# Intro to mixed effects models

## Analysis of Ecological and Environmental Data

## QERM 514

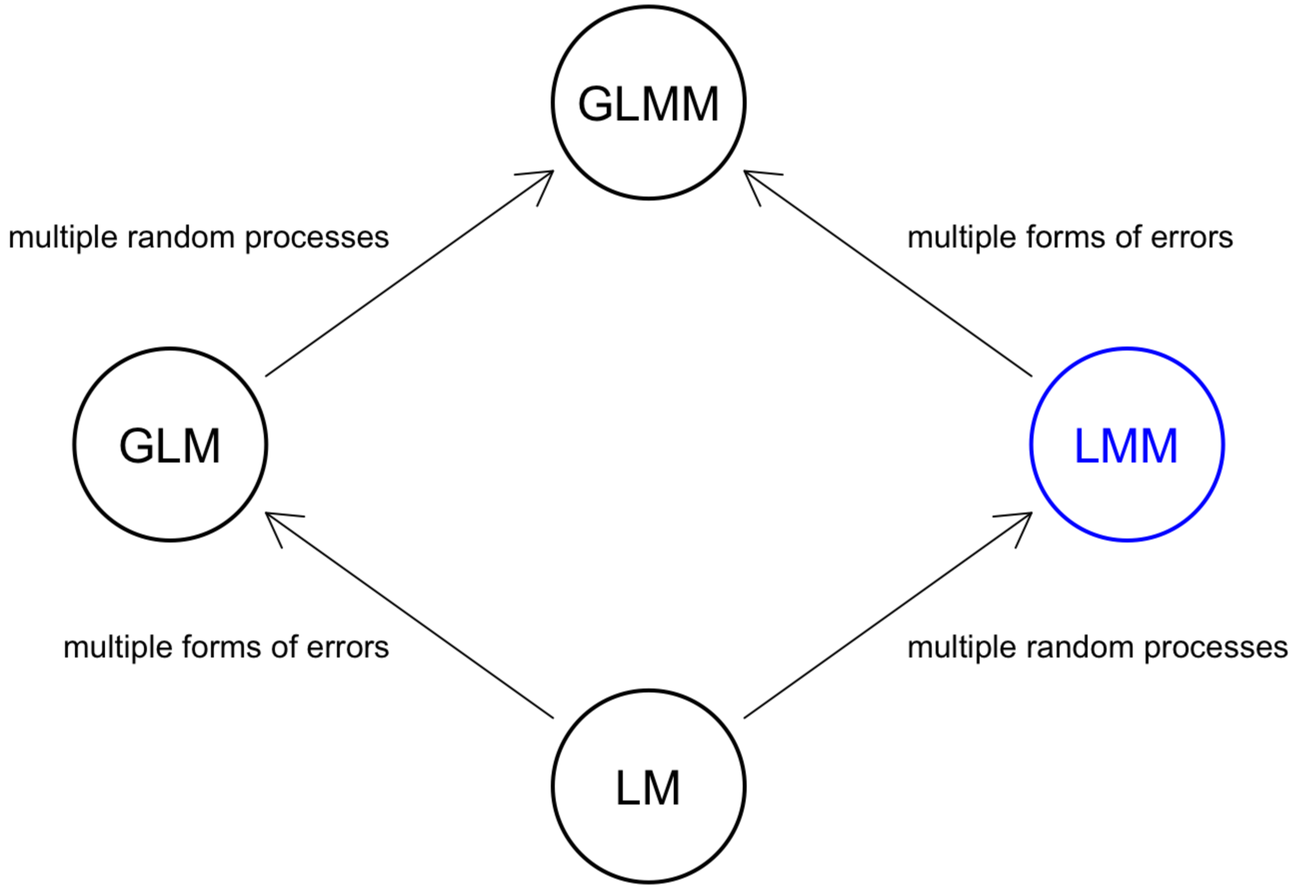
Mark Scheuerell

1 May 2020

# Goals for today

- Understand the difference between fixed and random effects
- Understand reasons to use random effects models
- Understand the benefits & costs of random effects models

# Forms of linear models



# Terminology

Mixed effects models are known by many names

- Variance components models
- Random effects models
- Varying coefficients models
- Hierarchical linear models
- Multilevel models

# Why use linear mixed models?

- Ecological data are often messy, complex, and incomplete
- Data are often grouped by location, species, etc
- May have multiple samples from the same individual
- Often small sample sizes for some locations, species, etc

# Fixed vs random effects

**fixed factor**: *qualitative* predictor (eg, sex)

**fixed effect**: *quantitative* change ("slope")

# Fixed vs random effects

**fixed factor**: *qualitative* predictor (eg, sex)

**fixed effect**: *quantitative* change (“slope”)

**random factor**: *qualitative* predictor whose levels are randomly sampled from a population (eg, age)

**random effect**: *quantitative* change whose levels are randomly sampled from a population

# Fixed vs random effects

Fixed effects describe *specific levels* of factors that are *not* part of a larger group

