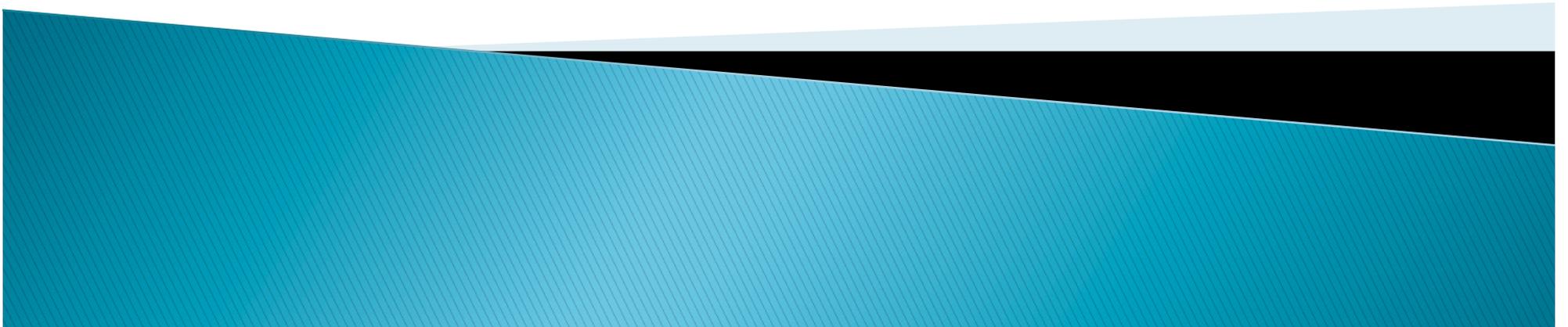


# 一般化線型混合モデル (GLMM)

2010/11/5

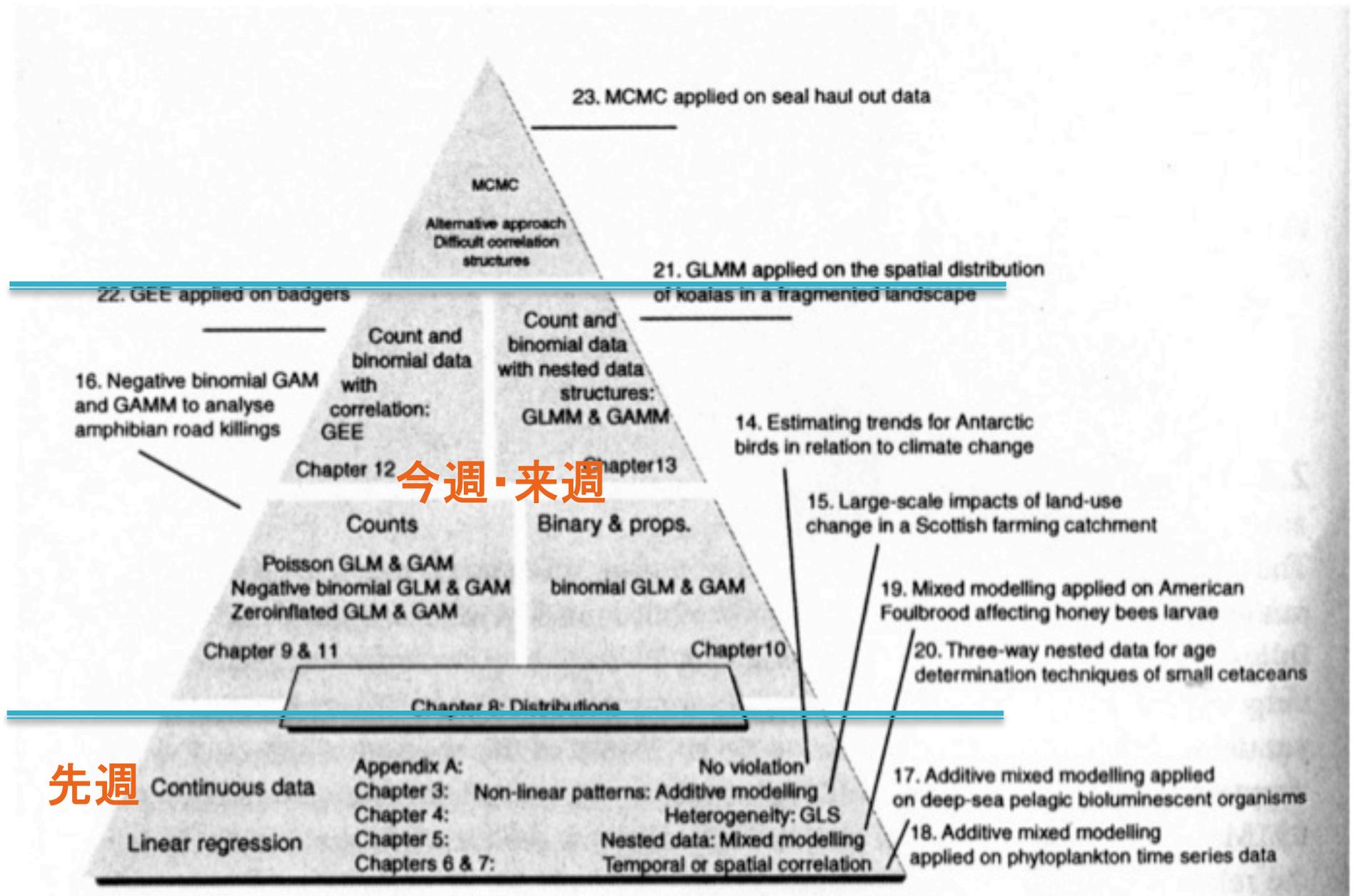
潮雅之・幸田良介



# 今日の内容

- ▶ 今までの線型モデルと一般化線型モデル、一般化線型混合モデルの関係
- ▶ 一般化線型混合モデルの注意点
- ▶ モデル選択、AICについて
- ▶ 一般化線型モデル
- ▶ 一般化線型混合モデル
  
- ▶ Rによる実演(幸田君)



