# Estimation and Model Selection in Mixed Effects Models Part I

Adeline Samson[1]

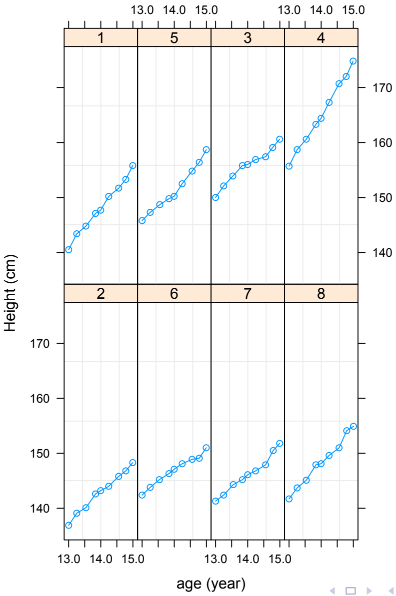
[1]University Paris Descartes

Summer school 2009 - Lipari, Italy

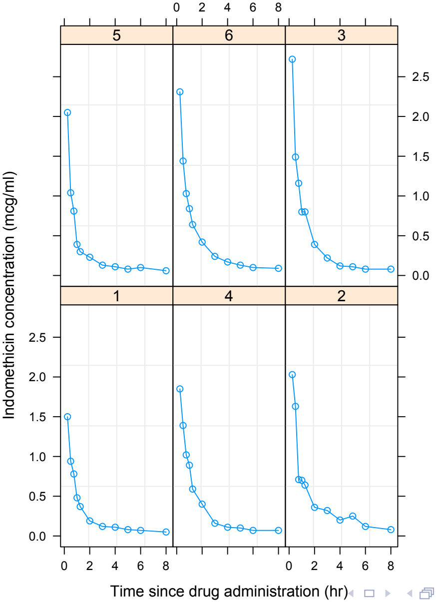
These slides are based on Marc Lavielle's slides

# Outline

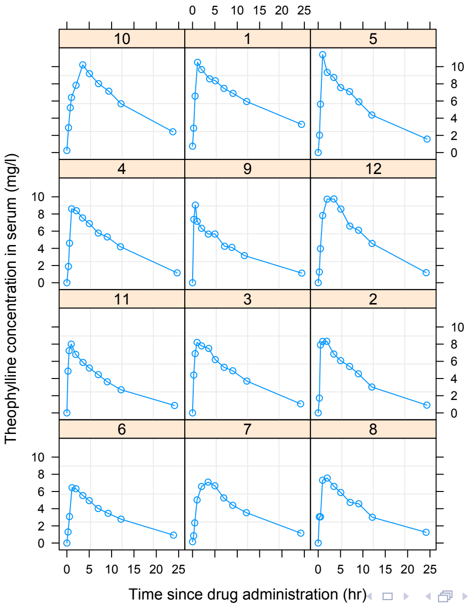
# Some examples of data
## Growth of children



# Some examples of data
## Pharmacokinetics of Indomethacin



# Some examples of data
## Pharmacokinetics of Theophylline



# The classical regression model

For one subject

$$\boxed{y_j = f(x_j, \beta) + \varepsilon_j \quad , \quad 1 \leq j \leq n}$$

$y_j \in \mathbb{R}$ is the $j$th observation of the subject,

$n$ is the number of observations

The regression variables, or design variables, $(x_j)$ are **known**,

The vector of parameters $(\beta)$ is **unknown**.

The measurement error variables $(\varepsilon_j)$ are **unknown**.

# The classical regression model

For one subject

$$y_j = f(x_j, \beta) + \varepsilon_j \quad , \quad 1 \leq j \leq n$$

$y_j \in \mathbb{R}$ is the $j$th observation of the subject,

$n$ is the number of observations

The regression variables, or design variables, $(x_j)$ are **known**,

The vector of parameters $(\beta)$ is **unknown**.

The measurement error variables $(\varepsilon_j)$ are **unknown**.

# The classical regression model

For one subject

$$y_j = f(x_j, \beta) + \varepsilon_j \quad , \quad 1 \leq j \leq n$$

$y_j \in \mathbb{R}$ is the $j$th observation of the subject,

$n$ is the number of observations

The regression variables, or design variables, $(x_j)$ are **known**,

The vector of parameters $(\beta)$ is **unknown**.

The measurement error variables $(\varepsilon_j)$ are **unknown**.

# The classical regression model

For one subject

$$y_j = f(x_j, \beta) + \varepsilon_j \quad , \quad 1 \leq j \leq n$$

$y_j \in \mathbb{R}$ is the $j$th observation of the subject,

$n$ is the number of observations

The regression variables, or design variables, $(x_j)$ are **known**,

The vector of parameters $(\beta)$ is **unknown**.

The measurement error variables $(\varepsilon_j)$ are **unknown**.

# The classical regression model

$$y_j = f(x_j, \beta) + \varepsilon_j \quad , \quad 1 \leq j \leq n$$

The $(\varepsilon_j)$ are modelized as sequences of *random variables*.

The goal of the modeler is to develop simultaneously two kinds of models:

(1) The structural model $f$

(2) The statistical model

# The classical regression model

$$y_j = f(x_j, \beta) + \varepsilon_j \quad , \quad 1 \leq j \leq n$$

(1) **The structural model** $f$**:** We are not interested with a purely *descriptive* model which nicely fits the data, but rather with a *mechanistic* model which has some biological meaning and which is a function of some physiological parameters.

Examples:

- compartimental PK models,
- viral dynamic models,
- ...

# The classical regression model

$$\boxed{y_j = f(x_j, \beta) + \varepsilon_j \quad , \quad 1 \leq j \leq n}$$

(2) **The statistical model** aims to explain the variability observed in the data:

- the residual error model: distribution of $(\varepsilon_j)$
- the model of covariates

# The classical regression model

$$y_j = f(x_j, \beta) + \varepsilon_j \quad , \quad 1 \leq j \leq n$$

Some statistical issues:

- **Estimation:**
  - estimate the parameters of the model
- **Model selection and model assessment:**
  - Select and assess the "best" structural model $f$,
  - Select and assess the "best" statistical model
- **Optimization of the design :**
  - Find the *optimal design* $(x_j)$

# Outline

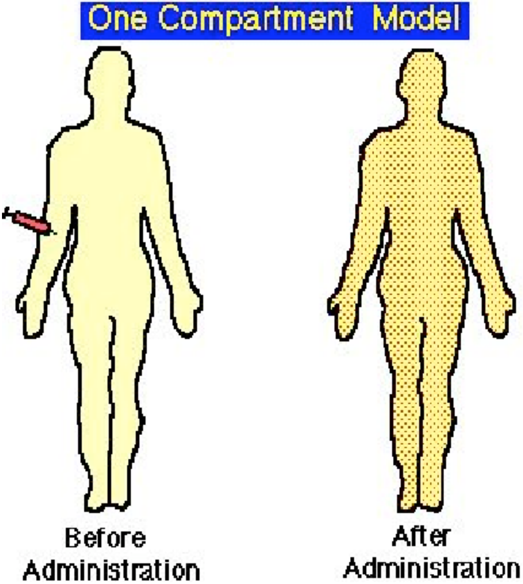
# Pharmacokinetics and Pharmacodynamics (PK/PD)

dose $\longrightarrow$ PK $\longrightarrow$ concentration $\longrightarrow$ PD $\longrightarrow$ response

- Pharmacokinetics (PK): "What the body does to the drug"
- Pharmacodynamics (PD): "What the drug does to the body"

# One compartment PK model

## intravenous administration



# intravenous administration and first-order elimination

dose D (t=0) $\rightarrow$ $\boxed{\text{DRUG AMOUNT } Q(t)}$ $\rightarrow$ elimination (rate $k_e$)

$$\frac{dQ}{dt}(t) = -kQ(t) \quad ; \quad Q(0) = D$$

$$\begin{aligned} Q(t) &= De^{-kt} \\ C(t) &= \frac{Q(t)}{V} = \frac{D}{V} e^{-k_e t} \end{aligned}$$

$C(t)$ : concentration of the drug,
$V$ : volume of the compartment

# intravenous administration and nonlinear elimination

dose D (t=0) $\rightarrow$ $\boxed{\text{DRUG AMOUNT } Q(t)}$ $\rightarrow$ nonlinear elimination

$$\frac{dQ(t)}{dt} = -\frac{V_m\, Q(t)}{V * K_m + Q(t)}$$

$$C(t) = \frac{Q(t)}{V}$$

$(V_m, K_m)$ : Michaelis-Menten elimination parameters,
$V$ : volume of the compartment.

