TMA4315 Generalized linear models H2018
Module 7: Linear mixed effects models, Week 1

Mette Langaas, Department of Mathematical Sciences, NTNU, with contibutions from Ingeborg G. Hem

01.11 and 08.11 [PL], 02.11 and 09.11 [IL]

(Latest changes: 08.11.2018 - typos).

(Warning: some changes may occur before the second week.)

Aim: Present methods for analysing correlated responses in a (normal/Gaussian) regression setting.

We will only consider *two-level models* and in particular focus on *random intercept and random slope models.*

Learning material

▶ Textbook: Fahrmeir et al (2013): Chapter 2.4, 7.1-7.3, 7.7. In greater detail: pages 349-354 (not "Alternative view on the random intercept model"), 356-365 (not 7.1.5 "Stochastic Covariates""), 368-377 (not"Bayesian Covariance Matrix" "), 379-380 (not"Testing Random Effects or Variance Parameters" ", only last part on page 383), 383 (middle), 401-409 (orange juice). Note: Bayesian solutions not on the reading list.

▶ Alternative readings: Zuur et al. (2009): "Mixed Effects Models and Extensions in Ecology with R", chapter 5 (pages 101-142). Available as free ebook from Springer for NTNU students. More explanations and less mathematics than Fahrmeir et al (2013), more focus on understanding. Link to ebook Chapter 5

▶ Classnotes 01.11.2018

▶ Classnotes 08.11.2018

Topics

First week

- ▶ correlated responses - when and why?
  - ▶ repeated measurements: clustered and longitudinal
  - ▶ example of clustered data from ecology: species richness
- ▶ notation
- ▶ random intercept models
  - ▶ intra class correlation (ICC)
- ▶ linear mixed effects models
  - ▶ measurement model and distributional assumptions
  - ▶ conditional and marginal formulation
- ▶ parameter estimation
  - ▶ with maximum likelihood for fixed effects
  - ▶ distribution of fixed parameter estimators

Jump to interactive (week 1)

Second week

- ▶ what did we do last week: beaches example
- ▶ parameter estimation (cont.)
  - ▶ (restricted) maximum likelihood for random effects
- ▶ predicting
  - ▶ random effects: method, formula, plots
  - ▶ random errors: two types of residuals
- ▶ random slope models
  - ▶ sleep study
  - ▶ interpretation of random effects
- ▶ hypothesis tests
- ▶ model selection
- ▶ fitting LMM with function `lmer` in package `lme4`
- ▶ what have we not covered?

Jump to interactive (week 2)

FIRST WEEK

Correlated responses

We may get correlated responses when we work with repeated measurements on a set of units. The units may be:

▶ subjects, patients, participants
▶ animal, plants
▶ families, towns, schools, classes, beaches

We will consider two types of repeated measurements: clustered and longitudinal

Clustered data

The data are nested in the sense that a lower level unit can only belong to one higher level unit (cluster), and there is in general no natural ordering of the units within each cluster.

Two-level clustered - examples

- ▶ patients in hospitals
- ▶ siblings in families
- ▶ pupils in schools

The data for each individual are observed at multiple points in time

Two-level longitudinal data - examples

▶ patients with drug A and drug B monitored over time

▶ metabolic rate for patients at fasting, then 15, 45, 75 and 135 minutes after a heavy meal

Q: Why not only analyse using linear models?

A: Assuming independent responses when they are correlated does not give the correct estimate for the standard error of the estimated parameters of interest.

Example from ecology: beaches and species

This example is taken from Zuur et al. (2009, chapter 5, pages 101-142), and data are referred to as RIKZ because they were collected by a Dutch institute with that name.

Data were collected on nine different beaches in the Netherlands with the aim to investigate if there is a relationship between the

▶ species richness (number of species observed) and
▶ NAP: the height of the sampling station compared to mean tidal level.

Data: 45 observations, taken at 9 beaches:

▶ beach: the beach that the samples were taken, for each beach 5 different samples were taken.
▶ exposure: an index describing how much the beach is exposed to the sea was measured, but we will not be used now.