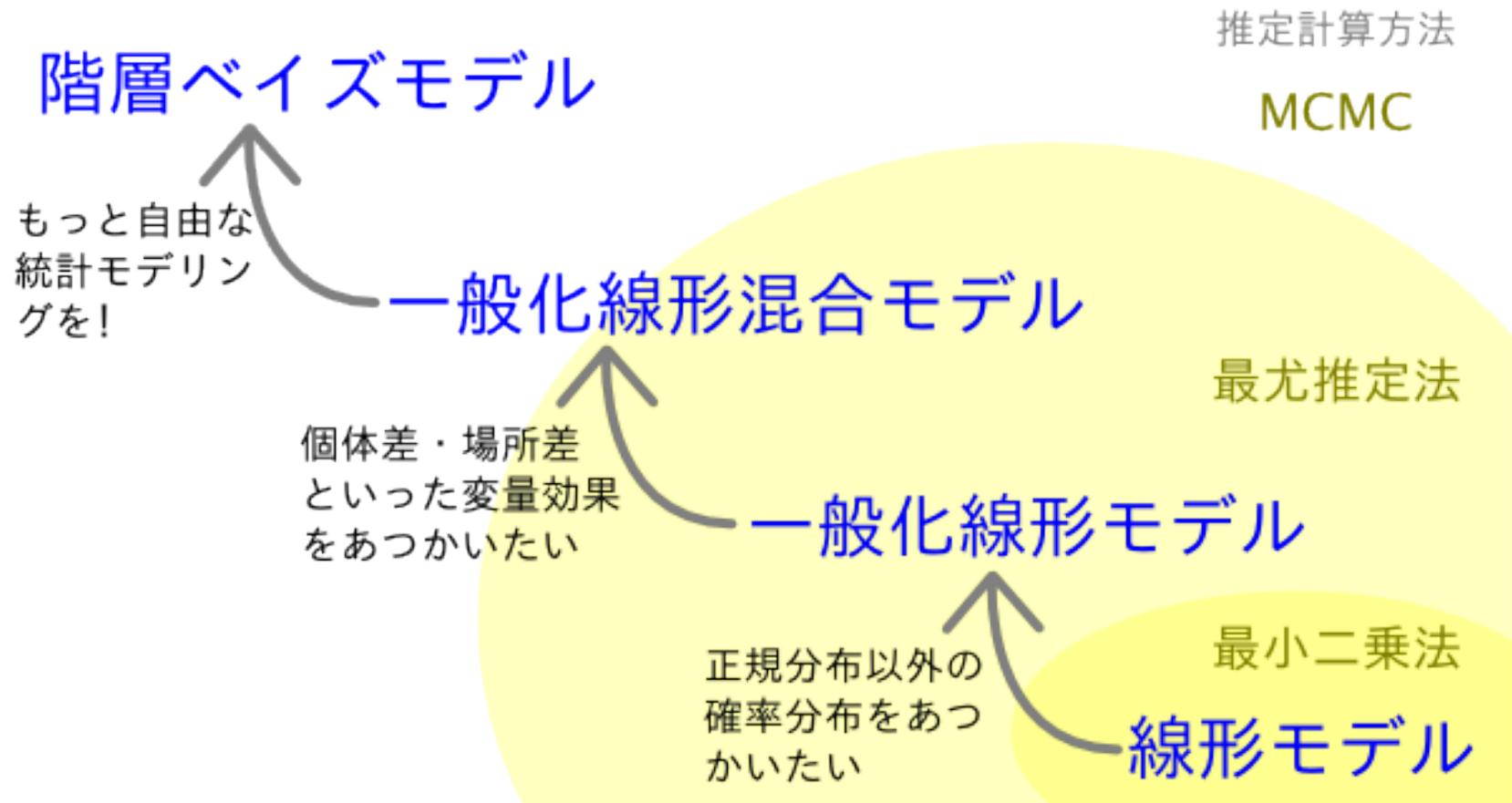


# 線型モデルの発展

- ▶ 変数間の関係は**線型**  
(非線型はGAM, GAMMで扱う、後日)
- ▶ 一般線型モデル (General Linear Model; GLM)  
→ 正規分布、混合効果無
- ▶ 一般化線型モデル (Generalized Linear Model; GLM)  
→ いろいろな分布、混合効果無
- ▶ 一般化線型混合モデル (Generalized Linear Mixed Model; GLMM)  
→ いろいろな分布、混合効果有



