

With many groups

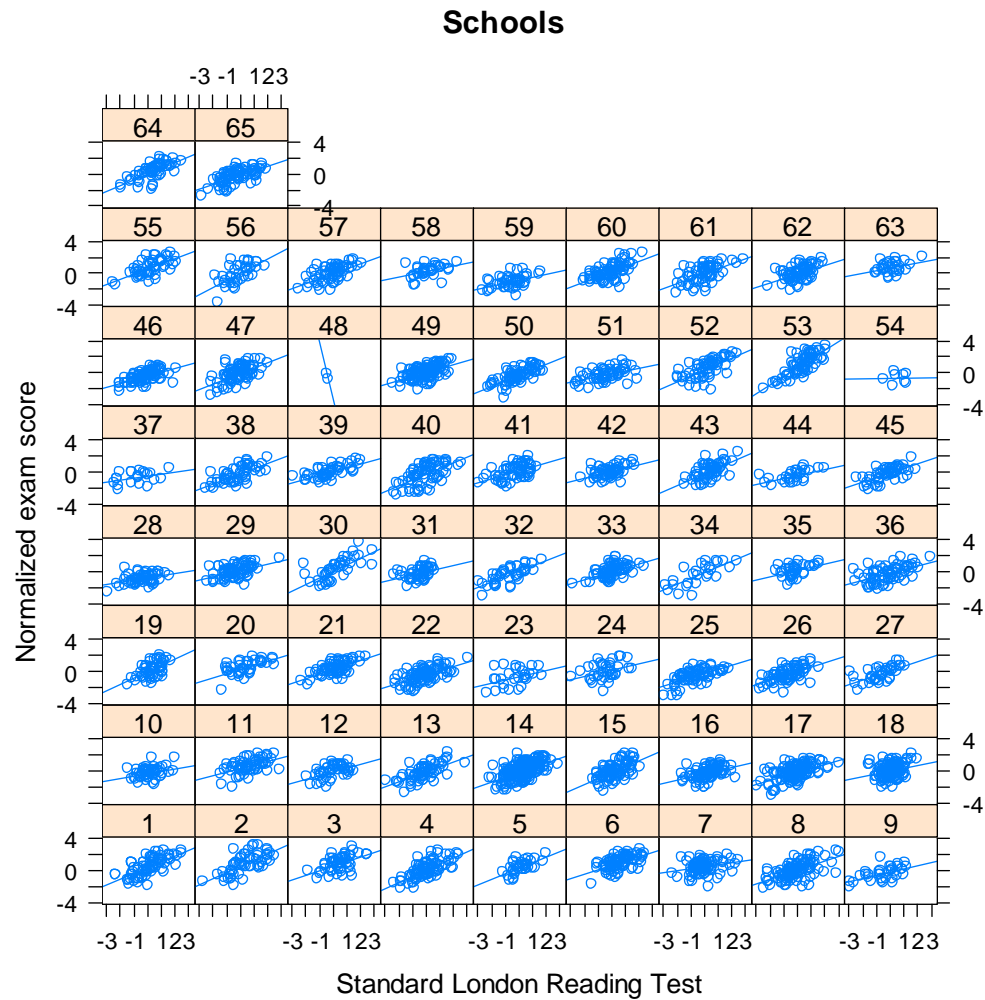
observations i in group j : $\eta_{ij} = \beta_{0j} + \beta_{1j}x_{ij}$

If we fit an intercept and slope for each group, the number of parameters increases with the number of groups.

often we are not primarily interested in the values of intercept and slope, but in their mean and variance between groups. We estimate these using (RE)ML.

Make extra assumptions and limit the number of parameters to be fitted: the parameter values per group are a sample from a population with a given sampling distribution. Using fitted parameters, we can estimate the group-specific parameters

Mixed models



(data in library mlmRev)

Table 1 Maximal model fitted to each dependent variable

| Grouping factor | Description (number of parameters) |
|----------------------------------|---|
| Fixed effect | Age (1) Weekday (3, Monday–Wednesday–Friday) Age \times Weekday Interaction (3) |
| Random effect | |
| Females | Age (1) |
| Females | Weekday (3) |
| Random effect | |
| Males | Age (1) |
| Males | Weekday (3) |
| Random effect | |
| Female \times male combination | Intercept (1) |
| Random effect | |
| Tank | Intercept (1) |
| Error term | (1, only for Reproductive effort and egg size) |

The model contains fixed effects, random effects, and for the reproductive effort and egg size traits also an error variance

Random effects can be estimated if there are repeated observations per level of a grouping factor (fish tank, females, males, or female-male combinations). The weekday variable is categorical with three levels. For the fixed effects, there is a separate parameter estimated per weekday level, for the random effects, a separate variance component is estimated per weekday

Fixed or random effects?

