
Applied Statistic

Chapter 12 - Random effects and mixed models

Christian Damgaard
Ecoscience
Aarhus University



Hierarchical data structures are common

It is normal to work with entities that naturally belong to groups

- Leaves on a plant
- Individuals in a population
- Ants in a colony
- Species in a community
- Populations in a geographical region

This hierarchical structures play a role in the statistical analysis
(chapter 8 – the assumption of independence)



Fixed and random effects

Fixed effects: the level of the explanatory variables are of interest
Fertilizer A, B or C

Typical question: Is yield higher when you use fertilizer A?

Random effects: random sample from a population

A sample of leaves from a plant

A sample of individuals in a population

A sample of populations in a geographical region

New type of question: How much of the observed variation is due to the randomly selected individual

Fixed and random effects

Sow rate	1	2	3	4
Variety 1	3	3	3	3
Variety 2	3	3	3	3

Sow rate and varieties are fixed effects

Sheep	1	2	3	4	5	6
Female	20	20	20			
Male				20	20	20

Sex is a fixed effect

Sheep is a random effect

Linear mixed model (both fixed and random effects)

Given a random sample (y_{ij}, x_{ij}, z_{ij})

y_{ij} is the observation j out of n_i observations from group i ;

$i = 1, 2, \dots, m$

$$y_{ij} = \alpha x_{ij} + \beta_i z_{ij} + \epsilon_{ij}$$

Fixed component:

x_{ij} is the fixed effect of observation y_{ij}

α is the fixed-effect parameter

ϵ_{ij} are normally distributed with mean 0 and variance σ^2

Random component:

z_{ij} are the random-effect regressors (group i)

β_i are the random-effect parameters for group i , that are assumed to be normally distributed with mean 0 and variance φ^2

Linear mixed model

$$y_{ij} = \alpha x_{ij} + \beta_i z_{ij} + \epsilon_{ij}$$

In a mixed model there are two sources of variation:

ϵ_{ij} are normally distributed with mean 0 and variance σ^2

= the sampling variance

β_i are the random-effect coefficients for group i , that are assumed to be normally distributed with mean 0 and variance φ^2

= the variation among groups

Fixed or random effects

Some factors may be considered to be either fixed or random effects

Fixed effects:

- large flexibility

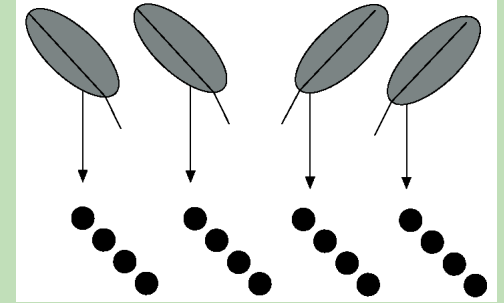
Random effects:

- assumption of normal distribution
- **possible to generalize outside the sampled entities**
- calculate variance components

Variance components - leaf example

4 leaves are taken as a random sample from a larger population of leaves

4 discs are cut from each leaves ($n_i = 4$)



GLM: Calcium concentration = Leaf

Source	DF	SS	MS	E(MS)
Leaf	3	0.281	0.093	$n_i \sigma_L^2 + \sigma_E^2$
Error	12	0.129	0.011	σ_E^2
Total	15			

$$4 \widehat{\sigma}_L^2 + \widehat{\sigma}_E^2 = 0.093 \text{ and } \widehat{\sigma}_E^2 = 0.011 \Rightarrow \widehat{\sigma}_L^2 = \frac{0.093 - 0.011}{4} = 0.021$$

$$\frac{0.021}{0.021 + 0.011} = 66\% \text{ of the total variance is due to the sampled leaves}$$

