard deviations per stratum  
Formula: ~1 | age  
Parameter estimates:

1 2 3 4 5 6 7  
1.0000000 0.9663973 0.7065238 0.4234375 0.2305685 0.1048387 0.2020408  
8  
0.3211019

Number of Observations: 224

Number of Groups:

group animal %in% group  
4 28



**age** is a fixed effect and it is modelled to have different blocks of residual variance for groups in relation to the different ages → variances according to the levels of the age factor

Variance Components	VC (default)
$\sigma_B^2$	0 0 0 0
0	$\sigma_B^2$ 0 0
0	0 $\sigma_{AB}^2$ 0
0	0 0 $\sigma_{AB}^2$



Standard deviations for each stratum (by age)

# Step 2: mixed models with different covariance structures: Autoregressive

```
model.c<-lme(bw ~ group + age + group*age, data = mydata, # Model  
               random = ~ 1|group/animal, # Random Effect  
               weights = varIdent(form = ~ 1|age), # Constant variance(s)  
               correlation = corAR1())
```

model.c

```
## Random effects:  
## Formula: ~1 | group
```

```
##           (Intercept)  
## StdDev:    7.962176
```

##

```
## Formula: ~1 | animal %in% group  
##           (Intercept) Residual
```

```
## StdDev:    22.58813 22.34425
```

##

```
## Correlation Structure: AR(1)  
## Formula: ~1 | group/animal
```

```
## Parameter estimate(s):  
##   Phi
```

```
## 0.8568647
```

## Variance function:

```
## Structure: Different standard deviations per stratum
```

```
## Formula: ~1 | age
```

```
## Parameter estimates:
```

```
##      1       2       3       4       5       6       7
```

```
## 1.0000000 1.0049240 0.8437745 0.6594190 0.5226454 0.3214862 0.4282820
```

```
##      8
```

```
## 0.4803856
```

```
## Number of Observations: 224
```

```
## Number of Groups:
```

```
##           group animal %in% group
```

```
##      4          28
```

First-Order  
Autoregressive      AR(1)

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$



## Step 2: mixed models with different covariance structures: Compound Symmetry

```
model.d<-lme(bw ~ group + age + group*age, data = mydata, # Model  
               random = ~ 1|group/animal, # Random Effect  
               correlation = corCompSymm()) # Compound Symmetry Correlation Structure
```

```
model.d
```

```
## Random effects:  
## Formula: ~1 | group  
##           (Intercept)  
## StdDev:    4.309886  
##  
## Formula: ~1 | animal %in% group  
##           (Intercept) Residual  
## StdDev:    15.5559 12.09459  
##  
## Correlation Structure: Compound symmetry
```

```
## Formula: ~1 | group/animal  
## Parameter estimate(s):  
  
## Rho  
## 0  
  
## Number of Observations: 224  
## Number of Groups:  
##           group animal %in% group  
##             4          28
```

Compound Symmetry

CS

Rho

$$\begin{bmatrix} \sigma^2 + \sigma_1 & \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma^2 + \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 \end{bmatrix}$$



## Step 2: mixed models with different covariance structures evaluated by their model fitting

```
anova(model.a,model.c,model.d)
```

# Model Comparisson

##	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
## model.a	1	42	1649.778	1786.593	-782.8890			
## model.c	2	43	1536.690	1676.762	-725.3449	1 vs 2	115.0882	<.0001
## model.d	3	36	1699.931	1817.201	-813.9656	2 vs 3	177.2414	<.0001

**Model.a:** costant variance (**variance components**)

**Model.c:** autoregressive variance matrix

**Model.d:** compound symmetry correlation structure

Lower AIC & BIC for Model.c: better fitting

**ANOVA on models:** t test used to compare model fitting estimates (P value)

# Summary of different matrix structures

TYPE=	Time points are		
	Equally spaced, same across subjects	Unequally spaced, same across subjects	Different across subjects
VC	n/a	n/a	n/a
UN	Yes	Yes	No
CS	Yes	Yes	Yes
AR(1)	Yes	No	No

# Correlated residuals & correlated values of random effect



- ✓ **Random effects** often used to model between-subject variation (G matrix, random effect)
- ✓ **Correlated residuals** often used to model within-subject variation (R matrix, residual term)

**Warning!!!! Possible confounding when the same covariance structure is used both for random effects & residuals!**

# Mixed models: repeated measures



Second

example of RM

- ✓ 67 maple seedling, equally divided in 4 growth solutions with
  - a) 0  $\mu\text{M}$ ; b) 100  $\mu\text{M}$ ; c) 300  $\mu\text{M}$ ; d) 600  $\mu\text{M}$  of alluminium