- Are you interested in the effect sizes? → Fixed
- Are factor levels informative? → Fixed
- Are the factor levels a sample from a population? → Random
- Enough levels to estimate variance among effects in the population? → Random
- Are factor levels just labels? → Random
- Main interest is in distribution/spread of effects → Random

Fixed or random effects? One exception to the previous slide

If you want to "borrow strength" from other groups to estimate a parameter for a group with limited sample size

~ When there are very many groups with few observations per group

- Are you interested in the effect sizes? → Random

→ Mixed models in bioinformatics (TMH)

Generalized linear model - recap

Linear predictor The structure of the model relating each observed value y_i of the response to a predicted value η_i .

$$\eta_i = \sum_{j=1}^p x_{ij} \beta_j \quad \text{or in matrix notation} \quad \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$$

x_{ij} are the values of the j -th explanatory variables for observation i .

Link function $h(\boldsymbol{\mu}) = \boldsymbol{\eta}$ relates the expected response to the predictor

The coefficients β_i are estimated by Maximum Likelihood Estimation.

At these $\hat{\beta}_i$ the (log)likelihood of the data given the coefficients and the covariates is maximized.

- For the normal distribution model, the error variance is also estimated or a dispersion parameter for other (quasi)distributions

Generalized linear model - recap

likelihood $L(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X})$

MLE $\hat{\boldsymbol{\beta}}$ are obtained using the IWLS algorithm.

Score $\frac{\partial}{\partial \boldsymbol{\beta}} \ln L(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X})$ is zero at the MLE $\hat{\boldsymbol{\beta}}$

Observed Fisher information $I(\boldsymbol{\beta}) = -\frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \ln L(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X})$

The standard errors of $\hat{\boldsymbol{\beta}}$ are often derived from this curvature of the likelihood surface at $\hat{\boldsymbol{\beta}}$, given by the observed Fisher information $I(\hat{\boldsymbol{\beta}})$.

Confidence intervals for β :

approximately or exactly $I(\hat{\beta})^{-1/2}(\hat{\beta} - \beta_{true}) \sim N(\mathbf{0}, \mathbf{1}_p)$ if beta contains p parameters.

or determined by **profile likelihood** methods.

Generalization: Mixed models for data with gaussian errors

Linear predictor. The structure of the model relating each observed value y_i of the response to a predicted value η_i given covariates and random effects.

$$\eta = X\beta + Zb$$

x_{ij} are the values of the j -th explanatory variables for observation i as above. b are unobserved random effects, which are assumed to be sampled from a (multivariate) normal distribution with zero mean vector and arbitrary variance-covariance matrix Σ_θ .

given the observed explanatory variables, observations are not always independent, because random effects can be correlated

As mentioned in Sect.~1.1, a mixed model incorporates two random variables: \mathcal{B} , the q -dimensional vector of random effects, and \mathcal{Y} , the n -dimensional response vector. In a linear mixed model the unconditional distribution of \mathcal{B} and the conditional distribution, $(\mathcal{Y}|\mathcal{B} = \mathbf{b})$, are both multivariate Gaussian (or “normal”) distributions,

$$\begin{aligned}(\mathcal{Y}|\mathcal{B} = \mathbf{b}) &\sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}, \sigma^2\mathbf{I}) \\ \mathcal{B} &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\theta}).\end{aligned}\tag{1.1}$$

The *conditional mean* of \mathcal{Y} , given $\mathcal{B} = \mathbf{b}$, is the *linear predictor*, $\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}$, which depends on the p -dimensional *fixed-effects parameter*, $\boldsymbol{\beta}$, and on \mathbf{b} . The *model matrices*, \mathbf{X} and \mathbf{Z} , of dimension $n \times p$ and $n \times q$, respectively, are determined from the formula for the model and the values of covariates. Although the matrix \mathbf{Z} can be large (i.e. both n and q can be large), it is sparse (i.e. most of the elements in the matrix are zero).

likelihood with random effects: an example with one level of grouping

for a group i with random effects \mathbf{b}_i (m such groups) $\boldsymbol{\eta}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i$

Integration over the distribution of the (i.i.d.) error with variance σ^2 , and over the distribution of random effects with parameters in $\boldsymbol{\Sigma}_\theta$.

$$L(\boldsymbol{\beta}, \boldsymbol{\Sigma}_\theta, \sigma^2 | \mathbf{y}) = \prod_{i=1}^m \int_{R_q} d\mathbf{b}_i p(\mathbf{y}_i | \mathbf{b}_i, \boldsymbol{\beta}, \sigma^2) p(\mathbf{b}_i | \boldsymbol{\Sigma}_\theta)$$

[p can be a complicated expression itself]

Estimation of variance parameters as independent as possible from estimation of mean parameters of the linear predictor:

REML estimation, restricted maximum likelihood (reduced, residual,..)

$$L_R(\boldsymbol{\Sigma}_\theta, \sigma^2 | \mathbf{y}) = \int_{R_p} d\boldsymbol{\beta} L(\boldsymbol{\beta}, \boldsymbol{\Sigma}_\theta, \sigma^2 | \mathbf{y})$$

REML maximizes the likelihood of the random effect parameters given observations \mathbf{y}

REML averages over uncertainty of parameter estimates of fixed effects, they are integrated out.

