Introduction to Generalised Linear Models

Dr Niamh Mimnagh

niamh@prstats.org

https://github.com/niamhmimnagh/glmm01

Hierarchical Data

- Many datasets have a nested or grouped structure:
  - Students within schools
  - Patients within hospitals
  - Plots within sites
- Observations within a group tend to be more similar than between groups.
- Ignoring this structure can lead to misleading conclusions.

Example: Bird Abundance Study

- Ecological survey: counts of a bird species across 15 sites, monthly for 2 years.

- Grouping structure:
  - Measurements nested within sites
  - Repeated measures over time

- Question: Does temperature influence abundance?

Why OLS Regression May Fail

- Ordinary Least Squares regression assumes independent observations.

- Hierarchical data violate this assumption:

  – Observations from the same group are correlated.

- Ignoring correlation:

  – Underestimates standard errors

  – Inflates Type I error rates (false positives)
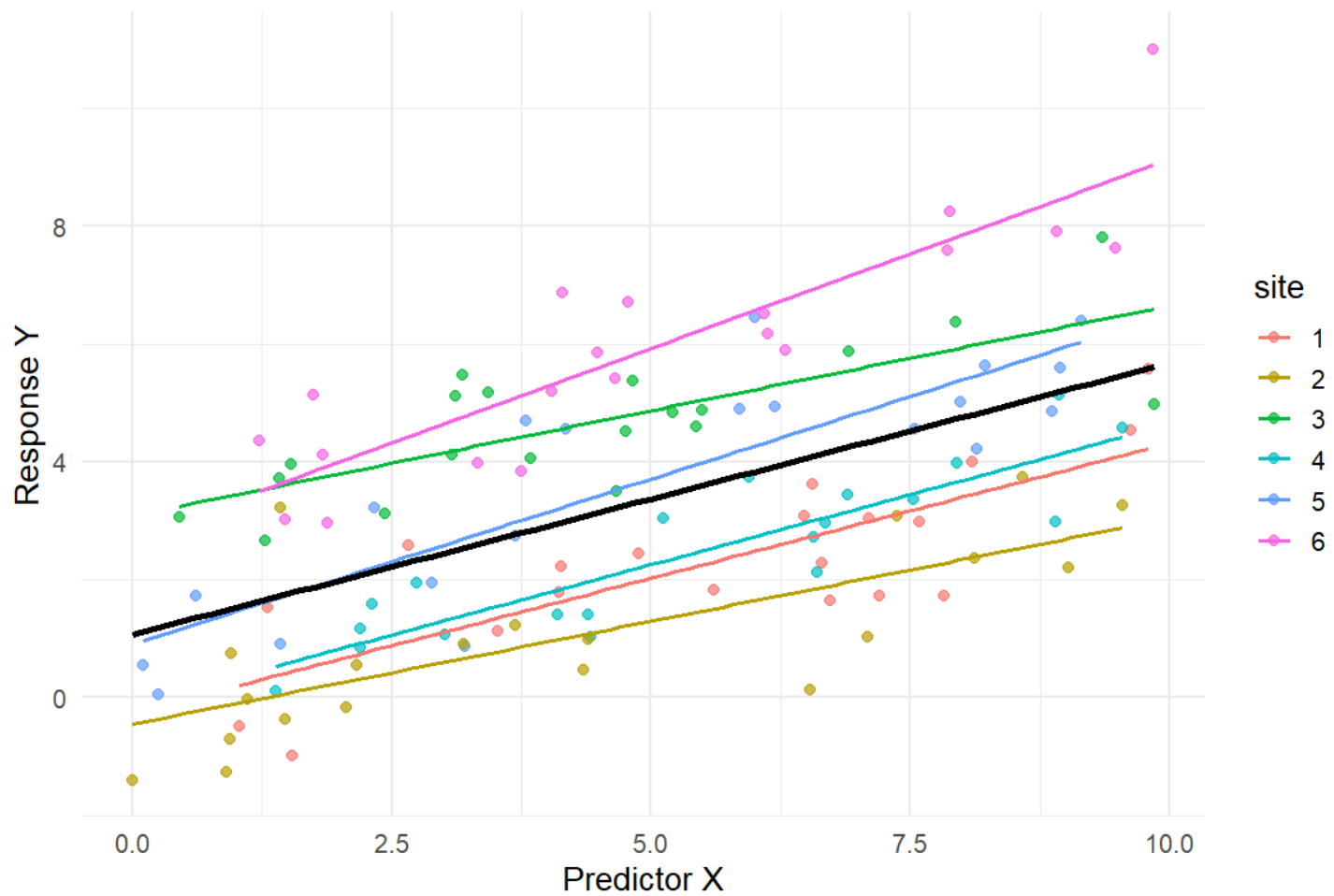
Consequences of Ignoring Grouping

- Treating clustered observations (e.g., sites, herds, subjects, years) as independent underestimates standard errors and produces artificially small p-values, leading to overconfident inference.

- Pooled models can distort effect sizes because they confound within-group and between-group variation, especially when groups have different baselines or slopes.

