Introduction to General and Generalized Linear Models

Mixed effects models - II

Henrik Madsen, Jan Kloppenborg Møller, Anders Nielsen

April 16, 2012

Today

- Estimation in general linear mixed models
- Longitudinal data analysis / repeated measurements

Remember the general linear mixed model

A general linear mixed model can be presented in matrix notation by:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{U} + \varepsilon, \quad \text{where } \mathbf{U} \sim N(0, \mathbf{\Psi}) \text{ and } \varepsilon \sim N(0, \mathbf{\Sigma}).$$

- $\mathbf{Y}$ is the observation vector
- $\mathbf{X}$ is the design matrix for the fixed effects
- $\boldsymbol{\beta}$ is the vector containing the fixed effect parameters
- $\mathbf{Z}$ is the design matrix for the random effects
- $\mathbf{U}$ is the vector of random effects
    - It is assumed that $\mathbf{U} \sim N(\mathbf{0}, \mathbf{\Psi})$
    - $\text{cov}(U_i, U_j) = G_{i,j}$ (typically $\mathbf{\Psi}$ has a very simple structure (for instance diagonal))
- $\boldsymbol{\varepsilon}$ is the vector of residual errors
    - It is assumed that $\varepsilon \sim N(\mathbf{0}, \mathbf{\Sigma})$
    - $\text{cov}(\varepsilon_i, \varepsilon_j) = R_{i,j}$ (typically $\mathbf{\Sigma}$ is diagonal, but we shall later see some useful exceptions for repeated measurements)

The distribution of $\mathbf{Y}$

From the model description:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{U} + \boldsymbol{\varepsilon}, \quad \text{where } \mathbf{U} \sim N(0, \boldsymbol{\Psi}) \text{ and } \boldsymbol{\varepsilon} \sim N(0, \boldsymbol{\Sigma}).$$

We can compute the mean vector $\boldsymbol{\mu} = E(\mathbf{Y})$ and covariance matrix $\mathbf{V} = \text{var}(\mathbf{Y})$:

$$\begin{aligned}
\boldsymbol{\mu} &= E(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{U} + \boldsymbol{\varepsilon}) = \mathbf{X}\boldsymbol{\beta} \quad \text{[All other terms have mean zero]} \\
\mathbf{V} &= \text{var}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{U} + \boldsymbol{\varepsilon}) \quad \text{[from model]} \\
&= \text{var}(\mathbf{X}\boldsymbol{\beta}) + \text{var}(\mathbf{Z}\mathbf{U}) + \text{var}(\boldsymbol{\varepsilon}) \quad \text{[all terms are independent]} \\
&= \text{var}(\mathbf{Z}\mathbf{U}) + \text{var}(\boldsymbol{\varepsilon}) \quad \text{[variance of fixed effects is zero]} \\
&= \mathbf{Z}\text{var}(\mathbf{U})\mathbf{Z}^T + \text{var}(\boldsymbol{\varepsilon}) \quad \text{[}\mathbf{Z}\text{ is constant]} \\
&= \mathbf{Z}\boldsymbol{\Psi}\mathbf{Z}^T + \boldsymbol{\Sigma} \quad \text{[from model]}
\end{aligned}$$

So $\mathbf{Y}$ follows a multivariate normal distribution:

$$\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{Z}\boldsymbol{\Psi}\mathbf{Z}^T + \boldsymbol{\Sigma})$$

General linear mixed effects models

It follows from the independence of $\boldsymbol{U}$ and $\boldsymbol{\epsilon}$ that

$$\mathrm{D}\left[\begin{pmatrix}\boldsymbol{\epsilon}\\ \boldsymbol{U}\end{pmatrix}\right] = \begin{pmatrix}\boldsymbol{\Sigma} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{\Psi}\end{pmatrix}$$

The model may also be interpreted as a *hierarchical model*

$$\boldsymbol{U} \sim N(\boldsymbol{0}, \boldsymbol{\Psi})$$
$$\boldsymbol{Y}|\boldsymbol{U} = \boldsymbol{u} \sim N(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u}, \boldsymbol{\Sigma})$$

One-way model with random effects - example

The one-way model with random effects

$$Y_{ij} = \mu + U_i + e_{ij}$$

We can formulate this as

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{Z}\boldsymbol{U} + \boldsymbol{\epsilon}$$

with

$$\begin{aligned}
\boldsymbol{X} &= \mathbf{1}_N \\
\beta &= \mu \\
\boldsymbol{U} &= (U_1, U_2, \dots, U_k)^T \\
\boldsymbol{\Sigma} &= \sigma^2 \boldsymbol{I}_N \\
\boldsymbol{\Psi} &= \sigma_u^2 \boldsymbol{I}_k
\end{aligned}$$

where $\mathbf{1}_N$ is a column of 1's. The i, j'th element in the $N \times k$ dimensional matrix $\boldsymbol{Z}$ is 1, if y_{ij} belongs to the i'th group, otherwise it is zero.

