

Introduction to Generalised Linear Mixed Models

Dr Niamh Mimmagh

niamh@prstats.org

[https://github.com/niamhmimmagh/FMME01---
Introduction-to-Generalised-Linear-Mixed-
Models](https://github.com/niamhmimmagh/FMME01---Introduction-to-Generalised-Linear-Mixed-Models)

Hierarchical Data

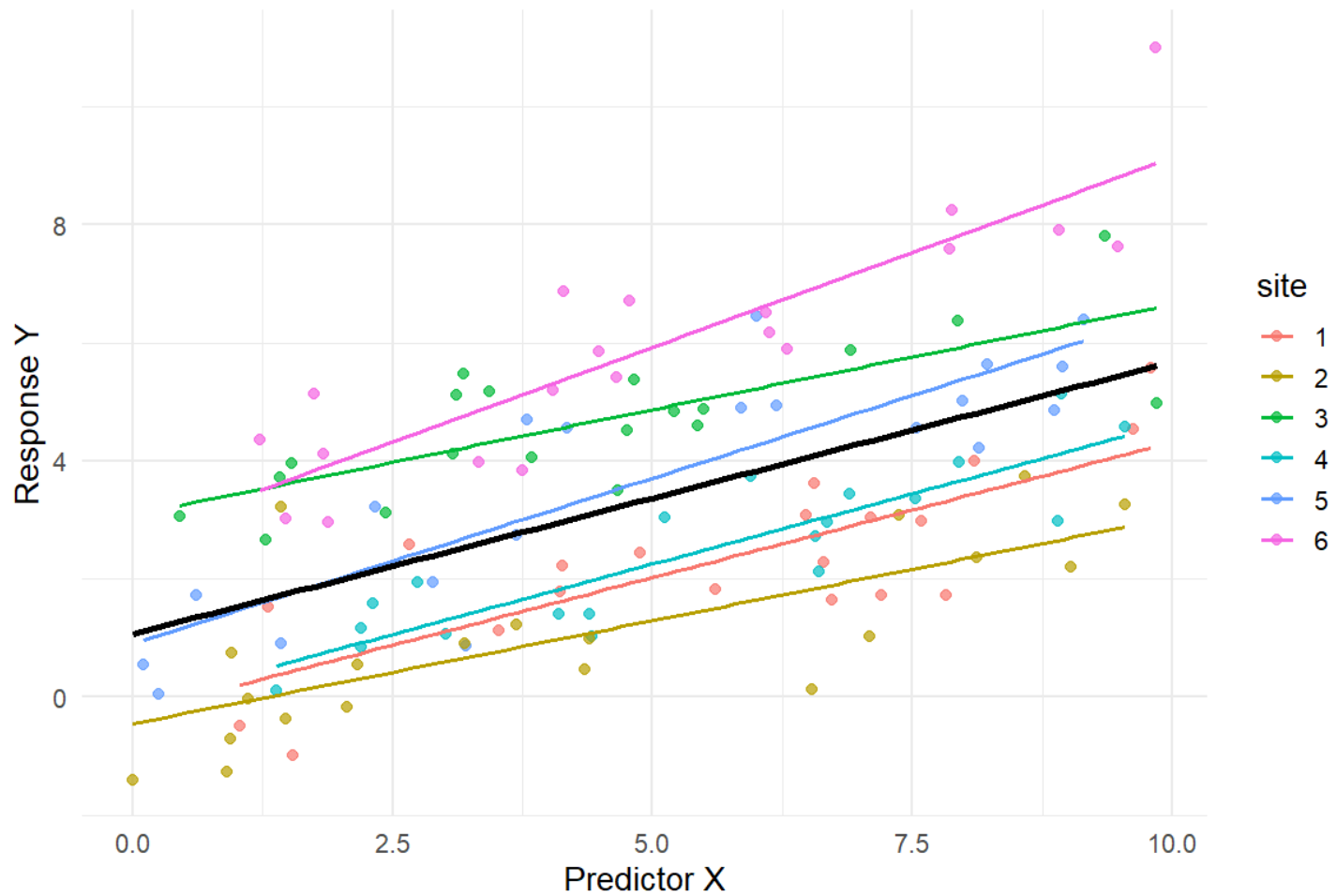
- Many datasets have a nested or grouped structure:
 - Students within schools
 - Patients within hospitals
 - Plots within sites
- Observations within a group tend to be more similar than between groups.
- Ignoring this structure can lead to misleading conclusions.

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

- Overconfident inference: artificially small p-values.
- Incorrect effect size estimates.
- Conclusions may not hold for individual groups.

