Plan

Models with fixed effects

1. Modelisation preambles and linear (fixed effects) models (LM)
2. Generalized linear (fixed effects) models (GLM)

Linear Mixed Models (LMM)

1. Definition and notations
2. Estimation and algorithms
3. Models selection

Generalized Linear Mixed Models (GLMM)

1. Definition and notations
2. Estimation and algorithms
3. Model selection

Longitudinal, overdispersion and non linear mixed models examples

Experimentation using R

Contents
Basics for model construction

- a set of **observations** for:
  - $y$ a **response** variable (i.e. **dependent** variable)
  - $x_1, x_2, \ldots$ **explanatory** variables

with matrix notation:

$$
y = \begin{pmatrix}
y_1 \\
\vdots \\
y_n
\end{pmatrix}
\quad X = \begin{pmatrix}
x_{1,1} & \cdots & x_{1,K} \\
\vdots & \ddots & \vdots \\
x_{n,1} & \cdots & x_{n,K}
\end{pmatrix}
$$

* $y$ is a realization of the $n \times 1$ response random vector $Y$
* $X$ is the observation of the $n \times K$ design matrix
A parametric density family is a collection of density functions $f_{\theta}(y)$, described by a finite-dimensional parameter $\theta$. The set of all allowable values for the parameter is denoted $\Theta \subseteq \mathbb{R}^K$, and the model itself is written as

$$\mathcal{F} = \{f_{\theta}(y) \mid \theta \in \Theta\}$$

### Example

- **Gaussian parametric model**
  $$\mathcal{F} = \left\{ f_{\mu,\sigma}(y) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \mid \mu \in \mathbb{R}, \sigma > 0 \right\}$$

- **Exponential parametric model**
  $$\mathcal{F} = \{ f_{\lambda}(y) = \lambda e^{-\lambda y} \mid \lambda > 0 \}$$

- **Binomial parametric model**
  $$\mathcal{F} = \{ f_p(y) = \binom{n}{y} p^y (1-p)^{n-y} \mid 0 \leq p \leq 1 \}$$
a link between $Y$ and the $X$’s

for 'linear models', this link is assumed between:
- the mean (i.e. expected value)
  $$\mu = \mathbb{E}(Y)$$
- a linear combination of the $X$’s through the linear predictor
  $$\eta = X\beta$$

by
$$g(\mu) = \eta$$

with $g(\mu) = \mu$ for the normality assumption!

Remark

Linearity is expressed in terms of linear combination of parameters

$$\eta = \mu + \beta_1 x + \beta_2 x^2$$

takes part of a linear model but relation between $y$ and $x$ is no more linear.
About explanatory variables

2 types of explanatory variables:

- **Factors**
  - Interest is in attributing variability in $y$ to various categories of the factor

  Example: patients classified by gender (M/F) and age group (A/B/C)

  $\eta_{ij} = \mu + \alpha_i + \beta_j \quad i = 1, 2 \quad j = 1, 2, 3$

  - Parameter values give the impact of factor’s levels on the response variable

  Factors may be **crossed** or **nested**
  Factors may have **main effect** and **interaction effect**
regressors

→ interest is in attributing variability in $y$ to changes in values of a continuous covariable

Example: changes due to weight $x$

$$\eta = \beta_0 + \beta_1 x$$

→ parameter values give the impact of an increase in $x$ on the response variable
Fixed vs Random effects

Example

4 loaves of bread are taken from each of 6 batches of bread baked at 3 different temperatures

- interest of course in each particular baking temperature used
- no interest in each batch which are very depending on the circumstances
- batches effect can be viewed as a sample of a random batch effect (levels are chosen at random from an infinite set of batches levels)
- interest in estimating the variance of the batch effect as a source of random variation in the data
- **fixed effect** (factor) is defined with a finite set of levels and when interest lies in the estimation of each particular level effect.

- **random effect** (factor) is defined with an infinite set of levels (with only a finite subset present in the data collection) and when interest lies more in the variance induced by these levels than in the estimation of the levels themselves.

Consequence: data collected within each level of the random effect factor are linked to a same realization of a random variable. This introduce dependency between this data.

**Example**

Data collected on the 4 loaves in each batch share something linked to the batch itself!
Example (Placebo and drug)

Clinical trial of treating epileptics with a drug. Patients randomly allocated to either drug or placebo. Response: number of seizures for patient \( k \) receiving treatment \( i \)

\[
\mathbb{E}(Y_{ik}) = \mu_i = \mu + \alpha_i
\]

The 2 treatments are the only 2 being used and there is no thought for any other treatments. Each level is of interest.

\( \rightarrow \) Treatment is a **fixed effect** factor.
Example (cont.)

Suppose the clinical trial is composed of repetition at 20 different clinics in the city.
Response: number of seizures for patient $k$ receiving treatment $i$ in clinics $j$

$$E(Y_{ijk}) = \mu_{ij} = \mu + \alpha_i + b_j$$

Clinics can be viewed as a random sample of clinics. Inference will be made on the population of clinics effects and in particular the variance (magnitude of the variations among clinics).

$\rightarrow$ Clinics is a **random effect** factor.
Example (cont.)

Assume the procedure is very specialized so that the trial is only conducted in a very few number of referral hospitals. We can no longer consider the clinics selected as a sample from a larger group of clinics. We want to make inference only to the clinics in the study.

 Clinics is a **fixed effect** factor.

The situation is determining whether clinics effect is to be considered fixed or random.
Help for decision random or fixed

Are the levels of the factor going to be considered as a random sample from a population of effects?

Is there enough information about a factor to decide that the levels of it in the data are like a random sample?

Does each level of the factor have a constant effect we want to estimate?

Do we want to identify the levels variation as a source of variability?
Mathematical notation for fixed effects

$$\mathbb{E}(Y_{ik}) = \mu + \alpha_i$$

Mathematical notation and properties for random effects

$$\mathbb{E}(Y_{ik}|a_i) = \mu + a_i$$

when $a_i$ are the levels of a random effect; $a_i$’s are treated as random variables.

2 usual assumptions:

- $a_i$’s are independently and identically distributed
- $a_i$’s have zero mean and all the same variance $\sigma_a^2$.

$$a_i \sim i.i.d.(0, \sigma_a^2)$$
Is height depending on sex and weight?

Observations

\((y, s, w)\) the height, the sex and the weight of 100 individuals sampled from a population.
Model construction

- \( y_{ij} \) is a realization for individual \( i \) of sex \( j \) of \( Y_{ij} \) from a parametric density family \( \mathcal{F} \)
  
  We assume:
  - \( \mathcal{F} \) is the gaussian parametric family,
  - \( Y_{ij} \) are independent
  - \( \mathbb{E}(Y_{ij}) = \mu + \alpha_j + \beta w_{ij} \) where \( \mu \) is the intercept, \( \alpha_j \) the sex fixed effects (male and female), \( \beta \) fixed effect of weight.
  - \( \mathbb{V}(Y_{ij}) = \sigma^2, \forall i, j \)

  this implies

1. \( Y_{ij} \sim \mathcal{N}(\mu + \alpha_j + \beta w_{ij}, \sigma^2) \)
2. \( \mu, \alpha_j, \beta \) and \( \sigma^2 \) are the unknown parameters.
Is number of foliar blotches depending on clone?

Observations

(y, c) number of foliar blotches and clone identifier sampled from a population. 25 individuals for each of 4 clones have been measured.
Is number of foliar blotches depending on clone?

Model construction

- \( y_{ij} \) is a realization for individual \( i \) of clone \( j \) of \( Y_{ij} \) from the parametric Poisson family \( \mathcal{F} \)
  \[
  \mathcal{F} = \left\{ e^{-\lambda} \frac{\lambda^y}{y!}, \lambda > 0 \right\} \text{ and } \mathbb{E}(Y) = \lambda.
  \]
- We assume:
  - \( Y_{ij} \) are independent
  - \( \eta_{ij} = \mu + \alpha_j \) where \( \mu \) is the intercept, \( \alpha_j \) the clone fixed effects
  - log link: \( \log[\mathbb{E}(Y_{ij})] = \log(\lambda_{ij}) = \eta_{ij} \)
- this implies
  1. \( \mathbb{V}(Y_{ij}) = \mathbb{E}(Y_{ij}) \)
  2. \( \mu \) and \( \alpha_j \)'s are unknown parameters.
Is number of foliar blotches depending on clone?

(y, c) number of foliar blotches and clone identifier sampled from a population. 25 individuals per clone have been measured.

$z = 1$ if tree is non-infected

$z = 0$ if tree is infected

<table>
<thead>
<tr>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>17</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>
Model construction

- \(z_{ij}\) is a realization for individual \(i\) of clone \(j\) of \(Z_{ij}\) from the parametric Bernoulli family \(\mathcal{F}\)
  \[ \mathcal{F} = \{ f_p(z) = p^z (1 - p)^{1-z} \mid 0 \leq p \leq 1 \} \]  and \(\mathbb{E}(Z) = p\).

We assume:

- \(Z_{ij}\)'s are independent
- \(\eta_{ij} = \mu + \alpha_j\) where \(\mu\) is the intercept, \(\alpha_j\) the clone fixed effects
- logit link: \(\text{logit}(p_{ij}) = \log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \eta_{ij}\)

this implies

1. \(\mathbb{V}(Z_{ij}) = p_{ij}(1 - p_{ij})\)
2. \(\mu\) and \(\alpha_j\)'s are unknown parameters.
30 families have been studied and 10 individuals per family have been measured. Let $(y, f)$ height and family of the 300 individuals sampled from a population.
Model construction 1: select the best families

- Estimate the value of each family $j$

  We assume:

  1. $Y_{ij}$ comes from the gaussian parametric density,
  2. $Y_{ij}$ are independent
  3. $\mathbb{E}(Y_{ij}) = \mu + \alpha_j$ where $\mu$ is the intercept, $\alpha_j$ the family fixed effects
  4. $\nabla(Y_{ij}) = \sigma^2, \forall i, j$

- this implies

  1. family fixed effects
  2. individuals belonging to a same family are assumed to be independent
Model construction 2: genetic breeding program

- estimate the family variability in the whole population
  
  We assume:
  1. $Y_{ij}$ comes from the gaussian parametric density,
  2. $\mathbb{E}(Y_{ij}|\alpha_j) = \mu + \alpha_j$ where $\mu$ is the intercept, $\alpha_j$ the family random effects
  3. conditionnally to $\alpha_j$, $Y_{ij}$ are independent
  4. $\mathbb{V}(Y_{ij}|\alpha_j) = \sigma^2, \forall i,j$
  5. $\alpha_j \sim \mathcal{N}(0, \sigma^2_{\alpha})$

- this implies
  1. marginal expectation $\mathbb{E}(Y_{ij}) = \mu$
  2. marginal correlation

$$\rho(Y_{ij}, Y_{i'j'}) = \begin{cases} \frac{\sigma^2_{\alpha}}{\sigma^2_{\alpha} + \sigma^2} & \text{if } j = j' \\ 0 & \text{if } j \neq j' \end{cases}$$

- individuals belonging to a same family are not independent
Linear Models definition

Let \((y_i, x_{i1}, \ldots, x_{iK})_{i=1}^n\) denote \(K + 1\) observations for \(n\) individuals. \((y_i)_{i=1}^n\) realizations of \(n\) random variables \((Y_i)_{i=1}^n\). \(Y_i\) are modeled by a linear model when

1. \(Y_i\) comes from the gaussian parametric density family
2. \(Y_i\) are independent
3. predictor: \(\eta_i = \beta_0 + \sum_{k=1}^K x_{ik} \beta_k\) (linear combination of parameters)
4. expectation with identity link: \(\mathbb{E}(Y_i) = \eta_i\)
5. variance: \(\mathbb{V}(Y_i) = \sigma^2, \forall i\)

\[
Y_i = \beta_0 + \sum_{k=1}^K x_{ik} \beta_k + \varepsilon_i \quad \text{fixed part}
\]

\[
\varepsilon_i \quad iid \quad \sim \mathcal{N}(0, \sigma^2)
\]
Matrix formulation

Whatever the nature of the explanatory variables (regressors, factors), linear model always could be defined as follows:

\[ Y = X\beta + \varepsilon \]

with \( \beta = (\beta_0, \beta_1, \ldots, \beta_K)' \)

\begin{align*}
\begin{pmatrix}
Y_1 \\
\vdots \\
Y_6
\end{pmatrix}
&= \begin{pmatrix}
1 & x_{1,1} & \cdots & x_{1,3} \\
1 & x_{6,1} & \cdots & x_{6,3}
\end{pmatrix}
\begin{pmatrix}
\beta_0 \\
\beta_1 \\
\beta_2 \\
\beta_3
\end{pmatrix} + \varepsilon

\begin{pmatrix}
Y_1 \\
\vdots \\
Y_6
\end{pmatrix}
&= \begin{pmatrix}
1 & 1 & 0 \\
1 & 1 & 0 \\
1 & 0 & 1 \\
1 & 0 & 1 \\
1 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
\beta_0 \\
\beta_1 \\
\beta_2
\end{pmatrix} + \varepsilon
\end{align*}
### Terminology:

#### Multiple Linear Regression/ANOVA/ANCOVA

- if matrix $X$ only contains regressors, models are called multiple regression models.
- if matrix $X$ only contains factors, models are called Analysis Of Variance (ANOVA) models ($X$ is a matrix composed with 0 or 1).
- if matrix $X$ contains both regressors and factors, models are called Analysis of Covariance (ANCOVA) models.
R procedure

> res.lm <- lm(formula=y ~ x+f+x:f, data = data, offset =rep(0,n), weights=rep(1,n))

> summary(res.lm)

> anova(res.lm)

> names(res.lm)

> res.lm$coefficients
Gaussian hypothesis

- Graphical
  - histogram, QQ-plot,
  > `hist(res.lm$residuals)`
  > `par(mfrow=c(2,2))`
  > `plot(res.lm)`

- Statistical test
  - Kolmogorov-Smirnov
  > `ks.test(res.lm$res, "pnorm")`
Homoscedasticity hypothesis

- Graphical residual/fitted

```r
> par(mfrow=c(1,2))
> plot(res.lm$fitted.values, res.lm$res) ;
> plot(res.lm$res)
> abline(v=0)
```
Independence hypothesis

Difficult to test !!!

have a look at the plot of the residuals
for time correlation : Durbin-Watson test ...
Estimation

Let’s assume a linear model:

\[ Y = X\beta + \varepsilon \]

with \( \varepsilon \sim \mathcal{N}(0, \sigma^2 I_d) \)

Parameters to be estimated are \( \beta, \sigma \)

In all the following, \( X \) is supposed of full rank: \( \text{rank}(X) = K \)

**Least squares approach**: \( \min(||Y - X\beta||^2) \)

\[ \hat{\beta}_{ls} = (X'X)^{-1}X'Y \]

- best linear unbiased estimator of \( \beta \)
- \( \hat{\beta}_{ls} \sim \mathcal{N}(\beta, \sigma^2 (X'X)^{-1}) \)
- best quadratic unbiased estimator of \( \sigma^2 \)

\[ \hat{\sigma}^2_{ls} = \frac{1}{n-K}(Y - X\hat{\beta})'(Y - X\hat{\beta}) \quad \text{and} \quad \hat{\sigma}^2_{ls} \sim \frac{\sigma^2}{n-K} \chi^2(n-K) \]
Preambles for estimation

Definition (Likelihood function)

The likelihood function is a function of the parameters of a statistical model and is defined as

\[ \Theta \mapsto \mathbb{R} \]

\[ \theta \mapsto f_{\theta}(y), \]

and will be denoted as \( \mathcal{L}(\theta; y) \)
Maximum likelihood approach

- Likelihood

\[
\mathcal{L}(\beta, \sigma, y) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i - x_i' \beta)'(y_i - x_i' \beta)}
\]

- Log-likelihood

\[
\ell(\beta, \sigma, y) = -\frac{n}{2} \log (2\pi\sigma^2) - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)
\]

- Maximum log-likelihood

\[
\partial_{\beta, \sigma} \ell(\beta, \sigma, y) = 0 \Rightarrow \begin{cases} 
\hat{\beta}_{ml} = (X'X)^{-1}X'Y \\
\hat{\sigma}_{ml}^2 = \frac{1}{n}(Y - X\hat{\beta})'(Y - X\hat{\beta})
\end{cases}
\]
\[ \hat{\beta}_{ls} = \hat{\beta}_{ml} \]

but \( \hat{\sigma}^2_{ls} \neq \hat{\sigma}^2_{ml} \)

\( \hat{\sigma}^2_{ls} \):
- is unbiased
- is calculated on the orthogonal of \( X \)
- it takes into account the difference between \( Y \) and its projection on \( X: X\hat{\beta} \)

\( \hat{\sigma}^2_{ml} \):
- is biased
- estimation of \( \sigma^2 \) and \( \beta \) are made jointly
- it does not take into account the lost in degrees of freedom due to the estimation of \( \beta \)
Goodness of fit criterion

- **Adjusted R-square**

\[
R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - x_i'\hat{\beta})^2 / (n - K)}{\sum_{i=1}^{n}(y_i - \bar{y})^2 / (n - 1)}
\]

- **Akaike’s Information Criterion**

\[
AIC = -2 \log L(\hat{\beta}_{ml}, \hat{\sigma}_{ml}, y) + 2K
\]

- **Bayesian Information Criterion**

\[
BIC = -2 \log L(\hat{\beta}_{ml}, \hat{\sigma}_{ml}, y) + K \log(n)
\]
Variable selection in multiple regression

The main approaches

- **Forward selection**, which involves starting with no variables in the model, trying out the variables one by one and including them if they are statistically significant.

- **Backward elimination**, which involves starting with all candidate variables and testing them one by one for statistical significance, deleting any that are not significant.

- **Stepwise methods** that are a combination of the above, testing at each stage for variables to be included or excluded.

step function

- For AIC: `> step(model, data, direction = c("both", "backward", "forward"), k=2, trace)`
  
k=2 is by default

- For BIC: `> step(model, data, direction, k=log(nrow(data)), trace)`
Generalized Linear Models – Introduction

- Agricultural Science - different types of data (responses):
  - continuous: weight, height, diameter
  - discrete: count, proportion

- Model choice - important part of the research: search for a simple model which explains well the data.

- All models envolve:
  - a systematic component - related to the explanatory variables (regression model, analysis of variance model, analysis of covariance model);
  - a random component - related to the distributions followed by the response variables;
  - a link between systematic and random components.
### Motivating example – Carnation meristem culture

<table>
<thead>
<tr>
<th></th>
<th>0.0</th>
<th>0.1</th>
<th>0.3</th>
<th>0.5</th>
<th>1.0</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>s</td>
<td>l</td>
<td>v</td>
<td>s</td>
<td>l</td>
<td>v</td>
</tr>
<tr>
<td>1</td>
<td>2.5</td>
<td>0</td>
<td>3</td>
<td>5.5</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>2.5</td>
<td>0</td>
<td>2</td>
<td>4.3</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>3.0</td>
<td>0</td>
<td>6</td>
<td>3.3</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>2.5</td>
<td>1</td>
<td>3</td>
<td>4.3</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>4.0</td>
<td>0</td>
<td>4</td>
<td>5.4</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>4.0</td>
<td>0</td>
<td>3</td>
<td>3.8</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>3.0</td>
<td>0</td>
<td>3</td>
<td>4.3</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>3.0</td>
<td>0</td>
<td>4</td>
<td>6.0</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>5.0</td>
<td>0</td>
<td>3</td>
<td>5.0</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>4.0</td>
<td>0</td>
<td>2</td>
<td>5.0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>2.0</td>
<td>1</td>
<td>3</td>
<td>4.5</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>4.0</td>
<td>0</td>
<td>3</td>
<td>4.0</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>3.0</td>
<td>0</td>
<td>4</td>
<td>3.3</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>3.5</td>
<td>1</td>
<td>3</td>
<td>4.3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>3.0</td>
<td>0</td>
<td>3</td>
<td>4.5</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>3.0</td>
<td>0</td>
<td>2</td>
<td>3.8</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>5.5</td>
<td>0</td>
<td>3</td>
<td>4.7</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>3.0</td>
<td>0</td>
<td>4</td>
<td>2.2</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>2.5</td>
<td>0</td>
<td>2</td>
<td>3.8</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>2.0</td>
<td>0</td>
<td>3</td>
<td>5.0</td>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>

- **Response variables:** number of shoots (s), average length of shoots (l), vitrification (v)
- **Distributions:** ??
- **Systematic component:** regression model, completely randomized experiment.
- **Links:** ??
Plan Models with fixed effects

Linear Mixed Models (LMM)  Generalized Linear Models

Overdispersion  Nonlinear models

Linear Models  Generalized Linear Models

BAP dose  Number of shoots

Average length of shoots

Vitrification

C. Demetrio, F. Mortier & C. Trottier  Mixed Models, theory and applications
Motivating example – Rotenon toxicity

<table>
<thead>
<tr>
<th>Dose ($d_i$)</th>
<th>$m_i$</th>
<th>$y_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>49</td>
<td>0</td>
</tr>
<tr>
<td>2.6</td>
<td>50</td>
<td>6</td>
</tr>
<tr>
<td>3.8</td>
<td>48</td>
<td>16</td>
</tr>
<tr>
<td>5.1</td>
<td>46</td>
<td>24</td>
</tr>
<tr>
<td>7.7</td>
<td>49</td>
<td>42</td>
</tr>
<tr>
<td>10.2</td>
<td>50</td>
<td>44</td>
</tr>
</tbody>
</table>

- Response variable: $Y_i$ – number of dead insects out of $m_i$ insects (Martin, 1942).
- Distribution: Binomial.
- Systematic component: regression model, completely randomized experiment.
- Aim: Lethal doses.
the curve of the observed proportions against doses has an S shape
Motivation for the link function

- Bioassay or biological assay – the measurement of the potency of a stimulus by means of the reactions it produces in living matter.
- The stimulus may be physiological, chemical, biological, etc.
  - insects exposed to chemical stimuli (insecticides or biological, eg a virus, stimuli)
  - other agricultural bioassays may involve large animals, fungi, plants or plant parts such as leaves.
- The response here is quantal or binary – just 2 possible responses (success or failure).
  - an insect dies or survives
  - a seed germinates or fails to germinate
  - a cutting roots or fails to root
- The rotenone data example

C. Demetrio, F. Mortier & C. Trottier
The tolerance of an individual (e.g., an insect) is the dose or concentration or intensity of the stimulus above which the individual responds (e.g., dies) and below which it does not respond.