# oral administration, first-order absorbtion and elimination

dose D at time t=0

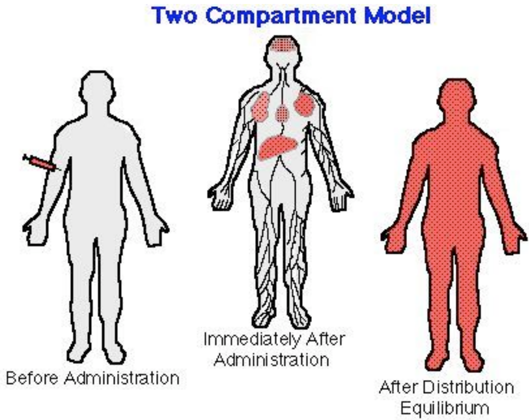
absorption (rate $k_a$) $\rightarrow$ $\boxed{\text{DRUG AMOUNT } Q(t)}$ $\rightarrow$ elimination (rate $k_e$)

$$\begin{aligned} \frac{dQ_a}{dt}(t) &= -k_a Q_a(t) \quad ; \quad Q_a(0) = D \\ \frac{dQ}{dt}(t) &= k_a Q_a(t) - k_e Q(t) \quad ; \quad Q(0) = 0 \end{aligned}$$

$Q_a(t)$: amount at absorption site.

$$C(t) = \frac{Q(t)}{V} = D \frac{k_a}{V(k_a - k_e)} \left( e^{-k_e t} - e^{-k_a t} \right)$$

# Two compartments PK model
## intravenous administration



# Two compartments PK model

$$
\begin{aligned}
\frac{dQ_a}{dt}(t) &= -k_a Q_a(t), \\
\frac{dQ_c}{dt}(t) &= k_a Q_a(t) - k_e Q_c(t) - k_{12} Q_c(t) + k_{21} Q_p(t), \\
\frac{dQ_p}{dt}(t) &= k_{12} Q_c(t) - k_{21} Q_p(t).
\end{aligned}
$$

$Q_a(t)$: amount at absorption site, $Q_a(0) = D$.

$Q_c(t)$: amount in the central compartment, $Q_c(0) = 0$.

$Q_p(t)$: amount in the peripheral compartment, $Q_p(0) = 0$.

# Outline

# The regression model

$$y_j \quad = \quad f(x_j, \beta) + \varepsilon_j \quad , \quad 1 \leq j \leq n$$

$n$ is the number of observations.

The regression variables, or design variables, $(x_j)$ are **known**,

The vector of parameters $\beta$ is **unknown**.

- linear model: $f$ is a linear function of the parameters $\beta$
- non linear model: $f$ is a non linear function of the parameters $\beta$

# The regression model

$$y_j \quad = \quad f(x_j, \beta) + \varepsilon_j \quad , \quad 1 \leq j \leq n$$

$n$ is the number of observations.

The regression variables, or design variables, $(x_j)$ are **known**,

The vector of parameters $\beta$ is **unknown**.

- linear model: $f$ is a linear function of the parameters $\beta$
- non linear model: $f$ is a non linear function of the parameters $\beta$

$$y_j = f(x_j, \beta) + \varepsilon_j$$

The simplest statistical model assumes that the $(\varepsilon_j)$ are independent and identically distributed (*i.i.d*) Gaussian random variables:

$$\varepsilon_j \sim_{i.i.d.} \mathcal{N}(0, \sigma^2)$$

**Problem:** estimate the parameters of the model $\theta = (\beta, \sigma^2)$.

# Maximum Likelihood Estimation

- Maximum likelihood estimation (MLE) is a popular statistical method used for fitting a statistical model to data, and providing estimates for the model's parameters.
- For a fixed set of data and underlying probability model, maximum likelihood picks the values of the model parameters that make the data "more likely" than any other values of the parameters would make them

# Maximum Likelihood Estimation

Consider a family of continuous probability distributions parameterized by an unknown parameter $\theta$, associated with a known probability density function $p_\theta$.

Draw a vector $y = (y_1, y_2, \ldots, y_n)$ from this distribution, and then using $p_\theta$ compute the probability density associated with the observed data,

$$p_\theta(y) = p_\theta(y_1, y_2, \ldots, y_n)$$

As a function of $\theta$ with $y_1, y_2, \ldots, y_n$ fixed, this is the likelihood function

$$\mathcal{L}(\theta; y) = p_\theta(y)$$

# Maximum Likelihood Estimation

Let $\theta^\star$ be the "true value" of $\theta$.

The method of maximum likelihood estimates $\theta^\star$ by finding the value of $\theta$ that maximizes $\mathcal{L}(\theta; y)$.

This is the maximum likelihood estimator (MLE) of $\theta$:

$$\hat{\theta} = \mathrm{Arg}\max_{\theta} \mathcal{L}(\theta; y)$$

Under certain (fairly weak) regularity conditions, the MLE is "asymptotically optimal":

- The MLE is asymptotically unbiased: $E(\hat{\theta}) \underset{n\to\infty}{\longrightarrow} \theta^\star$
- The MLE is a consistant estimate of $\theta^\star$ (LLN): $\hat{\theta} \underset{n\to\infty}{\longrightarrow} \theta^\star$
- The MLE is asymptotically normal (CLT)

$$\sqrt{n}(\hat{\theta} - \theta^\star) \underset{n\to\infty}{\longrightarrow} \mathcal{N}(0, \mathcal{I}(\theta^\star)^{-1})$$

$\mathcal{I}(\theta^\star) = -E\partial^2_\theta \log \mathcal{L}(\theta^\star; y)/n$ is the *Fisher Information Matrix*

- The MLE is asymptotically efficient, (Cramér-Rao)
  This means that no asymptotically unbiased estimator has lower asymptotic mean squared error than the MLE.

Under certain (fairly weak) regularity conditions, the MLE is "asymptotically optimal":

- The MLE is asymptotically unbiased: $E(\hat{\theta}) \underset{n\to\infty}{\longrightarrow} \theta^\star$
- The MLE is a consistant estimate of $\theta^\star$ (LLN): $\hat{\theta} \underset{n\to\infty}{\longrightarrow} \theta^\star$
- The MLE is asymptotically normal (CLT)

$$\sqrt{n}(\hat{\theta} - \theta^\star) \underset{n\to\infty}{\longrightarrow} \mathcal{N}(0, \mathcal{I}(\theta^\star)^{-1})$$

$\mathcal{I}(\theta^\star) = -E\partial_\theta^2 \log \mathcal{L}(\theta^\star; y)/n$ is the *Fisher Information Matrix*

- The MLE is asymptotically efficient, (Cramér-Rao)
  This means that no asymptotically unbiased estimator has lower asymptotic mean squared error than the MLE.

Under certain (fairly weak) regularity conditions, the MLE is "asymptotically optimal":

- The MLE is asymptotically unbiased: $E(\hat{\theta}) \underset{n \to \infty}{\longrightarrow} \theta^\star$
- The MLE is a consistant estimate of $\theta^\star$ (LLN): $\hat{\theta} \underset{n \to \infty}{\longrightarrow} \theta^\star$
- The MLE is asymptotically normal (CLT)

$$\sqrt{n}(\hat{\theta} - \theta^\star) \underset{n \to \infty}{\longrightarrow} \mathcal{N}(0, \mathcal{I}(\theta^\star)^{-1})$$

$\mathcal{I}(\theta^\star) = -E\partial_\theta^2 \log \mathcal{L}(\theta^\star; y)/n$ is the *Fisher Information Matrix*

- The MLE is asymptotically efficient, (Cramér-Rao)
  This means that no asymptotically unbiased estimator has lower asymptotic mean squared error than the MLE.

Under certain (fairly weak) regularity conditions, the MLE is "asymptotically optimal":

- The MLE is asymptotically unbiased: $E(\hat{\theta}) \underset{n\to\infty}{\longrightarrow} \theta^\star$
- The MLE is a consistant estimate of $\theta^\star$ (LLN): $\hat{\theta} \underset{n\to\infty}{\longrightarrow} \theta^\star$
- The MLE is asymptotically normal (CLT)

$$\sqrt{n}(\hat{\theta} - \theta^\star) \underset{n\to\infty}{\longrightarrow} \mathcal{N}(0, \mathcal{I}(\theta^\star)^{-1})$$

$\mathcal{I}(\theta^\star) = -E\partial_\theta^2 \log \mathcal{L}(\theta^\star; y)/n$ is the *Fisher Information Matrix*

- The MLE is asymptotically efficient, (Cramér-Rao)
  This means that no asymptotically unbiased estimator has lower asymptotic mean squared error than the MLE.

# The regression model
## Maximum likelihood estimation

$$\begin{aligned} y_j &= f(x_j, \beta) + \varepsilon_j \quad , \quad 1 \leq j \leq n \\ \varepsilon_j &\sim_{i.i.d.} \mathcal{N}(0, \sigma^2) \end{aligned}$$

$$y_j \sim \mathcal{N}(f(x_j, \beta), \sigma^2)$$

$$\begin{aligned} \mathcal{L}(\theta; y) &= \prod_{j=1}^{n} \mathcal{L}(\theta; y_j) = \prod_{j=1}^{n} p_\theta(y_j) \\ &= \prod_{j=1}^{n} \left(2\pi\sigma^2\right)^{-\frac{1}{2}} e^{-\frac{1}{2\sigma^2}(y_j - f(x_j,\beta))^2} \\ &= \left(2\pi\sigma^2\right)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{j=1}^{n}(y_j - f(x_j,\beta))^2} \end{aligned}$$

$$\begin{aligned}
\hat{\beta} &= \mathrm{Arg}\max_{\beta} \mathcal{L}(\beta; y) \\
&= \mathrm{Arg}\max_{\beta} \left\{ \left(2\pi\sigma^2\right)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{j=1}^{n}(y_j - f(x_j,\beta))^2} \right\} \\
&= \mathrm{Arg}\min_{\beta} \sum_{j=1}^{n} (y_j - f(x_j, \beta))^2
\end{aligned}$$

(Maximum Likelihood estimate of $\beta$ = Least-Square estimate of $\beta$)