# 線形モデルの発展



久保さん(北大)

# 注意点

- ▶ ... GLMM and GAMM are on the frontier of statistical research. This means that available documentation is rather technical, and there are only a few, if any, textbooks aimed at ecologists. There are multiple approaches for obtaining estimated parameters, and there are **at least four packages** in R that can be used for GLMM. **Sometimes these give the same results, but sometimes they give different results.** Some of these methods produce a deviance and AIC; others do not.

# モデルの説明の前に...

- ▶ モデル選択について
- ▶ AICについて

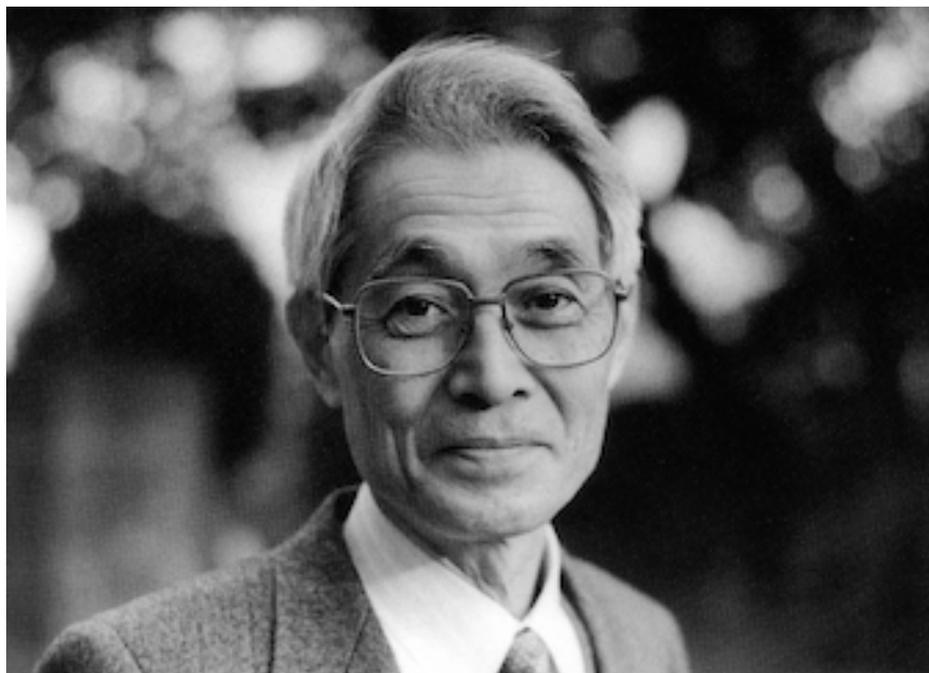


# モデル選択

- ▶ GLM, GLMMを使っているいろんなモデルが作られる。
- ▶ 例
  1. 根萌芽 ~ 親木 + シカ密度 + 親木\*シカ密度
  2. 根萌芽 ~ 親木 + シカ密度
  3. 根萌芽 ~ 親木
  4. 根萌芽 ~ シカ密度
- ▶ 1-4、どれが良いモデル?



