Linear Mixed Models Summary Sheet

Compiled by: Shravan Vasishth
(vasishth@uni-potsdam.de)

Version dated: May 15, 2013

Background and review of linear model theory

In linear modelling we model the mean of a response $Y_1, \dots, Y_n$ as a function of a vector of predictors $x_1, \dots, x_n$. We assume that the Y_i are conditionally independent given $\mathbf{x}, \beta$. When Y's are not marginally independent, we have $Cor(Y_1, Y_2) \neq 0$, or $P(Y_2 \mid Y_1) \neq P(Y_2)$.
Linear mixed models are useful for correlated data where $\mathbf{Y} \mid \mathbf{X}, \beta$ are not independently distributed.

Basic specification of LMMs

$$\underset{\substack{\uparrow\\ n\times 1}}{Y_i} = \underset{\substack{\uparrow\\ n\times p}}{X_i} \underset{\substack{\uparrow\\ p\times 1}}{\beta} + \underset{\substack{\uparrow\\ n\times q}}{Z_i} \underset{\substack{\uparrow\\ q\times 1}}{b_i} + \underset{\substack{\uparrow\\ n\times 1}}{\epsilon_i} \tag{1}$$

where $i = 1, \dots, m$, let n be total number of data points.
Distributional assumptions:
$b_i \sim N(0, D)$ and $\epsilon_i \sim N(0, \sigma^2 I)$. D is a $q \times q$ matrix that does not depend on i, and b_i and ϵ_i are assumed to be independent.
Y_i has a multivariate normal distribution:

$$Y_i \sim N(X_i\beta, V(\alpha)) \tag{2}$$

where $V(\alpha) = Z_i D Z_i^T + \sigma^2 I$, and α is the variance component parameters.
Note:

1. D has to be symmetric and positive definite.
2. The Z_i matrix columns are a subset of X_i. In the random intercept model, $Z_i = 1_i$.
3. In the varying intercepts and varying slopes model, $X_i = Z_i = (1_i, X_i)$. Then:

$$Y_i = X_i(\beta + b_i) + \epsilon_i \tag{3}$$

or

$$Y_i = X_i\beta_i + \epsilon_i \quad \beta_i \sim N(\beta, D) \tag{4}$$

$$\begin{aligned} D &= \begin{pmatrix} d_{00} & d_{01} \\ d_{10} & d_{11} \end{pmatrix} \\ &= \begin{pmatrix} d_{00} = Var(\beta_{i0}) & d_{01} = Cov(\beta_{i0}, \beta_{i1}) \\ d_{10} = Cov(\beta_{i0}, \beta_{i1}) & d_{11} = Var(\beta_{i1}) \end{pmatrix} \end{aligned} \tag{5}$$

4. The conditional mean of Y given the block effect b_i is:

$$E(Y \mid b_i) = X_i\beta + Z_i b_i \tag{6}$$

5. The marginal mean for Y:

$$\begin{aligned} E(Y_i) =& E(E(Y_i \mid b_i)) \\ =& E(X_i\beta + Z_i b_i) \\ =& X_i\beta \end{aligned} \tag{7}$$

6. The conditional variance of Y given the block effect b_i is:

$$Cov(Y_i \mid b_i) = Cov(\epsilon_i) = \sigma^2 I \tag{8}$$

7. The marginal variance of Y averaged over the distributions of b_i is

$$\begin{aligned} Cov(Y_i) =& Cov(Z_i b_i) + Cov(\epsilon_i) \\ =& Z_i Cov(b_i) Z_i^T + Cov(\epsilon_i) \\ =& Z_i D Z_i^T + \sigma^2 I \end{aligned} \tag{9}$$

