An Introduction to Linear Mixed Models in R and SAS

SCC Seminar on Data Analysis

Dan Hall, Director of the SCC

Department of Statistics
Franklin College of Arts and Sciences

Statistical Consulting Center
UNIVERSITY OF GEORGIA

Table of Contents

Mixed-Effect Models

Mixed-effect models (aka, "mixed models") are like classical statistical models, but with some regression parameters ("fixed effects") replaced by "random effects".

- Random effects can be thought of as random regression coefficients describing the effects of explanatory factors or covariates.
- Typically, fixed effects and random effects are used in the same model, hence *mixed effects models.*

Why use random effects?

- They account for and quantify multiple sources of variability.
  - Useful for multiple levels of sampling/randomization, accounting for measurement error, etc.

- They broaden the scope of inference.

- Useful for correlated data.
  - Observations that share the same random effect(s) are correlated, according to the model. So mixed models useful for longitudinal/panel/repeated measures data, spatial data, etc.
  - Some multivariate models use shared random effects to account for correlation between different response variables.

Why use random effects?

- They account for and quantify multiple sources of variability.
  - Useful for multiple levels of sampling/randomization, accounting for measurement error, etc.

- They broaden the scope of inference.

- Useful for correlated data.
  - Observations that share the same random effect(s) are correlated, according to the model. So mixed models useful for longitudinal/panel/repeated measures data, spatial data, etc.
  - Some multivariate models use shared random effects to account for correlation between different response variables.

Why use random effects?

- They account for and quantify multiple sources of variability.
    - Useful for multiple levels of sampling/randomization, accounting for measurement error, etc.

- They broaden the scope of inference.

- Useful for correlated data.
    - Observations that share the same random effect(s) are correlated, according to the model. So mixed models useful for longitudinal/panel/repeated measures data, spatial data, etc.
    - Some multivariate models use shared random effects to account for correlation between different response variables.

Why use random effects?

- They account for and quantify multiple sources of variability.
    - Useful for multiple levels of sampling/randomization, accounting for measurement error, etc.

- They broaden the scope of inference.

- Useful for correlated data.
    - Observations that share the same random effect(s) are correlated, according to the model. So mixed models useful for longitudinal/panel/repeated measures data, spatial data, etc.
    - Some multivariate models use shared random effects to account for correlation between different response variables.

Why use random effects?

- They account for and quantify multiple sources of variability.
    - Useful for multiple levels of sampling/randomization, accounting for measurement error, etc.

- They broaden the scope of inference.

- Useful for correlated data.
    - Observations that share the same random effect(s) are correlated, according to the model. So mixed models useful for longitudinal/panel/repeated measures data, spatial data, etc.
    - Some multivariate models use shared random effects to account for correlation between different response variables.

Why use random effects?

- They account for and quantify multiple sources of variability.
  - Useful for multiple levels of sampling/randomization, accounting for measurement error, etc.

- They broaden the scope of inference.

- Useful for correlated data.
  - Observations that share the same random effect(s) are correlated, according to the model. So mixed models useful for longitudinal/panel/repeated measures data, spatial data, etc.
  - Some multivariate models use shared random effects to account for correlation between different response variables.

Why use random effects?

- Shrinkage.
  - Predictors for factor levels modeled with random effects are closer to the fitted population mean than those from purely fixed effect models. Such "shrinkage" predictors typically have good statistical properties.
  - Usefull for small area estimation, other problems where it is sensible to "borrow strength" from the population when predicting/estimating characteristics of an individual in that population.
- Splines. Mixed models are useful for fitting nonparametric and semiparametric models where covariate effects are modeled flexibly with penalized splines.

Why use random effects?

- Shrinkage.
  - Predictors for factor levels modeled with random effects are closer to the fitted population mean than those from purely fixed effect models. Such "shrinkage" predictors typically have good statistical properties.
  - Usefull for small area estimation, other problems where it is sensible to "borrow strength" from the population when predicting/estimating characteristics of an individual in that population.
- Splines. Mixed models are useful for fitting nonparametric and semiparametric models where covariate effects are modeled flexibly with penalized splines.

Why use random effects?

- Shrinkage.
  - Predictors for factor levels modeled with random effects are closer to the fitted population mean than those from purely fixed effect models. Such "shrinkage" predictors typically have good statistical properties.
  - Usefull for small area estimation, other problems where it is sensible to "borrow strength" from the population when predicting/estimating characteristics of an individual in that population.
- Splines. Mixed models are useful for fitting nonparametric and semiparametric models where covariate effects are modeled flexibly with penalized splines.

Why use random effects?

- Shrinkage.
  - Predictors for factor levels modeled with random effects are closer to the fitted population mean than those from purely fixed effect models. Such "shrinkage" predictors typically have good statistical properties.
  - Usefull for small area estimation, other problems where it is sensible to "borrow strength" from the population when predicting/estimating characteristics of an individual in that population.
- Splines. Mixed models are useful for fitting nonparametric and semiparametric models where covariate effects are modeled flexibly with penalized splines.

Applications of mixed models abound:

- Longitudinal/panel/repeated measures data (aka clustered data).
- Spatial data.
- Small area estimation applications involving data from complex, multi-stage sampling designs.
- Study designs involving blocking factors and stratification variables.
- Analysis of data subject to variability from multiple hierarchically organized sources of variability.
  - E.g., educational studies involving data from students within classes, within schools, within school districts, etc.
- Applications in every field: education, epidemiology, agriculture, ecology, business, sociology, forestry, ...