We want to use the richness of species as the response in a regression model.

This is a count, so we should use *Poisson regression*, but to make things simpler we assume that the counts are such that we instead can assume a *normal distribution*.

(in the next module on generalized linear mixed effects models we will use the Poisson distribution).

```
# library('AED')
RIKZ <- read.csv("http://faculty.concordia.ca/pperesne/BIOL_422_680/RIK
summary(RIKZ)
```

```
##     Sample        Richness        Exposure         NAP              Min
## Min.   : 1    Min.   : 0.000   Min.   :1.000   Min.   :-1.3360   Min
## 1st Qu.:12    1st Qu.: 3.000   1st Qu.:1.000   1st Qu.:-0.3750   1st
## Median :23    Median : 4.000   Median :1.000   Median : 0.1670   Med
## Mean   :23    Mean   : 5.689   Mean   :1.444   Mean   : 0.3477   Mea
## 3rd Qu.:34    3rd Qu.: 8.000   3rd Qu.:2.000   3rd Qu.: 1.1170   3rd
## Max.   :45    Max.   :22.000   Max.   :2.000   Max.   : 2.2550   Max
```

We first consider models with reponse `Richness` and one covariate
`NAP`, and analyse data from each beach separately.
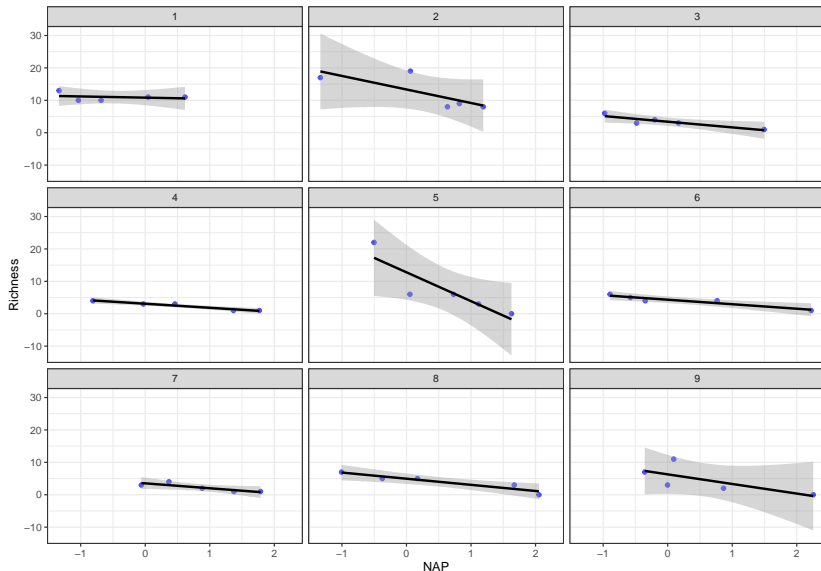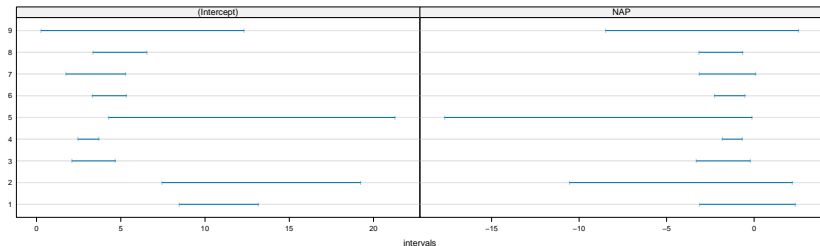
```
##                              1          2         3         4          5        4.32
## (Intercept) 10.8218944 13.345694  3.400702  3.087716 12.782828   4.32
## NAP         -0.3718279 -4.175271 -1.755353 -1.248577 -8.900178  -1.38
##                              7          8         9
## (Intercept)  3.520626  4.951455  6.295053
## NAP         -1.517613 -1.893066 -2.967530
```

The intercepts differ but the slopes are not that different.