σ_b^2 describes both between-block variance, and within block covariance

Consider the following model, a varying intercepts model:

$$Y_{ij} = \mu + b_i + e_{ij}, \tag{10}$$

with $b_i \sim N(0, \sigma_b^2)$, $e_{ij}\ N(0, \sigma^2)$.
Note that variance is a covariance of a random variable with itself, and then consider the model formulation. If we have

$$Y_{ij} = \mu + b_i + \epsilon_{ij} \tag{11}$$

where i is the group, j is the replication, if we *define* $b_i \sim N(0, \sigma_b^2)$, and refer to σ_b^2 as the between group variance, then we must have

$$\begin{aligned} Cov(Y_{i1}, Y_{i2}) =& Cov(\mu + b_i + \epsilon_{i1}, \mu + b_i + \epsilon_{i2}) \\ =& \underset{\substack{\uparrow\\ =0}}{Cov(\mu, \mu)} + \underset{\substack{\uparrow\\ =0}}{Cov(\mu, b_i)} + \underset{\substack{\uparrow\\ =0}}{Cov(\mu, \epsilon_{i2})} + \\ & \underset{\substack{\uparrow\\ =0}}{Cov(b_i, \mu)} + \underset{\substack{\uparrow\\ +ve}}{Cov(b_i, b_i)} \dots \\ =& Cov(b_i, b_i) = Var(b_i) = \sigma_b^2 \end{aligned} \tag{12}$$

Some basic types of linear mixed model and their variance components

Varying intercepts model

The model for a categorical predictor is:

$$Y_{ijk} = \beta_j + b_i + \epsilon_{ijk} \tag{13}$$

$i = 1, \dots, 10$ is subject id, $j = 1, 2$ is the factor level, k is the number of replicates (here 1). $b_i \sim N(0, \sigma_b^2), \epsilon_{ijk} \sim N(0, \sigma^2)$.

For a continuous predictor:

$$Y_{ijk} = \beta_0 + \beta_1 t_{ijk} + b_{ij} + \epsilon_{ijk} \tag{14}$$

The general form for any model in this case is:

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, V \right) \tag{15}$$

where

$$V = \begin{pmatrix} \sigma_b^2 + \sigma^2 & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 + \sigma^2 \end{pmatrix} = \begin{pmatrix} \sigma_b^2 + \sigma^2 & \rho\sigma_b\sigma_b \\ \rho\sigma_b\sigma_b & \sigma_b^2 + \sigma^2 \end{pmatrix} \tag{16}$$

Note also that the mean response for the subject, i.e., *conditional* mean of Y_{ij} given the subject-specific effect b_i is:

$$E(Y_{ij} \mid b_i) = X_{ij}^T \beta + b_i \tag{17}$$

The mean response in the population, i.e., the marginal mean of Y_{ij}:

$$E(Y_{ij}) = X_{ij}^T \beta \tag{18}$$

The marginal variance of each response is:

$$\begin{aligned} Var(Y_{ij}) =& Var(X_{ij}^T \beta + b_i + \epsilon_{ij}) \\ =& Var(\beta + b_i + \epsilon_{ij}) \\ =& \sigma_b^2 + \sigma^2 \end{aligned} \tag{19}$$

the covariance between any pair of responses Y_{ij} and $Y_{ij'}$ is given by

$$\begin{aligned} Cov(Y_{ij}, Y_{ij'}) =& Cov(X_{ij}^T \beta + b_i + \epsilon_{ij}, X_{ij'}^T \beta + b_i + \epsilon_{ij'}) \\ =& Cov(b_i + e_{ij}, b_i + e_{ij'}) \\ =& Cov(b_i, b_i) = \sigma_b^2 \end{aligned} \tag{20}$$

The correlation is

$$Corr(Y_{ij}, Y_{ij'}) = \frac{\sigma_b^2}{\sigma_b^2 + \sigma^2} \tag{21}$$

In other words, introducing a random intercept induces a correlation among repeated measurements.
$\hat{V}$ is therefore:

$$\begin{pmatrix} \hat{\sigma}_b^2 + \hat{\sigma}^2 & \hat{\rho}\hat{\sigma}_b\hat{\sigma}_b \\ \hat{\rho}\hat{\sigma}_b\hat{\sigma}_b & \hat{\sigma}_b^2 + \hat{\sigma}^2 \end{pmatrix} \tag{22}$$

Note: $\hat{\rho} = 1$. But this correlation is not *estimated* in the varying intercepts model.