Applications of mixed models abound:

- Longitudinal/panel/repeated measures data (aka clustered data).
- Spatial data.
- Small area estimation applications involving data from complex, multi-stage sampling designs.
- Study designs involving blocking factors and stratification variables.
- Analysis of data subject to variability from multiple hierarchically organized sources of variability.
  - E.g., educational studies involving data from students within classes, within schools, within school districts, etc.

- Applications in every field: education, epidemiology, agriculture, ecology, business, sociology, forestry, ...

Applications of mixed models abound:

- Longitudinal/panel/repeated measures data (aka clustered data).
- Spatial data.
- Small area estimation applications involving data from complex, multi-stage sampling designs.
- Study designs involving blocking factors and stratification variables.
- Analysis of data subject to variability from multiple hierarchically organized sources of variability.
  - E.g., educational studies involving data from students within classes, within schools, within school districts, etc.
- Applications in every field: education, epidemiology, agriculture, ecology, business, sociology, forestry, ...

Applications of mixed models abound:

- Longitudinal/panel/repeated measures data (aka clustered data).
- Spatial data.
- Small area estimation applications involving data from complex, multi-stage sampling designs.
- Study designs involving blocking factors and stratification variables.
- Analysis of data subject to variability from multiple hierarchically organized sources of variability.
  - E.g., educational studies involving data from students within classes, within schools, within school districts, etc.
- Applications in every field: education, epidemiology, agriculture, ecology, business, sociology, forestry, ...

Applications of mixed models abound:

- Longitudinal/panel/repeated measures data (aka clustered data).
- Spatial data.
- Small area estimation applications involving data from complex, multi-stage sampling designs.
- Study designs involving blocking factors and stratification variables.
- Analysis of data subject to variability from multiple hierarchically organized sources of variability.
  - E.g., educational studies involving data from students within classes, within schools, within school districts, etc.
- Applications in every field: education, epidemiology, agriculture, ecology, business, sociology, forestry, ...

Applications of mixed models abound:

- Longitudinal/panel/repeated measures data (aka clustered data).
- Spatial data.
- Small area estimation applications involving data from complex, multi-stage sampling designs.
- Study designs involving blocking factors and stratification variables.
- Analysis of data subject to variability from multiple hierarchically organized sources of variability.
  - E.g., educational studies involving data from students within classes, within schools, within school districts, etc.
- Applications in every field: education, epidemiology, agriculture, ecology, business, sociology, forestry, ...

Applications of mixed models abound:

- Longitudinal/panel/repeated measures data (aka clustered data).
- Spatial data.
- Small area estimation applications involving data from complex, multi-stage sampling designs.
- Study designs involving blocking factors and stratification variables.
- Analysis of data subject to variability from multiple hierarchically organized sources of variability.
    - E.g., educational studies involving data from students within classes, within schools, within school districts, etc.
- Applications in every field: education, epidemiology, agriculture, ecology, business, sociology, forestry, ...

Different names:

- Hierarchical models.
- Multilevel models.
- Random coefficient models.
- Variance component models.

Different classes of mixed models:

- Linear mixed models.
- Generalized linear mixed models.
- Nonlinear mixed effect models.
- Frailty models.
- Joint models.

We will focus on linear mixed-effect models or LMMs.

One-way random effects model—Rails

In a study of the stress experienced by railway rails, 6 randomly chosen rails were tested 3 times each by measuring travel time for an ultrasonic wave through the rail.
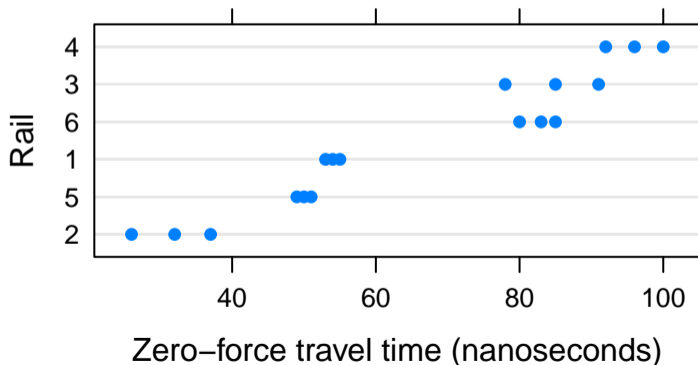
Rails

Clearly, these data are grouped, or clustered, by rail. This clustering has two closely related implications:

- Within-cluster correlation; and
- Between cluster heterogeneity.

These are really flip-sides of the same coin.

A Model for the Rails Data:

Let y_{ij} be the jth response from the ith rail. Assume

$$y_{ij} = \mu + b_i + e_{ij},$$

where

- $b_1, \ldots, b_6$ are independent random effects, where $b_i \sim N(0, \sigma_b^2)$,
- $e_{11} \ldots, e_{63}$ are independent random errors, where $e_{ij} \sim N(0, \sigma^2)$.
- and we assume the b_i's independent of the e_{ij}'s.

Model implies:

- $\operatorname{corr}(y_{ij}, y_{i'k}) = 0 \quad \Rightarrow \quad$ no b/w cluster correlation.
- $\operatorname{var}(y_{ij}) = \sigma_b^2 + \sigma^2 \quad \Rightarrow \quad$ total variance=sum of 2 variance components.
- $\operatorname{corr}(y_{ij}, y_{ik}) = \frac{\sigma_b^2}{\sigma_b^2 + \sigma^2} \equiv \rho \quad \Rightarrow \quad$ constant w/in cluster correlation.

A Model for the Rails Data:

Let y_{ij} be the jth response from the ith rail. Assume

$$y_{ij} = \mu + b_i + e_{ij},$$

where

- $b_1, \ldots, b_6$ are independent random effects, where $b_i \sim N(0, \sigma_b^2)$,
- $e_{11} \ldots, e_{63}$ are independent random errors, where $e_{ij} \sim N(0, \sigma^2)$.
- and we assume the b_i's independent of the e_{ij}'s.

