

# Fitting Models to Data in Ecology and Evolution

Samraat Pawar

*Department of Life Sciences (Silwood Park)*

**Imperial College**  
**London**

November 6, 2023

# MECHANISTIC VS. PHENOMENOLOGICAL MODELS

*What does “modelling data” mean to you?*

# MECHANISTIC VS. PHENOMENOLOGICAL MODELS

- *Mechanistic models* aim to explain the PROCESSES or MECHANISMS underlying PATTERNS or PHENOMENA in empirical data

# MECHANISTIC VS. PHENOMENOLOGICAL MODELS

- *Mechanistic models* aim to explain the PROCESSES or MECHANISMS underlying PATTERNS or PHENOMENA in empirical data
  - These models have a THEORETICAL BASIS

# MECHANISTIC VS. PHENOMENOLOGICAL MODELS

- *Mechanistic models* aim to explain the PROCESSES or MECHANISMS underlying PATTERNS or PHENOMENA in empirical data
  - These models have a THEORETICAL BASIS
- *Empirical/Phenomenological models* establish the existence of STATISTICALLY SIGNIFICANT, NON-RANDOM PATTERNS or PHENOMENA in empirical data

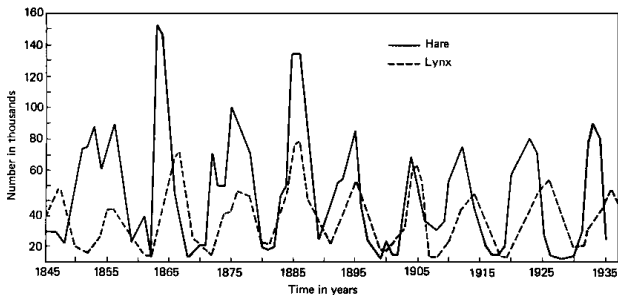
# MECHANISTIC VS. PHENOMENOLOGICAL MODELS

- *Mechanistic models* aim to explain the PROCESSES or MECHANISMS underlying PATTERNS or PHENOMENA in empirical data
  - These models have a THEORETICAL BASIS
- *Empirical/Phenomenological models* establish the existence of STATISTICALLY SIGNIFICANT, NON-RANDOM PATTERNS or PHENOMENA in empirical data
  - They make no assumptions about the processes or mechanisms that generate the patterns

# MECHANISTIC VS. PHENOMENOLOGICAL MODELS

- *Mechanistic models* aim to explain the PROCESSES or MECHANISMS underlying PATTERNS or PHENOMENA in empirical data
  - These models have a THEORETICAL BASIS
- *Empirical/Phenomenological models* establish the existence of STATISTICALLY SIGNIFICANT, NON-RANDOM PATTERNS or PHENOMENA in empirical data
  - They make no assumptions about the processes or mechanisms that generate the patterns
  - That is, these models lack a THEORETICAL BASIS

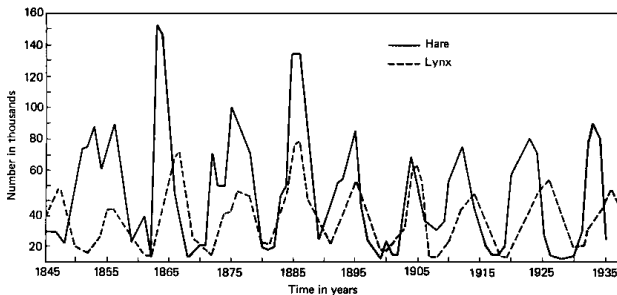
# MECHANISTIC VS. PHENOMENOLOGICAL MODEL FITTING



source: <https://www.cds.caltech.edu/~murray/amwiki/images/8/8f/LHgraph.gif>



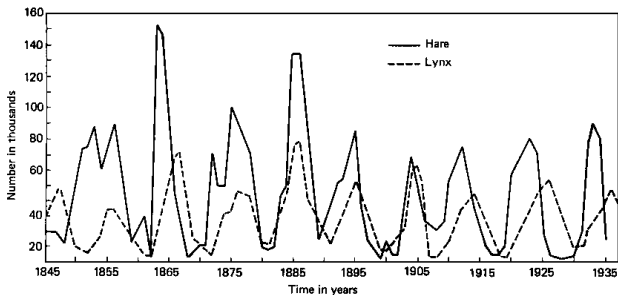
# MECHANISTIC VS. PHENOMENOLOGICAL MODEL FITTING



source: <https://www.cds.caltech.edu/~murray/amwiki/images/8/8f/LHgraph.gif>

- **Mechanistic model:** *The Lynx-Hare Cycle is driven by density-dependent population growth in hares*