# AICを使えば良い



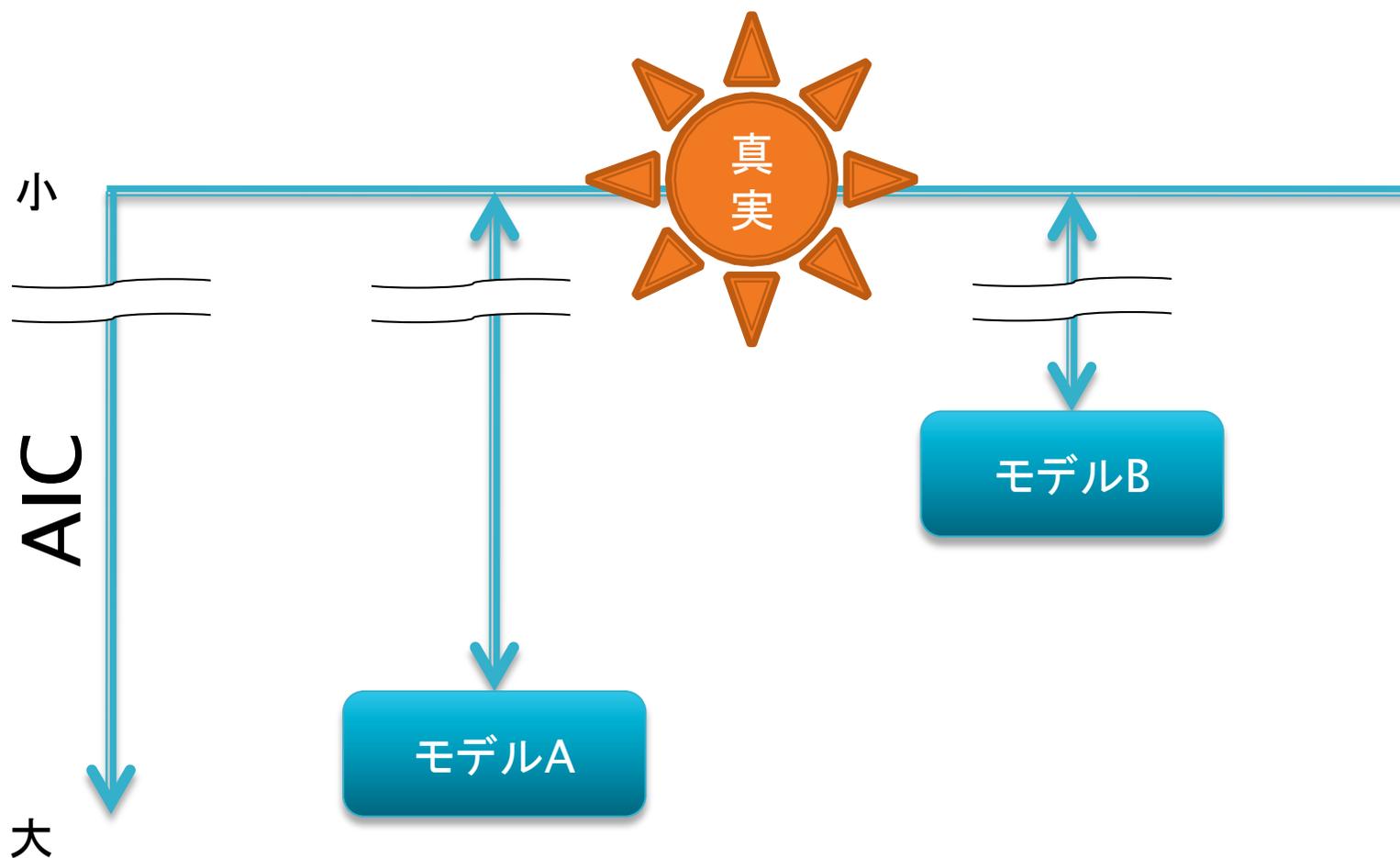
- ▶ 赤池博士が作ったAkaike Information Criteria

# AICとは？

- ▶ 赤池情報量基準 (Akaike Information Criteria)
- ▶ モデルの良さを評価する基準
  
- ▶  $AIC = -2 \times (\text{最大対数尤度}) + 2 (\text{変数の数})$
- ▶ ‘2’の意味は深いらしい...
  
- ▶ AICが低ければ低いほど良いモデル
- ▶ 他にもBIC, TIC, GIC, MDL...

以下に詳しく載っています。分かった方はぜひ解説して下さい。  
『赤池情報量基準—モデリング・予測・知識発見—』  
赤池弘次ほか(2007)

