# Advanced Methods in Biostatistics II

## Lecture 13

December 7, 2017

# General linear mixed model

- The general linear mixed model can be written as follows:

$$\mathbf{y}_i = \mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{u}_i + \varepsilon_i, \quad i = 1, \ldots m.$$

- Here:
  - $\mathbf{y}_i$ is an $n_i \times 1$ vector of observations.
  - $\mathbf{X}_i$ is an $n_i \times p$ design matrix for the fixed effects.
  - $\beta$ is a $p \times 1$ vector of fixed effects.
  - $\mathbf{Z}_i$ is an $n_i \times k$ design matrix for the random effects.
  - $\mathbf{u}_i$ is a $k \times 1$ vector of random effects.
  - $\varepsilon_i$ is an $n_i \times 1$ vector of error terms.

# Example - One-way random effects

- Consider the following one-way random effects model:

$$y_{ij} = \mu + u_i + \epsilon_{ij}$$

  with $i = 1, \ldots m$, $j = 1, \ldots n$, $u_i \sim N(0, \sigma_1^2)$, $\epsilon_{ij} \sim N(0, \sigma^2)$, and $\mathrm{cov}(u_j, \epsilon_{ij}) = 0$.

- Here

$$\mathbf{y}_i = \begin{pmatrix} \mathbf{y}_{i1} \\ \mathbf{y}_{i2} \\ \vdots \\ \mathbf{y}_{in} \end{pmatrix} \qquad \mathbf{X}_i = \mathbf{X} = \mathbf{J}_n \qquad \mathbf{Z}_i = \mathbf{Z} = \mathbf{J}_n$$

# Example - Random intercept and slope

- Consider the following model:

$$y_{ij} = \beta_0 + u_{1i} + \beta_1 x_j + u_{2i} x_j + \epsilon_{ij}$$

with $i = 1, \ldots m$, $j = 1, \ldots n$, $u_{1i} \sim N(0, \sigma_1^2)$, $u_{2i} \sim N(0, \sigma_2^2)$, $\epsilon_{ij} \sim N(0, \sigma^2)$, and all random effects are independent.

- Here

$$\mathbf{y}_i = \begin{pmatrix} \mathbf{y}_{i1} \\ \mathbf{y}_{i2} \\ \vdots \\ \mathbf{y}_{in} \end{pmatrix} \qquad \mathbf{X}_i = \mathbf{X} = [\mathbf{J}_n \;\; \mathbf{x}] \qquad \mathbf{Z}_i = \mathbf{Z} = [\mathbf{J}_n \;\; \mathbf{x}]$$

# Linear mixed models

- Now consider expressing the linear mixed model as follows:
$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}_{(1)}\mathbf{u}_1 + \cdots + \mathbf{Z}_{(k)}\mathbf{u}_k + \varepsilon$$
- Let $N = \sum_{i=1}^{m} n_i$.
- Here we assume each $\mathbf{Z}_{(h)}$ is an $N \times r_h$ matrix that specifies membership in the various clusters or subgroups, and the $\mathbf{u}_h$ are different random effects.

# Linear mixed models

- Let $E(\mathbf{u}) = E(\varepsilon) = \mathbf{0}$, and let us assume

$$\mathrm{var}(\varepsilon) = \sigma_\epsilon^2 \mathbf{I}_N$$

  and

$$\mathrm{var}(\mathbf{u}_h) = \sigma_h^2 \mathbf{I}_{r_h}.$$

- Here $r_h$ represents the number of elements in $\mathbf{u}_h$.
- In addition, $\mathrm{cov}(\mathbf{u}_i, \mathbf{u}_j) = \mathbf{0}\ \forall i \neq j$, and $\mathrm{cov}(\mathbf{u}, \varepsilon) = \mathbf{0}$.

# Example - One-way random effects

- Consider the following one-way random effects model:

$$y_{ij} = \mu + u_i + \epsilon_{ij}$$

with $i = 1, \ldots m$, $j = 1, \ldots n$, $u_i \sim N(0, \sigma_1^2)$, $\epsilon_{ij} \sim N(0, \sigma^2)$, and $\mathrm{cov}(u_j, \epsilon_{ij}) = 0$.

