

# 植生学でデータ解析ソフト”R”を利用する

松村俊和

## 1 はじめにのまえがき

みなさんは、植生データの解析にどれだけの時間をかけていますか？追加調査をしたら、また一からデータ解析をやり直していますか？

例えば、データの準備さえできていれば、DCAの解析自体から分散図の作成までががわずか数秒で終わるとすれば、解析に使っているソフトを見直しますか？しかも、クラスター分析が一瞬で終わってしまって、図表の作成までやってくれるとしたらどうですか？もちろん追加調査をした場合でも、解析作業にかかる時間は一瞬です。さらに、そのソフトが無料で使えるとしたらどうでしょうか？

これらは、ちょっと本気でRというデータ解析ソフトを勉強すれば、夢ではありません。そんな統計解析ソフトを紹介をします。

## 2 はじめに

植生学に限らず生態学のデータ解析にはほぼ全ての場合と言っていいほど何らかの形でパソコンを利用する。平均値や標準偏差などの算出・回帰分析・系列化手法やクラスター分析などの多変量解析から解析結果の図示に至るまで例をあげればきりが無い。

データ解析によく使われているソフトとしては表計算ソフトのMicrosoft ExcelやMathematica・STATISTICA・Matlabなどの統計解析ソフトなどがある。また、植生学や群集生態学のために作られたDecorana・Canoco・Pc-Ordなども広く使われていると思われる。Microsoft Excelは一般的にはワークシートを利用して、sumやaverageをはじめとする関数群や四則演算を使用して計算を行う。この場合、複数のデータ処理をするには何度もマウスを操作する必要があり、手間がかかるという欠点がある。Visual Basicを使ってマクロを作成すれば自分の解析用に特化した解析手法を自動化できるものの、データ構造の定義から全ての計算過程を記述する必要があるためかなりの開発時間と労力が必要である。そのため、実際にマクロを作成してデータ解析をしているという例はあまり多くないであろう。Decorana・Canoco・Pc-Ordなどのソフトは、それぞれの目的に特化した計算には便利ではあるものの、それ以外の計算はできない。また、Mathematica・STATISTICA・Matlabなどのソフトは高度な計算ができるものの、植生学で使用する解析手法が充実しているとは言いがたく、また高価でもある。以上のように、比較的安価に手に入る統計解析ソフトで、植生学全般にわたって網羅的にかつ容易に解析できるものはないといってよい。

しかし、データ解析ソフトRは多様な関数によって基礎的および応用的な統計解析が可能である点、S言語というプログラミング言語によるプログラムが可能である点、充実した植生学および群集生態学のための統計解析の機能を持っている点などの利点を持っている。さらに、Rはオープンソースかつフリーソフトウェアであることも評価できる。通常の統計解析ソフトはどのような処理手順でデータを解析しているかが分からな

いことがほとんどである。一方、Rのようにオープンソースであれば、どのような処理手順でデータを解析しているかを知ることが可能である。また、フリーソフトであるために無料で使用できるということは研究者の経済的な負担を軽減してくれる。

本稿はこのようなデータ解析ソフト R の基礎的な使用方法を紹介し広く普及させることで、植生学におけるデータ解析手法が向上することを目指している。

なお、組成表および常在度表の作成と表操作には神戸大学の武田義明氏作成の植生解析システム・環境設計株式会社の VX?・岡山理科大学の波田善夫氏作成の Veget・森林総合研究所の埴田宏氏作成の森林植生組成表作成・操作プログラムなどがある。本稿ではこれらの組成表および常在度表の作成と表操作以外の解析を対象としている。

### 3 R とは

R は Windows・Mac・Linux・Unix など多くの OS 上で実行可能な統計解析ソフトである。R は統計的検定・回帰分析・クラスター分析・モデル選択など一般的な統計手法の多くは使用可能であり、やや特殊な解析手法であってもオプションのパッケージをインストールすることで使用可能であることが多い。オプションのパッケージは R 本体の標準装備ではないが、追加インストール可能な関数などで、インターネット上からダウンロードすることができる。さらに、プログラミングが可能なので使いたい解析手法を自分で開発することもできる。また、グラフィック機能が非常に優れており、比較的容易に出版物並みの図表が作成できる。なお、R はインターネットを通じて無料でダウンロードして利用できる。