直感的に説明すると、



モデルBの方が、良いモデル！

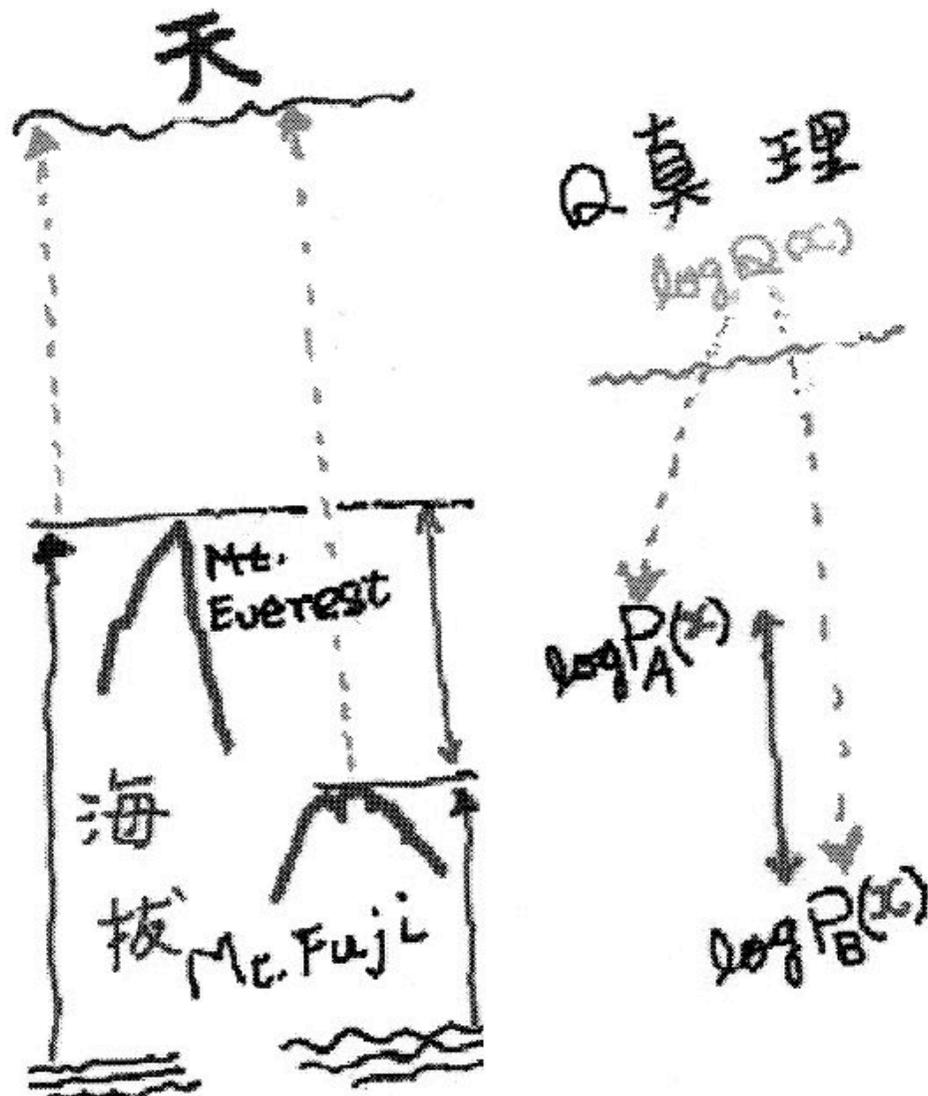


図 1.8

# AICでモデル選択

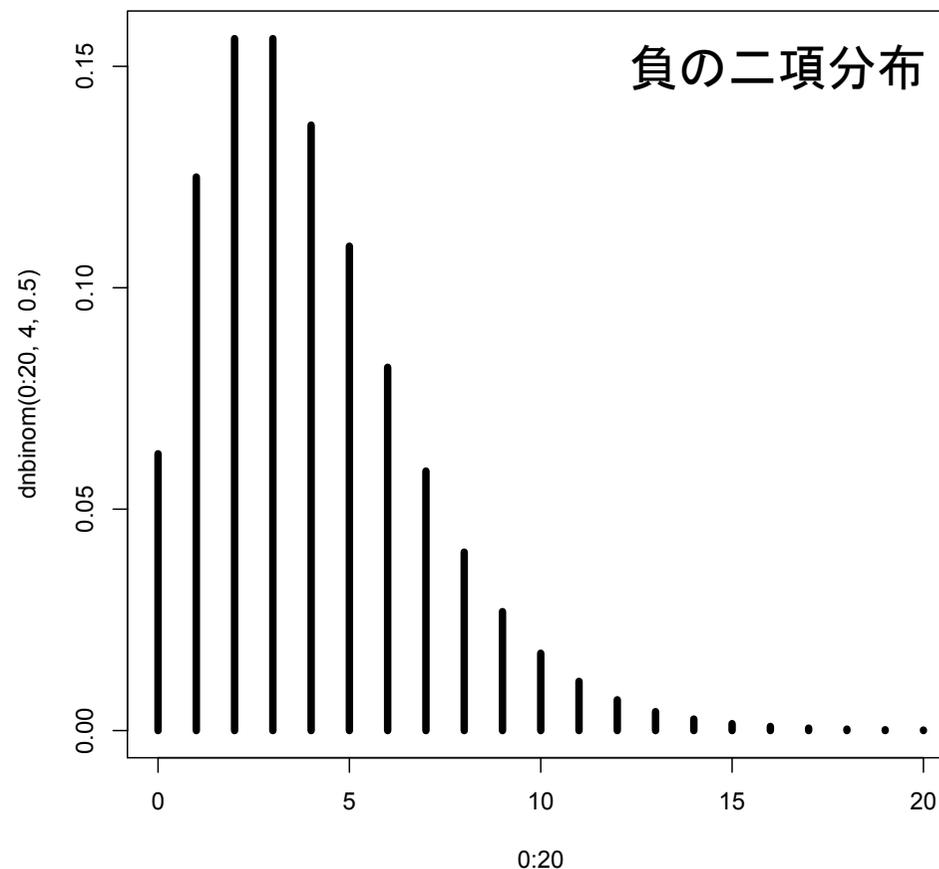