# MECHANISTIC VS. PHENOMENOLOGICAL MODEL FITTING



source: <https://www.cds.caltech.edu/~murray/amwiki/images/8/8f/LHgraph.gif>

- **Mechanistic model:** *The Lynx-Hare Cycle is driven by density-dependent population growth in hares*
- **Phenomenological model:** *The Lynx and Hare Cycles have a significant asynchrony (period shift) of  $x$  years*

# MECHANISTIC VS. PHENOMENOLOGICAL MODEL FITTING

- *It's not really one vs. the other*; Both types of models play a role in science (and Biology)

# MECHANISTIC VS. PHENOMENOLOGICAL MODEL FITTING

- *It's not really one vs. the other*; Both types of models play a role in science (and Biology)
- Phenomenological model-fitting reveals patterns in data that generate HYPOTHESES

# MECHANISTIC VS. PHENOMENOLOGICAL MODEL FITTING

- *It's not really one vs. the other*; Both types of models play a role in science (and Biology)
- Phenomenological model-fitting reveals patterns in data that generate HYPOTHESES
  - These can be tested using further model fitting

# MECHANISTIC VS. PHENOMENOLOGICAL MODEL FITTING

- *It's not really one vs. the other*; Both types of models play a role in science (and Biology)
- Phenomenological model-fitting reveals patterns in data that generate HYPOTHESES
  - These can be tested using further model fitting
  - Example: *Whether* climatic temperature affects the Lynx-Hare cycle (using Generalized Linear Model-fitting)

# MECHANISTIC VS. PHENOMENOLOGICAL MODEL FITTING

- *It's not really one vs. the other*; Both types of models play a role in science (and Biology)
- Phenomenological model-fitting reveals patterns in data that generate HYPOTHESES
  - These can be tested using further model fitting
  - Example: *Whether* climatic temperature affects the Lynx-Hare cycle (using Generalized Linear Model-fitting)
- Mechanistic model-fitting *tries* to validate a mechanistic model that can explain the observed phenomenological pattern and generate MORE ACCURATE, MECHANISTIC HYPOTHESES

# MECHANISTIC VS. PHENOMENOLOGICAL MODEL FITTING

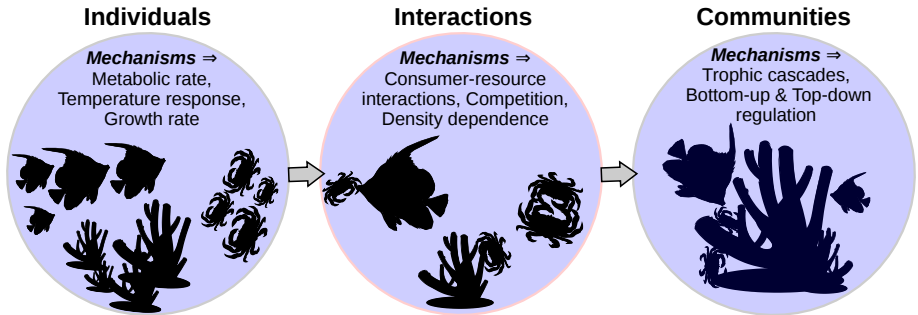
- *It's not really one vs. the other*; Both types of models play a role in science (and Biology)
- Phenomenological model-fitting reveals patterns in data that generate HYPOTHESES
  - These can be tested using further model fitting
  - Example: *Whether* climatic temperature affects the Lynx-Hare cycle (using Generalized Linear Model-fitting)
- Mechanistic model-fitting *tries* to validate a mechanistic model that can explain the observed phenomenological pattern and generate MORE ACCURATE, MECHANISTIC HYPOTHESES
  - Example: *How* climatic temperature *drives* the Lynx-Hare cycle