# The linear regression model

$$\begin{aligned} y_1 &= x_{11}\beta_1 + x_{12}\beta_2 + \ldots + x_{1p}\beta_p + \varepsilon_1 \\ y_2 &= x_{21}\beta_1 + x_{22}\beta_2 + \ldots + x_{2p}\beta_p + \varepsilon_2 \\ &\vdots \\ y_n &= x_{n1}\beta_1 + x_{n2}\beta_2 + \ldots + x_{np}\beta_p + \varepsilon_n \end{aligned}$$

$$\boxed{Y = X\beta + \varepsilon}$$

$$\begin{array}{rcl} y & = & X\beta + \varepsilon \\ \varepsilon_j & \sim & \mathcal{N}(0, \sigma^2) \\ \theta & = & (\beta, \sigma^2) \end{array}$$

$$y \sim \mathcal{N}(X\beta, \sigma^2 I_n)$$

$$\mathcal{L}(\theta; y) = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\|y - X\beta\|^2}$$

$$\hat{\beta} = \mathrm{Arg}\max_{\beta} \mathcal{L}(\beta; y) = \mathrm{Arg}\min_{\beta} \|y - X\beta\|^2$$

$$y = X\beta + \varepsilon$$

$$\begin{aligned}
\hat{\beta} &= \operatorname{Arg}\min_{\beta} \|y - X\beta\|^2 \\
&= (X'X)^{-1}X'y \\
&= (X'X)^{-1}X'(X\beta + \varepsilon) \\
&= \beta + (X'X)^{-1}X'\varepsilon \\
E(\hat{\beta}) &= \beta \\
Var(\hat{\beta}) &= \sigma^2(X'X)^{-1}
\end{aligned}$$

$$\begin{aligned}
-\log \mathcal{L}(\theta^\star; y) &= \frac{n}{2}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}\|y - X\beta\|^2 \\
\mathcal{I}(\beta) &= -\frac{1}{n}E\partial^2_\beta \log \mathcal{L}(\beta, \sigma^2; y) = \frac{1}{n\sigma^2}(X'X)
\end{aligned}$$

$$y = X\beta + \varepsilon$$

$$\begin{aligned}
\hat{\beta} &= \operatorname{Arg\,min}_{\beta} \|y - X\beta\|^2 \\
&= (X'X)^{-1}X'y \\
&= (X'X)^{-1}X'(X\beta + \varepsilon) \\
&= \beta + (X'X)^{-1}X'\varepsilon \\
E(\hat{\beta}) &= \beta \\
Var(\hat{\beta}) &= \sigma^2(X'X)^{-1}
\end{aligned}$$

$$\begin{aligned}
-\log \mathcal{L}(\theta^\star; y) &= \frac{n}{2}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}\|y - X\beta\|^2 \\
\mathcal{I}(\beta) &= -\frac{1}{n}E\partial^2_\beta \log \mathcal{L}(\beta, \sigma^2; y) = \frac{1}{n\sigma^2}(X'X)
\end{aligned}$$

$$y = X\beta + \varepsilon$$

Let $V = Var(\hat{\beta}) = \sigma^2(X'X)^{-1}$ be the variance-covariance matrix of $\hat{\beta}$.

The diagonal elements of V are the variances of the components of $\hat{\beta}$:

- $V_{k,k}$ is the variance of $\hat{\beta}_k$
- $\sqrt{V_{k,k}}$ is the *standard error* (*s.e.*) of $\hat{\beta}_k$
- 90% confidence interval for $\beta_k$:

$$[\hat{\beta}_k - 1.645\sqrt{V_{k,k}}; \hat{\beta}_k + 1.645\sqrt{V_{k,k}}]$$

# The non linear regression model

$$
\begin{aligned}
y_1 &= f(x_{11}\beta_1, x_{12}\beta_2, \ldots, x_{1p}\beta_p) + \varepsilon_1 \\
y_2 &= f(x_{21}\beta_1, x_{22}\beta_2, \ldots, x_{2p}\beta_p) + \varepsilon_2 \\
&\vdots \\
y_n &= f(x_{n1}\beta_1, x_{n2}\beta_2, \ldots, x_{np}\beta_p) + \varepsilon_n
\end{aligned}
$$

$$
Y = f(X, \beta) + \varepsilon
$$

$$\begin{aligned} y &= f(X, \beta) + \varepsilon \\ \varepsilon_j &\sim \mathcal{N}(0, \sigma^2) \end{aligned}$$

$$y \sim \mathcal{N}(f(X, \beta), \sigma^2 I_n)$$

$$\mathcal{L}(\theta; y) = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\|y - f(X,\beta)\|^2}$$

$$\hat{\beta} = \mathrm{Arg}\max_{\beta} \mathcal{L}(\beta; y) = \mathrm{Arg}\min_{\beta} \|y - f(X, \beta)\|^2$$

$\Rightarrow$ No explicit expression for $\hat{\beta}$, use of optimization algorithm (Newton-Raphson)

# Model selection

1. **Compute the likelihood of the different models**
   - Let $\hat{\theta}_{\mathcal{M}}$ be the maximum likelihood estimate of $\theta$ for model $\mathcal{M}$:

$$\hat{\theta}_{\mathcal{M}} = \mathrm{Arg}\max_{\theta} \mathcal{L}_{\mathcal{M}}(\theta; y)$$

   - Let $\mathcal{L}_{\mathcal{M}} = \mathcal{L}_{\mathcal{M}}(\hat{\theta}_{\mathcal{M}}; y)$ be the likelihood of model $\mathcal{M}$.

   Selecting the "most likely" models by comparing the likelihoods favor models of high dimension (with many parameters)!

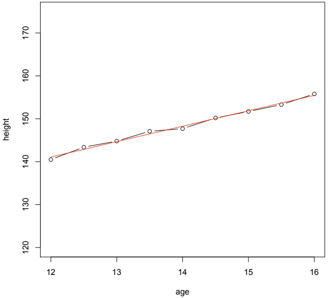
2. **Penalize the models of high dimension**

   Select the model $\hat{\mathcal{M}}$ that minimizes the penalized criteria

$$-2\mathcal{L}_{\mathcal{M}} + pen(\mathcal{M})$$

   Bayesian Information Criteria (BIC) : $pen(\mathcal{M}) = \log(n) \times \dim(\mathcal{M})$.
   Akaike Information Criteria (AIC) : $pen(\mathcal{M}) = 2\dim \times (\mathcal{M})$.

# Linear regression model
## Growth of children

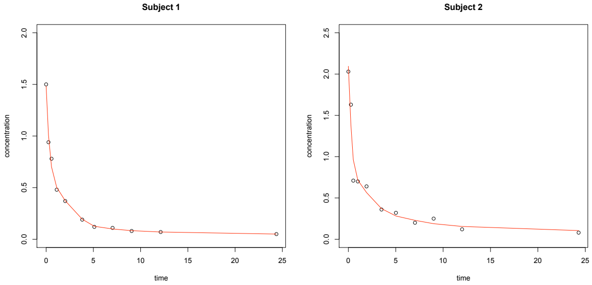



$$y_j = \beta_1 + t_j \beta_2 + \varepsilon_j$$

$$\hat{\beta}_1 = 97.97, \quad \hat{\beta}_2 = 3.59$$

# Non linear regression model

## Pharmacokinetic of Indomethacin

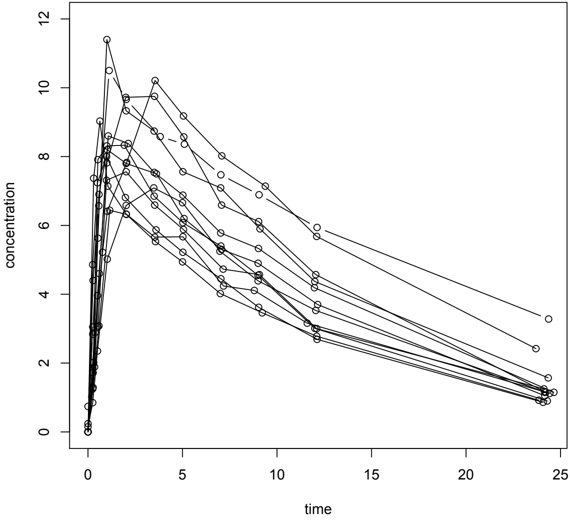



$$y_j = \beta_1 e^{-\beta_2 t_j} + \beta_3 e^{-\beta_4 t_j} + \varepsilon_j$$

| Subject | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\beta}_3$ | $\hat{\beta}_4$ |
|---|---|---|---|---|
| 1 | 2.03 | 1.78 | 0.19 | 0.17 |
| 2 | 2.83 | 2.23 | 0.50 | 0.19 |

# Outline

# Some examples of data
## Pharmacokinetics of Theophyllinee




Each individual curve is described by the same parametric model, with its own individual parameters

# Analysis of data from several subjects

Classical regression model for one subject

$$y_j = f(x_j, \psi) + \varepsilon_j \quad , \quad 1 \leq j \leq n$$

Classical regression model for several subjects

$$y_{ij} = f(x_{ij}, \psi) + \varepsilon_{ij} \quad , \quad 1 \leq i \leq N \quad , \quad 1 \leq j \leq n_i$$

$y_{ij} \in \mathbb{R}$ is the $j$th observation of subject $i$,

$N$ is the number of subjects

$n_i$ is the number of observations of subject $i$.

$x_{ij}$ are the regression variables, or design variables

# Analysis of data from several subjects

Classical regression model for one subject

$$y_j = f(x_j, \psi) + \varepsilon_j \quad , \quad 1 \leq j \leq n$$

Classical regression model for several subjects

$$y_{ij} = f(x_{ij}, \psi) + \varepsilon_{ij} \quad , \quad 1 \leq i \leq N \quad , \quad 1 \leq j \leq n_i$$

$y_{ij} \in \mathbb{R}$ is the $j$th observation of subject $i$,

$N$ is the number of subjects

$n_i$ is the number of observations of subject $i$.