- Conclusions drawn from the pooled fit may not hold for any individual group, and predictions will be overconfident - particularly for new or sparsely sampled groups.
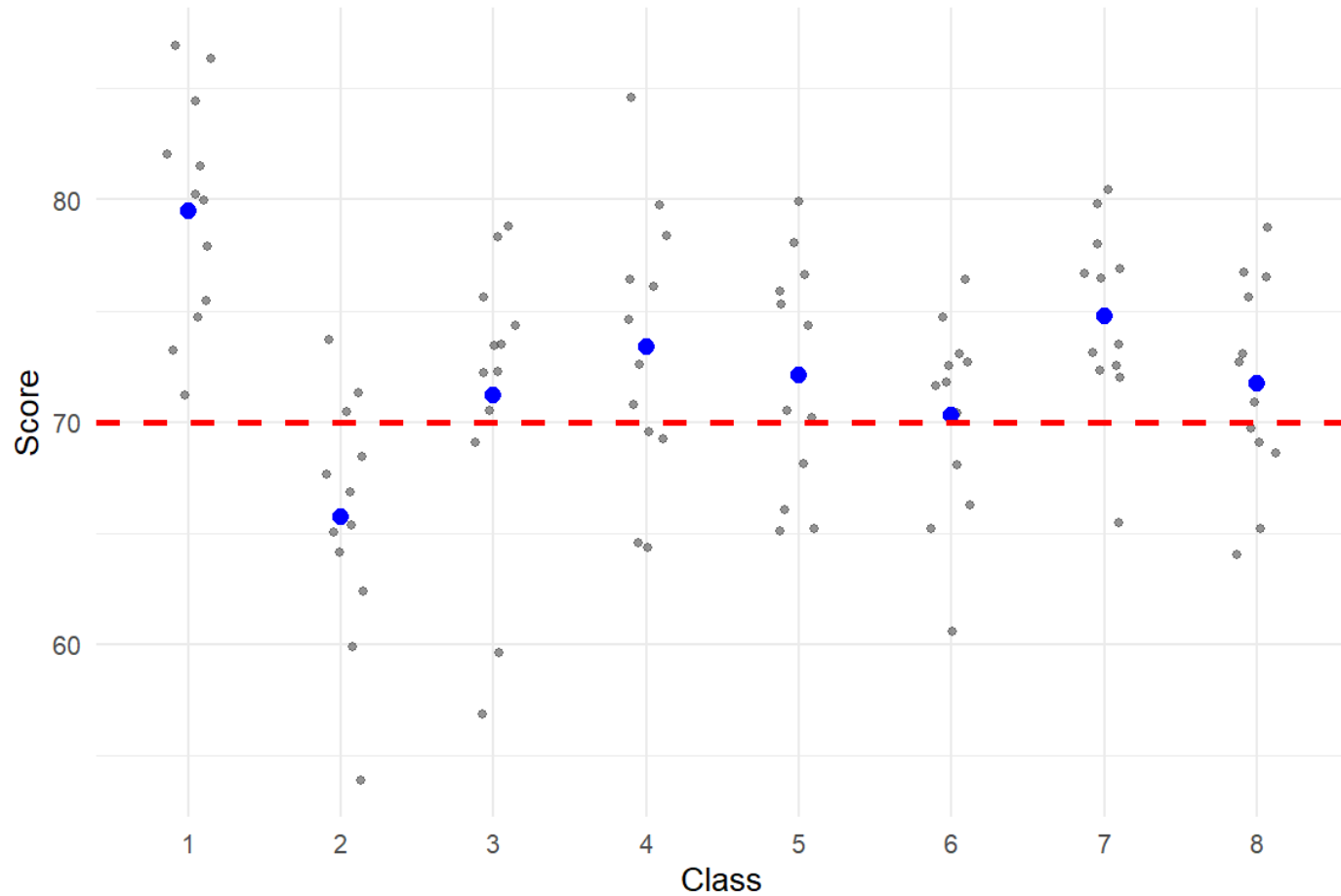
Fixed Effects

- Fixed effects are parameters that describe systematic relationships that are assumed to be the same across all observational units or tied to specific, repeatable conditions (such as treatments).

- They represent population-level averages: the coefficients answer, "on average, how does the outcome change when this predictor changes, holding other variables constant?"

- For example, the overall effect of temperature on bird abundance, or the mean difference between fertiliser A and fertiliser B, are fixed effects.

Random Effects

- Random effects describe group-specific deviations from the overall mean or slope, allowing each group to have its own baseline and/or response to predictors.

- We treat groups as a random sample from a larger population and model their effects as draws from a distribution—typically Normal with mean zero—so we also estimate a variance component for between-group variability.

- Using random effects yields correct standard errors for clustered data and supports prediction for new or sparsely observed groups via the estimated random-effects distribution.

- Example: site-specific baseline abundance (random intercept) and site-specific temperature responses (random slope) in a multi-site ecology study.

Class Exam Scores

- In this setting, the fixed effect is the overall mean exam score across all classes; it answers what the average student would score in the population of classes.

- A random intercept for class captures each class's deviation from that overall mean, acknowledging that some classes are consistently higher or lower for unmeasured reasons.

- This structure accounts for the natural variation between classes and avoids overstating precision that would occur if all students were treated as independent.
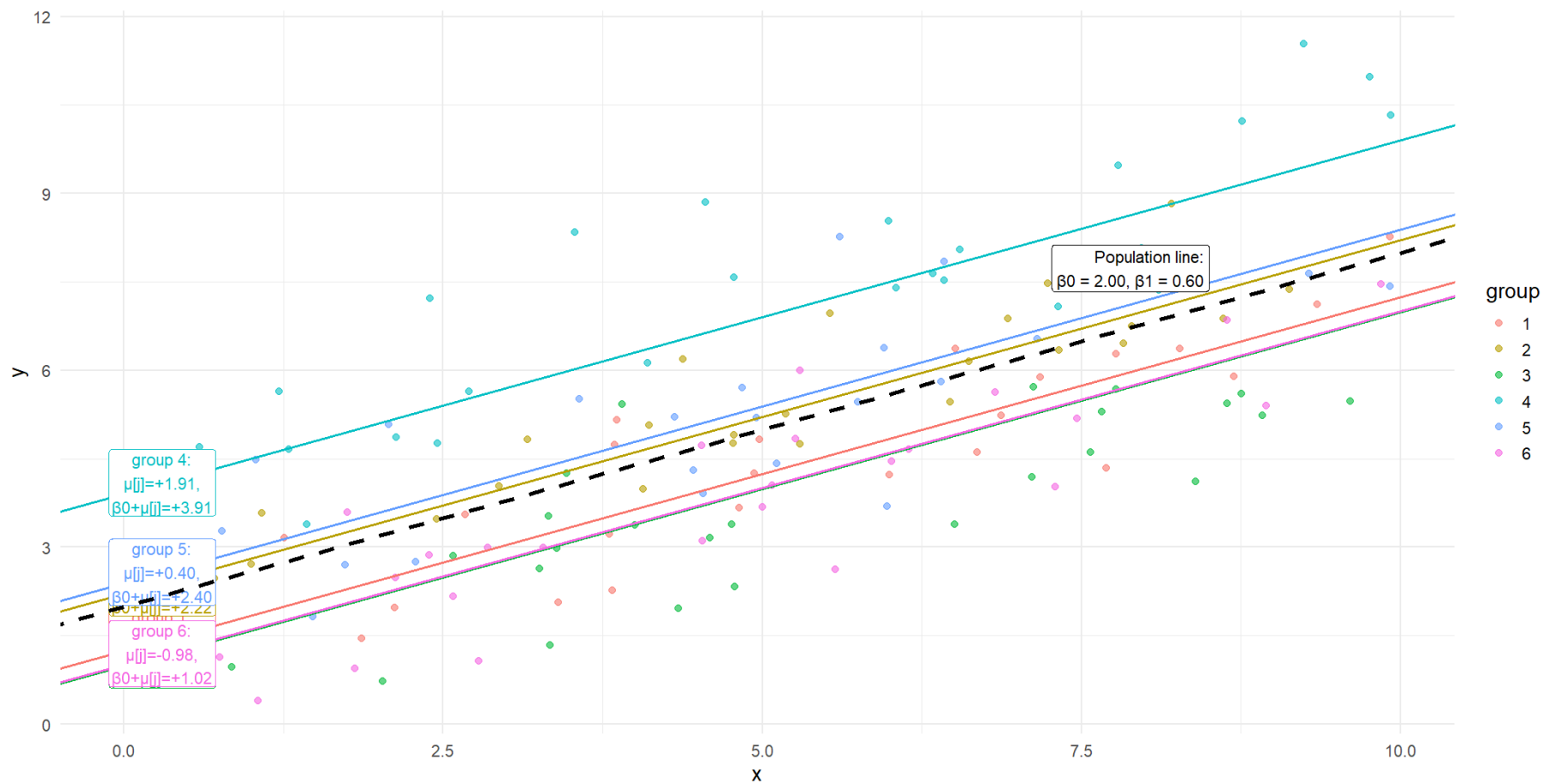
Fixed Effects Models