# MECHANISTIC VS. PHENOMENOLOGICAL MODEL FITTING

- *It's not really one vs. the other*; Both types of models play a role in science (and Biology)
- Phenomenological model-fitting reveals patterns in data that generate HYPOTHESES
  - These can be tested using further model fitting
  - Example: *Whether* climatic temperature affects the Lynx-Hare cycle (using Generalized Linear Model-fitting)
- Mechanistic model-fitting *tries* to validate a mechanistic model that can explain the observed phenomenological pattern and generate MORE ACCURATE, MECHANISTIC HYPOTHESES
  - Example: *How* climatic temperature *drives* the Lynx-Hare cycle
- *Ultimately, successful, EMPIRICALLY-GROUNDED mechanistic models are the best path towards a THEORY in any scientific discipline (including ecology and evolution)*

# MECHANISTIC VS. PHENOMENOLOGICAL MODEL FITTING



# MECHANISTIC MODELS IN ECOLOGY AND EVOLUTION?

- *Do most ecological studies perform phenomenological or mechanistic modelling (or model-fitting)?*

# MECHANISTIC MODELS IN ECOLOGY AND EVOLUTION?

- *Do most ecological studies perform phenomenological or mechanistic modelling (or model-fitting)?*
- The answer is mostly Phenomenological

# MECHANISTIC MODELS IN ECOLOGY AND EVOLUTION?

- *Do most ecological studies perform phenomenological or mechanistic modelling (or model-fitting)?*
- The answer is mostly Phenomenological

# MECHANISTIC MODELS IN ECOLOGY AND EVOLUTION?

- *Do most ecological studies perform phenomenological or mechanistic modelling (or model-fitting)?*
- The answer is mostly Phenomenological — *Why?*

# MECHANISTIC MODELS IN ECOLOGY AND EVOLUTION?

- *Do most ecological studies perform phenomenological or mechanistic modelling (or model-fitting)?*
- The answer is mostly Phenomenological — *Why?*
  - Partly because we are still establishing the existence of GENERAL PATTERNS/PHENOMENA,

# MECHANISTIC MODELS IN ECOLOGY AND EVOLUTION?

- *Do most ecological studies perform phenomenological or mechanistic modelling (or model-fitting)?*
- The answer is mostly Phenomenological — *Why?*
  - Partly because we are still establishing the existence of GENERAL PATTERNS/PHENOMENA,
  - ... and partly because we are (or are forced to be) interested in FORECASTING rather than EXPLAINING.



# MECHANISTIC MODELS IN ECOLOGY AND EVOLUTION?

- *Do most ecological studies perform phenomenological or mechanistic modelling (or model-fitting)?*
- The answer is mostly Phenomenological — *Why?*
  - Partly because we are still establishing the existence of GENERAL PATTERNS/PHENOMENA,
  - ... and partly because we are (or are forced to be) interested in FORECASTING rather than EXPLAINING.
- *So the big question is, can we FORECAST WITHOUT EXPLAINING?*

# MECHANISTIC MODELS IN ECOLOGY AND EVOLUTION?

- *Do most ecological studies perform phenomenological or mechanistic modelling (or model-fitting)?*
- The answer is mostly Phenomenological — *Why?*
  - Partly because we are still establishing the existence of GENERAL PATTERNS/PHENOMENA,
  - ... and partly because we are (or are forced to be) interested in FORECASTING rather than EXPLAINING.
- *So the big question is, can we FORECAST WITHOUT EXPLAINING?*
  - For example, disease outbreaks: Do we really need to care about the underlying mechanisms if we can predict a future event using phenomenological modelling (e.g., Machine-learning of time series patterns)?

# WHAT ARE MECHANISMS?

- Somewhat subjective!