- Here

$$m = 1, \quad \mathbf{X} = \mathbf{J}_N \quad \mathbf{Z}_{(1)} = \begin{pmatrix} \mathbf{J}_n & \mathbf{0}_n & \ldots & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{J}_n & \ldots & \mathbf{0}_n \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0}_n & \ldots & \mathbf{0}_n & \mathbf{J}_n \end{pmatrix}$$

# Example - Random intercept and slope

- Consider the following model:

$$y_{ij} = \beta_0 + u_{1i} + \beta_1 x_j + u_{2i} x_j \epsilon_{ij}$$

with $i = 1, \ldots m$, $j = 1, \ldots n$, $u_{1i} \sim N(0, \sigma_1^2)$, $u_{2i} \sim N(0, \sigma_2^2)$, $\epsilon_{ij} \sim N(0, \sigma^2)$, and all random effects are independent.

- Here

$$m = 2, \quad \mathbf{X} = \begin{pmatrix} \mathbf{J}_n & \mathbf{x} \\ \mathbf{J}_n & \mathbf{x} \\ \vdots & \vdots \\ \mathbf{J}_n & \mathbf{x} \end{pmatrix}$$

$$\mathbf{Z}_{(1)} = \begin{pmatrix} \mathbf{J}_n & \mathbf{0}_n & \ldots & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{J}_n & \ldots & \mathbf{0}_n \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0}_n & \ldots & \mathbf{0}_n & \mathbf{J}_n \end{pmatrix} \quad \mathbf{Z}_{(2)} = \begin{pmatrix} \mathbf{x} & \mathbf{0}_n & \ldots & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{x} & \ldots & \mathbf{0}_n \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0}_n & \ldots & \mathbf{0}_n & \mathbf{x} \end{pmatrix}$$

# Linear mixed models

- We summarize the model as follows:

$$\mathbf{y} = \mathbf{X}\beta + \sum_{h=1}^{k} \mathbf{Z}_{(h)}\mathbf{u}_h + \varepsilon$$

where

$$\mathbf{V} \equiv \mathrm{var}(\mathbf{y}) = \sum_{h=1}^{k} \sigma_h^2 \mathbf{Z}_{(h)}\mathbf{Z}'_{(h)} + \sigma_\epsilon^2 \mathbf{I}_N.$$

# Linear mixed models

- A useful extension is to make the following definition:

$$\mathbf{u}_0 \equiv \varepsilon \qquad \mathbf{Z}_{(0)} \equiv \mathbf{I}_N \qquad \sigma_0^2 \equiv \sigma_\epsilon^2$$

- We can now write the model:

$$\mathbf{y} = \mathbf{X}\beta + \sum_{h=0}^{k} \mathbf{Z}_{(h)}\mathbf{u}_h$$

with

$$\mathbf{V} = \sum_{h=0}^{k} \sigma_h^2 \mathbf{Z}_{(h)}\mathbf{Z}'_{(h)}.$$

# Estimation

- We seek to estimate the fixed effects and variance components.
- For a given $\mathbf{V}$ we can estimate $\beta$ using:

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

- The most common approaches towards estimating the parameters in the covariance matrices is to use maximum likelihood (ML) or restricted maximum likelihood (REML).

# REML

- Using REML we obtain a set of $k+1$ estimating equations for $\sigma_0^2, \ldots \sigma_k^2$ given by

$$\mathrm{tr}\left( (\mathbf{KVK}')^{-1}\mathbf{KZ}_{(h)}\mathbf{Z}'_{(h)}\mathbf{K}' \right)$$
$$= \mathbf{y}'\mathbf{K}'(\mathbf{KVK}')^{-1}\mathbf{KZ}_{(h)}\mathbf{Z}'_{(h)}\mathbf{K}'(\mathbf{KVK}')^{-1}\mathbf{Ky}$$

  where $\mathbf{K} = \mathbf{C}(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')$ and $\mathbf{C}$ is some full rank $(n-p) \times n$ matrix.