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

- Y_i Response for observation
- β_0: Fixed intercept (grand mean)
- β_1: Fixed slope for predictor x_i
- ϵ_i: Residual error, $\epsilon_i \sim N(0, \sigma^2)$

- We can also write $\epsilon_i = Y_i - \mu_i$ where $\mu_i = \beta_0 + \beta_1 x_i$
$$Y_i \sim Normal(\mu_i, \sigma^2)$$

Random Intercept Models

$$Y_{ij} \sim N\left(\mu_{ij}, \sigma^2\right)$$
$$\mu_{ij} = \beta_0 + a_{0j} + \beta_1 x_{1j}$$

- Y_{ij}: response for observation i in group j

- β_0: overall intercept

- $a_{0j} \sim N\left(0, \sigma_a^2\right)$: random intercept for group j

- β_1: fixed slope

Introduction to Generalised Linear Models
Dr Niamh Mimnagh, PR Stats

lme4

- The lme4 package is one of the most widely used tools for fitting LMMs and GLMMs

- Uses the lmer() function for continuous responses and glmer() for non-Gaussian responses (e.g., binomial, Poisson).

- Models include fixed effects (shared across all groups) and random effects (varying by group/cluster).

```
library(lme4)                Fixed Effect    Random Effect
model <- lmer(Y ~ X + (1 | Group), data = mydata)
summary(model)
```

Coding Demo

Model Output: Fixed Effects

- The intercept is the model's predicted mean outcome when all predictors are at their reference levels (for factors) or at zero (for numeric variables).

- The slope for a numeric predictor is the average change in the outcome associated with a one-unit increase in that predictor, holding the others constant.

- For categorical predictors, each coefficient is the mean difference from the reference category, with all other predictors held constant.

- If zero is not meaningful, centre numeric predictors so the intercept represents the mean outcome at typical values.

- When interactions are included, each main-effect coefficient is interpreted at the reference level (or at zero/mean) of the interacting variable, not as a global average effect.

Model Output: Random Effects

- Variance among site intercepts (between groups): σ_a^2
  - How much does group j's intercept deviate from the grand mean?
  - Bigger σ_a^2 means groups differ more in their baseline
- Residual variance (within groups): σ^2
  - How much do observations spread around their group's mean?
  - Bigger σ^2 means observations within a group are more scattered
- If $\sigma_a^2 \ll \sigma^2$, group effects are weak, and estimates shrink strongly toward the overall mean.
- Use these to compute ICC.

Intra-Class Correlation (ICC)

$$ICC = \frac{\sigma_a^2}{\sigma_a^2 + \sigma^2}$$

- Proportion of variance explained by grouping.

- The ICC answers the question: "If I pick two observations from the same group at random, how correlated will they be?"

  – ICC close to 0: little group effect. Knowing what group an observation is in tells you little about the outcome.

  – ICC close to 1: strong group-level similarity. Knowing what group an observation is in tells you a lot about the outcome – observations in the same group are similar to one another.

Estimating Random Effects

- A random effect a_j is the random deviation for group j from the overall mean. It is <u>not</u> the mean for group j.

- Best Linear Unbiased Predictions (BLUPs) are predictions of random effects in mixed models. They represent how much each group's intercept or slope differs from the overall fixed effect.

- In lme4: ranef() returns BLUPs; coef() combines BLUPs with fixed effects.

Shrinkage in Random Effects

$$\hat{a}_j \approx \frac{\sigma_a^2}{\sigma_a^2 + \frac{\sigma^2}{n_j}}(\bar{y}_j - \bar{y})$$

Shrinkage Factor

- In mixed models, shrinkage is the tendency for group-level estimates to be pulled toward the overall population mean.
- The amount of shrinkage depends on group sample size, variance of the random effects, and residual variance.
- Stronger shrinkage occurs for smaller groups.

Shrinkage in Random Effects

- Without shrinkage (no pooling):
  - Each group estimate is based only on its own data.
  - Groups with few observations have unstable, noisy estimates.
  - Large variability in estimates increases risk of overfitting.
- Shrinkage impacts groups differently:
  - Small groups: pulled strongly toward overall mean.
  - Large groups: minimal shrinkage, close to group mean.
- Reduces variance of estimates but introduces bias toward the mean.