Note that $\boldsymbol{\beta}$ has a multivariate uniform distribution in this integral

REML avoid using it for inference on the fixed effects model: use **ML**

The coefficients β_i of different models are preferably compared between models with Maximum Likelihood Estimation

The variances in Σ_θ are preferably estimated with REML

Random variation in intercept between schools

```
library(lme4)
```

```
mm1<-lmer(normexam~standLRT+sex+schgend+(1|school),Exam)
```

```
summary(mm1)
```

Linear mixed model fit by REML ['lmerMod']

Formula: normexam ~ standLRT + sex + schgend + (1 | school)

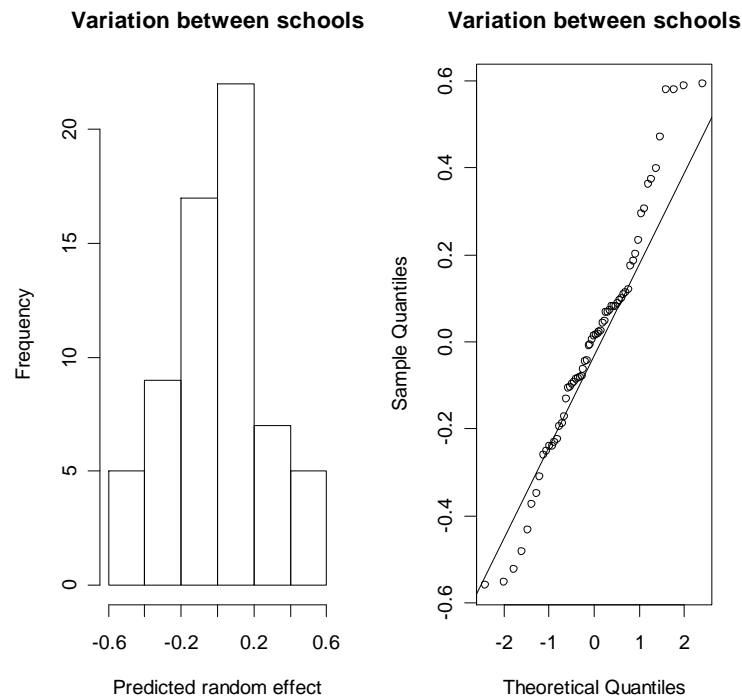
Random effects:

| Groups | Name | Variance | Std.Dev. |
|----------|-------------|----------|----------|
| school | (Intercept) | 0.08583 | 0.293 |
| Residual | | 0.56253 | 0.750 |

Number of obs: 4059, groups: school, 65

Fixed effects:

| | Estimate | Std. Error | t value |
|--------------|-----------|------------|---------|
| (Intercept) | -0.001049 | 0.055570 | -0.02 |
| standLRT | 0.559754 | 0.012450 | 44.96 |
| sexM | -0.167392 | 0.034100 | -4.91 |
| schgendboys | 0.177691 | 0.113476 | 1.57 |
| schgendgirls | 0.158997 | 0.089404 | 1.78 |



- This analysis of mixed models using `lmer()` assumes normally distributed random effects. Random effects can be **predicted** using the parameter estimates and covariates (best guess, Best Linear Unbiased Prediction).
- Fitting of this model with normally distributed errors is done using REML (Restricted maximum likelihood)

Random variation in intercept and slope between schools

```
mm2<-lmer(normexam~standLRT+sex+schgend+(1+standLRT|school),Exam)  
summary(mm2)
```

Random effects:

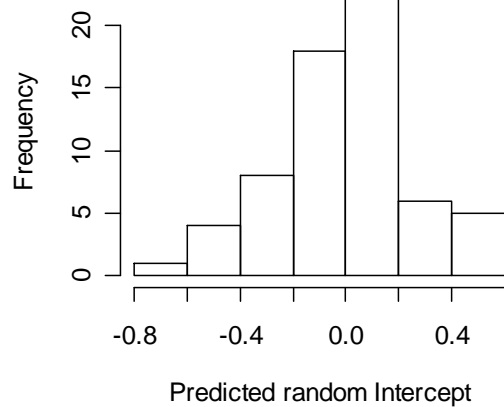
| Groups | Name | Variance | Std.Dev. | Corr |
|--------|-------------|----------|----------|------|
| school | (Intercept) | 0.08366 | 0.2892 | |
| | standLRT | 0.01508 | 0.1228 | 0.57 |
| | Residual | 0.55029 | 0.7418 | |

