

生態学基礎論 (生物多様性論 II)

5. 生物多様性解析法：統計モデリングの基礎

全部で 2 回講義の 2

統計ソフトウェア R のとりかかり

— R プログラミングの準備 —

<http://hosho.ees.hokudai.ac.jp/~kubo/stat/>

講釈: 久保拓弥 kubo@ees.hokudai.ac.jp

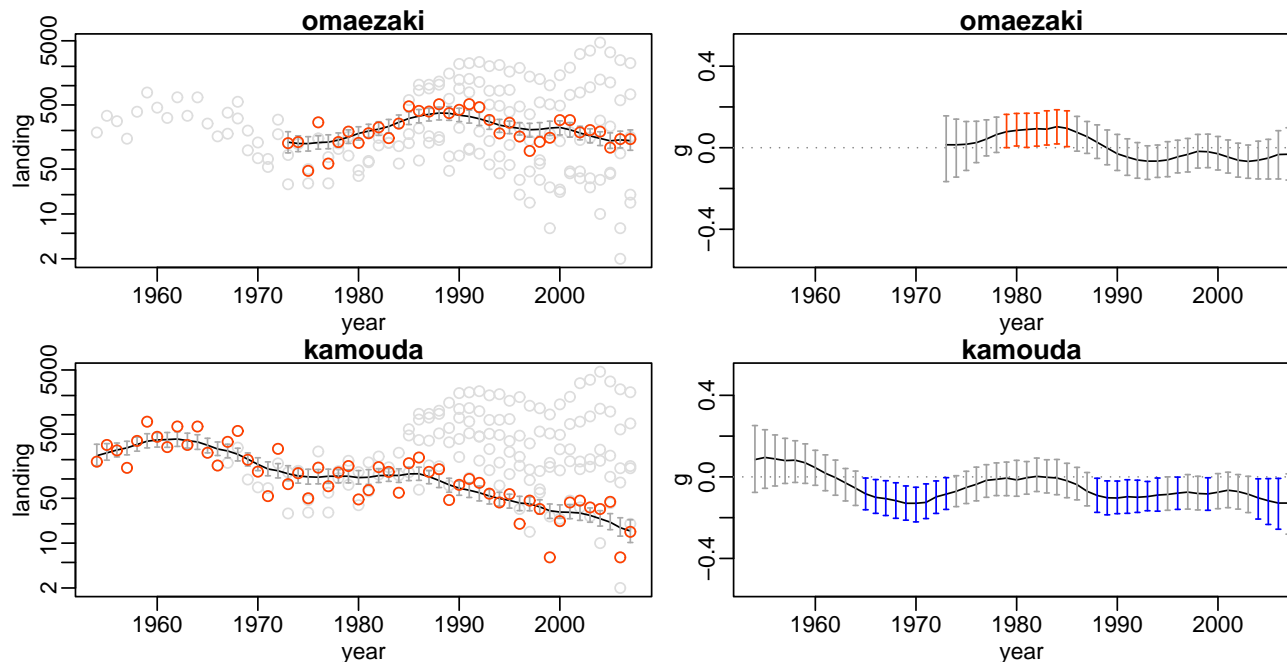
今日のハナシ: R をうごかしてみよう!

1. R インストール & プログラミング準備

とにかくインストールして使ってみる

2. 例題: ウミガメの上陸数データ 50 年ぶん

R で作図してみるウミガメ上陸数, 統計モデリング



これ使いましょう: 統計ソフトウェア R

<http://www.r-project.org/>

- いろいろな OS で使える **freeware**
- 使いたい機能が充実している
- **作図**機能も強力
- S 言語による **プログラミング**可能
- よい教科書が出版されつつある
 - 「R による保健医療データ解析演習」 中澤港 (2007)
 - 「The R-Tips」 舟尾暢男 (2005)
 - “Statistics: An Introduction Using R” M. Crawley (2005)
 - **ネット上**のあちこち