- Note, the expected value of the quadratic form on the right side is given by the left side of the equation.

# REML

- In certain special cases these equations can be simplified to yield closed-form solutions.
- However, in most cases, numerical methods are required to solve the system of equations.

# Inference for $\beta$

- Estimates of the variance components can be inserted into $\mathbf{V}$ to obtain:

$$\hat{\mathbf{V}} = \sum_{h=0}^{k} \hat{\sigma}_h^2 \mathbf{Z}_{(h)} \mathbf{Z}'_{(h)}.$$

- We can in turn use this to estimate:

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

- This is sometimes called the estimated generalized least-squares solution.

# Inference for $\beta$

- An approximate estimate of the variance-covariance matrix for $\hat{\beta}$ is given by:

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

- Note this ignores the variability due to the estimation of **V**.
- For large samples this extra variability is negligible, but it can be substantial for smaller samples.
- If the sample size is small it may be preferable to use a bootstrap procedure to approximate the distribution of the test statistics.

# Likelihood ratio statistic

## Definition

Consider testing:

$$H_0 : \theta \in \Theta_0 \ \ \textit{versus} \ \ H_1 : \theta \notin \Theta_0$$

The likelihood ratio statistic is

$$\lambda = -2 \log \left( \frac{\mathcal{L}(\hat{\theta}_0)}{\mathcal{L}(\hat{\theta})} \right),$$

where $\hat{\theta}$ is the MLE and $\hat{\theta}_0$ is the MLE when $\theta$ is restricted to lie in $\Theta_0$.

# Likelihood ratio test

**Theorem**

Let $\theta = (\theta_1, \ldots \theta_q, \theta_{q+1}, \ldots \theta_r)$ and

$$\Theta_0 = \{\theta : (\theta_{q+1}, \ldots \theta_r) = (0, \ldots 0)\}.$$

Let $\lambda$ be the likelihood ratio test statistic. Under $H_0 : \theta \in \Theta_0$ and very general regularity conditions, $\lambda$ is asymptotically $\chi^2$ with $r - q$ degrees of freedom.

The p-value for the test is $P(\chi^2_{r-q} > \lambda)$.

# Likelihood ratio tests

- Testing in linear mixed models can be performed using likelihood ratio tests.
- We can compare a null restricted model with an alternative unrestricted model using

$$\lambda = -2\log\left(\frac{\mathcal{L}(\hat{\theta}_0|\mathbf{y})}{\mathcal{L}(\hat{\theta}|\mathbf{y})}\right) = -2(\ell(\hat{\theta}_0|\mathbf{y}) - \ell(\hat{\theta}|\mathbf{y})).$$

- This test statistic is asymptotically $\chi^2$ with degrees of freedom equal to the difference in number of parameters between the two models.

# Restricted likelihood ratio tests

- When testing the variance components it is possible to use the restricted log-likelihood, instead of using the likelihood function, to form the test statistic.
- However, this is only possible if the models share the same fixed effects.
- Otherwise the two likelihood functions will not be comparable, as REML estimates the random effects by considering linear combinations of the data that remove the fixed effects.

# Likelihood ratio tests

- Note the asymptotic result for the LRT is based on some regularity assumptions that are not always satisfied in practice.
- In particular, the parameters under the null model cannot lie on the boundary of the parameter space.
- This is a problem for testing variance components when the null hypothesis is that they are equal to zero, as they are constrained to be positive (or positive definite).

# Testing variance components

- Consider the model:

$$y_{ij} = \beta_0 + u_i + \beta_1 x_{ij} + \epsilon_{ij}.$$

  where $u_i \sim_{iid} N(0, \sigma_u^2)$ and $\epsilon_{ij} \sim_{iid} N(0, \sigma_\epsilon^2)$.

- Suppose we want to test whether the intercepts of the groups are significantly different from one another.