Model implies:

- $\text{corr}(y_{ij}, y_{i'k}) = 0 \quad \Rightarrow \quad$ no b/w cluster correlation.
- $\text{var}(y_{ij}) = \sigma_b^2 + \sigma^2 \quad \Rightarrow \quad$ total variance=sum of 2 variance components.
- $\text{corr}(y_{ij}, y_{ik}) = \frac{\sigma_b^2}{\sigma_b^2 + \sigma^2} \equiv \rho \quad \Rightarrow \quad$ constant w/in cluster correlation.

A Model for the Rails Data:

Let y_{ij} be the jth response from the ith rail. Assume

$$y_{ij} = \mu + b_i + e_{ij},$$

where

- $b_1, \ldots, b_6$ are independent random effects, where $b_i \sim N(0, \sigma_b^2)$,
- $e_{11} \ldots, e_{63}$ are independent random errors, where $e_{ij} \sim N(0, \sigma^2)$.
- and we assume the b_i's independent of the e_{ij}'s.

Model implies:

- $\mathrm{corr}(y_{ij}, y_{i'k}) = 0 \quad \Rightarrow \quad$ no b/w cluster correlation.
- $\mathrm{var}(y_{ij}) = \sigma_b^2 + \sigma^2 \quad \Rightarrow \quad$ total variance=sum of 2 variance components.
- $\mathrm{corr}(y_{ij}, y_{ik}) = \frac{\sigma_b^2}{\sigma_b^2 + \sigma^2} \equiv \rho \quad \Rightarrow \quad$ constant w/in cluster correlation.

Fitting the Model in R with lme():

```r
library(nlme) # package with the lme() function
str(railData)
```

```
'data.frame':   18 obs. of  2 variables:
 $ Rail : Factor w/ 6 levels "1","2","3","4",..: 1 1 1 2 2 2 3 3 3 4 ...
 $ travel: num  55 53 54 26 37 32 78 91 85 92 ...
```

```r
(m1.Rails <- lme(travel~1,data=railData,random=~1|Rail))
```

```
Linear mixed-effects model fit by REML
  Data: railData
  Log-restricted-likelihood: -61.0885
  Fixed: travel ~ 1
(Intercept)
      66.5

Random effects:
 Formula: ~1 | Rail
         (Intercept) Residual
StdDev:    24.80547 4.020779

Number of Observations: 18
Number of Groups: 6
```

Some Results from the Fitted Model:

```
VarCorr(m1.Rails) # variance components
```

```
Rail = pdLogChol(1)
            Variance  StdDev
(Intercept) 615.31111 24.805465
Residual     16.16667  4.020779
```

```
emmeans(m1.Rails,specs=~1)  # from emmeans package: estimated marginal means (ls means)
```

```
 1       emmean   SE df lower.CL upper.CL
 overall   66.5 10.2  5     40.4     92.6
```

```
Degrees-of-freedom method: containment
Confidence level used: 0.95
```

Be careful!:

```
intervals(m1.Rails,which="fixed") # uses df=12, correct in fixed effect model, but not here
```

```
Approximate 95% confidence intervals

 Fixed effects:
              lower est.    upper
(Intercept) 44.33921 66.5 88.66079
attr(,"label")
[1] "Fixed effects:"
```

Notice that the big SE and wide CI for the mean respone. Compare to result from a fixed effect model—a one-way ANOVA model with fixed rail effects:

```
# Pop mean inference from random effects model:
emmeans(m1.Rails,specs=~1)


 1       emmean   SE df lower.CL upper.CL
 overall   66.5 10.2  5     40.4     92.6

Degrees-of-freedom method: containment
Confidence level used: 0.95
```

```
# Pop mean inference from fixed effects model:
m2.Rails <- lm(travel~Rail,data=railData); emmeans(m2.Rails,specs= ~1)


 1       emmean    SE df lower.CL upper.CL
 overall   66.5 0.948 12     64.4     68.6

Results are averaged over the levels of: Rail
Confidence level used: 0.95
```

Why such a big difference?

- Different scopes of inference!
- The mean in pop of all rails vs the mean for the 6 rails in the study.
- Our uncertainty is much greater when doing inference on the former quantity.

Notice that the big SE and wide CI for the mean respone. Compare to result from a fixed effect model—a one-way ANOVA model with fixed rail effects:

```
# Pop mean inference from random effects model:
emmeans(m1.Rails,specs=~1)


 1       emmean  SE df lower.CL upper.CL
 overall   66.5 10.2  5    40.4     92.6

Degrees-of-freedom method: containment
Confidence level used: 0.95
```

```
# Pop mean inference from fixed effects model:
m2.Rails <- lm(travel~Rail,data=railData); emmeans(m2.Rails,specs= ~1)


 1       emmean    SE df lower.CL upper.CL
 overall   66.5 0.948 12    64.4     68.6

Results are averaged over the levels of: Rail
Confidence level used: 0.95
```

Why such a big difference?

- Different scopes of inference!
- The mean in pop of all rails vs the mean for the 6 rails in the study.
- Our uncertainty is much greater when doing inference on the former quantity.