#### 3.1 R のインストール

R をインストールしなくても、General Rweb-jp interface(<http://r.nakama.ne.jp/Rweb-jp/>) で R を体験できる。パッケージの呼び出しなどには制限はあるものの、簡単な解析ならば可能である。例えば、次のように入力すると新しいウインドウが開いて様々な図が表示される。

デモンストレーションの実行方法

```
demo(graphics)
```

demo はデモンストレーションを見るのに使う関数でこの場合は R のグラフィック機能のデモンストレーションを見ることができる。本稿で紹介する解析のうちデータの入出力・ヘルプ機能・vegan および tree パッケージを使ったデータ解析は、General Rweb-jp interface では実行できない。しかし、それ以外の基礎統計量など多くの解析はウェブ上から実行可能なので、インストールを手間だと思えばウェブページによる R を一度試して欲しい。

R のインストールは OS ごとに手法が異なるものの、それほど難しくはない。インストール用のファイルをダウンロードして、それぞれのパソコン上で実行すればあとは質問に答える形で進めていけばよい。詳細は Rjpwiki(<http://www.okada.jp.org/RWiki/>) の「R のインストール」を参考にすること。

ところで、The Comprehensive R Archive Network(CRAN, <http://cran.r-project.org/>) からダウンロードした R では、変数名などに漢字・ひらがな・カタカナを使用することができない。そのため、変数名などに漢字などを使用する場合は、日本語化した R をインストールする必要がある。日本語化した R は中間氏のウェブページ (<http://r.nakama.ne.jp/>) からダウンロードすることができる。

なお、本稿では R のバージョンは R1.9.1、OS は Windows98SE、パッケージ cluster, tree, vegan をインストールしている状態を想定している。R のバージョンやパッケージによって基本パッケージの内容が異なるため、同じ命令を実行しても失敗する可能性があるので注意が必要である。

### 3.2 R の起動と終了

起動はスタートメニューのアイコンをクリックすると R が起動する。終了は「q()」と入力すると、「Save workspace image?」と聞いてくる。現在使用している関数や変数などを次回起動時に引き続いて使用したい場合は「yes」をクリックする。次回起動時に、新たな状態で使用したい場合は、「no」をクリックする。R では関数・変数・文字列など全てのものをオブジェクトと呼び、一度入力したオブジェクトは命令によって削除するか R を終了するまで保持される。計算の途中で全てのオブジェクトをなくして、起動直後の状態に戻りたい場合は、メニューの「Misc」をたどって、「Remove all objects」を実行するとよい。

なお、R では大文字と小文字を区別するので、命令を入力する時には注意する必要がある。また、半角スペースは無視される。本稿での命令で半角スペースを入れているのは、命令を読みやすくするためである。

### 3.3 各種パッケージのインストール

R の本体をインストールしただけでも基礎的な解析は可能である。だが、R は様々なオプションのパッケージが充実しており、それらをインストールすることで高度な解析が可能である。植生学や群集生態学で使用する頻度が高いと考えられるオプションのパッケージとしては、次のものがある。

- vegan(群集生態学)
- cluster(クラスター分析)
- tree(分類回帰樹木)

インターネットに繋がっていれば、メニューの「Packages - Install package(s) from CRAN」をたどり、該当するパッケージ名を選択すればインストールが可能である。インターネットに繋がっていないか安全面で直接繋ぐことができない場合は、該当するパッケージを CRAN(<http://cran.r-project.org/>) からダウンロードして任意の場所に保存し、メニューから「Packages - Install package(s) from local zip files」を実行すればインストールが可能である。

### 3.4 R を使った論文

R に関する情報源である Rjpwiki(<http://www.okada.jp.org/RWiki/>) の「R を使った学術論文」のページにいくつか紹介されている。論文自体に使用したことを明記していなくても実際にはより多くの論文で使われていると思われる。

## 4 R の使用方法

R の基本操作はテキストの入力である。これはマウスで操作する Microsoft Excel に慣れていると一見不便なように感じる。確かに全ての操作をキーボードから入力しようとするのが不便である。しかし、よく使うものはテキストファイルとして保存し、それを複写 + 貼付けすることで多くの操作を半自動化させることが可能で

ある。こういった操作に慣れれば、マウスよりも確実かつ迅速な操作が可能であることを実感できる。また、R Commander というパッケージを使用すればマウスで操作したりメニューを自分用に特化させられるので、どうしてもキーボードから操作するのが苦手な人は R Commander を使用すればよいだろう。