- This is equivalent to testing

$$H_0 : \sigma_u^2 = 0 \quad \textit{versus} \quad H_a : \sigma_u^2 > 0.$$

# Testing variance components

- Under certain independence assumptions, the asymptotic distribution when $H_0$ is true implies that there is a 50% chance that $\hat{\sigma}_u^2 = 0$.
- This leads to the following approximate result:

$$\lambda \sim 0.5\chi_0^2 + 0.5\chi_1^2$$

where $\chi_0^2$ is a point mass at 0.

# Testing variance components

- This result assumes that the data are independent and identically distributed both under the null and alternative hypothesis.
- This assumption does not necessarily hold for all mixed models.
- In the case of a linear mixed model with one variance component the finite sample and asymptotic distribution of the LRTs show that, under general conditions, the null distribution for testing $H_0$ is typically different from $0.5\chi^2_0 : 0.5\chi^2_1$.

# R

- Mixed models can be fit in `R` using the `lmer` function.
- It is part of the `lme4` package.

# Example

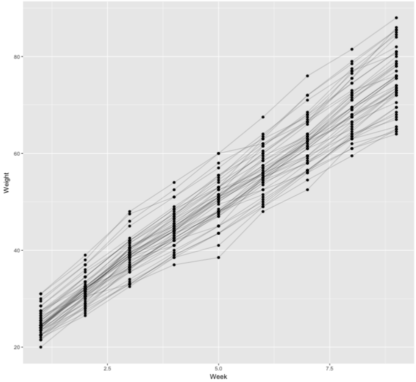
- Let us revisit the pig example.
- It consists of 9 repeated weight measures on 48 pigs.
- We use the following model:

$$y_{ij} = \beta_0 + u_i + \beta_1 x_{ij} + \epsilon_{ij}.$$

for $i = 1, \ldots 48$ and $j = 1, \ldots 9$ where $u_i \sim_{iid} N(0, \sigma_u^2)$ and $\epsilon_{ij} \sim_{iid} N(0, \sigma_\epsilon^2)$.

# Coding example

```
library(SemiPar)
library(lme4)
library(ggplot2)
data(pig.weights)

ggplot(pig.weights, aes(x = num.weeks, y = weight, group = id.num)) +
  geom_point() + geom_path(alpha = .2) +
  labs(x = "Week", y = "Weight")
```

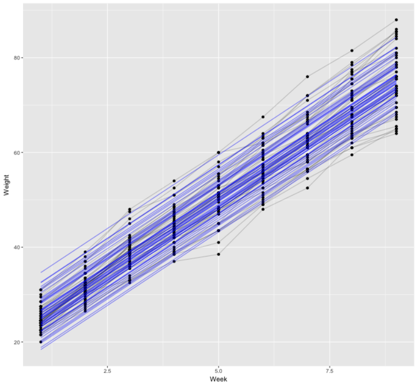


# Coding example

```
pig.mixed = lmer(weight ~ (1 | id.num) + num.weeks, data = pig.weights)

> summary(pig.mixed)
Linear mixed model fit by REML ['lmerMod']
Formula: weight ~ (1 | id.num) + num.weeks
   Data: pig.weights

REML criterion at convergence: 2033.8

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.7390 -0.5456  0.0184  0.5122  3.9313

Random effects:
 Groups   Name        Variance Std.Dev.
 id.num   (Intercept) 15.142   3.891
 Residual              4.395   2.096
Number of obs: 432, groups:  id.num, 48

Fixed effects:
            Estimate Std. Error t value
(Intercept) 19.35561    0.60314   32.09
num.weeks    6.20990    0.03906  158.97

Correlation of Fixed Effects:
          (Intr)
num.weeks -0.324
```

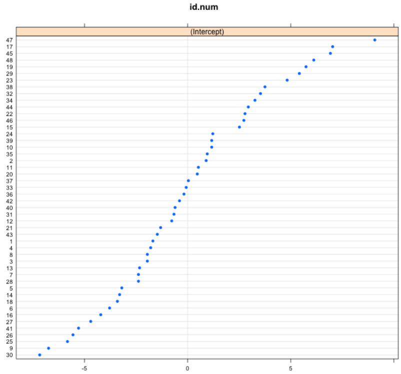
# Coding example

```
pig.weights$pig.mixed.fit = fitted(pig.mixed)
ggplot(pig.weights, aes(x = num.weeks, y = weight, group = id.num)) +
  geom_point() + geom_path(alpha = .2) +
  labs(x = "Week", y = "Weight") +
  geom_line(aes(y = pig.mixed.fit), color = "blue", alpha = .5)
```