$x_{ij}$ are the regression variables, or design variables

# Analysis of data from several subjects

Classical regression model for one subject

$$y_j = f(x_j, \psi) + \varepsilon_j \quad , \quad 1 \leq j \leq n$$

Classical regression model for several subjects

$$y_{ij} = f(x_{ij}, \psi) + \varepsilon_{ij} \quad , \quad 1 \leq i \leq N \quad , \quad 1 \leq j \leq n_i$$

$y_{ij} \in \mathbb{R}$ is the $j$th observation of subject $i$,

$N$ is the number of subjects

$n_i$ is the number of observations of subject $i$.

$x_{ij}$ are the regression variables, or design variables

# Analysis of data from several subjects

Classical regression model for one subject

$$y_j = f(x_j, \psi) + \varepsilon_j \quad , \quad 1 \leq j \leq n$$

Classical regression model for several subjects

$$y_{ij} = f(x_{ij}, \psi) + \varepsilon_{ij} \quad , \quad 1 \leq i \leq N \quad , \quad 1 \leq j \leq n_i$$

$y_{ij} \in \mathbb{R}$ is the $j$th observation of subject $i$,

$N$ is the number of subjects

$n_i$ is the number of observations of subject $i$.

$x_{ij}$ are the regression variables, or design variables

# Analysis of data from several subjects

Classical regression model for one subject

$$y_j = f(x_j, \psi) + \varepsilon_j \quad , \quad 1 \leq j \leq n$$

Classical regression model for several subjects

$$y_{ij} = f(x_{ij}, \psi) + \varepsilon_{ij} \quad , \quad 1 \leq i \leq N \quad , \quad 1 \leq j \leq n_i$$

$y_{ij} \in \mathbb{R}$ is the $j$th observation of subject $i$,

$N$ is the number of subjects

$n_i$ is the number of observations of subject $i$.

$x_{ij}$ are the regression variables, or design variables

# Analysis of data from several subjects

Classical regression model for one subject

$$y_j = f(x_j, \psi) + \varepsilon_j \quad , \quad 1 \leq j \leq n$$

Classical regression model for several subjects

$$y_{ij} = f(x_{ij}, \psi) + \varepsilon_{ij} \quad , \quad 1 \leq i \leq N \quad , \quad 1 \leq j \leq n_i$$

$y_{ij} \in \mathbb{R}$ is the $j$th observation of subject $i$,

$N$ is the number of subjects

$n_i$ is the number of observations of subject $i$.

$x_{ij}$ are the regression variables, or design variables

# The mixed effects model

[Pinheiro and Bates, 2002]

$$y_{ij} = f(x_{ij}, \psi_i) + \varepsilon_{ij} \quad , 1 \leq i \leq N \quad , \quad 1 \leq j \leq n_i$$

$y_{ij} \in \mathbb{R}$ is the $j$th observation of subject $i$,

$N$ is the number of subjects

$n_i$ is the number of observations of subject $i$.

The regression variables, or design variables, $(x_{ij})$ are **known**,

The individual parameters $(\psi_i)$ are **unknown**.

# The mixed effects model

$$\boxed{y_{ij} = f(x_{ij}, \psi_i) + \varepsilon_{ij} \quad , \; 1 \leq i \leq N \quad , \quad 1 \leq j \leq n_i}$$

The $(\psi_i)$ and the $(\varepsilon_{ij})$ are modelized as sequences of *random variables*.

The goal of the modeler is to develop simultaneously two kinds of models:

(1) The structural model $f$
(2) The statistical model

# The mixed effects model

$$y_{ij} = f(x_{ij}, \psi_i) + \varepsilon_{ij} \quad , \; 1 \leq i \leq N \quad , \quad 1 \leq j \leq n_i$$

(2) **The statistical model** aims to explain the variability observed in the data:

- the residual error model: distribution of $(\varepsilon_{ij})$
- the model of the individual parameters: distribution of $(\psi_i)$

$$\psi_i = h(C_i, \beta, \eta_i)$$

$C_i$ is a vector of covariates
$\beta$ is a vector of fixed effects
$\eta_i$ is a vector of random effects

# The individual parameters

$$\psi_i = h(C_i, \beta, \eta_i)$$

examples:

- additive random effects ($C_i = 1$)

$$\psi_i = \beta + \eta_i$$

- effect of a covariate ($\beta = [\beta_1, \beta_2]$, $C_i = [1, \text{sex}_i]$)

$$\psi_i = \beta_1 + \beta_2 \text{sex}_i + \eta_i$$

- multiplicative random effects

$$\psi_i = \beta \, e^{\eta_i}$$

- "partial" vector of random effects

$$\psi_i = (\psi_{i1}, \psi_{i2}) = (\beta_1 + \eta_{i1}, \beta_2)$$

# The individual parameters

$$\boxed{\psi_i = h(C_i, \beta, \eta_i)}$$

examples:

- additive random effects ($C_i = 1$)

$$\psi_i = \beta + \eta_i$$

- effect of a covariate ($\beta = [\beta_1, \beta_2]$, $C_i = [1, \text{sex}_i]$)

$$\psi_i = \beta_1 + \beta_2 \text{sex}_i + \eta_i$$

- multiplicative random effects

$$\psi_i = \beta\, e^{\eta_i}$$

- "partial" vector of random effects

$$\psi_i = (\psi_{i1}, \psi_{i2}) = (\beta_1 + \eta_{i1}, \beta_2)$$

# The individual parameters

$$\boxed{\psi_i = h(C_i, \beta, \eta_i)}$$

examples:

- additive random effects ($C_i = 1$)

$$\psi_i = \beta + \eta_i$$

- effect of a covariate ($\beta = [\beta_1, \beta_2]$, $C_i = [1, \text{sex}_i]$)

$$\psi_i = \beta_1 + \beta_2 \text{sex}_i + \eta_i$$

- multiplicative random effects

$$\psi_i = \beta\, e^{\eta_i}$$

- "partial" vector of random effects

$$\psi_i = (\psi_{i1}, \psi_{i2}) = (\beta_1 + \eta_{i1}, \beta_2)$$

# The individual parameters

$$\boxed{\psi_i = h(C_i, \beta, \eta_i)}$$

examples:

- additive random effects ($C_i = 1$)

$$\psi_i = \beta + \eta_i$$

- effect of a covariate ($\beta = [\beta_1, \beta_2],\ C_i = [1, \mathrm{sex}_i]$)

$$\psi_i = \beta_1 + \beta_2 \mathrm{sex}_i + \eta_i$$

- multiplicative random effects

$$\psi_i = \beta \, e^{\eta_i}$$

- "partial" vector of random effects

$$\psi_i = (\psi_{i1}, \psi_{i2}) = (\beta_1 + \eta_{i1}, \beta_2)$$

# The individual parameters

$$\boxed{\psi_i = h(C_i, \beta, \eta_i)}$$

examples:

- additive random effects ($C_i = 1$)

$$\psi_i = \beta + \eta_i$$

- effect of a covariate ($\beta = [\beta_1, \beta_2]$, $C_i = [1, \text{sex}_i]$)

$$\psi_i = \beta_1 + \beta_2 \text{sex}_i + \eta_i$$

- multiplicative random effects

$$\psi_i = \beta \, e^{\eta_i}$$

- "partial" vector of random effects

$$\psi_i = (\psi_{i1}, \psi_{i2}) = (\beta_1 + \eta_{i1}, \beta_2)$$

# The mixed effects model

$$
\begin{aligned}
y_{ij} &= f(x_{ij}, \psi_i) + \varepsilon_{ij} \quad , \; 1 \leq i \leq N \quad , \quad 1 \leq j \leq n_i \\
\psi_i &= h(C_i, \beta, \eta_i)
\end{aligned}
$$

$C_i$ is a **known** vector of covariates
$\beta$ is a **unknown** $p$-vector of fixed effects
$\eta_i$ is a **unknown** $q$-vector of random effects

$$
\begin{aligned}
\varepsilon_{ij} &\sim \mathcal{N}(0, \sigma^2) \\
\psi_i &\sim \mathcal{N}(0, \Omega)
\end{aligned}
$$

$\Omega$ is the $q \times q$ variance-covariance matrix of the random effects

(Hyper or population) parameters: $\theta = (\beta, \Omega, \sigma^2)$

Classical regression model

$$y_j = \beta_1 + \beta_2 t_j + \varepsilon_j$$

Mixed effects model

$$\begin{aligned} y_{ij} &= \psi_{i1} + \psi_{i2} t_{ij} + \varepsilon_{ij} \\ &= (\beta_1 + \eta_{i1}) + (\beta_2 + \eta_{i2}) t_{ij} + \varepsilon_{ij} \end{aligned}$$

where $\eta_i = (\eta_{i1}, \eta_{i2})' \sim \mathcal{N}(0, \Omega)$ means

$$\begin{pmatrix} \eta_{i1} \\ \eta_{i2} \end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \omega_1^2 & \omega_{12} \\ \omega_{12} & \omega_2^2 \end{pmatrix} \right)$$

# The mixed effects model

Some statistical issues:

- **Estimation:**
  - estimate the population parameters of the model $\theta$
  - estimate the individual parameters
  - compute confidence intervals
- **Model selection and model assessment:**
  - Determine if a parameter varies in the population
  - Select the best combination of covariates
  - Compare several treatments
- **Optimization of the design :**
  - Determine the design (the measurement times) that yields the most accurate estimation of the model