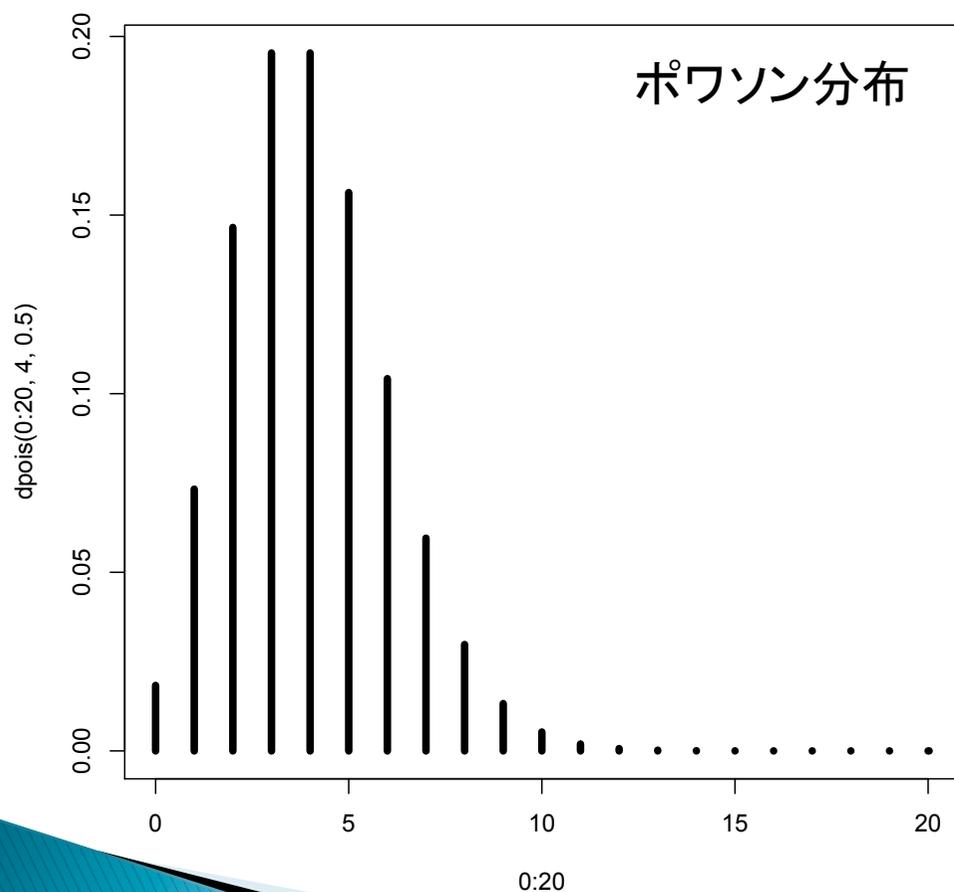
1. いろんなモデルを作って、
2. AICを比較。
3. 低いものがより良いモデルとして結果に書く。