# WHAT ARE MECHANISMS?

- Somewhat subjective!
- For example, the Ricker model can be thought of as mechanistic:

$$N_{t+1} = N_t e^{r\left(1 - \frac{N_t}{k}\right)}$$

# WHAT ARE MECHANISMS?

- Somewhat subjective!
- For example, the Ricker model can be thought of as mechanistic:

$$N_{t+1} = N_t e^{r\left(1 - \frac{N_t}{K}\right)}$$

- What is the mechanism?

# WHAT ARE MECHANISMS?

- Somewhat subjective!
- For example, the Ricker model can be thought of as mechanistic:

$$N_{t+1} = N_t e^{r\left(1 - \frac{N_t}{K}\right)}$$

- What is the mechanism?

# WHAT ARE MECHANISMS?

- Somewhat subjective!
- For example, the Ricker model can be thought of as mechanistic:

$$N_{t+1} = N_t e^{r\left(1 - \frac{N_t}{K}\right)}$$

- What is the mechanism? — Density dependence through scramble competition (Brannstrom & Sumpter 2005)
- If the Ricker model and another model with contest competition were compared with data — some would call it mechanistic modelling because one is trying to get at the underlying mechanism, scramble or contest competition

# WHAT ARE MECHANISMS?

- Somewhat subjective!
- For example, the Ricker model can be thought of as mechanistic:

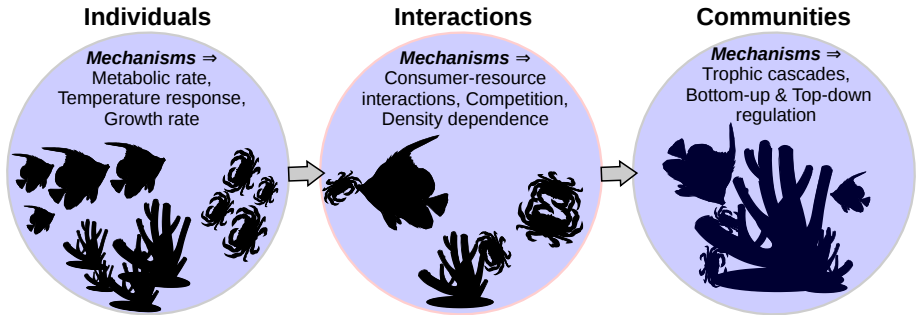
$$N_{t+1} = N_t e^{r(1 - \frac{N_t}{k})}$$

- What is the mechanism? — Density dependence through scramble competition (Brannstrom & Sumpter 2005)
- If the Ricker model and another model with contest competition were compared with data — some would call it mechanistic modelling because one is trying to get at the underlying mechanism, scramble or contest competition
- But is this REALLY mechanistic? What are  $r$  and  $k$  really?



# EXAMPLE OF A FUNDAMENTAL MECHANISM: METABOLIC RATE

- Proponents of *Ecological Metabolic Theory* (AKA “Metabolic Theory of Ecology”) argue that we have not progressed far enough towards mechanistic modelling because metabolism has been ignored

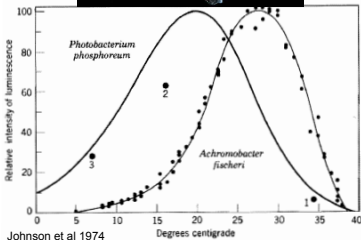


# EXAMPLE OF A FUNDAMENTAL MECHANISM: METABOLIC RATE

- The mechanistic basis of thermal performance curves  
(<https://youtu.be/6n8fCuDwn74>)

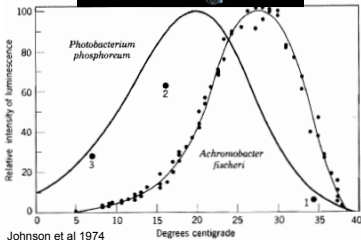
# EXAMPLE OF A FUNDAMENTAL MECHANISM: METABOLIC RATE

- The mechanistic basis of thermal performance curves (<https://youtu.be/6n8fCuDwn74>)