Fixed Effects

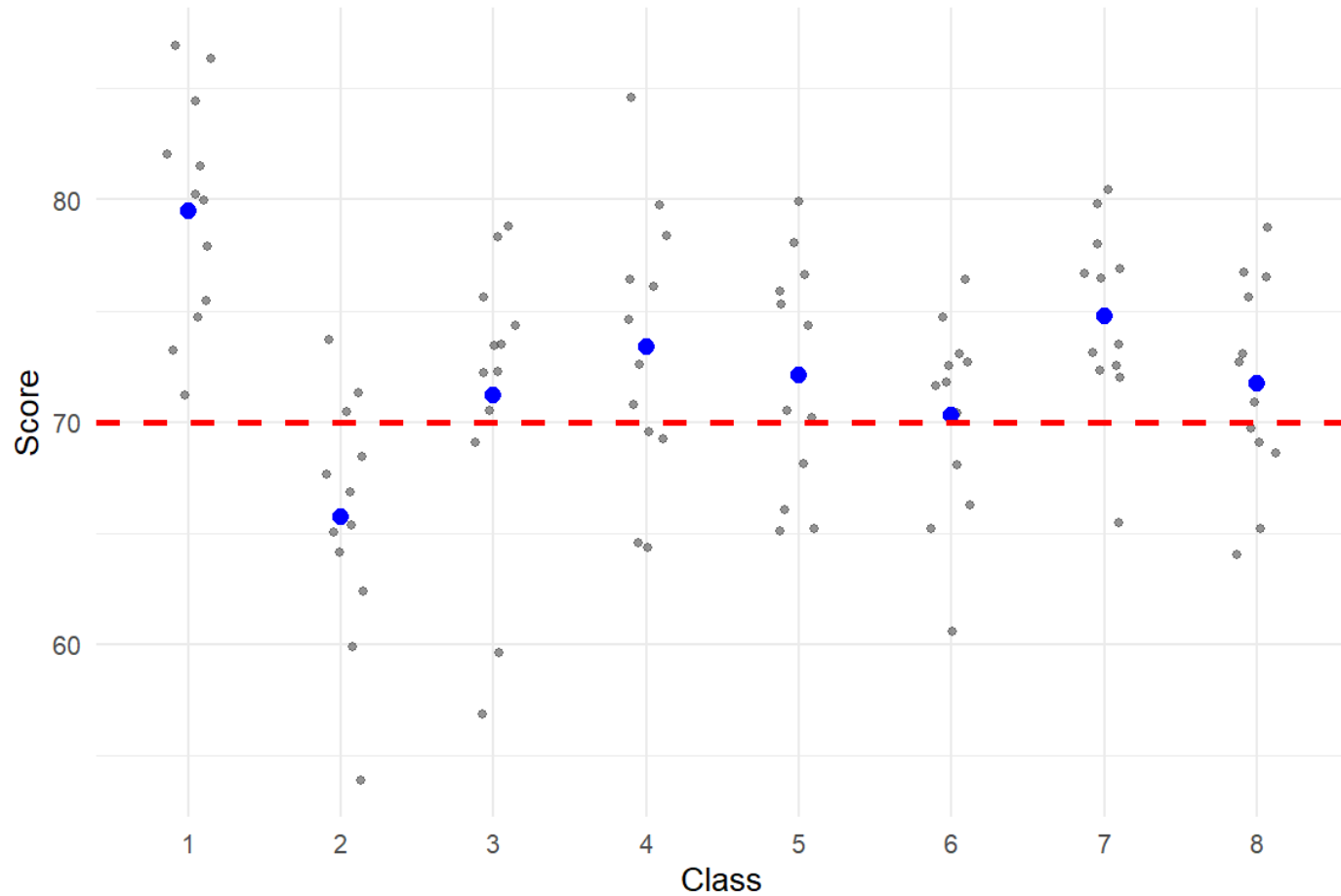
- Parameters associated with entire population or specific experimental conditions.
- Represent the average relationship across all groups.
- Example: Overall effect of temperature on bird abundance.

Random Effects

- Capture group-specific deviations from the overall mean or slope.
- Treated as a random sample from a larger population.
- Example: Site-specific baseline abundance.

Class Exam Scores

- Fixed effect: mean score across all classes.
- Random effect: deviation of each class mean from overall mean.
- Accounts for natural variation between classes.