幸田君の実演にも出てくる...はず。



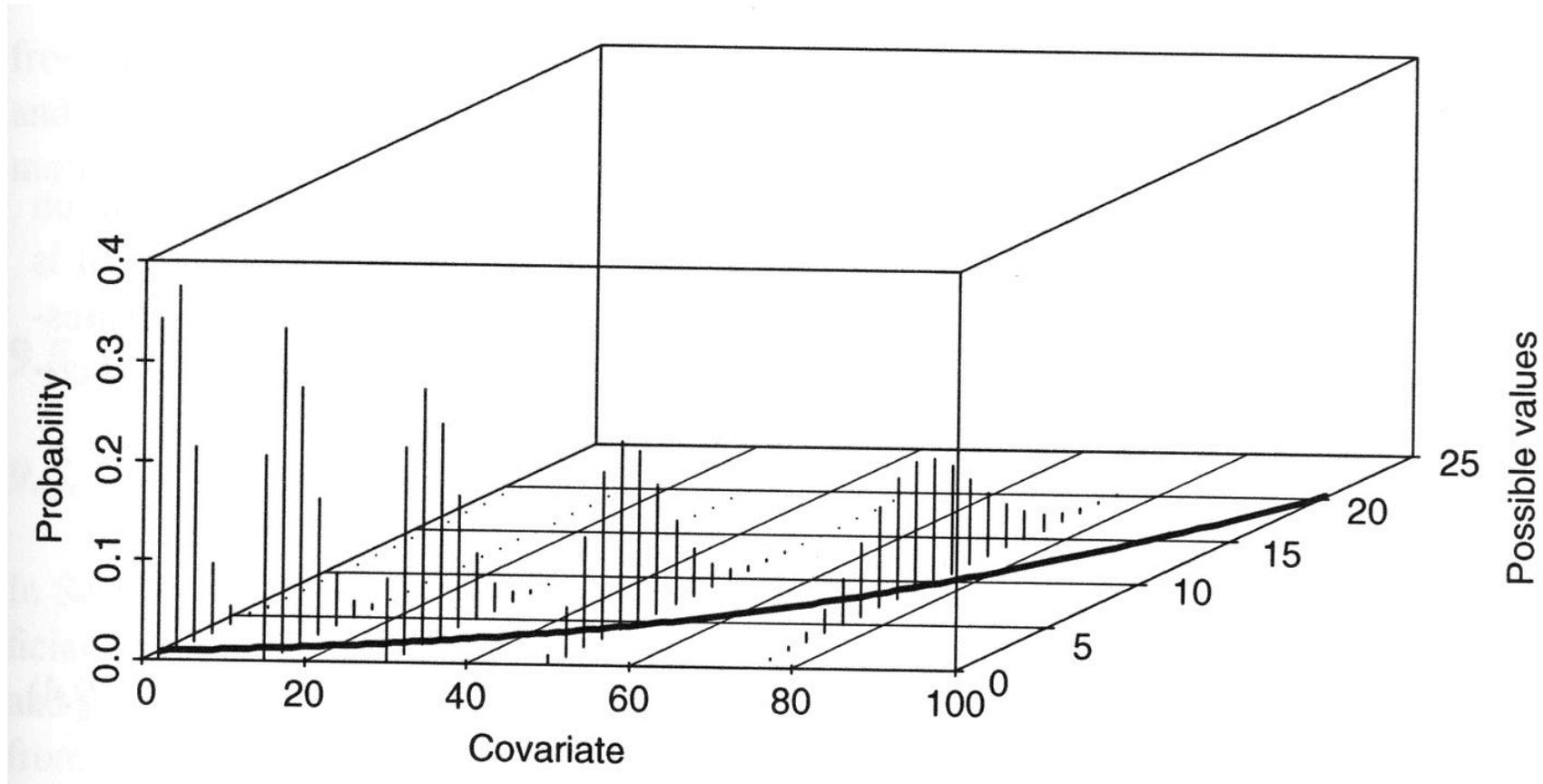
# 一般化線型モデル

- ▶ (残差が) 正規分布以外の分布を扱えるようになる



カウントデータなど、生態学にありがちなデータを解析するのに適している

# 一般線型モデル: イメージ



# 手順

1. 想定される被説明変数の分布型を考える  
例：ポワソン分布(カウントデータ)
2. 1の分布の平均値( $\mu$ )を予測するモデル式を決める  
例： $\mu \sim \beta_1 \times X_1 + \beta_2 \times X_2 + \varepsilon$

\* もう一つ

ポワソン分布の平均値は必ず0以上！

→2のモデル式から負の予測値が生成されるようでは困る！

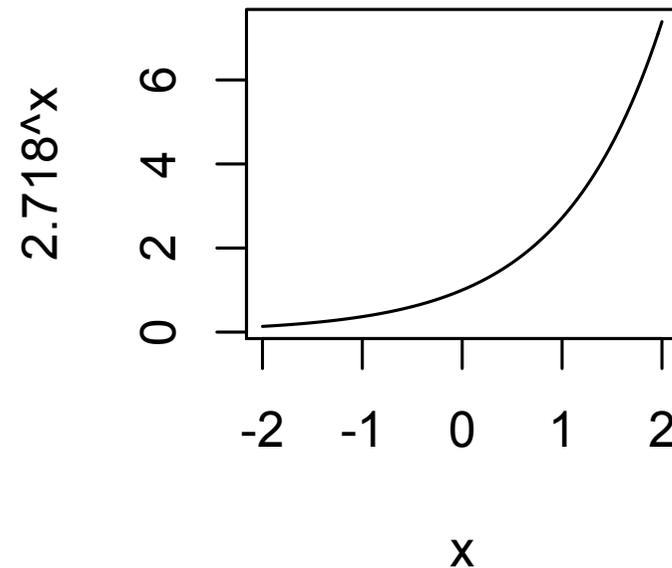


# どうしましょうか？

- ▶ どうやっても正の値しか出ないように2の式を変換。

$$\log(\mu) \sim \beta_1 \times X_1 + \beta_2 \times X_2 + \varepsilon$$

$$\mu \sim e^{(\beta_1 \times X_1 + \beta_2 \times X_2 + \varepsilon)} > 0$$



この $\log()$ のことをリンク関数と呼ぶ。

# 手順

1. 想定される被説明変数の分布型を考える
2. 1の分布の平均値( $\mu$ )を予測するモデル式を決める
3. リンク関数を決める。だいたい $\log()$



# 一般化線型混合モデル

- ▶ 混合効果を扱えるようになる。
- ▶ 個体内で反復測定をしたデータ、時系列データ、空間的に相関のあるデータなどを扱えるようになる。
- ▶ 異名がたくさん。  
混合効果 = ランダム効果 = 変量効果
- ▶ ちなみに今まで扱って来た効果は全部、固定効果と呼ばれるもの。



# 固定効果と混合効果（言葉で言うと）

- ▶ 区別は慣れないと難しい。いろいろ言い方がある。

固定効果	混合効果
予測値の平均に影響を与える	予測値の分散に影響を与える
測定できる説明変数による効果	測定できないもの全てを含んだ効果
どのような効果があるか事前に予測がたてられる	よくわからないけど結果に影響しそう
効果の大きさに興味がある	効果の大きさには興味無し

厳密ではないが、ここを一読するとちょっとつかめるかも。  
<http://hosho.ees.hokudai.ac.jp/~kubo/ce/RandomEffectsCrawley.html>

