

Mixed models in R

EdinbR April 20 2016

Random effects model

Hierarchical (multilevel) model:

(1) n observations in i -th group $N(m_i, \sigma_W^2)$,

(2) $m_1 \dots m_k$ is random sample from $N(m, \sigma_B^2)$

Variance of each observation is $\sigma_B^2 + \sigma_W^2$

Correlation between two obs in same group $\sigma_B^2/(\sigma_B^2 + \sigma_W^2)$

Expectations of mean squares

$$E(M_W) = \sigma_W^2, \quad E(M_B) = \sigma_W^2 + n\sigma_B^2,$$

where n is group size (if constant).

One-way ANOVA random model

Total variance $\sigma_B^2 + \sigma_W^2$ partitioned into components σ_B^2 and σ_W^2 .
One way to obtain estimates of these components is by equating observed and expected mean squares:

$$\hat{\sigma}_W^2 = M_W, \quad \hat{\sigma}_B^2 = (M_B - M_W)/n$$

In genetic applications, σ_W^2 and σ_B^2 are related to genetic components V_A and V_E :

$$\sigma_B^2 = \lambda V_A, \quad \sigma_W^2 = (1 - \lambda)V_A + V_E$$

$\lambda = 1$ for identical twins, $1/2$ for full sibs, $1/4$ for half sibs

Matrix notation

Mixed model $y = Xb + Zu + e$

b is vector of fixed effects,

u is a vector of random effects,

$E(u) = 0$, $\text{cov}(u) = \sigma_B^2 I$, $\text{cov}(e) = \sigma_W^2 I$.

Alternative formulation

$E(y) = Xb$, $\text{cov}(y) = \sigma_B^2 ZZ' + \sigma_W^2 I$

Example

Ten observations in five groups. First five observations on males, last five on females.

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \\ y_{10} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \\ e_8 \\ e_9 \\ e_{10} \end{bmatrix}$$

Fixed and random effects

Typical factor with fixed effects:

sex, age, herd (systematic effects)

Typical factor with random effects:

Variance between levels is of more interest than differences between particular levels

Repeat of the experiment would have same fixed effects, but different random effects.

Choice between fixed and random is not always clear-cut.

Predicting random effects

Given $y = m + U + e$, predicted value of U is

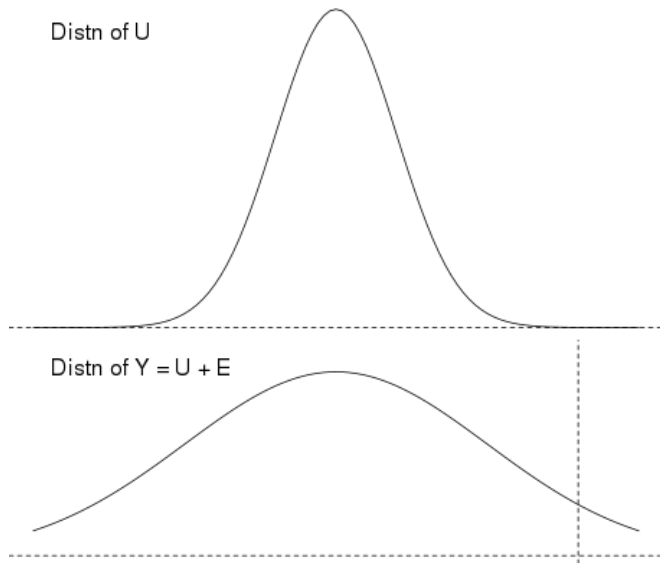
$$\frac{\sigma_B^2}{(\sigma_B^2 + \sigma_W^2)}(y - m)$$

The regression coefficient is less than 1 ('shrinkage').

Galton: 'Each peculiarity in a man is shared by his kinsman, but on the average in a less degree'.

Related to 'winner's curse': e.g. in genome-wide association tests.

Shrinkage



R packages for mixed models

Package	function
nlme	lme
lme4	lmer
asreml-R	asreml
regress	regress
coxme	lmekin
MCMCglmm	MCMCglmm
pedigreemm	pedigreemm (uses lmer)
mgcv	gamm (uses lme)
base	aov
...etc	

Syntax

f is a factor or covariate representing a fixed effect, g represents a factor with random effects.

```
aov( response ~ f + Error(g), data = ... )
```

```
lme( response ~ f, random = ~ 1 | g , data = ... )
```

```
lmer( response ~ f + (1 | g), data = ... )
```

ANOVA;

Between levels of g

Within levels of g

Two levels of nesting

g, h: factors with random effects.