#### 4.1 ヘルプと使用例を見る

本稿は R の使用方法の概略を説明するにとどめるため、詳しい関数の説明は省略している。詳細は後述 R の情報源の項目に掲載している書籍とホームページを見て勉強していただきたい。また、R のヘルプを参照することでも解決できる。「plot」という項目のヘルプを見るには次のように入力する。

ヘルプを見る

```
help(plot) # ?plot でも同じ
```

R での命令を入力する際、「」～「改行」までは注釈として無視されるので、メモ書きとして使うことができる。本稿ではここに注釈を記述する。また、関数（ここでは例として「plot」）の使用例を見たい場合は次のように入力する。入力後、別画面でグラフが次々と表示され、様々なグラフが作成可能であることがわかる。

使用例を見る

```
example(plot) #plot という関数の使用例を見る
```

#### 4.2 データの入力・出力

解析用データはキーボードからの入力とタブ区切りあるいはカンマ区切りのテキストファイルからの入力が可能である。通常はデータを Microsoft Excel などの表計算ソフトで入力、テキスト形式で保存しておき、R で読み込むのが手間のかからない方法である。テキスト形式のデータ作成の際は、データ中の半角スペースやカンマはデータの区切りとして認識されることがあるため注意する必要がある。

データの入力・出力

#データの入力

```
sample.data <- read.table("d:/data/sample.txt", header=TRUE) #header=T で見出し有り  
を指定
```

sample.data #データの確認

```
write.table(sample.data, "d:/data/output.txt", quote=FALSE, sep="\t", #データの出力  
#quote=FALSE で引用符「"」無しを指定、sep="\t"でタブ区切りを指定  
row.names=FALSE) #row.names=FALSE で列名無しを指定
```

上記の命令によりデータの入力と出力ができる。ただし、この場合の入出力データはタブ区切り・引用符「」は使用していないデータ形式であり、「d:/data/」はファイルのあるフォルダ名で、「sample.txt」はファイル名である。R の命令は関数と呼び、関数名の後ろに「()」で引数（ひきすう）としてデータ名やデータ処理の方法などを指定する。関数ごとに入力が必要の引数や引数の形式が決まっている。必須でない引数を省略すると、それぞれの既定値を入力したのと同じだとみなされる。

<-

を使うと右辺の式を左辺に代入する働きを持つ。R ではこのように式を代入するとその式を実行する。ただし、その結果は画面には表示されない。一方、sample.data のようにオブジェクトそのものを入力した場合はオブジェクトの内容を表示する。

ファイルから入力したデータの形式が 1 行ずつ独立したデータからクロス集計 (Microsoft Excel で言えばピボットテーブル) するには次のようにする。

#### クロス集計

```
data(trees) #アメリカ桜 (black cherry) の幹周・樹高・材積のデータの呼び出し (サンプルデータ)
volume.table <- xtabs(Volume ~ Height + Girth, data = trees) #テーブル形式へと変換
volume.table
volume.frame <- as.data.frame(volume.table) #データ・フレーム形式へと再変換
volume.frame
```

これを使えば、地点名・種名・被度が 1 行ずつ並んだ植生データの基礎データから素表を作成するあるいは逆のデータ変換が可能である。なお、data(trees) という命令では R にサンプルデータとして入っているものを呼び出している。以下、data(...) としてしているものは同様であり、data() とするとデータの一覧が表示される。

### 4.3 基礎統計量

平均値・標準偏差などの基礎統計量の計算は次のようにする。次のようなコマンドを打ち込めば瞬時にして結果を得ることができる (出力結果は省略)。また、ヒストグラムや散布図が簡単な命令で作成できる。

#### 基礎統計量の計算

```
x <- rnorm(100,5,1) #平均値 5、標準偏差 1 の正規分布に従う乱数を 100 個発生させて x に代入
x #発生させた乱数
mean(x) #平均値
sd(x) #標準偏差
var(x) #分散
median(x) #中央値
max(x) #最大値
min(x) #最小値
fivenum(x) #5 数要約 (最小値・下ヒンジ・中央値・上ヒンジ・最大値)
hist(x) #ヒストグラムの図示
plot(x) #散布図の図示
```

### 4.4 区間推定と検定

発芽率などの区間推定をするには次のようにする。

#### 母比率の区間推定

```
binom.test(30,100) #100 個の種子うち 30 が発芽した場合の区間推定 (95% 水準)
```

さらに、上記の結果と 200 個の種子のうち 20 個が発芽した結果とで発芽率の検定をするには、次のようにすればよい。

#### 母比率の検定

```
a <- c(30,20) #発芽数の 30 と 20 のベクトルを c という関数で作成  
b <- c(100,200) #実験数の 100 と 200 のベクトルを c という関数で作成  
prop.test(a, b) #発芽数と実験数のを入力して検定  
#結果は「p-value」の項目を見る
```