Fixed Effects Models

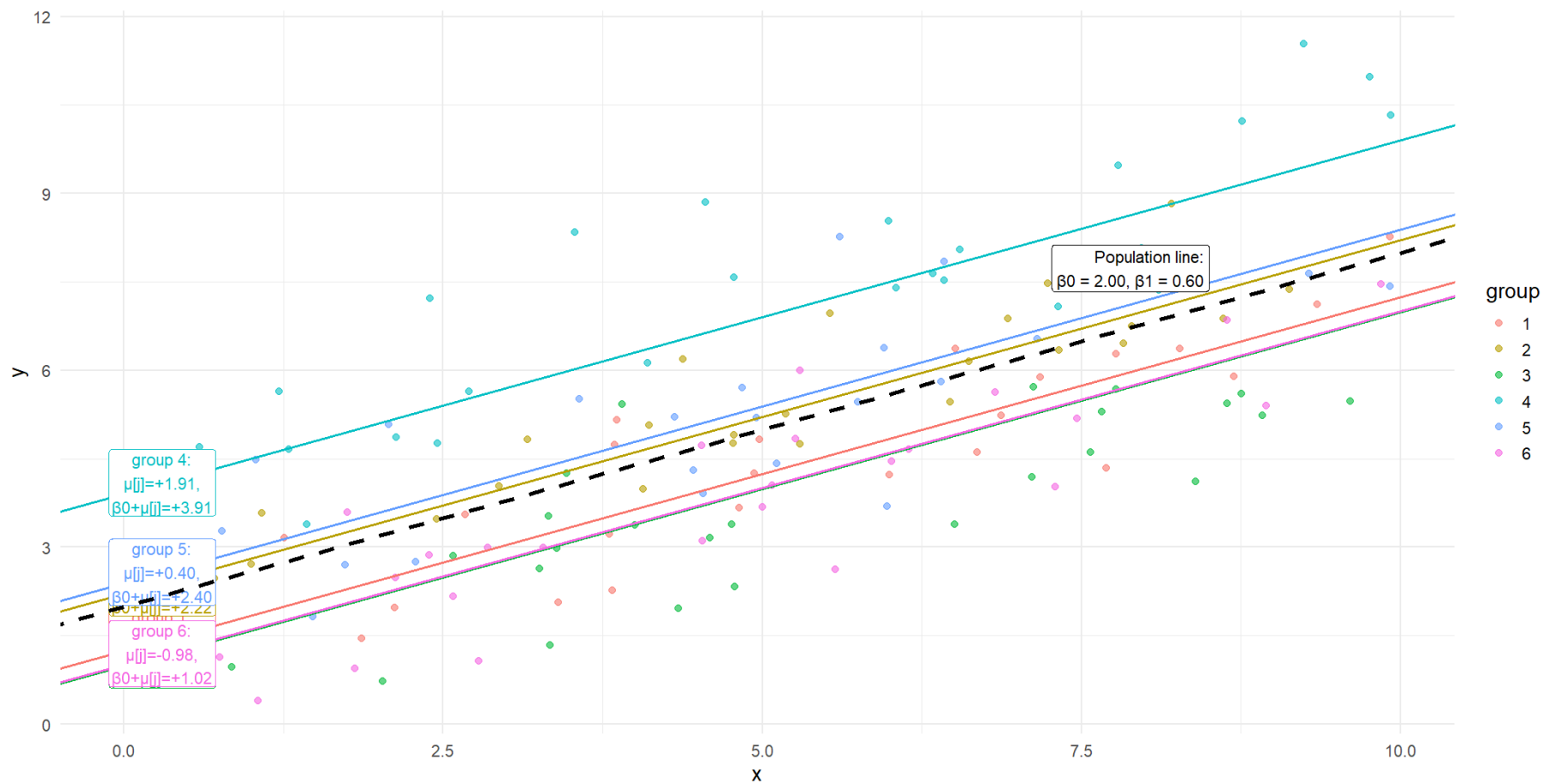
$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

- 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 $\mu_i = \beta_0 + \beta_1 x_i$
 $Y_i \sim \text{Normal}(\mu_i, \sigma^2)$

Random Intercept Models

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

- Y_{ij} : response for observation i in group j
- β_0 : overall intercept
- $a_{0j} \sim N(0, \sigma_a^2)$: random intercept for group j
- β_1 : fixed slope



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



PR
STATS

Introduction to Generalised Linear Mixed Models
Dr Niamh Mimmagh, PR Stats

Model Output: Fixed Effects

- Intercept: mean when predictors = 0.
- Slope: average change in response for one-unit change in predictor.

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 not 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.