One way ANOVA with random block effect

Consider again the model:

$$Y_{ij} = \mu + \alpha_i + B_j + \varepsilon_{ij},\ B_j \sim N(0, \sigma_B^2),\ \varepsilon_{ij} \sim N(0, \sigma^2),\ i = 1, 2,\ j = 1, 2, 3$$

Calculation of $\boldsymbol{\mu}$ and $\mathbf{V}$ gives:

$$\boldsymbol{\mu} = \begin{pmatrix} \mu + \alpha_1 \\ \mu + \alpha_2 \\ \mu + \alpha_1 \\ \mu + \alpha_2 \\ \mu + \alpha_1 \\ \mu + \alpha_2 \end{pmatrix}, \mathbf{V} = \begin{pmatrix} \sigma^2 + \sigma_B^2 & \sigma_B^2 & 0 & 0 & 0 & 0 \\ \sigma_B^2 & \sigma^2 + \sigma_B^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2 + \sigma_B^2 & \sigma_B^2 & 0 & 0 \\ 0 & 0 & \sigma_B^2 & \sigma^2 + \sigma_B^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 + \sigma_B^2 & \sigma_B^2 \\ 0 & 0 & 0 & 0 & \sigma_B^2 & \sigma^2 + \sigma_B^2 \end{pmatrix}$$

Notice that two observations from the same block are correlated.

The likelihood function

- The *likelihood* L is a function of model parameters and observations
- For given parameter values L returns a measure of the probability of observing $\mathbf{y}$
- The *log likelihood* ℓ for a mixed linear model is:

$$\ell(\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\psi}) \propto -\frac{1}{2}\left\{\log|\mathbf{V}(\boldsymbol{\psi})| + (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{V}(\boldsymbol{\psi}))^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right\}$$

- Here $\boldsymbol{\psi}$ is the variance parameters (σ^2 and σ_B^2 in our example)
- A natural estimate is to choose the parameters that make our observations most likely:

$$(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\psi}}) = \underset{(\boldsymbol{\beta},\boldsymbol{\psi})}{\operatorname{argmax}}\, \ell(\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\psi})$$

- This is the *maximum likelihood (ML)* method

The restricted/residual maximum likelihood method

- The maximum likelihood method tends to give (slightly) too low estimates of the random effects parameters. We say it is *biased downwards*
- The simplest example is:

  $(x_1, \ldots, x_N) \sim N(\mu, \sigma^2)$ i.i.d.
  $\hat{\sigma}^2 = \frac{1}{n} \sum (x_i - \overline{x})^2$ is the maximum likelihood estimate, but
  $\hat{\sigma}^2 = \frac{1}{n-1} \sum (x_i - \overline{x})^2$ is generally preferred, because it is *unbiased*
- The *restricted/residual maximum likelihood (REML)* method modifies the maximum likelihood method by maximizing:

$$\ell_{re}(\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\psi}) \propto -\frac{1}{2} \left\{ \log |\mathbf{V}(\boldsymbol{\psi})| + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{V}(\boldsymbol{\psi}))^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \log |\mathbf{X}^T (\mathbf{V}(\boldsymbol{\psi}))^{-1} \mathbf{X}| \right\}$$

  which gives unbiased estimates (at least in balanced cases)
- The REML method is generally preferred in mixed models

ML vs. REML, simple example

Consider again the model:

$$Y_{ij} = \mu + B_j + \varepsilon_{ij},\ B_j \sim N(0, \sigma_B^2),\ \varepsilon_{ij} \sim N(0, \sigma^2),\ i = 1, 2,\ j = 1, 2, 3$$

Calculation of $\boldsymbol{\mu}$ and $\mathbf{V}$ gives:

$$\boldsymbol{\mu} = \begin{pmatrix} \mu \\ \mu \\ \mu \\ \mu \\ \mu \\ \mu \end{pmatrix},\ \mathbf{V} = \begin{pmatrix} \sigma^2+\sigma_B^2 & \sigma_B^2 & 0 & 0 & 0 & 0 \\ \sigma_B^2 & \sigma^2+\sigma_B^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2+\sigma_B^2 & \sigma_B^2 & 0 & 0 \\ 0 & 0 & \sigma_B^2 & \sigma^2+\sigma_B^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2+\sigma_B^2 & \sigma_B^2 \\ 0 & 0 & 0 & 0 & \sigma_B^2 & \sigma^2+\sigma_B^2 \end{pmatrix}$$

Fixed effect parameters

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

$$l_\beta(\beta, \psi; \mathbf{y}) = \frac{1}{2}\mathbf{X}^T(\mathbf{V}^{-1}\mathbf{y} - \mathbf{V}^{-1}(\mathbf{X}\beta)$$

$$\mathbf{V}^{-1} = \frac{1}{\sigma^4 + 2\sigma^2\sigma_B^2}\begin{pmatrix} \sigma^2+\sigma_B^2 & -\sigma_B^2 & 0 & 0 & 0 & 0 \\ -\sigma_B^2 & \sigma^2+\sigma_B^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2+\sigma_B^2 & -\sigma_B^2 & 0 & 0 \\ 0 & 0 & -\sigma_B^2 & \sigma^2+\sigma_B^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2+\sigma_B^2 & -\sigma_B^2 \\ 0 & 0 & 0 & 0 & -\sigma_B^2 & \sigma^2+\sigma_B^2 \end{pmatrix}$$

$$\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y} = \frac{\sigma^2}{\sigma^4 + 2\sigma^2\sigma_B^2}\sum_i\sum_j y_{ij}$$

$$\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X} = \frac{6\sigma^2}{\sigma^4 + 2\sigma^2\sigma_B^2}$$