平均値・標準偏差・95% 水準での信頼区間の値を求める方法は次のとおりである。

#### 母平均の区間推定

```
data(iris) #エドガー・アンダーソンのアヤメのデータ (サンプルデータ)  
attach(iris) #データの呼び出しをやすくする関数  
tapply(Sepal.Length, Species, mean) #種ごとに萼片長の平均値を計算  
tapply(Sepal.Length, Species, sd) #種ごとに萼片長の標準誤差を計算  
tapply(Sepal.Length, Species, t.test) #種ごとに萼片長の 95% 水準での信頼区間を計算  
detach(iris) #attach の解除
```

tapply はデータをカテゴリーデータに区分して集計する関数で、集計対象、集計したい区分、集計方法の順に記述する。この場合では、種名を集計区分として萼片長の平均値・標準誤差・95

上に引き続き、種間での値を t 検定とウィルコクソン検定 (マンホイットニーの U 検定) をする方法は以下のとおりである。

#### t 検定・ウィルコクソン検定 (マンホイットニーの U 検定)

```
t.test(Sepal.Length[Species == "setosa"], Sepal.Length[Species == "versicolor"])  
wilcox.test(Sepal.Length[Species == "setosa"], #次の行に続く  
Sepal.Length[Species == "versicolor"])
```

なお、平均値などをグラフ表示するには次の命令を実行する。

#### グラフ表示

```
plot(Sepal.Length ~ Species, data = iris) #species ごとの平均値をグラフ表示する
```

$\chi^2$  検定と Fisher の正確確率検定による独立性の検定は次のとおりである。

#### $\chi^2$ 検定

```
sample <- matrix(c(4, 2, 1, 6), ncol=2, byrow=T) #4,2,1,6 の 2 × 2 の行列を作成  
#matrix は行列を作る関数 (ncol=2 で列数を 2 に指定、byrow=T で行ごとの入力を指定)  
chisq.test(sample) #検定結果
```

#### Fisher の正確確率検定

```
sample <- matrix(c(4, 2, 1, 6), ncol=2, byrow=T) #4,2,1,6 の 2 × 2 の行列を作成
fisher.test(sample) #検定結果
```

多群間での差の検定である一元配置の分散分析やテューキーの方法による多重比較の方法は次のとおりである。

#### 一元配置の分散分析

```
data(iris) #エドガー・アンダーソンのアヤメのデータ (サンプルデータ)
oneway.test(Sepal.Length ~ Species, data=iris, var=TRUE) #種によって萼片の長さが異なるかの検定
#「~」の左辺には従属変数、右辺には独立変数を記述する
#var=TRUE で等分散性を仮定,var=FALSE とすると等分散性を仮定せずに検定する
anova(lm(iris$Sepal.Length ~ iris$Species))
summary(aov(iris$Sepal.Length ~ iris$Species))
```

#### テューキーの方法による多重比較

```
data(iris) #エドガー・アンダーソンのアヤメのデータ (サンプルデータ)
summary(fm <- aov(Sepal.Length ~ Species, data=iris)) #種ごとの分散
TukeyHSD(fm, "Species") #テューキーの方法による多重比較
```

## 4.5 回帰分析

単回帰分析と重回帰分析は R 本体のみで実行可能である。重回帰分析では AIC(Akaike's Index of Criterion) によるモデル選択が使用可能である。

#### 単回帰分析

```
x <- c(44.4, 45.9, 41.9, 53.3, 44.7, 44.1, 50.7, 45.2, 60.1) #データの手入力
y <- c( 2.6,  3.1,  2.5,  5.0,  3.6,  4.0,  5.2,  2.8,  3.8) #データの手入力
cor.test(x, y)
plot(x,y) #プロットの図示 (図は省略)
abline(lsfite(x,y), col="red") #回帰直線の図示 (図は省略)
```

#### 重回帰分析

```
data(swiss) #スイスの 1888 年の社会指標のデータ (サンプルデータ)
lm1 <- lm(Fertility ~ ., data = swiss) #lm という関数で、Fertility をそれ以外のデータで
線形重回帰
#「~」の左辺には従属変数、右辺には独立変数を記述する
#「.」はその他のデータの省略した表示
summary(lm1) #重回帰分析結果の要約を表示
sml1 <- step(lm1) #step という関数で、AIC によるモデル選択
summary(sml1) #変数選択の結果の要約を表示
```