Notice that the big SE and wide CI for the mean respone. Compare to result from a fixed effect model—a one-way ANOVA model with fixed rail effects:

```
# Pop mean inference from random effects model:
emmeans(m1.Rails,specs=~1)


 1        emmean   SE df lower.CL upper.CL
 overall   66.5 10.2  5     40.4     92.6

Degrees-of-freedom method: containment
Confidence level used: 0.95
```

```
# Pop mean inference from fixed effects model:
m2.Rails <- lm(travel~Rail,data=railData); emmeans(m2.Rails,specs= ~1)


 1        emmean    SE df lower.CL upper.CL
 overall   66.5 0.948 12     64.4     68.6

Results are averaged over the levels of: Rail
Confidence level used: 0.95
```

Why such a big difference?

- Different scopes of inference!
- The mean in pop of all rails vs the mean for the 6 rails in the study.
- Our uncertainty is much greater when doing inference on the former quantity.

More Results from the Fitted Model:

Recall:
- Model implies constant variance (equal to between-rail + within-rail variances).
- Model implies responses are uncorrelated across different rails, but implies (constant) correlation for repeated responses on the same rail.

```
getVarCov(m1.Rails,individuals=c("1","2"),type="marginal")
```

```
Rail 1
Marginal variance covariance matrix
        1      2      3
1 631.48 615.31 615.31
2 615.31 631.48 615.31
3 615.31 615.31 631.48
  Standard Deviations: 25.129 25.129 25.129
Rail 2
Marginal variance covariance matrix
        1      2      3
1 631.48 615.31 615.31
2 615.31 631.48 615.31
3 615.31 615.31 631.48
  Standard Deviations: 25.129 25.129 25.129
```

This covariance structure is called **compound symmetry**. Here, $\hat{\rho} = \frac{615.31}{631.48} = .97$.

An Alternative Model for the Rails Data:

Assume

$$y_{ij} = \mu + e_{ij},$$

where

- $e_1 \ldots, e_6$ are independent random error vectors, where
  $$e_i = \begin{pmatrix} e_{i1} \\ e_{i2} \\ e_{i3} \end{pmatrix} \sim N_3(\mathbf{0}, \mathbf{\Sigma}).$$
- and we assume $\mathbf{\Sigma}$ is compound symmetric.

Fitting the model in R

This model can be fitted with gls(), also from the nlme package:

```
m1a.Rails <- gls(travel~1,data=railData,corr=corCompSymm(form=~1|Rail))
logLik(m1.Rails);logLik(m1a.Rails) # maximized objective functions (restricted loglikelihoods)
```

'log Lik.' -61.0885 (df=3)

'log Lik.' -61.0885 (df=3)

They are the same model!

Actually, not quite.

- The random effect model implies a model of compound symmetry, but not vice versa.
- The latter model admits negative within-cluster correlation (rare), which the former does not.

Random Effects or Error Covariance?

The previous example illustrates that correlation can be built into a model through

- the inclusion of random effects;
- the assumption of an error covariance structure;
- or both (but not redundantly, which induces non-identifiability).

General form of LMM for clustered data:

$$y_{ij} = x_{1ij}\beta_1 + x_{2ij}\beta_2 + \cdots + x_{pij}\beta_p + z_{1ij}b_{1i} + \cdots + z_{qij}b_{qi} + e_{ij}$$

for i indexes clusters; j indexes obs w/in clusters.

Equivalently, writing with vectors:

$$\mathbf{y}_i = \mathbf{x}_i'\beta + \mathbf{z}_i'\mathbf{b}_i + \mathbf{e}_i$$

where

$$\{\mathbf{b_i}\} \overset{ind}{\sim} N(\mathbf{0}, \mathbf{D}), \quad \{\mathbf{e_i}\} \overset{ind}{\sim} N(\mathbf{0}, \mathbf{R}_i) \quad \Rightarrow \quad \mathrm{var}(\mathbf{y}_i) = \mathbf{z}_i'\mathbf{D}\mathbf{z}_i + \mathbf{R}_i \equiv \mathbf{\Sigma}_i.$$

More Examples:

- RCBD Model: $y_{ij} = \mu + \alpha_j + b_i + e_{ij}$, i for blocks, j for treatments.

- Split Plot Model:

$$y_{hij} = \mu + \alpha_h + \beta_j + (\alpha\beta)_{hj} + b_{i(h)} + e_{hij},$$

  h for whole plot factor, i for whole plots, j for split plot factor.

- Repeated Measures ANOVA Model:

$$y_{hij} = \mu + \alpha_h + \beta_j + (\alpha\beta)_{hj} + b_{i(h)} + e_{hij}$$

  h for b/w subjects factor, i for subjects, j for time factor.

  - Classical RM-ANOVA: assumes $(e_{hij}) \overset{ind}{\sim} N(0, \sigma^2)$ (same as split plot model) and "fixes" the analysis for non-compound symmetry.

  - Modern approach: assume a more suitable covariance structure in the specification of $b_{i(h)} + e_{hij}$, so no need for ad hoc correction.

More Examples:

- RCBD Model: $y_{ij} = \mu + \alpha_j + b_i + e_{ij}$, i for blocks, j for treatments.

- Split Plot Model:

$$y_{hij} = \mu + \alpha_h + \beta_j + (\alpha\beta)_{hj} + b_{i(h)} + e_{hij},$$

  h for whole plot factor, i for whole plots, j for split plot factor.

- Repeated Measures ANOVA Model:

$$y_{hij} = \mu + \alpha_h + \beta_j + (\alpha\beta)_{hj} + b_{i(h)} + e_{hij}$$

  h for b/w subjects factor, i for subjects, j for time factor.

  - Clasical RM-ANOVA: assumes $(e_{hij}) \overset{ind}{\sim} N(0, \sigma^2)$ (same as split plot model) and "fixes" the analysis for non-compound symmetry.

  - Modern approach: assume a more suitable covariance structure in the specification of $b_{i(h)} + e_{hij}$, so no need for ad hoc correction.

