Linear mixed models fitted with `lmer()` in R: p–values based on a Kenward-Roger modification of the F-statistic or on parametric bootstrap methods.

Ulrich Halekoh and Søren Højsgaard

Department of Molecular Biology and Genetics
Aarhus University, Denmark

10. august 2011

Dependence of yield [kg] and sugar percentage of sugar beets on harvest time and sowing time is investigated.

Five sowing times and two harvesting times were used.

The experiment was laid out in three blocks.

```
Experimental plan for sugar beets experiment

Sowing times:
  1: 4/4, 2: 12/4, 3: 21/4, 4: 29/4, 5: 18/5

Harvest times:
  1: 2/10, 2: 21/10

Plot allocation:
           |  Block 1  |  Block 2  |  Block 3  |
           +-----------|-----------|-----------+
     Plot  | 1 1 1 1 1 | 2 2 2 2 2 | 1 1 1 1 1 | Harvest time
      1-15 | 3 4 5 2 1 | 3 2 4 5 1 | 5 2 3 4 1 | Sowing time
           |-----------|-----------|-----------|
     Plot  | 2 2 2 2 2 | 1 1 1 1 1 | 2 2 2 2 2 | Harvest time
     16-30 | 2 1 5 4 3 | 4 1 3 2 5 | 1 4 3 2 5 | Sowing time
           +-----------|-----------|-----------+
```

Let h denote harvest time ($h = 1, 2$), b denote block ($b = 1, 2, 3$) and s denote sowing time ($s = 1, \dots, 5$). Let $H = 2$, $B = 3$ and $S = 5$.

For simplicity we assume that there is no interaction between sowing and harvesting times.

A typical model for such an experiment would be:

$$y_{hbs} = \mu + \alpha_h + \beta_b + \gamma_s + U_{hb} + \epsilon_{hbs}, \tag{1}$$

where $U_{hb} \sim N(0, \omega^2)$ and $\epsilon_{hbs} \sim N(0, \sigma^2)$.

Notice that U_{hb} describes the random variation between whole–plots (within blocks).

As the design is balanced we may make F–tests for each of the effects as:

R-code

```
> beets$bh <- with(beets, interaction(block, harvest))
> summary(aov(sugpct~block+sow+harvest+Error(bh), beets))

Error: bh
          Df Sum Sq Mean Sq F value Pr(>F)
block      2 0.0327  0.0163    2.58   0.28
harvest    1 0.0963  0.0963   15.21   0.06
Residuals  2 0.0127  0.0063

Error: Within
          Df Sum Sq Mean Sq F value  Pr(>F)
sow        4   1.01  0.2525     101 5.7e-13
Residuals 20   0.05  0.0025
```

Notice: the F–statistics are $F_{1,2}$ for harvest time and $F_{4,20}$ for sowing time.

Using lmer() from lme4 we can fit the models and test for no effect of sowing and harvest time as follows:

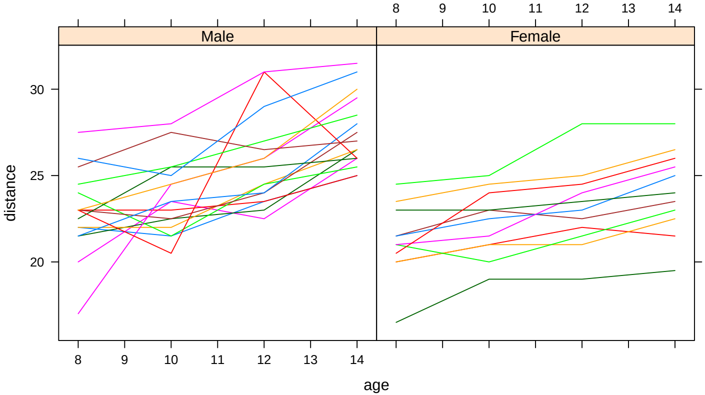
R-code

```
> beetLarge<-lmer(sugpct~block+sow+harvest+(1|block:harvest),
+             data=beets, REML=FALSE)
> beet_no.harv <- update(beetLarge, .~.-harvest)
> beet_no.sow  <- update(beetLarge, .~.-sow)
> as.data.frame(anova(beetLarge, beet_no.sow))

            Df     AIC     BIC logLik Chisq Chi Df Pr(>Chisq)
beet_no.sow  6  -2.795   5.612  7.398    NA     NA         NA
beetLarge   10 -79.997 -65.985 49.999  85.2      4  1.374e-17

> as.data.frame(anova(beetLarge, beet_no.harv))

             Df    AIC    BIC logLik Chisq Chi Df Pr(>Chisq)
beet_no.harv  9 -69.08 -56.47  43.54    NA     NA         NA
beetLarge    10 -80.00 -65.99  50.00 12.91      1  0.0003262
```

The LRT based p–values are anti–conservative: the effect of harvest appears stronger than it is.