# EXAMPLE OF A FUNDAMENTAL MECHANISM: METABOLIC RATE

- The mechanistic basis of thermal performance curves  
(<https://youtu.be/6n8fCuDwn74>)



$$B = B_0 \left[ e^{-\frac{E}{kT}} \right] f(T, T_{pk}, E_D)$$

$T$  = temperature (K)

$k$  = Boltzmann constant ( $\text{eV K}^{-1}$ )

$E$  = Activation energy (eV)

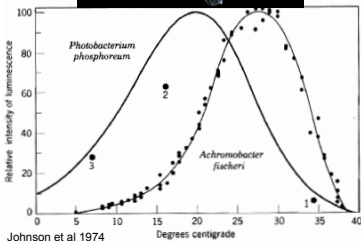
$T_{pk}$  = Temperature of peak performance

$E_D$  = Deactivation energy (eV)

(J H van't Hoff 1884, S Arrhenius 1889)

# EXAMPLE OF A FUNDAMENTAL MECHANISM: METABOLIC RATE

- The mechanistic basis of thermal performance curves (<https://youtu.be/6n8fCuDwn74>)



$$B = B_0 \left[ e^{-\frac{E}{kT}} \right] f(T, T_{pk}, E_D)$$

$T$  = temperature (K)

$k$  = Boltzmann constant ( $\text{eV K}^{-1}$ )

$E$  = Activation energy (eV)

$T_{pk}$  = Temperature of peak performance

$E_D$  = Deactivation energy (eV)

(J H van't Hoff 1884, S Arrhenius 1889)

- Surely there is more to thermal responses?
- *What about alternative models?*

# MODELLING, AND FITTING MODELS TO DATA: WHAT'S THE BIG IDEA?

- *If possible*, use biological knowledge to construct models

# MODELLING, AND FITTING MODELS TO DATA: WHAT'S THE BIG IDEA?

- *If possible*, use biological knowledge to construct models
- See if the models “agree well” with data

# MODELLING, AND FITTING MODELS TO DATA: WHAT'S THE BIG IDEA?

- *If possible*, use biological knowledge to construct models
- See if the models “agree well” with data
- Whichever model “agrees best” is most likely to have the right mechanisms



# MODELLING, AND FITTING MODELS TO DATA: WHAT'S THE BIG IDEA?

- *If possible*, use biological knowledge to construct models
- See if the models “agree well” with data
- Whichever model “agrees best” is most likely to have the right mechanisms
- That's the one that's best for predictions (e.g. population cycles), estimating rates (e.g. population or individual growth rates), etc

# MODELLING, AND FITTING MODELS TO DATA: WHAT'S THE BIG IDEA?

- *If possible*, use biological knowledge to construct models
- See if the models “agree well” with data
- Whichever model “agrees best” is most likely to have the right mechanisms
- That's the one that's best for predictions (e.g. population cycles), estimating rates (e.g. population or individual growth rates), etc
- Don't use models you already know have the wrong mechanisms just because they are popular!

# MODELLING, AND FITTING MODELS TO DATA: WHAT'S THE BIG IDEA?

- *If possible*, use biological knowledge to construct models
- See if the models “agree well” with data
- Whichever model “agrees best” is most likely to have the right mechanisms
- That's the one that's best for predictions (e.g. population cycles), estimating rates (e.g. population or individual growth rates), etc
- Don't use models you already know have the wrong mechanisms just because they are popular!
- Phenomenological/statistical models often perform better than mechanistic ones.

# MODELLING, AND FITTING MODELS TO DATA: WHAT'S THE BIG IDEA?

- *If possible*, use biological knowledge to construct models
- See if the models “agree well” with data
- Whichever model “agrees best” is most likely to have the right mechanisms
- That's the one that's best for predictions (e.g. population cycles), estimating rates (e.g. population or individual growth rates), etc
- Don't use models you already know have the wrong mechanisms just because they are popular!
- Phenomenological/statistical models often perform better than mechanistic ones.