Q: What if we now want to combine the regression models for the nine beaches into one model (for the population of beaches), so we can answer the question about relationship between species richness and NAP on the population level. What can we do then?

  1. Use all 45 observations in a regression - with a common intercept and linear effect in NAP.

Problem: violation of assumption of independent observations lead to wrong estimates for variances of parameter estimates.

  2. Add beach as covariate to regression - then we estimate one regression coefficient for each beach (intercepts) in addition to the linear effect in NAP.

Problem: why do we want to add 8 extra parameters to estimate (why 8?) values for the 9 beaches? Loss of power and what would we use the beach estimates for?

3. Add beach as a *random* covariate to the regression: this is called *random intercept models*.

Problem: new stuff - slightly complicated. We do this because beaches not of interest in themselves, only random sample from population of beaches, and therefore we only need to account for beaches, not estimate separate parameters.

Solution 1: all observations together - standard errors not correct

```
fitall = lm(Richness ~ NAP, data = RIKZ)
summary(fitall)
```

```
##
## Call:
## lm(formula = Richness ~ NAP, data = RIKZ)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.0675 -2.7607 -0.8029  1.3534 13.8723
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.6857     0.6578  10.164 5.25e-13 ***
## NAP          -2.8669     0.6307  -4.545 4.42e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.16 on 43 degrees of freedom
## Multiple R-squared:  0.3245, Adjusted R-squared:  0.3088
## F-statistic: 20.66 on 1 and 43 DF,  p-value: 4.418e-05
```

Solution 2: fixed effects for each beach - many estimates not so much of interest when population is in focus.

```
RIKZ$beachfactor = as.factor(RIKZ$Beach)
fitbeach = lm(Richness ~ NAP + beachfactor, data = RIKZ, contrasts = li
summary(fitbeach)
```

```
##
## Call:
## lm(formula = Richness ~ NAP + beachfactor, data = RIKZ, contrasts =
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.8518 -1.5188 -0.1376  0.7905 11.8384
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    6.5556     0.4885  13.421 2.30e-15 ***
## NAP           -2.4928     0.5023  -4.963 1.79e-05 ***
## beachfactor1   3.2503     1.3554   2.398   0.0220 *
## beachfactor2   6.3284     1.2908   4.903 2.15e-05 ***
## beachfactor3  -3.1546     1.3020  -2.423   0.0207 *
## beachfactor4  -2.7826     1.2943  -2.150   0.0386 *
## beachfactor5   2.3520     1.2967   1.814   0.0783 .
## beachfactor6  -1.9728     1.2915  -1.528   0.1356
```

Notation

We have n_i repeated observations ($j = 1, \ldots, n_i$) from each of $i = 1, \ldots, m$ clusters (or individuals).

▶ Responses: $Y_{i1}, Y_{i2}, \ldots, Y_{in_i}$ (e.g. species richness at beach i sample j)

▶ Covariates: $\mathbf{x}_{i1}, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i}$ (e.g. NAP for beach i sample j)

The covariates $\mathbf{x}_{ij}$ are $p \times 1$ vectors (as before, k covariates and one intercept so $p = k + 1$).

Random intercept models

Model

Simple linear regression

We start with a simple linear regression (one fixed effect).
First, only one covariate (in addition to the intercept), observed for cluster (beach) i on occation j we have x_{ij}

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \varepsilon_{ij} \text{ where } \varepsilon_{ij} \text{ i.i.d. } N(0, \sigma^2)$$

but, we know that Y_{i1} and Y_{i2} are observed for the same cluster and should not be independent (i.e. from same beach), to fix that we insert a *random intercept*.