Random coefficient model

The change with age of the distance between two cranial fissures was observed for 16 boys and 11 girls from age 8 until age 14.



Random coefficient model

Plot suggests:

$$dist_{[i]} = \alpha_{sex[i]} + \beta_{sex[i]} age_{[i]} + A_{Subj[i]} + B_{Subj[i]} age_{[i]} + e_{[i]}$$

with $(A, B) \sim N(0, \mathbf{S})$.

ML-test of $\beta_{boy} = \beta_{girl}$:

R-code

```
> ort1ML<- lmer(distance ~ age + Sex + age:Sex + (1 + age | Subject),
+                         REML = FALSE, data=Orthodont)
> ort2ML<- update(ort1ML, .~.-age:Sex)
> as.data.frame(anova(ort1ML, ort2ML))

       Df   AIC   BIC logLik Chisq Chi Df Pr(>Chisq)
ort2ML  7 446.8 465.6 -216.4    NA     NA         NA
ort1ML  8 443.8 465.3 -213.9 5.029      1    0.02492
```

Our goal...

Our goal is to improve on the tests provided by `lmer()`.

There are two issues here:

- The choice of test statistic and
- The reference distribution in which the test statistic is evaluated.

Setting the scene

For multivariate normal data

$$Y_{n\times 1} \sim N(\mathbf{X}_{n\times p}\boldsymbol{\beta}_{p\times 1}, \boldsymbol{\Sigma})$$

we consider the test of the hypothesis

$$\mathbf{L}_{l\times p}\boldsymbol{\beta} = \boldsymbol{\beta}_0$$

where $\mathbf{L}$ is a regular matrix of estimable functions of $\boldsymbol{\beta}$.

The linear hypothesis can be tested via the Wald-type statistic

$$F = \frac{1}{l}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top \mathbf{L}^\top (\mathbf{L}^\top \boldsymbol{\Phi}(\hat{\boldsymbol{\sigma}})\mathbf{L})^{-1}\mathbf{L}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \qquad (2)$$

$\boldsymbol{\Phi} = (\mathbf{X}^\top \boldsymbol{\Sigma}\mathbf{X})^{-1}$: the asymptotic covariance matrix of the REML estimate $\hat{\boldsymbol{\beta}}$,

$\hat{\boldsymbol{\sigma}}$: vector of REML estimates of the elements of $\boldsymbol{\Sigma}$

Kenward and Roger's modification

Kenward and Roger (1997) modify the test statistic

- Φ is replaced by an improved small sample approximation Φ_A

Furthermore

- the statistic is scaled by a factor λ
- denominator degrees of freedom m are determined

such that the approximate expectation and variance are those of a $F_{l,m}$ distribution.

Restriction on covariance

If $\mathbf{\Sigma}$ is linear combination of known matrices $\mathbf{G}_i$

$$\mathbf{\Sigma} = \sum_i \sigma_i \mathbf{G}_i \tag{3}$$

then $\Phi_A(\hat{\boldsymbol{\sigma}})$ is dependent only on the first partial deriviatives of $\mathbf{\Sigma}^{-1}$: $\frac{\partial \mathbf{\Sigma}^{-1}}{\partial \sigma_i} = -\mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \sigma_i} \mathbf{\Sigma}^{-1}$.

Notice: Variance component and random coefficient models satisfy this restriction.

$\Phi_A(\hat{\boldsymbol{\sigma}})$ depends also on $Var(\hat{\boldsymbol{\sigma}})$.

Kenward and Roger propose to estimate $Var(\hat{\boldsymbol{\sigma}})$ via the inverse expected information matrix.

R package lme4

The R package lme4 (Bates, D., Maechler, M, Bolker, B., 2011) provides efficient estimation of linear mixed models.

The package provides all necessary matrices and estimates to implement the Kenward-Roger approach.

Properties of the Kenward–Roger adjustment

The modification of the F-statistic by Kenward and Roger

- yields the exact F-statistic in case of Hotelling multivariate T–test and for ANOVA-models which allow exact F–statistics.
- Simulation studies (e.g. Spilke, J. et al.(2003)) indicate that the Kenward-Roger approach perform mostly better than alternatives (like Satterthwaite or containment method) for blocked experiments even with missing data.

Kenward–Roger: split-plot (sugar-beets)

The Kenward–Roger approach yields the same results as the anova-test:

R-code

```
> beetLarge <- update(beetLarge, REML=TRUE)
> beet_no.harv <- update(beet_no.harv, REML=TRUE)
```

Test for harvest effect:

R-code

```
> KRmodcomp(beetLarge,beet_no.harv)$stats[c('df2','Fstat','pval')]

     df2    Fstat     pval
 2.00038 15.20898  0.05988
```

Kenward–Roger: random regression (cranial change)

For the data on change in cranial distances the Kenward and Roger modified F-test yields

R-code

```
> ort1<- update(ort1ML, .~., REML = TRUE)
> ort2<- update(ort2ML, .~., REML = TRUE)
> KRmodcomp(ort1,ort2)$stats[c('df2','pval')]

     df2     pval
24.99863  0.03262
```

The p-value form the ML-test was 0.0249.

