GLMM workshop 7 July 2016

Instructors: David Schneider, with Louis Charron, Devin Flawd, Kyle Millar, Anne St. Pierre Provencher, Sam Trueman

First session 1 PM Room SN2109 Writing the model
Break
Second session 2 PM SN 2018/2025 F-ratios from Expected Mean Squares
Break
Third session 3:30 SN 2067/2071 Executing the analysis

Goal of the first session – Writing Statistical Models

GLM	The General Linear Model	Fixed Effects + Normal Error
GzLM	The Generalized Linear model	Fixed Effects + Non-normal Errors
GLMM	The General Linear Mixed Model	Fixed + Random + Normal
GzLMM	The Generalized Linear Mixed Model	Fixed + Random + Non-normal

Goal of the second session - Writing out the expected mean squares
Forming unambiguous likelihood ratio tests (F, t, χ^2)

Goal of the third session - Executing a GLMM in a statistical package
Interpreting the output

Review of fixed effect, random effect, and mixed models.
Write the model and construct the ANOVA table.
<u>Step 1.</u> Assign symbols to the response variable and explanatory variables and write the model (Session 1)
<u>Step 2.</u> Identify each explanatory variable as Factor or Covarate (Session 1)
<u>Step 3.</u> For explanatory variable and each interaction term, identify as Random or Fixed. (Session 1)
<u>Step 4.</u> For explanatory variable and each interaction term, identify as Crossed or Nested. (Session 1)
Step 5. Rewrite the model as an ANOVA table, showing results of Steps 2, 3, and 4.

Here is an example. We have data on response of plants to CO_2 at three levels (none, medium, high). The response variable is plant size at 5 successive times to produce an estimate of growth rate. The model term of interest is Trt× Time - does growth rate depend on Trt? The experiment is repeated at 6 locations taken to be representative of a larger area. The full model has three main effects, three pairwise interaction terms, and a three way interaction. .

Note that Fixed × Fixed = Fixed
Fixed × Random = Mixed
Random × Random = Random

Time is a covariate – we expect monotonic response.

*In this example TIME was not measured at the same intervals in each Block, so TIME×Block fails the cross test.

TIME×Block cannot be estimated.
However, this mixed term is still there, lurking.
We cannot assume it is small, with little effect.
To cancel out its effect we need to keep track of it.

To do this we use the "random within" rule.

Source	df	Factor or Covariate	Random or Fixed	Crossed or Nested
Trt	3-1	Factor	Fixed	NA
TIME (n=5)	1	Cov	Fixed	NA
Block	6-1	Factor	Random	NA
Trt×TIME	2×1		Fixed	Crossed
Trt×Block	2×5		Mixed	Crossed
TIME×Block	1×5		Mixed	Nested*
Trt×Time×Bl	2×1×5		Mixed	Nested
Model	35			
Residual	89-35			
Total	(3×5×6)-1			

TIME×Block: Block(Time) is random within. Time(Block) is not random within.
We assign TIME×Block to Block(Time)

Second session 2 PM SN 2018/2025 F-ratios from Expected Mean Squares

Write the model and construct the ANOVA table.

Step 1. Assign symbols to the response variable and explanatory variables and write the model (Session 1)
Step 2. Identify each explanatory variable as Factor or Covarate (Session 1)
Step 3. For explanatory variable and each interactive effect term, identify as Random or Fixed. (Session 1)
Step 4. For explanatory variable and each interactive effect term, identify as Crossed or Nested. (Session 1)
Step 5. Rewrite the model as an ANOVA table, showing results of Steps 2, 3, and 4.
Step 6. Revise ANOVA table for nested terms.

Source	df
Trt	3-1
TIME (n=5)	1
Block	6-1
Trt×TIME	2×1
Trt×Block	2×5
TIME×Block	1×5
Trt×Time×Bl	2×1×5
Model	35
Residual	89-35
Total	(3×5×6)-1

Source	df
Trt	3-1
TIME (n=5)	1
TIME×Block	1×5
Block	6-1
Trt×TIME	2×1
Trt×Block	2×5
Trt×Time×Bl	2×1×5
Model	35
Residual	89-35
Total	(3×5×6)-1

Source	df
Trt	3-1
TIME (n=5)	1
Block(Time)	1×5+6-1
Trt×TIME	2×1
Trt×Block	2×5
Trt×Time×Bl	2×1×5
Model	35
Residual	89-35
Total	(3×5×6)-1

12. Fisher (1925 *Statistical Methods for Research Workers*) introduced the latin square design, which controls for two sources of random variation, rows and columns in a square array in field experiment.