The maximum likelihood estimator of $\theta = (\beta, \Omega, \sigma^2)$ maximizes

$$\begin{aligned} \mathcal{L}(\theta; y) &= \prod_{i=1}^{N} \mathcal{L}_i(\theta; y_i) \\ \mathcal{L}_i(\theta; y_i) &= \int p(y_i, \eta_i; \theta) d\eta_i \\ &= \int p(y_i|\eta_i; \theta) p(\eta_i; \theta) d\eta_i \end{aligned}$$

We know that

$$\begin{aligned} y_i|\eta_i &\sim \mathcal{N}\left(f(x_i, h(C_i, \beta, \eta_i)), \sigma^2 I_{n_i}\right) \\ \eta_i &\sim \mathcal{N}(0, \Omega) \end{aligned}$$

The maximum likelihood estimator of $\theta = (\beta, \Omega, \sigma^2)$ maximizes

$$\begin{aligned} \mathcal{L}(\theta; y) &= \prod_{i=1}^{N} \mathcal{L}_i(\theta; y_i) \\ \mathcal{L}_i(\theta; y_i) &= \int p(y_i, \eta_i; \theta) d\eta_i \\ &= \int p(y_i|\eta_i; \theta) p(\eta_i; \theta) d\eta_i \end{aligned}$$

We know that

$$\begin{aligned} y_i|\eta_i &\sim \mathcal{N}\left(f(x_i, h(C_i, \beta, \eta_i)), \sigma^2 I_{n_i}\right) \\ \eta_i &\sim \mathcal{N}(0, \Omega) \end{aligned}$$

Thus

$$\mathcal{L}_i(\theta; y_i) = \int (2\pi\sigma^2)^{-\frac{n_i}{2}} e^{-\frac{1}{2\sigma^2}||y_i - f(x_i, h(C_i, \beta, \eta_i))||^2} \times (2\pi|\Omega|)^{-\frac{1}{2}} e^{-\frac{1}{2}\eta_i' \Omega^{-1} \eta_i} d\eta_i$$

Example: $\psi_i = \beta + \eta_i$

$$\mathcal{L}_i(\theta; y_i) = C \int \sigma^{-n_i} |\Omega|^{-\frac{1}{2}} e^{-\frac{1}{2\sigma^2}||y_i - f(x_i, \psi_i)||^2 - \frac{1}{2}(\psi_i - \beta)' \Omega^{-1} (\psi_i - \beta)} d\psi_i$$

Thus

$$\mathcal{L}_i(\theta; y_i) = \int (2\pi\sigma^2)^{-\frac{n_i}{2}} e^{-\frac{1}{2\sigma^2}||y_i - f(x_i, h(C_i, \beta, \eta_i))||^2} \times (2\pi|\Omega|)^{-\frac{1}{2}} e^{-\frac{1}{2}\eta_i'\Omega^{-1}\eta_i} d\eta_i$$

Example: $\psi_i = \beta + \eta_i$

$$\mathcal{L}_i(\theta; y_i) = C \int \sigma^{-n_i} |\Omega|^{-\frac{1}{2}} e^{-\frac{1}{2\sigma^2}||y_i - f(x_i, \psi_i)||^2 - \frac{1}{2}(\psi_i - \beta)'\Omega^{-1}(\psi_i - \beta)} d\psi_i$$

# The mixed effects model
## Estimation of the individual parameters

Assume that $\theta = (\beta, \Omega, \sigma^2)$ is known (or previously estimated) $\hat{\psi}_i$ maximizes the conditional distribution $p(\psi_i|y_i;\theta)$

$$\begin{aligned} p(\psi_i|y_i;\theta) &= \frac{p(\psi_i, y_i;\theta)}{p(y_i : \theta)} \\ &= \frac{p(y_i|\psi_i;\theta)p(\psi_i;\theta)}{p(y_i : \theta)} \\ &\propto p(y_i|\psi_i;\theta)p(\psi_i;\theta) \end{aligned}$$

Example: $\psi_i = \beta + \eta_i$

$$p(\psi_i|y_i;\theta) = Ce^{-\frac{1}{2}||y_i - f(x_i,\psi_i)||^2 - \frac{1}{2}(\psi_i-\beta)'\Omega^{-1}(\psi_i-\beta)}$$

$\hat{\psi}_i$ minimizes a penalized least-square criteria:

$$\hat{\psi}_i = \arg\min_{\psi} \left( ||y_i - f(x_i, \psi_i)||^2 - \frac{1}{2}(\psi_i - \beta)'\Omega^{-1}(\psi_i - \beta) \right)$$

Assume that $\theta = (\beta, \Omega, \sigma^2)$ is known (or previously estimated) $\hat{\psi}_i$ maximizes the conditional distribution $p(\psi_i | y_i; \theta)$

$$\begin{aligned} p(\psi_i | y_i; \theta) &= \frac{p(\psi_i, y_i; \theta)}{p(y_i : \theta)} \\ &= \frac{p(y_i | \psi_i; \theta) p(\psi_i; \theta)}{p(y_i : \theta)} \\ &\propto p(y_i | \psi_i; \theta) p(\psi_i; \theta) \end{aligned}$$

Example: $\psi_i = \beta + \eta_i$

$$p(\psi_i | y_i; \theta) = C e^{-\frac{1}{2} ||y_i - f(x_i, \psi_i)||^2 - \frac{1}{2}(\psi_i - \beta)' \Omega^{-1} (\psi_i - \beta)}$$

$\hat{\psi}_i$ minimizes a penalized least-square criteria:

$$\hat{\psi}_i = \arg\min_{\psi} \left( ||y_i - f(x_i, \psi_i)||^2 - \frac{1}{2}(\psi_i - \beta)' \Omega^{-1} (\psi_i - \beta) \right)$$

# The linear mixed effects model
## Maximum likelihood estimate

$$\begin{array}{rcl} y_i & = & X_i \psi_i + \varepsilon_i \\ \varepsilon_i & \sim & \mathcal{N}(0, \sigma^2 I_{n_i}) \\ \psi_i & \sim & \mathcal{N}(\beta, \Omega) \end{array}$$

We have $y_i = X_i \beta + X_i \eta_i + \varepsilon_i$ thus by linearity

$$y_i \sim \mathcal{N}(X_i\beta, X_i\Omega X_i' + \sigma^2 I_{n_i})$$

Set $V_i = X_i\Omega X_i'/\sigma^2 + I_{n_i}$, thus the likelihood is explicit

$$\mathcal{L}(\theta; y) = \prod_{i=1}^{N} (2\pi|\sigma^2 V_i|)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - X_i\beta)' V_i^{-1}(y_i - X_i\beta)\right)$$

Computation of the MLE $\hat{\theta}$ via an optimization routine (Newton-Raphson iterations or EM algorithm)

$$
\begin{aligned}
y_i &= X_i \psi_i + \varepsilon_i \\
\varepsilon_i &\sim \mathcal{N}(0, \sigma^2 I_{n_i}) \\
\psi_i &\sim \mathcal{N}(\beta, \Omega)
\end{aligned}
$$

We have $y_i = X_i \beta + X_i \eta_i + \varepsilon_i$ thus by linearity

$$y_i \sim \mathcal{N}(X_i\beta, X_i\Omega X_i' + \sigma^2 I_{n_i})$$

Set $V_i = X_i\Omega X_i'/\sigma^2 + I_{n_i}$, thus the likelihood is explicit

$$\mathcal{L}(\theta; y) = \prod_{i=1}^{N} (2\pi|\sigma^2 V_i|)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - X_i\beta)' V_i^{-1}(y_i - X_i\beta)\right)$$

Computation of the MLE $\hat{\theta}$ via an optimization routine (Newton-Raphson iterations or EM algorithm)

# The linear mixed effects model
## Maximum likelihood estimate

$$\begin{aligned} y_i &= X_i \psi_i + \varepsilon_i \\ \varepsilon_i &\sim \mathcal{N}(0, \sigma^2 I_{n_i}) \\ \psi_i &\sim \mathcal{N}(\beta, \Omega) \end{aligned}$$

We have $y_i = X_i \beta + X_i \eta_i + \varepsilon_i$ thus by linearity

$$y_i \sim \mathcal{N}(X_i\beta, X_i\Omega X_i' + \sigma^2 I_{n_i})$$

Set $V_i = X_i\Omega X_i'/\sigma^2 + I_{n_i}$, thus the likelihood is explicit

$$\mathcal{L}(\theta; y) = \prod_{i=1}^{N} (2\pi|\sigma^2 V_i|)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - X_i\beta)' V_i^{-1}(y_i - X_i\beta)\right)$$

Computation of the MLE $\hat{\theta}$ via an optimization routine (Newton-Raphson iterations or EM algorithm)

$$\begin{aligned} y_i &= X_i \psi_i + \varepsilon_i \\ \varepsilon_i &\sim \mathcal{N}(0, \sigma^2 I_{n_i}) \\ \psi_i &\sim \mathcal{N}(\beta, \Omega) \end{aligned}$$

We have $y_i = X_i \beta + X_i \eta_i + \varepsilon_i$ thus by linearity

$$y_i \sim \mathcal{N}(X_i\beta, X_i\Omega X_i' + \sigma^2 I_{n_i})$$

Set $V_i = X_i\Omega X_i'/\sigma^2 + I_{n_i}$, thus the likelihood is explicit

$$\mathcal{L}(\theta; y) = \prod_{i=1}^{N} (2\pi|\sigma^2 V_i|)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - X_i\beta)' V_i^{-1}(y_i - X_i\beta)\right)$$