まずは簡単に

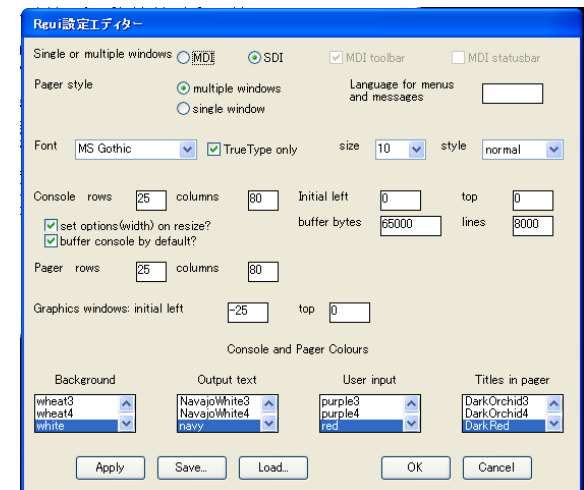
1. R インストール & プログラミング準備

とにかくインストールして使ってみる

R のインストール (例: んどーづにおいて)

<http://hosho.ees.hokudai.ac.jp/~kubo/ce/HowtoInstall.html>

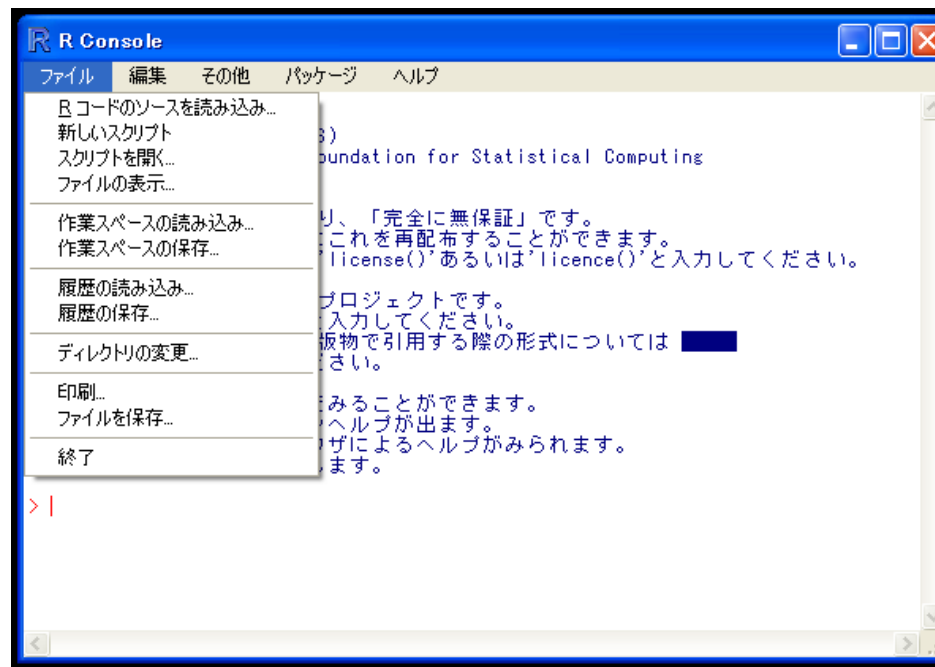
1. 上記 URL ページの「最新版の R 本体のインストール」からダウンロードサイトへ
2. 最新版 installer (本日の場合, R-2.6.1-win32.exe) をダウンロード
3. installer を起動
4. R を起動 (Rconsole file がない場合は文字化けしてるはず)
5. 「編集」 「GUI プリファレンス」
 - **Single or multiple windows: SDI** (とくに Tinn-R 使用時は必須)
 - **Font:** なにか日本語フォント
 - **Apply** して **Save ...** (My Document に Rconsole file を保存)
6. いったん R を終了して再起動



R とプログラム (スクリプト)