Shrinkage in Random Effects

- Advantages:
    - Improves predictive accuracy for new data.
    - Reduces overfitting in small groups.
    - Makes estimates more stable and reliable.
    - Handles imbalance in group sizes naturally.
- Disadvantages:
    - May underestimate true group differences.
    - Can mask real effects in small but important groups.
    - Estimates are not simply raw group means.

Shrinkage in Random Effects

Shrinkage Effect by Sample Size
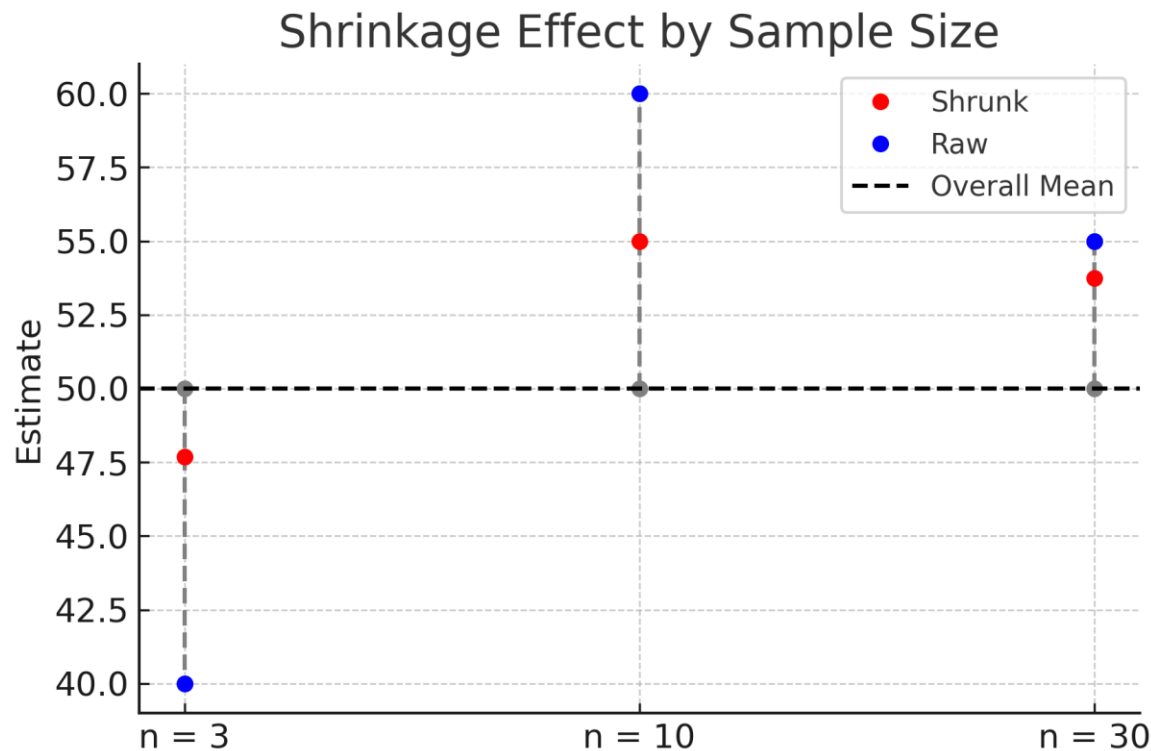
ML vs. REML

- Linear mixed-effects model (LMM) has two types of parameters:
  - Fixed effects – population-average trends.
  - Random effects and variance components – variability among groups and residual noise.
- Two estimation methods:
  - Maximum Likelihood (ML) – estimates all parameters together.
  - Restricted Maximum Likelihood (REML) – adjusts for uncertainty in estimating fixed effects.

Differences Between Them

- What is maximised:
  - ML: Likelihood of full data given all parameters (estimates fixed effects and variance components together)
  - REML: Likelihood of residuals (data after removing fixed effects)
- Variance components:
  - ML: Biased downward (esp. in small samples)
  - REML: Unbiased
- Model comparison:
  - ML: Useful for comparing models with different fixed effects
  - REML: Useful for comparing models with the same fixed effects but different random effects

When to Use Random and Fixed Effects

- Use random effects when:
  - Groups are random sample from a larger population.
  - Interested in population-level effect, not each group specifically.
- Use fixed effects when:
  - Groups are exhaustive and represent all possible levels.
  - Aim is to compare specific groups.

Going Beyond Random Intercepts

- Random intercept models assume the same slope across all groups.

- In practice, predictor-response relationships may vary by group.