Fixed effect parameter

$$
\begin{aligned}
l_\beta(\beta, \psi; \mathbf{y}) =& 0 \Rightarrow \\
\frac{\sigma^2}{\sigma^4 + 2\sigma^2\sigma_B^2} \sum_i \sum_j y_{ij} =& \frac{6\sigma^2\beta}{\sigma^4 + 2\sigma^2\sigma_B^2} \Rightarrow \\
\hat{\beta} =& \frac{1}{6} \sum_i \sum_j y_{ij} \\
=& \frac{1}{6} \bar{\bar{y}}
\end{aligned}
$$

also

$$
\begin{aligned}
E[l_\beta(\beta, \psi; \mathbf{y})] =& \mathbf{X}^T(\mathbf{V}^{-1}E[\mathbf{y}] - \mathbf{V}^{-1}\mathbf{X}\beta) \\
=& \mathbf{X}^T(\mathbf{V}^{-1}\mathbf{X}\beta - \mathbf{V}^{-1}\mathbf{X}\beta) = 0
\end{aligned}
$$

Estimation of σ^2

$$l_{\sigma^2}(\beta, \psi; \mathbf{y}) = -\frac{1}{2}\frac{\partial}{\partial \sigma^2} \log|\mathbf{V}| - \frac{1}{2}(\mathbf{y} - \mathbf{X}\beta)^T \frac{\partial}{\partial \sigma^2}\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta)$$

$$\begin{aligned} |\mathbf{V}| &= (\sigma^4 + 2\sigma^2\sigma_B^2)^3 \\ \log|\mathbf{V}| &= 3\log(\sigma^4 + 2\sigma^2\sigma_B^2) \\ \frac{\partial}{\partial \sigma^2}\log|\mathbf{V}| &= 6\frac{\sigma^2 + \sigma_B^2}{\sigma^2 + 2\sigma^2\sigma_B^2} \end{aligned}$$

Estimation of σ^2

$$\frac{\partial \mathbf{V}^{-1}}{\partial \sigma^2} = -\frac{2(\sigma^2+\sigma_B^2)}{(\sigma^4+2\sigma^2\sigma_B^2)^2}\begin{pmatrix} \sigma^2+\sigma_B^2 & -\sigma_B^2 & 0 & 0 & 0 & 0 \\ -\sigma_B^2 & \sigma^2+\sigma_B^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2+\sigma_B^2 & -\sigma_B^2 & 0 & 0 \\ 0 & 0 & -\sigma_B^2 & \sigma^2+\sigma_B^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2+\sigma_B^2 & -\sigma_B^2 \\ 0 & 0 & 0 & 0 & -\sigma_B^2 & \sigma^2+\sigma_B^2 \end{pmatrix}$$

$$+\frac{1}{\sigma^4+2\sigma^2\sigma_B^2}\mathbf{I}$$

with $e_{ij} = y_{ij} - x_i\hat{\beta}$ we get

$$\mathbf{e}^T\frac{\partial \mathbf{V}^{-1}}{\partial \sigma^2}\mathbf{e} = -\frac{2(\sigma^2+\sigma_B^2)}{(\sigma^4+2\sigma^2\sigma_B^2)^2}\left((\sigma^2+\sigma_B^2)\sum_{i,j} e_{ij}^2 - 2\sigma_B^2\sum_j (e_{1j}\,e_{2j}\right)$$

$$+\frac{1}{\sigma^4+2\sigma^2\sigma_B^2}\sum_{i,j} e_{ij}^2$$

Estimation of σ^2

$$
\begin{aligned}
E[e_{ij}^2] =& E[(y_{ij} - \hat{\beta})^2] \\
=& \sigma^2 + \sigma_B^2 + V[\hat{\beta}] - \frac{1}{3}(\sigma^2 + 2\sigma_B^2) \\
E[e_{1j}e_{2j}] =& E[(y_{1j} - \hat{\beta})(y_{2j} - \hat{\beta})] \\
=& \sigma_B^2 + V[\hat{\beta}] - \frac{1}{3}(\sigma^2 + 2\sigma_B^2) \\
V[\hat{\beta}] =& (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1} \\
=& \frac{1}{6}(\sigma^2 + 2\sigma_B^2)
\end{aligned}
$$

and

$$
E\left[\mathbf{e}^T \frac{\partial \mathbf{V}^{-1}}{\partial \sigma^2}\mathbf{e}\right] = -\frac{6(\sigma^2 + \sigma_B^2)}{(\sigma^4 + 2\sigma^2\sigma_B^2)} - \frac{6(V[\hat{\beta}] - \frac{1}{3}(\sigma^2 + 2\sigma_B^2))}{(\sigma^2 + 2\sigma_B^2)^2}
$$

Estimation of σ^2

$$
\begin{aligned}
E[l_{\sigma^2}(\sigma^2, ..)] &= -\frac{6}{2}\frac{\sigma^2+\sigma_B^2}{\sigma^4+2\sigma^2\sigma_B^2} + \frac{6}{2}\frac{\sigma^2+\sigma_B^2}{\sigma^4+2\sigma^2\sigma_B^2} - \frac{1}{2}\frac{1}{\sigma^2+2\sigma_B^2} \\
&= -\frac{1}{2}\frac{1}{\sigma^2+2\sigma_B^2} < 0
\end{aligned}
$$