Computation of the MLE $\hat{\theta}$ via an optimization routine (Newton-Raphson iterations or EM algorithm)

$$\begin{aligned} y_i &= X_i \psi_i + \varepsilon_i \\ \varepsilon_i &\sim \mathcal{N}(0, \sigma^2 I_{n_i}) \\ \psi_i &\sim \mathcal{N}(\beta, \Omega) \end{aligned}$$

We have $y_i = X_i \beta + X_i \eta_i + \varepsilon_i$ thus by linearity

$$y_i \sim \mathcal{N}(X_i\beta, X_i\Omega X_i' + \sigma^2 I_{n_i})$$

Set $V_i = X_i\Omega X_i'/\sigma^2 + I_{n_i}$, thus the likelihood is explicit

$$\mathcal{L}(\theta; y) = \prod_{i=1}^{N} (2\pi|\sigma^2 V_i|)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - X_i\beta)' V_i^{-1}(y_i - X_i\beta)\right)$$

Computation of the MLE $\hat{\theta}$ via an optimization routine (Newton-Raphson iterations or EM algorithm)

$$
\begin{aligned}
y_i &= X_i \psi_i + \varepsilon_i \\
\varepsilon_i &\sim \mathcal{N}(0, \sigma^2 I_{n_i}) \\
\psi_i &\sim \mathcal{N}(\beta, \Omega)
\end{aligned}
$$

We have $y_i = X_i \beta + X_i \eta_i + \varepsilon_i$ thus by linearity

$$y_i \sim \mathcal{N}(X_i\beta, X_i\Omega X_i' + \sigma^2 I_{n_i})$$

Set $V_i = X_i\Omega X_i'/\sigma^2 + I_{n_i}$, thus the likelihood is explicit

$$\mathcal{L}(\theta; y) = \prod_{i=1}^{N} (2\pi|\sigma^2 V_i|)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - X_i\beta)' V_i^{-1}(y_i - X_i\beta)\right)$$

Computation of the MLE $\hat{\theta}$ via an optimization routine (Newton-Raphson iterations or EM algorithm)

Optimization much simpler using *concentrated* or *profiled* likelihood, ie likelihood as a function of $\Omega$

From

$$y_i \sim \mathcal{N}(X_i\beta, \sigma^2 V_i)$$

one can deduce

$$\hat{\beta}(\Omega) = \left(\sum_{i=1}^{N} X_i' V_i X_i\right)^{-1} \sum_{i=1}^{N} X_i' V_i^{-1} y_i$$

$$\hat{\sigma}^2(\Omega) = \frac{\sum_{i=1}^{N} \left(y_i - X_i\hat{\beta}(\Omega)\right)' V_i^{-1} \left(y_i - X_i\hat{\beta}(\Omega)\right)}{\sum_{i=1}^{N} n_i}$$

Optimization much simpler using *concentrated* or *profiled* likelihood, ie likelihood as a function of $\Omega$

From

$$y_i \sim \mathcal{N}(X_i\beta, \sigma^2 V_i)$$

one can deduce

$$\hat{\beta}(\Omega) = \left(\sum_{i=1}^{N} X_i' V_i X_i\right)^{-1} \sum_{i=1}^{N} X_i' V_i^{-1} y_i$$

$$\hat{\sigma}^2(\Omega) = \frac{\sum_{i=1}^{N} \left(y_i - X_i\hat{\beta}(\Omega)\right)' V_i^{-1} \left(y_i - X_i\hat{\beta}(\Omega)\right)}{\sum_{i=1}^{N} n_i}$$

Using these expressions, derive the profiled log-likelihood $\mathcal{L}(\Omega; y)$ as a function of $\Omega$:

$$\mathcal{L}(\Omega; y) = \mathcal{L}(\hat{\beta}(\Omega), \Omega, \hat{\sigma}^2(\Omega); y)$$

Estimator of $\Omega$ is obtained by maximizing $\mathcal{L}(\Omega; y)$

$$\hat{\Omega} = \arg\max_{\Omega} \mathcal{L}(\Omega; y)$$

Plug in estimators of $\beta$ and $\sigma^2$

$$\begin{aligned} \hat{\beta} &= \hat{\beta}(\hat{\Omega}) \\ \hat{\sigma}^2 &= \hat{\sigma}^2(\hat{\Omega}) \end{aligned}$$

Using these expressions, derive the profiled log-likelihood $\mathcal{L}(\Omega; y)$ as a function of $\Omega$:

$$\mathcal{L}(\Omega; y) = \mathcal{L}(\hat{\beta}(\Omega), \Omega, \hat{\sigma}^2(\Omega); y)$$

Estimator of $\Omega$ is obtained by maximizing $\mathcal{L}(\Omega; y)$

$$\hat{\Omega} = \arg\max_{\Omega} \mathcal{L}(\Omega; y)$$

Plug in estimators of $\beta$ and $\sigma^2$

$$\begin{aligned} \hat{\beta} &= \hat{\beta}(\hat{\Omega}) \\ \hat{\sigma}^2 &= \hat{\sigma}^2(\hat{\Omega}) \end{aligned}$$

MLE $\hat{\Omega}$ and $\hat{\sigma}^2$ underestimate the parameters$\Omega$ and $\sigma^2$

[Patterson and Thompson, 1971] proposes the *Restricted maximum likelihood* (REML) estimates by maximizing

$$\mathcal{L}_R(\Omega, \sigma^2; y) = \int \mathcal{L}(\beta, \Omega, \sigma^2; y) d\beta$$

Equivalent in a Bayesian framework to assume a uniform prior distribution for the fixed effects $\beta$

$$
\begin{aligned}
\mathcal{L}(\theta; y) &= \prod_{i=1}^{N} \int p(y_i, \eta_i; \theta) d\eta_i \\
&= \prod_{i=1}^{N} \int p(y_i | \eta_i; \theta) p(\eta_i; \theta) d\eta_i \\
&= \prod_{i=1}^{N} \int \frac{1}{(2\pi|\sigma^2|)^{\frac{n_i}{2}}} e^{-\frac{1}{2\sigma^2}(y_i - X_i\beta - X_i\eta_i)'(y_i - X_i\beta - X_i\eta_i)} \\
&\quad \frac{1}{(2\pi|\Omega|)^{\frac{1}{2}}} e^{-\frac{1}{2}\eta_i'\Omega^{-1}\eta_i} d\eta_i
\end{aligned}
$$

Introduce

$$\Delta'\Delta = \frac{\Omega^{-1}}{\sigma^2}, \tilde{y}_i = \begin{bmatrix} y_i \\ 0 \end{bmatrix}, \tilde{X}_i = \begin{bmatrix} X_i \\ 0 \end{bmatrix}, \tilde{Z}_i = \begin{bmatrix} X_i \\ \Delta \end{bmatrix}$$

$$\begin{aligned} \mathcal{L}(\theta; y) &= \prod_{i=1}^{N} \int \frac{1}{(2\pi|\sigma^2|)^{\frac{n_i}{2}}} e^{-\frac{1}{2\sigma^2}(y_i - X_i\beta - X_i\eta_i)'(y_i - X_i\beta - X_i\eta_i)} \\ &\quad \frac{1}{(2\pi|\Omega|)^{\frac{1}{2}}} e^{-\frac{1}{2}\eta_i'\Omega^{-1}\eta_i} d\eta_i \\ &= \prod_{i=1}^{N} C \int e^{-\frac{1}{2\sigma^2}(\tilde{y}_i - \tilde{X}_i\beta - \tilde{Z}_i\eta_i)'(\tilde{y}_i - \tilde{X}_i\beta - \tilde{Z}_i\eta_i)} d\eta_i \end{aligned}$$

then by linearity of the model

$$\hat{\eta}_i = (\tilde{Z}_i'\tilde{Z}_i)^{-1}\tilde{Z}_i'(\tilde{y}_i - \tilde{X}_i\hat{\beta})$$

Introduce

$$\Delta'\Delta = \frac{\Omega^{-1}}{\sigma^2}, \tilde{y}_i = \begin{bmatrix} y_i \\ 0 \end{bmatrix}, \tilde{X}_i = \begin{bmatrix} X_i \\ 0 \end{bmatrix}, \tilde{Z}_i = \begin{bmatrix} X_i \\ \Delta \end{bmatrix}$$

$$\begin{aligned} \mathcal{L}(\theta; y) &= \prod_{i=1}^{N} \int \frac{1}{(2\pi|\sigma^2|)^{\frac{n_i}{2}}} e^{-\frac{1}{2\sigma^2}(y_i - X_i\beta - X_i\eta_i)'(y_i - X_i\beta - X_i\eta_i)} \\ &\quad \frac{1}{(2\pi|\Omega|)^{\frac{1}{2}}} e^{-\frac{1}{2}\eta_i'\Omega^{-1}\eta_i} d\eta_i \\ &= \prod_{i=1}^{N} C \int e^{-\frac{1}{2\sigma^2}(\tilde{y}_i - \tilde{X}_i\beta - \tilde{Z}_i\eta_i)'(\tilde{y}_i - \tilde{X}_i\beta - \tilde{Z}_i\eta_i)} d\eta_i \end{aligned}$$

then by linearity of the model

$$\hat{\eta}_i = (\tilde{Z}_i'\tilde{Z}_i)^{-1}\tilde{Z}_i'(\tilde{y}_i - \tilde{X}_i\hat{\beta})$$