- Random slopes are essential when there is strong evidence that the effect size varies by group, or that prediction accuracy depends on group-specific slopes
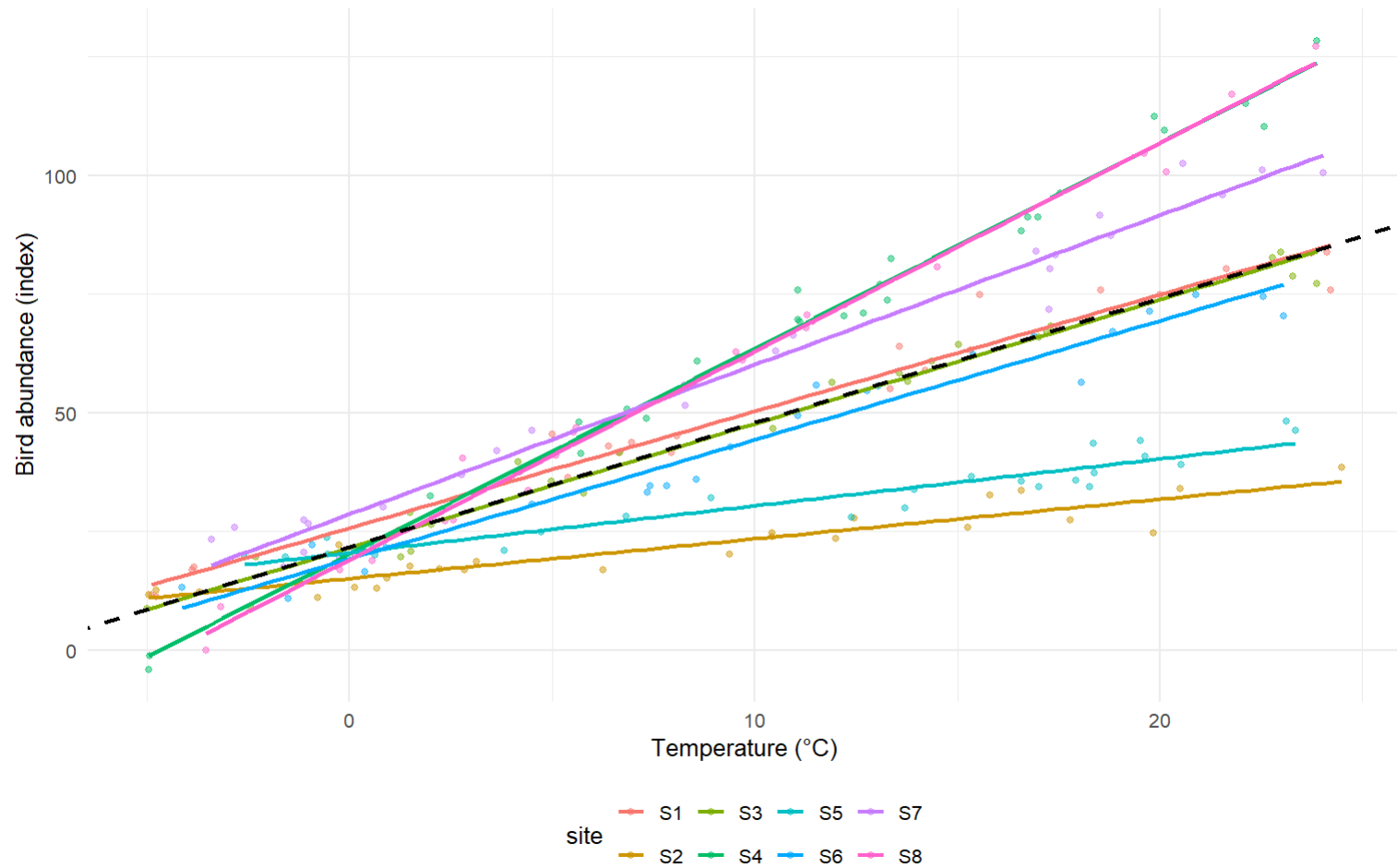
Consequences of Ignoring Slope Variation

- If the effect of a predictor truly differs across groups, omitting random slopes forces a single common slope and underestimates the uncertainty around that effect and your predictions.

- When slopes vary and are correlated with group characteristics (or group intercepts), the fixed-effect slope can become biased, often appearing stronger than it really is.

- Standard errors and p-values for the fixed slope are then anti-conservative, inflating Type I error and encouraging spurious significance.

- Predictions for new or sparsely sampled groups will be overconfident and may miss systematic group-specific trends.

Random Slopes Model

$$Y_{ij} \sim N\left(\mu_{ij}, \sigma^2\right)$$
$$\mu_{ij} = \beta_0 + a_{0j} + \left(\beta_1 + a_{1j}\right)x_{ij}$$

- a_{0j} : random intercept for group j. This tells us how much higher or lower the group's baseline is compared to the overall average

- a_{1j} : random slope for predictor x in group j. This tells us how much stronger or weaker the group's response is to the predictor, compared to the overall slope.

Intercept-Slope Correlation

- Groups often differ not only in baseline (intercept) but also in responsiveness (slope).

- High-baseline groups may systematically respond differently than low-baseline groups.

- Allowing a covariance captures this coupling and avoids misallocating variance. Constraining it to zero can distort fixed-effect SEs and shrinkage.

- Positive correlation: high-baseline groups have strong positive slopes.

- Negative correlation: high-baseline groups have weaker slopes.

Intercept-Slope Correlation

$$Y_{ij} \sim N(\mu_{ij}, \sigma^2)$$

$$\mu_{ij} = \beta_0 + a_{0j} + (\beta_1 + a_{1j})x_{ij}$$

$$\begin{bmatrix} a_{0j} \\ a_{1j} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma \right), \Sigma = \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{bmatrix}$$
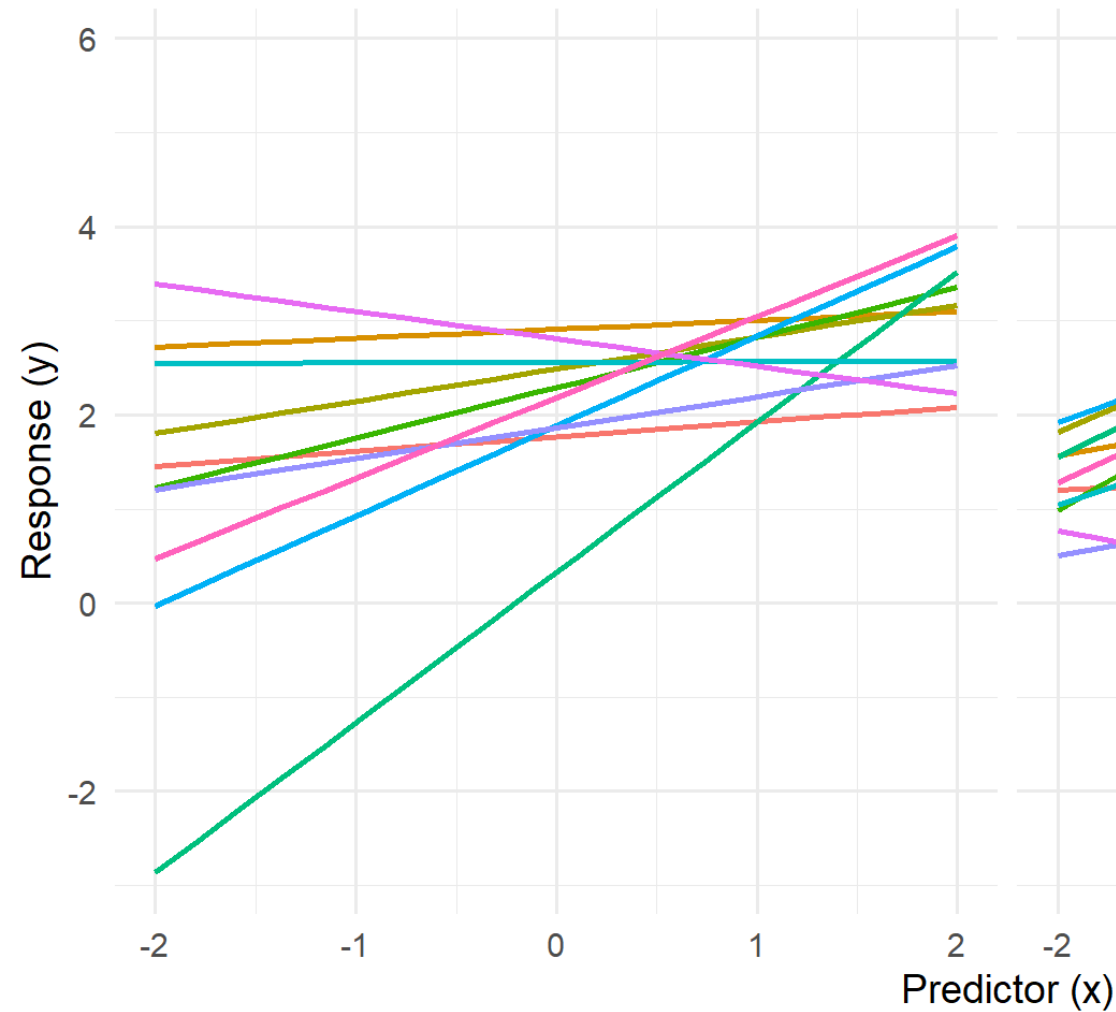
- Intercept and slope deviations are jointly normally distributed across groups. The mean is (0,0) because the fixed effects already capture the overall average intercept and slope.