Variance components - leaf example - in R

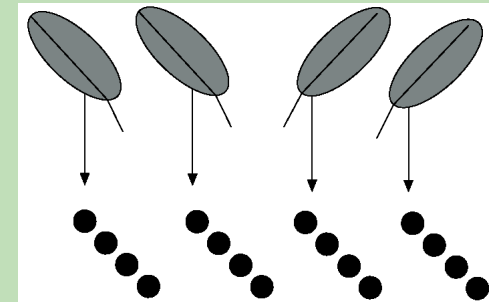
```
> library(nlme)
> mixedmodel <- lme(CACONC ~ 1, random= ~1|LEAF, data = Dataset)
> summary(mixedmodel)
Linear mixed-effects model fit by REML
Data: Dataset
      AIC      BIC    logLik
-10.14963 -8.025484 8.074817

Random effects:
Formula: ~1 | LEAF
      (Intercept)  Residual
StdDev:   0.1439292 0.1037225

Fixed effects: CACONC ~ 1
              Value Std.Error DF   t-value p-value
(Intercept) 0.55545 0.07649378 12  7.261374     0

Standardized Within-Group Residuals:
      Min          Q1          Med          Q3          Max
-1.44815697 -0.84333106  0.07807443  0.62609351  1.46056604

Number of Observations: 16
Number of Groups: 4
> VarCorr(mixedmodel)
LEAF = pdLogChol(1)
      Variance StdDev
(Intercept) 0.02071561 0.1439292
Residual    0.01075836 0.1037225
> |
```



Variance components - plant and leaf example - in R

```
> mixedmodel2 <- lme(CACONC2 ~ 1, random = ~ 1|PLANT2/LEAF2, data = Dataset2)
> summary(mixedmodel2)
Linear mixed-effects model fit by REML
Data: Dataset2
      AIC      BIC    logLik
23.38662 27.92859 -7.693308

Random effects:
Formula: ~1 | PLANT2
(Intercept)
StdDev:  0.3528517

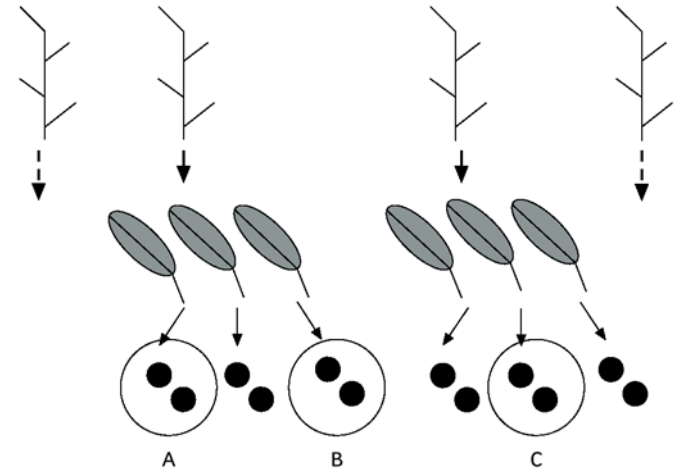
Formula: ~1 | LEAF2 %in% PLANT2
(Intercept) Residual
StdDev:  0.2403798 0.2053887

Fixed effects: CACONC2 ~ 1
              Value Std.Error DF   t-value p-value
(Intercept)  2.37825  0.1941622 12  12.24878      0

Standardized Within-Group Residuals:
      Min       Q1       Med       Q3       Max
-1.4245896 -0.4910138  0.1003876  0.6086579  1.8406515

Number of Observations: 24
Number of Groups:
      PLANT2 LEAF2 %in% PLANT2
         4      12

> VarCorr(mixedmodel2)
      Variance StdDev
PLANT2 = pdLogChol(1)
(Intercept) 0.12450430 0.3528517
LEAF2 = pdLogChol(1)
(Intercept) 0.05778247 0.2403798
Residual    0.04218451 0.2053887
```



Relative large variation
among the four plants
(increase the number of
plants in future experiments)

Sheep example (Exercise 8-1)

The lookup rate was observed in 3 male sheep and 3 female sheep in 20 time periods

Did SEX have an affect on the lookup rate?

```
> lm1 <- lm(LUPRATE ~ SEX + SHEEP, data=sheepdata)
> Anova(lm1, type="II")
Anova Table (Type II tests)

Response: LUPRATE
      Sum Sq  Df F value    Pr(>F)
SEX      0.36130   1 131.849 < 2.2e-16 ***
SHEEP    0.29477   4  26.893 1.015e-15 ***
Residuals 0.31239 114
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
```

LUPRATE=SEX+ SHEEP

But SEX is nested within SHEEP, so

LUPRATE=SEX+ SHEEP is a **wrong** model
SHEEP has 4 Df instead of 5 Df
114 residuals Df !!!