R のプログラムを「どこで」書けばよいか

1. Tinn-R : R と便利に連携可能
2. その他のテキストエディター (notepad とか?)
3. R の「ファイル」 「新しいスクリプト」 (今日はこれを解説)

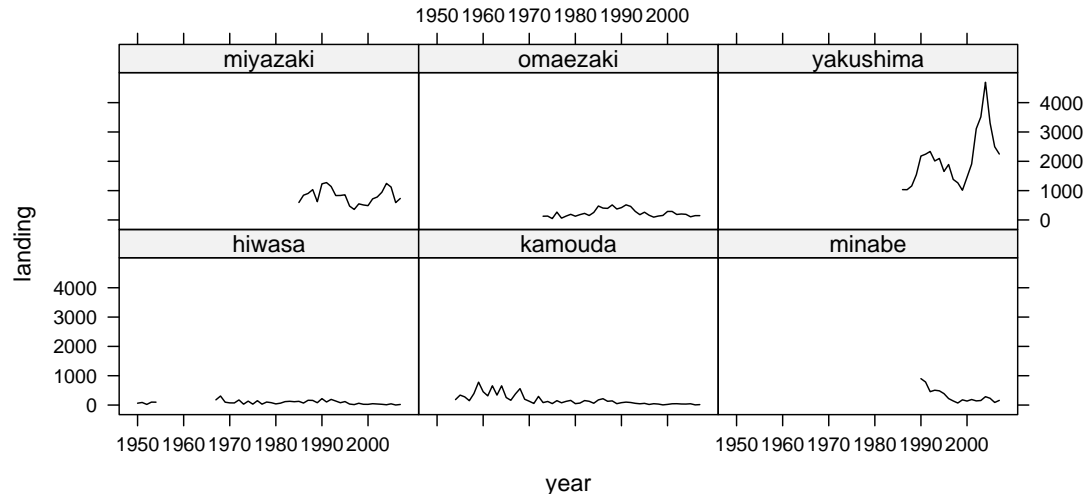


データを「観る」ことからすべては始まる

3. 例題: ウミガメの上陸数データ 50 年ぶん

R で作図してみるウミガメ上陸数 , 統計モデリング

ウミガメ上陸数の観測データ (1950年代 - 2007)



- 環境科学院・環境起学の M2 の重田さんの研究 (の下うけ担当・久保)
- ウミガメ上陸数観測データをとったのは日本ウミガメ協議会
- 今回つかうデータは 6 観測地点から得られた (静岡・瀬戸内・南九州)
- 一番ふるいデータは 1950 年代からある (観測地点によってデータの長さが異なる)
- 毎年, 観測地点に一定の幅の観測区間を設定, 6-9 月に上陸してくるウミガメ数をカウント