Number of obs: 4059, groups: school, 65

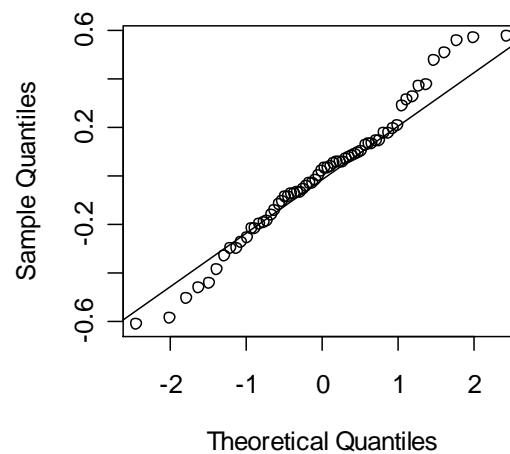
Fixed effects:

| | Estimate | Std. Error | t value |
|--------------|----------|------------|---------|
| (Intercept) | -0.02038 | 0.05316 | -0.383 |
| standLRT | 0.55400 | 0.02012 | 27.537 |
| sexM | -0.16851 | 0.03384 | -4.980 |
| schgendboys | 0.17974 | 0.10171 | 1.767 |
| schgendgirls | 0.17443 | 0.08079 | 2.159 |

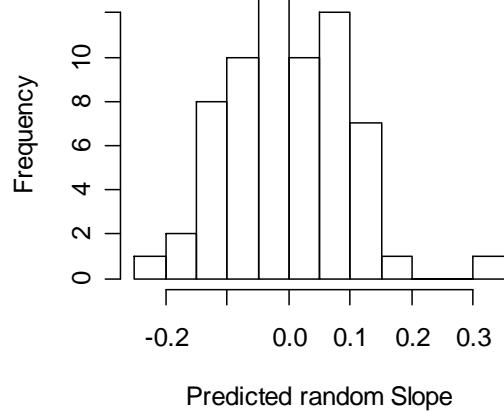
Variation between schools



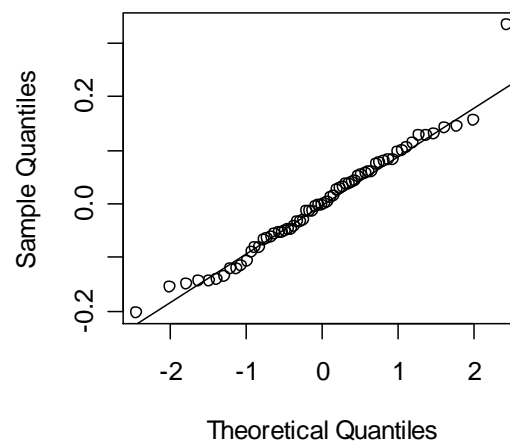
Variation between schools



Variation between schools



Variation between schools



Random effects [**BestLinearUnbiasedPrediction**]

Distinction between prediction and estimation of random effects -
overdone

In lmer you can get the BLUPs.

Shrinkage towards the mean: magnitudes of random effects are in
between the value of the difference ($\mathbf{y} - \mathbf{X}\boldsymbol{\beta}$) and zero.

$$\hat{\mathbf{b}} = \left(\mathbf{Z}^T \mathbf{Z} + \sigma^2 \boldsymbol{\Sigma}_{\theta}^{-1} \right)^{-1} \mathbf{Z}^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

Borrowing strength: average response per random effect group is a
weighted average of overall mean and residual