Instead make a mixed model

Sheep example - mixed model

```
> mixed2 <- lme(LUPRATE ~ SEX, random=~1|SHEEP,data=sheepdata)
> summary(mixed2)
Linear mixed-effects model fit by REML
Data: sheepdata
      AIC      BIC logLik
-331.9401 -320.8573 169.97

Random effects:
Formula: ~1 | SHEEP
      (Intercept)  Residual
StdDev:  0.05956224 0.05234731

Fixed effects: LUPRATE ~ SEX
              Value Std.Error DF t-value p-value
(Intercept) 0.17836817 0.03504603 114 5.089540 0.0000
SEX2         0.06653721 0.04956257   4 1.342489 0.2506
Correlation:
      (Intr)
SEX2 -0.707

Standardized within-Group Residuals:
      Min          Q1          Med          Q3          Max
-2.290672415 -0.741203992  0.000282715  0.654437380  2.166736038

Number of Observations: 120
Number of Groups: 6
> anova(mixed2)
      numDF denDF F-value p-value
(Intercept) 1 114 72.93477 <.0001
SEX         1  4  1.80228 0.2506
> varCorr(mixed2)
SHEEP = pdLogChol(1)
      Variance StdDev
(Intercept) 0.003547660 0.05956224
Residual    0.002740241 0.05234731
> |
```

Since it is a balanced design we get the same result ($P = 0.251$) as in Exercise 8-1 when using mean values

More information:
variance components

Pig growth example - mixed model

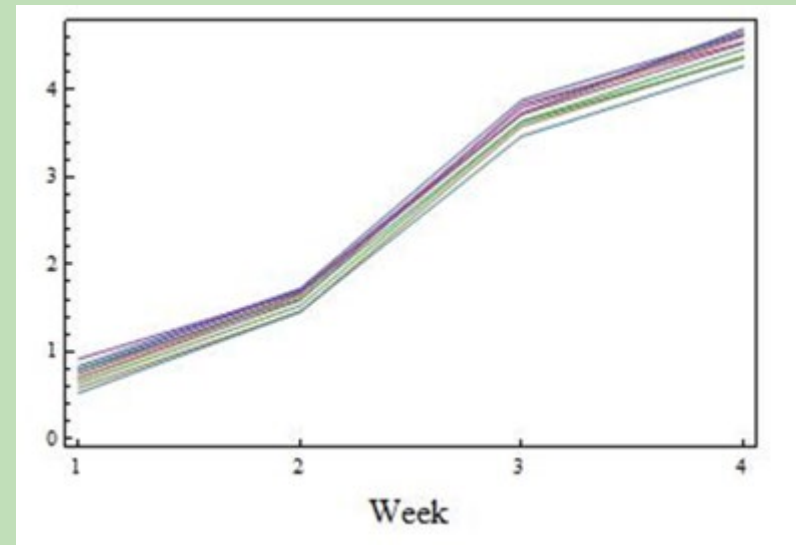
```
> mixedmodel4 <- lme(LGWT ~ DIET+WEEK, random = ~ 1|PIG,
> summary(mixedmodel4)
Linear mixed-effects model fit by REML
Data: Dataset4
      AIC      BIC    logLik
-61.85757 -50.97013 37.92878

Random effects:
Formula: ~1 | PIG
      (Intercept)  Residual
StdDev:  0.09716379 0.05070658

Fixed effects: LGWT ~ DIET + WEEK
              Value Std.Error DF   t-value p-value
(Intercept)  0.778050 0.04700591 27   16.55218 0.0000
DIET[T.2]    -0.113174 0.06350934  8   -1.78201 0.1126
WEEK[T.2]     0.885764 0.02267667 27   39.06058 0.0000
WEEK[T.3]     2.976736 0.02267667 27  131.26865 0.0000
WEEK[T.4]     3.786529 0.02267667 27  166.97905 0.0000
Correlation:
      (Intr) DIET[T WEEK[T.2 WEEK[T.3
DIET[T.2] -0.676
WEEK[T.2] -0.241  0.000
WEEK[T.3] -0.241  0.000  0.500
WEEK[T.4] -0.241  0.000  0.500  0.500

Standardized within-Group Residuals:
      Min      Q1      Med      Q3      Max
-1.2877507 -0.5054114 -0.0703765  0.2354116  2.0352143

Number of observations: 40
Number of Groups: 10
> anova(mixedmodel4)
              numDF denDF   F-value p-value
(Intercept)     1     27  6878.980 <.0001
DIET             1     8    3.176 0.1126
WEEK            3     27 12129.981 <.0001
```