More Examples:

- RCBD Model: $y_{ij} = \mu + \alpha_j + b_i + e_{ij}$, i for blocks, j for treatments.

- Split Plot Model:

$$y_{hij} = \mu + \alpha_h + \beta_j + (\alpha\beta)_{hj} + b_{i(h)} + e_{hij},$$

  h for whole plot factor, i for whole plots, j for split plot factor.

- Repeated Measures ANOVA Model:

$$y_{hij} = \mu + \alpha_h + \beta_j + (\alpha\beta)_{hj} + b_{i(h)} + e_{hij}$$

  h for b/w subjects factor, i for subjects, j for time factor.

  - Clasical RM-ANOVA: assumes $\{e_{hij}\} \overset{ind}{\sim} N(0, \sigma^2)$ (same as split plot model) and "fixes" the analysis for non-compound symmetry.
  - Modern approach: assume a more suitable covariance structure in the specification of $b_{i(h)} + e_{hij}$, so no need for ad hoc correction.

More Examples:

- RCBD Model: $y_{ij} = \mu + \alpha_j + b_i + e_{ij}$, i for blocks, j for treatments.

- Split Plot Model:

$$y_{hij} = \mu + \alpha_h + \beta_j + (\alpha\beta)_{hj} + b_{i(h)} + e_{hij},$$

  h for whole plot factor, i for whole plots, j for split plot factor.

- Repeated Measures ANOVA Model:

$$y_{hij} = \mu + \alpha_h + \beta_j + (\alpha\beta)_{hj} + b_{i(h)} + e_{hij}$$

  h for b/w subjects factor, i for subjects, j for time factor.

  - Clasical RM-ANOVA: assumes $\{e_{hij}\} \overset{ind}{\sim} N(0, \sigma^2)$ (same as split plot model) and "fixes" the analysis for non-compound symmetry.
  - Modern approach: assume a more suitable covariance structure in the specification of $b_{i(h)} + e_{hij}$, so no need for ad hoc correction.

More Examples:

- RCBD Model: $y_{ij} = \mu + \alpha_j + b_i + e_{ij}$, i for blocks, j for treatments.

- Split Plot Model:

$$y_{hij} = \mu + \alpha_h + \beta_j + (\alpha\beta)_{hj} + b_{i(h)} + e_{hij},$$

  h for whole plot factor, i for whole plots, j for split plot factor.

- Repeated Measures ANOVA Model:

$$y_{hij} = \mu + \alpha_h + \beta_j + (\alpha\beta)_{hj} + b_{i(h)} + e_{hij}$$

  h for b/w subjects factor, i for subjects, j for time factor.

  - Clasical RM-ANOVA: assumes $\{e_{hij}\} \overset{ind}{\sim} N(0, \sigma^2)$ (same as split plot model) and "fixes" the analysis for non-compound symmetry.
  - Modern approach: assume a more suitable covariance structure in the specification of $b_{i(h)} + e_{hij}$, so no need for ad hoc correction.

More Examples:

Many LMMs are suitable for clustered data with cluster-specific random effects, but not all.

- Latin Square Model (crossed random effects):

$$y_{ijk} = \mu + \alpha_k + r_i + c_j + e_{ijk}$$

i for row factor, j for column factor, k for treatment factor.

- Nested Random Effects: We select three leading grocery chains, four cities to study, sample 5 stores per chain within each city, and sample 10 bags of grapes per store. Response is proportion of soft or rotted grapes per bag.

$$y_{hijk} = \mu + \alpha_h + c_i + s_{j(hi)} + e_{hijk}$$

h for chains, i for cities, j for stores, k for bags.

- c_i and $s_{j(hi)}$ are city and store-within-city random effects.

More Examples:

Many LMMs are suitable for clustered data with cluster-specific random effects, but not all.

- Latin Square Model (crossed random effects):

$$y_{ijk} = \mu + \alpha_k + r_i + c_j + e_{ijk}$$

i for row factor, j for column factor, k for treatment factor.

- Nested Random Effects: We select three leading grocery chains, four cities to study, sample 5 stores per chain within each city, and sample 10 bags of grapes per store. Response is proportion of soft or rotted grapes per bag.

$$y_{hijk} = \mu + \alpha_h + c_i + s_{j(hi)} + e_{hijk}$$

h for chains, i for cities, j for stores, k for bags.

  - c_i and $s_{j(hi)}$ are city and store-within-city random effects.

More Examples:

Many LMMs are suitable for clustered data with cluster-specific random effects, but not all.

- Latin Square Model (crossed random effects):

$$y_{ijk} = \mu + \alpha_k + r_i + c_j + e_{ijk}$$

  i for row factor, j for column factor, k for treatment factor.

- Nested Random Effects: We select three leading grocery chains, four cities to study, sample 5 stores per chain within each city, and sample 10 bags of grapes per store. Response is proportion of soft or rotted grapes per bag.

$$y_{hijk} = \mu + \alpha_h + c_i + s_{j(hi)} + e_{hijk}$$

  h for chains, i for cities, j for stores, k for bags.
    - c_i and $s_{j(hi)}$ are city and store-within-city random effects.

More Examples:

- Penalized Quadratic Spline Model (random effects at the observation level):

$$y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^2 + b_1 [\max(x_i - \kappa_1, 0)]^2 + \cdots + b_q [\max(x_i - \kappa_q, 0)]^2 + e_i$$

$\kappa_1, ..., \kappa_q$ are knots; $b_1, ..., b_q \overset{ind}{\sim} N(0, \sigma_b^2)$.