Cluster-specific parameters γ_{0i}

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \gamma_{0i} + \varepsilon_{ij} \text{ where } \varepsilon_{ij} \text{i.i.d.} N(0, \sigma^2)$$

- ▶ β_0: population intercept (fixed)
- ▶ γ_{0i}: deviation (for members of cluster i) from the population intercept β_0 - not a parameter but a random variable!
- ▶ $\beta_0 + \gamma_{0i}$: random intercept for cluster i
- ▶ β_1: population slope (fixed), common to all clusters

The clusters are assumed to be random samples from a large population, and for the cluster deviation intercept we assume

$$\gamma_{0i} \sim N(0, \tau_0^2)$$

and that the γ_{0i}s and the ε_{ij}s are mutually independent random variables. So, we have now two error terms.

Beach-example: Parameter estimates

(will talk about parameter estimation with ML and REML later)

Q: Try to identify $\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2, \hat{\tau}_0^2$ in the print-out.

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Richness ~ NAP + (1 | Beach)
##    Data: RIKZ
##
## REML criterion at convergence: 239.5
##
## Scaled residuals:
##    Min     1Q Median     3Q    Max
## -1.4227 -0.4848 -0.1576  0.2519  3.9794
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Beach    (Intercept) 8.668    2.944
##  Residual             9.362    3.060
## Number of obs: 45, groups:  Beach, 9
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   6.5819     1.0958   6.007
## NAP          -2.5684     0.4947  -5.192
##
## Correlation of Fixed Effects:
##     (Intr)
## NAP -0.157
```

A: $\hat{\beta}_0 = 6.5818929$, $\hat{\beta}_1 = -2.5683996$, $\hat{\sigma}^2 = 9.3621916$, and $\hat{\tau}_0^2 = 8.6675183$

Intra class correlation (ICC)

The *conditional distribution* of Y_{ij} given the value of γ_{0i} (=regarding γ_{0i} as known) is

$$Y_{ij} \mid \gamma_{0i} \sim N(\beta_0 + \beta_1 x_{ij} + \gamma_{0i}, \sigma^2)$$

One motivation for inserting this new random intercept was to make sure that observations from the same cluster are dependent, but between clusters are independent. This means that we need to look at $\text{Cov}(Y_{ij}, Y_{kl})$ when $i = k$ and when $i \neq k$. To do that we need the (joint) *marginal distribution* of the responses.

What is the marginal distribution for Y_{ij}?

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \gamma_{0i} + \varepsilon_{ij}$$

▶ marginal over the random effects
▶ what is the distribution of Y_{ij} for a new beach?
▶ Treat the model as $Y_{ij} = \beta_0 + \beta_1 x_{ij} + \varepsilon_{ij}^*$, where $\varepsilon_{ij}^* = \gamma_{0i} + \varepsilon_{ij}$ are not iid

What is the marginal distribution for Y_{ij}?

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \gamma_{0i} + \varepsilon_{ij}$$

$$Y_{ij} \sim N(\beta_0 + \beta_1 x_{ij}, \sigma^2 + \tau_0^2)$$

We will consider Y_{ij}, and Y_{kl}, where

$$Y_{kl} = \beta_0 + \beta_1 x_{kl} + \gamma_{0k} + \varepsilon_{kl} \sim N(\beta_0 + \beta_1 x_{kl}, \sigma^2 + \tau_0^2)$$

Now to the covariance between Y_{ij} and Y_{kl}.

$$\mathsf{Cov}(Y_{ij}, Y_{kl}) = \mathsf{E}[(Y_{ij} - \mu_{ij})(Y_{kl} - \mu_{kl})]$$

And

$$\begin{aligned} Y_{ij} - \mu_{ij} &= (\beta_0 + \beta_1 x_{ij} + \gamma_{0i} + \varepsilon_{ij}) - (\beta_0 + \beta_1 x_{ij}) \\ &= \gamma_{0i} + \varepsilon_{ij} \end{aligned}$$

Note that we assume independence, so $Cov(\gamma_{0i}, \gamma_{0k}) = 0$ etc.

$$\mathsf{Cov}(Y_{ij}, Y_{kl}) = \left\{ \begin{array}{ll} \tau_0^2 + \sigma^2 = \mathsf{Var}(Y_{ij}) & \text{for } i = k, j = l \\ \tau_0^2 & \text{for } i = k, j \neq l \\ 0 & \text{for } i \neq k, j \neq l \end{array} \right\}$$

$\oplus$ derivations on the board in class

If we put this into a covariance matrix for the vector of responses for cluster i this type of structure is called *compound symmetry*.

$\oplus$ write this out on the board in class

$$\text{Cov}(\mathbf{Y}_i) = \tau_0^2 \mathbf{1}\mathbf{1}^T + \sigma^2 \mathbf{I}$$

where as before $\mathbf{1}$ is a $n_i \times 1$ vector of 1s and $\mathbf{I}$ is a $n_i \times n_i$ identity matrix.

The correlation between Y_{ij} and Y_{il} (two observations in the same cluster that is - same beach) is called the *within subject* or *within cluster* correlation coefficient, and is for our random intercept model

$$\text{Corr}(Y_{ij}, Y_{il}) = \frac{\text{Cov}(Y_{ij}, Y_{il})}{\sqrt{\text{Var}(Y_{ij})\text{Var}(Y_{il})}} = \frac{\tau_0^2}{\tau_0^2 + \sigma^2} \text{ for } j \neq l$$

Inserted parameter estimates this is called the *intra class correlation (ICC)* for the random intercept model.

Beach-example: ICC

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Richness ~ NAP + (1 | Beach)
##    Data: RIKZ
##
## REML criterion at convergence: 239.5
##
## Scaled residuals:
##     Min     1Q  Median     3Q     Max
## -1.4227 -0.4848 -0.1576  0.2519  3.9794
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Beach    (Intercept) 8.668    2.944
##  Residual             9.362    3.060
## Number of obs: 45, groups:  Beach, 9
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   6.5819     1.0958   6.007
## NAP          -2.5684     0.4947  -5.192
##
## Correlation of Fixed Effects:
##     (Intr)
## NAP -0.157
```