ウミガメデータを読みこんで作図

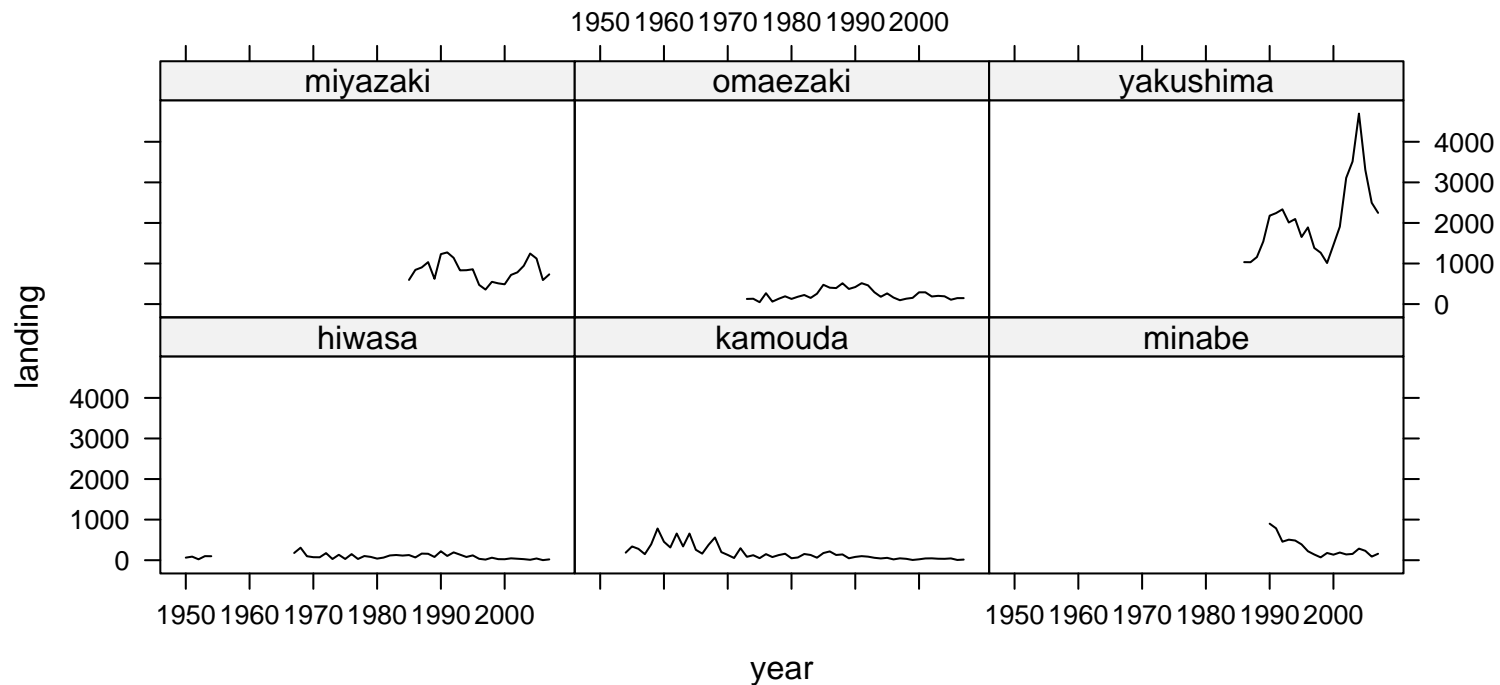
```
d <- read.csv("data.csv")  
dsub <- d[d$location == "omaezaki",]  
plot(dsub$year, dsub$landing, type = "b")
```



ウミガメデータ, 6 地点の比較

(さっきのつづき)

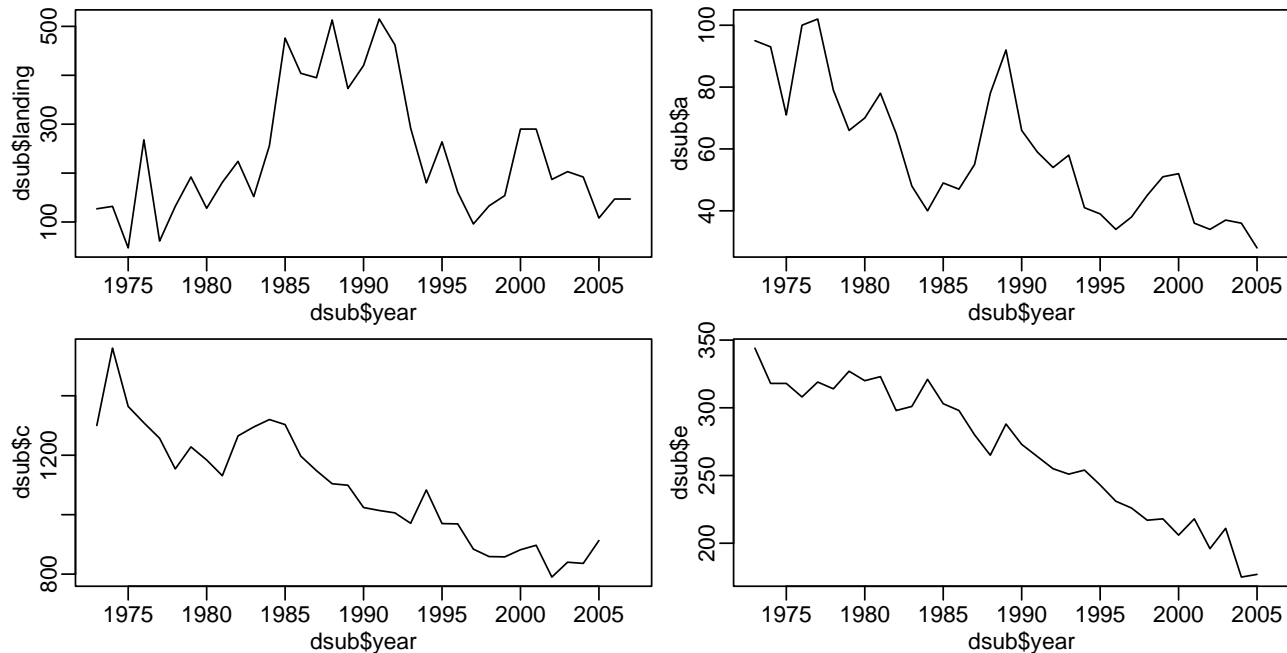
```
library(lattice)
xyplot(landing ~ year | location, data = d, type = "l")
```