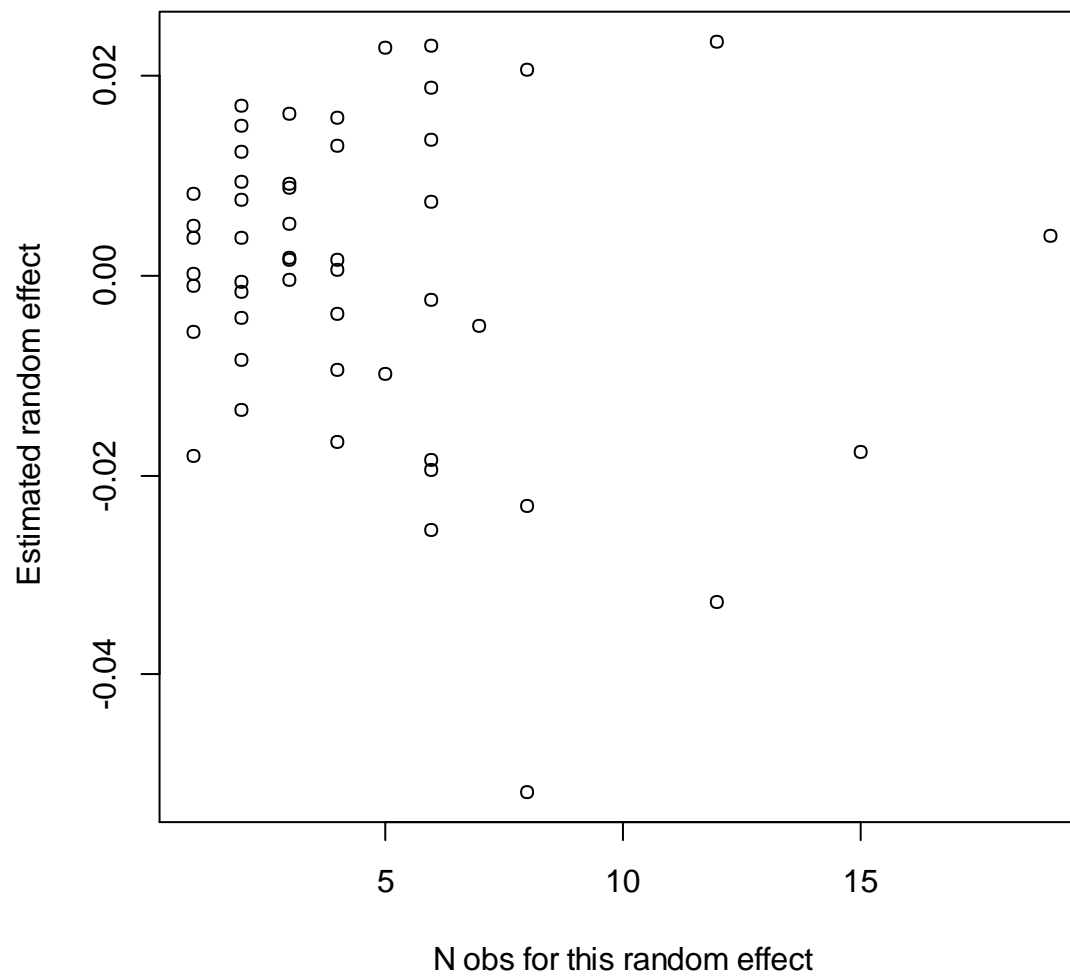
Example

```
nrandoms<-50
randomeffects<-rnorm(nrandoms,sd=.3)
obsperindividual<-rgeom(nrandoms,0.3)+1 #
randomeffectspersobservation<-rep(randomeffects,obsperindividual)
levels<-rep(c(1:nrandoms),obsperindividual)

obs<- randomeffectspersobservation+
rnorm(length(randomeffectspersobservation))

shrinkexample<-lmer(obs~(1|levels))

qqnorm(randomeffects)
qqnorm(ranef(shrinkexample)$levels[[1]])
plot(ranef(shrinkexample)$levels[[1]]~obsperindividual[obsperindividual>0],xlab="N obs for this random effect",ylab="Estimated random effect")
```



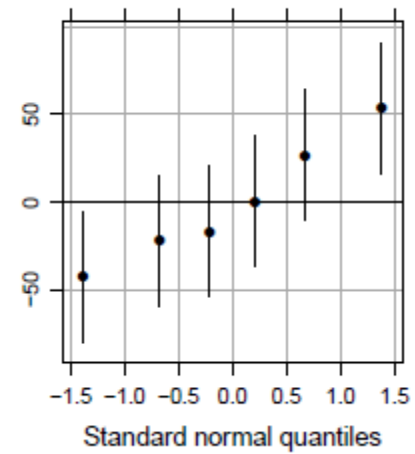
Two types of post-fit diagnosis:

Diagnosis of model **assumptions**: check distributions, is the fixed effects model appropriate?

Diagnosis of **appropriateness of inference**: simulations of tests/confidence intervals and their assumptions

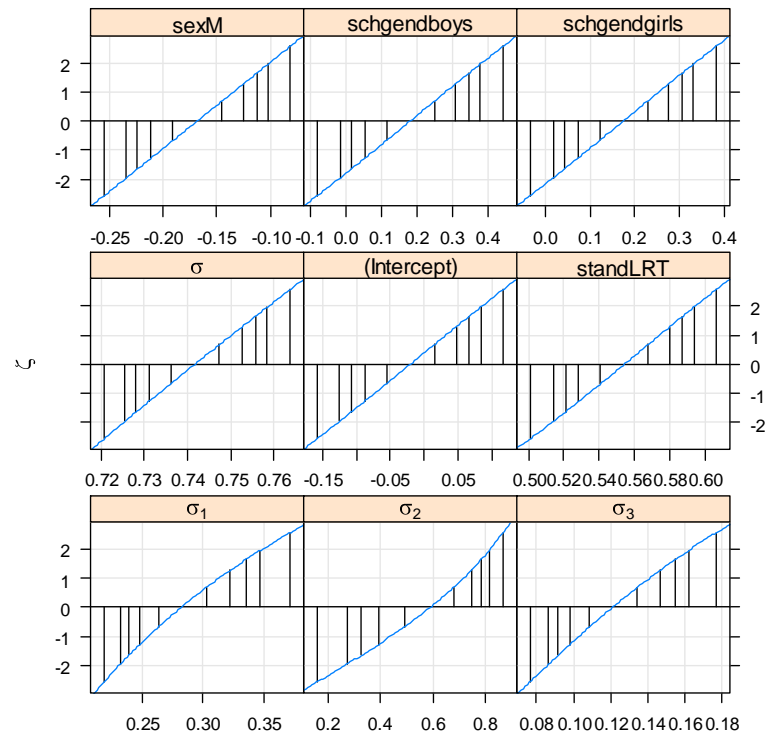
Assess precision and distribution of random effects using qqplots (above) and prediction intervals