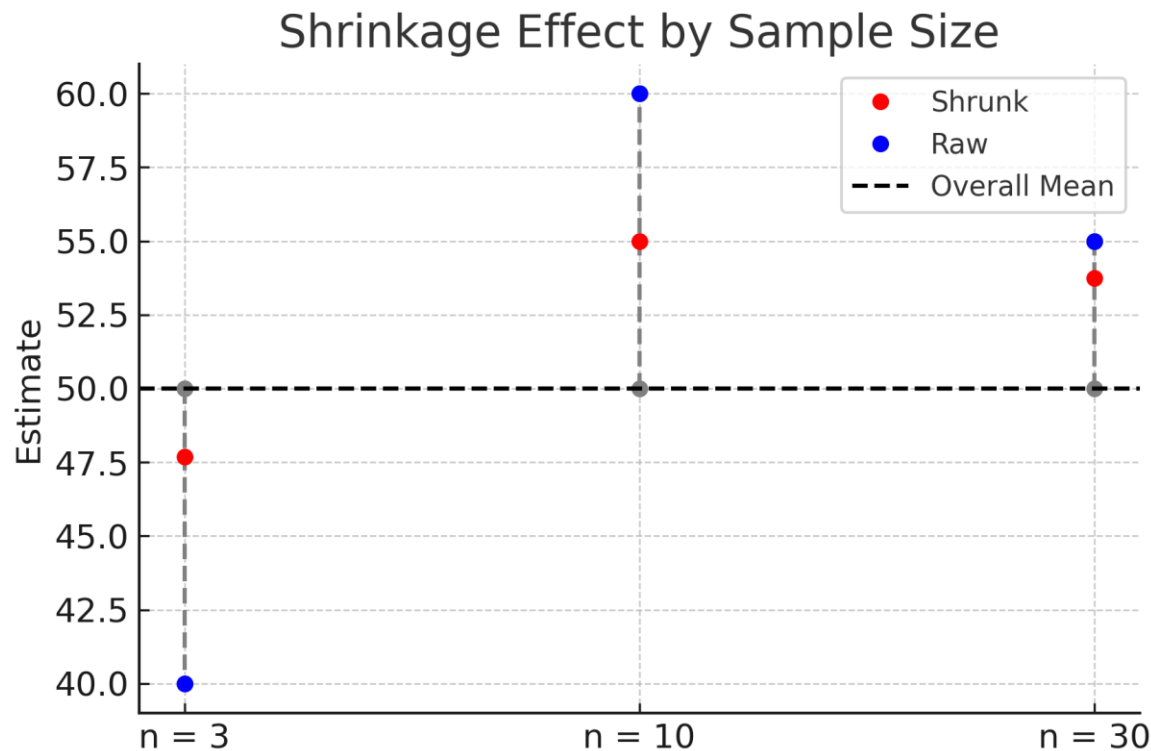
Shrinkage in Random Effects

- 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



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

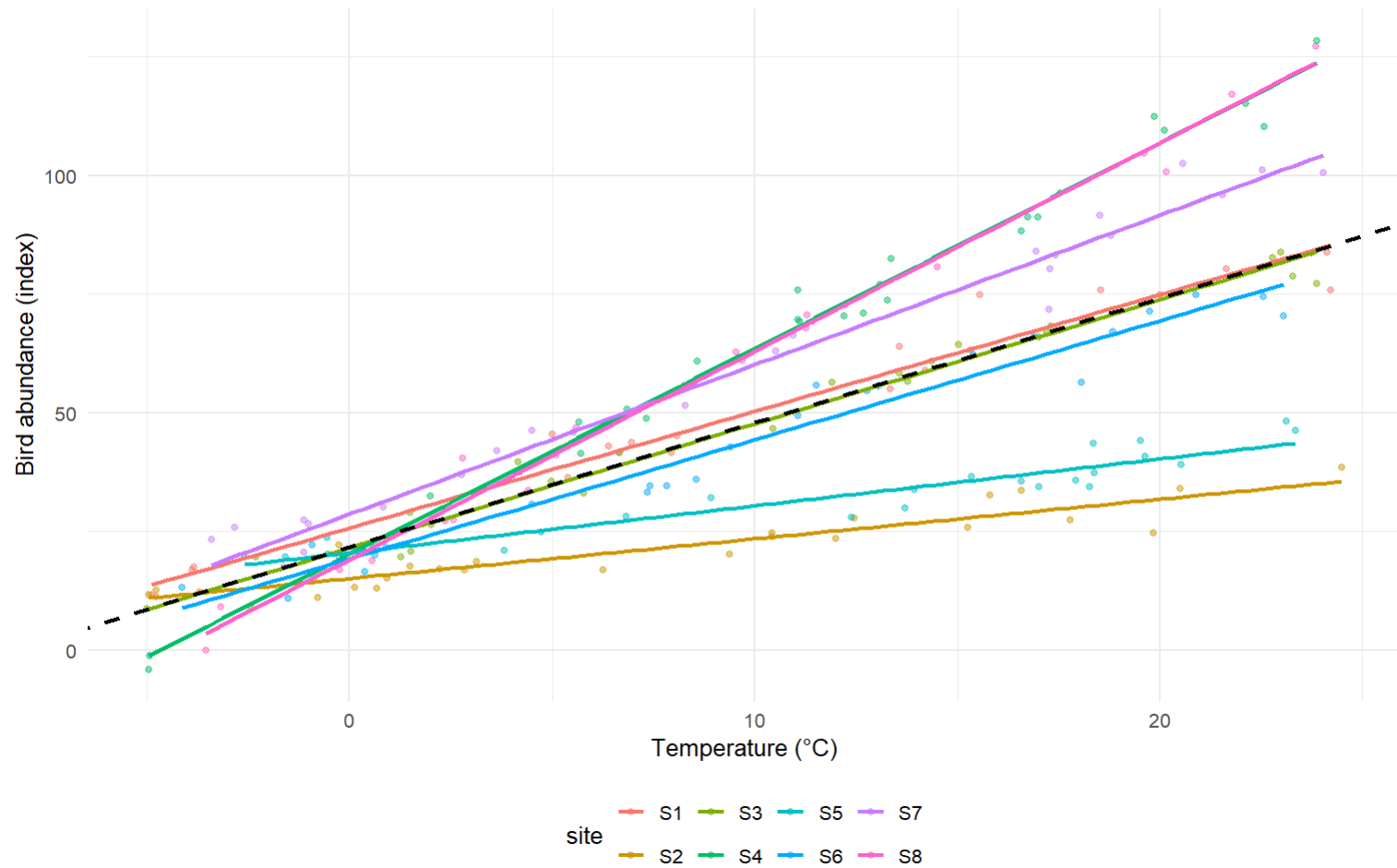
- What is maximised:
 - ML: Likelihood of full data given all parameters (estimates fixed effects and variance components together)
 - REML: Likelihood of residuals (data after removing fixed effects)
- Variance components:
 - ML: Biased downward (esp. in small samples)
 - REML: Unbiased
- Model comparison: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