Varying intercepts and slopes (with correlation)

The model for a categorical predictor is:

$$Y_{ij} = \beta_1 + b_{1i} + (\beta_2 + b_{2i}) x_{ij} + \epsilon_{ij} \quad i = 1, ..., M, j = 1, ..., n_i \tag{23}$$

with $b_{1i} \sim N(0, \sigma_1^2), b_{2i} \sim N(0, \sigma_2^2)$, and $\epsilon_{ij} \sim N(0, \sigma^2)$.
Note: I have seen this presentation elsewhere:

$$Y_{ijk} = \beta_j + b_{ij} + \epsilon_{ijk} \tag{24}$$

$b_{ij} \sim N(0, \sigma_b)$. The variance σ_b must be a 2×2 matrix:

$$\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \tag{25}$$

The general form for the model is:

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, V \right) \tag{26}$$

where

$$V = \begin{pmatrix} \sigma_{b,A}^2 + \sigma^2 & \rho\sigma_{b,A}\sigma_{b,B} \\ \rho\sigma_{b,A}\sigma_{b,B} & \sigma_{b,B}^2 + \sigma^2 \end{pmatrix} \tag{27}$$

No varying intercepts, only slopes for each level

The model is

$$Y_{ijk} = \beta_j + b_{ij} + \epsilon_{ijk} \tag{28}$$

The random effects are:

$b_{ij} = \begin{pmatrix} b_{i1} \\ b_{i12} \end{pmatrix} \sim N(0, \sigma_b^2)$, where $\sigma_b^2 = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$.
Here, V is

$$V = \begin{pmatrix} \sigma_{b,A}^2 + \sigma^2 & \rho\sigma_{b,A}\sigma_{b,B} \\ \rho\sigma_{b,A}\sigma_{b,B} & \sigma_{b,B}^2 + \sigma^2 \end{pmatrix} \tag{29}$$

Note that here, a random effect is computed for each material separately.

One insight is that V can be derived from the random effects variance components, and the error term's variance component:

$$V = \begin{pmatrix} \sigma_{b,A}^2 & \rho\sigma_{b,A}\sigma_{b,B} \\ \rho\sigma_{b,A}\sigma_{b,B} & \sigma_{b,B}^2 \end{pmatrix} + \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} \tag{30}$$

Nested models (e.g., Worker/Machine)

The model is:

$$Y_{ijk} = \beta_j + b_i + b_{ij} + \epsilon_{ijk} \tag{31}$$

Here, we force all random effects to be independent. Observations between workers are independent, but observations on the same worker are correlated.
$b_i \sim N(0, \sigma_1^2), b_{ij} \sim N(0, \sigma_2^2)$, and $\epsilon \sim N(0, \sigma^2)$. i is Worker, j is machine, and k is replicate.

```
> fm1<-lmer(score~Machine-1+(1|Worker/Machine),
+ data=Machines)
```

The term Worker/Machine is estimating machine variance within worker:

```
xyplot(score~Machine|Worker,Machines)
```

The variance components in fm1:

Comp.	Groups	Name	Var
$\hat{\sigma}_2^2$	Machine:Worker	(Int)	13.909
$\hat{\sigma}_1^2$	Worker	(Int)	22.858
$\hat{\sigma}^2$	Res		0.925

Number of obs: 54, groups: Machine:Worker, 18; Worker, 6.

For observations on Worker i,

$$Var(Y_{ijk}) = \sigma_1^2 + \sigma_2^2 + \sigma^2 \tag{32}$$

Variance between machines within workers:

$$Covar(Y_{ijk}, Y_{ijk'}) = \sigma_1^2 + \sigma_2^2 \tag{33}$$

Variance between workers:

$$Covar(Y_{ijk}, Y_{ij'k'}) = \sigma_1^2 \tag{34}$$

Note:
1. $\hat{\sigma}_1^2$ all observations have the same variance;
2. $\hat{\sigma}_2^2$: the covariance between observations corresponding to the same worker using different machines is the same, for any pair of machines.

```
> ranef(fm1)
$`Machine:Worker`     $Worker
    (Intercept)
A:6       1.91609     6 -7.514666
A:2       1.55253     2 -1.375925
```