Levels of h usually 'nested' within levels of g (each level of h is combined with exactly one level of g)

'Nested' factors

	h_1	h_2	h_3	h_4	h_5	h_6
g_1	x	x	x	.	.	.
g_2	.	.	.	x	x	x

'Crossed' factors

	h_1	h_2	h_3
g_1	x	x	x
g_2	x	x	x

Syntax for nested random factors

`lmer(response ~ f + (1| g/h) , data = ...)`

which is equivalent to either

`lmer(response ~ f + (1| g) + (1| h) , data = ...)`

or

`lmer(response ~ f + (1| g) + (1| g:h) , data = ...)`

depending on how levels of h are numbered.

ANOVA (nested groups)

Between levels of g

Between levels of h within levels of g

Within levels of h

Finger ridges

Number of finger ridges in 12 pairs of female identical twins.

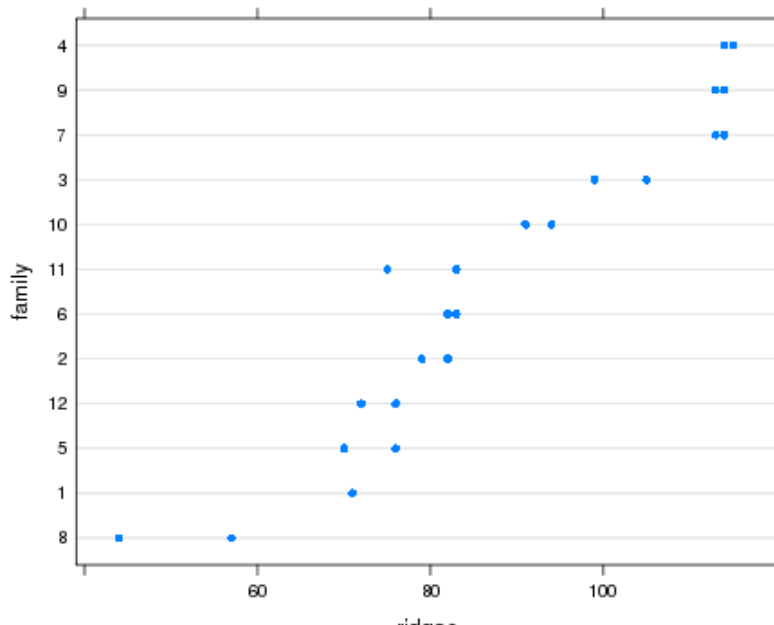
Family											
1	2	3	4	5	6	7	8	9	10	11	12
71	79	105	115	76	83	114	57	114	94	75	76
71	82	99	114	70	82	113	44	113	91	83	72

Source	DF	SSQ	MSQ	F
Between families	11	8990.5	817.31	57.19
Within families	12	171.5	14.29	

$$\hat{\sigma}_W^2 = 14.29, \quad \hat{\sigma}_B^2 = (817.31 - 14.29)/2 = 401.51$$

Estimated heritability of the trait is equal to intraclass correlation (0.97).

twins results



R code for anova

aov provides mean squares M_B and M_W :

```
ridges <- c(71,71,79,82,105,99,115,114,76,70,83,82,  
           114,113,57,44,114,113,94,91,75,83,76,72)  
family <- gl(12,2)  
summary(aov(ridges ~ family))
```

	Df	Sum Sq	Mean Sq	F value
family	11	8990	817.3	57.19
Residuals	12	172	14.3	

lmer code

lmer estimates variance components σ_B^2 and σ_W^2 directly:

```
library(lme4)
summary(lmer(ridges ~ 1 + (1|family)))
```

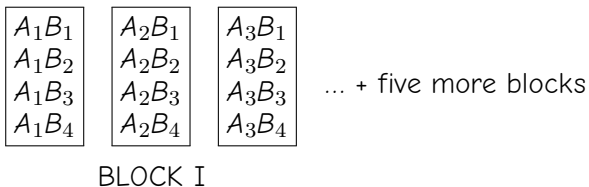
Random effects:

Groups	Name	Variance	Std.Dev.
family	(Intercept)	401.51	20.04
Residual		14.29	3.78