# 固定効果と混合効果(式を使うと)

▶  $Y \sim \beta_1 \times X + \beta_2 \times Z + \alpha + \varepsilon$

ただし、

$$Z \sim N(0, d_1)$$

$$\varepsilon \sim N(0, d_2)$$

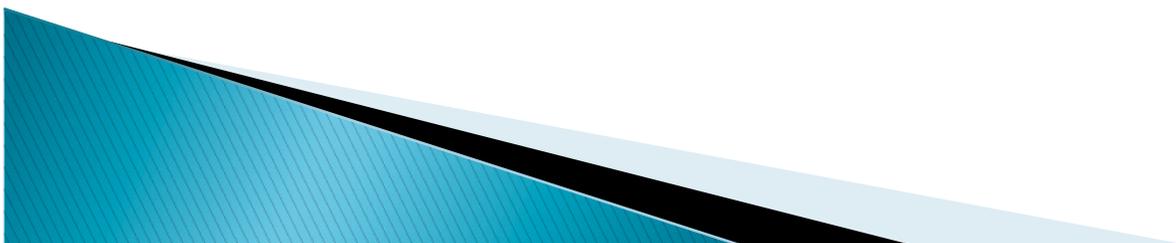
Nは正規分布を表す

Y : 被説明変数

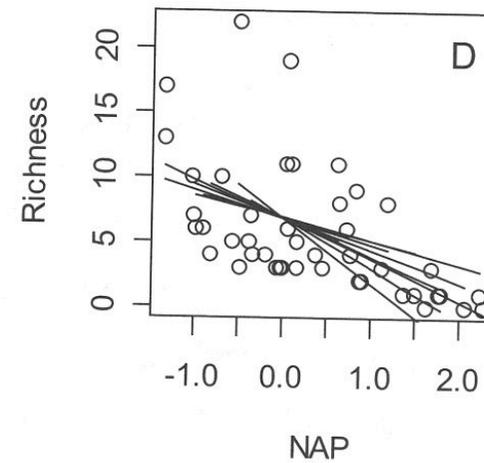
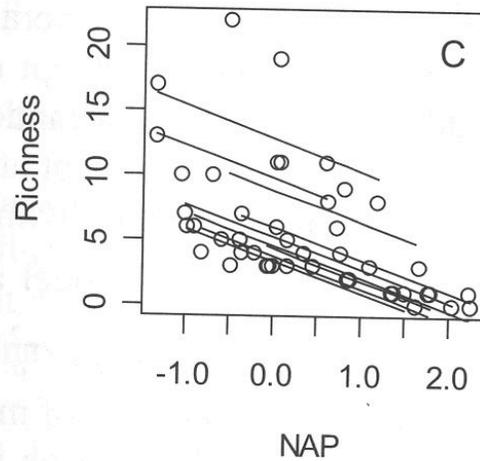
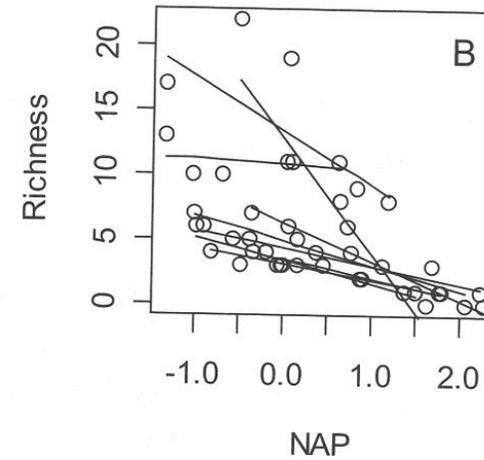
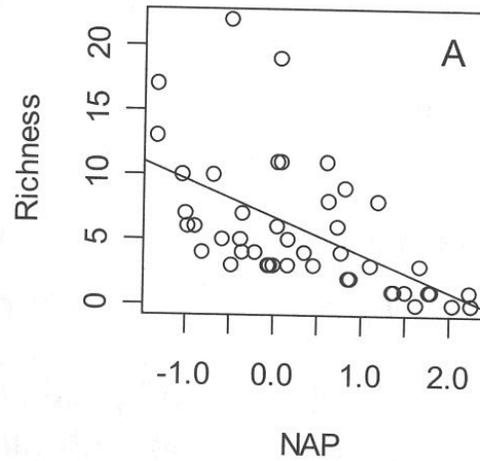
X : 固定効果

Z : 混合効果

$\alpha$  : 切片、 $\beta_i$  : 係数



# ランダム切片、ランダム傾き



# 利点

- ▶ 固定効果として扱うと水準の数だけパラメータを推定しないといけない  
→ 自由度がその分減る
- ▶ 混合効果だと、分散を一つ推定するだけでOK  
→ 自由度減らない
- ▶ 自由度が減らない→サンプル数をいっぱい採ってるようなもの。サンプルを無駄にしない。差が検出しやすい。



# 参考文献

- ▶ Analyzing Ecological Data. Zuur et al. (2007)
- ▶ Mixed Effects Models and Extensions in Ecology with R. Zuur et al. (2009)
- ▶ 『AIC -モデリング・予測・知識発見-』 赤池弘次ほか (2007)



あとは、

▶ 幸田君、よろしく！