Fig. 1.11 95% prediction intervals on the random effects in `fm1ML` versus quantiles of the standard normal distribution.



Assess difference in likelihood ratio statistic when one parameter is fixed and others are again optimized at MLE - zeta is the signed square root of this difference

The **profile zeta** plot: compare zeta to quantiles of normal distribution. Inference using s.e. from curvature of likelihood is ok if straight lines are seen (= symmetry of likelihood at MLE, quadratic approximation OK).



Inference

Progress every year.

Read Bolker (2009), <http://glmm.wikidot.com/>

Fitting models is going quite OK

AIC available, but discussion on what the number of parameters should be

Confidence intervals, hypothesis tests:

Examples given below. Differences between fixed and random effect procedures, look at the wiki first, try different routes

Large-sample results (on distributions of statistics) increasingly abandoned in favour of simulated approaches (e.g. **parametric bootstrap simulations**).

Random effects test

Conservative approach, fifteen years ago:

anova(mm2,mm1)

Data: Exam

Models:

mm1: normexam ~ standLRT + sex + schgend + (1 | school)

mm2: normexam ~ standLRT + sex + schgend + (1 + standLRT | school)

| | Df | AIC | BIC | 1 | ogLik | Chisq | ChiDf | Pr(>Chisq) |
|------------|-----------|---------------|---------------|----------|----------------|---------------|--------------|----------------------|
| mm1 | 6 | 9337.5 | 9375.4 | | -4662.8 | | | |
| mm2 | 8 | 9297.2 | 9347.7 | | -4640.6 | 44.294 | 2 | 2.409e-10 *** |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Now, simulated approach:

for a single random effect, not easy to program. Read ?exactRLRT

```
h0<-mm1 # Null hypothesis with one random effect
```

```
hA<-lmer(normexam~standLRT+sex+schgend+(1|school)+(0+standLRT|school),Exam)
```

```
# alternative with two random effects, the function can only test one effect at a time
```

```
# Null hypothesis
```

```
mmeff<- lmer(normexam~standLRT+sex+schgend+(0+standLRT|school),Exam) #
```

```
model containing only the random effect that will be removed
```

```
library(RLRsim)
```

```
exactRLRT(mmeff,hA,h0)
```

simulated finite sample distribution of RLRT.

(p-value based on 10000 simulated values)

data:

RLRT = 33.6928, p-value < 2.2e-16

Fixed effects: Fit with Maximum Likelihood!

```
mm3<-
```

```
lmer(normexam~standLRT+sex+schgend+(1+standLRT|school),Exam,REML=FALSE)
```

```
mm4<-lmer(normexam~standLRT+sex+(1+standLRT|school),Exam,REML=FALSE)
```

```
anova(mm4,mm3)
```

Data: Exam

Models:

```
mm3: normexam ~ standLRT + sex + (1 + standLRT | school)
```

```
mm2: normexam ~ standLRT + sex + schgend + (1 + standLRT | school)
```

| | Df | AIC | BIC | logLik | Chisq | Chi Df | Pr(>Chisq) |
|------------|-----------|---------------|---------------|----------------|---------------|---------------|----------------------|
| mm3 | 6 | 9299.4 | 9337.2 | -4643.7 | | | |
| mm2 | 8 | 9297.1 | 9347.6 | -4640.6 | 6.2671 | 2 | 0.04356 * |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

An alternative is using a simulation approach or adjusted F tests

Confidence intervals (SLOW)

`confint(mm2)` # uses the profile likelihood for each parameter

Computing profile confidence intervals ...

| | 2.5 % | 97.5 % |
|--------------|-------------|-------------|
| .sig01 | 0.23182865 | 0.34751250 |
| .sig02 | 0.27119698 | 0.81385819 |
| .sig03 | 0.08654243 | 0.16216781 |
| .sigma | 0.72564450 | 0.75846533 |
| (Intercept) | -0.12414665 | 0.08277617 |
| standLRT | 0.51411259 | 0.59385700 |
| sexM | -0.23458950 | -0.10193721 |
| schgendboys | -0.01768457 | 0.37720828 |
| schgendgirls | 0.01720522 | 0.33168112 |

`confint(mm2, method = "boot")` # uses parametric bootstrap = simulating the model

Computing bootstrap confidence intervals ...

| | 2.5 % | 97.5 % |
|---------------------------------|------------|----------------|
| sd_(Intercept) school | 0.22036809 | 0.32942345 |
| cor_standLRT.(Intercept) school | 0.33429582 | 0.85579285 |
| sd_standLRT school | 0.07844876 | 0.15286694 |
| sigma | 0.72528513 | 0.75794169 ... |

lme4 is the most commonly used library

lme4a is/was underway; see the online drafts by Bates.