General LMM:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{b} + \mathbf{e}.$$

where

$$\mathbf{b} \sim N(\mathbf{0}, \mathbf{D}), \quad \mathbf{e} \sim N(\mathbf{0}, \mathbf{R}) \quad \Rightarrow \quad \mathrm{var}(\mathbf{y}) = \mathbf{Z}\mathbf{D}\mathbf{Z}' + \mathbf{R} \equiv \mathbf{\Sigma}.$$

Software

Many software packages implement LMMs, but I primarily use SAS and R. The main tools for LMMs in these packages are

- PROCs MIXED and HPMIXED in SAS
- the lme() function and others in the nlme package for R.
- the lmer() function and others in the lme4 package for R.

Feature	MIXED	HPMIXED	lme	lmer
Non-nested random effects	Yes	Yes	No	Yes
Non-spherical error covariance	Yes	No	Yes	No
Heterogeneous error covariance	Yes	No	No	No
Kenward-Roger inference	Yes	No	No	Yes
Handles large problems	No	Yes	No	Yes

Software

Other Software Notes:

- PROC GLIMMIX in SAS handles generalized linear mixed models (GLMMS). But also useful for LMMs because of tools that MIXED lacks (e.g., for penalized splines).
- PROC HPLMIXED has similar capabilities to PROC MIXED, but runs in distributed and multicore computing environments for greater speed.
- NLMMs but not GLMMs implemented in nlme package for R.
- GLMMs and NLMMs implemented in lme4 package for R.
- Routines in lme4 and HPMIXED use sparse matrix techniques and other computationally efficient methods to efficiently handle big problems.

Software

Other Software Notes:

- PROC GLIMMIX in SAS handles generalized linear mixed models (GLMMS). But also useful for LMMs because of tools that MIXED lacks (e.g., for penalized splines).
- PROC HPLMIXED has similar capabilities to PROC MIXED, but runs in distributed and multicore computing environments for greater speed.
- NLMMs but not GLMMs implemented in nlme package for R.
- GLMMs and NLMMs implemented in lme4 package for R.
- Routines in lme4 and HPMIXED use sparse matrix techniques and other computationally efficient methods to efficiently handle big problems.

Software

Other Software Notes:

- PROC GLIMMIX in SAS handles generalized linear mixed models (GLMMS). But also useful for LMMs because of tools that MIXED lacks (e.g., for penalized splines).
- PROC HPLMIXED has similar capabilities to PROC MIXED, but runs in distributed and multicore computing environments for greater speed.
- NLMMs but not GLMMs implemented in nlme package for R.
- GLMMs and NLMMs implemented in lme4 package for R.
- Routines in lme4 and HPMIXED use sparse matrix techniques and other computationally efficient methods to efficiently handle big problems.

Software

Other Software Notes:

- PROC GLIMMIX in SAS handles generalized linear mixed models (GLMMS). But also useful for LMMs because of tools that MIXED lacks (e.g., for penalized splines).
- PROC HPLMIXED has similar capabilities to PROC MIXED, but runs in distributed and multicore computing environments for greater speed.
- NLMMs but not GLMMs implemented in `nlme` package for R.
- GLMMs and NLMMs implemented in `lme4` package for R.
- Routines in `lme4` and HPMIXED use sparse matrix techniques and other computationally efficient methods to efficiently handle big problems.

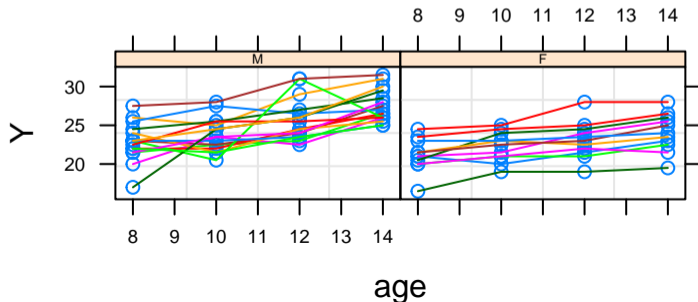# Software

Other Software Notes:

- PROC GLIMMIX in SAS handles generalized linear mixed models (GLMMS). But also useful for LMMs because of tools that MIXED lacks (e.g., for penalized splines).
- PROC HPLMIXED has similar capabilities to PROC MIXED, but runs in distributed and multicore computing environments for greater speed.
- NLMMs but not GLMMs implemented in nlme package for R.
- GLMMs and NLMMs implemented in lme4 package for R.
- Routines in lme4 and HPMIXED use sparse matrix techniques and other computationally efficient methods to efficiently handle big problems.

Orthodontic Growth Data

Data below are skull measurements taken every 2 years for 11 females, 16 males.
Y = distance between pituitary and the pterygomaxillary fissure (mm).

```
Grouped Data: Y ~ age | id
      id age    Y sexFac male fem
1.8   1   8 21.0     F     0   1
1.10  1  10 20.0     F     0   1
1.12  1  12 21.5     F     0   1
1.14  1  14 23.0     F     0   1
2.8   2   8 21.0     F     0   1
```

Orthodontic Growth Data

Random intercept model.

$$y_{hij} = (\alpha_h + b_{hi}) + \beta_h \text{age}_{hij} + e_{hij}$$

where h for sex, i for subject, j for time point, and

$$\{b_{hi}\} \overset{ind}{\sim} N(0, \psi_h), \quad \{e_{hij}\} \overset{ind}{\sim} N(0, \sigma_h^2).$$

- Model mean is linear in time with different intercepts and slopes for each sex.
- Intercepts vary randomly by subject, with different variability for each sex.
- Error variance differs by sex.