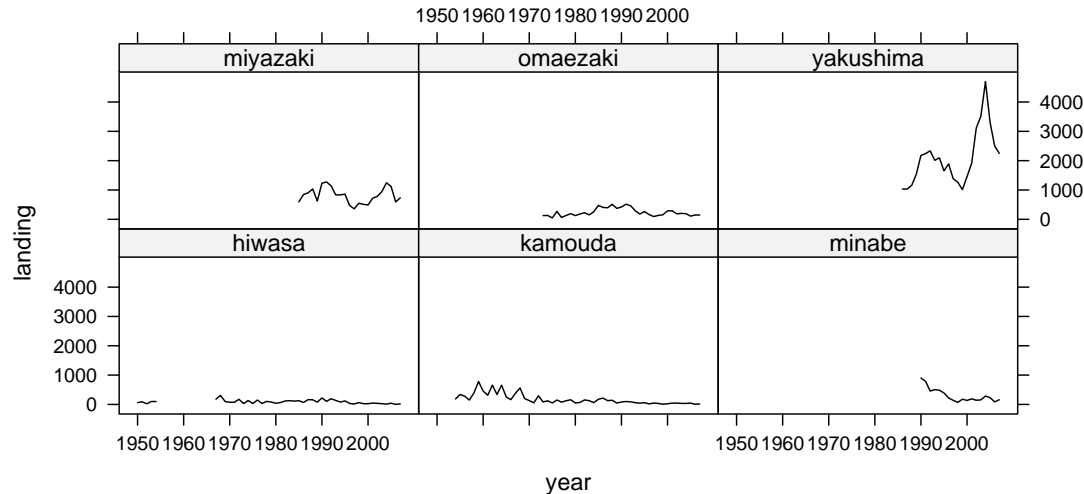
ウミガメデータ, 上陸数と漁業努力

(さっきのつづき)

```
par(mfrow = c(2, 2))  
plot(dsub$year, dsub$landing, type = "l")  
plot(dsub$year, dsub$a, type = "l")  
plot(dsub$year, dsub$c, type = "l")  
plot(dsub$year, dsub$e, type = "l")
```



ウミガメ上陸数の統計モデリング: 統計モデルとは何か?