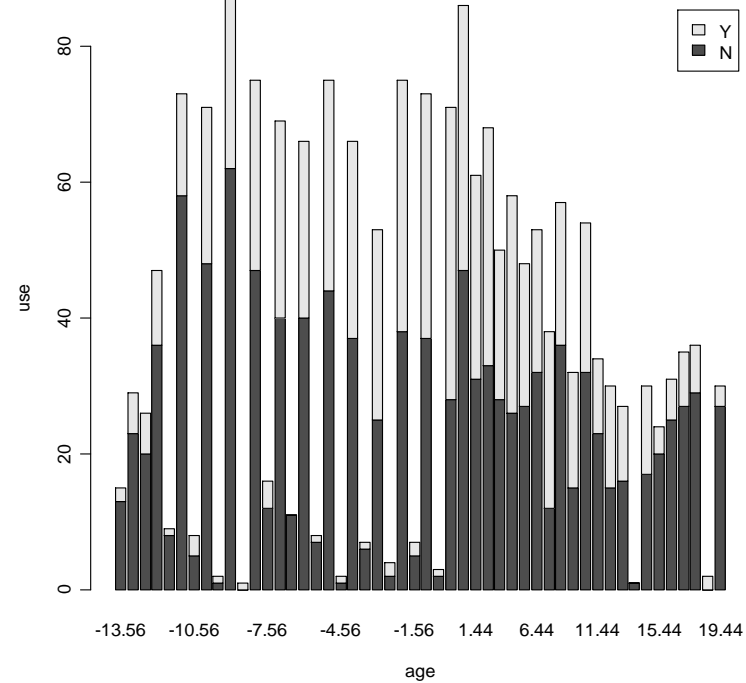
Everyone is getting better at mixed models...development of mixed modelling continues....but there are a limited number people doing the actual programming...which is often very model-specific....so many resort to bayesian model fitting for involved models.

Bates introduces new types of plots such as the zeta profile plot, to assess the appropriateness of models and inference based on them.

Diagnostics are increasingly better developed.

Generalized linear mixed models

Use of contraception in Bangladesh



data from 60 districts, urban/rural areas, # living children

Generalized linear mixed models

model specification as with mixed models, but now with non-normal families and non-identity link functions

inference more approximate, fewer exact results, read the wiki in as much detail as possible

next to random effect variances and covariances, often **no** dispersion parameter estimated

quasi-distributions recently abandoned

<http://glmm.wikidot.com/basic-glmm-simulation>

```
data(Contraception)
```

```
binmm<-glmer(use~age+urban+livch+(urban|district),  
family=binomial,data=Contraception, nAGQ = 1L)
```

```
summary(binmm)
```

Generalized linear mixed model fit by maximum likelihood ['glmerMod']

Family: binomial (logit)

Formula: use ~ age + urban + livch + (urban | district)

Data: Contraception

Random effects:

| Groups | Name | Variance | Std.Dev. | Corr |
|----------|-------------|----------|----------|-------|
| district | (Intercept) | 0.3812 | 0.6174 | |
| | urbanY | 0.6418 | 0.8011 | -0.80 |

Model fitting

The likelihood involves integration over the probability distributions of the random effects.

In glmm, several approximations for the likelihood:

PQL - AGC – Laplace

The "**PQL**" method is fastest but least accurate.

The "**Laplace**" method is intermediate in speed and accuracy (**nAGQ=1L**).

The "**AGQ**" method is the most accurate but can be considerably slower than the others.

Example Exercise

Contagious bovine pleuropneumonia (CBPP) dataset

serological incidence of CBPP in zebu cattle

15 commercial herds located in the Boji district of Ethiopia.

Blood samples were quarterly collected from all animals of these herds.