In this model, the sum of the random effects for Worker 1 on Machine A is
$s_1 = b_1 + b_{11}$

```
> ranef(fm1)
...
$`Machine:Worker` $Worker
(Intercept)
s1 = A:1 -0.75012+1.044598=0.29448
```

and for Worker 1 on machine B,
$s_2 = b_1 + b_{21}$.

```
> ranef(fm1)
...
$`Machine:Worker` $Worker
(Intercept)
s2 = B:1 1.50002 + 1.044598 = 2.5446
```

For all Workers and machines, we can obtain these random effects s from this matrix:

```
> mat<-matrix(
+ unlist(ranef(fm1)$`Machine:Worker`),6,3)
> +
+ matrix(unlist(ranef(fm1)$Worker),6,3)

          [,1]      [,2]      [,3]
[1,] -7.514666 -7.514666 -7.514666
[2,] -1.375925 -1.375925 -1.375925
[3,] -0.059823 -0.059823 -0.059823
[4,]  1.044598  1.044598  1.044598
[5,]  5.361045  5.361045  5.361045
[6,]  2.544771  2.544771  2.544771

> mat

         A        B        C
6  1.91609 -8.97590  2.48677
2  1.55253  0.60682 -2.99667
4 -1.03937  2.41736 -1.41440
1 -0.75012  1.50002 -0.11421
3  1.77775  2.29952 -0.81481
5 -3.45687  2.15218  2.85331
```

Using lmer, we have b_i and b_{ij} independent, but s_1 and s_2 are correlated via the common term b_1. We can recover the correlations between machine through the vcov matrix of the random effects (BLUPs) (**but note that we never see this in the lmer output—what's the significance of the fact that these are correlated?**):

```
> var(mat)

        A       B       C
A  4.5670 -4.6492 -1.9288
B -4.6492 19.7897 -4.6925
C -1.9288 -4.6925  5.1966
```

Varying slopes, no varying intercept

$$\boxed{Y_{ijk} = \beta_j + b_{ij} + \epsilon_{ijk} \tag{35}}$$

```
> fm3<-lmer(score~Machine-1+
+               (Machine-1|Worker),
+             data=Machines)

> head(ranef(fm3))

$Worker
  MachineA  MachineB MachineC
6 -5.59160 -16.58381  -5.0305
2  0.18387  -0.80332  -4.2823
4 -1.02388   2.32846  -1.4144
1  0.31199   2.55323   0.9304
3  6.96922   7.77935   4.4733
5 -0.84961   4.72610   5.3235
```

The random effects for Worker 1 on Machine A is
$s_1 = b_{11} = 0.31199$
and for Worker 1 on Machine B,
$s_2 = b_{12} = 2.55323$.
The 'Machine independent' Worker random effect (varying intercept) b_i has been dropped. We have b_{11} correlated with b_{12}. We can see this when we recover the (co-)variances between machines from the random effects:

```
> var(ranef(fm3)$Worker)

         MachineA MachineB MachineC
MachineA   16.347   28.239   11.146
MachineB   28.239   74.093   29.181
MachineC   11.146   29.181   18.972
```

Also, the variances for each machine (16, 74, 18) are also allowed to be different. Here are the variance components:

Comp.	Groups	Name	Variance	Corr$_{1,\cdot}$	Corr$_{2,\cdot}$
$\hat{\sigma}^2_{j=1}$	Worker	A	16.640		
$\hat{\sigma}^2_{j=2}$		B	74.395	$\hat{\rho}_{1,2}$ 0.803	
$\hat{\sigma}^2_{j=3}$		C	19.268	$\hat{\rho}_{1,3}$ 0.623	$\hat{\rho}_{2,3}$ 0.771
$\hat{\sigma}^2$	Res		0.925		

