# Kenward-Roger modification of the F-statistic for some linear mixed models fitted with lmer

Ulrich Halekoh [1] Søren Højsgaard [2]

[1]Department of Molecular Biology and Genetics
Aarhus University, Denmark
*ulrich.halekoh@agrsci.dk*

[2]Department of Mathematical Sciences
Aalborg University, Denmark
*sorenh@math.aau.dk*

9. november 2011

## Motivation: Sugar beets - A split–plot experiment

- Dependence of sugar percentage of sugar beets on harvest time and sowing time is investigated.
- Five sowing times ($s$) and two harvesting times ($h$).
- Experiment was laid out in three blocks ($b$).

```
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  | h1  h1  h1  h1  h1  | h2  h2  h2  h2  h2  | h1  h1  h1  h1  h1  | Harvest time
 1-15 | s3  s4  s5  s2  s1  | s3  s2  s4  s5  s1  | s5  s2  s3  s4  s1  | Sowing time
      |---------------------|---------------------|---------------------|
Plot  | h2  h2  h2  h2  h2  | h1  h1  h1  h1  h1  | h2  h2  h2  h2  h2  | Harvest time
16-30 | s2  s1  s5  s4  s3  | s4  s1  s3  s2  s5  | s1  s4  s3  s2  s5  | Sowing time
      +---------------------|---------------------|---------------------+
```

- For simplicity we assume that there is no interaction between sowing and harvesting times.
- A typical model for such an experiment would be:

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

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:

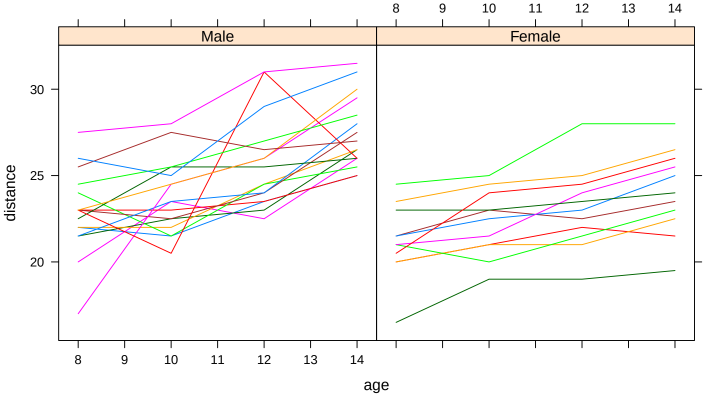
R-code

```
> 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 distances 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 extend 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

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

- $\boldsymbol{\Phi}(\boldsymbol{\sigma}) = (\mathbf{X}^\top \boldsymbol{\Sigma}(\boldsymbol{\sigma})\mathbf{X})^{-1} \approx \mathbb{C}\text{ov}(\hat{\boldsymbol{\beta}})$, $\hat{\boldsymbol{\beta}}$ REML estimate of $\boldsymbol{\beta}$
- $\hat{\boldsymbol{\sigma}}$: vector of REML estimates of the elements of $\boldsymbol{\Sigma}$

# Kenward and Roger's modification

Kenward and Roger (1997) modify the test statistic

- $\Phi$ is replaced by an improved small sample approximation $\Phi_A$

Furthermore

- the statistic $F$ is scaled by a factor $\lambda$,
- 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

- Consider only situations where

$$\boldsymbol{\Sigma} = \sum_i \sigma_i \mathbf{G}_i, \quad \mathbf{G}_i \text{ known matrices}$$

- Variance component and random coefficient models satisfy this restriction.
- $\boldsymbol{\Phi}_A(\hat{\boldsymbol{\sigma}})$ depends now only on the first partial derivatives of $\boldsymbol{\Sigma}^{-1}$:

$$\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_i} = -\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_i} \boldsymbol{\Sigma}^{-1}.$$

- $\boldsymbol{\Phi}_A(\hat{\boldsymbol{\sigma}})$ depends also on $\mathbb{V}\text{ar}(\hat{\boldsymbol{\sigma}})$.
- Kenward and Roger propose to estimate $\mathbb{V}\text{ar}(\hat{\boldsymbol{\sigma}})$ via the inverse expected information matrix.

# Properties of the Kenward–Roger adjustment

The modification of the F-statistic by Kenward and Roger

- yields the exact F-statistic for balanced mixed classification nested models or balanced split plot models (Alnosaier, 2007).
- 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.

# 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.

1. The implementation uses a straightforward transcription of the description in the article of Kenward and Roger, 1997.
2. Matrix operations use sparse matrices representation.
3. Matrices are extracted from `lmer` objects via their slots (using `@`).

# 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)

F-test with Kenward-Roger approximation
large : sugpct ~ block + sow + harvest + (1 | block:harvest)
small : sugpct ~ block + sow + (1 | block:harvest)
 Fstat df1 df2 p.value F.scaling
 15.21   1   2  0.0599         1
```