Using parametric bootstrap

We consider two models M_0 and M_1 where $M_0 \subset M_1$. We have linear mixed effects models with difference in the fixed effect space in mind but the approach here applies more generally. The p-value for testing the small against the large model is

$$p = \sup_{\theta \in \Theta_0} P_\theta(T \geq t_{obs})$$

where t_{obs} the observed value of a test statistic T.

Using the log-likelihood ratio test statistic T the large sample approximation uses

$$p^{LRT} = P_{\chi^2_f}(T > t_{obs})$$

where f is the difference in parameters of the two models and

We consider additionally the parametric bootstrap p-value

$$p^{PB} = P_{\hat{\theta}_0}(T \geq t_{obs})$$

Parametric bootstrapping

For the parametric bootstrap we simulate under the hypothesis.

- To calculate p^{PB} we draw B (say $B = 1000$) parametric bootstrap samples $y^1, \dots, y^B$ by simulating from $f_0(y|\hat{\theta}_0)$ and calculate the corresponding values $t^1, \dots, t^M$ of T.
- The values $t^1, \dots, t^M$ provide a reference distribution in which t_{obs} can be evaluated.

Improve limiting χ^2 distribution of T by Bartlett–type correction.

That is we want to find a value K such that for

$$T' = K \cdot T \text{ we have } \mathbf{E}(T') = f.$$

We propose the estimate

$$K = \frac{f}{\bar{T}}$$

where $\bar{T}$ denotes the average of the bootstrap sample $t^1, \dots, t^M$.

Typically, the distribution of T will have a heavier tail than a χ^2_f distribution such that $\bar{T} > f$. Hence adjusting T by the factor $f/\bar{T}$ will "shrink"T towards zero.

Extension: Assume T follows a gamma distribution with mean and variance determined by the estimated mean and variance of the parametric bootstrap samples $t^1, \dots, t^M$

Finding a Bartlett correction using PB

Results from sugar beets:

Tabel: p-values ($\times$ 100) for removing the harvest or sow effect.

	LRT	KR	ParmBoot	Bartlett	Gamma
harvest	<0.001	6	4.1	8.3	4.9
sow	<0.001	<0.001	<0.001	<0.001	<0.001

Results for cranial distance data:

Tabel: p-values ($\times$ 100) testing the sex:age interaction.

	LRT	KR	ParmBoot	Bartlett	Gamma
sex:age	2.5	3.3	4.2	4.0	4.2

Random coefficient model

We consider the simulation from a simple random coefficient model (cf. Kenward and Roger (1997, table 4)):

$$y_{it} = (\beta_0 + \epsilon_i^0) + (\beta_1 + \epsilon_i^1)t_i + \epsilon_{it} \tag{4}$$

with $cov(\epsilon_i^0, \epsilon_i^1) = \begin{bmatrix} 0.250 & -0.133 \\ -0.133 & 0.250 \end{bmatrix}$ and $var(\epsilon_{it}) = 0.25$.

There are observed $i = 1, \ldots, 24$ subjects divided in groups of 8. For each group observations are at the non overlapping times $t = 0, 1, 2; t = 3, 4, 5$ and $t = 6, 7, 8$.

Results from random coefficient model

Tabel: Observed test sizes ($\times 100$) for $H_0 : \beta_k = 0$ for random coefficient model.

	LR	Wald	ParmBoot	Bartlett	Gamma	KR(R)	KR(SAS)
β_0	6.8	8.8	5.6	5.4	5.8	4.0	4.8
β_1	7.1	6.6	5.6	5.4	5.7	5.4	5.0

Summary

The functions described here are available in the doBy package on CRAN.

The Kenward–Roger approach requires fitting by REML; the parametric bootstrapping approaches requires fitting by ML.

The required fitting scheme is set by the relevant functions, so the user needs not worry about this.

Literature

- Bates, D., Maechler, M. and Bolker, B. (2011) *lme4: Linear mixed-effects models using S4 classes*, R package version 0.999375-39.
- Kenward, M. G. and Roger, J. H. (1997) *Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood*, Biometrics, Vol. 53, pp. 983–997
- Spilke J., Piepho, H.-P. and Hu, X. Hu (2005) *A Simulation Study on Tests of Hypotheses and Confidence Intervals for Fixed Effects in Mixed Models for Blocked Experiments With Missing Data* Journal of Agricultural, Biological, and Environmental Statistics, Vol. 10,p. 374-389