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

- Underestimates uncertainty in predictions.
- Can bias fixed effect estimates if slopes vary systematically with group characteristics.

Random Slopes Model

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

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

- 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

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

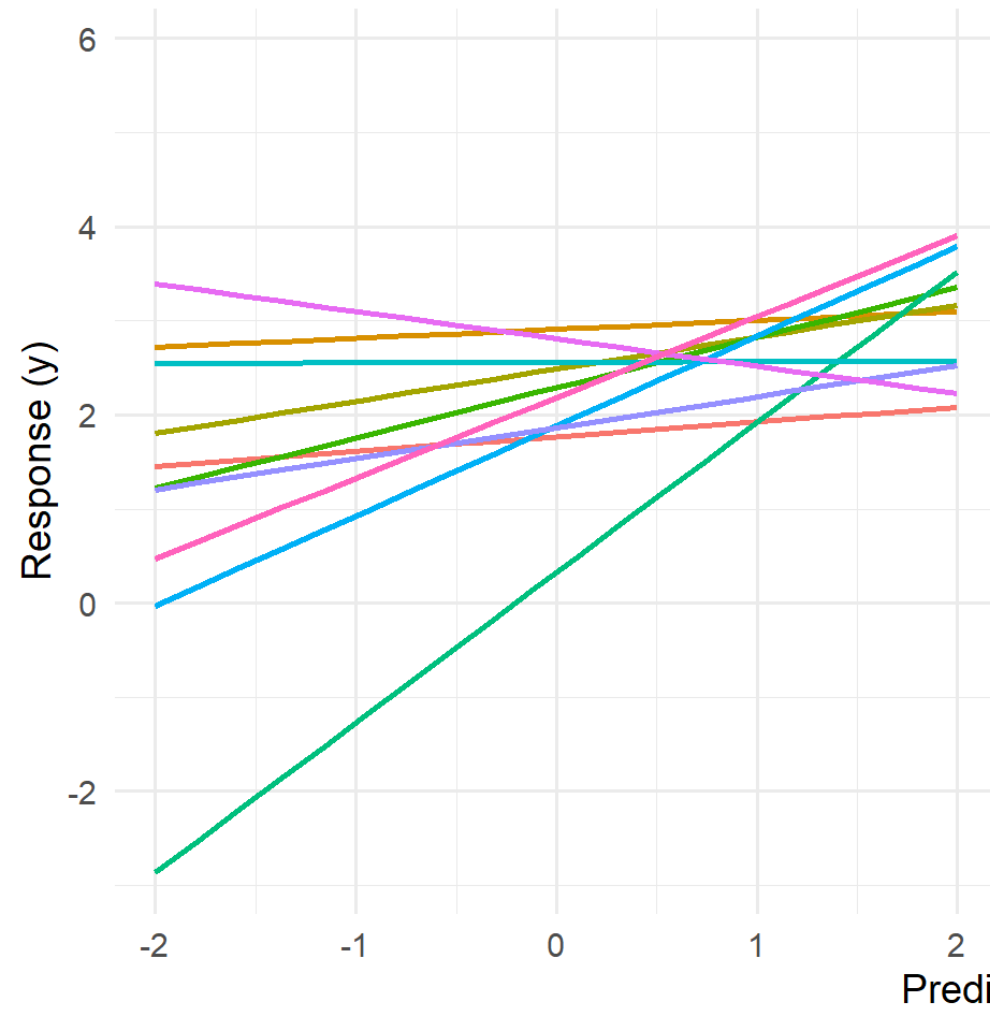
- Intercept and slope deviations are jointly normally distributed across groups. The mean is (0,0) because the fixed effects already capture the overall average intercept and slope.