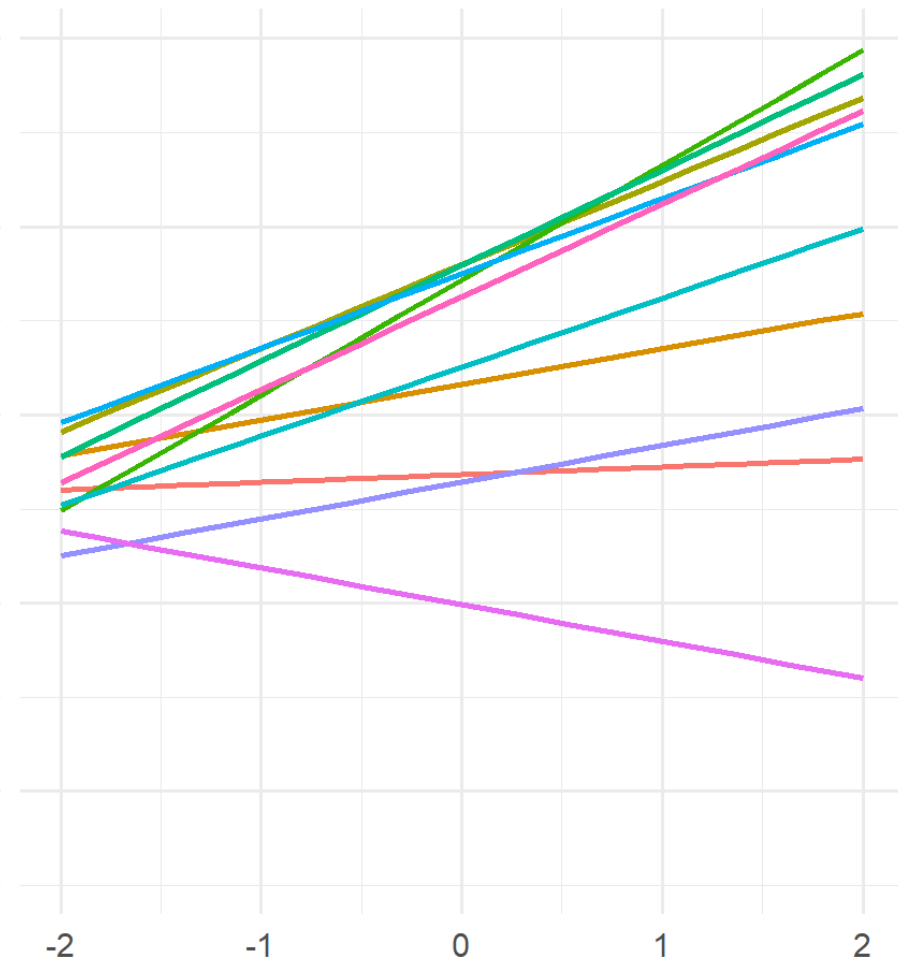$$\rho = \frac{\sigma_{01}}{\sigma_0 \sigma_1}$$

- ρ lies between -1 and +1. If $\rho > 0$, groups with higher intercepts tend to have more positive slopes. If $\rho < 0$, groups with higher intercepts tend to have weaker or more negative slopes.

Negative correlation (ρ = -0.7)   Positive correlation (ρ = +0.7)

Introduction to Generalised Linear Models
Dr Niamh Mimnagh, PR Stats

ICC Revisited

- With a random slope, the ICC is no longer constant - it depends on the value of the predictor x.

- Groups differ not only in their intercepts but also in their slopes, so the between-group variance changes across x.

- The ICC now answers: "If I pick two observations from the same group at random, at a given x, how correlated will they be?"

- Formula for between-group variance at x:

$$Var_{between(x)} = \sigma_0^2 + 2x\sigma_{01} + x^2\sigma_1^2$$

ICC at x:

$$ICC(x) = \frac{Var_{between(x)}}{Var_{between(x)} + \sigma^2}$$

ICC Revisited

- Random intercept: group differences are the same for all x: ICC constant.

- Random slope: group differences vary with x: ICC changes.

- At some x values, groups may be similar: low ICC.

- At other x values, groups may be far apart: high ICC.