$$Var(Y_{ijk}) = \sigma_j^2 + \sigma^2 \tag{36}$$

$$Covar(Y_{ijk}, Y_{ijk'}) = \sigma_j^2 \tag{37}$$

$$Covar(Y_{ijk}, Y_{ij'k'}) = \rho_{j,j'}\sigma_j\sigma_{j'} \tag{38}$$

Note that the BLUPs' vcov matrix reflects the estimated values:

```
> diag(var(ranef(fm3)$Worker))

MachineA MachineB MachineC
  16.347   74.093   18.972
```

```
> cor(ranef(fm3)$Worker)

         MachineA MachineB MachineC
MachineA  1.00000  0.81141  0.63292
MachineB  0.81141  1.00000  0.77832
MachineC  0.63292  0.77832  1.00000
```

```
> # look at the fm3 output
> ## (the random effects table)
```

1. $\hat{\sigma}_j^2$ the variance of an observation depends on the machine being used;

2. $\rho_{j,j'}\sigma_j\sigma_{j'}$ the covariance between observations corresponding to the same worker using different machines is different, for different pairs of machines.

```
> var(ranef(fm3)$Worker)
         MachineA MachineB MachineC
MachineA   16.347   28.239   11.146
MachineB   28.239   74.093   29.181
MachineC   11.146   29.181   18.972
```

$$\begin{pmatrix} \sigma_A^2 & Cov_{A,B} & Cov_{A,C} \\ & \sigma_B^2 & Cov_{B,C} \\ & & \sigma_C^2 \end{pmatrix} \tag{39}$$

Note that, for given machines j and j', say A, B:
$Covar(Y_{ijk}, Y_{ij'k'}) = Cov_{A,B} = 28.239 \approx \rho_{A,B}\sigma_A\sigma_B = .803 \times \sqrt{16.347} \times \sqrt{74.093} = 27.946$.

Comparing fm1 and fm3

The sum of fm1's (Worker/Machine) ranefs ($b_{ij} + b_i$) are roughly the same as fm3's (Machine-1| Worker) random effects b_{ij} for each machine. **In other words, the random effect b_i is folded into b_{ij} in fm3.**

```
> fm1<-lmer(score~Machine-1+
+              (1|Worker/Machine),
+             data=Machines)
> fm3<-lmer(score~Machine-1+
+              (Machine-1|Worker),
+             data=Machines)
```

fm1's ranefs summed up:

```
> #fm1's ranefs summed up are
> ## roughly the same as the fm3 ranefs:
> matrix(unlist(ranef(fm1)$`Machine:Worker`),
+         6,3) +
+ matrix(unlist(ranef(fm1)$Worker),6,3)

         [,1]      [,2]     [,3]
[1,] -5.59858 -16.49057 -5.02789
[2,]  0.17661  -0.76911 -4.37259
[3,] -1.09920   2.35754 -1.47422
[4,]  0.29448   2.54462  0.93039
[5,]  7.13879   7.66056  4.54624
[6,] -0.91210   4.69695  5.39808
```

The fm3 ranefs:

```
> ranef(fm3)
```

```
$Worker
  MachineA  MachineB MachineC
6 -5.59160 -16.58381  -5.0305
2  0.18387  -0.80332  -4.2823
4 -1.02388   2.32846  -1.4144
1  0.31199   2.55323   0.9304
3  6.96922   7.77935   4.4733
5 -0.84961   4.72610   5.3235
```

Parameter estimation

Likelihood based model fitting procedure

Recall:

1. If we have two continuous random variables Y and Z, with density functions $f_Y(y)$ and $f_Z(z)$ and joint density $f_{Y,Z}(y,z)$, then

$$f_Y(y) = \int f_{Y,Z}(y,z)dz. \tag{40}$$

2. The conditional density of $Y \mid Z$ is defined as

$$f_{Y|Z}(y \mid z) = \frac{f_{Y,Z}(y,z)}{f_Z(z)} \tag{41}$$

so we can write

$$f_{Y,Z}(y,z) = f_{Y|Z}(y \mid z) \times f_Z(z). \tag{42}$$

3. Combining equations 40 and 41, we have

$$f_Y(y) = \int f_{Y|Z}(y \mid z) * f_Z(z)dz \tag{43}$$

Equation 43, where we condition on a second random variable Z (note that Z could be "non-observable") can be helpful in deriving $f_Y(y)$, if the two densities on the RHS are easy to write down, and the integral can be solved.