# Kenward–Roger: random regression (cranial change)

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

R-code

```
> ort1<- update(ort1ML, .~., REML = TRUE)
> ort2<- update(ort2ML, .~., REML = TRUE)
> KRmodcomp(ort1,ort2)

F-test with Kenward-Roger approximation
large : distance ~ age + Sex + (1 + age | Subject) + age:Sex
small : distance ~ age + Sex + (1 + age | Subject)
 Fstat df1 df2 p.value F.scaling
 5.118   1  25  0.0326         1
```

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

# Using parametric bootstrap

- The Kenward–Roger approach is no panacea.
- Additionally, we provide the parametric bootstrap p-value $P_{\hat{\theta}_0}(T \geq t_{obs})$ based on the log-LR statistic $T$.

We draw $B$ parametric bootstrap samples $t^1, \dots, t^B$ under the estimated null model $\hat{\theta}_0$ and provide three choices to calculate the p-value.

1. directly via the proportion of sampled $t_i$ exceeding $t_{obs}$,
2. approximating the distribution of the scaled statistic $\frac{f}{\bar{t}} \cdot T$ by a $\chi^2_f$ distribution (Bartlett type correction)
   ($\bar{t}$ is the sample average and $f$ the difference in the number of parameters between the null and the alternative model)
3. approximating the bootstrap distribution by a $\Gamma(\alpha, \beta)$ distribution which mean and variance match the moments of the bootstrap sample.

**R-code**

```
> PBmodcomp(beetLarge,beet_no.harv)

large : sugpct ~ block + sow + harvest + (1 | block:harvest)
small : sugpct ~ block + sow + (1 | block:harvest)
Number of parametric bootstrap samples: 200
          stat df p.value
LRT      11.815  1  0.0006
PBtest   11.815 NA  0.0550
Bartlett  2.447  1  0.1178
Gamma    11.815 NA  0.0782
```

Results from sugar beets:

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

| | LRT | KR | ParmBoot | Bartlett | Gamma |
|---|---|---|---|---|---|
| harvest | 0.03 | 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 $\beta_{boy} = \beta_{girl}$.

| LRT | KR | ParmBoot | Bartlett | Gamma |
|---|---|---|---|---|
| 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)):

$$y_{it} = \beta_0 + \beta_1 \cdot t_i + A_i + B_i \cdot t_i + \epsilon_{it}$$

with $cov(A_i, B_i) = \begin{bmatrix} 0.250 & -0.133 \\ -0.133 & 0.250 \end{bmatrix}$ and $var(\epsilon_{it}) = 0.25$.
There are observed $i = 1, \dots, 24$ 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) |
|---|---|---|---|---|---|---|---|
| $\beta_0$ | 6.8 | 8.8 | 5.6 | 5.4 | 5.8 | 4.0 | 4.8 |
| $\beta_1$ | 7.1 | 6.6 | 5.6 | 5.4 | 5.7 | 5.4 | 5.0 |

# Summary

- The functions `KRmodcomp()` and `PBmodcomp()` described here are available in the `pbkrtest` package.
- The Kenward–Roger approach requires fitting by REML; the parametric bootstrap approaches requires fitting by ML.
- The required fitting scheme is set by the relevant functions, so the user needs not worry about this.
- Parametric bootstrap is parallelized using the `snow` package.

# Literature


- Alnosaier, W. (2007) *Kenward-Roger Approximate F Test for Fixed Effects in Mixed Linear Models*, Dissertation, Oregon State University
- 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