The REML correction term is

$$
\begin{aligned}
\log|\mathbf{X}\mathbf{V}^{-1}\mathbf{X}| &= \log\left(\frac{1}{V[\hat{\beta}]}\right) = \log(6) - \log(\sigma^2 + 2\sigma_B^2) \\
\frac{\partial}{\partial\sigma^2}\log|\mathbf{X}\mathbf{V}^{-1}\mathbf{X}| &= -\frac{1}{\sigma^2+2\sigma_B^2}
\end{aligned}
$$

Estimation of σ_B^2

By similar calculation

$$E[l_{\sigma_B^2}(\sigma_B^2, ..)] = -\frac{1}{\sigma^2 + 2\sigma_B^2} < 0$$

The REML correction term is

$$\log|\mathbf{X}\mathbf{V}^{-1}\mathbf{X}| = \log\left(\frac{1}{V[\hat{\beta}]}\right) = \log(6) - \log(\sigma^2 + 2\sigma_B^2)$$

$$\frac{\partial}{\partial\sigma_B^2}\log|\mathbf{X}\mathbf{V}^{-1}\mathbf{X}| = -\frac{2}{\sigma^2 + 2\sigma_B^2}$$

Estimation of random effects

- Formally, the random effects, $\boldsymbol{U}$ are not parameters in the model, and the usual likelihood approach does not make much sense for "estimating" these random quantities.
- It is, however, often of interest to assess these "latent", or "state" variables.
- We formulate a so-called *hierarchical likelihood* by writing the joint density for observable as well as unobservable random quantities.

$$
\begin{aligned}
f(\mathbf{y}, \mathbf{u}; \beta, \psi) =& f_{Y|u}(\mathbf{y}; \beta) f_U(\mathbf{u}; \psi) \\
=& \frac{1}{(\sqrt{2})^N \sqrt{|\boldsymbol{\Sigma}|}} e^{-\frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\boldsymbol{u})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\boldsymbol{u})} \times \\
& \frac{1}{(\sqrt{2})^q \sqrt{|\boldsymbol{\psi}|}} e^{-\frac{1}{2}\mathbf{u}^T \boldsymbol{\psi}^{-1} \mathbf{u}}
\end{aligned}
$$

Estimation of random effects

- *Hierarchical likelihood*

$$
\begin{aligned}
l(\boldsymbol{\beta}, \boldsymbol{\psi}, \boldsymbol{u}) = & -\frac{1}{2}\log(|\boldsymbol{\Sigma}|) - \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{u})^T \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{u}) \\
& -\frac{1}{2}\log(|\boldsymbol{\psi}|) - \frac{1}{2}\mathbf{u}^T \boldsymbol{\psi}^{-1}\mathbf{u} \\
l_{\boldsymbol{u}}(\boldsymbol{\beta}, \boldsymbol{\psi}, \boldsymbol{u}) = & \mathbf{Z}^T \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{u}) - \boldsymbol{\psi}^{-1}\mathbf{u}
\end{aligned}
$$

- By putting the derivative of the hierarchical likelihood equal to zero and solving with respect to $\boldsymbol{u}$ one finds that the estimate $\widehat{\boldsymbol{u}}$ is solution to

$$
(\boldsymbol{Z}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{Z} + \boldsymbol{\Psi}^{-1})\boldsymbol{u} = \boldsymbol{Z}^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})
$$

  where the estimate $\widehat{\boldsymbol{\beta}}$ is inserted in place of $\boldsymbol{\beta}$.

- The solution is termed the *best linear unbiased predictor*

REML or ML

- When we want to estimate model parameters - especially variance parameters - we should use REML
- But when we want to compute the likelihood ratio test we should use ML.
- The `lme()` function defaults to REML, and can use ML by specifying

```
> fit<-lme(y~A, random=~1|B, method="ML")
```

The repeated measurements setup

- Several "individuals"
- Several measurements on each individual
- Two measurements on the same individual might be correlated
- Might even be highly correlated if "close" and less correlated if "far apart"
- Typical example:
  - 20 individuals from relevant population
  - Half get drug A and half get drug B
  - Measured every week for two months

To pretend all observations are independent can lead to wrong conclusions

		Month									
Dose	Cage	1	2	3	4	5	6	7	8	9	10
1	1	20584	15439	17376	14785	11189	10366	8725	9974	9576	6849
1	2	23265	16956	16200	12934	13763	11893	9949	10490	8674	7153
1	3	17065	12429	14757	10524	11783	8828	9016	9635	8028	8099
1	4	19265	19316	20598	16619	16092	13422	10532	10614	9466	9494
1	5	21062	14095	13267	12543	12734	12268	12219	11791	10379	8463
1	6	23456	10939	13270	14089	12986	13723	11878	13338	12442	10094
1	7	13383	11899	12531	15081	14295	13650	9988	11518	11915	7844
1	8	22717	22434	23151	13163	10029	10408	9119	10188	9549	11153
1	9	17437	13950	15535	14199	11540	9568	8481	9143	8117	5765
1	10	18546	12520	15394	10137	9218	7343	6702	7173	7257	5708
2	11	18536	16827	19185	12445	13227	10412	9855	9169	9639	6853
2	12	18831	14043	16493	12562	10397	8568	8599	8818	6011	5062
2	13	15016	13765	16648	14537	13929	10778	9897	9225	9491	5523
2	14	22276	15497	22024	15616	12440	11454	10290	9456	9567	7003
2	15	18943	14834	18403	16232	13085	12679	10489	9495	10896	8836
2	16	13598	10233	13392	10457	9236	8847	9445	9501	8509	5656
2	17	20498	22136	22094	19825	18157	11452	14809	14564	14503	10643
2	18	19586	12710	12745	7294	15757	15296	14097	14308	13933	10210
2	19	11474	8108	17714	16795	17364	16766	15016	13475	14349	8698
2	20	10284	10760	15628	10692	8420	5842	6138	10271	8435	4486
3	21	18459	15805	19924	18337	24197	18790	19333	22234	18291	11595
3	22	16186	11750	16470	18637	14862	14695	14458	14228	12909	9079
3	23	9614	8319	11375	9446	13157	11153	10540	11476	8976	6123
3	24	15688	15016	20929	12706	17351	15089	14605	15952	14795	10434
3	25	15864	13169	20991	20655	19763	19180	19003	18172	15025	11790
3	26	17721	14489	19085	21333	17011	16148	15280	14762	15745	10477
3	27	17606	7558	15646	15194	13036	10316	8172	8977	8378	3962
3	28	34907	29247	35831	15093	9754	10061	9042	11732	8716	4922
3	29	15189	14046	14909	14713	14999	14201	13184	13073	14639	10330
3	30	16388	14538	17548	19416	22034	17761	14488	16068	14773	10595