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

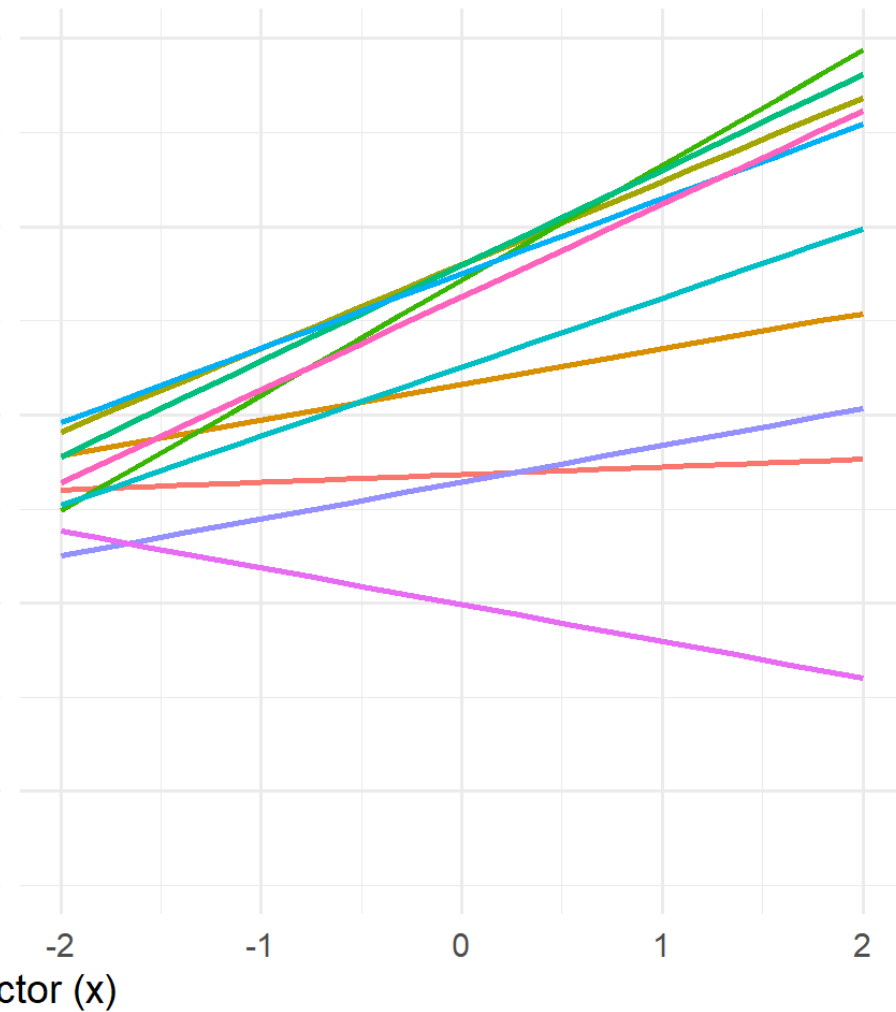
- ρ lies between -1 and +1. If $\rho > 0$, groups with higher intercepts tend to have more positive slopes. If $\rho < 0$, groups with higher intercepts tend to have weaker or more negative slopes.



Negative correlation ($\rho = -0.7$)



Positive correlation ($\rho = +0.7$)



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 :