Q: What is the ICC for our fit? Hint:
$\mathsf{Corr}(Y_{ij}, Y_{il}) = \frac{\tau_0^2}{\tau_0^2 + \sigma^2}$ for $j \neq l$.

A: $8.688/(8.688+9.362)=0.48$.

Summing up - so far

We have now looked at models with a random intercept, which is a special case of linear mixed models (LMM). We have now see that we have used two components in our model:

▶ fixed effects - like we have used so far in our GLM-course.
  This can be gender, age, time, experimental condition.
▶ random effects - used to model correlated responses.
  Remember the correlation (ICC) in the previous example.

But, what if we also need the different clusters to have different slopes? The generalization of the random intercept model is the *random slope* model which can be written as:

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \gamma_{0i} + \gamma_{1i} x_{ij} + \varepsilon_{ij}$$

and where the new parts are

- ▶ β_1: population slope (fixed), common to all clusters
- ▶ γ_{1i}: deviation (for members of cluster i) from the population slope β_0 - not a parameter but a random variable!
- ▶ $\beta_1 + \gamma_{1i}$: random slope for cluster i.

More about this model in week 2.

Linear mixed effects model (LMM)

We now define the LMM from a measurement model and distributional assumption for clusters $i = 1, \ldots, m$.

Measurement model
for the ith cluster:

$$\mathbf{Y}_i = \mathbf{X}_i \beta + \mathbf{U}_i \gamma_i + \varepsilon_i$$

▶ $\mathbf{Y}_i$: $n_i \times 1$ random vector of responses
▶ $\mathbf{X}_i$: $n_i \times p$ design matrix
▶ $\mathbf{U}_i$: $n_i \times (q + 1)$ design matrix for random effects
▶ β: $p \times 1$ vector of fixed coefficients (common for all clusters)
▶ γ_i: $(q + 1) \times 1$ random vector
▶ ε_i: $n_i \times 1$ random vector
p fixed effect covariates, q random effect covariates

Distributional assumptions

for the ith cluster:

$$\gamma_i \sim N(\mathbf{0}, \mathbf{Q})$$
$$\varepsilon_i \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$$

All elements of $\gamma_1, \gamma_2, \dots, \gamma_m$ and $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m$ are mutually independent. The dimension of $\mathbf{I}$ is $n_i \times n_i$, and $\mathbf{Q}$ is $(q+1) \times (q+1)$.