# Fixed vs random effects

Fixed effects describe *specific levels* of factors that are *not* part of a larger group

Random effects describe *varying levels* of factors drawn from a larger group

# Fixed vs random effects

### Fixed effects

- nutrient added or not
- female vs male
- wet vs dry

### Random effects

- genotype
- plot within a forest
- genus within family

# Random effects

Random effects occur in 3 circumstances

1. nested (hierarchical) studies

(eg, fish within lakes, multiple lakes within a state)

# Random effects

Random effects occur in 3 circumstances

1. nested (hierarchical) studies

2. time series (longitudinal) studies

(eg, repeated measurements from the same place or individual)

# Random effects

Random effects occur in 3 circumstances

1. nested (hierarchical) studies
2. time series (longitudinal) studies
3. spatial studies

(eg, multiple trees within a plot)

# Fixed vs random effects

Fixed effects influence only the **mean** of $y$

Random effects influence only the **variance** of $y$

# A linear model (ANCOVA)

Fish mass as a function of its length and *specific* lake

$$y_{i,j} = \underbrace{\alpha + \beta x_{i,j} + \delta_j}_{\text{fixed}} + \underbrace{\epsilon_{i,j}}_{\text{random}}$$

$y_i$ is the log(mass) for fish *i* in lake $j$

$x_i$ is the log(length) for fish *i* in lake $j$

$\delta_j$ is the mean log(mass) of fish in lake $j$

$\epsilon_{i,j} \sim \text{N}(0, \sigma_\epsilon)$

# A linear mixed model

Fish mass as a function of its length and *general* lake

$$y_{i,j} = \underbrace{\alpha + \beta x_{i,j}}_{\text{fixed}} + \underbrace{\delta_j + \epsilon_{i,j}}_{\text{random}}$$

$y_i$ is the log(mass) for fish *i* in lake $j$

$x_i$ is the log(length) for fish *i* in lake $j$

$\delta_j$ is the mean log(mass) of fish in lake $j$

$\epsilon_{i,j} \sim \text{N}(0, \sigma_\epsilon)$ and $\delta_j \sim \text{N}(0, \sigma_\delta)$

# Michael Freeman’s visualization

# Five fundamental assumptions

- Within-group errors are *independent* with mean zero and variance $\sigma^2$
- Within-group errors are *independent of the random effects*
- Random effects are normally distributed with mean zero and covariance $\Psi$
- Covariance matrix $\Psi$ *does not depend* on the level
- Random effects are *independent* among different levels

# Levels of random effects

In many cases, we can have multiple *levels* of random effects

trees within plots within forests within regions within states

# Tricks to random effects

- learning which variables are random effects
- correctly specifying the fixed and random effects in a model
- getting the nesting structure correct

# Questions about random effects

## Experimental design

Where does most of the variation occur & where would increased replication help?

# Questions about random effects

## Hierarchical structure

What are the different levels of variation?

# Pseudoreplication

To qualify as *true replicates*, measurements must

- be independent
- not be part of a time series
- not be grouped in together in one place
- not be repeated on the same subject

# Pseudoreplication

## An example

Imagine a field experiment to test insecticide effects on plant2

- 20 plots: 10 sprayed & 10 unsprayed
- 50 plants within each plot
- each plant is measured 5 times

# Pseudoreplication

## An example

Imagine a field experiment to test insecticide effects on plants

- 20 plots: 10 sprayed & 10 unsprayed
- 50 plants within each plot
- each plant is measured 5 times

What are the degrees of freedom?

$20 \times 50 \times 5 = 5000$ (?)

# Pseudoreplication

## An example

Imagine a field experiment to test insecticide effects on plants

- 20 plots: 10 sprayed & 10 unsprayed
- 50 plants within each plot
- each plant is measured 5 times

What are the degrees of freedom?

~~20~~ × ~~50~~ × ~~5~~ = ~~5000~~ (?)

$2 \times 9 = 18$ (!)