# The non-linear mixed effects model

$$y_i = f(X_i, \psi_i) + \varepsilon_i = f(X_i, \beta + \eta_i) + \varepsilon_i$$

$$\mathcal{L}(\theta; y_i) = \int \frac{(2\pi|\sigma^2|)^{-\frac{n_i}{2}}}{(2\pi|\Omega|)^{\frac{1}{2}}} e^{-\frac{1}{2\sigma^2}(y_i - f(X_i, \beta+\eta_i))'(y_i - f(X_i, \beta+\eta_i)) - \frac{1}{2}\eta_i' \Omega^{-1} \eta_i} d\eta_i$$

The likelihood has no explicit form because of the non linearity of the regression function $f$ with respect to $\eta_i$

Existing methods are based on approximations or numerical computations of the likelihood

Principle: linearization of $f$ to come down to a linear mixed effects model

- First order methods (FO) [Beal and Sheiner, 1982]
  - linearization of $f$ around $\beta$
  - NONMEM software
- First order conditional methods (FOCE) [Lindstrom and Bates, 1990]
  - linearization of $f$ around $\psi_i$
  - NONMEM, SAS, proc NLMIXED, Splus/R function nlme

Principle: linearization of $f$ to come down to a linear mixed effects model

- First order methods (FO) [Beal and Sheiner, 1982]
  - linearization of $f$ around $\beta$
  - NONMEM software
- First order conditional methods (FOCE) [Lindstrom and Bates, 1990]
  - linearization of $f$ around $\psi_i$
  - NONMEM, SAS, proc NLMIXED, Splus/R function nlme

Principle: linearization of $f$ to come down to a linear mixed effects model

- First order methods (FO) [Beal and Sheiner, 1982]
  - linearization of $f$ around $\beta$
  - NONMEM software
- First order conditional methods (FOCE) [Lindstrom and Bates, 1990]
  - linearization of $f$ around $\psi_i$
  - NONMEM, SAS, proc NLMIXED, Splus/R function nlme

# The non-linear mixed effects model
## First order method

Linearization of $f$ around $\beta$

$$\begin{aligned} f(x_{ij}, \psi_i) &= f(x_{ij}, \beta + \eta_i) \\ &= f(x_{ij}, \beta) + \frac{\partial f}{\partial \psi}(x_{ij}, \beta)\, \eta_i + o(\eta_i^2) \end{aligned}$$

An (approximated) model is deduced

$$y_{ij} = f(x_{ij}, \beta) + \frac{\partial f}{\partial \psi}(x_{ij}, \beta)\, \eta_i + \varepsilon_{ij}$$

A linear mixed effects model is defined by plugging in a previously estimated value of $\beta$

$$y_{ij} = f(x_{ij}, \hat{\beta}) + \frac{\partial f}{\partial \psi}(x_{ij}, \hat{\beta})\, \eta_i + \varepsilon_{ij}$$

Linearization of $f$ around $\beta$

$$\begin{aligned} f(x_{ij}, \psi_i) &= f(x_{ij}, \beta + \eta_i) \\ &= f(x_{ij}, \beta) + \frac{\partial f}{\partial \psi}(x_{ij}, \beta)\, \eta_i + o(\eta_i^2) \end{aligned}$$

An (approximated) model is deduced

$$y_{ij} = f(x_{ij}, \beta) + \frac{\partial f}{\partial \psi}(x_{ij}, \beta)\, \eta_i + \varepsilon_{ij}$$

A linear mixed effects model is defined by plugging in a previously estimated value of $\beta$

$$y_{ij} = f(x_{ij}, \hat{\beta}) + \frac{\partial f}{\partial \psi}(x_{ij}, \hat{\beta})\, \eta_i + \varepsilon_{ij}$$

Linearization of $f$ around $\beta$

$$\begin{aligned} f(x_{ij}, \psi_i) &= f(x_{ij}, \beta + \eta_i) \\ &= f(x_{ij}, \beta) + \frac{\partial f}{\partial \psi}(x_{ij}, \beta)\, \eta_i + o(\eta_i^2) \end{aligned}$$

An (approximated) model is deduced

$$y_{ij} = f(x_{ij}, \beta) + \frac{\partial f}{\partial \psi}(x_{ij}, \beta)\, \eta_i + \varepsilon_{ij}$$

A linear mixed effects model is defined by plugging in a previously estimated value of $\beta$

$$y_{ij} = f(x_{ij}, \hat{\beta}) + \frac{\partial f}{\partial \psi}(x_{ij}, \hat{\beta})\, \eta_i + \varepsilon_{ij}$$

Iterative algorithm

1. Penalized nonlinear least squares (PNLS) step: with current estimate $\hat{\Omega}$ and $\hat{\sigma}^2$, conditional modes of $\beta$ and $\eta_i$ obtained by minimizing

$$\sum_{i=1}^{N}(y_i - f(X_i, \beta + \eta_i)'(y_i - f(X_i, \beta + \eta_i)) + \hat{\sigma}^2 \eta_i' \hat{\Omega}^{-1} \eta_i$$

2. Linear mixed effects (LME) step: first order Taylor expansion of $f$ around $\hat{\beta}$

$$y_i \approx f(X_i, \hat{\beta} + \hat{\eta}_i) + \frac{\partial f}{\partial \psi_i}(X_i, \hat{\beta})\, \eta_i + \varepsilon_i$$

$\Rightarrow$ MLE estimates of $\Omega$ and $\sigma^2$

Iterative algorithm

1. Penalized nonlinear least squares (PNLS) step: with current estimate $\hat{\Omega}$ and $\hat{\sigma}^2$, conditional modes of $\beta$ and $\eta_i$ obtained by minimizing

$$\sum_{i=1}^{N}(y_i - f(X_i, \beta + \eta_i)'(y_i - f(X_i, \beta + \eta_i)) + \hat{\sigma}^2 \eta_i' \hat{\Omega}^{-1} \eta_i$$

2. Linear mixed effects (LME) step: first order Taylor expansion of $f$ around $\hat{\psi}_i = \hat{\beta} + \hat{\eta}_i$

$$y_i \approx f(X_i, \hat{\psi}_i) + \frac{\partial f}{\partial \psi_i}(X_i, \hat{\psi}_i)\,(\psi_i - \hat{\psi}_i) + \varepsilon_i$$

$\Rightarrow$ MLE estimates of $\Omega$ and $\sigma^2$

Drawbacks

- Theoretical drawbacks: no well-known statistical properties of the algorithm
- Practical drawbacks: very sensitive to the initial guess, does not always converge, poor estimation of some parameters

Methods based on numerical approximations of the likelihood

- Laplace method [Wolfinger, 1993]
- Gaussian quadrature method [Davidian and Gallant, 1993] (SAS proc NLMIXED)

Properties

- Theoretical: maximum likelihood estimate is performed
- Practical: limited to few random effects

# Model selection

Several steps:

- Choice of the regression function $f$
- Choice of the covariate model $C_i$
- Choice of the random effects model $\Omega$: diagonal matrix, block-diagonal matrix, plain matrix, etc

# Model selection

Several steps:

- Choice of the regression function $f$
- Choice of the covariate model $C_i$
- Choice of the random effects model $\Omega$: diagonal matrix, block-diagonal matrix, plain matrix, etc

# Model selection

Several steps:

- Choice of the regression function $f$
- Choice of the covariate model $C_i$
- Choice of the random effects model $\Omega$: diagonal matrix, block-diagonal matrix, plain matrix, etc

# Model selection

Several steps:

- Choice of the regression function $f$
- Choice of the covariate model $C_i$
- Choice of the random effects model $\Omega$: diagonal matrix, block-diagonal matrix, plain matrix, etc

# Model selection

1. **Compute the likelihood of the different models**
   - Let $\hat{\theta}_{\mathcal{M}}$ be the maximum likelihood estimate of $\theta$ for model $\mathcal{M}$:

$$\hat{\theta}_{\mathcal{M}} = \mathrm{Arg}\max_{\theta} \mathcal{L}_{\mathcal{M}}(\theta; y)$$

   - Let $\mathcal{L}_{\mathcal{M}} = \mathcal{L}_{\mathcal{M}}(\hat{\theta}_{\mathcal{M}}; y)$ be the likelihood of model $\mathcal{M}$.

   Selecting the "most likely" models by comparing the likelihoods favor models of high dimension (with many parameters)!

2. **Penalize the models of high dimension**

   Select the model $\hat{\mathcal{M}}$ that minimizes the penalized criteria

$$-2\mathcal{L}_{\mathcal{M}} + pen(\mathcal{M})$$

   Bayesian Information Criteria (BIC) : $pen(\mathcal{M}) = \log(n) \times \dim(\mathcal{M})$.
   Akaike Information Criteria (AIC) : $pen(\mathcal{M}) = 2\dim \times (\mathcal{M})$.