- The intercept–slope covariance term (σ_{01}) can increase or decrease ICC depending on its sign.

region
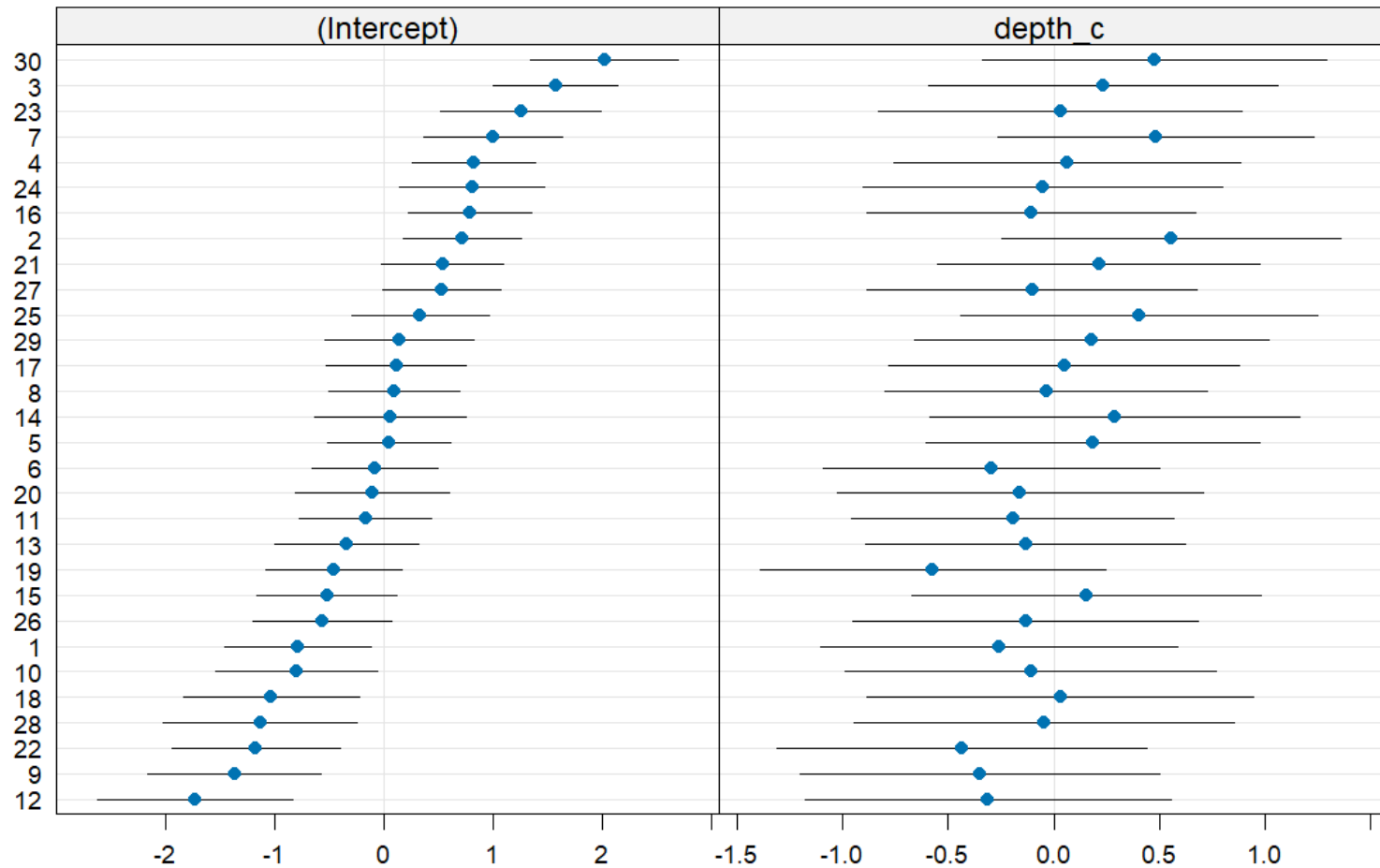
Coding Demo

Model Complexity and Overfitting

- Random slope models use more parameters: a random intercept model uses one variance term for the intercept. A model with both a random slope and intercept uses three parameters.

- Random slope models require more data per group: A random intercept can be stable with a much smaller number of groups (5-10 groups and 3-4 observations per group can sometimes work), whereas the additional parameters in a random slope model requires more groups (the rule of thumb is at least 20 groups), and at least 5-6 observations per group.

- Overfitting can lead to convergence warnings.

- To simplify an overfitted model you can remove correlation between intercept and slope: (1 + temperature || site). This reduces parameters while keeping slope variation.

Moving Beyond LMMs

- LMMs assume normally distributed residuals and constant variance.

- Many ecological/biological data types are:

  – Binary (presence/absence)

  – Counts (number of events)

  – Proportions (successes/trials)

- These require different distributions and link functions.

GLMMs

Generalised Linear Models (GLMMs) extend LMMs to handle non-normal response distributions while accounting for random effects.

Structure:

1. Random component: a distribution belonging to the exponential family
2. Systematic component: fixed effects and random effects
3. Link function: a function that links the mean to the linear predictor

GLMMs

Random
Component

$$Y_{ij} \sim Normal(\mu_{ij}, \sigma^2)$$
$$\mu_{ij} = \beta_o + \beta_1 x_i + a_{0j}$$

Link
Function:
The identity link ($g(\mu_i) = \mu_i$)

Systematic
Component

The Bernoulli GLMM

$$Y_i \sim Bernoulli(\pi_i)$$

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + a_0 + \beta_1 x_{1i} + \ldots + \beta_p x_{pi}$$

The function $\log\left(\frac{\pi_i}{1-\pi_i}\right)$ can also be written as $logit(\pi_i)$

We want to model the success probabilities π_i as a function of predictors, but since the β coefficients are unbounded, this would result in unbounded π_i values

But π_i are probabilities, and so have to be bounded in the $(0,1)$ interval

The logit is simply the natural logarithm of the odds. This is how we ensure predicted probabilities stay between 0 and 1.

Example:
Frog Presence vs. Water Depth

- Imagine you're an ecologist studying the habitat preferences of a frog species in a series of wetland sites.