Returning to parameter estimation in LMMs, the model is:

$$Y_i = X_i\beta + Z_i\beta_i + \epsilon_i, \quad i = 1, \ldots, M \tag{44}$$

where $b_i \sim N(0, \Psi), \epsilon_i \sim N(0, \sigma^2 I)$. Let θ be the parameters that determine Ψ.

$$\begin{aligned} L(\beta, \theta, \sigma^2 \mid y) =& p(y : \beta, \theta, \sigma^2) \\ =& \prod_i^M p(y_i : \beta, \theta, \sigma^2) \\ =& \prod_i^M \int p(y_i \mid b_i, \beta, \sigma^2) p(b_i : \theta, \sigma^2)\, db_i \end{aligned} \tag{45}$$

we want the density of the observations (y_i) given the parameters β, θ and σ^2 only. In this case, using equation 43 above, with $Y = y_i$ and $Z = b_i$ is helpful for deriving the density for y_i, because $f(y_i \mid b_i)$ (or, in the notation of (4.9), $p(y_i \mid b_i, \beta, \sigma^2)$) has a simple form, and so we can get a closed form expression for the integral.

REML estimation (REstricted/REsidual ML)

To estimate variance parameters, **first fit fixed effects using least squares**, and then focus attention on residuals.

The **residuals**' distribution depends on σ^2 and variance parameters θ of random effects.

A **likelihood** for these parameters is formed based on the residuals alone. Maximization of this **marginal likelihood** gives estimates of σ^2 and the other variance-covariance parameters which are less biased than the full maximum likelihood estimates.

Once the REML variance-covariance estimates are obtained the **fixed effects are re-estimated by maximum likelihood assuming the random effects parameters are known**, a procedure equivalent to generalized least squares.

Alternatively, define a restricted likelihood:

$$L_R(\theta, \sigma^2 \mid y) = \int L(\beta, \theta, \sigma^2 \mid y)\, d\beta \tag{46}$$

and maximize this to obtain estimates of these parameters.

Unlike full (max.) likelihood, restricted lik. is not invariant to parameterization, so we cannot compare models with different fixed effects.

How the random effects are 'predicted' when using the ranef() command

In linear mixed models, we fit models like these (the Ware-Laird formulation–see Pinheiro and Bates 2000, for example):

$$Y = X\beta + Zu + \epsilon \tag{47}$$

Let $u \sim N(0, \sigma_u^2)$, and this is independent from $\epsilon \sim N(0, \sigma^2)$.
Given Y, the "minimum mean square error predictor" of u is the conditional expectation:

$$\hat{u} = E(u \mid Y) \tag{48}$$

We can find $E(u \mid Y)$ as follows. We write the joint distribution of Y and u as:

$$\begin{pmatrix} Y \\ u \end{pmatrix} = N\left(\begin{pmatrix} X\beta \\ 0 \end{pmatrix}, \begin{pmatrix} V_Y & C_{Y,u} \\ C_{u,Y} & V_u \end{pmatrix} \right) \tag{49}$$

$V_Y, C_{Y,u}, C_{u,Y}, V_u$ are the various variance-covariance matrices. It is a fact that

$$u \mid Y\ N(C_{u,Y}V_Y^{-1}(Y - X\beta)), Y_u - C_{u,Y}V_Y^{-1}C_{Y,u}) \tag{50}$$

This allows you to derive the BLUPs:

$$\hat{u} = C_{u,Y}V_Y^{-1}(Y - X\beta)) \tag{51}$$

Substituting $\hat{\beta}$ for β, we get:

