CPUE standardisation

Carl Smith
Centre for Analysis, Modelling and Computational Science
University of Łódź
Poland

carl.smith@biol.uni.lodz.pl

Organisation

- 2 days
- 8 sessions
- Lots of breaks

- Interrupt
- Ask questions
- Discuss

Czech bitterling

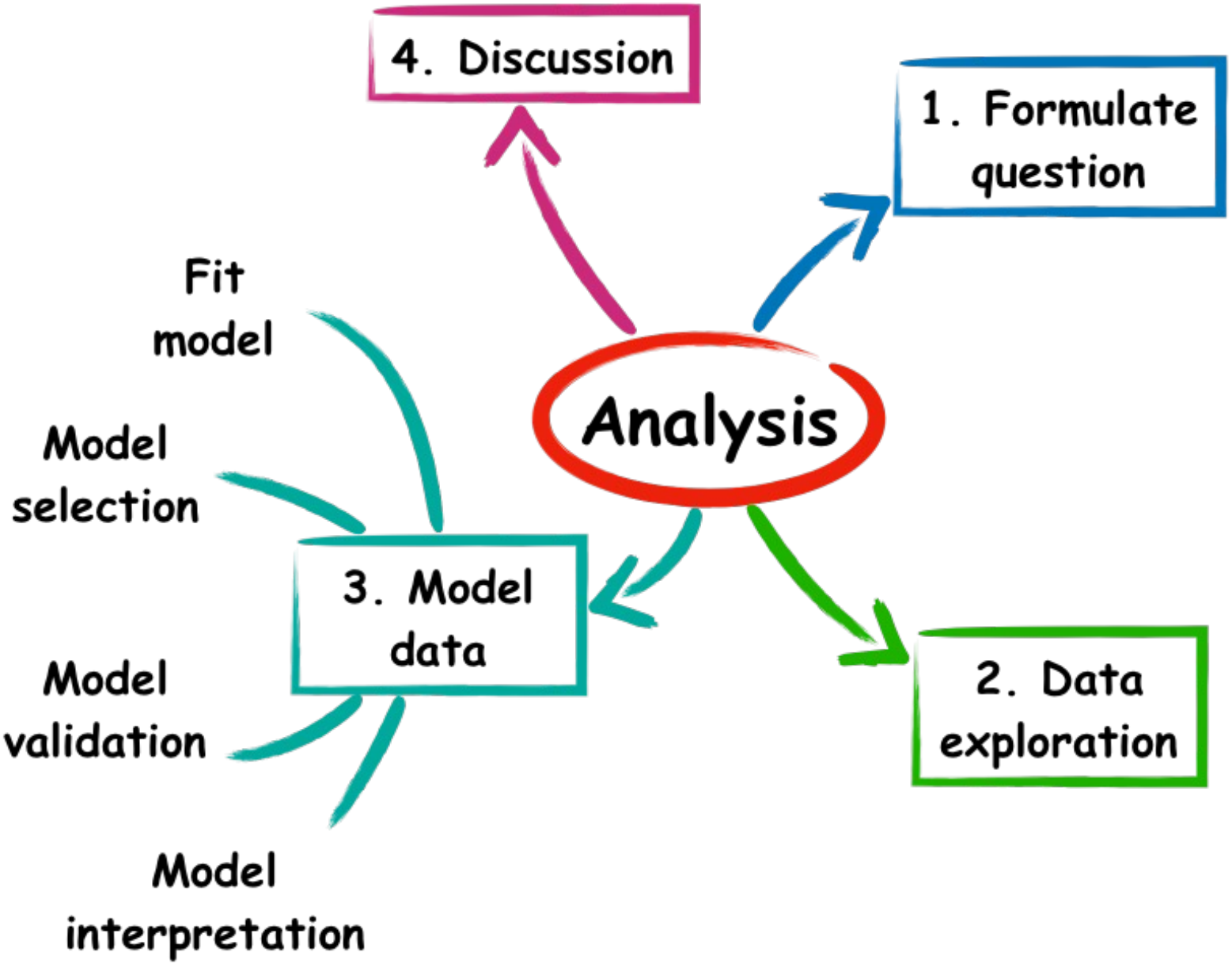
Data set (*bitterling.csv*) comprising:

- **Catch** of bitterling as number of fish
- **Effort** as number of sampling points (points)
- Variation due to different **habitat** (vegetated, unvegetated), **season** (spring, autumn), **year** (1995-1998)
- Samples from 8 different **oxbow** lakes
- R Script in file *Bitterling.R*

Approach

1. Formulate the question

What variables predict variance in bitterling catches?


4. Discussion
1. Formulate question
Fit model
Model selection
Analysis
3. Model data
2. Data exploration
Model validation
Model interpretation

CPUE standardisation

Multiple samples from 8 oxbow lakes

- Catches *within* oxbow lakes will be more similar than *among* lakes
- We cannot treat samples from the same lake as ‘independent’
- ‘Pseudoreplication’
- We have **dependency**

Dependency

- A common feature of biological data
- *Correlation of observations within a group*
- Analyses assume independence
- Dependency due to:
  - Repeated measures of same individual
  - Membership of a group
  - Experimental nesting
- *Fixed* effects and *random* effects

Dependency

- *Fixed* effect
  - Unknown constant
  - Try to estimate from data
- *Random* effect
  - No benefit to estimate
  - Instead, estimate parameters for its distribution
- Sometimes difficult to decide if a variable should be classed as random or fixed

Terminology

- Misleading
  - *Random* does not imply random!
  - Think of random effects as grouping variables
- Mixed effect models (GLMMs, GAMMs) are really just hierarchical models
- Some special types of dependency
  - Spatial
  - Temporal
  - Both common in ecological/fisheries data

Is it a fixed or random effect?

- Much debated by statisticians
- Only use for factors/categories
  - Not continuous data
- Is there dependency?
  - In the data
  - In the design
- Number of levels
  - >4 (ideally >10), treat as random
  - 2 or 3 levels, treat as fixed
- Don’t use the same variable as fixed and random in the same model (except in very special cases)

Drawback of mixed models

The P-values are not reliable!

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

(We will use a Bayesian approach with zander data)

Dependency by design

- Catches from the same oxbow more similar than adjacent lake
- Why?
  - Food, habitat, mussels, predators, water quality, etc.
- Any other dependency in the data?
  - Year
- Population size in 1996 not independent of size in 1995
- But only 4 years of data...so treat as fixed effect (not ideal)

Model these data with GLMM