`lme()`:

```
m1.lme <-
  lme(Y ~ 0+fem+male+fem:age+male:age, # formula for fixed effects
      data=orthData.grpd, # data frame to use (convenient if a groupedData object)
      method="REML", # the default and recommended fitting method
      random= list(id=pdDiag(form= ~ -1+fem+male)), # random effects specification
      weights = varIdent(form = ~ 1|sexFac), # specifies error heteroscedasticity model
      corr=NULL  # specifies error correlation model (not used in this model)
      )
```

Orthodontic Growth Data

Random intercept model.

$$y_{hij} = (\alpha_h + b_{hi}) + \beta_h \text{age}_{hij} + e_{hij}$$

where h for sex, i for subject, j for time point, and

$$\{b_{hi}\} \overset{ind}{\sim} N(0, \psi_h), \quad \{e_{hij}\} \overset{ind}{\sim} N(0, \sigma_h^2).$$

- Model mean is linear in time with different intercepts and slopes for each sex.
- Intercepts vary randomly by subject, with different variability for each sex.
- Error variance differs by sex.

`lme()`:

```
m1.lme <-
  lme(Y ~ 0+fem+male+fem:age+male:age, # formula for fixed effects
      data=orthData.grpd, # data frame to use (convenient if a groupedData object)
      method="REML", # the default and recommended fitting method
      random= list(id=pdDiag(form= ~ -1+fem+male)), # random effects specification
      weights = varIdent(form = ~ 1|sexFac), # specifies error heteroscedasticity model
      corr=NULL  # specifies error correlation model (not used in this model)
      )
```

Orthodontic Growth Data

Random intercept model.

$$y_{hij} = (\alpha_h + b_{hi}) + \beta_h \text{age}_{hij} + e_{hij}$$

where h for sex, i for subject, j for time point, and

$$\{b_{hi}\} \overset{ind}{\sim} N(0, \psi_h), \quad \{e_{hij}\} \overset{ind}{\sim} N(0, \sigma_h^2).$$

- Model mean is linear in time with different intercepts and slopes for each sex.
- Intercepts vary randomly by subject, with different variability for each sex.
- Error variance differs by sex.

`lme()`:

```
m1.lme <-
  lme(Y ~ 0+fem+male+fem:age+male:age, # formula for fixed effects
      data=orthData.grpd, # data frame to use (convenient if a groupedData object)
      method="REML", # the default and recommended fitting method
      random= list(id=pdDiag(form= ~ -1+fem+male)), # random effects specification
      weights = varIdent(form = ~ 1|sexFac), # specifies error heteroscedasticity model
      corr=NULL  # specifies error correlation model (not used in this model)
      )
```

`lmer()`:

- `lmer()` only handles errors that are i.i.d. $N(0, \sigma^2)$ (uncorrelated, homoscedastic).
- So syntax below is for simpler model with constant error variance.

```
m2.lmer <-
  lmer(Y ~ 0+fem+fem:age+male+male:age+ # single formula specifies fixed...
        (0+fem|id)+(0+male|id),         # and random effects
       REML=TRUE, # REML fitting method is the default
       data=dentT.grpd) # data frame with model variables (need not be a groupedData object)
)
```

- Random effects in parentheses with a bar and grouping factor specifying the level at which they operate.
  - (0+fem|id) specifies subject-level random intercepts for females
  - (0+male|id) specifies subject-level random intercepts for males
  - (1+age|id) would specify subject-level random intercepts and slopes for each subject. By this syntax they would be assumed to by correlated.

`lmer()`:

- `lmer()` only handles errors that are i.i.d. $N(0, \sigma^2)$ (uncorrelated, homoscedastic).
- So syntax below is for simpler model with constant error variance.

```
m2.lmer <-
  lmer(Y ~ 0+fem+fem:age+male+male:age+ # single formula specifies fixed...
        (0+fem|id)+(0+male|id),          # and random effects
      REML=TRUE, # REML fitting method is the default
      data=dentT.grpd) # data frame with model variables (need not be a groupedData object)
)
```

- Random effects in parentheses with a bar and grouping factor specifying the level at which they operate.
  - (0+fem|id) specifies subject-level random intercepts for females
  - (0+male|id) specifies subject-level random intercepts for males
  - (1+age|id) would specify subject-level random intercepts and slopes for each subject. By this syntax they would be assumed to by correlated.

`lmer()`:

- `lmer()` only handles errors that are i.i.d. $N(0, \sigma^2)$ (uncorrelated, homoscedastic).
- So syntax below is for simpler model with constant error variance.

```
m2.lmer <-
  lmer(Y ~ 0+fem+fem:age+male+male:age+ # single formula specifies fixed...
       (0+fem|id)+(0+male|id),          # and random effects
     REML=TRUE, # REML fitting method is the default
     data=dentT.grpd) # data frame with model variables (need not be a groupedData object)
)
```

- Random effects in parentheses with a bar and grouping factor specifying the level at which they operate.
  - (0+fem|id) specifies subject-level random intercepts for females
  - (0+male|id) specifies subject-level random intercepts for males
  - (1+age|id) would specify subject-level random intercepts and slopes for each subject. By this syntax they would be assumed to by correlated.

`lmer()`:

- `lmer()` only handles errors that are i.i.d. $N(0, \sigma^2)$ (uncorrelated, homoscedastic).
- So syntax below is for simpler model with constant error variance.

```
m2.lmer <-
  lmer(Y ~ 0+fem+fem:age+male+male:age+ # single formula specifies fixed...
       (0+fem|id)+(0+male|id),         # and random effects
       REML=TRUE, # REML fitting method is the default
       data=dentT.grpd) # data frame with model variables (need not be a groupedData object)
)
```

- Random effects in parentheses with a bar and grouping factor specifying the level at which they operate.
  - `(0+fem|id)` specifies subject-level random intercepts for females
  - `(0+male|id)` specifies subject-level random intercepts for males
  - `(1+age|id)` would specify subject-level random intercepts and slopes for each subject. By this syntax they would be assumed to by correlated.