An individual dies if his tolerance is smaller than a given dose.

Tolerance varies between individuals – it is a random variable.

Suppose

- $z$: tolerance of a randomly chosen individual
- $f_Z(z)$: the probability density function of $Z$
- $x$: dose of the stimulus
Then, for an individual chosen at random from the population

\[ P(\text{death}|x) = P(Z < x) = \int_{-\infty}^{x} f(z)\,dz \]

It is often reasonable to assume that the tolerance \( Z \) has a normal distribution, that is, \( Z \sim N(\mu, \sigma^2) \). Then, the cumulative normal distribution is

\[ \pi = P(\text{death}|x) = P(Z < x) = P \left( \frac{Z - \mu}{\sigma} < \frac{x - \mu}{\sigma} \right) = \Phi \left( \frac{x - \mu}{\sigma} \right) \]

The inverse of the cumulative normal function, \( \text{probit}(\pi) = \Phi^{-1}(\pi) = \alpha + \beta x \), where \( \alpha = -\frac{\mu}{\sigma} \) and \( \beta = \frac{1}{\sigma} \), is called the \textit{probit} transformation.
Alternatively, we can assume that tolerance $Z$ has a logistic distribution with mean $\mathbb{E}(Z) = \mu$, variance $\text{Var}(Z) = \pi^2 \tau^2 / 3$ and pdf

$$f_Z(z; \mu, \tau) = \frac{1}{\tau} \exp \left( \frac{z - \mu}{\tau} \right) \left[ 1 + \exp \left( \frac{z - \mu}{\tau} \right) \right]^{-2}, \quad \mu \in \mathbb{R}, \quad \tau > 0,$$

Then, the cumulative logistic distribution is

$$\pi = P(Z \leq x) = F(x) = \int_{-\infty}^{x} \frac{\beta e^{\alpha + \beta z}}{(1 + e^{\alpha + \beta z})^2} dz = \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}}$$

where $\alpha = -\mu / \tau$ and $\beta = 1 / \tau$.

The inverse of the cumulative distribution of the logistic distribution is called the \textbf{logit} transformation

$$\text{logit}(\pi) = \log \left( \frac{\pi}{1 - \pi} \right) = \alpha + \beta x$$

The normal and logistic distributions are symmetrical around the mean.
A different distribution, which is asymmetrical, for the tolerance \( Z \) is the extreme value (Gumbel) distribution with mean \( \mathbb{E}(Z) = \alpha + \gamma \tau \), variance \( \text{Var}(Z) = \pi^2 \tau^2 / 6 \), where \( \gamma \approx 0.577216 \) is the Euler number defined by \( \gamma = -\psi(1) = \lim_{n \to \infty} (\sum_{i=1}^{n} i^{-1} - \log n) \), \( \psi(p) = d \log \Gamma(p) / dp \) is the digamma function, and pdf

\[
f_Z(z; \alpha, \tau) = \frac{1}{\tau} \exp \left( \frac{z - \alpha}{\tau} \right) \exp \left[ -\exp \left( \frac{z - \alpha}{\tau} \right) \right], \quad \alpha \in \mathbb{R}, \quad \tau > 0,
\]

Then, the cumulative Gumbel distribution is

\[
\pi = F(x) = \int_{-\infty}^{x} \beta \exp (\alpha + \beta z - e^{\alpha + \beta z}) \, dz = 1 - \exp \left[ -\exp(\alpha + \beta x) \right]
\]

where \( \alpha = -\mu / \tau \) and \( \beta = 1 / \tau \).

The inverse of the cumulative Gumbel distribution is called complementary log-log transformation

\[
c\text{-loglog}(\pi) = \log(-\log(1 - \pi)) = \alpha + \beta x
\]

The exposed theory shows one type of biological justification for the link function.
Generalized Linear Models – Definition

The three components of a generalized linear model (Nelder and Wedderburn, 1972; McCullagh and Nelder, 1989) are:

- independent random variables $Y_i$, $i = 1, \ldots, n$, from an exponential family distribution with means $\mu_i$ and constant dispersion (scale) parameter $\phi$,

$$f(y) = \exp \left\{ \frac{y\theta - b(\theta)}{\phi} + c(y, \phi) \right\}$$

where $\mu = \mathbb{E}[Y] = b'(\theta)$ and $\text{Var}(Y) = \phi b''(\theta) = \phi V(\mu)$, $V(\mu)$ called variance function.

- a linear predictor vector $\eta$ given by

$$\eta = X\beta$$

where $\beta$ is a vector of $p$ unknown parameters and $X = [x_1, \ldots, x_n]'$ is the $n \times p$ design matrix;

- a link function $g(\cdot)$ relating the mean to the linear predictor, i.e.

$$g(\mu_i) = \eta_i = x'_i \beta$$
Normal Models

\(Y_i, \, i = 1, \ldots, n\), a continuous response variable, \(Y_i \sim N(\mu_i, \sigma^2)\) with mean \(\mu_i\) and constant variance \(\sigma^2\). We model the mean \(\mu_i\) in terms of the explanatory variables \(x_i\).

As a glm

- **Random component:**

\[
f(y_i; \mu_i, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{(y_i - \mu_i)^2}{2\sigma^2} \right] = \exp \left[ \frac{1}{\sigma^2} \left( y_i \mu_i - \frac{\mu_i^2}{2} \right)^2 - \frac{1}{2} \log(2\pi\sigma^2) - \frac{y_i^2}{2\sigma^2} \right]
\]

where

\[
\theta_i = \mu_i, \quad \phi = \sigma^2, \quad b(\theta_i) = \frac{\mu_i^2}{2} = \frac{\theta_i^2}{2}, \quad c(y_i, \phi) = -\frac{1}{2} \left[ \frac{y_i^2}{\sigma^2} + \log(2\pi\sigma^2) \right], \quad V(\mu_i) = 1
\]

- **Systematic component:** \(\eta_i = x_i'\beta\)

- **Link function:** \(\eta_i = \mu_i\) (identity link, canonical link)
Binomial regression model

$Y_i$, count of successes out of a sample of size $m_i$, $i = 1, \ldots, n$

$Y_i \sim \text{Bin}(m_i, \pi_i)$ with mean $\mathbb{E}[Y_i] = \mu_i = m_i\pi_i$, and

$\text{Var}(Y_i) = m_i\pi_i(1 - \pi_i)$

We model the expected proportions $\pi_i \in [0, 1]$ in terms of the explanatory variables $x_i$. As a glm

- **Random component:**

  $$f(y_i; \pi_i) = \left(\frac{m}{y_i}\right)\pi_i^{y_i}(1 - \pi_i)^{m_i - y_i} = \exp \left[ y_i \log \left( \frac{\pi_i}{1 - \pi_i} \right) + m_i \log(1 - \pi_i) + \log \left( \frac{m_i}{y_i} \right) \right],$$

  where $y_i = 0, 1, \ldots, m_i$,

  $$\phi = 1, \quad \theta_i = \log \left( \frac{\pi_i}{1 - \pi_i} \right) = \log \left( \frac{\mu_i}{m_i - \mu_i} \right) \Rightarrow \mu_i = \frac{me_i^\theta}{(1 + e_i^\theta)},$$

  $$b(\theta) = -m_i \log(1 - \pi_i) = m_i \log (1 + e^{\theta_i}), \quad c(y_i, \phi) = \log \left( \frac{m}{y_i} \right), \quad V(\mu_i) = \mu_i(m_i - \mu_i)/m_i$$

- **Systematic component:**

  $$\eta_i = x_i'\beta$$
**Link function:**

The canonical link function is the logit

\[ \eta_i = g(\pi_i) = g\left(\frac{\mu_i}{m_i}\right) = \log \left(\frac{\mu_i}{m_i - \mu_i}\right) = \log \left(\frac{\pi_i}{1 - \pi_i}\right) \]

Other common choices are

- probit \( \eta_i = g(\pi_i) = g\left(\frac{\mu_i}{m_i}\right) = \Phi^{-1}(\mu_i/m_i) = \Phi^{-1}(\pi_i) \)
- complementary log-log link

\[ \eta_i = g(\pi_i) = g\left(\frac{\mu_i}{m_i}\right) = \log\{-\log(1 - \pi_i)\}. \]
Poisson regression models

$Y_i, \, i = 1, \ldots, n$, are counts with means $\mu_i$

$Y_i \sim \text{Pois}(\mu_i)$ with mean $\mu_i$ and variance $\text{Var}(Y_i) = \mu_i$

As a glm

- **Random component:**

$$f(y_i; \mu_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} = \exp[y_i \log(\mu_i) - \mu_i - \log(y_i!)]$$

where

$\phi = 1, \quad \theta_i = \log(\mu_i) \Rightarrow \mu_i = \exp(\theta_i), \quad b(\theta) = \mu_i = \exp(\theta_i), \quad c(y_i, \phi) = \log(y_i!), \quad V(\mu_i) = \mu_i$

- **Systematic component:**

$$\eta_i = x_i' \beta$$

- **Link function:**

The canonical link function is the log

$$\eta_i = g(\mu_i) = \log(\mu_i)$$
For different observation periods/areas/volumes:

\[ Y_i \sim \text{Pois}(t_i \lambda_i) \]

Taking a log-linear model for the rates,

\[ \log(\lambda_i) = x_i' \beta \]

results in the following log-linear model for the Poisson means

\[ \log(\mu_i) = \log(t_i \lambda_i) = \log(t_i) + x_i' \beta, \]

where the \( \log(t_i) \) is included as a fixed term, or offset, in the model.
## Components of some distributions from exponential family

<table>
<thead>
<tr>
<th>Distribution</th>
<th>$\phi$</th>
<th>$\theta$</th>
<th>$b(\theta)$</th>
<th>$c(y, \phi)$</th>
<th>$\mu(\theta)$</th>
<th>$V(\mu)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal: $N(\mu, \sigma^2)$</td>
<td>$\sigma^2$</td>
<td>$\mu$</td>
<td>$\frac{\theta^2}{2}$</td>
<td>$-\frac{1}{2} \left[ \frac{y^2}{\sigma^2} + \log(2\pi \sigma^2) \right]$</td>
<td>$\theta$</td>
<td>1</td>
</tr>
<tr>
<td>Poisson: $P(\mu)$</td>
<td>1</td>
<td>$\log(\mu)$</td>
<td>$e^\theta$</td>
<td>$-\log(y!)$</td>
<td>$\frac{e^\theta}{\mu}$</td>
<td>$\frac{\mu}{m(m - \mu)}$</td>
</tr>
<tr>
<td>Binomial: $B(m, \pi)$</td>
<td>1</td>
<td>$\log\left(\frac{\frac{\mu}{m - \mu}}{1 - \mu + \pi} \right)$</td>
<td>$m \log(1 + e^\theta)$</td>
<td>$\log\left(\frac{m}{y} \right)$</td>
<td>$\frac{m e^\theta}{1 + e^\theta}$</td>
<td>$\frac{\mu}{m} \left(\frac{\mu}{k} + 1\right)$</td>
</tr>
<tr>
<td>Negative Binomial: $BN(\mu, k)$</td>
<td>1</td>
<td>$\log\left(\frac{\frac{\mu}{\mu + k}}{1 - \mu + \pi} \right)$</td>
<td>$-k \log(1 - e^\theta)$</td>
<td>$\log\left(\frac{\Gamma(k + y)}{\Gamma(k) y!}\right)$</td>
<td>$\frac{ke^\theta}{1 - e^\theta}$</td>
<td>$\frac{\mu^2}{\pi}$</td>
</tr>
<tr>
<td>Gamma: $G(\mu, \nu)$</td>
<td>$\nu^{-1}$</td>
<td>$-\frac{1}{\mu}$</td>
<td>$-\log(-\theta)$</td>
<td>$\nu \log(\nu y) - \log(y) - \log(\Gamma(\nu))$</td>
<td>$\frac{1}{\theta}$</td>
<td>$\mu^2$</td>
</tr>
<tr>
<td>Inverse Gaussian: $IG(\mu, \sigma^2)$</td>
<td>$\sigma^2$</td>
<td>$\frac{1}{2\mu^2}$</td>
<td>$-(2\theta)^{1/2}$</td>
<td>$-\frac{1}{2} \left[ \log(2\pi \sigma^2 y^3) + \frac{1}{\sigma^2 y} \right]$</td>
<td>$(-2\theta)^{-1/2}$</td>
<td>$\mu^3$</td>
</tr>
</tbody>
</table>
## Canonical link functions for some distributions

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Canonical link functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>Identity: $\eta = \mu$</td>
</tr>
<tr>
<td>Poisson</td>
<td>Logarithmic: $\eta = \log(\mu)$</td>
</tr>
<tr>
<td>Binomial</td>
<td>Logistic: $\eta = \log \left( \frac{\pi}{1 - \pi} \right) = \log \left( \frac{\mu}{m - \mu} \right)$</td>
</tr>
<tr>
<td>Gamma</td>
<td>Reciprocal: $\eta = \frac{1}{\mu}$</td>
</tr>
<tr>
<td>Inverse Gaussian</td>
<td>Reciprocal squared: $\eta = \frac{1}{\mu^2}$</td>
</tr>
</tbody>
</table>

C. Demétrio, F. Mortier & C. Trottier

*Mixed Models, theory and applications*
Maximum likelihood estimation.

Estimation algorithm (Nelder and Wedderburn, 1972) – Iteratively weighted least squares (IWLS)

\[ X'W^{[t]}X\beta^{[t+1]} = X'W^{[t]}z^{[t]} \]

where

- \( t \) denotes the iteration index
- \( X = [x_1, x_2, \ldots, x_n]' \) is a design matrix \( n \times p \),
- \( W = \text{diag}\{W_i\} \) – depends on the known (prior) weights (\( w_i \)), variance function (distribution) and link function
  \[ W_i = \frac{w_i}{V(\mu_i)} \left( \frac{d\mu_i}{d\eta_i} \right)^2 \]
- \( \beta \) – parameter vector \( p \times 1 \)
- \( z \) – a vector \( n \times 1 \) (adjusted response variable) – depends on \( y \) and link function
  \[ z_i = \eta_i + (y_i - \mu_i) \frac{d\eta_i}{d\mu_i} \]
When the dispersion parameter is unknown, it may be estimated by the Pearson Estimator

$$\hat{\phi} = \frac{1}{n - p} \sum_{i=1}^{n} \frac{w_i(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}$$

where $\hat{\mu}_i = g^{-1}(x_i'\hat{\beta})$ is the $i$th fitted value.

Some computer packages estimate $\phi$ by the deviance estimator $D(\hat{\beta})/(n - p)$; but it cannot be recommended because of problems with bias and inconsistency in the case of a non-constant variance function.

For positive data, the deviance may also be sensitive to rounding errors for small values of $y_i$.

The asymptotic variance of $\hat{\beta}$ is estimated by the inverse (Fisher) information matrix, giving

$$\text{Var}(\hat{\beta}) = K = \phi(X'WX)^{-1},$$

where $W$ is calculated from $\hat{\beta}$.

The standard error $\text{se}(\hat{\beta}_j)$ is calculated as the square-root of the $j$th diagonal element of this matrix, for $j = 1, \ldots, p$. 
When \( \phi \) is known, a \( 1 - \alpha \) confidence interval for \( \beta_j \) is defined by the endpoints
\[
\hat{\beta}_j \pm \text{se}(\hat{\beta}_j) z_{1-\alpha/2}
\]
where \( z_{1-\alpha/2} \) is the \( 1 - \alpha/2 \) standard normal quantile.

For \( \phi \) unknown, we replace \( \phi \) by \( \hat{\phi} \) in \( K \) and a \( 1 - \alpha \) confidence interval for \( \beta_j \) is defined by the endpoints
\[
\hat{\beta}_j \pm \text{se}(\hat{\beta}_j) t(1-\alpha/2)(n - p)
\]
where \( t(1-\alpha/2)(n - p) \) is the \( 1 - \alpha/2 \) quantile of Student’s t distribution with \( n - p \) degrees of freedom.
Analysis of Deviance – Goodness of fitting and model selection

- Analysis of deviance is the method of parameter inference for generalized linear models based on the deviance, generalizing ideas from ANOVA, and first introduced by Nelder and Wedderburn (1972).

- The situation is similar to regression analysis, in the sense that model terms must be eliminated sequentially, and the significance of a term may depend on which other terms are in the model.

- The deviance $D$ measures the distance between $y$ and $\hat{\mu}$, given by

$$S = \frac{D(\hat{\beta})}{\phi} = -2[\log L(\hat{\mu}, y) - \log L(y, y)] = 2\phi^{-1} \sum_{i=1}^{n} w_i [y_i(\tilde{\theta}_i - \hat{\theta}_i) + b(\hat{\theta}_i) - b(\tilde{\theta}_i)]$$

where $L(\hat{\mu}, y)$ and $L(y, y)$ are the likelihood function values for the current and saturated models, $\tilde{\theta}_i = \theta(y_i)$, $\hat{\theta}_i = \theta(\hat{\mu}_i)$ and $D(\hat{\beta}) = \sum_{i=1}^{n} w_id(y_i; \hat{\mu}_i)$. 
## Deviance for some models

<table>
<thead>
<tr>
<th>Model</th>
<th>Deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>( D_p = \sum_{i=1}^{n} (y_i - \hat{\mu}_i)^2 )</td>
</tr>
<tr>
<td>Binomial</td>
<td>( D_p = 2 \sum_{i=1}^{n} \left[ y_i \log \left( \frac{y_i}{\hat{\mu}_i} \right) + (m_i - y_i) \log \left( \frac{m_i - y_i}{m_i - \hat{\mu}_i} \right) \right] )</td>
</tr>
<tr>
<td>Poisson</td>
<td>( D_p = 2 \sum_{i=1}^{n} \left[ y_i \log \left( \frac{y_i}{\hat{\mu}_i} \right) + (\hat{\mu}_i - y_i) \right] )</td>
</tr>
<tr>
<td>Negative Binomial</td>
<td>( D_p = 2 \sum_{i=1}^{n} \left[ y_i \log \left( \frac{y_i}{\hat{\mu}_i} \right) + (y_i + k) \log \left( \frac{\hat{\mu}_i + k}{y_i + k} \right) \right] )</td>
</tr>
<tr>
<td>Gamma</td>
<td>( D_p = 2 \sum_{i=1}^{n} \left[ \log \left( \frac{\hat{\mu}_i}{y_i} \right) + \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i} \right] )</td>
</tr>
<tr>
<td>Inverse Gaussian</td>
<td>( D_p = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{y_i \hat{\mu}_i^2} )</td>
</tr>
</tbody>
</table>
We consider separately the cases where $\phi$ is known and unknown, but first we introduce some notation.

Let $M_1$ denote a model with $p$ parameters, and let $D_1 = D(\hat{\beta})$ denote the minimized deviance under $M_1$.

Let $M_2$ denote a sub-model of $M_1$ with $q < p$ parameters, and let $D_2$ denote the corresponding minimized deviance, where $D_2 \geq D_1$. 
Known dispersion $\phi$ parameter

- Mainly relevant for discrete data, for which, in general, $\phi = 1$.
- The deviance $D_1$ is a measure of goodness-of-fit of the model $M_1$; and is also known as the $G^2$ statistic in discrete data analysis.
- A more traditional goodness-of-fit statistic is Pearson’s $X^2$ statistic

$$X^2 = \sum \frac{w_i(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}$$

- Asymptotically, for large $w$ the statistics $D_1$ and $X^2$ are equivalent and distributed as $\chi^2(n - p)$ under $M_1$.
- Various numerical and analytical investigations have shown that the limiting $\chi^2$ distribution is approached faster for the $X^2$ statistic than for $D_1$, at least for discrete data.
- A formal level $\alpha$ goodness-of-fit test for $M_1$ is obtained by rejecting $M_1$ if $X^2 > \chi^2_{(1-\alpha)}(n - p)$
- This test may be interpreted as a test for overdispersion.
- The fit of a model is a complex question, cannot be summarized in a single number – supplement with an inspection of residuals.
To test the sub-model $M_2$ with $q < p$ we use the log likelihood ratio statistic

$$D_2 - D_1 \sim \chi^2(p - q)$$

$M_2$ is rejected at level $\alpha$ if $D_2 - D_1 > \chi^2_{1-\alpha}(p - q)$

In the case where $\phi \neq 1$ we use the scaled deviance $D/\phi$ instead of $D$; and the scaled Pearson statistic $X^2/\phi$ instead of $X^2$ and so on.
The dispersion parameter is usually unknown for continuous data.

In the discrete case we may prefer to work with unknown dispersion parameter, if evidence of overdispersion has been found in the data.

There is no formal goodness-of-fit test available based on $X^2$ – the fit of the model $M_1$ to the data must be checked by residual analysis.

$X^2$ is used to estimate the dispersion parameter

$$\hat{\phi} = \frac{1}{n - p} \sum \frac{w_i (y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}$$

where $\hat{\mu}_i = g^{-1}(\hat{\beta}'x_i)$ is the $i$th fitted value.

To test the sub-model $M_2$ with $q < p$ parameters inference may be based on $F$–statistic,

$$F = \frac{(D_2 - D_1)/(p - q)}{\hat{\phi}} \sim F(p - q, n - p)$$

We reject $M_2$ at level $\alpha$ if $F > F_{1-\alpha}(p - q, n - p)$.
Suppose a completely randomized experiment with two factors $A$ (with $a$ levels) and $B$ (with $b$ levels) and $r$ replications.