# Coding example

```
> u = ranef(pig.mixed)
> dotplot(u)
```



# Coding example

```
> fixef(pig.mixed)
(Intercept)   num.weeks
  19.355613    6.209896

> VarCorr(pig.mixed)
 Groups   Name        Std.Dev.
 id.num   (Intercept) 3.8913
 Residual             2.0964

> 3.8913^2/(3.8913^2 + 2.0964^2)
[1] 0.775049
```

# R

From Bates et al. (2015).

| Formula | Alternative | Meaning |
|---|---|---|
| `(1 | g)` | `1 + (1 | g)` | Random intercept with fixed mean. |
| `0 + offset(o) + (1 | g)` | `-1 + offset(o) + (1 | g)` | Random intercept with *a priori* means. |
| `(1 | g1/g2)` | `(1 | g1) + (1 | g1:g2)` | Intercept varying among `g1` and `g2` within `g1`. |
| `(1 | g1) + (1 | g2)` | `1 + (1 | g1) + (1 | g2)` | Intercept varying among `g1` and `g2`. |
| `x + (x | g)` | `1 + x + (1 + x | g)` | Correlated random intercept and slope. |
| `x + (x || g)` | `1 + x + (1 | g) + (0 + x | g)` | Uncorrelated random intercept and slope. |

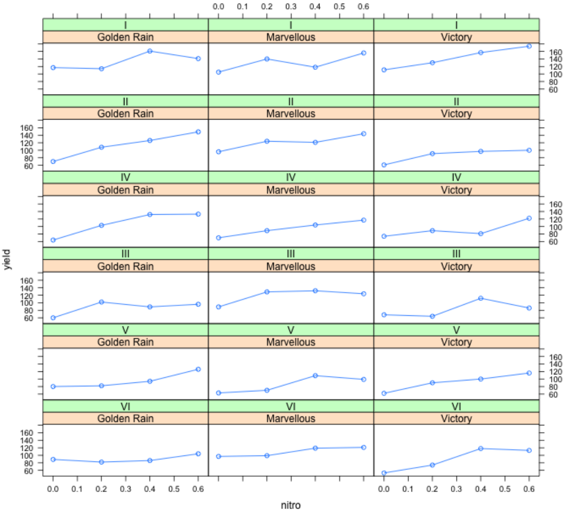
Table 2: Examples of the right-hand-sides of mixed-effects model formulas. The names of grouping factors are denoted `g`, `g1`, and `g2`, and covariates and *a priori* known offsets as `x` and `o`.

# R

- The Oats data from the `nlme` package comes from an experiment in which fields in 6 different locations were each divided into three plots and each of these 18 plots was further subdivided into four subplots.
- Three varieties of oats were randomly assigned to the three plots in each block and four concentrations of fertilizer (measured as nitrogen concentration) were randomly assigned to the subplots in each plot.

# Coding example

```
> xyplot(yield ~ nitro | Variety + Block, data = Oats, type = "o")
```



# Coding example

```
> fit <- lmer(yield ~ nitro + Variety + (1 | Block/Variety), dat=Oats)
> fit
Linear mixed model fit by REML ['lmerMod']
Formula: yield ~ nitro + Variety + (1 | Block/Variety)
   Data: Oats
REML criterion at convergence: 578.8918
Random effects:
 Groups        Name        Std.Dev.
 Variety:Block (Intercept) 10.44
 Block         (Intercept) 14.65
 Residual                  12.87
Number of obs: 72, groups:  Variety:Block, 18; Block, 6
Fixed Effects:
      (Intercept)              nitro  VarietyMarvellous     VarietyVictory
           82.400             73.667              5.292             -6.875
```