Some data are missing (lost to follow-up)

```
data(cbpps)
```

```
summary(cbpp)
```

```
help(cbpp)
```

```
m1 <- glmer(cbind(incidence, size - incidence) ~ period + (1 | herd),  
            family = binomial, data = cbpp)
```

```
# this is not necessarily the maximal model
```

Let's organize this inference procedure together

Quasi-likelihood in generalized (families with fixed dispersion) mixed models:

<http://glmm.wikidot.com/faq#overdispersion>

Fitting quasi- families in lmer is abandoned.
Now: fit a random effect per observation.

```
obs<-gl(length(cbpp$incidence),1) # group names  
mq1 <- glmer(cbind(incidence, size - incidence) ~ period + (1 | herd)  
+(1|obs), family = binomial, data = cbpp)
```

Examples in bioinformatics

either allow for very many mixed models to be fitted independently [or with shared initial values]

BIOINFORMATICS APPLICATIONS NOTE Vol. 30 no. 21 2014, pages 3134–3135
doi:10.1093/bioinformatics/btu435

Genetics and population analysis

Advance Access publication July 16, 2014

lrgpr: interactive linear mixed model analysis of genome-wide association studies with composite hypothesis testing and regression diagnostics in R

Gabriel E. Hoffman^{1,2,3,*}, Jason G. Mezey^{3,4} and Eric E. Schadt^{1,2}

In the help for glmApply():

"family gaussian() for a continuous response, and binomial() to fit a logit model for a binary response"

Other distributions than normal/gaussian for random effects?

Possible but little used

Some links:

Paper with nice overview of references in introduction:

www.jstatsoft.org/v56/c02/paper

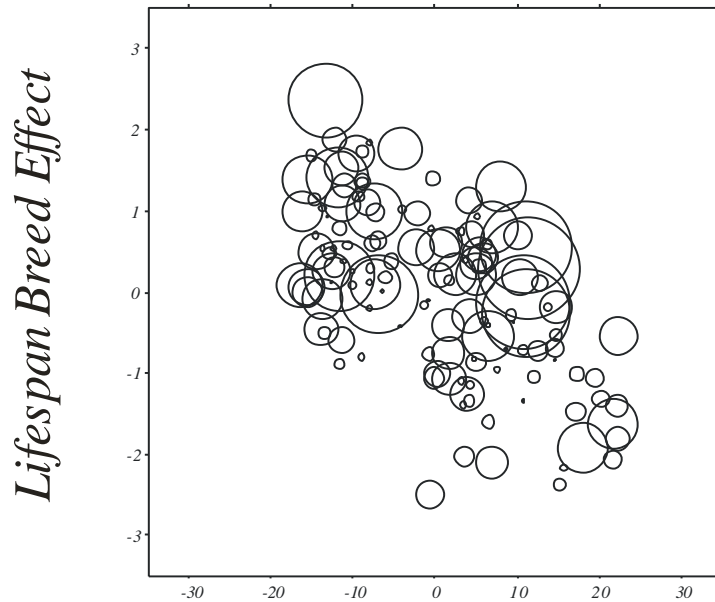
a package:

<http://cran.r-project.org/web/packages/spaMM/index.html>

Multivariate Mixed Models

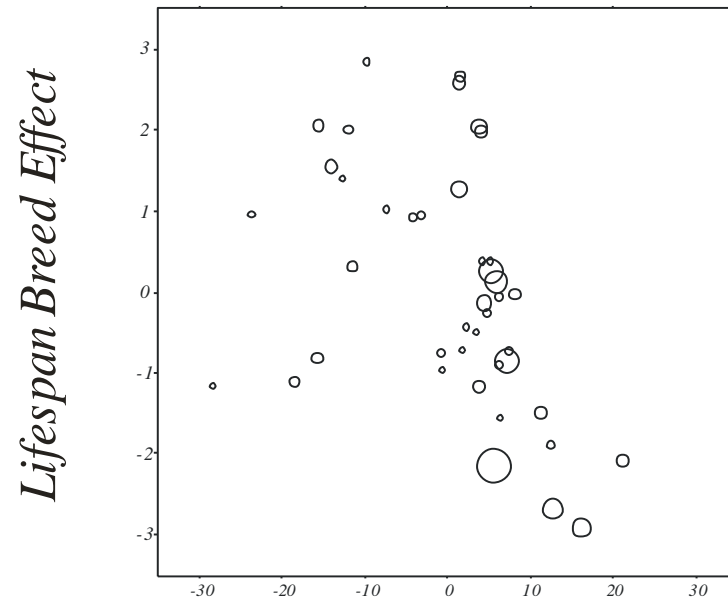
Use specialized software, here ASReml
or recode data with an extra variable "Trait"

VMDB dataset



Weight Breed Effect

NMBE dataset



Length Breed Effect

References

JC Pinheiro and DM Bates (2000) Mixed-Effects Models in S and S-PLUS. Springer Verlag.

<http://lme4.r-forge.r-project.org/lmmwR/lrgprt.pdf>

mlmRev.pdf

appendix-mixed-models.pdf [lme syntax, predecessor or lmer]

BM Bolker et al. (2009) Generalized linear mixed models: a practical  guide for ecology and evolution. TREE 24: 127-135.