Remark: one possible generalization is to assume $\varepsilon_i \sim N(\mathbf{0}, \Sigma)$ where Σ is general, but we will not look into that in our course. This can be needed for example if time series structure (like AR1) is present.

Questions

$$\mathbf{Y}_i = \mathbf{X}_i\beta + \mathbf{U}_i\gamma_i + \varepsilon_i$$
$$\gamma_i \sim N(\mathbf{0}, \mathbf{Q})$$
$$\varepsilon_i \sim N(\mathbf{0}, \sigma^2\mathbf{I})$$

Q:
▶ What is $\mathbf{U}_i$, γ_i and $\mathbf{Q}$ for the random intercept model?
▶ General: what is the marginal distribution of $\mathbf{Y}_i$?

A:

- We had $\mathbf{U}_i = \mathbf{1}$ ($n_i \times 1$) and scalar $\gamma_i = \gamma_{0i}$, and $\mathbf{Q} = \tau_0^2$.
- $\mathbf{Y}_i \sim N(\mathbf{X}_i\beta, \mathbf{U}_i\mathbf{Q}\mathbf{U}_i^T + \sigma^2\mathbf{I})$.

Marginal and Conditional Formulatios

We can write a conditional formulation if we condition on γ_i, or a margunal if we do not (in which case we marginalise over the distribution of γ_i

Conditional formulation

Conditional Gaussian model for the response $\mathbf{Y}_i$ given the random effect γ_i:

$$\mathbf{Y}_i \mid \gamma_i \sim N(\mathbf{X}_i \beta + \mathbf{U}_i \gamma_i, \sigma^2 \mathbf{I})$$

Marginal Gaussian model

for the response for cluster 8, $\mathbf{Y}_i$ (Laird and Ware (1982) formulation)

$$\mathbf{Y}_i = \mathbf{X}_i\beta + \mathbf{U}_i\gamma_i + \varepsilon_i = \mathbf{X}_i\beta + \varepsilon_i^*$$
$$\varepsilon_i^* = \mathbf{U}_i\gamma_i + \varepsilon_i$$
$$\mathsf{E}(\varepsilon_i^*) = \mathbf{0}$$
$$\mathbf{V}_i = \mathsf{Cov}(\varepsilon_i^*) = \mathsf{Cov}(\mathbf{U}_i\gamma_i) + \mathsf{Cov}(\varepsilon_i) = \mathbf{U}_i\mathbf{Q}\mathbf{U}_i^T + \sigma^2\mathbf{I}$$
$$\varepsilon_i^* \sim N(\mathbf{0}, \mathbf{V}_i)$$

which gives

$$\mathbf{Y}_i \sim N(\mu_i = \mathbf{X}_i\beta, \mathbf{V}_i = \sigma^2\mathbf{I} + \mathbf{U}_i\mathbf{Q}\mathbf{U}_i^T)$$

Global model

From the cluster specific model:

$$\mathbf{Y}_i = \mathbf{X}_i\beta + \mathbf{U}_i\gamma_i + \varepsilon_i$$

into the global model for all clusters:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{U}\gamma + \varepsilon$$

where

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_m \end{pmatrix}, \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_m \end{pmatrix}, \mathbf{U} = \begin{pmatrix} \mathbf{U}_1 & \mathbf{0} & ... & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_2 & ... & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & ... & \mathbf{U}_m \end{pmatrix}, \gamma = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_m \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{pmatrix}$$

Let $N = \sum_{i=1}^{m} n_i$, then dimensions are:

▶ $\mathbf{Y}, \varepsilon$: $N \times 1$
▶ $\mathbf{X}$: $N \times p$
▶ β: $p \times 1$
▶ $\mathbf{U}$: $N \times m(q+1)$
▶ γ: $m(q+1) \times 1$

Conditional Gaussian model
for the response $\mathbf{Y}$ given the random effect γ:

$$\mathbf{Y} \mid \gamma \sim N(\mathbf{X}\beta + \mathbf{U}\gamma, \sigma^2\mathbf{I})$$

Now $\mathbf{I}$ is $N \times N$ where $N = \sum_{i=1}^{m} n_i$.