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

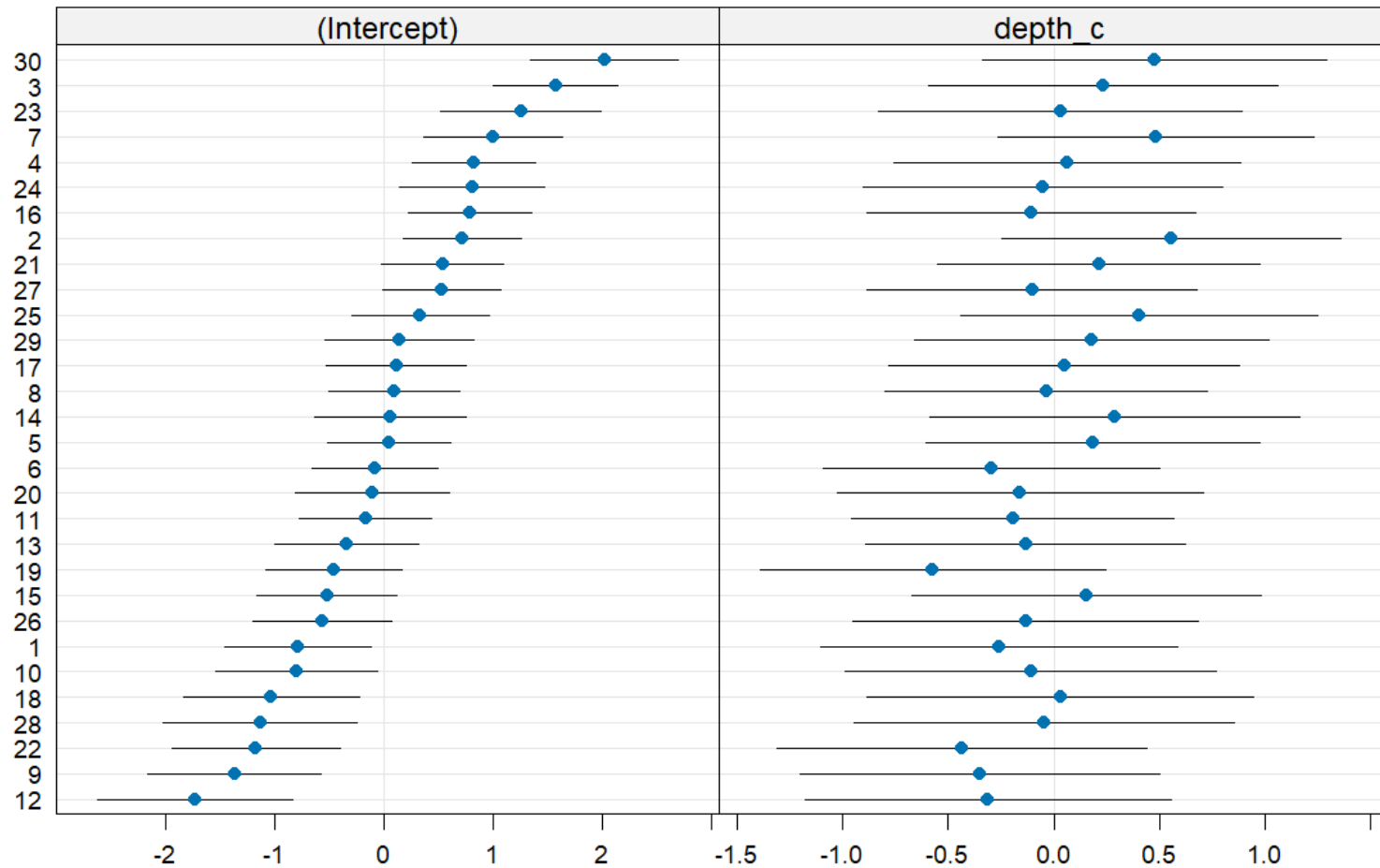
ICC at x :

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

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 & 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 Mixed 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

$$Y_{ij} \sim \text{Normal}(\mu_{ij}, \sigma^2)$$
$$\underbrace{\mu_{ij}}_{\text{Link}} = \underbrace{\beta_0 + \beta_1 x_i + a_{0j}}_{\text{Systematic Component}}$$

Link
Function: Systematic
Component

The identity link ($g(\mu_i) = \mu_i$)

Common GLMM Families and Link Functions

- Binomial - binary/proportion data (logit link).
- Poisson - count data (log link).
- Negative binomial - overdispersed counts.
- A link function connects the mean of the response to the linear predictor. It ensures that predictions stay in an allowed range.
 - Identity: $g(\mu) = \mu$
 - Logit: $g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$
 - Log: $g(\mu) = \log(\mu)$

Binary Data

Many real-world problems involve binary (yes/no, success/failure) outcomes:

- Did a patient survive? (yes/no)
- Was the animal infected? (yes/no)
- Did the student pass the course? (yes/no)

$$Y_i = \begin{cases} 1, & \text{if success} \\ 0, & \text{if failure} \end{cases}$$

What is 'success'?

It's whatever you want it to be! It's not necessarily the 'best' outcome.

Binary Data

When data is binary, there are only two possible outcomes, and so when we talk about the probability of each outcome, we have:

$$\begin{aligned}P(\text{success}) &= P(Y_{ij} = 1) = \pi_{ij} \\P(\text{failure}) &= P(Y_{ij} = 0) = 1 - \pi_{ij}\end{aligned}$$

Y_{ij} has a Bernoulli distribution:

$$Y_{ij} \sim \text{Bernoulli}(\pi_{ij})$$

Why Not Use a Linear Model?