WEEK is a categorical variable

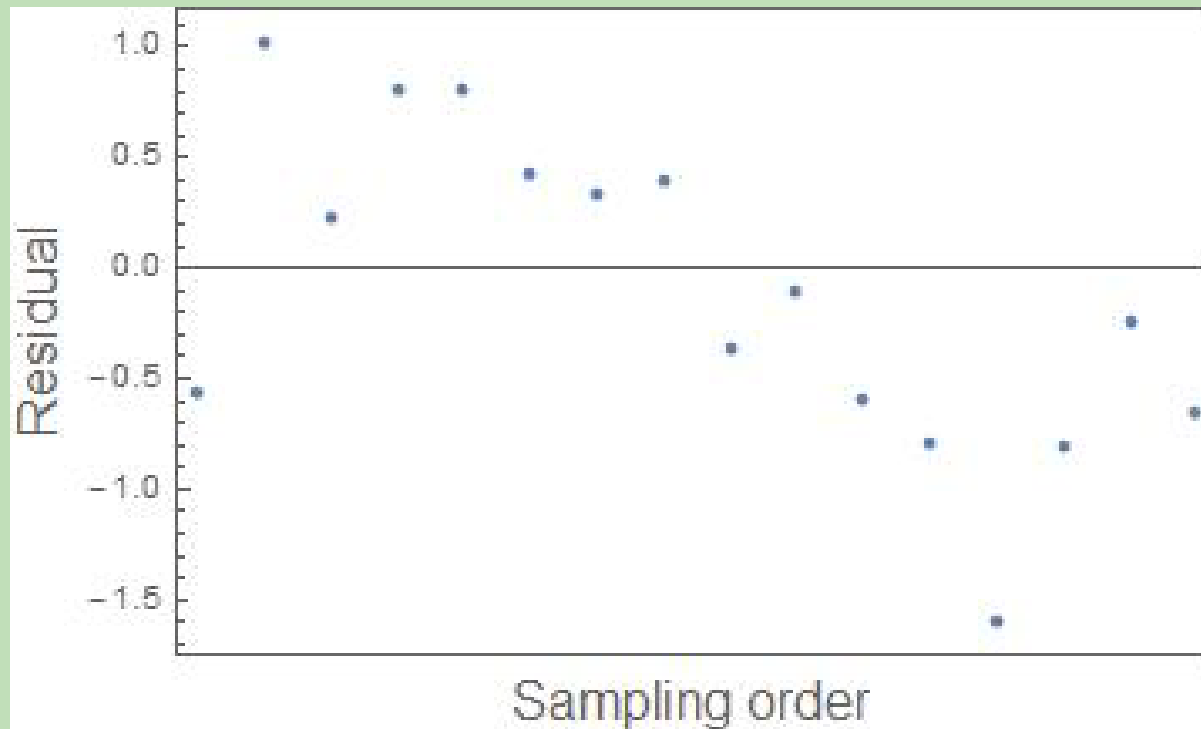
PIG is a random effect

No effect of diet

Time series data - repeated measures

Independent residuals?

