GLM Model Selection and Validation

Cóilín Minto, Olga Lyashevska

Marine and Freshwater Research Centre
Atlantic Technological University
Galway, Ireland

July 15th 2022

Ollscoil
Teicneolaíochta
an Atlantaigh

Atlantic
Technological
University

Outline

All models are wrong but some are useful

- Statistical models are mathematical approximations to reality that represent the important features of data for the task at hand.

All models are wrong but some are useful

- Statistical models are mathematical approximations to reality that represent the important features of data for the task at hand.
- The purpose of a statistical model effects how it is developed: prediction versus interpretation

All models are wrong but some are useful

- Statistical models are mathematical approximations to reality that represent the important features of data for the task at hand.
- The purpose of a statistical model effects how it is developed: prediction versus interpretation
- For any set of data, there are numerous components that could be chosen. How do we choose a statistical model? Statistical models are based on underlying theory, or from an understanding of the biological features, and are built with this knowledge in mind.

Criteria

An adequate statistical model balances two criteria:

- **Accuracy:** The model should accurately describe both the systematic (fixed) and random components.
- **Parsimony:** The model should be as simple as possible. The simplest accurate model is the preferred model. Complex models may fit the given data well but usually do not generalize well to other data sets (over-fitting).

Outline

Outline

Outline

Four principles for improved statistical ecology

ISEC 2022, Session 18. Statistical Theory. Gordana Popovic

1. First define a focused research question, then plan sampling and analysis to answer it;
2. Develop a model that accounts for the distribution and characteristics (dependencies) of your data;
3. Emphasise effect sizes to replace statistical significance (p-values) with ecological relevance;
4. Report you methods and finding in sufficient detail so that your research is valid and reproducible;