Marginal Gaussian model

for the response $\mathbf{Y}$

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{U}\gamma + \varepsilon = \mathbf{X}\beta + \varepsilon^*$$
$$\varepsilon^* = \mathbf{U}\gamma + \varepsilon$$
$$\mathbf{V} = \mathsf{Cov}(\varepsilon^*) = \mathsf{Cov}(\varepsilon) + \mathsf{Cov}(\mathbf{U}\gamma) = \sigma^2\mathbf{I} + \mathbf{U}\mathbf{G}\mathbf{U}^T$$
$$\varepsilon^* \sim N(\mathbf{0}, \mathbf{V})$$

Here $\mathbf{G}$ is a $m(q+1)$ block-diagonal matrix with $\mathbf{Q}$ m times on the diagonal (see below). This gives

$$\mathbf{Y} \sim N(\mathbf{X}\beta, \mathbf{V} = \sigma^2\mathbf{I} + \mathbf{U}\mathbf{G}\mathbf{U}^T)$$

$$\mathbf{G} = \begin{pmatrix} \mathbf{Q} & \mathbf{0} & ... & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} & ... & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & ... & \mathbf{Q} \end{pmatrix}$$

Parameter estimation

- ▶ Fixed effects β: estimated using maximum likelihood
- ▶ Random effects parameters σ^2 and $\mathbf{Q}$ (in $\mathbf{V}$): estimated using restricted maximum likelihood (REML).

For the random effects γ_i and ε_i we also provide predictions

- ▶ Predicted values for the random effects γ_i using *best linear unbiased predictors* (BLUP).
- ▶ Prediction values for the random effects ε_i are our *residuals*. Two types of residuals possible "response minus fixed effects"", or "response minus fixed and predicted random effects".

Parameter estimation with maximum likelihood for fixed effects

Cluster specific model:

$$\mathbf{Y}_i \sim N(\mu_i = \mathbf{X}_i\beta, \mathbf{V}_i = \sigma^2\mathbf{I} + \mathbf{U}_i\mathbf{Q}\mathbf{U}_i^T)$$

Global model:

$$\mathbf{Y} \sim N(\mathbf{X}\beta, \mathbf{V} = \sigma^2\mathbf{I} + \mathbf{U}\mathbf{G}\mathbf{U}^T)$$

The log-likelihood function is then ($\pm$ some constants)

$$l(\beta, \mathbf{V}) = -\frac{1}{2}\ln|\mathbf{V}| - \frac{1}{2}(\mathbf{y} - \mathbf{X}\beta)^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta)$$

We assume that the parameters in $\mathbf{V}$ are **known**, and transform our problem with $\mathbf{V}^{-1/2}$ to get the standard multiple linear regression model.

In TMA4267 Exam 2014 Problem 4, the weighted least squares estimator was derived from a similar situation. Now we have

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon^*$$

where

$$\varepsilon^* \sim N(\mathbf{0}, \mathbf{V})$$

We then define $\mathbf{V}^{-1/2}$ based on eigenvalue-vector decomposition of $\mathbf{V}$, and premultiply the above equation to get:

$$\mathbf{V}^{-1/2}\mathbf{Y} = \mathbf{V}^{-1/2}\mathbf{X}\beta + \mathbf{V}^{-1/2}\varepsilon^*$$

$$\mathbf{Y}^\bullet = \mathbf{X}^\bullet\beta + \varepsilon^\bullet$$

where $\varepsilon^\bullet \sim N(\mathbf{0}, \mathbf{I})$. We then know that the ML estimator for β is given as

$$\widehat{\beta} = (\mathbf{X}^{\bullet T}\mathbf{X}^\bullet)^{-1}\mathbf{X}^{\bullet T}\mathbf{Y}^\bullet$$