## Homework!

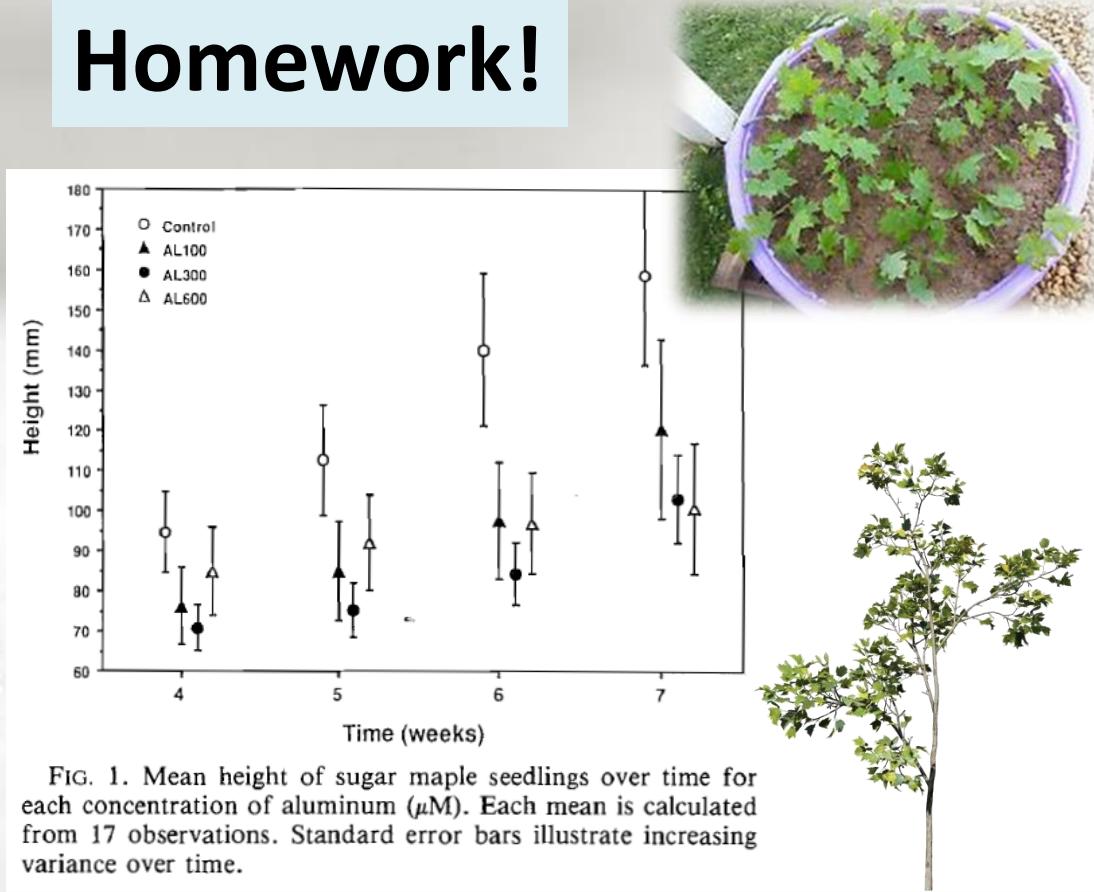
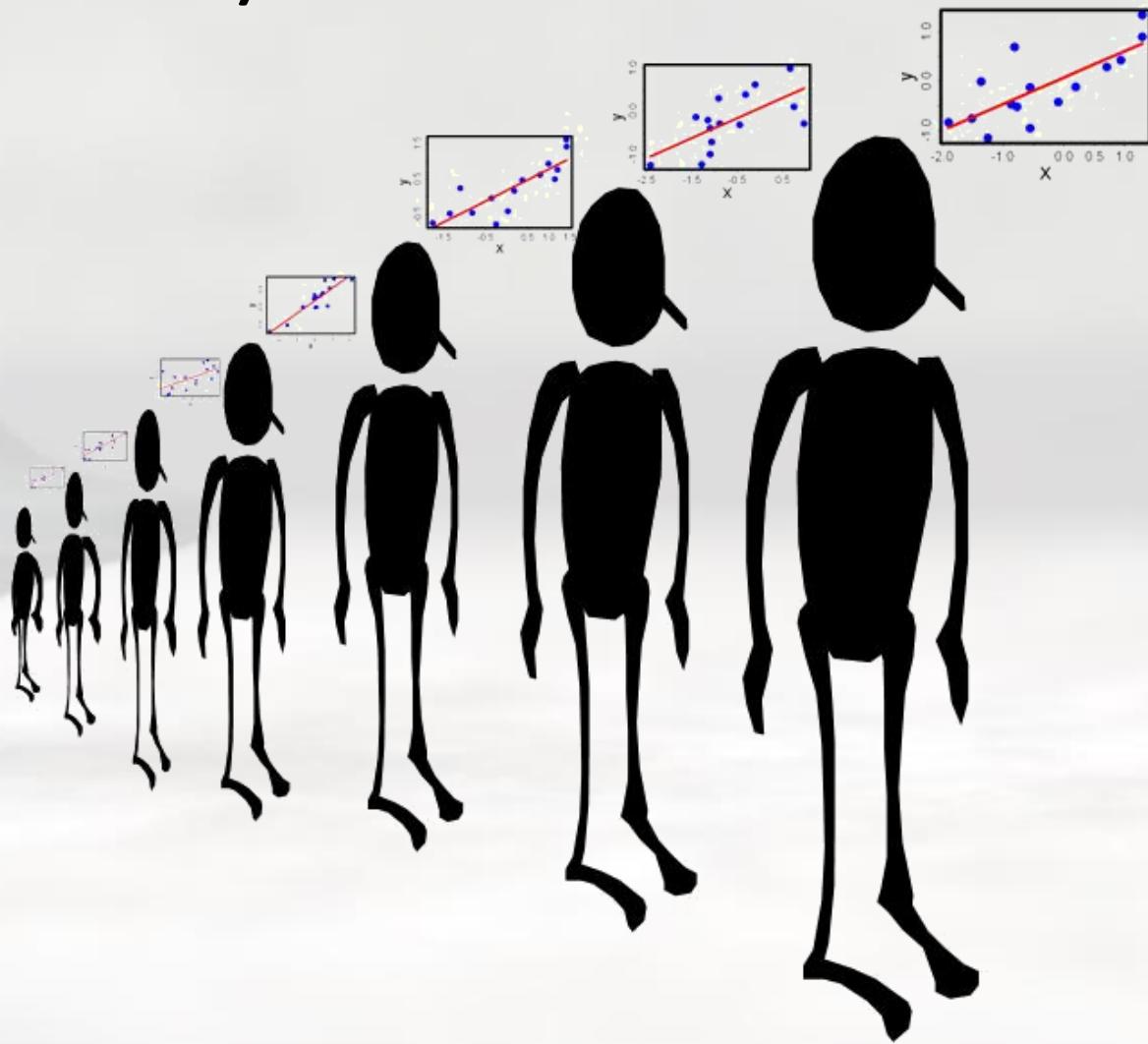