"The following root weights for mangolds were found by Mercer and Hall in 25 plots…
Then out of the 24 degrees of freedom…. 12 will remain for the estimation of error.
The 12 will provide an unbiased estimate of the errors in the comparison of treatments, provided that every pair of plots, not in the same row or column, belong equally frequently to the same treatment.."
Here is the ANOVA table as reported by Fisher (p267 in 13th ed, 1958, Table 61)

Note that all three factors can be crossed but there are insufficient df to estimate any of the interaction terms.

	df	SS	MS	sd
Columns	4	701.84		
Rows	4	4240.24		
Trt	4	330.24	130.29	11.41
Remainder	12	1754.32		
Total	24	7026.64	292.78	17.11

Complete Steps 1-6 for Fisher's three factor table. For step 5 show zero df for all interactive effect terms.

Writing out the expected mean squares.

Once the model is written and its components identified we can write out the expected means squares. From there we can write the correct (unambiguous) F-ratios.

With fixed effect models we do not need to do this because we always use the residual MS to form F-ratios.

With mixed models we need to write out expected mean squares to form the correct F-ratio, because the residual is **NOT** always the error term that isolates the each term from all others.

To see why the residual MS is not always the correct F-ratio denominator, we will write out the expected mean squares for Example 11, Mosquito Winglengths.

In this example there are 3 factors, all are random, and all are nested within another term. That is, all three pairs of terms fail the cross test.

Writing out the expected mean squares to form F-ratios - Random Effects

List the terms in the model, as in the ANOVA table
List the same term horizontally. In each row, show the row term.

		Cage	Fly%in% Cage	Error	
EMS	Cage	Cage			Each EMS includes itself
EMS	Fly%in% Cage		Fly%in% Cage		
EMS	Error			Error	

		Cage	Fly%in% Cage	Error	
EMS	Cage	Cage		Error	Each EMS includes the fixed error term
EMS	Fly%in% Cage		Fly%in% Cage	Error	
EMS	Error			Error	

		Cage	Fly%in% Cage	Error	
EMS	Cage	Cage	Fly%in% Cage	Error	Each EMS includes crossed (or nested) random terms
EMS	Fly%in% Cage		Fly%in% Cage	Error	
EMS	Error			Error	

		Cage	Fly%in% Cage	Error	Correct Denominator MS	
EMS	Cage	Cage	Fly%in% Cage	Error	Fly%in% Cage	Identify the denominator MS for the F-ratio
EMS	Fly%in% Cage		Fly%in% Cage	Error	Error	The denominator MS cancels all but the term of interest
EMS	Error			Error		

		Cage	Fly%in% Cage	Error	Incorrect Denominator MS	
EMS	Cage	Cage	Fly%in% Cage	Error		Cage MS / Error results in
	Fly%in% Cage					*2* uncancelled terms.
EMS	Error			Error	Error	The F-test is ambiguous

Writing out the expected mean squares to form F-ratios - Mixed Effects

List the terms in the model, as in the ANOVA table (Example 14 Sleep data)
List the same term horizontally

		Subj	Drug	Subj×Drug	Error
EMS	Subject	Subj			
EMS	Drug		Drug		
EMS	Subj×Drug			Subj×Drug	
EMS	Error				Error

Each EMS includes itself

		Subj	Drug	Subj×Drug	Error
EMS	Subject	Subj			Error
EMS	Drug		Drug		Error
EMS	Subj×Drug			Subj×Drug	Error
EMS	Error				Error

Each EMS includes the constant error term

		Subj	Drug	Subj×Drug	Error
EMS	Subject	Subj		Drug(Subj)	Error
EMS	Drug		Drug	Subj(Drug)	Error
EMS	Subj×Drug			Subj×Drug	Error
EMS	Error				Error

For each mixed term, display as if nested
Display Drug within Subj for EMS Subj
Display Subj within Drug for EMS Drug

		Subj	Drug	Subj×Drug	Error
EMS	Subject	Subj			Error
EMS	Drug		Drug	Subj(Drug)	Error
EMS	Subj×Drug			Subj×Drug	Error
EMS	Error				Error

Drop fixed terms, Retain random terms
Drug is fixed, Drug(Subj) is dropped,
Subj is random, Subj(Drug) is retained,

		Subj	Drug	Subj×Drug	Error
EMS	Subject	Subj			Error
EMS	Drug		Drug	Subj(Drug)	Error
EMS	Subj×Drug			Subj×Drug	Error
EMS	Error				Error

Identify denominator MS that isolates a single term

Error	Isolates Subj : F = 1 if Subj = 0
Subj(Drug) + Error	Isolates Drug : F = 1 if Drug = 0
Error	

Writing out the expected mean squares to form F-ratios - Mixed Effects

Gossett data has only one measurement per subject with each drug

There are too few df to estimate both the error term and the Drug×Subject term.

The residual term includes the error and the mixed term . Residual = Subj×Drug + Error

We test over the residual term.

		Subj	Drug	Subj×Drug	Residual	Denominator MS
EMS	Subject	Subj			Residual	(no test)
EMS	Drug		Drug	Subj(Drug)	Residual	Residual = Subj(Drug) + Error
EMS	Residual				Residual	