<table>
<thead>
<tr>
<th>Model</th>
<th>DF</th>
<th>Deviance</th>
<th>Deviance Diff.</th>
<th>DF Diff.</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null</td>
<td>$rab - 1$</td>
<td>$D_1$</td>
<td>$D_1 - D_A$</td>
<td>$a - 1$</td>
<td>$A$ ignoring $B$</td>
</tr>
<tr>
<td>A</td>
<td>$a(rb - 1)$</td>
<td>$D_A$</td>
<td>$D_A - D_{A+B}$</td>
<td>$b - 1$</td>
<td>$B$ including $A$</td>
</tr>
<tr>
<td>A+B</td>
<td>$a(rb - 1) - (b - 1)$</td>
<td>$D_{A+B}$</td>
<td>$D_{A+B} - D_{A*B}$</td>
<td>$(a - 1)(b - 1)$</td>
<td>Interaction $AB$ included $A$ and $B$</td>
</tr>
<tr>
<td>A+B+A.B</td>
<td>$ab(r - 1)$</td>
<td>$D_{A*B}$</td>
<td>$D_{A*B}$</td>
<td>$ab(r - 1)$</td>
<td>Residual</td>
</tr>
<tr>
<td>Saturated</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Residual analysis

- **Pearson residual**

\[ r_{Pi} = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}} \]

reflect the skewness of the underlying distribution.

- **Deviance residual**

\[ r_{Di} = \text{sign}(y_i - \hat{\mu}_i) \sqrt{d(y_i; \hat{\mu}_i)} \]

which is much closer to being normal than the Pearson residual, but has a bias (Jorgensen, 2011).

- **Modified deviance residual** (Jorgensen, 1997)

\[ r_{Di}^* = r_{Di} + \frac{\phi}{r_{Di}} \log \frac{r_{Wi}}{r_{Di}} \]

where \( r_{Wi} \) is the Wald residual defined by

\[ r_{Wi} = [g_0(y_i) - g_0(\mu_i)] \sqrt{V(y_i)} \]

where \( g_0 \) is the canonical link.
- All those residuals have approximately mean zero and variance $\phi(1 - h_i)$, where $h_i$ is the $i$th diagonal element of the matrix $H = W^{1/2}X(X'WX)^{-1}X'W^{1/2}$.

- Use standardized residuals such as $r_{Di}^*(1 - h_i)^{1/2}$, which are nearly normal with variance $\phi$.

- Plot residuals against the fitted values – to check the proposed variance function.

- Normal Q-Q plot (or normal Q-Q plot with simulated envelopes) for the residuals – to check the correctness of the distributional assumption.
Checking for the link function

Suppose the link function $g_0(\mu) = g(\mu, \lambda_0) = X\beta$, nested in a parametric family $g(\mu, \lambda)$, indexed by the parameter $\lambda$, for example,

$$g(\mu, \lambda) = \begin{cases} \frac{\mu^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log(\mu) & \lambda = 0 \end{cases}$$

which includes the identity, logarithmic links, or the Aranda-Ordaz family,

$$\mu = \begin{cases} 1 - (1 + \lambda e^\eta)^{-\frac{1}{\lambda}} & \lambda e^\eta > -1 \\ 1 & \text{c.c.} \end{cases}$$

which includes the identity, complementary log-log links. The Taylor expansion for $g(\mu, \lambda)$ around $\lambda_0$, gives

$$g(\mu, \lambda) \simeq g(\mu, \lambda_0) + (\lambda - \lambda_0)u(\lambda_0) = X\beta + \gamma u(\lambda_0)$$

where $u(\lambda_0) = \frac{\partial g(\mu, \lambda)}{\partial \lambda} \bigg|_{\lambda=\lambda_0}$.

In general it is used $u(\lambda_0) = \hat{\eta}^2$. 
Justification

Suppose the link function used was \( \eta = g(\mu) \) and that the true link is \( g^*(\mu) \). Then,

\[
g(\mu) = g[g^*-1(\eta)] = h(\eta)
\]

The null hypothesis is \( H_0 : h(\eta) = \eta \) and \( H_a : h(\eta) = \text{non linear} \).

Using Taylor expansion for \( g(\mu) \) we have:

\[
g(\mu) \approx h(0) + \eta h'(0) + \eta^2 \frac{h''(0)}{2}
\]

then, the added variable is \( \hat{\eta}^2 \), assuming that the model has terms for which the mean is marginal.

**Formal tests**

- Likelihood ratio test
- Score Test
- Wald Test

**Graphics**

- Added variable plot
- Partial residual plot
R commands for GLM

glm(resp ~ linear predictor + offset(of), weights = w, family=familyname(link ="linkname" ))

The $resp$ is the response variable $y$. For a binomial regression model it is necessary to create:

$resp<-cbind(y,n-y)$

The possible families ("canonical link") are:

- binomial(link = "logit")
- gaussian(link = "identity")
- Gamma(link = "inverse")
- inverse.gaussian(link = "1/mu^2")
- poisson(link = "log")
- quasi(link = "identity", variance = "constant")
- quasibinomial(link = "logit")
- quasipoisson(link = "log")

The default family is the gaussian family and default links are the canonical links (don't need to be declared). Other possible links are "probit", "cloglog", "cauchit", "sqrt",etc. To see more, type

? glm
**Example**

Average daily fat yields (kg/day) from milk from a single cow for each of 35 weeks (McCulloch, 2001)

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.31</td>
<td>0.39</td>
<td>0.50</td>
<td>0.58</td>
<td>0.59</td>
<td>0.64</td>
</tr>
<tr>
<td>0.68</td>
<td>0.66</td>
<td>0.67</td>
<td>0.70</td>
<td>0.72</td>
<td>0.68</td>
</tr>
<tr>
<td>0.65</td>
<td>0.64</td>
<td>0.57</td>
<td>0.48</td>
<td>0.46</td>
<td>0.45</td>
</tr>
<tr>
<td>0.31</td>
<td>0.33</td>
<td>0.36</td>
<td>0.30</td>
<td>0.26</td>
<td>0.34</td>
</tr>
<tr>
<td>0.29</td>
<td>0.31</td>
<td>0.29</td>
<td>0.20</td>
<td>0.15</td>
<td>0.18</td>
</tr>
<tr>
<td>0.11</td>
<td>0.07</td>
<td>0.06</td>
<td>0.01</td>
<td>0.01</td>
<td></td>
</tr>
</tbody>
</table>
A typical model:
Fat yield $\approx \alpha t^\beta e^{\gamma t}$ where $t=$week

Transform
\[
Y_i = \alpha t_i^\beta e^{\gamma t_i} e^{\epsilon_i}
\]
\[
\log(Y_i) = \mu_i + \epsilon_i = \log \alpha + \beta \log(t_i) + \gamma_i t + \epsilon_i
\]
\[
\log(Y_i) \sim N(\log \alpha + \beta \log(t_i) + \gamma_i t, \sigma^2)
\]
\[
E[\log(Y_i)] = \log \alpha + \beta \log(t_i) + \gamma_i t
\]

Link
\[
Y_i = \mu_i + \xi_i = \alpha t_i^\beta e^{\gamma t_i} + \xi_i
\]
\[
Y_i \sim N(\alpha t_i^\beta e^{\gamma t_i}, \tau^2)
\]
\[
E[Y_i] = \alpha t_i^\beta e^{\gamma t_i}
\]
\[
\log(E[Y_i]) = \log \alpha + \beta \log(t_i) + \gamma_i t
\]
Plot of fat yield for each week – Observed values and fitted curve

C. Demetrio, F. Mortier & C. Trottier

Mixed Models, theory and applications
R program

# Average daily fat yields (kg/day) from milk
# from a single cow for each of 35 weeks
# McCulloch (2001)
# Ruppert, Cressie, Carroll (1989), Biometrics, 45:637-656

fatyield.dat<-scan(what=list(yield=0))
0.31 0.39 0.50 0.58 0.59 0.64
0.68 0.66 0.67 0.70 0.72 0.68
0.65 0.64 0.57 0.48 0.46 0.45
0.31 0.33 0.36 0.30 0.26 0.34
0.29 0.31 0.29 0.20 0.15 0.18
0.11 0.07 0.06 0.01 0.01

fatyield.dat$week=1:35
attach(fatyield.dat)
lweek<-log(week)

## Normal model for log(fat)
lyield<-log(yield)
mod1<-lm(lyield~week+lweek)
summary(mod1)
anova(mod1)
## Normal model with log link

```r
mod2<-glm(yield~week+lweek, (family=gaussian(link="log")))
fit2<-fitted(mod2)
summary(mod2)
anova(mod2)
```

# Plotting

```r
# Plotting
plot(c(0,35), c(0,0.9), type="n", xlab="Weeks", ylab="Fat yield (kg/day)")
points(week,yield, pch="*")
w<-seq(1,35,0.1)
lines(w, predict(mod2,data.frame(week=w, lweek=log(w)), type="response"), col="green", lty=1)
lines(w, exp(predict(mod1,data.frame(week=w, lweek=log(w)), type="response")), col="red", lty=2)
legend(20,0.9,c("Observed","Log transformation", "Log link"), lty=c(-1,2,1), pch=c("*"," "," "), col=c("black","red","green"),cex=.6)
title(sub="Figure 1. Fat yield (kg/day) for each week")
```
Binomial regression model

Example

Batches of 20 pyrethroid-resistant moths (*Heliothis virescens*), a cotton crop pest) of each sex were exposed to a range of doses of cypermethrin two days after emergence from pupation. The number of moths which were either knocked down or dead was recorded after 72h.

<table>
<thead>
<tr>
<th>Doses ($d_i$)</th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2.0</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>4.0</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>8.0</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>16.0</td>
<td>18</td>
<td>12</td>
</tr>
<tr>
<td>32.0</td>
<td>20</td>
<td>16</td>
</tr>
</tbody>
</table>

- **Response variable**: $Y_i$ – number of dead insects out of 20 insects.
- **Distribution**: Binomial.
- **Systematic component**: regression model, completely randomized experiment.
- **Aim**: Lethal doses.
Cypermethrin toxicity example

Deviance residuals

<table>
<thead>
<tr>
<th>Terms fitted in model</th>
<th>d.f.</th>
<th>Deviance</th>
<th>$X^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>11</td>
<td>124.9</td>
<td>101.4</td>
</tr>
<tr>
<td>Sex</td>
<td>10</td>
<td>118.8</td>
<td>97.4</td>
</tr>
<tr>
<td>Dose</td>
<td>6</td>
<td>15.2</td>
<td>12.9</td>
</tr>
<tr>
<td>Sex + Dose</td>
<td>5</td>
<td>5.0</td>
<td>3.7</td>
</tr>
</tbody>
</table>

Analysis of deviance

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>Deviance</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>1</td>
<td>6.1</td>
<td>0.0144</td>
</tr>
<tr>
<td>Sex</td>
<td>Dose</td>
<td>1</td>
<td>10.2</td>
</tr>
<tr>
<td>Dose</td>
<td>5</td>
<td>109.7</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Dose</td>
<td>Sex</td>
<td>5</td>
<td>113.8</td>
</tr>
<tr>
<td>Residual</td>
<td>5</td>
<td>5.0</td>
<td>0.5841</td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
<td>124.9</td>
<td></td>
</tr>
</tbody>
</table>
Cypermethrin toxicity example

Models

\[ \text{logit}(p) = \alpha_j + \beta_j \ \log_2(\text{dose}) \] – different logistic regression lines

\[ \text{logit}(p) = \alpha_j + \beta \ \log_2(\text{dose}) \] – common slope

\[ \text{logit}(p) = \alpha + \beta_j \ \log_2(\text{dose}) \] – common intercept

\[ \text{logit}(p) = \alpha + \beta \ \log_2(\text{dose}) \] – same logistic regression line.
## Residual Deviances

<table>
<thead>
<tr>
<th>Terms fitted in model</th>
<th>d.f.</th>
<th>Deviance</th>
<th>$X^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>11</td>
<td>124.9</td>
<td>101.4</td>
</tr>
<tr>
<td>Sex + Sex $\log_2$(dose)</td>
<td>8</td>
<td>4.99</td>
<td>3.51</td>
</tr>
<tr>
<td>Sex + $\log_2$(dose)</td>
<td>9</td>
<td>6.75</td>
<td>5.31</td>
</tr>
<tr>
<td>Const. + Sex $\log_2$(dose)</td>
<td>9</td>
<td>5.04</td>
<td>3.50</td>
</tr>
<tr>
<td>Const. + $\log_2$(dose)</td>
<td>10</td>
<td>16.98</td>
<td>14.76</td>
</tr>
</tbody>
</table>

## Analysis of deviance

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>Deviance</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>1</td>
<td>6.1</td>
<td>0.0144</td>
</tr>
<tr>
<td>Linear Regression</td>
<td>1</td>
<td>112.0</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Residual</td>
<td>9</td>
<td>6.8</td>
<td>0.7473</td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
<td>124.9</td>
<td></td>
</tr>
</tbody>
</table>
Cypermethrin toxicity example

Regression equations
males
\[ \log \frac{\hat{p}}{1 - \hat{p}} = -2.372 + 1.535 \log_2(\text{dose}) \]
females
\[ \log \frac{\hat{p}}{1 - \hat{p}} = -3.473 + 1.535 \log_2(\text{dose}) \]

Lethal Doses
males
\[ \log_2(\hat{LD}_{50}) = \frac{2.372}{1.535} = 1.55 \Rightarrow \hat{LD}_{50} = 4.69 \]
females
\[ \log_2(\hat{LD}_{50}) = \frac{3.473}{1.535} = 2.26 \Rightarrow \hat{LD}_{50} = 9.61 \]
Cypermethrin toxicity example

Plot of observed proportions and fitted curves
```r
# R program

y <- c(1, 4, 9, 13, 18, 20, 0, 2, 6, 10, 12, 16)
sex<-factor(rep(c("M","F"), c(6,6)))
ldose<-rep(0:5,2)
dose<-2**ldose
dose<-factor(dose)
Cyper.dat <- data.frame(sex, dose, ldose, y)
attach(Cyper.dat)

plot(ldose,y/20, pch=c(rep("*",6),rep("+",6)),
col=c(rep("green",6), rep("red",6)),
xlab="log(dose)", ylab="Proportion killed")

resp<-cbind(y,20-y)

mod1<-glm(resp~1, family=binomial)
mod2<-glm(resp~dose, family=binomial)
mod3<-glm(resp~sex, family=binomial)
mod4<-glm(resp~dose+sex, family=binomial)
anova(mod1, mod2, mod4, test="Chisq")
anova(mod1, mod3, mod4, test="Chisq")
```
mod5<-glm(resp~ldose, family=binomial)
mod6<-glm(resp~sex+ldose-1, family=binomial)
mod7<-glm(resp~ldose/sex, family=binomial)
mod8<-glm(resp~ldose*sex, family=binomial)
anova(mod1, mod5, mod6, mod8, test="Chisq")
anova(mod1, mod5, mod7, mod8, test="Chisq")
summary(mod6)

## Plotting
plot(c(1,32), c(0,1), type="n", xlab="log(dose)", ylab="Proportions", log="x")
points(2**ldose,y/20, pch=c(rep("*",6),rep("+",6)), col=c(rep("green",6),rep("red",6)))
ld<-seq(0,5,0.1)
lines(2**ld, predict(mod6,data.frame(ldose=ld, sex=factor(rep("M",length(ld)),levels=levels(sex))), type="response"), col="green")
lines(2**ld, predict(mod6,data.frame(ldose=ld, sex=factor(rep("F",length(ld)),levels=levels(sex))), type="response"), col="red")
CS2 toxicity

Mortality of adult beetles after five hours’ exposure to gaseous carbon disulphid (Bliss, 1935).

<table>
<thead>
<tr>
<th>Log dosage</th>
<th>Number exposed</th>
<th>Number killed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.6907</td>
<td>59</td>
<td>6</td>
</tr>
<tr>
<td>1.7242</td>
<td>60</td>
<td>13</td>
</tr>
<tr>
<td>1.7552</td>
<td>62</td>
<td>18</td>
</tr>
<tr>
<td>1.7842</td>
<td>56</td>
<td>28</td>
</tr>
<tr>
<td>1.8113</td>
<td>63</td>
<td>52</td>
</tr>
<tr>
<td>1.8369</td>
<td>59</td>
<td>53</td>
</tr>
<tr>
<td>1.8610</td>
<td>62</td>
<td>61</td>
</tr>
<tr>
<td>1.8839</td>
<td>60</td>
<td>60</td>
</tr>
</tbody>
</table>

Considerations

- **Response variable**: $Y_i$ – number of killed beetles out of $m_i$ exposed beetles (Pregibon, 1980).
- **Distribution**: Binomial.
- **Systematic component**: regression model, completely randomized experiment.
- **Aim**: Lethal doses.
- Binomial logistic model with $\eta = \beta_0 + \beta_1 \log(\text{dose})$ gives deviance 11.23 on 6 d.f. $\Rightarrow$ lack of fit
- Deviance for link test 8.037 on 1 d.f. $\Rightarrow$ need for different link
- Added variable plot for $U = \hat{\eta}^2$ $\Rightarrow$ evidence that $\gamma \neq 0$, spread throughout data

![Added variable plot](image)

**Figure:** CS2 - Added variable plot

- C-loglog link with a linear lp gives good fit
- Logistic link with a quadratic lp gives good fit
Figure: Beetle mortality - Observed proportions and fitted curves
# *** Mortality of adult beetle Example ***

```r
y <- c(6, 13, 18, 28, 52, 53, 61, 60)
m <- c(59, 60, 62, 56, 63, 59, 62, 60)
ldose <- c(1.6907, 1.7242, 1.7552, 1.7842, 1.8113, 1.8369, 1.8610, 1.8839)
beetle.dat <- data.frame(ldose, m, y)
attach(beetle.dat)
resp<-cbind(y,m-y)

## Logistic link function
modl<-glm(resp ~ poly(I(ldose))+poly(I(ldose),2), family=binomial(link="logit"))
anova(modl, test="Chisq")

mod1l<-glm(resp ~ 1, family=binomial(link="logit"))
1-pchisq(deviance(mod1l), df.residual(mod1l))
print(sum(residuals(mod1l, 'pearson')^2))
mod2l<-glm(resp ~ I(ldose), family=binomial(link="logit"))
1-pchisq(deviance(mod2l), df.residual(mod2l))
print(sum(residuals(mod2l, 'pearson')^2))

## Link function test
LP2l <- (predict(mod2l,type = c("link")))^2
mod3l<-update(mod2l , .~. +LP2l, family=binomial(link="logit"))
anova(mod3l, test="Chisq")

mod4l<-glm(resp ~ poly(I(ldose),2), family=binomial(link="logit"))
1-pchisq(deviance(mod4l), df.residual(mod4l))
print(sum(residuals(mod4l, 'pearson')^2))
```
## Link function test
LP2q <- (predict(mod4l,type = c("link")))^2
mod5l<-update(mod4l , .~. +LP2q, family=binomial(link="logit"))
anova(mod5l, test="Chisq")

## Added variable plot
rp <- residuals(mod2l, 'pearson')
W <- mod2l$fitted*(1- mod2l$fitted)*m
U <- LP2l
mod1n <- glm(U~I(ldose), family=gaussian, weights=W)
rU <- (U - mod1n$fitted)
plot(rU,rp, xlab="Ordinary residuals", ylab="Pearson residuals")

# Partial residual plot
resparc <- residuals(mod3l, 'pearson') + mod3l$coef[3]*U
plot(U,resparc, xlab="U", ylab="Residual + component")

summary(mod6l<-glm(resp ~ I(ldose)+I(ldose^2), family=binomial(link="logit")))
anova(mod6l, test="Chisq")
library(MASS)
# variance-covariance matrix of estimated parameters
vcov(mod6l)
R program

## Complementary log-log link function
modc<-glm(resp ~ poly(I(ldose))+poly(I(ldose),2), family=binomial(link="cloglog"))
anova(modc, test="Chisq")

mod1c<-glm(resp ~ 1, family=binomial(link="cloglog"))
1-pchisq(deviance(mod1c), df.residual(mod1c))
print(sum(residuals(mod1c, 'pearson')^2))
mod2c<-glm(resp ~ I(ldose), family=binomial(link="cloglog"))
1-pchisq(deviance(mod2c), df.residual(mod2c))
print(sum(residuals(mod2c, 'pearson')^2))

## Link function test
LP2l <- (predict(mod2c,type = c("link")))^2
mod3c<-update(mod2c , .~. +LP2l, family=binomial(link="cloglog"))
anova(mod3c, test="Chisq")

## Added variable plot
rp <- residuals(mod2c, 'pearson')
W <- mod2c$fitted*(1- mod2c$fitted)*m
U <- LP2l
mod1n <- glm(U~I(ldose), family=gaussian, weights=W)
rU <- (U - mod1n$fitted)
plot(rU,rp, xlab="Ordinary residuals", ylab="Pearson residuals")

# Partial residual plot
resparc <- residuals(mod3c, 'pearson') + mod3c$coef[3]*U
plot(U,resparc, xlab="U", ylab="Residual + component")
R program

summary(mod5c<-glm(resp ~ I(ldose), family=binomial(link="cloglog")))

library(MASS)

# variance-covariance matrix of estimated parameters
vcov(mod5c)

# Doses LD50 and LD90
dose.p(mod5c); dose.p(mod5c, p=0.9)

# Doses LD25, LD50, LD75
dose.p(mod5c, p = 1:3/4)

# Plotting
plot(c(1.69,1.89), c(0,1), type="n", xlab="Log(dose)", ylab="Proportion of beetles killed")
points(ldose, y/m, pch="*")
ld<-seq(1.69,1.89,0.01)
lines(ld, predict(mod4l,data.frame(ldose=ld),type="response"), col="blue", lty=1)
lines(ld, predict(mod2c,data.frame(ldose=ld),type="response"), col="red", lty=2)
legend(1.7,1.0,c("Observed","Logit", "Cloglog"), lty=c(-1,1,2), pch=c("*"," "," "), col=c("black","blue", "red"))
Germination of Orobanche seed

Example

<table>
<thead>
<tr>
<th>O. aegyptiaca 75</th>
<th>O. aegyptiaca 73</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bean</td>
<td>Cucumber</td>
</tr>
<tr>
<td>10/39</td>
<td>5/6</td>
</tr>
<tr>
<td>23/62</td>
<td>53/74</td>
</tr>
<tr>
<td>23/81</td>
<td>55/72</td>
</tr>
<tr>
<td>26/51</td>
<td>32/51</td>
</tr>
<tr>
<td>17/39</td>
<td>46/79</td>
</tr>
<tr>
<td>10/13</td>
<td></td>
</tr>
</tbody>
</table>

- Response variable: $Y_i$ – number of germinated seeds out of $m_i$ seeds.
- Distribution: Binomial.
- Systematic component: factorial $2 \times 2$ (2 species, 2 extracts), completely randomized experiment (Crowder, 1978).
- Aim: to see how germination is affected by species and extracts.
- Problem: overdispersion.
Germination of Orobanche seed

Simple Binomial Model

- No of seeds germinating \( y_i \) as y-variable
- Binomial logit model
- Species \* Extract interaction model

Analysis of deviance

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>Deviance</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>1</td>
<td>55.97</td>
<td></td>
</tr>
<tr>
<td>E * S</td>
<td>1</td>
<td>56.49</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>1</td>
<td>2.54</td>
<td></td>
</tr>
<tr>
<td>S * E</td>
<td>1</td>
<td>3.06</td>
<td></td>
</tr>
<tr>
<td>S.E</td>
<td>1</td>
<td>6.41</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>17</td>
<td>33.28</td>
<td>0.0096</td>
</tr>
<tr>
<td>Total</td>
<td>20</td>
<td>98.72</td>
<td></td>
</tr>
</tbody>
</table>

- Interaction significant
- Some evidence of overdispersion?
Germination of Orobanche seed

Normal plot with simulated envelope
```R
Species <- factor(c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
                   2, 2, 2, 2, 2, 2, 2, 2, 2, 2))
Extract <- factor(c(1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2,
                    1, 1, 1, 1, 1, 2, 2, 2, 2, 2))
y <- c(10, 23, 23, 26, 17, 5, 53, 55, 32, 46, 10, 8,
       10, 8, 23, 0, 3, 22, 15, 32, 3)
m <- c(39, 62, 81, 51, 39, 6, 74, 72, 51, 79, 13, 16,
       30, 28, 45, 4, 12, 41, 30, 51, 7)

orobanch <- data.frame(Species, Extract, m, y)
attach(orobanch)
resp<-cbind(y,m-y)
p<-y/m

# Binomial fit
orobanchB.fit<-glm(resp~Species*Extract, family=binomial)
summary(orobanchB.fit)
anova(orobanchB.fit, test="Chisq")

orobanchB.fit<-glm(resp~Extract*Species, family=binomial)
summary(orobanchB.fit)
anova(orobanchB.fit, test="Chisq")
```
Example

Storing of micro-organisms

Bacterial concentrations (counts per fixed area) measured at initial freezing (−70°C) and at 1, 2, 6, 12 months.