- Your main research question: Does water depth influence the probability of finding frogs at a given location?

- Survey area: Several distinct wetland regions (e.g., separate ponds or marsh complexes)

- Sampling method: Multiple sites per region, along a depth gradient (0.1-2.5 m)

- Outcome variable (presence): 1 if frogs detected, 0 if not

- Predictor variable (depth): Measured in metres

Example:
Frog Presence vs. Water Depth

- Why water depth might matter:
  - Retains moisture longer into the dry season
  - Offers more stable temperatures
  - Provides refuge from predators
- Alternatively, very deep water may be unsuitable for some life stages - the relationship could be positive or negative.

GLM Without Random Effects

- We assume the binary presence/absence variable follows a Bernoulli distribution:

$$Y_i \sim Bernoulli(\pi_i)$$
$$logit(\pi_i) = \beta_0 + \beta_1 depth_i$$

Where:

$Y_i = 1$ if frogs are detected at site i, 0 otherwise

π_i = probability of frog presence at site

= intercept (log-odds at zero depth)

β_1 = slope (change in log-odds per metre depth)

GLM Without Random Effects

- This model might be too simplistic. Different regions may have baseline differences due to:
  - Local climate
  - Vegetation cover
  - Pollution or predator presence
  - Surveyor differences
- Ignoring this risks underestimating uncertainty or biasing slope estimates.

Adding Random Intercepts for Region

Lets try accounting for some baseline variation:

$$Y_{ij} \sim Bernoulli(\pi_{ij})$$
$$logit(\pi_{ij}) = \beta_0 + \beta_1 depth_i + a_j$$
$$a_j \sim N(0, \sigma^2_{region})$$

Where:

Y_{ij} = presence at site i in region

π_{ij} = probability of presence at site i in region

a_j = random intercept for region j, capturing unobserved variation in baseline presence probability

Adding Random Intercepts for Region

- A random intercept assumes regions differ only in baseline occupancy, not in how the depth affects frog presence. This can be unrealistic when:
  - Depth range differs within regions - if region A has mostly shallow sites and region B spans the full gradient, the depth-presence relationship is learned differently across regions
  - Unmeasured modifiers - e.g., predator fish appear only in the deepest pools of certain regions, vegetation or temperature stratification changes with depth locally.

Adding Random Slopes for Depth

$$Y_{ij} \sim Bernoulli(\pi_{ij})$$
$$logit(\pi_{ij}) = \beta_0 + a_{0j} + (\beta_1 + a_{1j})depth_i$$
$$\begin{bmatrix} a_{0j} \\ a_{1j} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma\right), \Sigma = \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{bmatrix}$$

Region-specific baseline: intercept in region j is $\beta_0 + a_{0j}$

Region-specific depth effect: slope in region j is $\beta_1 + a_{1j}$

Covariance σ_{01} lets intercepts and slopes co-vary. If $\sigma_{01} > 0$, regions with higher baseline occupancy tend to have more positive depth effects.

We typically centre depth so β_0 is the average log-odds at mean depth, and to reduce intercept-slope collinearity, improving estimation

Coding Demo

Choosing Appropriate Models

- If there is a reason that a predictor's effect differs by group and that predictor varies within groups, a random slope may be needed. Otherwise, a random-intercept model is often sufficient.

- Random slopes introduce extra parameters, increase computation, and are more prone to convergence issues; they also require enough groups and within-group data to estimate reliably.

- A practical workflow is to start with a random-intercept model, examine group-wise scatterplots/slopes, then test whether adding random slopes improves fit or prediction using AIC/LOO or a likelihood-ratio test for nested models.

- If the slope variance estimates near zero, you see "singular fit" warnings, or predictions do not improve, prefer the simpler random-intercept model.

Reporting Effects

Communicating Random Effects

- Report variance and SD for random effects in GLMMs.

Communicating Fixed Effects

- Provide effect size, standard error, and 95% CI.
- Back-transform where necessary (odds ratios).

Model Comparison Metrics

- AIC for model selection:

  - When comparing two models, a lower AIC value is better,

  - AIC tends to pick the model with the best expected out-of-sample predictive performance

- BIC for model selection:

  - BIC uses a stronger penalty than AIC

  - Similar to AIC, when comparing two models, a lower BIC value is better,

  - BIC is good for picking the 'truest' model structure from a candidate set.

- Likelihood Ratio Test for nested models:

  - H_o: the simpler model provides a better fit.

  - H_a: the fuller model provides a better fit to the data.

  - $p < 0.05 \rightarrow$ fuller model is better

PR STATS

Introduction to Generalised Linear Models
 Dr Niamh Mimnagh, PR Stats

Model Simplification

- Drop predictors if they are not improving fit:
  - If the models are nested compare with LRT, otherwise try AIC
- Prefer parsimony
- Keep ecologically meaningful variables
  - Keep variables that are essential for interpretation/causal control even if their effect is weak

Overfitting Risks:

- Too many fixed effects will lead to large SEs/instability, and weak prediction
- Over-complex random part can lead to convergence issues
- Always check the biological plausibility of your parameters

Common Errors in GLMMs

- Treating grouping as fixed:
  - This will inflate Type-I errors, because repeated measures within groups are correlated, and if we don't account for grouping, we treat all observations as independent.

- Ignoring overdispersion
  - Extra variance that is unaccounted for will lead to SEs that are too small.

- Not scaling covariates
  - This can lead to huge SEs, because if predictors are on different scales, optimisation can be unstable

- Blind AIC use
  - Picking models on the smallest AIC without checking assumptions, or whether AIC works with the model in question. For nested models, confirm with LRT. If difference in AIC values is less than 2, treat it as a tie and prefer the parsimonious model.

What if GLMMs are Not Enough?

- If you suspect your predictors have a non-linear relationship with the response, then Generalised Additive Models (GAMs) should be tried.

- If your data has an excess number of zero counts, then zero-inflated models are a good approach

- If your data is small and you suspect grouping, then Bayesian GLMMs can handle this.

Path for Further Learning

- Hierarchical Bayesian models

- GAMMs

- Spatial GLMMs

- Multivariate models