Inserting then for $\mathbf{X}^{\bullet T}$ and $\mathbf{Y}^{\bullet}$ leads to the weighted least squares solution for β

$$\hat{\beta} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y}$$

Since we have independence between clusters, we had block-diagonal $\mathbf{V}$ and then

$$\hat{\beta} = (\sum_{i=1}^{m} \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i)^{-1} \sum_{i=1}^{m} \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{Y}_i$$

But, we really don't know $\mathbf{V}$ so an estimate must be inserted (week 2 with REML).

Beach-example: parameter estimation for fixed effects

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Richness ~ NAP + (1 | Beach)
##    Data: RIKZ
##
## REML criterion at convergence: 239.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.4227 -0.4848 -0.1576  0.2519  3.9794
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Beach    (Intercept) 8.668    2.944
##  Residual             9.362    3.060
## Number of obs: 45, groups: Beach, 9
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   6.5819     1.0958   6.007
## NAP          -2.5684     0.4947  -5.192
##
## Correlation of Fixed Effects:
##     (Intr)
## NAP -0.157
```

Q: Where is the parameter estimat for effect of NAP, with standard deviation? What is the interpretation of "Correlation of Fixed Effects"?

A:"Fixed effects, NAP, Estimate and Std. Error". Correlation of fixed effects is the off-diagonal elements of $\mathrm{Cov}(\hat{\beta})$ which here is only $\mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1)$.

Properties of parameters estimators

Since $\hat{\beta}$ is a linear function of $\mathbf{Y}_i$, it has a multivariate normal distribution - let us write with

$$\hat{\beta} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y} = \mathbf{A} \mathbf{Y}$$

Mean:

$$\mathsf{E}(\hat{\beta}) = \mathbf{A}\mathsf{E}(\mathbf{Y}) = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}\beta = \beta$$

so $\hat{\beta}$ is unbiased.

Variance-covariance matrix:

$$\mathsf{Cov}(\widehat{\beta} = \mathbf{A}\mathsf{Cov}(\mathbf{Y})\mathbf{A}^T = \mathbf{A}\mathbf{V}\mathbf{A}^T$$

$$= (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{V}\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}$$

Since $\mathbf{V}$ is block diagonal this can be written as a sum with the clusters:

$$\mathsf{Cov}(\widehat{\beta}) = (\sum_{i=1}^{m} \mathbf{X}_i^T\mathbf{V}_i^{-1}\mathbf{X}_i)^{-1}$$

Also, if we insert $\hat{\mathbf{V}}$ as an estimate for $\mathbf{V}$ then

$$\hat{\beta} = (\sum_{i=1}^{m} \mathbf{X}_i^T \hat{\mathbf{V}}_i^{-1} \mathbf{X}_i)^{-1} \sum_{i=1}^{m} \mathbf{X}_i^T \hat{\mathbf{V}}_i^{-1} \mathbf{Y}_i$$

is asymptotic multivariate normal (under regularity conditions).

Remark: this is only asymptotically, so we need large samples for this to hold. And, we do not arrive at a t-distribution here - however approximations for t (and F) exists, but the problem is the number of degrees of freedom (in particular when we have time varying fixed effects).

In this course we will only use the asymptotic normality for confidence intervals (and when a Wald type test is desired). We will consider hypothesis testing with likelihood ratio test under "Model selection" in the end of this module.

Beach-example: confidence interval for fixed effects

```
##                   2.5 %     97.5 %
## .sig01         1.484204   5.100814
## .sigma         2.435429   3.877572
## (Intercept)    4.324560   8.824909
## NAP           -3.566901  -1.599779
```

Q: Interpret! First the last two rows. The first two we have not covered yet - since we don't know how to estimate the parameters in $\mathbf{V}$ but what do you think this is?

A: Estimate $\pm$ 1.96 times standard error of estimate.