<table>
<thead>
<tr>
<th>Time</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>6</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>31</td>
<td>26</td>
<td>19</td>
<td>15</td>
<td>20</td>
</tr>
</tbody>
</table>

Hypothesized that count decays over time, i.e.

$$\text{average count } \propto \frac{1}{(\text{Time})^\gamma}$$

Model

$$\text{Count } \sim \text{Pois}(\mu)$$

$$\log \mu = \beta_0 + \beta_1 \log(\text{Time})$$

Avoid problems at time 0 by using

$$\log \mu = \beta_0 + \beta_1 \log(\text{Time} + 0.1)$$
Storing of micro-organisms example

Residual deviances

<table>
<thead>
<tr>
<th>Model</th>
<th>d.f.</th>
<th>Deviance</th>
<th>$X^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>4</td>
<td>7.0672</td>
<td>7.1532</td>
</tr>
<tr>
<td>log(Time)</td>
<td>3</td>
<td>1.8338</td>
<td>1.8203</td>
</tr>
</tbody>
</table>

Analysis of deviance

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>Deviance</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear Regression</td>
<td>1</td>
<td>5.2334</td>
<td>0.0222</td>
</tr>
<tr>
<td>Error</td>
<td>3</td>
<td>1.8338</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>4</td>
<td>7.0672</td>
<td></td>
</tr>
</tbody>
</table>

$log(\hat{\mu}) = 3.149 - 0.1261 \log(\text{Time})$
Storing of micro-organisms example

Plot of bacterial concentration: observed values and fitted curve
ltime <- log((tim <- c(0, 1, 2, 6, 12))+ 0.1)
lcount <- log(count <- c(31, 26, 19, 15, 20))
bacteria.dat <- data.frame(tim, count, ltime, lcount)
with(bacteria.dat,{
  par(mfrow=c(1,2))
  plot(tim, count, xlab="Time in months", ylab="Counts")
  plot(ltime,lcount, xlab="Log(time in months)", ylab="Log(counts)")
  par(mfrow=c(1,1))
})
mod1<-glm(count ~ tim, family=poisson, bacteria.dat)
anova(mod1, test="Chisq")
mod2<-glm(count ~ ltime, family=poisson)
anova(mod2, test="Chisq")

plot(c(0,12), c(15,31), type="n", xlab="Time in months", ylab="Counts")
points(tim,count,pch="*")
x<-seq(0,12,0.1)
lp<-predict(mod2,data.frame(ltime=log(x+0.1)), type="response")
lines(x,lp,lty=1)
title(sub="Figure 1. Data and Fitted curve")
Exercise Relative potency - Toxicity of insecticide to flour beetles

<table>
<thead>
<tr>
<th>Insecticide</th>
<th>Deposit of Insecticide</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.00</td>
</tr>
<tr>
<td>DDT</td>
<td>3/50</td>
</tr>
<tr>
<td>$\gamma$-BHC</td>
<td>2/50</td>
</tr>
<tr>
<td>DDT + $\gamma$-BHC</td>
<td>28/50</td>
</tr>
</tbody>
</table>

Considerations

- **Response variable**: $Y_i$ – number of dead insects out of $m_i$ insects.
- **Distribution**: Binomial.
- **Systematic component**: ANOVA with regression model, completely randomized experiment.
- **Aim**: Lethal doses and comparison of insecticides.
Exercise

Count of the number of plant species on plots that have different biomass (a continuous explanatory variable with three levels: high, mid and low) (Venables and Ripley, 1994)

Tabela: Number of plant species (Y), quantity of biomass (X) and levels of pH of the soil.

<table>
<thead>
<tr>
<th>pH level</th>
<th>Y</th>
<th>X</th>
<th>Y</th>
<th>X</th>
<th>Y</th>
<th>X</th>
<th>Y</th>
<th>X</th>
<th>Y</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>18</td>
<td>0.1008</td>
<td>15</td>
<td>2.6292</td>
<td>13</td>
<td>0.6526</td>
<td>8</td>
<td>3.6787</td>
<td>9</td>
<td>1.5079</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>0.1385</td>
<td>9</td>
<td>3.2522</td>
<td>9</td>
<td>1.5553</td>
<td>2</td>
<td>4.8315</td>
<td>8</td>
<td>2.3259</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.8635</td>
<td>3</td>
<td>4.4172</td>
<td>8</td>
<td>1.6716</td>
<td>17</td>
<td>0.2897</td>
<td>12</td>
<td>2.9957</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>1.2929</td>
<td>2</td>
<td>4.7808</td>
<td>14</td>
<td>2.8700</td>
<td>14</td>
<td>0.0775</td>
<td>14</td>
<td>3.5381</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>2.4691</td>
<td>18</td>
<td>0.0501</td>
<td>13</td>
<td>2.5107</td>
<td>15</td>
<td>1.4290</td>
<td>7</td>
<td>4.3645</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>2.3665</td>
<td>19</td>
<td>0.4828</td>
<td>4</td>
<td>3.4976</td>
<td>17</td>
<td>1.1207</td>
<td>3</td>
<td>4.8705</td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>0.1757</td>
<td>30</td>
<td>1.3767</td>
<td>21</td>
<td>2.5510</td>
<td>18</td>
<td>3.0002</td>
<td>13</td>
<td>4.9056</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>5.3433</td>
<td>9</td>
<td>7.7000</td>
<td>24</td>
<td>0.5536</td>
<td>26</td>
<td>1.9902</td>
<td>26</td>
<td>2.9126</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>3.2164</td>
<td>21</td>
<td>4.9798</td>
<td>15</td>
<td>5.6587</td>
<td>8</td>
<td>8.1000</td>
<td>31</td>
<td>0.7395</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>1.5269</td>
<td>18</td>
<td>2.2321</td>
<td>16</td>
<td>3.8852</td>
<td>19</td>
<td>4.6265</td>
<td>20</td>
<td>5.1209</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>8.3000</td>
<td>25</td>
<td>0.5112</td>
<td>23</td>
<td>1.4782</td>
<td>25</td>
<td>2.9345</td>
<td>22</td>
<td>3.5054</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>4.6179</td>
<td>11</td>
<td>5.6969</td>
<td>17</td>
<td>6.0930</td>
<td>24</td>
<td>0.7300</td>
<td>27</td>
<td>1.1580</td>
</tr>
<tr>
<td>Mid</td>
<td>30</td>
<td>0.4692</td>
<td>39</td>
<td>1.7308</td>
<td>44</td>
<td>2.0897</td>
<td>35</td>
<td>3.9257</td>
<td>25</td>
<td>4.2667</td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>5.4819</td>
<td>23</td>
<td>6.6846</td>
<td>18</td>
<td>7.5116</td>
<td>19</td>
<td>8.1322</td>
<td>12</td>
<td>9.5721</td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>0.0866</td>
<td>35</td>
<td>1.2369</td>
<td>30</td>
<td>2.5320</td>
<td>30</td>
<td>3.4079</td>
<td>33</td>
<td>4.6050</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>5.3677</td>
<td>26</td>
<td>6.5608</td>
<td>36</td>
<td>7.2420</td>
<td>18</td>
<td>8.5036</td>
<td>7</td>
<td>9.3909</td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>0.7648</td>
<td>39</td>
<td>1.1764</td>
<td>34</td>
<td>2.3251</td>
<td>31</td>
<td>3.2228</td>
<td>24</td>
<td>4.1361</td>
</tr>
</tbody>
</table>
Linear Mixed Models

Violation of independence in Linear Models

- sampling process
- cluster trials (randomized units: families, countries, factories)
- multilevel (classes into schools into towns, repetitions into block into trials)
- ... 
- longitudinal or repeated data
- spatial data

Then, $\mathbf{Y}$ comes from the multivariate gaussian parametric family density:

$$
\mathcal{F} = \left\{ \frac{1}{(2\pi)^{n/2}} \det(V)^{-1/2} e^{-\frac{1}{2}(\mathbf{Y} - \mathbf{X}\beta)'V^{-1}(\mathbf{Y} - \mathbf{X}\beta)} \right\}
$$

dimension of $V$ is $n(n + 1)/2$

difficult to specify directly these elements

$\Rightarrow$ with random effects we can deal with variances and covariances.
Eucalyptus example

Example

Breeding program of Eucalypt. Ten families of trees full-sibs have been measured each year. There are three blocks, and 36 individuals per family in each block. Let assume that

\[ Y_{ijk} = \mu + B_k + F_j + \varepsilon_{ijk} \]

with \( B_k, k = 1, 2, 3 \) random block effects, \( F_j, j = 1, \ldots, 10 \) random family effects. Then

- **variances**: \( \text{Var}(Y_{ijk}) = \sigma_B^2 + \sigma_F^2 + \sigma_\varepsilon^2 \)
- **covariances**:
  \[
  \begin{align*}
  \text{cov}(Y_{ijk}, Y_{i'j'k}) &= \sigma_B^2 + \sigma_F^2 \\
  \text{cov}(Y_{ijk}, Y_{i'j'k}) &= \sigma_B^2 \quad \text{and} \quad \text{cov}(Y_{ijk}, Y_{i'j'k'}) = \sigma_F^2 \\
  \text{cov}(Y_{ijk}, Y_{i'j'k'}) &= 0
  \end{align*}
  \]
Definition

Let $\mathbf{y}$ denote the observation vector, for $n$ individuals ($\mathbf{y}$ a realization of random vector $\mathbf{Y}$).
Let $\mathbf{\beta}$ be a vector of fixed unknown parameters and $\mathbf{u}$ be the unobserved realized vector of the random effect vector $\mathbf{U}$.
Let $\mathbf{X}$ and $\mathbf{Z}$ be two known model matrices.

$\mathbf{Y}$ is modeled by a linear mixed model when

1. $\mathbf{Y}$ comes from the multivariate gaussian parametric density family
2. the conditional expectation $\mathbb{E}(\mathbf{Y}|\mathbf{U} = \mathbf{u})$ is related to the predictor $\eta = \mathbf{X}\mathbf{\beta} + \mathbf{Zu}$ through the identity link: $\mathbb{E}(\mathbf{Y}|\mathbf{U} = \mathbf{u}) = \eta$
3. the conditional variance: $\mathbb{V}\
\text{(}\mathbf{Y}|\mathbf{U} = \mathbf{u}) = R$
4. $\mathbf{U} \sim \mathcal{N}(0, D)$ and $D$ is an unknown covariance matrix

This is commonly written:

$$\mathbf{Y} = \mathbf{X}\mathbf{\beta} + \mathbf{Zu} + \mathbf{\varepsilon}$$

with $\mathbf{U}$ and $\mathbf{\varepsilon}$ independent.
Marginal moments

- marginal expectation

\[ \mathbb{E}(\mathbf{Y}) = \mathbb{E}_u[\mathbb{E}(\mathbf{Y}|\mathbf{U})] = \mathbb{E}_u(X\beta + Z\mathbf{U}) = X\beta \]

- marginal variance

\[ \mathbb{V}(\mathbf{Y}) = \mathbb{E}_u(\mathbb{V}(\mathbf{Y}|\mathbf{U})) + \mathbb{V}_u(\mathbb{E}(\mathbf{Y}|\mathbf{U})) = R + \mathbb{V}_u(X\beta + Z\mathbf{U}) = R + ZDZ' \]

- marginal distribution of \( \mathbf{Y} \)

\[ \mathbf{Y} \sim \mathcal{N}(X\beta, ZDZ' + R) \]
Estimation for $V$ known

Coming back to the very general case $\mathbf{Y} \sim \mathcal{N}(\mathbf{X}\beta, \mathbf{V})$

Maximum likelihood estimator of $\beta$?

- Log-likelihood
  \[
  \ell(\beta; \mathbf{y}) = -\frac{1}{2} (\mathbf{y} - \mathbf{X}\beta)' \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta) - \frac{1}{2} \log [\text{det}(\mathbf{V})] - \frac{n}{2} \log(2\pi)
  \]

- Score function
  \[
  U(\beta) = \frac{\partial \ell(\beta; \mathbf{y})}{\partial \beta} = \mathbf{X}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta)
  \]

- Solution of $U(\beta) = 0$
  \[
  \beta^0 = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}
  \]

  where $A^{-}$ is the generalized inverse of $A$

- Theorem (Kruskal):
  \[
  \beta^0 = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} \quad \text{iff} \quad \mathbf{V}\mathbf{X} = \mathbf{X}
  \]
Properties of $\beta^0$ and tests

**Properties**

- **expectation**
  \[ E(\beta^0) = \beta \]

- **variance**
  \[ V(\beta^0) = (X'V^{-1}X)^{-1} \]

- **multivariate gaussian distribution**
  \[ \beta^0 \sim \mathcal{N}_K [\beta, (X'V^{-1}X)^{-1}] \]

**Tests**

- **$H_0 : L\beta = m$**

- **Test statistics**:
  \[ X^2 = (L\beta^0 - m)'[L(X'V^{-1}X)^{-1}L']^{-1}(L\beta^0 - m) \]

- **Distribution**:
  \[ X^2 \sim \chi^2(r(L)) \]
Estimation for $V$ known up to a constant

Let $V = \sigma^2 W$, $W$ known, $\sigma^2$ unknown

Maximum likelihood estimator of $\beta$ and $\sigma^2$?

- Log-likelihood

$$
\ell(\beta; y) = -\frac{1}{2\sigma^2} (y - X\beta)' W^{-1} (y - X\beta) - \frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2} \log[\det(W)]
$$

- Score function for $\beta$

$$
U(\beta; \hat{\sigma}^2) = \frac{\partial \ell(\beta, \sigma^2; y)}{\partial \beta} = X'(\hat{\sigma}^2 W)^{-1} (y - X\beta)
$$

- Score function for $\sigma^2$

$$
U(\sigma^2; \hat{\beta}) = \frac{\partial \ell(\beta, \sigma^2; y)}{\partial \sigma^2} = \frac{1}{2(\sigma^2)^2} (y - X\hat{\beta})' W^{-1} (y - X\hat{\beta}) - \frac{n}{2\sigma^2}
$$

- Solution of $U(\beta; \hat{\sigma}^2) = 0$ and $U(\sigma^2; \hat{\beta}) = 0$

$$
\beta^0 = (X' W^{-1} X)^{-1} X' W^{-1} Y
$$

$$
\hat{\sigma}^2_{ml} = \frac{1}{n} (Y - X\beta^0)' W^{-1} (Y - X\beta^0)
$$
Properties and tests

Properties

- $\beta^0 \sim \mathcal{N}_K [\beta, \sigma^2 (X'W^{-1}X)^{-}]$
- $E(\hat{\sigma}^2_{ml}) = \frac{\sigma^2 (n - r(X))}{n}$
  - $\hat{\sigma}^2_{ml}$ is biased but asymptotically unbiased
- $\hat{\sigma}^2_{ls} = \frac{n}{n - r(X)} \hat{\sigma}^2_{ml}$ is unbiased
- $\hat{\sigma}^2_{ls} \sim \frac{\sigma^2}{n - r(X)} \chi^2(n - r(X))$

Tests

$H_0 : L\beta = m$

Test statistics: $F = \frac{(L\beta^0 - m)'[L(X'W^{-1}X)^{-}L']^{-1}(L\beta^0 - m)}{r(L)\hat{\sigma}^2_{ls}}$

Distribution: $F \sim F(r(L), n - r(X))$
Geometrical interpretation with $W = I_d$

$\hat{Y} = X\hat{\beta} \in \text{Vec}(X)$

$\dim(\text{Vec}(X)) = r(X)$

$\hat{Y}$ projection of $Y$ on $\text{Vec}(X)$

$\hat{\epsilon} = (Y - \hat{Y}) \in \text{Vec}^\perp(X)$

$\dim(\text{Vec}^\perp(X)) = n - r(X)$

$P = (Id - X(X'X)^{-1}X')$

$\hat{\sigma}^2_{ls} = \frac{\hat{\epsilon}'\hat{\epsilon}}{n-r(X)}$

$\hat{\sigma}^2_{ml} = \frac{\hat{Y}'P\hat{Y}}{n} = \frac{\hat{\epsilon}'\hat{\epsilon}}{n}$

$\leftrightarrow$ ML does not take into account $\dim(\text{Vec}^\perp(X))$

$\leftrightarrow$ REML estimation
Restricted likelihood

Let $L$ such as $LX = 0$

- Restricted log-likelihood of $y$ (log-likelihood of $Ly$)

$$-2\ell_{\text{reml}}(\sigma^2; y) = y' L'(L\sigma^2 WL')^{-1} Ly + \log [\det(L\sigma^2 WL')] + (n - r(X)) \log (2\pi)$$

$$= C_1 + y' Py + \log [\det(\sigma^2 W)] + \log \{\det[X'(\sigma^2 W)^{-1} X]\}$$

$$= C_2 + y' Py + n \log(\sigma^2) - r(X) \log(\sigma^2)$$

$$= C_2 + y' Py + (n - r(X)) \log(\sigma^2)$$

$$= C_2 + \frac{1}{\sigma^2} y' \tilde{P} y + (n - r(X)) \log(\sigma^2)$$

where

$$C_1 = (n - r(X)) \log(2\pi) - \log \det(X'X) + \log \det(LL')$$

$$C_2 = C_1 + \log \det(W) + \log \{\det[X' W^{-1} X]\}$$

$$P = V^{-1} - V^{-1} X (X' V^{-1} X)^{-1} X' V^{-1}$$

$$\tilde{P} = W^{-1} - W^{-1} X (X' W^{-1} X)^{-1} X' W^{-1}$$
• Score function

\[ U_{reml}(\sigma^2) = \frac{1}{2(\sigma^2)^2} y^\prime \tilde{P} y - \frac{n - r(X)}{2\sigma^2} \]

• Solution

\[ \hat{\sigma}^2_{reml} = \frac{y^\prime \tilde{P} y}{n - r(X)} \]
ML estimation for $V$ unknown

Let $Y \sim \mathcal{N}(X\beta, V)$
Let assume that $V$ depends on a parameter vector $\phi$: $V = V(\phi)$

Maximum likelihood estimator of $\beta$ and $\phi$?

- **Score functions**

$$U(\beta) = \frac{\partial \ell(\beta; \hat{\phi}, y)}{\partial \beta} = X'\hat{V}^{-1}(y - X\beta), \quad \text{where} \quad \hat{V} = V(\hat{\phi})$$

$$-2U(\phi) = -2\frac{\partial \ell(\phi; \hat{\beta}, y)}{\partial \phi} = tr(V^{-1}\frac{\partial V}{\partial \phi}) - (y - X\hat{\beta})'V^{-1}\frac{\partial V}{\partial \phi}V^{-1}(y - X\hat{\beta})$$

- **Solutions**

1. $$\hat{\beta} = (X'\hat{V}^{-1}X)^{-}X'\hat{V}^{-1}Y$$

2. $$tr(V^{-1}\frac{\partial V}{\partial \phi}) = y'P\frac{\partial V}{\partial \phi}Py$$

where $P = V^{-1} - V^{-1}X(X'V^{-1}X)^{-}X'V^{-1}$.