$$BLUP(u) = \hat{u}(\hat{\beta})C_{u,Y}V_Y^{-1}(Y - X\hat{\beta})) \tag{52}$$

Here's an example with R:

```
> ## Calculate the predicted random
> ## effects by hand for the
> ## ergoStool data
> fm1<-lmer(effort~Type-1 +
+                (1|Subject),ergoStool)
> ## Here are the BLUPs we will
> ## estimate by hand:
> head(ranef(fm1))

$Subject
  (Intercept)
1  1.7088e+00
2  1.7088e+00
3  4.2720e-01
4 -8.5439e-01
5 -1.4952e+00
6 -1.3546e-14
7  4.2720e-01
8 -1.7088e+00
9 -2.1360e-01

> ## this gives us all the
> ## variance components:
> ## this could have been done
> ## ``by hand'',
> ## or at least an approximation:
> VarCorr(fm1)

$Subject
             (Intercept)
(Intercept)       1.7755
attr(,"stddev")
(Intercept)
     1.3325
attr(,"correlation")
             (Intercept)
(Intercept)            1

attr(,"sc")
[1] 1.1003
```

First, calculate the predicted random effect for subject 1:

```
> ## The variance for the random
> ## effect subject is the term C_{u,Y}:
> (covar.u.y<-VarCorr(fm1)$Subject[1])

[1] 1.7755
```

Estimated covariance between u_1 and Y_1 make up a var-covar matrix from this:

```
> (cov.u.Y<-matrix(covar.u.y,1,4))

       [,1]   [,2]   [,3]   [,4]
[1,] 1.7755 1.7755 1.7755 1.7755
```

Estimated variance matrix for Y_1:

```
> (V.Y<-matrix(1.7755,4,4)+
+    diag(1.2106,4,4))

       [,1]   [,2]   [,3]   [,4]
[1,] 2.9861 1.7755 1.7755 1.7755
[2,] 1.7755 2.9861 1.7755 1.7755
[3,] 1.7755 1.7755 2.9861 1.7755
[4,] 1.7755 1.7755 1.7755 2.9861
```

Extract observations for subject 1:

```
> (Y<-matrix(ergoStool$effort[1:4],4,1))
```

```
     [,1]
[1,]   12
[2,]   15
[3,]   12
[4,]   10
```

Estimated fixed effects:

```
> (beta.hat<-matrix(fixef(fm1),4,1))
```

```
        [,1]
[1,]  8.5556
[2,] 12.4444
[3,] 10.7778
[4,]  9.2222
```

Predicted random effect:

```
> cov.u.Y %*% solve(V.Y)%*%(Y-beta.hat)
```

```
       [,1]
[1,] 1.7087
```

Compare with ranef command:

```
> ranef(fm1)$Subject[1,1]
```

```
[1] 1.7088
```

Calculate predicted random effects for all subjects:

```
> #t(cov.u.Y %*% solve(V.Y)%*%
> #    (matrix(ergoStool$effort,4,9)-
> #       matrix(fixef(fm1),4,9)))
> #ranef(fm1)
```

Correlation of fixed effects

For an ordinary linear model, the covariance matrix (from which we can get the correlation matrix) of $\hat{\beta}$ is:

$$\sigma^2 \times (X^T X)^{-1}. \tag{53}$$

For a mixed effects model, the standard deviations (standard errors) and correlations for the fixed effects estimators are listed at the end of the lmer output.

```
> lm.full<-lmer(wear~material-1+
+                   (1|Subject),
+                  data = BHHshoes)
```

```
Correlation of Fixed Effects:
          matrlA
materialB 0.988
```

$$\hat{\beta}_1 = (Y_{1,1} + Y_{2,1} + \cdots + Y_{10,1})/10 \tag{54}$$

$$\hat{\beta}_2 = (Y_{1,2} + Y_{2,2} + \cdots + Y_{10,2})/10 \tag{55}$$

```
> b1.vals<-subset(BHHshoes,
+                  material=="A")$wear
> b2.vals<-subset(BHHshoes,
+                  material=="B")$wear
> vcovmatrix<-var(cbind(b1.vals,b2.vals))
> ## get covariance from off-diagonal:
> covar<-vcovmatrix[1,2]
> sds<-sqrt(diag(vcovmatrix))
> ## correlation of fixed effects:
> covar/(sds[1]*sds[2])
```

```
b1.vals
0.98823
```

```
> #cf:
> covar/((0.786*sqrt(10))^2)
```

```
[1] 0.98752
```