Linear regression assumes:

- The response variable is continuous and unbounded.
- The relationship between predictors and the response is linear.

Problems with using it on binary data:

- Predictions can fall outside $[0,1]$
- Error terms are heteroscedastic (non-constant variance)
- Residuals are not normally distributed

This leads to poor model performance and invalid inference.

The Bernoulli GLMM

$$Y_{ij} \sim \text{Bernoulli}(\pi_{ij})$$

We want to model the success probabilities π_i as a function of predictors.

Can we simply write $\pi_{ij} = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi} + a_0$?

No!

Since the β coefficients are unbounded (they can take any real value from $-\infty$ to ∞), this would result in unbounded π_i values

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

So we need a link function that maps the (0,1) interval to the real line.

The Bernoulli GLMM

$$Y_i \sim \text{Bernoulli}(\pi_i)$$
$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi}$$

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

‘logit’ stems from the words **logistic unit**, since its based on the cumulative distribution function of the logistic distribution

It is simply the natural logarithm of the odds

This is how we ensure predicted probabilities stay between 0 and 1.

Probability vs. Odds

- Probability (π) is the chance of an event occurring (range: 0 to 1)
e.g., $\pi = 0.8$ means an 80% chance of success
- Odds compare the probability of success to the probability of failure.

$$Odds = \frac{\pi}{1 - \pi}$$

- For example, if $\pi = 0.8$ then $odds = \frac{0.8}{0.2} = 4$ (4 to 1 odds of success)

$$\pi = \frac{Odds}{1 + Odds}$$

$$\text{If odds} = 4, \text{ then } \pi = \frac{4}{1+4} = \frac{4}{5} = 0.8$$

Interpreting Fixed-Effect Coefficients

$$\text{logit}(\pi) = \beta_0 + \beta_1 x = -0.5 + 1.2x$$

- $\beta_0 = -0.5$: when $x = 0$, the log-odds of success are -0.5 .
- Odds = $e^{-0.5} \approx 0.607$: when $x = 0$, the odds of success are 0.61 to 1 (success is less likely than failure).
- $\pi = \frac{0.607}{1+0.607} \approx 0.38$: when $x = 0$, the probability of success is about 38%.
- $\beta_1 = 1.2$: for each 1-unit increase in x , the log-odds increase by 1.2.
- Odds ratio = $e^{1.2} \approx 3.32$: each 1-unit increase in x multiplies the odds of success by about 3.3.
- At $x = 1$: $\text{logit}(\pi) = -0.5 + 1.2(1) = 0.7 \rightarrow \text{Odds} = e^{0.7} \approx 2.01$
- $\pi = \frac{2.01}{1+2.01} \approx 0.67$: increasing x by one unit raises probability from 38% to 67%.

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:

$$Y_i \sim \text{Bernoulli}(\pi_i)$$
$$\text{logit}(\pi_i) = \beta_0 + \beta_1 \text{depth}_i$$

Where:

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

π_i = probability of frog presence at site i

β_0 = 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:

$$\begin{aligned}Y_{ij} &\sim \text{Bernoulli}(\pi_{ij}) \\ \text{logit}(\pi_{ij}) &= \beta_0 + \beta_1 \text{depth}_i + a_j \\ a_j &\sim N(0, \sigma_{\text{region}}^2)\end{aligned}$$

Where:

Y_{ij} = presence at site i in region j

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

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

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

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

- Random slopes add more parameters which makes them harder models to fit, and they are more likely to cause convergence issues.
- Try starting with a random intercept model, then test slope variation.

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).

Further Topics

- A wider range of GLMM families: moving beyond binomial models to Poisson, negative binomial, and zero-inflated models for ecological count data.
- Complex diagnostics and model evaluation: including overdispersion checks, DHARMA residuals, and comparing models with REML, ML, and AIC.
- Bayesian multilevel modelling: why and when to use Bayesian approaches, fitting models with brms
- Full case studies: working through realistic ecological datasets combining multiple random effects, complex predictors, and mixed outcome types.



Thank You

- Thank you for attending this introduction to generalised linear mixed models course. I hope this session has helped you:
 - Understand when and why to use mixed models.
 - Interpret fixed and random effects in the GLMM framework
 - Visualise and communicate model results effectively.
- This was just the starting point. In the full 5-day module, we build on these foundations with:
 - More GLMM families for different data types.
 - More diagnostics and model comparison tools
 - Practical case studies and Bayesian approaches.
- If you'd like to explore any of these topics further, keep an eye out for upcoming courses!