FIG. 1. Mean height of sugar maple seedlings over time for each concentration of aluminum ( $\mu\text{M}$ ). Each mean is calculated from 17 observations. Standard error bars illustrate increasing variance over time.

seedling	alluminium	t1	t2	t3	t4
1	0	60	62	78	104
2	0	41	50	60	60
3	0	85	97	115	120
4	0	88	87	90	80
5	0	66	65	80	95
6	0	106	100	133	172
7	0	61	65	65	54
8	0	52	52	65	50
9	0	194	210	250	300
10	0	86	90	105	163
11	0	140	183	283	310
12	0	41	45	83	135
13	0	99	190	280	262
14	0	144	208	220	221
15	0	117	155	208	323
16	0	102	100	105	100
17	100	70	86	100	90
18	100	167	190	230	320
19	100	116	120	157	291
20	100	85	100	125	180
21	100	90	85	90	150
22	100	44	45	70	82
23	100	71	68	70	65
24	100	40	35	37	34
25	100	29	30	32	32

# Random regression models

(just a brief look)



# Analysis of longitudinal data

✓ Different strategies for analysing longitudinal data have been developed over time; nowadays **mixed models accounting for different (co)variance structures for residuals** are preferred

Univariate analysis of single temporal lags (linear model)

Multivariate analysis – each time is a different trait (linear model)