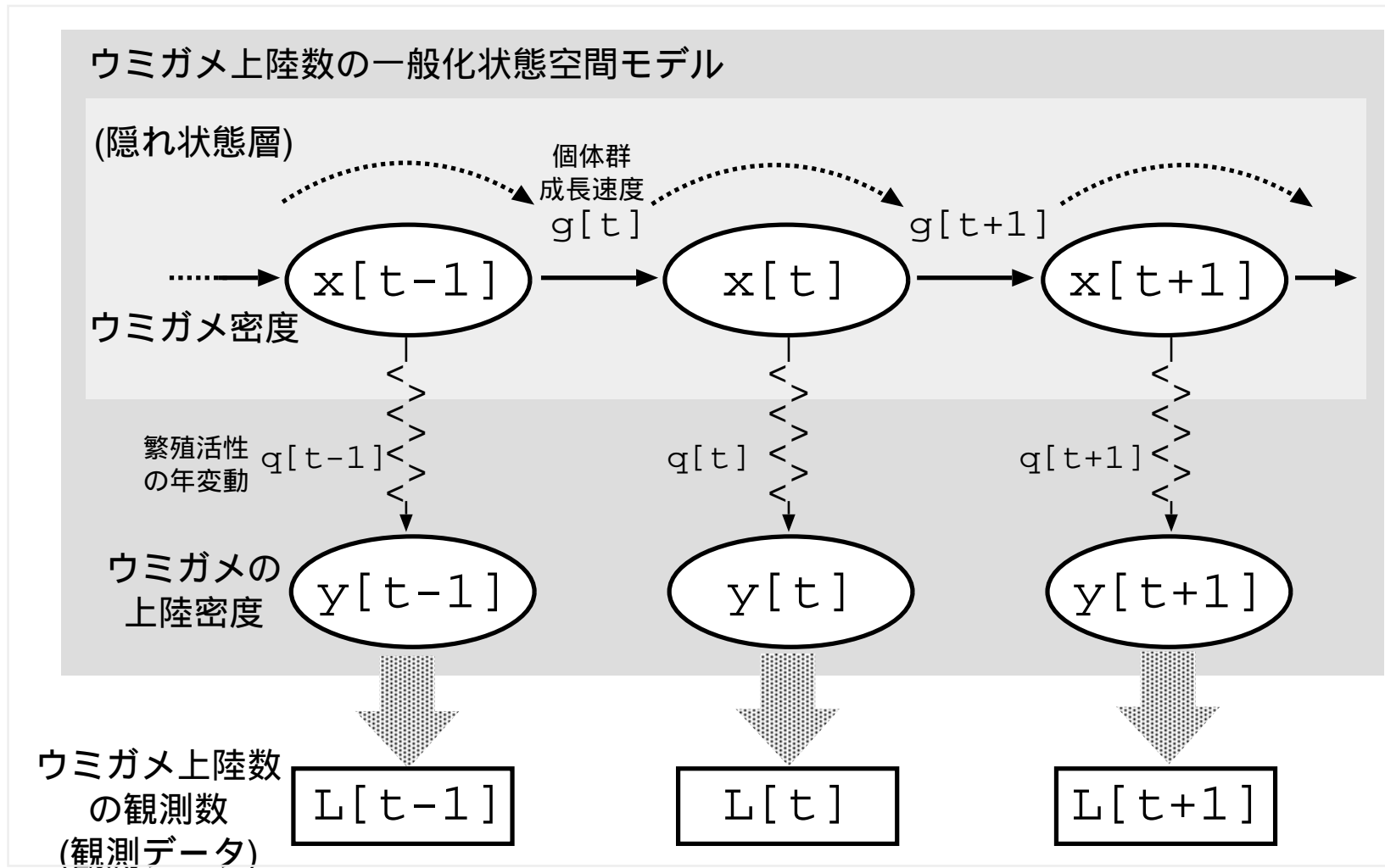


- 観測されたデータを「説明」できるような数理モデル
- 観測された値にみられるばらつきを確率分布で説明する
- 観測された値の背後に存在する関係・構造・機構を仮定する
- モデルの挙動を決めるパラメーターをもつ
- 観測されたデータを使ってパラメーターの値を推定できる
- 観測されたデータに「似ている」パターンを生成できる