# Model for means

Consider a simple one-way ANOVA model

$$y_{ij} = \mu + \alpha_j + \epsilon_{ij}$$
$$\epsilon_{ij} \sim \text{N}(0, \sigma_\sigma^2)$$

where the group-level means $\alpha_j$ are *fixed*

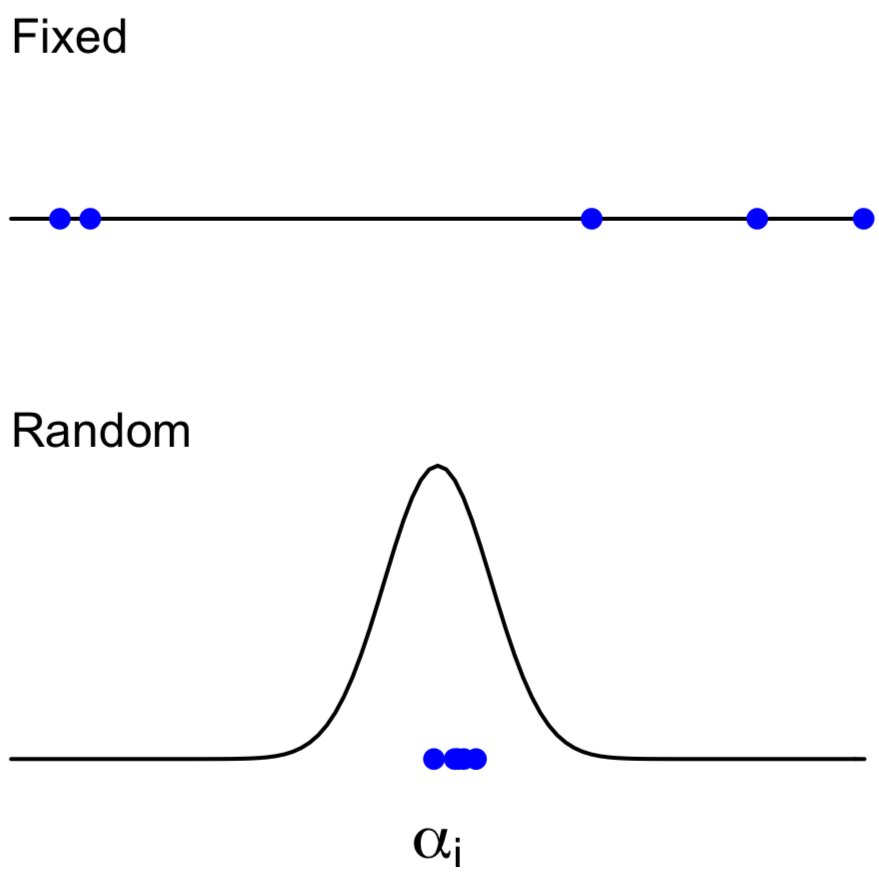
# Model for means

Now consider this one-way ANOVA model

$$
\begin{aligned}
y_{ij} &= \mu + \alpha_j + \epsilon_{ij} \\
\epsilon_{ij} &\sim \text{N}(0, \sigma_\sigma^2) \\
\alpha_j &\sim \text{N}(0, \sigma_\alpha^2)
\end{aligned}
$$

where the group-level means $\alpha_j$ are *random*

# Distribution of means



# Intraclass correlation

The means in the fixed effect model are independent

The means in the random effects model are correlated

# Intraclass correlation

The means in the fixed effect model are independent

The means in the random effects model are correlated

$$\rho = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\epsilon^2}$$

The correlation depends on the relative size of $\sigma_\alpha^2$ vs $\sigma_\epsilon^2$

# Group means

In fixed effects models, the group means are

$$\alpha_j = \bar{y} - \mu$$

# Shrinkage of group means

In fixed effects models, the group means are

$$\alpha_j = \bar{y} - \mu$$

In random effects models, the group means "shrink" towards one another

$$\alpha_j = (\bar{y} - \mu)\left(\frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\epsilon^2}\right)$$

# Shrinkage

Consider what happens to $\alpha_j$ as $\sigma_\alpha^2 \to \infty$

$$\alpha_j = (\bar{y} - \mu)\left(\frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\epsilon^2}\right)$$

$$\Downarrow$$

$$\begin{aligned} \alpha_j &= (\bar{y} - \mu)\left(\frac{\infty}{\infty + \sigma_\epsilon^2}\right) \\ &= \bar{y} - \mu \end{aligned}$$

# Shrinkage

As $\sigma_\alpha^2 \to \infty$, our random effects become increasingly independent

$$\alpha_j \sim \mathrm{N}(0, \sigma_\alpha^2)$$

$$\Downarrow$$

$$\alpha_j \sim \mathrm{Unif}(-\infty, \infty)$$

# Benefits

- Broadens our inference to a larger population
- Larger groups inform smaller groups (“Robin Hood Effect”)

# Benefits

- Broadens our inference to a larger population
- Larger groups inform smaller groups (“Robin Hood Effect”)
- Represents a “compromise” in terms of information used
  - fixed effects: no grouping (“no pooling”)
  - random effects: some grouping (“partial pooling”)
  - none: all one group (“complete pooling”)

# Costs

- Precision decreases with number of levels
- Random effects perhaps more difficult to explain