Example: Activity of rats

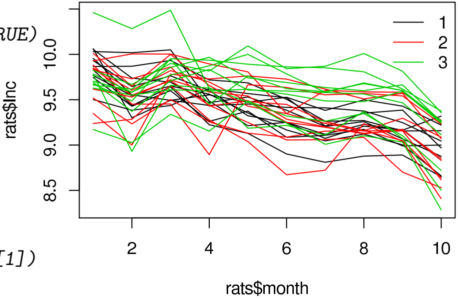
Summary of experiment:

- 3 treatments: 1, 2, 3 (concentration)
- 10 cages per treatment
- 10 contiguous months
- The response is activity ($\log$(count) of intersections of light beam during 57 hours)

```
> rats<-read.csv('rep/rats.csv', header=TRUE)
> head(rats,2)

  treatm cage month    lnc
1      1    1     1 9.9323
2      1    1     2 9.6447

> plot(rats$lnc~rats$month, type='n')
> by(rats,rats$cage,function(X){
+       lines(X$month,X$lnc, col=X$treatm[1])
+    } )->out
> legend("topright", bty='n', legend=1:3, lty='solid', col=1:3)
```



Separate analysis for each time–point

- Select a fixed time point
- The observations at that time (one from each individual) are independent
- Do a separate analysis for the observations at that time
- This is not wrong, but (possibly) a lot of information is waisted
- This can be done for several time–points, but
  - Difficult to reach a coherent conclusion
  - Sub–tests are not independent
  - Tempting to select time–points that supports out preference
  - Mass significance: If many tests are carried out at 5% level some might be significant by chance. (Bonferroni correction: Use significance level $0.05/n$ instead of 0.05)

Separate analysis of rats data

- The model at each time–step is:

$$\texttt{lnc}_i = \mu + \alpha(\texttt{treatm}_i) + \varepsilon_i \ , \quad \varepsilon_i \sim \text{i.i.d. } N(0, \sigma^2), \quad i = 1 \ldots 30$$

- To analyze this in R we, write:

```
> rats$treatm<-factor(rats$treatm)
> rats$month<-factor(rats$month)
> doone<-function(X){
+   anova(lm(lnc~treatm,data=X))
+ }
> results<-by(rats,rats$month,doone)
```

- The result of the ten tests for no treatment effect:

Month	1	2	3	4	5	6	7	8	9	10
F–value	1.22	0.27	1.02	2.30	3.87	4.10	4.70	7.29	4.09	0.88

Compare with $F_{95\%;2,27} = 3.35$ or $F_{99.5\%;2,27} = 6.49$ if Bonferroni correction is used

Analysis of summary statistic

- Choose a single measure to summarize the individual curves
- This again reduces the data set to independent observations
- Popular choices of summary measures:
  - Average over time
  - Slope in regression with time (or higher order polynomial coefficients)
  - Total increase (last point minus first point)
  - Area under curve (AUC)
  - Maximum or minimum point
- Good method with few and easily checked assumptions
- Information may be lost
- Important to choose a good summary measure

Rats data analyzed via summary measure

- The log of the total activity is chosen as summary measure
  $\texttt{lnTot} = \log(\text{Total count})$
- The one way ANOVA model becomes:

$$\texttt{lnTot}_i = \mu + \alpha(\texttt{treatm}_i) + \varepsilon_i, \quad \varepsilon_i \sim \text{ i.i.d. } N(0, \sigma^2), \quad i = 1 \ldots 30$$

- Which is easily implemented in R:

```
> fun<-function(x){
+   log(sum(exp(x)))
+ }
> ratsSum<-aggregate(lnc ~ cage+treatm, data = rats,fun)
> names(ratsSum)<-c('cage','treatm','lnTot')
> fit<-lm(lnTot~treatm,data=ratsSum)
> anova(fit)
```

- The P–value for no treatment effect in this summary model is 5.23%
- Notice the simplicity of the model and the relative few assumptions

Simple mixed model

- Add “individual” (here cage) as a random effect
- Makes measurements on same individual correlated (as we have seen)
- This model uses all observations instead of reducing to one observation per individual
- Unfortunately equally correlated no matter if they are “close” or “far apart”
- Can be considered first step in modelling the actual covariance structure
- Usually only good for short series
- This model is also known as the split–plot model for repeated measurements (with “individuals” as main–plots and the single measurements as sub–plots)

Rats data analyzed via the simple mixed model approach

- The model can now be enhanced to:

  $$\texttt{lnc}_i = \mu + \alpha(\texttt{treatm}_i) + \beta(\texttt{month}_i) + \gamma(\texttt{treatm}_i, \texttt{month}_i) + d(\texttt{cage}_i) + \varepsilon_i,$$

  with $\varepsilon_i \sim N(0, \sigma^2)$ and $d(\text{cage}_i) \sim N(0, \sigma_d^2)$ all independent.