ウミガメ上陸数の統計モデリング: 仮定あれこれ

- ウミガメ上陸数の「背後」にあるウミガメ密度の時間変化を説明できるようなモデル: カウントデータであるウミガメ上陸数 (以下, ウミガメの観測数や個体群密度はメスについて述べたものである) はウミガメ個体群の密度を反映したものであり, この時間変化を説明できるような 個体群増加 (減少) 速度の時間変化を観測データから推定できるような統計モデルを構築した .
- ウミガメ密度の増減だけに着目した統計モデル: 上陸数観測データと漁業努力それぞれの時間変化を図示してみると両者のあいだに関係があるとは思えないので, 漁業努力は説明変数として使わないことにした .
- 独立した個体群を仮定: 単純化のために各観測地点の個体群が独立していると仮定している .
- 全観測地点に共通する年変動を仮定: 上陸数の年変動の図から離れた地点間で年ごとの上陸数の増減が一致するケースが多いことからこの仮定をくみこんだ .
- 上陸地点ごとの random effects も仮定: しかしながら年変動は常に地点間で一致しているわけではないので, 上陸地点ごとの random effects も考慮した .
- ウミガメ個体群密度はゆっくり変化すると仮定: 人間が直接観測できないウミガメ密度 (とその増減速度) はゆっくり変化していると仮定している . 理由はウミガメ個体群の回転速度がそれほど速くない, と考えられるためである .

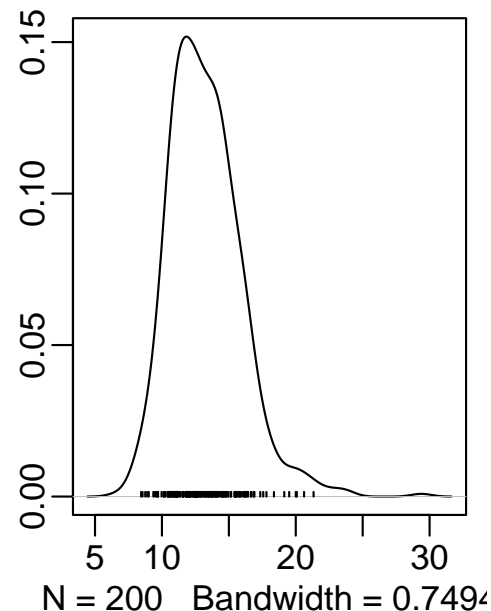
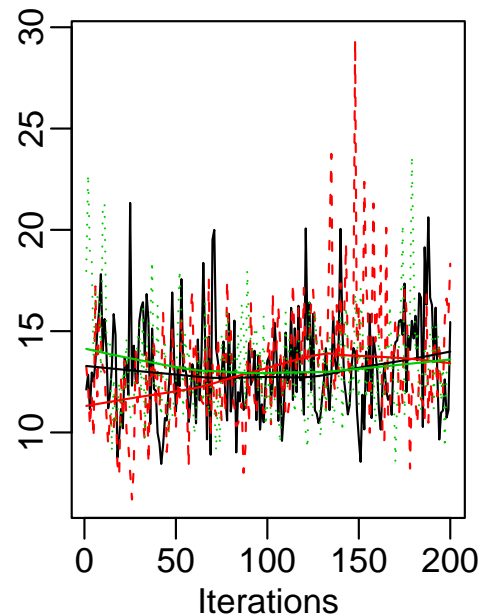
ウミガメ上陸数の統計モデリング: モデルの構造



観測された値の背後にある関係を模索する

ウミガメ上陸数の統計モデリング: 階層ベイズモデル化

- 多数のパラメーターを推定しなければならないので、柔軟性の高い階層ベイズモデルという考えかたでウミガメ上陸数の統計モデルを作った
- 階層ベイズモデルでは事前分布に (観測地点間などの) 制約をくわえることによって、多数のパラメーターがちょうどよい「観測データへのあてはまり」と「モデルの簡潔さ」をバランスさせるように調整できる
- Markov chain Monte Carlo (MCMC) 法という試行錯誤的な方法で階層ベイズモデルのパラメーター推定した



ウミガメ上陸数の統計モデルの推定結果

```
source("plot.xg.R")  
plot.all(c("omaezaki", "kamouda"), plot.q = F)
```