$\leftrightarrow$ not a closed form; use iterative algorithms
**Properties**

1. \((X' \hat{V}^{-1} X)\) is not an estimator of \(\text{Var}(\hat{\beta})\): we should take into account variability in \(\hat{V}\)

\[
\text{Var}(\hat{\beta} - \beta) = \text{Var}(\beta^0 - \beta) + \text{Var}(\hat{\beta} - \beta^0) \geq \text{Var}(\beta^0 - \beta)
\]

2. \(\text{Var}_\infty(\hat{\beta}, \hat{\phi}) = \mathcal{I}^{-1}
\]

with

\[
\mathcal{I} = \begin{pmatrix}
X' \hat{V}^{-1} X & 0 \\
0 & \left(\frac{1}{2} \text{tr} \left( V^{-1} \frac{\partial V}{\partial \phi_m} V^{-1} \frac{\partial V}{\partial \phi_{m'}} \right) \right)_{m,m' = 1,\ldots,M}
\end{pmatrix}
\]

the Fisher information matrix \(\mathcal{I} = \mathbb{E} \left( - \frac{\partial^2 l(\beta, \phi; y)}{\partial \gamma \partial \gamma'} \right)\) with \(\gamma = (\beta, \phi)'

and \(\hat{\beta}\) assymptotically gaussian
Test

$H_0 : L\beta = m$

Test statistics: $F = \frac{(L\beta^0 - m)'[L(X' \hat{V}^{-1}X) - L']^{-1}(L\beta^0 - m)}{r(L)}$

Approximate distribution $F \sim F(r(L), \tilde{df})$

- $\tilde{df} = n - r(X)$ if $V$ known up to a constant
- many corrections of degrees of freedom $\tilde{df}$ have been proposed
  - Satterthwaite (1941)
  - Kenward and Roger (1997)
  - ...
REML estimation for $V$ unknown

Restricted maximum likelihood estimator of $\phi$?

- Restricted likelihood
  $$-2\ell_{\text{reml}}(\sigma^2; y) = y'Py + \log(\det(V)) + \log(\det(X'V^{-1}X))$$

- Score functions
  $$-2U_{\text{reml}}(\phi) = -2\frac{\partial \ell_{\text{reml}}(\phi; \hat{\beta}, y)}{\partial \phi} = \text{tr}(P \frac{\partial V}{\partial \phi}) - y'P \frac{\partial V}{\partial \phi}Py$$

- Solution
  $$\text{tr}(P \frac{\partial V}{\partial \phi}) = y'P \frac{\partial V}{\partial \phi}Py$$

where $P = V^{-1} - V^{-1}X(X'V^{-1}X)^{-}X'V^{-1}$.

$\leftrightarrow$ not a closed form; use iterative algorithms

Remark 1: $\mathbb{E}(U_{\text{reml}}(\phi; Y)) = 0$

Remark 2: $\hat{\beta} = (X'\hat{V}_{\text{reml}}^-1X)^{-}X'\hat{V}_{\text{reml}}^-1y$
Properties

\[ \text{Var}_\infty(\hat{\beta}, \hat{\phi}) = \mathcal{I}^{-1} \]

with

\[ \mathcal{I} = \begin{pmatrix} X'V^{-1}X & 0 \\ 0 & \left( \frac{1}{2} \text{tr}(P \frac{\partial V}{\partial \phi_m} P \frac{\partial V}{\partial \phi_{m'}}) \right)_{m,m' = 1, \ldots, M} \end{pmatrix} \]

and \( \hat{\beta} \) asymptotically gaussian.
Estimation in the linear mixed model

Coming back to the linear mixed model definition:

\[ Y = X\beta + \sum_{l=1}^{L} Z_l u_l + \epsilon \]

with \( L \) different random effects \( u_l \sim \mathcal{N}(0, \sigma^2_l I_{q_l}) \)
and \( \epsilon \sim \mathcal{N}(0, \sigma^2_0 I_n) \)

Parameters to be estimated: \( \beta \) and \( \sigma^2_0, \sigma^2_1, \ldots, \sigma^2_L \).

Conditional moments

\[
\mathbb{E}(Y|u_1, \ldots, u_L) = X\beta + \sum_{l=1}^{L} Z_l u_l \\
\mathbb{V}(Y|u_1, \ldots, u_L) = \sigma^2_0 I_n
\]
Marginal moments

\[ E(Y) = X\beta \]

\[ V(Y) = \sum_{l=1}^{L} \sigma_l^2 Z_l Z_l' + \sigma_0^2 I_n \]

\[ = \sum_{l=0}^{L} \sigma_l^2 V_l \text{ with } V_0 = I_n \text{ and } V_l = Z_l Z_l' \]

\( \sigma^2 = (\sigma_0^2, \sigma_1^2, \ldots, \sigma_L^2) \) are called \textbf{variance components}

\( \leftrightarrow V \) is a linear combination of variance components

\( \leftrightarrow \frac{\partial V}{\partial \sigma_l^2} = V_l \text{ for all } l \in \{1, \ldots, L\}. \)
ML estimation

1. \( \hat{\beta}_{ml} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}Y \)

2. \( tr(V^{-1}V_l) = y'PV_lPy \)

i.e. \( \sum_{m=1}^{L} tr(V^{-1}V_lV^{-1}V_m)\sigma_m^2 = y'PV_lPy \)

\( \leftrightarrow \) iterative linear system of equations:

\[
\begin{pmatrix}
tr(\hat{V}^{-1}V_l\hat{V}^{-1}V_m) \\
\sigma_m^2
\end{pmatrix}_{l,m=1,...,L}
= 
\begin{pmatrix}
y'\hat{P}V_m\hat{P}y
\end{pmatrix}_{m=1,...,L}
\]

REML estimation

1. \( \hat{\beta}_{reml} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}Y \)

2. \( tr(PV_l) = y'PV_lPy \)

i.e. \( \sum_{m=1}^{L} tr(PV_lPV_m)\sigma_m^2 = y'PV_lPy \)

\( \leftrightarrow \) iterative linear system of equations:

\[
\begin{pmatrix}
tr(\hat{P}V_l\hat{P}V_m) \\
\sigma_m^2
\end{pmatrix}_{l,m=1,...,L}
= 
\begin{pmatrix}
y'\hat{P}V_m\hat{P}y
\end{pmatrix}_{m=1,...,L}
\]
Properties

- **ML estimation**
  
  $\text{Var}_\infty(\hat{\beta}_{ml}, \hat{\sigma}_{ml}^2) = \mathcal{I}^{-1}$

  with
  
  $\mathcal{I} = \begin{pmatrix}
  X'V^{-1}X & 0 \\
  0 & \left(\frac{1}{2} \text{tr}(V^{-1}V_lV^{-1}V_m)\right)_{l,m=1,...,L}
  \end{pmatrix}$

- **REML estimation**
  
  $\text{Var}_\infty(\hat{\beta}_{reml}, \hat{\sigma}_{reml}^2) = \mathcal{I}^{-1}$

  with
  
  $\mathcal{I} = \begin{pmatrix}
  X'V^{-1}X & 0 \\
  0 & \left(\frac{1}{2} \text{tr}(PV_lPV_m)\right)_{l,m=1,...,L}
  \end{pmatrix}$
Test model $M_0$ embeded in $M_1$

($M_0$ and $M_1$ have the same variance structure)

with $m_0 = \text{dim}(\text{Vec}(X_0))$ and $m_1 = \text{dim}(\text{Vec}(X_1))$ ($m_0 < m_1$).

Let $\ell_{M_0}$ and $\ell_{M_1}$ denote the associated log-likelihood calculated at the ML-estimated parameter values (not REML).

Then

$$X^2 = -2\ell_{M_0} + 2\ell_{M_1} \sim_{H_0} \chi^2_{m_1-m_0}$$
Comparison of models

- Akaike’s Information Criterion
  \[ AIC = -2 \log L(\hat{\beta}_{ml}, \hat{\sigma}_{ml}, y) + 2K \]

- Bayesian Information Criterion
  \[ BIC = -2 \log L(\hat{\beta}_{ml}, \hat{\sigma}_{ml}, y) + K \log(n) \]

- Other criterions with other penalty terms

**Danger !!!** Softwares often give AIC or BIC calculated at reml-estimated parameter values.
\( \sigma^2 = (\sigma_0^2, \ldots, \sigma_L^2) \) are solution of the non-linear equations:

\[
\hat{F}(\hat{\sigma}^2)\hat{\sigma}^2 = \hat{g}(\hat{\sigma}^2)
\]

where \( \hat{F}(\sigma)^2 \) is \((L + 1) \times (L + 1)\) matrix

\[
\hat{F}\hat{\sigma}^2 = \sum_{l}^L tr(\hat{V}^{-1}V_l'\hat{V}^{-1}V_l), \quad l' = 0; \ldots L
\]

and

\[
\hat{g}(\hat{\sigma}^2) = y'\hat{P}V_l'\hat{P}y
\]

Iterative algorithm should be employed
β are unknown but fixed parameters to be estimated

\( u \) are unobserved realized values of the latent random vector \( \mathbf{U} \)

\( \mathbf{U} \rightarrow \) estimation of \( \mathbf{U} \) has no sense but prediction has!

As the joint distribution \( (Y, \mathbf{U}) \) is gaussian

\[
\mathbf{\hat{U}} = E(\mathbf{U}|Y) = DZ'V^{-1}(Y - X\beta)
\]

is the best linear predictor (BLP) in the sense of minimizing the mean square error

- if \( V \) is known,
  \[
  \mathbf{U}^0 = DZ'V^{-1}(Y - X\beta^0)
  \]
  is the best Unbiased linear predictor (BLUP)

- if \( V \) is unknown
  \[
  \mathbf{\hat{U}} = \hat{D}Z'\hat{V}^{-1}(Y - X\hat{\beta})
  \]
  is the estimate of the best linear predictor
ANOVA with random effects

Example

2 varieties (Resistant/Susceptible)
4 pesticide doses
5 blocks
40 observations of a response $Y$

$Y_{ijk}$ response for the $i^{th}$ dose $j^{th}$ variety and $k^{th}$ block.

Evaluation of the influence of varieties, pesticide doses and interaction on the response.

Take into account the variability between blocks and between interactions block $\times$ dose.
Example

\[ Y_{ijk} = \mu + \delta_i + \gamma_j + (\delta \gamma)_{ij} + r_k + w_{ik} + \varepsilon_{ijk} \]

\( r_k \sim \mathcal{N}(0, \sigma_r^2 I_5), \ w_{ik} \sim \mathcal{N}(0, \sigma_w^2 I_{20}) \) and \( \varepsilon_{ijk} \sim \mathcal{N}(0, \sigma^2 I_{40}) \) mutually independent.

\[
\text{cov}(Y_{ijk}, Y_{i'j'k'}) = \begin{cases} 
\sigma_r^2 + \sigma_w^2 + \sigma^2 & \text{if } k = k', \ i = i' \text{ and } j = j' \\
\sigma_r^2 + \sigma_w^2 & \text{if } k = k' \text{ and } i = i' \\
\sigma_r^2 & \text{if } k = k' \\
0 & \text{otherwise}
\end{cases}
\]
```r
> variety <- read.table("DataSets/Variety_eval.csv", sep=";",
+ header=T, colClasses=c(rep("factor",3),rep("numeric",2)))

> interaction.plot(variety$dose, variety$type, variety$y)
> plot(variety$y ~ I(variety$dose:variety$type))
```

![Graph showing interaction plot and box plots for variety data.](image)
Basic R code

```r
> res.reml <- lmer(y ~ dose*type + (1|block) + (1|block:dose),
+ data=variety,REML = TRUE)
```

res.reml is a mer object

- methods are available (e.g. `fixef(res.reml), VarCorr(res.reml))`
- slots are available (e.g. `res.reml@ranef, res.reml@mu`)

Summary

```r
> res.sum <- summary(res.reml)
> print(c(slotNames(res.sum)[1:5],"...."))

[1] "methTitle" "logLik" "ngrps" "sigma" "coefs" "...."
```
Summary

> res.sum@coefs

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>20.00</td>
<td>1.476828</td>
</tr>
<tr>
<td>dose2</td>
<td>7.84</td>
<td>1.879586</td>
</tr>
<tr>
<td>dose4</td>
<td>8.18</td>
<td>1.879586</td>
</tr>
<tr>
<td>dose8</td>
<td>4.80</td>
<td>1.879586</td>
</tr>
<tr>
<td>types</td>
<td>-2.94</td>
<td>1.314363</td>
</tr>
<tr>
<td>dose2:types</td>
<td>0.30</td>
<td>1.858791</td>
</tr>
<tr>
<td>dose4:types</td>
<td>3.14</td>
<td>1.858791</td>
</tr>
<tr>
<td>dose8:types</td>
<td>3.94</td>
<td>1.858791</td>
</tr>
</tbody>
</table>

> data.frame(res.sum@REmat)

<table>
<thead>
<tr>
<th>Groups</th>
<th>Name</th>
<th>Variance</th>
<th>Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 block:dose</td>
<td>(Intercept)</td>
<td>4.5132</td>
<td>2.1244</td>
</tr>
<tr>
<td>2 block</td>
<td>(Intercept)</td>
<td>2.0735</td>
<td>1.4400</td>
</tr>
<tr>
<td>3 Residual</td>
<td></td>
<td>4.3189</td>
<td>2.0782</td>
</tr>
</tbody>
</table>
### Random effects

<table>
<thead>
<tr>
<th>Groups</th>
<th>var. ML</th>
<th>std. ML</th>
<th>var. REML</th>
<th>std. REML</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 block:dose</td>
<td>3.6106</td>
<td>1.9002</td>
<td>4.5132</td>
<td>2.1244</td>
</tr>
<tr>
<td>2 block</td>
<td>1.6588</td>
<td>1.2879</td>
<td>2.0735</td>
<td>1.4400</td>
</tr>
<tr>
<td>3 Residual</td>
<td>3.4551</td>
<td>1.8588</td>
<td>4.3189</td>
<td>2.0782</td>
</tr>
</tbody>
</table>

### Fixed effects

<table>
<thead>
<tr>
<th></th>
<th>Est. ML</th>
<th>t.value. ML</th>
<th>Est. REML</th>
<th>t.value. REML</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>20.00</td>
<td>15.14</td>
<td>20.00</td>
<td>13.54</td>
</tr>
<tr>
<td>dose2</td>
<td>7.84</td>
<td>4.66</td>
<td>7.84</td>
<td>4.17</td>
</tr>
<tr>
<td>dose4</td>
<td>8.18</td>
<td>4.87</td>
<td>8.18</td>
<td>4.35</td>
</tr>
<tr>
<td>dose8</td>
<td>4.80</td>
<td>2.86</td>
<td>4.80</td>
<td>2.55</td>
</tr>
<tr>
<td>types</td>
<td>-2.94</td>
<td>-2.50</td>
<td>-2.94</td>
<td>-2.24</td>
</tr>
<tr>
<td>dose2:types</td>
<td>0.30</td>
<td>0.18</td>
<td>0.30</td>
<td>0.16</td>
</tr>
<tr>
<td>dose4:types</td>
<td>3.14</td>
<td>1.89</td>
<td>3.14</td>
<td>1.69</td>
</tr>
<tr>
<td>dose8:types</td>
<td>3.94</td>
<td>2.37</td>
<td>3.94</td>
<td>2.12</td>
</tr>
</tbody>
</table>
Tests for fixed effects: each parameters

Use maximum likelihood: REML=FALSE

- t-test approach, with n-p degrees of freedom (40-8)

<table>
<thead>
<tr>
<th></th>
<th>Est. ML</th>
<th>t.value.ML</th>
<th>p.val</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>20.00</td>
<td>15.14</td>
<td>0.0000</td>
</tr>
<tr>
<td>dose2</td>
<td>7.84</td>
<td>4.66</td>
<td>0.0001</td>
</tr>
<tr>
<td>dose4</td>
<td>8.18</td>
<td>4.87</td>
<td>0.0000</td>
</tr>
<tr>
<td>dose8</td>
<td>4.80</td>
<td>2.86</td>
<td>0.0075</td>
</tr>
<tr>
<td>types</td>
<td>-2.94</td>
<td>-2.50</td>
<td>0.0177</td>
</tr>
<tr>
<td>dose2:types</td>
<td>0.30</td>
<td>0.18</td>
<td>0.8579</td>
</tr>
<tr>
<td>dose4:types</td>
<td>3.14</td>
<td>1.89</td>
<td>0.0680</td>
</tr>
<tr>
<td>dose8:types</td>
<td>3.94</td>
<td>2.37</td>
<td>0.0240</td>
</tr>
</tbody>
</table>
Tests for fixed effects: each parameters

- resampling approach: available for simple linear mixed models

```r
> ic <- mcmcsamp(res.ml,n = 100)
> ic <- ic@fixef
> ic <- t(apply(ic,1,function(x) quantile(x,prob=c(0.025,0.975))))
> xtable(ic)

<table>
<thead>
<tr>
<th></th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>16.83</td>
<td>23.12</td>
</tr>
<tr>
<td>dose2</td>
<td>4.12</td>
<td>11.44</td>
</tr>
<tr>
<td>dose4</td>
<td>4.30</td>
<td>11.64</td>
</tr>
<tr>
<td>dose8</td>
<td>0.77</td>
<td>8.50</td>
</tr>
<tr>
<td>types</td>
<td>-5.78</td>
<td>1.14</td>
</tr>
<tr>
<td>dose2:types</td>
<td>-4.16</td>
<td>5.12</td>
</tr>
<tr>
<td>dose4:types</td>
<td>-1.73</td>
<td>7.88</td>
</tr>
<tr>
<td>dose8:types</td>
<td>-1.03</td>
<td>8.52</td>
</tr>
</tbody>
</table>
```
Tests for fixed effects: global test

- F-test approach, with n-p degrees of freedom (40-8) using \texttt{anova(res.reml)} R command

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F.value</th>
<th>p.val</th>
</tr>
</thead>
<tbody>
<tr>
<td>dose</td>
<td>3</td>
<td>176.54</td>
<td>58.85</td>
<td>13.625</td>
<td>0.0000</td>
</tr>
<tr>
<td>type</td>
<td>1</td>
<td>11.99</td>
<td>11.99</td>
<td>2.776</td>
<td>0.1054</td>
</tr>
<tr>
<td>dose:type</td>
<td>3</td>
<td>29.64</td>
<td>9.88</td>
<td>2.288</td>
<td>0.0974</td>
</tr>
</tbody>
</table>

- balanced case: exact F test

\begin{verbatim}
Error: block
  Df  Sum Sq Mean Sq F value Pr(>F)
Residuals 4 119.73  29.933

Error: block:dose
  Df Sum Sq Mean Sq F value Pr(>F)
dose  3 545.50 181.835 13.625 0.0003595 ***
Residuals 12 160.14  13.345
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
\end{verbatim}

\begin{verbatim}
Error: block:dose:type
  Df Sum Sq Mean Sq F value Pr(>F)
type  1  11.99  11.990  2.7762 0.1151
dose:type  3  29.64  9.8809  2.2878 0.0974
Residuals 16  69.10  4.3189
\end{verbatim}
we want to compare different nested models: use ML (not REML)

- full model: dose*type
- additive model (dose+type) versus full model
- an approximated test using chi-square distribution:
  ```r
  > compar <- anova(res.ml, additif)
  ```

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>additif</td>
<td>8</td>
<td>212.85</td>
<td>226.36</td>
<td>-98.43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>res.ml</td>
<td>11</td>
<td>211.71</td>
<td>230.29</td>
<td>-94.86</td>
<td>7.14</td>
<td>3</td>
<td>0.0676</td>
</tr>
</tbody>
</table>

- AIC or BIC criterion
  ```r
  > AICTab <- rbind(summary(additif)@AICtab, summary(res.ml)@AICtab)
  > rownames(AICTab) <- c("additif", "res.ml")
  ```
Predictions

- conditional: \( E(Y|U) = X\beta + Zu \)
  > \texttt{pred.cond <- res.sum@mu}

- marginal: \( E(Y|U) = X\beta \)
  > \texttt{pred.mar <- res.sum@X%*%res.sum@coefs[,1]}
ANCOVA with random effects

Example

32 steers dispatched into
8 farms
4 diet treatments
each steer in farm received one treatment

influence of the initial weight of steers

32 observations of a response $Y$

$Y_{ij}$ response for the $i^{th}$ diet treatment $j^{th}$ farm.