- The covariance structure of this model is:

  $$\text{cov}(y_{i_1}, y_{i_2}) = \begin{cases} 0 & , \text{ if cage}_{i_1} \neq \text{cage}_{i_2} \\ \sigma_d^2 & , \text{ if cage}_{i_1} = \text{cage}_{i_2} \text{ and } i_1 \neq i_2 \\ \sigma_d^2 + \sigma^2 & , \text{ if } i_1 = i_2 \end{cases}$$

- This model is implemented in R by:

```
> library(nlme)
> fit.mm<-lme(lnc~month+treatm+month:treatm, random = ~1|cage, data=rats)
```

- The P–value for the interaction term is 0.0059. Significant, but is the model still too simple?

```
> fit.mm

Linear mixed-effects model fit by REML
  Data: rats
  Log-restricted-likelihood: -4.307319
  Fixed: lnc ~ month + treatm + month:treatm
    (Intercept)          month2          month3          month4          month5
       9.874280       -0.282870       -0.199010       -0.381285       -0.464289
         month6          month7          month8          month9         month10
      -0.574140       -0.711864       -0.638082       -0.724398       -0.901828
        treatm2         treatm3  month2:treatm2  month3:treatm2  month4:treatm2
      -0.168020       -0.137266        0.080240        0.243762        0.160360
 month5:treatm2  month6:treatm2  month7:treatm2  month8:treatm2  month9:treatm2
       0.217248        0.153330        0.265727        0.202875        0.246120
month10:treatm2  month2:treatm3  month3:treatm3  month4:treatm3  month5:treatm3
       0.049884        0.055665        0.285257        0.332030        0.412954
 month6:treatm3  month7:treatm3  month8:treatm3  month9:treatm3 month10:treatm3
       0.410386        0.472180        0.465124        0.443817        0.202159

Random effects:
 Formula: ~1 | cage
        (Intercept)  Residual
StdDev:   0.1657654 0.1946757

Number of Observations: 300
Number of Groups: 30
```

Pros and cons of simple approaches

Separate analysis for each time–point

+ Not wrong
– Can be confusing
– Difficult to reach coherent conclusion
– In general not very informative

Analysis of summary statistic

+ Good method with few and easily checked assumptions
– Important to choose good summary measure(s)

Simple mixed model approach

+ Good method for short series
+ Uses all observations
– Usually not good for long series

Different view on the mixed model approach

- Any linear mixed model can be expressed as:

$$\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{Z}\boldsymbol{\Psi}\mathbf{Z}^T + \boldsymbol{\Sigma}),$$

- The total covariance of all observations are described by

$$\mathbf{V} = \mathbf{Z}\boldsymbol{\Psi}\mathbf{Z}^T + \boldsymbol{\Sigma}$$

- The $\mathbf{Z}\boldsymbol{\Psi}\mathbf{Z}^T$ part is specified through the random effects of the model
- The $\boldsymbol{\Sigma}$ part has so far been $\sigma^2\mathbf{I}$, but now we will put some structure into $\boldsymbol{\Sigma}$
- For instance the structure known from the simple mixed model

$$\text{cov}(y_{i_1}, y_{i_2}) = \begin{cases} 0 & , \text{ if } \texttt{individual}_{i_1} \neq \texttt{individual}_{i_2} \\ \sigma^2_{\texttt{individual}} & , \text{ if } \texttt{individual}_{i_1} = \texttt{individual}_{i_2} \text{ and } i_1 \neq i_2 \\ \sigma^2_{\texttt{individual}} + \sigma^2, & \text{ if } i_1 = i_2 \end{cases}$$

- This structure is known as *compound symmetry*

Activity of rats analyzed via compound symmetry model

- The model is the same as the random effects model, but specified directly

$$
\begin{aligned}
\textbf{lnc} &\sim N(\boldsymbol{\mu}, \mathbf{V}), \quad \text{where} \\
\mu_i &= \mu + \alpha(\texttt{treatm}_i) + \beta(\texttt{month}_i) + \gamma(\texttt{treatm}_i, \texttt{month}_i), \text{ and} \\
V_{i_1,i_2} &= \begin{cases} 0 & \text{, if } \mathsf{cage}_{i_1} \neq \mathsf{cage}_{i_2} \\ \sigma_d^2 & \text{, if } \mathsf{cage}_{i_1} = \mathsf{cage}_{i_2} \text{ and } i_1 \neq i_2 \\ \sigma_d^2 + \sigma^2 & \text{, if } i_1 = i_2 \end{cases}
\end{aligned}
$$

- Implemented in R by:

```
> fit.cs<-gls(lnc~month+treatm+month:treatm,
+             correlation=corCompSymm(form=~1|cage),
+             data=rats)
```