# Model checking

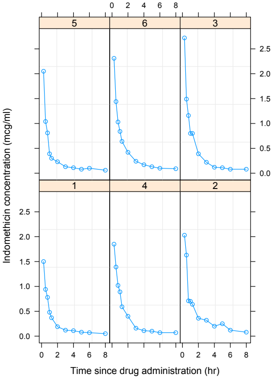
Computation of

- Population predictions: $f(x_{ij}, \hat{\beta})$
- Individual predictions: $f(x_{ij}, \hat{\psi}_i)$
- Population residuals: $y_{ij} - f(x_{ij}, \hat{\beta})$
- Individual residuals: $y_{ij} - f(x_{ij}, \hat{\psi}_i)$

Plots

- Population/Individual predictions vs observations
- Population/Individual residuals vs population/Individual predictions
- Normality of the residuals

# Pharmacokinetic of Indomethacin



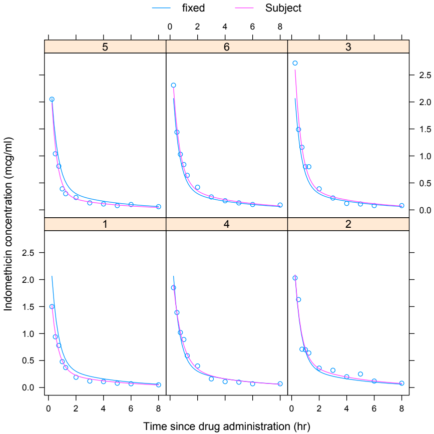

$$
\begin{aligned}
y_{ij} &= \psi_{i1} e^{-\psi_{i2} t_{ij}} + \psi_{i3} e^{-\psi_{i4} t_{ij}} + \varepsilon_{ij} \\
&= (\beta_1 + \eta_{i1}) e^{-(\beta_2 + \eta_{i2}) t_{ij}} + (\beta_3 + \eta_{i3}) e^{-(\beta_4 + \eta_{i4}) t_{ij}} + \varepsilon_{ij}
\end{aligned}
$$

$\Longrightarrow$ Choice of the covariance matrix $\Omega$

# Pharmacokinetic of Indomethacin

## Diagonal covariance matrix $\Omega$

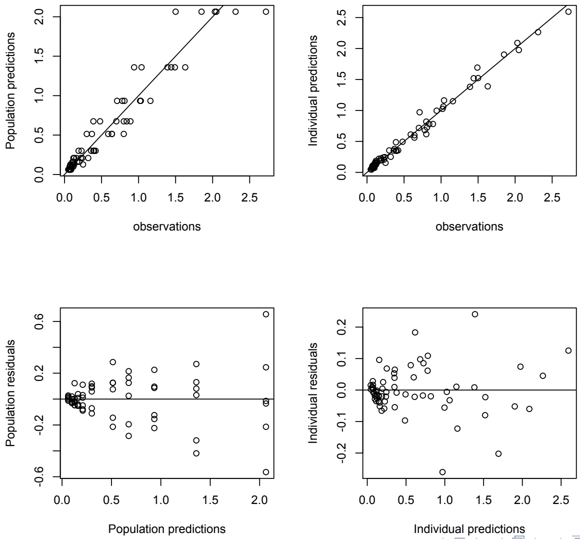
AIC = -90.24




$$\hat{\beta}_1 = 2.83, \quad \hat{\beta}_2 = 2.10, \quad \hat{\beta}_3 = 0.41, \quad \hat{\beta}_4 = 0.24$$

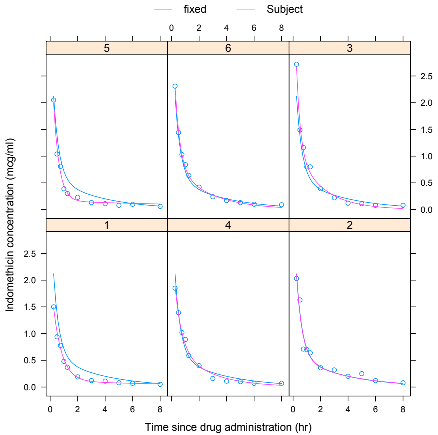
$$\hat{\omega}_{\beta_1} = 0.56, \quad \hat{\omega}_{\beta_2} = 0.34, \quad \hat{\omega}_{\beta_3} = 0.10, \quad \hat{\omega}_{\beta_4} = 10^{-6}$$

# Pharmacokinetic of Indomethacin
## Diagonal covariance matrix $\Omega$



# Pharmacokinetic of Indomethacin
## Plain covariance matrix Ω

AIC = -95.28

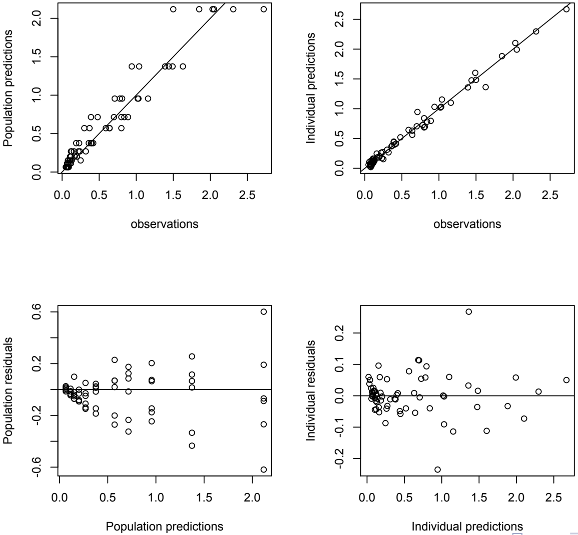



$$\hat{\beta}_1 = 2.84, \quad \hat{\beta}_2 = 2.46, \quad \hat{\beta}_3 = 0.63, \quad \hat{\beta}_4 = 0.28$$

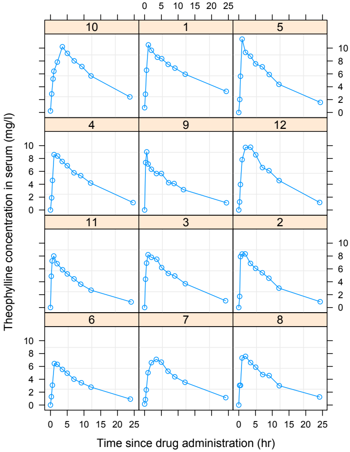
$$\hat{\omega}_{\beta_1} = 0.74, \quad \hat{\omega}_{\beta_2} = 0.61, \quad \hat{\omega}_{\beta_3} = 0.40, \quad \hat{\omega}_{\beta_4} = 0.17$$

# Pharmacokinetic of Indomethacin

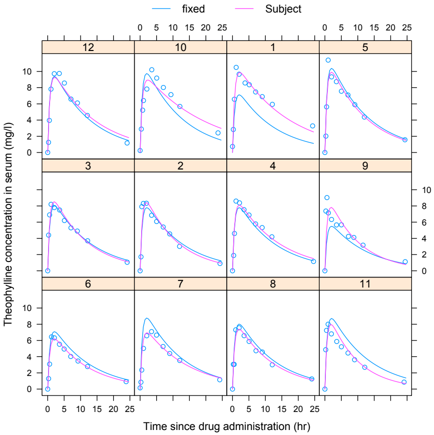
## Plain covariance matrix Ω



# Pharmacokinetic of Theophylline



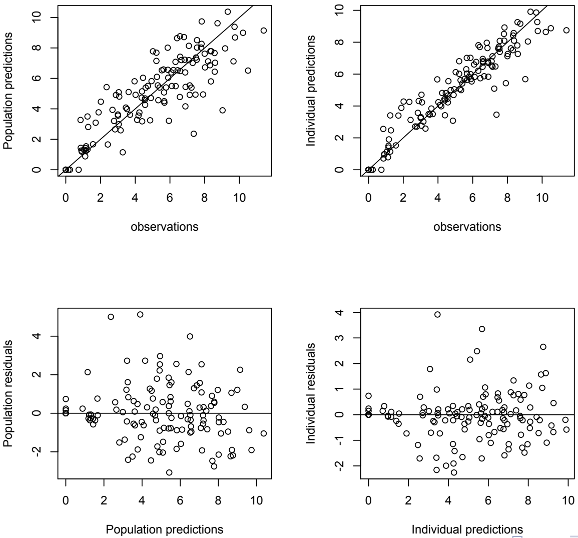

$$y_{ij} = \frac{Dose\, k_{a_i}}{V_i\,(k_{a_i} - k_{e_i})}\left(e^{-k_{e_i}\,t_{ij}} - e^{-k_{a_i}\,t_{ij}}\right) + \varepsilon_{ij}$$




$$\hat{k}_e = 0.08, \quad \hat{k}_a = 1.53, \quad \hat{V} = 0.48$$

$$\hat{\omega}_{k_e} = 0.02, \quad \hat{\omega}_V = 0.08$$

# Non linear mixed model
## Pharmacokinetic of Theophylline



# Pharmacokinetic of Theophylline

$$y_{ij} = \frac{Dose\ k_{a_i}}{V_i\,(k_{a_i} - k_{e_i})}\,(e^{-k_{e_i}\,t_{ij}} - e^{-k_{a_i}\,t_{ij}}) + \varepsilon_{ij}$$

Log parametrisation

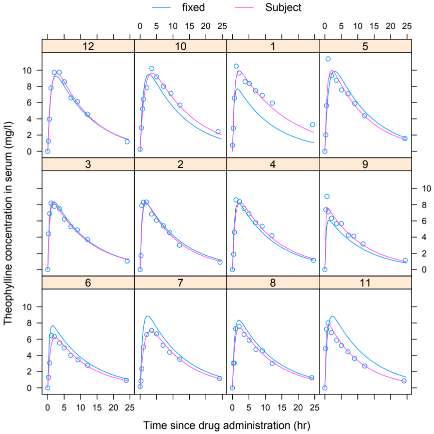
$$y_{ij} = \frac{Dose\ e^{lk_{a_i}}}{e^{lV_i}\,(e^{lk_{a_i}} - e^{lk_{e_i}})}\,(e^{-e^{lk_{e_i}}\,t_{ij}} - e^{-e^{lk_{a_i}}\,t_{ij}}) + \varepsilon_{ij}$$

$\Longrightarrow$ Final model

- Random effects on $lk_e$, $lk_a$ and $V$
- Covariate effect (weight) on $lka$

# Non linear mixed model

## Pharmacokinetic of Theophylline



# Non linear mixed model
## Pharmacokinetic of Theophylline