PROC MIXED:

- MODEL statement specifies fixed effects in the model.
- RANDOM specifies random effects and their assumed covariance structure.
- REPEATED statement specifies the error covariance structure for the model.

```
proc mixed data=dentT method=reml;    /* REML is the default */
  class id;                            /* specifies the factors */
  model y= fem male fem*age male*age   /* fixed effect specification */
    / ddfm=kr;                         /* this option requests Kenward-Roger inference
                                          for fixed effects (use it!) */
  random intercept / subject=id        /* specifies random effects structure */
                     group=sex;        /* allows random effect cov parms to differ by group*/
  repeated / group=sex;                /* allows error cov parms to differ by group*/
run;
```

or

```
proc mixed data=dentT method=reml;
  class id time;
  model y= fem male fem*age male*age / ddfm=kr;
  repeated time / subject=id /* the clustering structure, order of obs w/in a cluster */
                  type=cs    /* specifies compound symmetric covariance in each cluster */
                  group=sex; /* allows error cov parms to differ by group */
run;
```

PROC MIXED:

- MODEL statement specifies fixed effects in the model.
- RANDOM specifies random effects and their assumed covariance structure.
- REPEATED statement specifies the error covariance structure for the model.

```
proc mixed data=dentT method=reml;    /* REML is the default */
  class id;                           /* specifies the factors */
  model y= fem male fem*age male*age  /* fixed effect specification */
    / ddfm=kr;                        /* this option requests Kenward-Roger inference
                                         for fixed effects (use it!) */
  random intercept / subject=id       /* specifies random effects structure */
                     group=sex;       /* allows random effect cov parms to differ by group*/
  repeated / group=sex;               /* allows error cov parms to differ by group*/
run;
```

or

```
proc mixed data=dentT method=reml;
  class id time;
  model y= fem male fem*age male*age / ddfm=kr;
  repeated time / subject=id /* the clustering structure, order of obs w/in a cluster */
                  type=cs    /* specifies compound symmetric covariance in each cluster */
                  group=sex; /* allows error cov parms to differ by group */
run;
```

PROC MIXED:

- MODEL statement specifies fixed effects in the model.
- RANDOM specifies random effects and their assumed covariance structure.
- REPEATED statement specifies the error covariance structure for the model.

```sas
proc mixed data=dentT method=reml;      /* REML is the default */
  class id;                             /* specifies the factors */
  model y= fem male fem*age male*age    /* fixed effect specification */
    / ddfm=kr;                          /* this option requests Kenward-Roger inference
                                           for fixed effects (use it!) */
  random intercept / subject=id         /* specifies random effects structure */
                     group=sex;         /* allows random effect cov parms to differ by group*/
  repeated / group=sex;                 /* allows error cov parms to differ by group*/
run;
```

or

```sas
proc mixed data=dentT method=reml;
  class id time;
  model y= fem male fem*age male*age / ddfm=kr;
  repeated time / subject=id /* the clustering structure, order of obs w/in a cluster */
               type=cs    /* specifies compound symmetric covariance in each cluster */
               group=sex; /* allows error cov parms to differ by group */
run;
```

Fitted Growth Curves

Estimated mean (fixed) and predicted subject-specific (random) growth curves for four female subjects (IDs 1–4) and 4 males (IDs 12–15):
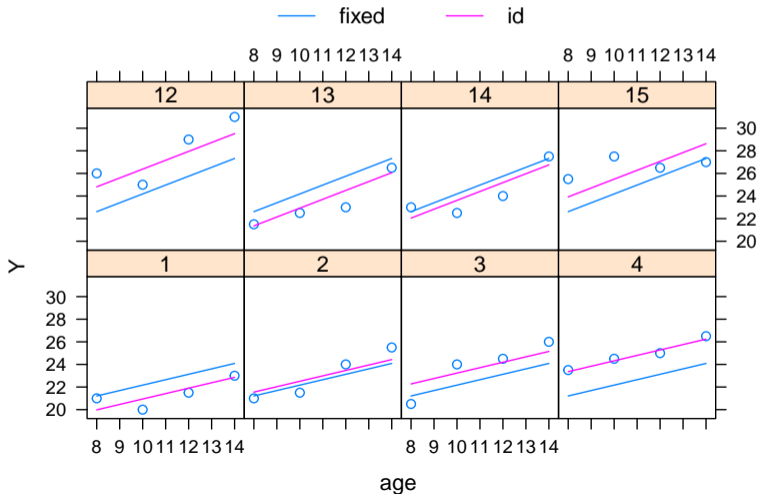
References—Good Applied Books on LMMs:

Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge University Press.

Pinheiro, J.C. and Bates, D.M. (2000). *Mixed-effects Models in S and S-PLUS.* Springer.

Rabe-Hesketh, S. and Skrondal, A. (2012). *Multilevel and Longitudinal Modeling Using Stata, Volumnes I and II, Third Edition.* Stata Press.

Stroup, W.W., Milliken, G.A., Claasen, E.A., and Wolfinger, R.D. (2018). *SAS for Mixed Models, Introduction and Basic Applications.* SAS Institute Inc.

West, B.T., Welch, K.B., Galecki, A.T. (2015). *Linear Mixed Models: A Practical Guide Using Statistical Software, Second Edition.* CRC Press.

Thank You!

If you need help using mixed effect models in your research, contact the SCC!
`www.stat.uga/consulting`