# MODELLING, AND FITTING MODELS TO DATA: WHAT'S THE BIG IDEA?

- *If possible*, use biological knowledge to construct models
- See if the models “agree well” with data
- Whichever model “agrees best” is most likely to have the right mechanisms
- That's the one that's best for predictions (e.g. population cycles), estimating rates (e.g. population or individual growth rates), etc
- Don't use models you already know have the wrong mechanisms just because they are popular!
- Phenomenological/statistical models often perform better than mechanistic ones. *Why?*

# MODELLING, AND FITTING MODELS TO DATA: WHAT'S THE BIG IDEA?

- *If possible*, use biological knowledge to construct models
- See if the models “agree well” with data
- Whichever model “agrees best” is most likely to have the right mechanisms
- That's the one that's best for predictions (e.g. population cycles), estimating rates (e.g. population or individual growth rates), etc
- Don't use models you already know have the wrong mechanisms just because they are popular!
- Phenomenological/statistical models often perform better than mechanistic ones. *Why? — because they have less restrictive assumptions*

# BUILDING MODELS

- It's an art, takes practice (Levins' paper on the strategy of model building in biology)
- Build models one mechanism at a time — in biology, it means start at the right level of organization!
- Always consider an alternative that is more parsimonious, even if it is phenomenological!
- For example, the Boltzmann-Arrhenius model is a good first try describe and uncover mechanisms underlying individual level “traits” that are rates (e.g., fecundity or development rate)
- The next step would be to include species interactions with temperature dependence of individuals (or go in an evolutionary direction)

# FITTING MODELS (TO DATA)

- Least Squares methods
  - Linear
  - Non-linear
- Likelihood-based methods
  - Maximum Likelihood Estimation (MLE)
  - Bayesian
- Machine learning and Artificial intelligence



# FITTING MODELS (TO DATA)

- Linear and non-linear least squares model fitting: (mathematically /algorithmically simple) approaches, useful in many scenarios in biology
  - Many mechanisms in biology are inherently non-linear (i.e.,  $r$  data are better-explained by a non-linear mathematical model)
- Bayesian (and MLE) methods: Versatile and powerful when data are limited and your (e.g., mechanistic) models are complex (many parameters). Bayesian more accurate results than MLE if you have prior info (mechanistic models!).
- AI/machine Learning: most versatile and powerful for large amounts of noisy data, but the focus on maximizing ability to discover pattern and predict comes at the cost of mechanistic insights

# SUMMARY: MODEL SELECTION IS THE KEY

- Ideally, several competing (meaningful, not just null) hypotheses (mathematical models) should be fitted to data and compared using statistical theory

# SUMMARY: MODEL SELECTION IS THE KEY

- Ideally, several competing (meaningful, not just null) hypotheses (mathematical models) should be fitted to data and compared using statistical theory
- This is an advance over the traditional “null hypothesis” approach in Biology

# SUMMARY: MODEL SELECTION IS THE KEY

- Ideally, several competing (meaningful, not just null) hypotheses (mathematical models) should be fitted to data and compared using statistical theory
- This is an advance over the traditional “null hypothesis” approach in Biology
- Necessary for the advancement of Biology from an observational and axiomatic discipline to one with general theories

# SUMMARY: MODEL SELECTION IS THE KEY

- Ideally, several competing (meaningful, not just null) hypotheses (mathematical models) should be fitted to data and compared using statistical theory
- This is an advance over the traditional “null hypothesis” approach in Biology
- Necessary for the advancement of Biology from an observational and axiomatic discipline to one with general theories
- Necessary for understanding the mechanisms underlying biological patterns/phenomena

# COMPARING AND SELECTING MODELS

- It's all about the “Likelihood” of a model:  
the set of parameter values of the model ( $\theta$ ) given outcomes ( $x$ ), equals the probability of those observed outcomes given those parameter values, that is,

$$\mathcal{L}(\theta|x) = P(x|\theta)$$

- The easiest thing to do for you is to use information theory (including AIC and BIC) to compare models.
- Both AIC and BIC use the *estimated (log-) likelihood of a model*:
  - AIC:  $-2 \ln[\mathcal{L}(\theta|x)] + 2p$
  - BIC (Schwartz criterion):  $-2 \ln[\mathcal{L}(\theta|x)] + p \ln(n)$   
( $n$  = sample size,  $p$  = number of free parameters)
- The lower the AIC or BIC, the better