Determine the optimal level of feed additive to maximize the average daily gain of steer.
ANCOVA with random farm effects

\[ Y_{ij} = dtrt_i + \gamma w_{ij} + \text{farm}_j + \varepsilon_{ij} \]

\( \text{farm}_j \sim \mathcal{N}(0, \sigma_f^2 l_8) \), \( \varepsilon_{ij} \sim \mathcal{N}(0, \sigma_i^2 l_{32}) \) mutually independent.

\[
\text{cov}(Y_{ij}, Y_{i'j'}) = \begin{cases} 
\sigma_f^2 + \sigma^2 & \text{if } i = i' \text{ and } j = j' \\
\sigma_f^2 & \text{if } j = j' \\
0 & \text{otherwise}
\end{cases}
\]
Pairwise comparison of treatments via Tukey

```r
> steers.ml <- lmer(adg~dtrt+iwt + (1|farm), data=steers,REML=FALSE)
> library(multcomp)
> summary(glht(steers.ml, linfct = mcp(dtrt = "Tukey")))
```

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: lmer(formula = adg ~ dtrt + iwt + (1 | farm), data = steers, REML = FALSE)

Linear Hypotheses:

| Estimate | Std. Error | z value | Pr(>|z|) |
|----------|------------|---------|----------|
| 10 - 0 == 0 | 0.48366    | 0.10308 | 4.692    | <1e-04 *** |
| 20 - 0 == 0 | 0.46401    | 0.10235 | 4.534    | <1e-04 *** |
| 30 - 0 == 0 | 0.55180    | 0.10487 | 5.262    | <1e-04 *** |
| 20 - 10 == 0 | -0.01966   | 0.10251 | -0.192   | 0.998    |
| 30 - 10 == 0 | 0.06814    | 0.10867 | 0.627    | 0.923    |
| 30 - 20 == 0 | 0.08779    | 0.10622 | 0.827    | 0.412    |

---

Signif. codes:  0 Š***ˇ Š**ˇ Š*ˇ Š.ˇ Š 1
(Adjusted p values reported -- single-step method)
ANCOVA with random farm effects and specific fixed weight slope

\[ Y_{ij} = dtrt_i + \gamma_i w_{ij} + \text{farm}_j + \varepsilon_{ij} \]

\[ \text{farm}_j \sim \mathcal{N}(0, \sigma_f^2 I_8), \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2 I_{32}) \text{ mutually independent.} \]

\[ \text{cov}(Y_{ij}, Y_{i'j'}) = \begin{cases} 
\sigma_f^2 + \sigma^2 & \text{if } i = i' \text{ and } j = j' \\
\sigma_f^2 & \text{if } j = j' \\
0 & \text{otherwise}
\end{cases} \]
> steers2.ml <- lmer(adg~dtrt*iwt + (1|farm), data=steers, REML=FALSE)
> anova(steers.ml,steers2.ml)

Data: steers
Models:
steers.ml:  adg ~  dtrt +  iwt + (1 |  farm)
steers2.ml:  adg ~  dtrt *  iwt + (1 | farm)

Df    AIC   BIC  logLik  Chisq Chi Df Pr(>Chisq)
steers.ml 7 27.61737.877 -6.8086
steers2.ml 10 29.97144.628 -4.9855  3.6463 3 0.3023

> library(ggplot2)
> datapred <- data.frame(iwt= steers$iwt,pred=
+   steers.ml$X%*%summary(steers.ml)$coefs[,1],
+   dtrt=steers$dtrt)
> steers.plot <- ggplot(datapred, aes(y=pred,x=iwt,group=dtrt,
+   colour=dtrt))+geom_line()
Plan Models with fixed effects

Linear Mixed Models (LMM)

Generalized Linear Mixed Models (GLMM)

Longitudinal case

Overdispersion

Nonlinear models

C. Demetrio, F. Mortier & C. Trottier

Mixed Models, theory and applications
Random coefficient model

Example

10 varieties of winter wheat sampled in population
6 clones per varieties
60 combinations completely randomized in the field
pre-planting moisture at each plot have been measured
60 observations of a response $Y$ of wheat yield (in bushel)
$Y_{ij}$ response for the $i^{th}$ variety $j^{th}$ clone
moisture influences the wheat yield.
```r
> library(ggplot2)
> wheat <- read.table("DataSets/wheat.csv", sep=";", + colClasses=c("factor","numeric","numeric"),header=T)
> wheat.plot <- ggplot(wheat, aes(y=yield,x=moisture,group=wheat, + colour=wheat))+geom_line()
```
1. **fixed coefficients**

\[ Y_{ij} = \mu + \alpha w_{ij} + v_i + \gamma_i w_{ij} + \epsilon_{ij} \]

but varieties sampled from population,

2. **random variety effects**

\[ Y_{ij} = \mu + \alpha w_{ij} + v_i + \gamma_i w_{ij} + \epsilon_{ij} \]

where

\[
\begin{align*}
    v & \sim \mathcal{N}(0, \sigma_v^2 I_{10}) \\
    \gamma & \sim \mathcal{N}(0, \sigma_\gamma^2 I_{10})
\end{align*}
\]

→ are \( \gamma \) and \( v \) independent?
Plan Models with fixed effects

Linear Mixed Models (LMM)

Generalized Linear Mixed Models (GLMM)

Longitudinal case

Overdispersion

Nonlinear models

R code: independent case

```r
> wheat.ind <- lmer(yield~moisture + (1|wheat)
+   + (0+moisture|wheat), data=wheat, REML=FALSE)
> xtable(data.frame(summary(wheat.ind)@REmat))
```

<table>
<thead>
<tr>
<th>Groups</th>
<th>Name</th>
<th>Variance</th>
<th>Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>wheat</td>
<td>(Intercept)</td>
<td>16.3594335</td>
</tr>
<tr>
<td>2</td>
<td>wheat</td>
<td>moisture</td>
<td>0.0019991</td>
</tr>
<tr>
<td>3</td>
<td>Residual</td>
<td></td>
<td>0.3551692</td>
</tr>
</tbody>
</table>

R code: dependent case

```r
> wheat.dep <- lmer(yield~moisture + (1+moisture|wheat),
+   + data=wheat, REML=FALSE)
```

<table>
<thead>
<tr>
<th>Groups</th>
<th>Name</th>
<th>Variance</th>
<th>Std.Dev.</th>
<th>Corr</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>wheat</td>
<td>(Intercept)</td>
<td>16.9590713</td>
<td>4.118139</td>
</tr>
<tr>
<td>2</td>
<td>wheat</td>
<td>moisture</td>
<td>0.0021012</td>
<td>0.045839</td>
</tr>
<tr>
<td>3</td>
<td>Residual</td>
<td></td>
<td>0.3524533</td>
<td>0.593678</td>
</tr>
</tbody>
</table>
Comparing two models

\[
\text{xtable(anova(wheat.ind, wheat.dep))}
\]

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>Chisq</th>
<th>Chi</th>
<th>Df</th>
<th>Pr(&gt;Chi)</th>
</tr>
</thead>
<tbody>
<tr>
<td>wheat.ind</td>
<td>5</td>
<td>193.10</td>
<td>203.57</td>
<td>-91.55</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wheat.dep</td>
<td>6</td>
<td>194.06</td>
<td>206.62</td>
<td>-91.03</td>
<td>1.04</td>
<td>1</td>
<td></td>
<td>0.3076</td>
</tr>
</tbody>
</table>
Is random slope to be included?

\> xtable(data.frame(summary(wheat.ind)@REmat))

<table>
<thead>
<tr>
<th>Groups</th>
<th>Name</th>
<th>Variance</th>
<th>Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>wheat</td>
<td>(Intercept)</td>
<td>16.3594335</td>
</tr>
<tr>
<td>2</td>
<td>wheat</td>
<td>moisture</td>
<td>0.0019991</td>
</tr>
<tr>
<td>3</td>
<td>Residual</td>
<td></td>
<td>0.3551692</td>
</tr>
</tbody>
</table>

\> wheat.int <- lmer(yield~moisture + (1|wheat),
+ data=wheat, REML=FALSE)
\> xtable(anova(wheat.int,wheat.ind))

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>wheat.int</td>
<td>4</td>
<td>208.26</td>
<td>216.64</td>
<td>-100.13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wheat.ind</td>
<td>5</td>
<td>193.10</td>
<td>203.57</td>
<td>-91.55</td>
<td>17.16</td>
<td>1</td>
<td>0.0000</td>
</tr>
</tbody>
</table>
GLMM construction

Flexibility to account for:
- categorical, discrete ... non gaussian data ... distributions in the exponential family
  - mean non linearly related to the model parameters
  - variances of the data related to the mean
- random effect factors
- correlation structure in the data ... repeated data
- over-dispersion
Example

8 clinics sampled from a larger population

At clinic $j$, - $n_{j1}$ receive treatment 1
  - $n_{j2}$ receive treatment 2

The response $Y_{jk}$ at treatment $k$ and clinic $j$ is the number of ‘favorable’ responses.

$$f_{jk} = \frac{Y_{jk}}{n_{jk}}$$

is the proportion of ‘favorable’ responses.

$\longleftrightarrow$ binomial data in a multi-center clinical trial.
Model definition

- Two random sources in the data:
  - $u_j$ $j = 1, \ldots, q$ ($q$ clusters) $j^{th}$ random effect
  - $Y_{ijk}$ $k = 1, \ldots, n_{ij}$ $k^{th}$ observation for subject $i$ in cluster $j$

- Conditional building of the model:
  - conditional distribution of $Y$ given random effect:
    $$Y_i|U \sim \text{indep. } \text{ExpFam}(\mu, \phi)$$
    with $\forall i$
    $$\mathbb{E}(Y_i|U) = \mu_i = g^{-1}(\eta_i) = g^{-1}(x_i'\beta + z_i' u)$$
    $$\mathbb{V}(Y_i|U) = \phi v(\mu_i)$$
    with $g$ the link function
    and $v$ the variance function
  - random effects distribution $U$:
    $$U_j \sim \mathcal{N}_{q_j}(0, \sigma_j^2 I_{q_j})$$
Example (cont.)

- \( Y_{jk} | U \sim \text{indep. BinomFam}(n_{jk}, p_{jk}) \)
- \( \mathbb{E}(Y_{jk} | U) = \mu_{jk} = n_{jk} \times p_{jk} = n_{jk} \times g^{-1}(\eta_{jk}) \)
- \( \mathbb{V}(Y_{jk} | U) = \phi v(\mu_{jk}) \)
- \( U \sim \mathcal{N}_8(0, \sigma^2 I_8) \)
Marginal moments

- \( \mathbb{E}(Y_i) = \mathbb{E}(\mathbb{E}(Y_i|U)) = \mathbb{E}(g^{-1}(x_i^T \beta + z_i^T u)) = ? \)
  \( \hookrightarrow \) cannot in general be simplified due to the non linear function \( g^{-1} \).

- \( \mathbb{V}(Y_i) = \mathbb{V}(\mathbb{E}(Y_i|U)) + \mathbb{E}(\mathbb{V}(Y_i|U)) = \mathbb{V}(g^{-1}(x_i^T \beta + z_i^T u)) + \mathbb{E}(\phi v(g^{-1}(x_i^T \beta + z_i^T u))) = ? \)
  \( \hookrightarrow \) cannot in general be simplified.

- \( \text{cov}(Y_i, Y_j) = \text{cov}(\mathbb{E}(Y_i|U), \mathbb{E}(Y_j|U)) + \mathbb{E}(\text{cov}(Y_i, Y_j|U)) = \text{cov}(g^{-1}(x_i^T \beta + z_i^T u), g^{-1}(x_j^T \beta + z_j^T u)) + 0 = ? \)
  \( \hookrightarrow \) cannot in general be simplified.
Likelihood definition

- parameters to be estimated: $\beta$, $\sigma^2$
- observed data: $y$

$$\mathcal{L}(\beta, \sigma^2; y) = \int_{\mathbb{R}^q} f_{Y|U}(y) f_U(u) \, du$$
$$= \int_{\mathbb{R}^q} \prod_i f_{Y_i|U}(y_i) f_U(u) \, du$$

**Problem:** This integral cannot be formally expressed in closed form with exceptions e.g. gaussian distribution, beta-binomial model, Poisson-gamma model ...

$\rightarrow$ no analytic expression for the marginal distribution of $Y$
$\rightarrow$ need for approximation
2 main estimation approaches:

- numerical integration for calculating the likelihood and hence numerical maximisation of the likelihood
- linearization methods using Taylor expansions to approximate the model
Estimation by numerical integration

\[ \mathcal{L}(\beta, \sigma^2; y) = \int_{\mathbb{R}^q} \prod_i f_{Y_i|U}(y_i) f_U(u) \, du \]

▷ Gaussian quadrature:
integral approximated by a weighted sum:

*cf. Abramowitz & Stegun (1964)*

If \( U_j \) are independent:

\[
\mathcal{L}(\beta, \sigma^2; y) = \int_{\mathbb{R}^q} \prod_i f_{Y_i|U}(y_i) f_U(u) \, du
= \prod_j \int_{\mathbb{R}} \prod_i f_{Y_{ij}|U_j}(y_{ij}) f_{U_j}(u_j) \, du_j
= \prod_j \int_{\mathbb{R}} \prod_i f_{Y_{ij}|U_j^*}(y_{ij}) f_{U_j^*}(u_j^*) \, du_j^*
\]

\[
\ell(\beta, \sigma^2; y) \simeq \sum_j \ln \left[ \sum_l \left( \prod_i f_{Y_{ij}|U_j^*}(y_{ij}, u_{j,l}^*) \right) w_{j,l} \right]
\]

with  
- quadrature points: \( u_{j,l}^* \)
- weights: \( w_{j,l} \)
Difficulties:

- choice of the quadrature points
- low dimension of random effects
- a single random effects or two, no more
  - nested: OK
  - crossed: impossible

→ works only in very simple situations

→ implemented in SAS macro *nlmixed* devoted to non linear mixed models

Other numerical integration techniques:
MCMC, EM, MCEM ... suffer from nearly the same difficulties
Estimation by linearization

▷ Simple approach cf. Schall (1991)

2 steps method:

⇒ conditionally on $U$: the GLMM model is a GLM

- recall estimation of $\beta$ in a GLM: weighted least squares

$$X'W_\beta^{-1}X\beta = X'W_\beta^{-1}z_\beta$$

- on pseudo data (or working variates): $z_\beta = \eta + (y - \mu_\beta)g'(\mu_\beta)$

⇔ seen as an expansion of the link function around the mean
- with covariance matrix: $W_\beta = diag\{g'(\mu_\beta)^2 \nu(y)\}$

- linearized model $M^{[t]}$ (iterative procedure):

$$Z^{[t]} = X\beta + \varepsilon \quad \text{where} \quad \nabla(\varepsilon) = W_{\beta^{[t]}}$$
given $U$, linearized model $\mathcal{M}_u^{[t]}$:

$$Z^{[t]} = X\beta + ZU + \varepsilon \quad \text{where} \quad \nabla(\varepsilon) = W_{\beta^{[t]}, u^{[t]}} = W^{[t]}$$

⇒ estimation in the LMM

iterative estimation of parameters in the LMM (random effects recovering their random nature)

↩ double iterative algorithm

Other justifications to similar computational algorithms:


▷ Laplace approximation \textit{cf. Wolfinger (1993)}
Advantages:
- easy to compute
- numerically fast
- no limit on the number or type of random effects

Drawbacks:
- small bias (correction proposed by Lin & Breslow (1996))
- normality assumption of the random effects
- small value of variance components

implemented in SAS macro glimmix
Multi-center clinical trial

<table>
<thead>
<tr>
<th>clinic</th>
<th>trt</th>
<th>fav</th>
<th>unfav</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>drug</td>
<td>11.00</td>
<td>25.00</td>
</tr>
<tr>
<td>2</td>
<td>cntl</td>
<td>10.00</td>
<td>27.00</td>
</tr>
<tr>
<td>3</td>
<td>drug</td>
<td>16.00</td>
<td>4.00</td>
</tr>
<tr>
<td>4</td>
<td>cntl</td>
<td>22.00</td>
<td>10.00</td>
</tr>
<tr>
<td>5</td>
<td>drug</td>
<td>14.00</td>
<td>5.00</td>
</tr>
<tr>
<td>6</td>
<td>cntl</td>
<td>7.00</td>
<td>12.00</td>
</tr>
<tr>
<td>7</td>
<td>drug</td>
<td>2.00</td>
<td>14.00</td>
</tr>
<tr>
<td>8</td>
<td>cntl</td>
<td>1.00</td>
<td>16.00</td>
</tr>
<tr>
<td>9</td>
<td>drug</td>
<td>6.00</td>
<td>11.00</td>
</tr>
<tr>
<td>10</td>
<td>cntl</td>
<td>0.00</td>
<td>12.00</td>
</tr>
<tr>
<td>11</td>
<td>drug</td>
<td>1.00</td>
<td>10.00</td>
</tr>
<tr>
<td>12</td>
<td>cntl</td>
<td>0.00</td>
<td>10.00</td>
</tr>
<tr>
<td>13</td>
<td>drug</td>
<td>1.00</td>
<td>4.00</td>
</tr>
<tr>
<td>14</td>
<td>cntl</td>
<td>1.00</td>
<td>8.00</td>
</tr>
<tr>
<td>15</td>
<td>drug</td>
<td>4.00</td>
<td>2.00</td>
</tr>
<tr>
<td>16</td>
<td>cntl</td>
<td>6.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Do drug and clinic have an effect on the probability of favorable response?

→ each clinic is measured twice for drug and control: dependency between the 2 responses for each clinic

→ clinics sampled from a target population: between clinics variability
Modeling approach

Denote $p_{jk}$ the probability of favorable response for treatment $k$ and clinic $j$.

A linear model would be described:

$$f_{jk} = p_{jk} + \varepsilon_{jk}$$

with

$$p_{jk} = \mu + \alpha_k + c_j + (\gamma)_{jk}$$

Problems:

- non normal distribution of the data
- values of linear combination of effects are outside of the range for probabilities
- interaction $\gamma$ and error $\varepsilon$ are confounded in the model
Binomial model with canonical logit link

\[ \eta_{jk} = \log \frac{p_{jk}}{1 - p_{jk}} = \mu + \alpha_k + c_j + (\gamma)_{jk} \]

with \( c_j \sim \mathcal{N}(0, \sigma_c^2) \) and \((\gamma)_{jk} \sim \mathcal{N}(0, \sigma_\gamma^2)\).

```r
> clinics <- read.table("DataSets/clinics.txt", header=T, colClasses = c(rep("factor",2),rep("numeric",2)))
> clinics$trt <- relevel(clinics$trt,"drug") #change reference level
> clinics.glmm <- glmer(cbind(fav,unfav)~ trt + (1|clinic/trt),
                      data=clinics, family=binomial(link="logit"))
```
Binomial model with canonical logit link

```R
> summary(clinics.glmm)

Generalized linear mixed model fit by the Laplace approximation
Formula: cbind(fav, unfav) ~ trt + (1 | clinic/trt)
Data: clinics

AIC  BIC  logLik deviance
43.81 46.9 -17.9  35.81

Random effects:
Groups     Name      Variance  Std.Dev.
trt:clinic (Intercept) 0.0057396 0.07576
clinic     (Intercept) 1.9317511 1.38987
Number of obs: 16, groups: trt:clinic, 16; clinic, 8

Fixed effects:
                Estimate Std. Error  z value Pr(>|z|)
(Intercept)   -0.4573     0.5442  -0.840  0.4008
trtcntl      -0.7423     0.3001  -2.473  0.0134 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
        (Intr)
trtcntl -0.263
```

C. Demetrio, F. Mortier & C. Trottier
Mixed Models, theory and applications
Binomial model with other link functions

```r
> clinicsProb.glmm <- glmer(cbind(fav, unfav) ~ trt + (1|clinic/trt),
+ data=clinics, family=binomial(link="probit"))
> summary(clinicsProb.glmm)

Generalized linear mixed model fit by the Laplace approximation
Formula: cbind(fav, unfav) ~ trt + (1 | clinic/trt)
Data: clinics

AIC  BIC  logLik  deviance
44.12 47.21 -18.06 36.12

Random effects:
Groups   Name        Variance   Std.Dev.
trt:clinic (Intercept) 0.00045787 0.021398
clinic    (Intercept) 0.66611291 0.816157
Number of obs: 16, groups: trt:clinic, 16; clinic, 8

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.2680    0.3181  -0.842  0.3996
trtcntl      -0.4425    0.1735  -2.550  0.0108 *

---
Signif. codes:  0 Š***Š 0.001 Š**Š 0.01 Š*Š 0.05 Š.Š 0.1 Š Š 1

Correlation of Fixed Effects:
   (Intr)
trtcntl -0.266
```

C. Demetrio, F. Mortier & C. Trottier
Mixed Models, theory and applications
You can get **predicted values**:
- on the predictor scale / on the inverse link scale (here probability scale)
- with BLUP of random effects or not.

\[
\begin{array}{ccc}
\text{Inv.pred} & \text{BLUP} & \text{NOBLUP} \\
\hline
\text{pred} & g^{-1}(x'\hat{\beta} + z'\hat{u}) & g^{-1}(x'\hat{\beta}) \\
\end{array}
\]

```r
> xtable(data.frame(logits=clinics.glmm@eta[1:5], ilogits=clinics.glmm@mu[1:5], + ilogit2 = exp(clinics.glmm@eta[1:5])/(1+exp(clinics.glmm@eta[1:5]))))
```

<table>
<thead>
<tr>
<th></th>
<th>logits</th>
<th>ilogits</th>
<th>ilogit2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.57</td>
<td>0.36</td>
<td>0.36</td>
</tr>
<tr>
<td>2</td>
<td>-1.29</td>
<td>0.22</td>
<td>0.22</td>
</tr>
<tr>
<td>3</td>
<td>1.39</td>
<td>0.80</td>
<td>0.80</td>
</tr>
<tr>
<td>4</td>
<td>0.65</td>
<td>0.66</td>
<td>0.66</td>
</tr>
<tr>
<td>5</td>
<td>0.54</td>
<td>0.63</td>
<td>0.63</td>
</tr>
</tbody>
</table>
Longitudinal (or repeated) data are successive observations on each of a collection of observational units.

Example

- traits are measured at different times
  - tree or animal growth
  - change in survivorship over time among populations
- individuals are exposed to different level of a same treatment
  - same plants exposed to varying nitrogen dose (discrete or continuous)
  - patients receiving several treatments along time periods

General objectives

→ Modeling the evolution among repetitions of traits depending on co-variables (factor or regressor) taking into account dependencies within subjects
→ Modeling the between subjects variability
Advantages:
- an experiment more efficient
- helps keep the variability low
- helps keep the validity of the results higher

Drawbacks:
- may not be possible for each participant to be in all conditions of the experiment
- unbalanced design
- missing data
Main question

observations on same subject

⇓

dependencies within subjects

but consecutive observations can be more correlated than more distant observations

⇓

How to model within subject dependencies?
Simple treatment \( \times \) time model with repeated measures

**Example**

Pharmaceutical company tests a new drug on respiratory ability

A = standard, C = new, P = placebo

24 patients

each patient will receive each drug in a randomly assigned order

respiratory ability (FEV1) measured hourly for 8 hours following treatment
Hypothesis

- How does the response mean differ among treatment?
  - Is there a treatment main effect?
- How does the response mean change over time?
  - Is there a time main effect?
- How do response differences among treatment change over time?
  - Is there treatment × time interactions?

