# Nested by design: Model fitting and interpretation in a mixed model era

Holger Schielzeth1 & Shinichi Nakagawa2

# Supplementary Information:

# Worked example with R code

The following example assumes a treatment with four levels applied to 10 study sites. The simulated includes four sources of variance (treatment main effect, study site main effect, treatment x site interaction variance and residual variance). We use a very low residual variance in order to demonstrate more clearly how the other three sources of variance are estimated.

We first generate the true underlying effects:

# Seed set for the random number generator to ensure reproducibility

set.seed(50000)

# Definition of the treatment main effects

treatment.main.effect = rnorm(5, 0, 5)

names(treatment.main.effect) = LETTERS[1:5]

# Definition of the site main effect

site.main.effect = rnorm(10, 0, 5)

names(site.main.effect) = letters[1:10]

# Definition of the treatment x sites interaction effect

interaction.effect = rnorm(50, 0, 5)

names(interaction.effect) = paste(rep(LETTERS[1:5], each = 10),

 rep(letters[1:10], 5), sep = "")

# Summary of the true underlying effect

true.effects = rep(treatment.main.effect, each = 10) +

rep(site.main.effect, 5) +

interaction.effect

names(true.effects) = paste(rep(LETTERS[1:5], each = 10),

rep(letters[1:10], 5), sep = "")

We then simulate sampling using a nested experimental design:

# Simulation of response values for nested sampling

nested.sampling = rep(true.effects[c(1:2, 13:14, 25:26, 37:38, 49:50)],

each = 50) + rnorm(500, 0, 0.1)

table(treatment <- substr(names(nested.sampling), 1, 1),

sites.nested <- substr(names(nested.sampling), 2, 2))

## a b c d e f g h i j

## A 50 50 0 0 0 0 0 0 0 0

## B 0 0 50 50 0 0 0 0 0 0

## C 0 0 0 0 50 50 0 0 0 0

## D 0 0 0 0 0 0 50 50 0 0

## E 0 0 0 0 0 0 0 0 50 50

# Simulation of response values for crossed sampling

crossed.sampling = rep(true.effects[c(1:50)], each = 10) +

rnorm(500, 0, 0.1)

table(treatment <- substr(names(crossed.sampling), 1, 1),

sites.crossed <- substr(names(crossed.sampling), 2, 2))

## a b c d e f g h i j

## A 10 10 10 10 10 10 10 10 10 10

## B 10 10 10 10 10 10 10 10 10 10

## C 10 10 10 10 10 10 10 10 10 10

## D 10 10 10 10 10 10 10 10 10 10

## E 10 10 10 10 10 10 10 10 10 10

Next we fit the models with treatment as a fixed effect and site as a random effect:

# package lme4 might need to be downloaded: install.packages('lme4')

require(lme4)

# Fit of the model for the nested dataset

nested.mod = lmer(nested.sampling ~ factor(treatment) +

(1 | sites.nested))

summary(nested.mod)

## Linear mixed model fit by REML

## Formula: nested.sampling ~ factor(treatment) + (1 | sites.nested)

## AIC BIC logLik deviance REMLdev

## -789 -759 401 -775 -803

## Random effects:

## Groups Name Variance Std.Dev.

## sites.nested (Intercept) 86.97093 9.3258

## Residual 0.00968 0.0984

## Number of obs: 500, groups: sites.nested, 10

##

## Fixed effects:

## Estimate Std. Error t value

## (Intercept) -9.80 6.59 -1.49

## factor(treatment)B 7.35 9.32 0.79

## factor(treatment)C 5.41 9.32 0.58

## factor(treatment)D 13.77 9.32 1.48

## factor(treatment)E 20.77 9.32 2.23

##

## Correlation of Fixed Effects:

## (Intr) fct()B fct()C fct()D

## fctr(trtm)B -0.707

## fctr(trtm)C -0.707 0.500

## fctr(trtm)D -0.707 0.500 0.500

## fctr(trtm)E -0.707 0.500 0.500 0.500

# Fit of the model for the crossed dataset

crossed.mod = lmer(crossed.sampling ~ factor(treatment) +

(1 | sites.crossed))

summary(crossed.mod)

## Linear mixed model fit by REML

## Formula: crossed.sampling ~ factor(treatment) + (1 | sites.crossed)

## AIC BIC logLik deviance REMLdev

## 2769 2798 -1377 2759 2755

## Random effects:

## Groups Name Variance Std.Dev.

## sites.crossed (Intercept) 45.0 6.71

## Residual 13.3 3.65

## Number of obs: 500, groups: sites.crossed, 10

##

## Fixed effects:

## Estimate Std. Error t value

## (Intercept) -10.664 2.152 -5.0

## factor(treatment)B 7.988 0.516 15.5

## factor(treatment)C 8.548 0.516 16.6

## factor(treatment)D 13.649 0.516 26.5

## factor(treatment)E 18.096 0.516 35.1

##

## Correlation of Fixed Effects:

## (Intr) fct()B fct()C fct()D

## fctr(trtm)B -0.120

## fctr(trtm)C -0.120 0.500

## fctr(trtm)D -0.120 0.500 0.500

## fctr(trtm)E -0.120 0.500 0.500 0.500

We see that the interaction variance inflates the site variance estimate in the nested design, whereas it is assigned to the residual variance in the crossed model. Note that both models do not explicitly model the interaction variance.