- A `random=...` statement adds random effects, but a `correlation=...` statement writes a structure directly into the $\mathbf{\Sigma}$–matrix

Comparing

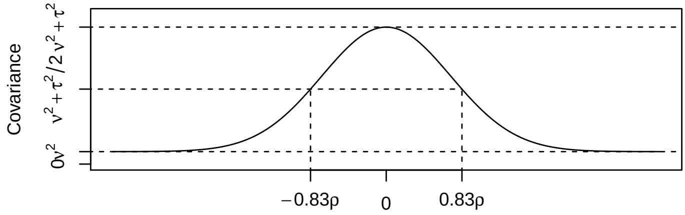
- Notice I had to use `gls()` instead of `lme()`, but only because `lme()` does not allow models with no random effects.
- But `lme()` also has a `correlation=...` argument
- Is it the same model?

```
> fit.cs<-gls(lnc~month+treatm+month:treatm,
+             correlation=corCompSymm(form=~1|cage),
+             data=rats, method="ML")
> logLik(fit.cs)
'log Lik.' 49.39459 (df=32)
> fit.mm<-lme(lnc~month+treatm+month:treatm,
+             random = ~1|cage,
+             data=rats, method="ML")
> logLik(fit.mm)
'log Lik.' 49.39459 (df=32)
```

Gaussian model of spatial correlation

- Covariance structures depending on "how far" observations are apart are known as *spatial*
- The following covariance structure has been proposed for repeated measurements

$$V_{i_1,i_2} = \begin{cases} 0 & \text{, if } \texttt{individual}_{i_1} \neq \texttt{individual}_{i_2} \\ \nu^2 + \tau^2 \exp\left\{ \frac{-(t_{i_1} - t_{i_2})^2}{\rho^2} \right\} & \text{, if } \texttt{individual}_{i_1} = \texttt{individual}_{i_2} \text{ and } i_1 \neq i_2 \\ \nu^2 + \tau^2 + \sigma^2 & \text{, if } i_1 = i_2 \end{cases}$$



Rats data via spatial Gaussian correlation model

- The entire model is:

$$
\begin{aligned}
\textbf{Inc} &\sim N(\boldsymbol{\mu}, \mathbf{V}), \text{ where} \\
\mu_i &= \mu + \alpha(\texttt{treatm}_i) + \beta(\texttt{month}_i) + \gamma(\texttt{treatm}_i, \texttt{month}_i), \text{ and} \\
V_{i_1,i_2} &= \begin{cases} 0 & , \text{ if cage}_{i_1} \neq \text{cage}_{i_2} \\ \nu^2 + \tau^2 \exp\left\{ \frac{-(\texttt{month}_{i_1} - \texttt{month}_{i_2})^2}{\rho^2} \right\} & , \text{ if cage}_{i_1} = \text{cage}_{i_2} \\ & \quad \text{and } i_1 \neq i_2 \\ \nu^2 + \tau^2 + \sigma^2 & , \text{ if } i_1 = i_2 \end{cases}
\end{aligned}
$$

- This model is implemented by:

```
> fit.gau <- lme(lnc~month+treatm+month:treatm,
+                random=~1|cage,
+                correlation=corGaus(form=~as.numeric(month)|cage,nugget=TRUE),
+                data=rats)
```

Important output

```
> fit.gau
Linear mixed-effects model fit by REML
  Data: rats
  Log-restricted-likelihood: 52.6567
  Fixed: lnc ~ month + treatm + month:treatm
    (Intercept)           month2           month3           month4           month5
       9.874280        -0.282870        -0.199010        -0.381285        -0.464289
         month6           month7           month8           month9          month10
      -0.574140        -0.711864        -0.638082        -0.724398        -0.901828
        treatm2          treatm3   month2:treatm2   month3:treatm2   month4:treatm2
      -0.168020        -0.137266         0.080240         0.243762         0.160360
 month5:treatm2   month6:treatm2   month7:treatm2   month8:treatm2   month9:treatm2
       0.217248         0.153330         0.265727         0.202875         0.246120
month10:treatm2   month2:treatm3   month3:treatm3   month4:treatm3   month5:treatm3
       0.049884         0.055665         0.285257         0.332030         0.412954
 month6:treatm3   month7:treatm3   month8:treatm3   month9:treatm3 month10:treatm3
       0.410386         0.472180         0.465124         0.443817         0.202159

Random effects:
 Formula: ~1 | cage
        (Intercept)  Residual
StdDev:   0.1404056 0.2171559

Correlation Structure: Gaussian spatial correlation
 Formula: ~as.numeric(month) | cage
 Parameter estimate(s):
    range    nugget
2.3863954 0.2186743
Number of Observations: 300
Number of Groups: 30
```

Parametrization

- The model outputs are not exactly how we set up the model:

$$\begin{aligned}
\texttt{(Intercept)} &= \nu \\
\texttt{(Residual)} &= \sqrt{\tau^2 + \sigma^2} \\
\texttt{(range)} &= \rho^2 \\
\texttt{(nugget)} &= \sigma^2/(\tau^2 + \sigma^2)
\end{aligned}$$

