9. Random regression models (random slopes)
Random intercepts model

\[ R_{ij} = (\beta_0 + \beta_{0j}) + \beta_1 \times NAP_{ij} + \epsilon_{ij} \]

- \( R_{ij} \) = individual observation (coded for Beach ID)
- \( \beta_0 \) = grand intercept
- \( \beta_1 \) = overall relationship between Richness & NAP
- \( \beta_{0j} \) = variation in the intercepts among the beaches

\( \beta_0 \) = grand intercept
\( \beta_{0j} \) = variation in the intercepts among the beaches
\( \beta_1 \) = overall relationship between Richness & NAP
Random slopes (& intercepts) model

\[ R_{ij} = (\beta_0 + \beta_{0j}) + (\beta_1 + \beta_{1j}) \times NAP_{ij} + \epsilon_{ij} \]
Linear model

\[ y_{ij} = \beta_0 + \beta_1 \times x_{ij} + \epsilon_{ij} \]

\[ \epsilon_{ij} \sim N(0, \sigma_e^2) \]
Linear mixed-effect model (random intercepts)

\[ y_{ij} = (\beta_0 + \beta_{0j}) + \beta_1 \times x_{ij} + \epsilon_{ij} \]

\[ \epsilon_{ij} \sim N(0, \sigma_e^2) \]

\[ \beta_{0j} \sim N(0, \sigma_{int}^2) \]
Linear *mixed-effects* model
(random intercept and slope model, aka. random regression)

\[ y_{ij} = (\beta_0 + \beta_{0j}) + (\beta_1 + \beta_{1j}) x_{1ij} + \varepsilon_{ij} \]

- \( \beta_0 + \beta_{0j} \): intercept
- \( \beta_1 + \beta_{1j} \): slope of \( x_1 \)
- \( \varepsilon_{ij} \): error

\[ \varepsilon_{ij} \sim N(0, \sigma^2) \]
\[ \beta_{0j} \sim N(0, \sigma_{int}^2) \]
\[ \beta_{1j} \sim N(0, \sigma_{slope[1]}^2) \]
When do we use random regression?

• When we are interested in a general trend of the relationship (intercept and slope).

• But when observations are not independent, and slopes are not equal among groups.

• When we are interested in whether among-group slopes are equal.
RIKZ data set

• Include random slopes in the RIKZ data set
Do it on your own!

- Use the Cetaceans dataset and work through the model selection
- Start with this model:
  - \( \text{Age} \sim \text{fStain} \times \text{fSex} + \text{fStain} \times \text{fLocation} + + \)
  - \((1 | \text{fSpecies/fDolphinID})\)
- Try to validate this model and see what needs to be changed about the random structure
- Then finish the model selection (Stain is our fixed effect of interest, i.e. ‘treatment’)

Another variance het data set

• The ‘Biodiversity’ data set has similar variance heterogeneity problems

• You can work through this dataset too if you want
  – Concentration is the response variable
  – Biomass, Treatment, Nutrient are all predictors
  – Just make sure to make Treatment and Nutrient levels *factors* first
BONUS – generalized linear mixed models

• Data (specifically, the residuals from your data) are often not normally distributed
  – Poisson (counts)
  – Binomial (presence/absence)
  – Negative binomial
  – Gamma

• Mixed models can handle these too

• But this is an area of active research and there is lots of disagreement about the best way to run these models

• Read Chapter 13 in Zuur!
BONUS – generalized linear mixed models

• Packages that can run GLMMs (there are others):
  – lme4 (using the glmer() command)
  – glmmPQL
  – glmmML

• Unfortunately, because behind-the-scenes math is not agreed upon, they can sometimes give different results (& so be very cautious about strongly believing results close to $p = 0.05$)
Really impressive supplementary too, with worked examples!
Using observation-level random effects to model overdispersion in count data in ecology and evolution

Xavier A. Harrison
Institute of Zoology, Zoological Society of London, London, UK

ABSTRACT
Overdispersion is common in models of count data in ecology and evolutionary biology, but researchers have recently highlighted how it can also be a feature of models of binomial data. Here we present a comparison of two types of model designed to account for overdispersion in count data: a discrete negative binomial model and a continuous Beta-Binomial model.

A comparison of observation-level random effect and Beta-Binomial models for modelling overdispersion in Binomial data in ecology & evolution

Xavier A. Harrison
Institute of Zoology, Zoological Society of London, UK

ABSTRACT
Overdispersion is a common feature of models of biological data, but researchers have recently highlighted how it can also be a feature of models of binomial data. Here we present a comparison of two types of model designed to account for overdispersion in binomial data: a continuous Beta-Binomial model and a discrete negative binomial model.
Analyzing count data in R

1. Plot your data and calculate summary statistics.
2. Is there zero inflation? (Yes/No)
   - Yes: What type of zeros do you have?
     - Only true zeros: Run ZAP/ZANB models using the 'hurdle' function in the 'pscl' package.
     - True zeros & false zeros: Run ZIP/ZINB models using the 'zeroinfl' function in the 'pscl' package.
   - No: Try a Poisson GLM. Is there overdispersion? (Yes/No)
     - Yes: Run a quasipoisson GLM. Is there still overdispersion? (Yes/No)
       - Yes: Use a negative binomial GLM. Success!
       - No: Lucky you, you're done!
     - No: The P model is probably more appropriate.
3. Is there overdispersion in the non-zero part of the data? (Yes/No)
   - Yes: The NB model is probably more appropriate.
   - No: Compare P/NB models using likelihood ratio tests and AIC values.

*True (structural) zeros arise when there is nothing to count. False (sampling) zeros arise when there could be things to be counted but they were not detected.*

From: Natalie Pilakouta