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	87.208	5.836	14.94

Split-plot experiment



A_1, A_2, A_3 three varieties of oats

B_1, B_2, B_3, B_4 four levels of nitrogen fertilizer

Top-level experiment: 6 blocks \times 3 varieties

Nested experiment: 18 plots \times 4 fertilizer treatments

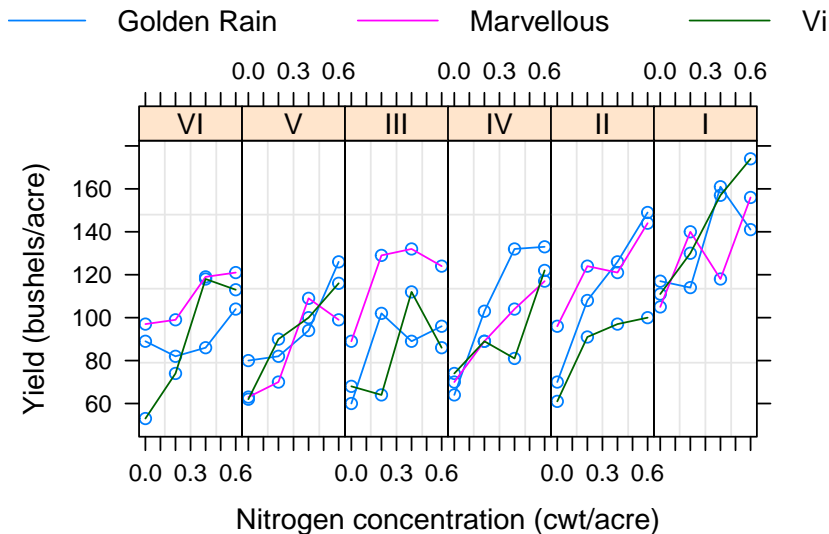
More generally, 'split-unit' experiment.

Oats data (block I)

```
data(oats, package = 'MASS')  
head(oats, n = 12)
```

Variety	Nitrogen	Yield
Victory	0.0cwt	111
Victory	0.2cwt	130
Victory	0.4cwt	157
Victory	0.6cwt	174
Golden.rain	0.0cwt	117
Golden.rain	0.2cwt	114
Golden.rain	0.4cwt	161
Golden.rain	0.6cwt	141
Marvellous	0.0cwt	105
Marvellous	0.2cwt	140
Marvellous	0.4cwt	118
Marvellous	0.6cwt	156

Oats results



Analysis of oats data

```
oats.aov <- aov(Y ~ N*V + Error(B/V), data = oats)
```

```
oats.lme <- lme(Y ~ N*V, random = ~1| B/V, data = oats)
```

```
oats.lmer <- lmer(Y ~ N*V + (1| B/V), data = oats)
```

Anova for split-plot expt

```
summary(oats.aov)
```

Error: B

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Residuals	5	15875	3175		

Error: B:V

V	2	1786	893.2	1.485	0.272
Residuals	10	6013	601.3		

Error: Within N	3	20020	6674	37.686	2.46e-12
N:V	6	322	54	0.303	0.932
Residuals	45	7969	177		

Obvious differences among levels of nitrogen

No evidence of differences among varieties, no interaction

If required, estimate σ_B^2 and σ_W^2 (between and within plots) from the two residual mean squares.

lme(r) summary output

```
summary(oats.lmer)
```

Random effects:

Groups	Name	Variance	Std.Dev.
V:B	(Intercept)	106.1	10.30
B	(Intercept)	214.5	14.65
Residual		177.1	13.31

Fixed effects:

	Estimate	Std. Error	t value
N0.2cwt	18.5000	7.6830	2.408
N0.4cwt	34.6667	7.6830	4.512
N0.6cwt	44.8333	7.6830	5.835
VMarvellous	6.6667	9.7150	0.686
VVictory	-8.5000	9.7150	-0.875

...

Sequential anova for fixed effects

```
anova(oats.lme)
```

	numDF	denDF	F-value	p-value
N	3	45	37.68560	<.0001
V	2	10	1.48534	0.2724
N:V	6	45	0.30282	0.9322

```
anova(oats.lmer)
```

	Df	Sum Sq	Mean Sq	F value
N	3	20020.5	6673.5	37.6856
V	2	526.1	263.0	1.4853
N:V	6	321.8	53.6	0.3028

lme analysis agrees exactly with anova.

p-values

One of the more controversial design decisions of lme4 was to omit p-values associated with sequential anova of fixed effects

...

Null distributions are asymptotically normal, but not t distributions for finite sample sizes. Methods for calculating approximate degrees of freedom are at best *ad hoc*.

– Douglas Bates et al

See page 35 of

vignette('lmer', 'lme4')

Pseudo replication

Standard error of the difference between two variety means is proportional to

$$\sqrt{\frac{\sigma_B^2}{n_1} + \frac{\sigma_W^2}{n_2}}$$

n_1 = number of mainplots per variety (6)

n_2 = number of subplots per variety (24)

Usually $\sigma_B^2 \gg \sigma_W^2$ so for example doubling the number of subplots and keeping the number of mainplots constant might have a negligible effect on the standard error.

Other applications

Incomplete block designs (BIB, PBIB, 'Alpha')

Spatial (or time series) processes (AR, MA, etc)

Repeated measures (allows for pseudo-replication).

Random regression models

Genetic evaluation (the animal model)

GLMM for binomial, Poisson, or gamma distributed data.

Special interest group

For discussion about lme/lmer/MCMCglmm sign up to
r-sig-mixed-models@r-project.org