Mixed model analysis including the experimental unit as random factor

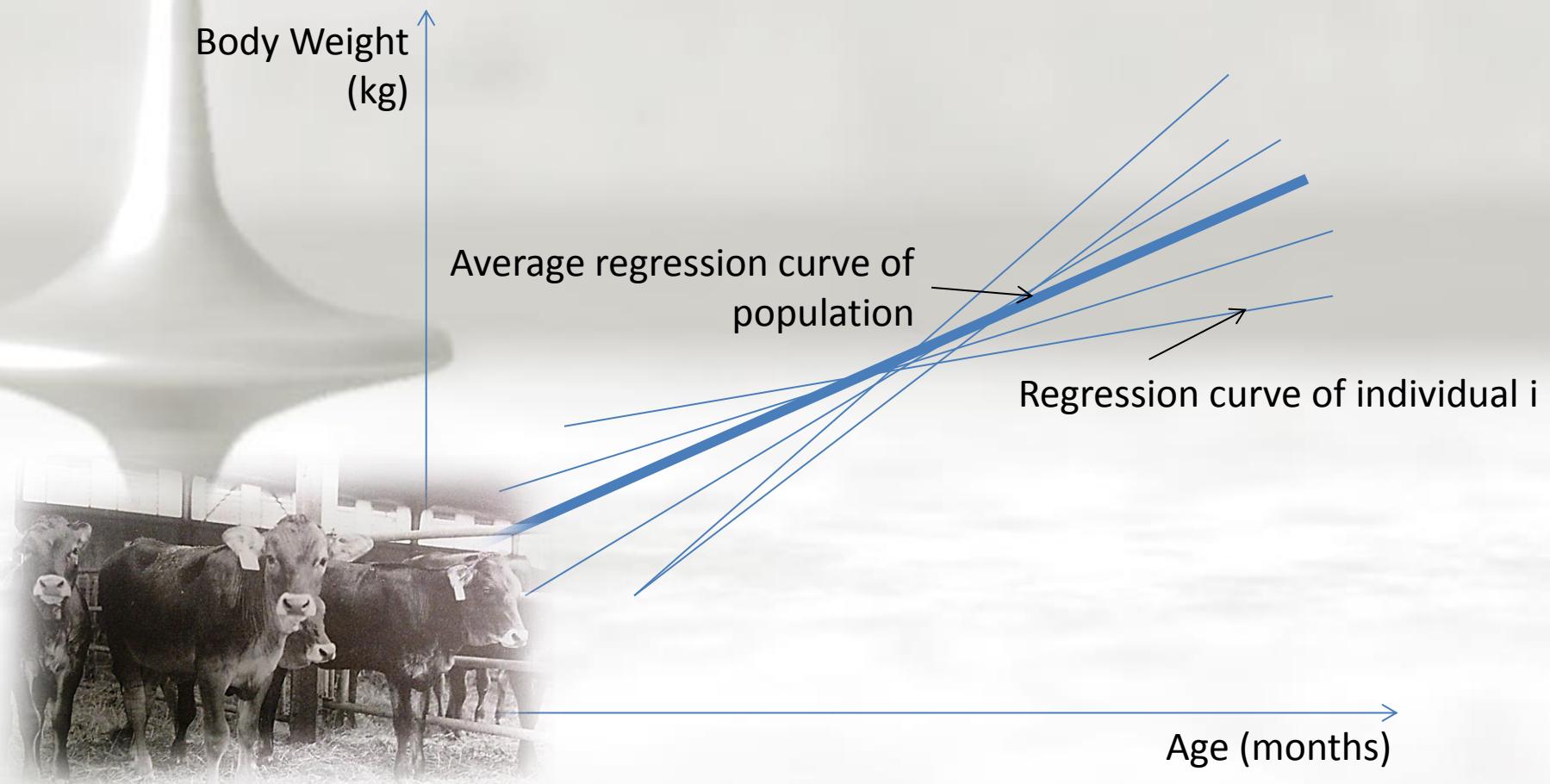
Mixed model analysis accounting for different (co)variance structures for residual variance

Random regression analysis (mixed model analysis)



# Random regression models

Another way to model the **variation over time** of individual records is to treat them as **individual covariates (random term)**



# Random regression models



Random regression models allow researchers to measure changes over time in a phenomenon of interest (e.g., response to treatment) at both the aggregate (i.e., population) and individual (i.e., study participant) levels.

In RRM, two levels of analysis (Bryk & Raudenbush, 1987; Singer & Willett, 2003):

- ✓ **LEVEL 1:** within-person or intra-individual change model.
  - It captures **person-specific (i.e., individual) variation** over time
  - **Time-varying predictor variables** (e.g., age, time elapsed since treatment) can be included in the level 1 model.
- ✓ **LEVEL 2:** between-person or inter-individual change model
  - It captures **between-person variability** in the variation over time.
  - **Time-invariant predictors** (e.g., sex, diet, treatment...) can be included in the level 2 model.

# Thanks for your attention!



- Exam at the end of the course
- Some extra lessons with examples and exercises are possible (to be organized)
- Please, let's write an email or something similar if you have any doubt