- Treatment is called the between-subjects factor
- Time is called within-subjects factor
Data and reshape

```r
> respir <- read.table("DataSets/respiratory.csv", + colClasses=c("factor", rep("numeric",9), "factor"), + header=TRUE, sep=";")
> colnames(respir) <-c("patient", 0:8,"drug")
> library(reshape)
> respir.lon <- melt(respir,id=c("patient","drug"))
> respir.lon$time <- as.numeric(respir.lon[,3])-1
> colnames(respir.lon)[4] <- "ability"
```
Data and reshape

> `qplot(time, ability, data=resp.lon, facets =~drug,` 
> `+ group=patient, geom="line", colour=patient)`
Fixed model

- gaussian family

\[ Y = \mu + \varepsilon \]

- expectation for individual \( i \) in the treatment \( j \) at time \( k \)

\[ \mathbb{E}[Y_{ijk}] = \mu + \alpha_j + \gamma_k + (\alpha\gamma)_{jk} \]

where \( \mu \) the intercept, \( \alpha \) the treatment effects, \( \gamma \) the time effect and \( (\alpha\gamma) \) the interaction time \( \times \) treatment

- variance

\[ \mathbb{V}(Y) = \sigma^2 I_d \]
Simple ANCOVA model

```r
> Subrespir.lon <- respir.lon[ respir.lon$variable != 0, ]
> respir.lm <- lm(ability ~ drug * variable, data = Subrespir.lon)
> anova(respir.lm)

Analysis of Variance Table

Response: ability

<table>
<thead>
<tr>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>drug</td>
<td>2</td>
<td>25.783</td>
<td>12.8913</td>
<td>25.6060</td>
</tr>
<tr>
<td>variable</td>
<td>7</td>
<td>17.170</td>
<td>2.4529</td>
<td>4.8722</td>
</tr>
<tr>
<td>drug:variable</td>
<td>14</td>
<td>6.280</td>
<td>0.4486</td>
<td>0.8910</td>
</tr>
<tr>
<td>Residuals</td>
<td>552</td>
<td>277.903</td>
<td>0.5034</td>
<td></td>
</tr>
</tbody>
</table>

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```
> residus <- matrix(respir.lm$res, ncol=8)
> pairs(residus[,1:5])

Simple plots

![Simple plots](image-url)
Linear mixed model

```r
> respir1.lmm <- lmer(ability~drug*variable+(1|patient),
+ data=Subrespir.lon, REML=FALSE)
> respir2.lmm <- lmer(ability~drug*variable+(1|patient/drug),
+ data=Subrespir.lon, REML=FALSE)
> anova(respir1.lmm,respir2.lmm)

Data: Subrespir.lon
Models:
respir1.lmm: ability ~ drug * variable + (1 | patient)
respir2.lmm: ability ~ drug * variable + (1 | patient/drug)

Df AIC BIC logLik Chisq Chi Df Pr(>Chisq)
respir1.lmm 26 456.37 569.63 -202.19
respir2.lmm 27 294.02 411.63 -120.01 164.35 1 < 2.2e-16 ***

---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1  1

> data.frame(summary(respir2.lmm)@REmat)

Groups Name Variance Std.Dev.
1 drug:patient (Intercept) 0.053485  0.23127
2 patient (Intercept) 0.368488  0.60703
3 Residual 0.060497  0.24596
```
Comparing fixed parameters estimation

```r
> aic.lm <- c(AIC(respir.lm), BIC(respir.lm))
> aic.lm
[1] 1264.807 1373.710

> aic.lmm <- summary(respir2.lmm)$AICtab
> aic.lmm

       AIC      BIC   logLik deviance REMLdev
respir2.lmm 294.0151 411.630 -120.0075 240.0151 329.7807
```
Comparing fixed part given covariance structure

```r
> respir3.lmm <- lmer(ability~drug+variable+(1|patient/drug),
+ data=Subrespir.lon, REML=FALSE)
> anova(respir2.lmm,respir3.lmm)

Data: Subrespir.lon
Models:
  respir3.lmm: ability ~ drug + variable + (1 | patient/drug)
  respir2.lmm: ability ~ drug * variable + (1 | patient/drug)

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>respir3.lmm</td>
<td>13</td>
<td>360.41</td>
<td>417.04</td>
<td>-167.20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>respir2.lmm</td>
<td>27</td>
<td>294.02</td>
<td>411.63</td>
<td>-120.01</td>
<td>94.391</td>
<td>14</td>
<td>5.584e-14 ***</td>
</tr>
</tbody>
</table>
```

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Balanced experimental design

\[ \rightarrow \text{estimator of fixed effects do not depend on variance estimations} \]

\[ \rightarrow \text{estimation of fixed effects are equals for both cases} \]

\[ \textbf{> coefs} \]

\[
\begin{bmatrix}
[,1] & [,2] \\
(Intercept) & 3.48875000 & 3.48875000 \\
drugc & 0.19625000 & 0.19625000 \\
drugp & -0.67375000 & -0.67375000 \\
variable2 & -0.07666667 & -0.07666667 \\
variable3 & -0.28958333 & -0.28958333 \\
\end{bmatrix}
\]

Residual variances are different

0.5 without random effects and 0.06 taking into account random effects

\[ \rightarrow \text{interaction becomes "significant".} \]
Modelling covariance function

- are residual variances constant over time?
- are covariances between times depending on lag?
- are covariances depending on treatments?
- ...

Plan Models with fixed effects  Linear Mixed Models (LMM)  Generalized Linear Models (GLMM)  Longitudinal case  Overdispersion  Nonlinear models
From here, we consider:

\[
V = \begin{pmatrix}
V_{11} & 0 & 0 & 0 \\
0 & V_{12} & 0 & 0 \\
0 & 0 & \ddots & 0 \\
0 & 0 & 0 & V_{83}
\end{pmatrix}
\]

We assume no patient dependency from one drug to the other.
We work on the $8 \times 8$-covariance sub-block matrix
Studying relation between estimated covariance and time

- Let assume an unstructured covariance model: the most parameterized model

\[
\mathbf{V}(\mathbf{Y}_{ij}) = \begin{pmatrix}
\sigma_1^2 & \sigma_{12} & \sigma_{13} & \cdots \\
\sigma_{12} & \sigma_2^2 & \sigma_{23} & \cdots \\
\sigma_{13} & \sigma_{23} & \ddots & \ddots \\
\cdots & \cdots & \cdots & \ddots
\end{pmatrix}
\]

```r
> Subrespir.lon <- respir.lon[respir.lon$variable!=0,]
> Subrespir.lon$indic <- I(Subrespir.lon$patient:Subrespir.lon$drug)
> respir.un <- gls(ability~drug*variable, data=Subrespir.lon,
+ correlation = corSymm(form = ~1 | indic),
+ weights= varIdent( form = ~1 | time),method="ML")
> VarCov
```

```
[1,] 0.4351987 0.4395978 0.4256334 0.3981156 0.4168162 0.3769659 0.3413459 0.3679788
[2,] 0.4395978 0.4947652 0.4607254 0.4492399 0.4737156 0.4077123 0.3825432 0.4079240
[3,] 0.4256334 0.4607254 0.4717923 0.4491765 0.4641391 0.4085276 0.3853456 0.4078965
[4,] 0.3981156 0.4492399 0.4491765 0.4732017 0.4635553 0.4004393 0.3855756 0.4073768
[5,] 0.4168162 0.4737156 0.4641391 0.4635553 0.5538420 0.4738645 0.4449387 0.4744100
[6,] 0.3769659 0.4077123 0.4085257 0.4004393 0.4738645 0.4701743 0.4267930 0.4438868
[7,] 0.3413459 0.3825432 0.3853456 0.3855756 0.4449387 0.4267930 0.4786082 0.4308936
[8,] 0.3679788 0.4079240 0.4078965 0.4073768 0.4744100 0.4438868 0.4308936 0.4821444
```
Let assume autoregressive covariance matrix:

$$\nabla(Y_{ij}) = \sigma^2 R \quad \text{where} \quad r_{kk'} = \text{cor}(Y_{ijk}, Y_{ijk'}) = \rho^{|k-k'|}$$

```r
> respir.ar <- gls(ability~drug*variable, data = Subrespir.lon,
+ correlation = corAR1(0.2, form= ~time|indic), method="ML")

\[ \phi = 0.9219 \quad \text{and} \quad \sigma^2 = 0.4685 \]```
Plot between estimated unstructured and AR(1) covariances.

C. Demetrio, F. Mortier & C. Trottier
Mixed Models, theory and applications
Random effect added to autoregressive structure

Autoregressive **conditional** covariance matrix for 1 subject within drug:

$$\nabla(Y_{ij}|u) = \sigma^2 R \quad \text{where} \quad r_{kk'} = \text{cor}(Y_{ijk}, Y_{ijk'}|u) = \rho^{|k-k'|}$$

**Marginal** covariance matrix for 1 subject within drug:

$$V_{ij} = \begin{pmatrix} 
\sigma^2 + \sigma_u^2 & \sigma^2 \rho^{|k-k'|} + \sigma_u^2 \\
\sigma^2 \rho^{|k-k'|} + \sigma_u^2 & \sigma^2 + \sigma_u^2 
\end{pmatrix}$$

```r
> respir.lmear <- lme(ability~drug*variable, data=Subrespir.lon,
+ random= ~1|indic, correlation = corAR1(0.2, form= ~time|indic))
```
Plot between estimated unstructured and AR(1)+RE covariances
Impact of the patient RE

```r
> respir2.lme <- lme(ability~drug*variable, data=Subrespir.lon, 
+ random= ~1|patient/drug, 
+ correlation = corAR1(0.2, form= ~time|patient/drug),method="ML") 
> anova(respir.lme, respir2.lme)

<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>Test</th>
<th>L.Ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>respir.lme</td>
<td>1</td>
<td>27</td>
<td>258.8141</td>
<td>-102.4070</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>respir2.lme</td>
<td>2</td>
<td>28</td>
<td>190.0226</td>
<td>-67.0113</td>
<td>1 vs 2</td>
<td>70.79148</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>
```
Some remarks

- taking into account dependencies within subjects can increase test for fixed effects
- random effects are a way to model within subjects dependencies but are not necessarily sufficient
- defining appropriate residual correlation form is a difficult task
  - plots are useful tools but not sufficient
  - different models should be compared but:

  → conclusions may differ according to different criterion
Overdispersion in glms (Hinde and Demétrio, 1998a,b)

Residual Deviance $\approx$ Residual d.f.

What if Residual Deviance $\gg$ Residual d.f.?

1. Badly fitting model
   - omitted terms/variables
   - incorrect relationship (link)
   - outliers

2. variation greater than predicted by model: $\Rightarrow$ Overdispersion
   - count data: $\text{Var}(Y) > \mu$
   - counted proportion data:
     $$\text{Var}(Y) > n\pi(1 - \pi)$$
Overdispersion in glms (Hinde and Demétrio, 1998a,b)

Causes of Overdispersion

- variability of experimental material
  - individual level variability
- correlation between individual responses
  e.g. litters of rats
- cluster sampling
  e.g. areas; schools; classes; children
- aggregate level data
- omitted unobserved variables
- excess zero counts (structural and sampling zeros)

Consequences

With correct mean model we have consistent estimates of $\beta$ but:

- incorrect standard errors
- selection of overly complex models
### Example

**Worldwide Airline Fatalities, 1976-85**

<table>
<thead>
<tr>
<th>Year</th>
<th>Fatal accidents</th>
<th>Passenger deaths</th>
<th>Passenger miles (100 million)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1976</td>
<td>24</td>
<td>734</td>
<td>3863</td>
</tr>
<tr>
<td>1977</td>
<td>25</td>
<td>516</td>
<td>4300</td>
</tr>
<tr>
<td>1978</td>
<td>31</td>
<td>754</td>
<td>5027</td>
</tr>
<tr>
<td>1979</td>
<td>31</td>
<td>877</td>
<td>5481</td>
</tr>
<tr>
<td>1980</td>
<td>22</td>
<td>814</td>
<td>5814</td>
</tr>
<tr>
<td>1981</td>
<td>21</td>
<td>362</td>
<td>6033</td>
</tr>
<tr>
<td>1982</td>
<td>26</td>
<td>764</td>
<td>5877</td>
</tr>
<tr>
<td>1983</td>
<td>20</td>
<td>809</td>
<td>6223</td>
</tr>
<tr>
<td>1984</td>
<td>16</td>
<td>223</td>
<td>7433</td>
</tr>
<tr>
<td>1985</td>
<td>22</td>
<td>1066</td>
<td>7107</td>
</tr>
</tbody>
</table>

C. Demetrio, F. Mortier & C. Trottier
Worldwide Airline Fatalities, 1976-85

**Simple Models**

- Passenger miles ($m_i$) as exposure variable
- Poisson log-linear model
- Linear time trend

\[
Y_i \sim \text{Pois}(m_i \lambda_i) \\
\log \lambda_i = \beta_0 + \beta_1 \text{year}_i
\]  \(1\)

**Fatal accidents:**

Deviance(time trend) = 20.68
Residual Deviance = 5.46 on 8 d.f.

**Passenger deaths:**

Deviance(time trend) = 202.1
Residual Deviance = 1051.5 on 8 d.f.

\[\Rightarrow\] compounding with aircraft size
Two broad categories

- Assume some more general form for the variance function, possibly with additional parameters.

- Assume a two-stage model for the response with the response model parameter following some distribution.
  Maximum likelihood estimation (conjugate distribution models) or approximate methods (e.g. using first two moments as above)
  Full hierarchical model – Bayesian methods
Mean-variance Models (Hinde and Demétrio, 1998a,b)

**Overdispersed Proportion Data**

$Y_i$ successes out of $m_i$ trials, $i = 1, \ldots, n.$

Model expected proportions $\pi_i$ with link function $g$ and

$$g(\pi_i) = \beta'x_i$$

- Constant overdispersion

$$\text{Var}(Y_i) = \phi m_i \pi_i (1 - \pi_i)$$

- A general variance function:

  Overdispersion allowed to depend upon both $m_i$ and $\pi_i$.

$$\text{Var}(Y_i) = m_i \pi_i (1 - \pi_i) \times 
\left[1 + \phi (m_i - 1)^{\delta_1} \{\pi_i (1 - \pi_i)\}^{\delta_2}\right]$$
Mean-variance Models (Hinde and Demétrio, 1998a,b)

**Overdispersed Count data**

Random variables $Y_i$ represent counts with means $\mu_i$.

- Constant overdispersion

\[
\text{Var}(Y_i) = \phi \mu_i
\]

...can arise through a simple compounding process.

Suppose that $N_i \sim \text{Pois}(\mu_N)$ and $T = \sum_{i=1}^{N_i} X_i$, $X_i$ are iid random variables.

\[
\mathbb{E}[T] = \mu_T = \mathbb{E}_N(\mathbb{E}[T|N]) = \mu_N\mu_X
\]

\[
\text{Var}(T) = \mathbb{E}_N[\text{Var}(T|N)] + \text{Var}_N(\mathbb{E}[T|N])
\]

\[
= \mu_T \left( \frac{\sigma_X^2}{\mu_X} + \mu_X \right) = \mu_T \frac{\mathbb{E}[X^2]}{\mathbb{E}[X]}
\]

- A general variance function

\[
\text{Var}(Y_i) = \mu_i \left\{ 1 + \phi \mu_i^\delta \right\}
\]
Two-stage Models – Binomial (Hinde and Demétrio, 1998a,b)

Beta-Binomial

\[ Y_i | P_i \sim \text{Bin}(m_i, P_i) \]

\[ \mathbb{E}(P_i) = \pi_i \quad \text{Var}(P_i) = \phi \pi_i (1 - \pi_i) \]

Unconditionally, \( \mathbb{E}(Y_i) = m_i \pi_i \) and \( \text{Var}(Y_i) = m_i \pi_i (1 - \pi_i) [1 + (m_i - 1) \phi] \)

Taking \( P_i \sim \text{Beta}(\alpha_i, \beta_i) \), with \( \alpha_i + \beta_i \) fixed, gives beta-binomial distribution for \( Y_i \) with the same variance function.
Two-stage Models – Binomial (Hinde and Demétrio, 1998a,b)

The same variance function results from assuming that individual binary responses are not independent but have a constant correlation. Writing \( Y_i = \sum_{j=1}^{m_i} R_{ij} \), where \( R_{ij} \) are Bernoulli random variables with

\[
\mathbb{E}[R_{ij}] = \pi_i \text{ and } \text{Var}(R_{ij}) = \pi_i(1 - \pi_i)
\]

then, assuming a constant correlation \( \rho \) between the \( R_{ij} \)'s for \( j \neq k \), we have

\[
\text{Cov}(R_{ij}, R_{ik}) = \rho \pi_i(1 - \pi_i)
\]

and

\[
\begin{align*}
\mathbb{E}[Y_i] &= m_i \pi_i \\
\text{Var}(Y_i) &= \sum_{j=1}^{m_i} \text{Var}(R_{ij}) + \sum_{j=1}^{m_i} \sum_{k \neq j} \text{Cov}(R_{ij}, R_{ik}) \\
&= m_i \pi_i(1 - \pi_i) + m_i (m_i - 1)[\rho \pi_i(1 - \pi_i)] \\
&= m_i \pi_i(1 - \pi_i)[1 + \rho(m_i - 1)],
\end{align*}
\]
Two-stage Models – Binomial (Hinde and Demétrio, 1998a,b)

**Logistic-normal and related models**
Random effect in the linear predictor

\[ \eta_i = \beta' x_i + \sigma z_i \]

- assume \( z_i \sim N(0, 1) \)
- Probit-normal model - a convenient interpretation as a threshold model for a normally distributed latent variable (McCulloch, 1994).
- Logistic-normal using EM algorithm with Gaussian quadrature.
- Approximate approach using a Williams type III model with

\[
\text{Var}(Y_i) = m_i \pi_i (1 - \pi_i) [1 + \phi(m_i - 1) \pi_i (1 - \pi_i)]
\]

- make no specific distributional assumption about \( z \) - estimate a discrete mixing distribution by non-parametric maximum likelihood (NPML).
Two-stage Models – Binomial (Hinde and Demétrio, 1998a,b)

Considered as the two-stage model, the logit($P_i$) have a normal distribution with variance $\sigma^2$, i.e. logit($P_i$) $\sim$ N($x'_i \beta$, $\sigma^2$). Writing

$$U_i = \text{logit}(P_i) = \log \frac{P_i}{1 - P_i} \Rightarrow P_i = \frac{e^{U_i}}{1 + e^{U_i}}$$

and using Taylor series for $P_i$, around $U_i = \mathbb{E}[U_i] = x'_i \beta$, we have

$$P_i = \frac{e^{x'_i \beta}}{(1 + e^{x'_i \beta})} + \frac{e^{x'_i \beta}}{(1 + e^{x'_i \beta})^2} (U_i - x'_i \beta) + o(U_i - x'_i \beta).$$

Then

$$\mathbb{E}(P_i) \approx \frac{e^{x'_i \beta}}{(1 + e^{x'_i \beta})} := \pi_i$$

and

$$\text{Var}(P_i) \approx \left[ \frac{e^{x'_i \beta}}{(1 + e^{x'_i \beta})^2} \right]^2 \text{Var}(U_i) = \sigma^2 \pi_i^2 (1 - \pi_i)^2$$

Consequently the variance function for the logistic-normal model can be approximated by

$$\text{Var}(Y_i) \approx m_i \pi_i (1 - \pi_i) [1 + \sigma^2 (m_i - 1) \pi_i (1 - \pi_i)]$$

which Williams (1982) refers to as a type III variance function.
Two-stage models – Count data (Hinde and Demétrio, 1998a,b)

**Negative Binomial**

- Variation in Poisson rate parameter:
  \[ Y_i | \theta_i \sim \text{Pois}(\theta_i), \quad \theta_i \sim \Gamma(k, \lambda_i) \]

  leads to negative binomial distribution with

  \[ \mathbb{E}[Y_i] = \mu_i = \frac{k}{\lambda_i} \]

  and

  \[ \text{Var}(Y_i) = \mu_i + \frac{\mu_i^2}{k} \]

  For known \( k \), in the 1-parameter exponential family so still in glm framework.