Plot residuals as a function of their sampling order



Autoregressive model of order 1

Given a random sample (y_t, x_t) measured repeatedly at times t

$$y_t = f(x_t) + \varphi (y_{t-1} - f(x_{t-1})) + \epsilon_t$$

The parameter φ measure the degree of autocorrelation

Relevant R functions:

Calculate **the residuals**: `res=resid(model)`

Calculate the degree of autocorrelation: `ar(res)`

Update a mixed model: `update(model, correlation =
corCAR1(form = ~ t))`

Pig growth example - update mixed model with AR1

```
> mixedmodel4a=update(mixedmodel4, correlation = corCAR1(form = ~WEEK))
> summary(mixedmodel4a)
Linear mixed-effects model fit by REML
Data: Dataset4
      AIC      BIC    logLik
-60.40491 -47.96213 38.20245

Random effects:
Formula: ~1 | PIG
      (Intercept)  Residual
StdDev:  0.09654958 0.05405284

Correlation Structure: Continuous AR(1)
Formula: ~WEEK | PIG
Parameter estimate(s):
      Phi
0.2203406

Fixed effects: LGWT ~ DIET + WEEK
              value Std.Error DF   t-value p-value
(Intercept)  0.778226 0.04750482 27   16.38204 0.0000
DIET[T.2]    -0.113526 0.06426057  8   -1.76665 0.1153
WEEK[T.2]     0.885764 0.02134450 27   41.49847 0.0000
WEEK[T.3]     2.976736 0.02357906 27  126.24489 0.0000
WEEK[T.4]     3.786529 0.02404352 27  157.48646 0.0000

Correlation:
      (Intr) DIET[T WEEK[T.2 WEEK[T.3
DIET[T.2] -0.676
WEEK[T.2] -0.225 0.000
WEEK[T.3] -0.248 0.000 0.552
WEEK[T.4] -0.253 0.000 0.465 0.598

Standardized within-Group Residuals:
      Min      Q1      Med      Q3      Max
-1.21872124 -0.49403443 -0.02169107  0.23667080  1.95143905

Number of Observations: 40
Number of Groups: 10
> anova.lme(mixedmodel4a, type="m")
      numDF denDF F-value p-value
(Intercept)  1  27  268.371 <.0001
DIET         1   8   3.121 0.1153
WEEK        3  27 9832.470 <.0001
> anova(mixedmodel4, mixedmodel4a)
      Model df      AIC      BIC    logLik    Test    L.Ratio p-value
mixedmodel4  1  7 -61.85757 -50.97013 37.92878
mixedmodel4a  2  8 -60.40491 -47.96213 38.20245 1 vs 2 0.5473415 0.4594
```

Autoregressive parameter
(Phi)

No significant
autocorrelation

Still no effect of diet

Useful functions in Library(nlme)

```
mixedmodel1=lme(y ~ x, random= ~1 | ind)
```

```
summary(mixedmodel1)
```

```
anova.lme(mixedmodel1, type="m")
```

```
mixedmodel2=update(mixedmodel1, correlation = corCAR1(form =  
~time))
```

```
summary(mixedmodel2)
```

```
anova.lme(mixedmodel2, type="m")
```

```
anova(mixedmodel1,mixedmodel2)
```

```
plot(mixedmodel) – plot of residuals
```

```
plot(ranef(mixedmodel)) – plot of random effects
```

```
res=resid(mixedmodel) – calculate residuals
```

```
plot(res) – another way to plot residuals
```

Mixed models in R

There are several R packages that fit mixed models:

nlme (used in the shown examples)

lme4

Crawley (2013): The R book

R-INLA (Bayesian statistics)