We can also fit the model with two random effects. The model can be fitted to the nested and the crossed dataset using the exact same syntax, since the coding employed here reflects implicit nesting. Note the shift in the assignment of the interaction variance to the site variance in the nested dataset and to the residual component in the crossed model.

# Fit of a random effect model for the nested dataset (using the syntax of

# crossed model fitting)

nested.raneff.mod <- lmer(nested.sampling ~ (1 | treatment)

+ (1 | sites.nested))

summary(nested.raneff.mod)

## Linear mixed model fit by REML

## Formula: nested.sampling ~ (1 | treatment) + (1 | sites.nested)

## AIC BIC logLik deviance REMLdev

## -765.2 -748.3 386.6 -768.8 -773.2

## Random effects:

## Groups Name Variance Std.Dev.

## sites.nested (Intercept) 87.97 9.33

## treatment (Intercept) 20.74 4.55

## Residual 0.01 0.10

## Number of obs: 500, groups: sites.nested, 10; treatment, 5

## Fixed effects:

## Estimate Std. Error t value

## (Intercept) -0.335 3.583 -0.093

# Fit of a random effect model for the crossed dataset

crossed.raneff.mod <- lmer(crossed.sampling ~ (1 | treatment)

+ (1 | sites.crossed))

summary(crossed.raneff.mod)

## Linear mixed model fit by REML

## Formula: crossed.sampling ~ (1 | treatment) + (1 | sites.crossed)

## AIC BIC logLik deviance REMLdev

## 2791 2808 -1392 2788 2783

## Random effects:

## Groups Name Variance Std.Dev.

## sites.crossed (Intercept) 44.98 6.71

## treatment (Intercept) 45.97 6.78

## Residual 13.30 3.65

## Number of obs: 500, groups: sites.crossed, 10; treatment, 5

##

## Fixed effects:

## Estimate Std. Error t value

## (Intercept) -1.008 3.703 -0.272

We want to demonstrate that the model fit for the nested dataset using the syntax of crossed random effects gives the correct answer for a nested data structure. It is currently not even possible to use a special syntax for nested effects in lmer(). But we can use the lme() for a comparison (compare nested.raneff.mod to nested.raneff.mod.lme). The estimates are essentially the same (up to slight differences due to different implementations). Note that random effect variances are presented as random effect standard deviations in summary(lme()).

# This might require downloading package nlme: install.packages('nlme'))

require(nlme)

nested.raneff.mod.lme <- lme(nested.sampling ~ 1, random = ~ 1 | treatment/

sites.nested)

summary(nested.raneff.mod.lme)

## Linear mixed-effects model fit by REML

## AIC BIC logLik

## -765.2 -748.3 386.5836

## Random effects:

## Formula: ~1 | treatment

## (Intercept)

## StdDev: 4.55

## Formula: ~1 | sites.nested %in% treatment

## (Intercept) Residual

## StdDev: 9.33 0.098

## Fixed effects: nested.sampling ~ 1

## Value Std.Error DF t-value p-value

## (Intercept) -0.335 3.584 490 -0.09 0.93

## Number of Observations: 500

## Number of Groups:

## treatment sites.nested %in% treatment

## 5 10

So far, we have not explicitly modeled the interaction variance component. But if the experimental design is crossed and there is replication within cells, we can easily extend the random effect model to estimate the interaction variance. Note that this reduces primarily the residual variance component (that was simulated very low).

# Fit of a random effect model for the crossed dataset including the

 interaction variance

interaction.raneff.mod <- lmer(crossed.sampling ~ (1 | treatment)

+ (1 | sites.crossed) + (1 | names(crossed.sampling)))

summary(interaction.raneff.mod)

## Linear mixed model fit by REML

## Formula: crossed.sampling ~ (1 | treatment) + (1 | sites.crossed)

## + (1 | names(crossed.sampling))

## AIC BIC logLik deviance REMLdev

## -362.2 -341.1 186.1 -367.7 -372.2

## Random effects:

## Groups Name Variance Std.Dev.

## names(crossed.sampling) (Intercept) 17.94 4.24

## sites.crossed (Intercept) 41.66 6.45

## treatment (Intercept) 44.31 6.66

## Residual 0.01 0.10

## Number of obs: 500, groups: names(crossed.sampling), 50;

## sites.crossed, 10; treatment, 5

##

## Fixed effects:

## Estimate Std. Error t value

## (Intercept) -1.008 3.659 -0.276