- Different assumptions for the \( \Gamma \)-distribution lead to different parameterizations with different overdispersed variance functions, e.g. \( \theta_i \sim \Gamma(k_i, \lambda) \) gives

  \[ \text{Var}(Y_i) = \mu_i \left(1 + \frac{1}{\lambda} \right) = \phi \mu_i \]
Two-stage models – Count data (Hinde and Demétrio, 1998a,b)

Negative binomial distribution

\[ Y_i | \theta_i \sim \text{Pois}(\theta_i) \]
\[ \theta_i \sim \text{Gamma}(k, \lambda_i), \ i = 1, \ldots, n \]

This leads to a negative binomial distribution for the \( Y_i \) with

\[
f_{Y_i}(y_i; \mu_i, k) = \frac{\Gamma(k + y_i)}{\Gamma(k)y!} \frac{\mu_i^{y_i}k^k}{(\mu_i + k)^{k+y_i}}, \quad y_i = 0, 1, \ldots
\]

and

\[
\mathbb{E}(Y_i) = k/\lambda_i = \mu_i
\]
\[
\text{Var}(Y_i) = \mathbb{E}_{\theta_i}[\text{Var}(Y_i|\theta_i)] + \text{Var}_{\theta_i}(\mathbb{E}[Y_i|\theta_i])
\]
\[
= \mathbb{E}[\theta_i] + \mathbb{V}(\theta_i) = \frac{k}{\lambda_i} + \frac{k}{\lambda_i^2}
\]

\[
\text{Var}(Y_i) = \mu_i + \frac{\mu_i^2}{k}
\]
Two-stage models – Count data (Hinde and Demétrio, 1998a,b)

**Poisson-normal and related models**

Individual level random effect in the linear predictor

\[ \eta_i = \beta' x_i + \sigma Z_i \]

- assume \( Z_i \sim N(0, 1) \), so

\[ Y_i | Z_i \sim \text{Pois}(\lambda_i) \quad \text{with} \quad \log \lambda_i = x_i' \beta + \sigma Z_i \]

where \( Z_i \sim N(0, 1) \), which gives

\[
\begin{align*}
\mathbb{E}[Y_i] &= \mathbb{E}_Z_i (\mathbb{E}[Y_i | Z_i]) = \mathbb{E}_Z_i [e^{x_i' \beta + \sigma Z_i}] \\
&= e^{x_i' \beta + \frac{1}{2} \sigma^2} := \mu_i \\
\text{Var}(Y_i) &= \mathbb{E}_Z_i [\text{Var}(Y_i | Z_i)] + \text{Var}_Z_i (\mathbb{E}[Y_i | Z_i]) \\
&= e^{x_i' \beta + \frac{1}{2} \sigma^2} + \text{Var}_Z_i (e^{x_i' \beta + \sigma Z_i}) \\
&= e^{x_i' \beta + \frac{1}{2} \sigma^2} + e^{2x_i' \beta + \sigma^2} (e^{\sigma^2} - 1).
\end{align*}
\]

i.e. a variance function of the form

\[ \text{Var}(Y_i) = \mu_i + k' \mu_i^2 \]
Germination of Orobanche seed

Example

<table>
<thead>
<tr>
<th></th>
<th>O. aegyptiaca 75</th>
<th>O. aegyptiaca 73</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bean</td>
<td>10/39</td>
<td>8/16</td>
</tr>
<tr>
<td>Cucumber</td>
<td>5/6</td>
<td>3/12</td>
</tr>
<tr>
<td>Bean</td>
<td>23/62</td>
<td>10/30</td>
</tr>
<tr>
<td>Cucumber</td>
<td>53/74</td>
<td>22/41</td>
</tr>
<tr>
<td>Bean</td>
<td>23/81</td>
<td>8/28</td>
</tr>
<tr>
<td>Cucumber</td>
<td>55/72</td>
<td>15/30</td>
</tr>
<tr>
<td>Bean</td>
<td>26/51</td>
<td>23/45</td>
</tr>
<tr>
<td>Cucumber</td>
<td>32/51</td>
<td>32/51</td>
</tr>
<tr>
<td>Bean</td>
<td>17/39</td>
<td>0/4</td>
</tr>
<tr>
<td>Cucumber</td>
<td>46/79</td>
<td>3/7</td>
</tr>
<tr>
<td></td>
<td>10/13</td>
<td></td>
</tr>
</tbody>
</table>

- Response variable: $Y_i$ – number of germinated seeds out of $m_i$ seeds.
- Distribution: Binomial.
- Systematic component: factorial $2 \times 2$ (2 species, 2 extracts), completely randomized experiment (Crowder, 1978).
- Aim: to see how germination is affected by species and extracts.
- Problem: overdispersion.
Models for Orobanche Data (Hinde and Demétrio, 1998a,b)

Binomial:

\[ \text{Var}(Y_i) = m_i \pi_i (1 - \pi_i) \]

- residual deviance for the interaction model is 33.28 on 17 df. – overdispersion

Quasi-likelihood:

\[ \text{Var}(Y_i) = m_i \pi_i (1 - \pi_i) \]

- constant overdispersion \( \tilde{\phi} = 1.862 \)
- only marginal evidence of interaction
- extract only important factor

Williams:

\[ \text{Var}(Y_i) = m_i \pi_i (1 - \pi_i)[1 + \phi(m_i - 1)] \]

- moment estimate \( \tilde{\phi} = 0.0249 \)
- only marginal evidence of interaction
- extract only important factor
- note similarity to QL, even though \( m_i \) not equal.
**Table:** Deviances with overdispersion estimated from maximal model.

<table>
<thead>
<tr>
<th></th>
<th>Bin (ML)</th>
<th>Const (QL)</th>
<th>Beta-Binomial (ML)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>56.49</td>
<td>30.34</td>
<td>32.69</td>
</tr>
<tr>
<td>S</td>
<td>3.06</td>
<td>1.64</td>
<td>2.88</td>
</tr>
<tr>
<td>S.E</td>
<td>6.41</td>
<td>3.44</td>
<td>4.45</td>
</tr>
<tr>
<td>$\hat{\phi}$</td>
<td>1.862</td>
<td>0.012</td>
<td></td>
</tr>
</tbody>
</table>

**Table:** Deviances with overdispersion re-estimated for each model.

<table>
<thead>
<tr>
<th>Source</th>
<th>Beta-Binomial (ML)</th>
<th>Logistic-Normal (ML)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>15.44</td>
<td>15.40</td>
</tr>
<tr>
<td>S</td>
<td>2.73</td>
<td>2.71</td>
</tr>
<tr>
<td>S.E</td>
<td>4.13</td>
<td>4.17</td>
</tr>
</tbody>
</table>
Half-normal plots with simulation envelopes (Hinde and Demétrio, 1998a,b)
Species <- factor(c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
                   2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2))
Extract <- factor(c(1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2,
                    1, 1, 1, 1, 1, 2, 2, 2, 2, 2))
y <- c(10, 23, 23, 26, 17, 5, 53, 55, 32, 46, 10, 8,
      10, 8, 23, 0, 3, 22, 15, 32, 3)
m <- c(39, 62, 81, 51, 39, 6, 74, 72, 51, 79, 13, 16,
      30, 28, 45, 4, 12, 41, 30, 51, 7)

orobanch <- data.frame(Species, Extract, m, y)
rm(Species, Extract, m, y)
attach(orobanch)
resp<-cbind(y,m-y)

# Binomial fit
oro.Bin<-glm(resp~Species*Extract, family=binomial)
anova(oro.Bin, test="Chisq")
summary(oro.Bin, cor=FALSE)

oro.Bin<-glm(resp~Extract*Species, family=binomial)
anova(oro.Bin, test="Chisq")
summary(oro.Bin, cor=FALSE)

## Pearson estimate of phi
(X2<-sum(residuals(oro.Bin, 'pearson')^2))
(phi<-X2/df.residual(oro.Bin))
# Quasilikelihood fit
oro.QL<-glm(resp~Species*Extract, family=quasibinomial)
summary(oro.QL, cor=FALSE)
summary(oro.QL)$dispersion
anova(orobanchQL.fit, test="F")

##### Beta-binomial model
library(aod)
# re-estimating phi
oro.BBin4 <- betabin(cbind(y,m-y) ~Extract,~1, data=orobanch)
oro.BBin3 <- betabin(cbind(y,m-y) ~Species,~1, data=orobanch)
oro.BBin2 <- betabin(cbind(y,m-y) ~Extract+Species,~1, data=orobanch)
oro.BBin1 <- betabin(cbind(y,m-y) ~Extract+Species+Extract:Species,~1, data=orobanch)
anova(oro.BBin4,oro.BBin2, oro.BBin1)
anova(oro.BBin3,oro.BBin2, oro.BBin1)
summary(oro.BBin2)

# fixing phi estimated by the maximum model
f1 <- betabin(cbind(y, m - y) ~ Extract+Species+Extract:Species, ~ 1, data=orobanch, fixpar = list(5, 0.01238))
f2 <- betabin(cbind(y, m - y) ~ Extract+Species, ~ 1, data=orobanch, fixpar = list(4, 0.01238))
f3 <- betabin(cbind(y, m - y) ~ Species, ~ 1, data=orobanch, fixpar = list(3, 0.01238))
f4 <- betabin(cbind(y, m - y) ~ Extract, ~ 1, data=orobanch, fixpar = list(3, 0.01238))
anova(f4,f2,f1)
anova(f3,f2,f1)
# logistic normal
library(lme4)
ind <- (1:length(y))
# re-estimating phi
oro.LN1<-glmer(resp~Extract*Species + (1|ind), family=binomial(link="logit"))
oro.LN1
summary(oro.LN1)@coefs
summary(oro.LN1)@REmat
oro.LN2<-glmer(resp~Extract+Species + (1|ind), family=binomial(link="logit"))
oro.LN2
oro.LN4<-glmer(resp~Extract + (1|ind), family=binomial(link="logit"))
oro.LN4
oro.LN3<-glmer(resp~Species + (1|ind), family=binomial(link="logit"))
oro.LN3
anova(oro.LN4,oro.LN2,oro.LN1)
anova(oro.LN3,oro.LN2,oro.LN1)
### Example

<table>
<thead>
<tr>
<th>Year</th>
<th>Fatal accidents</th>
<th>Passenger deaths</th>
<th>Passenger miles (100 million)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1976</td>
<td>24</td>
<td>734</td>
<td>3863</td>
</tr>
<tr>
<td>1977</td>
<td>25</td>
<td>516</td>
<td>4300</td>
</tr>
<tr>
<td>1978</td>
<td>31</td>
<td>754</td>
<td>5027</td>
</tr>
<tr>
<td>1979</td>
<td>31</td>
<td>877</td>
<td>5481</td>
</tr>
<tr>
<td>1980</td>
<td>22</td>
<td>814</td>
<td>5814</td>
</tr>
<tr>
<td>1981</td>
<td>21</td>
<td>362</td>
<td>6033</td>
</tr>
<tr>
<td>1982</td>
<td>26</td>
<td>764</td>
<td>5877</td>
</tr>
<tr>
<td>1983</td>
<td>20</td>
<td>809</td>
<td>6223</td>
</tr>
<tr>
<td>1984</td>
<td>16</td>
<td>223</td>
<td>7433</td>
</tr>
<tr>
<td>1985</td>
<td>22</td>
<td>1066</td>
<td>7107</td>
</tr>
</tbody>
</table>
Simple Models

- Passenger miles \( (m_i) \) as exposure variable
- Poisson log-linear model
- Linear time trend

\[
Y_i \sim \text{Pois}(m_i \lambda_i) \quad (2)
\]
\[
\log \lambda_i = \beta_0 + \beta_1 \text{year}_i
\]

Fatal accidents:
Deviance(time trend) = 20.68
Residual Deviance = 5.46 on 8 d.f.

Passenger deaths:
Deviance(time trend) = 202.1
Residual Deviance = 1051.5 on 8 d.f.

⇒ compounding with aircraft size
Confidence Intervals – airline data
R program

```r
airline.dat <- scan(what=list(acc=0, death=0, miles=0))
24 734 3863
25 516 4300
31 754 5027
31 877 5481
22 814 5814
21 362 6033
26 764 5877
20 809 6223
16 223 7433
22 1066 7107

airline <- data.frame(airline.dat)
airline.dat$year<-seq(1976,1985,by=1)
attach(airline.dat)
plot(year,acc, main="Number of Fatal Accidents vs Year")

acc1.fit<-glm(acc~year+offset(log(miles)), family=poisson)
summary(acc1.fit)
anova(acc1.fit,test="Chisq")

plot(c(1976,1985), c(15,35), type="n", xlab="Year", ylab="Number of Accidents")
points(year,acc)
x<-seq(1976,1985,1)
lp<-predict(acc1.fit,data.frame(year=x))
fv<-exp(lp)
lines(x,fv,lty=1)
title(sub="Observed values and fitted curve")

qqnorm(resid(acc1.fit),ylab="Residuo")
qqline(resid(acc1.fit))
```
# Number of deaths

# Poisson fit
plot(year, death, main="Number of Fatal Accidents vs Year")

death1.fit <- glm(death ~ year + offset(log(miles)), family = poisson)
summary(death1.fit)
anova(death1.fit, test="Chisq")

inflim <- with(predict(death1.fit, se = TRUE, type="response"),
    fit - pnorm(0.975) * se.fit)
suplim <- with(predict(death1.fit, se = TRUE, type="response"),
    fit + pnorm(0.975) * se.fit)

plot(c(1976, 1985), c(350, 1200), type="n", xlab="Year", ylab="Number of deaths")
points(year, death)
x <- seq(1976, 1985, 1)
fv1 <- predict(death1.fit, data.frame(year = x), type="response")
lines(x, fv1, lty = 1)
lines(x, inflim, lty = 2)
lines(x, suplim, lty = 2)

# Negative binomial fit
library(MASS)
death2.fit <- glm.nb(death ~ year + offset(log(miles)), link = log)
summary(death2.fit)
anova(death2.fit, test="F")

inflim <- with(predict(death2.fit, se = TRUE, type="response"),
    fit - qt(0.975, 8) * se.fit)
suplim <- with(predict(death2.fit, se = TRUE, type="response"),
    fit + qt(0.975, 8) * se.fit)
fv2<-predict(death2.fit,data.frame(year=x),type="response")
#lines(x,fv2,lty=1)
#lines(x,fv1,lty=1,col="blue")
lines(x,inflim,lty=3, col="blue")
lines(x,suplim,lty=3, col="blue")

# Quasilikelihhod fit
death3.fit<-glm(death~year+offset(log(miles)), family=quasipoisson)
summary(death3.fit)
anova(death3.fit, test="F")
inflim <- with(predict(death3.fit, se = TRUE,type="response"),
fit -qt(0.975,8)*se.fit)
suplim <- with(predict(death3.fit, se = TRUE,type="response"),
fit + qt(0.975,8)*se.fit)

cv3<- predict(death3.fit, data.frame(year=x),type="response")
#lines(x,fv3,lty=1,col="red")
lines(x,inflim,lty=4, col="red")
lines(x,suplim,lty=4, col="red")
legend(1976,1200.0,c("Observed","Poisson", "Negative binomial","Quasi-Poisson"), lty=c(-1,2,3,4), pch=c("*", " ", " ", " "), col=c("black","black","blue","red"))

# Poisson log-normal
library(lme4)
ind <- (1:length(death))
death.LN1<-glmer(death~year+offset(log(miles)) + (1|ind),
family=poisson(link="log"))
death.LN1
summary(death.LN1)@coefs
summary(death.LN1)@REmat
**Orange data (Pinheiro & Bates, 2000; Molenberghs & Verbeke, 2008)**

**Example**

The data are the trunk circumference (mm) of each of five orange trees at seven different occasions.

<table>
<thead>
<tr>
<th>Day</th>
<th>Tree 1</th>
<th>Tree 2</th>
<th>Tree 3</th>
<th>Tree 4</th>
<th>Tree 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>118</td>
<td>30</td>
<td>33</td>
<td>30</td>
<td>32</td>
<td>30</td>
</tr>
<tr>
<td>484</td>
<td>58</td>
<td>69</td>
<td>51</td>
<td>62</td>
<td>49</td>
</tr>
<tr>
<td>664</td>
<td>87</td>
<td>111</td>
<td>75</td>
<td>112</td>
<td>81</td>
</tr>
<tr>
<td>1004</td>
<td>115</td>
<td>156</td>
<td>108</td>
<td>167</td>
<td>125</td>
</tr>
<tr>
<td>1231</td>
<td>120</td>
<td>172</td>
<td>115</td>
<td>179</td>
<td>142</td>
</tr>
<tr>
<td>1372</td>
<td>142</td>
<td>203</td>
<td>139</td>
<td>209</td>
<td>174</td>
</tr>
<tr>
<td>1582</td>
<td>145</td>
<td>203</td>
<td>140</td>
<td>214</td>
<td>177</td>
</tr>
</tbody>
</table>

All trees measured on the same occasions – balanced, longitudinal data.
- Much variability between trees
- Much less variability within trees
- Non-linear trend
- Possible correlation along the time
Models for Orange data

**Model 1**: Logistic model for tree $i$ at age $x_{ij}$, $i = 1, \ldots, 5$, $j = 1, \ldots, 7$ without random effects (4 parameters)

$$Y_{ij} = \frac{\phi_1}{1 + \exp[-(x_{ij} - \phi_2)/\phi_3]} + \varepsilon_{ij}$$

$\phi_1$ is an asymptotic trunk circumference,  
$\phi_2$ is the age at which the tree attains half of it asymptotic trunk circumference,  
$\phi_3$ represents the growth scale,  
the error term $\varepsilon_{ij} \sim N(0, \sigma^2)$

**Model 2**: Separate logistic curve to each tree (16 parameters)

$$Y_{ij} = \frac{\phi_{1i}}{1 + \exp[-(x_{ij} - \phi_{2i})/\phi_{3i}]} + \varepsilon_{ij}$$

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$
Models for Orange data (cont.)

Model 3: Logistic model with random effects (10 parameters)

\[
Y_{ij} = \frac{\phi_{1i}}{1 + \exp[-(x_{ij} - \phi_2)/\phi_3]} + \varepsilon_{ij}
\]

\[
\phi_i = \begin{bmatrix}
\psi_{1i} \\
\psi_{2i} \\
\psi_{3i}
\end{bmatrix} = \begin{bmatrix}
\beta_1 \\
\beta_2 \\
\beta_3
\end{bmatrix} + \begin{bmatrix}
b_{1i} \\
b_{2i} \\
b_{3i}
\end{bmatrix} = \boldsymbol{\beta} + \boldsymbol{b}_i, \quad \psi = \begin{bmatrix}
\sigma_1^2 & \sigma_{12} & \sigma_{13} \\
\sigma_{12} & \sigma_2^2 & \sigma_{23} \\
\sigma_{13} & \sigma_{23} & \sigma_3^2
\end{bmatrix}
\]

\[
\boldsymbol{b}_i \sim N(0, \psi), \quad \varepsilon_{ij} \sim N(0, \sigma^2)
\]

Model 4: Logistic model with random effect on the asymptote (5 parameters)

\[
Y_{ij} = \frac{\phi_{1i}}{1 + \exp[-(x_{ij} - \phi_2)/\phi_3]} + \varepsilon_{ij}
\]

\[
\phi_{1i} = \beta_1 + b_i, \quad b_i \sim N(0, \sigma_b^2), \quad \varepsilon_{ij} \sim N(0, \sigma^2)
\]
Model 1: Fixed logistic model

- The variability in the residuals increases with the fitted values, due to correlation among observations in the same tree.
- The residuals are mostly negative for trees 1 and 3 and mostly positive for trees 2 and 4.
Model 2: Separate logistic curve to each tree

- The estimates for all the parameters vary with tree, there appears to be relatively more variability in the Asym estimates.
- Asym – the only parameter for which all the confidence intervals do not overlap, suggesting that it is the only parameter for which random effects are needed to account for variation among trees.
### Models 3 and 4: Random logistic models

<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>Test</th>
<th>L.Ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>10</td>
<td>279.98</td>
<td>295.53</td>
<td>-129.99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>273.17</td>
<td>280.95</td>
<td>-131.58</td>
<td>3 vs 4</td>
<td>3.1896</td>
<td>0.6708</td>
</tr>
</tbody>
</table>

**Note:** The default estimation method in *nlme* in R is maximum likelihood (ML), while in *lme* is restricted maximum likelihood (REML).

The large p-value for the likelihood ratio test confirms that the simpler model (random effect only for the asymptote) is to be preferred.
<table>
<thead>
<tr>
<th>Par.</th>
<th>Value</th>
<th>Std. Error</th>
<th>DF</th>
<th>t-value</th>
<th>Value</th>
<th>Std. Error</th>
<th>DF</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asym</td>
<td>192.68</td>
<td>20.24</td>
<td>32</td>
<td>9.52</td>
<td>191.05</td>
<td>16.15</td>
<td>28</td>
<td>11.83</td>
</tr>
<tr>
<td>xmid</td>
<td>728.71</td>
<td>107.27</td>
<td>32</td>
<td>6.79</td>
<td>722.56</td>
<td>35.15</td>
<td>28</td>
<td>20.56</td>
</tr>
<tr>
<td>scal</td>
<td>353.49</td>
<td>81.46</td>
<td>32</td>
<td>4.34</td>
<td>344.17</td>
<td>27.15</td>
<td>28</td>
<td>12.68</td>
</tr>
<tr>
<td>Residual</td>
<td>546.26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>61.56</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- The fixed-effects estimates are similar, but the standard errors are much smaller in the nlme fit.
- The estimated within-group standard error is also considerably smaller in the non-linear mixed model.
- This is because the between-group variability is not incorporated in the non-linear fixed model, being absorbed in the standard error.
- This pattern is generally observed when comparing mixed-effects versus fixed-effects fits.
Plan Models with fixed effects

Linear Mixed Models (LMM)

Generalized Linear Mixed Models (GLMM)

Longitudinal case

Overdispersion

Nonlinear models

Fitted values (mm)

Standardized residuals

-1

0

1

50 100 150 200

TIME since December 31, 1968 (days)

T runk circumference (mm)

50 100 150 200

C. Demetrio, F. Mortier & C. Trottier

Mixed Models, theory and applications
library(nlme)
data() # lists all datasets
data(Orange, package='datasets')
str(Orange)
head(Orange)
summary(Orange)
par(mfrow=c(1,2))
plot(Orange)  ## five curves in five different panels
plot(Orange, outer=~1)  ## all five curves in one panel

## no random effects - Model 1
(fm1Oran.nls <- nls(circumference ~ SSlogis(age,Asym,xmid,scal),data = Orange))
summary(fm1Oran.nls)
plot(fm1Oran.nls)
plot(fm1Oran.nls, Tree ~ resid(.), abline=0)

## separate logistic curve to each tree - Model 2
(fm1Oran.lis <- nlsList(circumference ~ SSlogis(age, Asym, xmid, scal)|Tree,
data = Orange))
summary(fm1Oran.lis)
plot(intervals(fm1Oran.lis), layout = c(3,1))
plot(fm1Oran.lis, Tree ~ resid(.), abline = 0)
R program (cont.)

```r
## random effect model - Model 3
## no need to specify groups, as Orange is a groupedData object
# random is ommitted - by default is equal to fixed
(fm1Oran.nlme <- nlme(circumference ~ SSlogis(age,Asym,xmid,scal),data = Orange,
fixed=Asym+xmid+scal~1, start=fixef(fm1Oran.lis)))
## or simply
fm1Oran.nlme <- nlme(fm1Oran.lis)
summary(fm1Oran.nlme) ## compare with next (look at the standard errors)
summary(fm1Oran.nls)

## random effect model - Model 4
fm2Oran.nlme <- update(fm1Oran.nlme, random = Asym ~ 1)
anova(fm1Oran.nlme,fm2Oran.nlme) ## compares Model 3 and 4
summary(fm2Oran.nlme)
plot(fm1Oran.nlme) ## plots the standardized residuals versus the fitted values
plot(augPred(fm2Oran.nlme, level = 0:1), layout = c(5,1))
qqnorm(fm2Oran.nlme, abline = c(0,1),cex=0.7)
```
References