- So we can get our estimates by:

```
> nu.sq<-0.1404056^2
> sigma.sq<-0.2171559^2*0.2186743
> tau.sq<-0.2171559^2-sigma.sq
> rho.sq<-2.3863954
> c(nu.sq=nu.sq, sigma.sq=sigma.sq, tau.sq=tau.sq, rho.sq=rho.sq)

     nu.sq   sigma.sq     tau.sq     rho.sq
0.01971373 0.01031196 0.03684473 2.38639540
```

Comparing variance structures

- Comparing the three different variance structures
  - independent
  - simple correlation within cage
  - spatial Gaussian correlation structure

```
> fit.id <- lm(lnc~month+treatm+month:treatm, data=rats) # WRONG independent mo
> fit.mm <- lme(lnc~month+treatm+month:treatm,
+               random = ~1|cage,
+               data=rats, method="ML")
> fit.gau<- lme(lnc~month+treatm+month:treatm,
+               random=~1|cage,
+               correlation=corGaus(form=~as.numeric(month)|cage,nugget=TRUE),
+               data=rats, method="ML")
> anova(fit.gau,fit.mm,fit.id)
```

```
        Model df        AIC       BIC    logLik   Test  L.Ratio p-value
fit.gau     1 34 -157.37587 -31.44726 112.68793
fit.mm      2 32  -34.78917  83.73187  49.39459 1 vs 2 126.5867  <.0001
fit.id      3 31   63.47860 178.29585  -0.73930 2 vs 3 100.2678  <.0001
```

- Which shows that spatial Gaussian correlation structure is preferable.

Other spatial correlation structures

- R has a lot of build–in correlation structures. A few examples are:

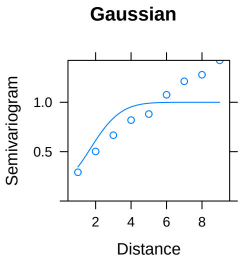
Write in R	Name	Correlation term
`corGaus`	Gaussian	$\tau^2 \exp\{\frac{-(t_{i_1}-t_{i_2})^2}{\rho^2}\}$
`corExp`	exponential	$\tau^2 \exp\{\frac{-\lvert t_{i_1}-t_{i_2}\rvert}{\rho}\}$
`corAR1`	autoregressive(1)	$\rho^{\lvert i_1-i_2\rvert}$
`corSymm`	unstructured	$\tau^2_{i_1,i_2}$

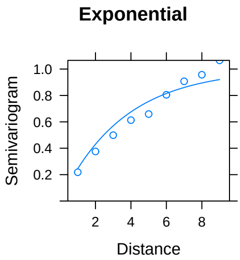
- Unfortunately is can be very difficult to choose — especially for "short" individual series
- General advice:
  - Keep it simple: Numerical problems often occur with (too) complicated structures
  - Graphical methods: Especially for "long" series the variogram is useful
  - Information criteria: AIC or BIC can be used as guideline
  - Try to cross–validate your main conclusion(s) by one of the "simple" methods

The semi-variogram

- A variogram compares the model predicted correlation (or rather one minus) to empirical estimates of the correlation at different distances.
- The empirical estimates will be uncertain at large distances

```
> fit.gau<- lme(lnc~month+treatm+month:treatm, random=~1|cage,
+               correlation=corGaus(form=~as.numeric(month)|cage,nugget=TRUE),
+               data=rats)
> fit.exp<- lme(lnc~month+treatm+month:treatm, random=~1|cage,
+               correlation=corExp(form=~as.numeric(month)|cage,nugget=TRUE),
+               data=rats)
> plot(Variogram(fit.gau), main='Gaussian')
> plot(Variogram(fit.exp), main='Exponential')
```





Comparing by AIC

- Remember to run with `method="ML"`

```
> fit.gau<- lme(lnc~month+treatm+month:treatm, random=~1|cage,
+               correlation=corGaus(form=~as.numeric(month)|cage,nugget=TRUE),
+               data=rats, method="ML")
> fit.exp<- lme(lnc~month+treatm+month:treatm, random=~1|cage,
+               correlation=corExp(form=~as.numeric(month)|cage,nugget=TRUE),
+               data=rats, method="ML")
> anova(fit.gau,fit.exp)

        Model df       AIC       BIC   logLik
fit.gau     1 34 -157.3759 -31.44726 112.6879
fit.exp     2 34 -163.3743 -37.44572 115.6872
```

- So also in favor of exponential structure.

Reducing mean value structure

- Remember to run with `method="ML"`

```
> fit.exp.1<-lme(lnc~month+treatm+month:treatm, random=~1|cage,
+                correlation=corExp(form=~as.numeric(month)|cage,nugget=TRUE),
+                data=rats, method="ML")
> fit.exp.2<-lme(lnc~month+treatm, random=~1|cage,
+                correlation=corExp(form=~as.numeric(month)|cage,nugget=TRUE),
+                data=rats, method="ML")
> anova(fit.exp.2,fit.exp.1)
```

```
          Model df       AIC        BIC    logLik   Test  L.Ratio p-value
fit.exp.2     1 16 -167.1792 -107.91867  99.58959
fit.exp.1     2 34 -163.3743  -37.44572 115.68716 1 vs 2 32.19514  0.0208
```

- So interaction term is significant.

Diagram of analysis

- Select covariance structure from
  - knowledge about the experiment
  - guided by information criteria
  - guided by variogram
- Covariance parameters are tested by likelihood ratio test
- The green arrow is often omitted by the argument that a non–significant simplification of the mean structure should not change the covariance structure much