# AIC AND BIC

- In models fitted with least squares and normally-distributed errors,  
 $\ln[\mathcal{L}(\theta|x)] = -\frac{n}{2} \ln\left(\frac{RSS}{n}\right)$
- Thus

$$\begin{aligned} AIC &= -2 \ln[\mathcal{L}(\theta|x)] + 2p \\ &= n + 2 + n \ln\left(\frac{2\pi}{n}\right) + n \ln(RSS) + 2p \end{aligned}$$

- And

$$\begin{aligned} BIC &= -2 \ln[\mathcal{L}(\theta|x)] + p \ln(n) \\ &= n + 2 + n \ln\left(\frac{2\pi}{n}\right) + n \ln(RSS) + p \ln(n) \end{aligned}$$

- *The small-sample AIC can also be calculated similarly (see Johnson & Omland 2004)*

# COMPARING AND SELECTING MODELS

This is how you calculate AIC and BIC (using python syntax):



# COMPARING AND SELECTING MODELS

This is how you calculate AIC and BIC (using python syntax):

- $\text{residuals} = \text{Observations} - \text{Predictions}$
- $\text{rss} = \text{sum}(\text{residuals} ** 2)$
- Then,  $\text{AIC} = n + 2 + n * \log((2 * \pi) / n) + n * \log(\text{rss}) + 2 * p$   
(note  $n$  and  $p$ !)
- And  $\text{BIC} = n + 2 + n * \log((2 * \pi) / n) + n * \log(\text{rss}) + (\log(n)) * (p + 1)$
- For both AIC and BIC, If model **A** has AIC lower by 2-3 or more than model **B**, it's better — Differences of less than 2-3 don't really matter

# COMPARING AND SELECTING MODELS

This is how you calculate AIC and BIC (using python syntax):

- $\text{residuals} = \text{Observations} - \text{Predictions}$
- $\text{rss} = \text{sum}(\text{residuals} ** 2)$
- Then,  $\text{AIC} = n + 2 + n * \log((2 * \pi) / n) + n * \log(\text{rss}) + 2 * p$   
(note  $n$  and  $p$ !)
- And  $\text{BIC} = n + 2 + n * \log((2 * \pi) / n) + n * \log(\text{rss}) + (\log(n)) * (p + 1)$
- For both AIC and BIC, If model **A** has AIC lower by 2-3 or more than model **B**, it's better — Differences of less than 2-3 don't really matter

Also note that:

- $R^2 = 1 - (\text{rss}/\text{tss})$ , where tss is total sum of squares:  
 $\text{tss} = \text{sum}((\text{Observations} - \text{mean}(\text{Predictions})) ** 2)$   
(a useful measure of goodness of fit)

# COMPARING AND SELECTING MODELS: MORE STUFF

- You can also calculate Akaike Weights, which is very useful/important when comparing  $> 2$  models. These weights can then be used to perform *model averaging*
- Model selection using the Likelihood-Ratio test (LRT) is another option when you are comparing 2 models
- Adjusted  $R^2$  can be used to get a rigorous “idea” about how alternative models are performing
- Very often, you can do step-wise model simplification, especially in *for linear least squares model fitting*: Start with a complex model and drop terms till you have found a the most *parsimonious* simpler version of the original model
  - There are ready-made functions in R to do this (of course!)

# READINGS

- Levins, R. (1966) The strategy of model building in population biology. *Am. Sci.* 54, 421–431.
- Johnson, J. B. & Omland, K. S. (2004) Model selection in ecology and evolution. *Trends Ecol. Evol.* 19, 101–108.
- Bolker, B. M. et al. (2013) Strategies for fitting nonlinear ecological models in R, AD Model Builder, and BUGS. *Methods Ecol. Evol.* 4, 501–512 .
- Additional readings on the TheMulQuaBio git repository