#### 4.6 多様性指数と類似度指数の計算

多様性指数および類似度指数を計算させるには `vegan` というパッケージをあらかじめインストールしておく、使用する前にライブラリを呼び出す必要がある。ライブラリとはパッケージ内に含まれる関数などの集合であり、1つのパッケージはパッケージと同じ名前のライブラリ1つだけから構成されていることが多い。ただし、パッケージによっては複数のライブラリを含んでいることがある。

なお、多様性指数および類似度の計算方法は、小林 (1995) の第 2 章「群集サンプル間の類似性指数」を参照していただきたい。

#### 多様性指数の計算例

```
library(vegan) #ライブラリの呼び出し (あらかじめインストールが必要)
data(BCI) #パロー・コロラド島の樹木数データ (サンプルデータ)
specnumber(BCI) #specnumber という関数で種数を計算
diversity(BCI) #Shannon-Weaver 指数の計算
diversity(BCI, "simpson") #Simpson の 1-D の計算
diversity(BCI, "inv") #Simpson の 1/D の計算
```

#### 類似度指数の計算例

```
library(vegan) #ライブラリの呼び出し (あらかじめインストールが必要)
data(varespec) #地衣類の植生データ (サンプルデータ)
%vegdist(varespec) #? の計算
%vegdist(varespec, method="jaccard") #jaccard の群集係数みたいなもの?
pa.varespec <- decostand(varespec, method="pa") #decostand という関数で量的データを有
無データに変換
vegdist(pa.varespec) #vegdist という関数で Sorensen の類似商
vegdist(pa.varespec, method="jaccard") #Jaccard の群集係数
```

類似度指数の計算は他にも各種手法があるので、ヘルプを参照して欲しい。当然ながら、非類似度は各指数によって異なるものの、例えば 1-類似度指数とすることで算出可能である。

## 4.7 序列化手法 (PCA・RA・DCA・CCA)

序列化手法である主成分分析 (PCA)・交互平均法 (RA)・除歪対応分析 (DCA)・正準対応分析 (CCA) のうち PCA は追加パッケージ無しで使用可能で、他のものはパッケージ `vegan` のインストールと使用前の呼び出しにより使用可能である。R では序列化手法の計算だけでなく、結果の図示も可能である。なお、序列化手法については CCA を除いて、小林 の第 5 章「生物群集の座標づけ」に掲載されている。CCA は Ter Braak(1986) や Palmer(1993) や McCune(1997) に解析手法が紹介されている。

### PCA の解析例

```
data(USArrests) #アメリカの州ごとの暴力犯罪率データ (サンプルデータ)
result.pca <- prcomp(USArrests, scale = TRUE) #prcomp という関数で PCA 解析
result.pca #結果を表示
summary(result.pca) #結果の要約を表示
biplot(result.pca) #結果の図示
```

### RA および DCA の解析例

```
library(vegan) #ライブラリーの呼び出し
data(varespec) #地衣類の植生データ (サンプルデータ)
ra <- decorana(varespec, ira=1) #decorana という関数で RA の実行
ra #RA の結果の表示
summary(ra) #RA の結果の要約を表示
plot(ra) #RA の結果を図示
dca <- decorana(varespec, ira=0) #DCA の実行「ira=0」は省略可能
dca #DCA の結果の表示
summary(dca) #DCA の結果の要約を表示
plot(dca) #DCA の結果を図示
```

### CCA の解析例

```
library(vegan) #ライブラリーの呼び出し
data(varespec) #地衣類の植生データ (サンプルデータ)
data(varechem) #地衣類の環境データ (サンプルデータ)
result.cca <- cca(varespec, varechem) #CCA 解析
result.cca #結果を表示
summary(result.cca) #結果の要約を表示
plot(result.cca) #結果の図示
```

## 4.8 クラスタ分析

ここでは、生態学で一般的に用いられている群平均法 (非加重平均法 : UPGMA) を紹介する。クラスタ分析をするにはあらかじめ `cluster` パッケージのインストールと呼び出しをする必要がある。解析結果の図示

が即座にできる。なお、クラスター分析の詳細については、小林の第4章「生物群集の分類」を参照していただきたい。

#### クラスター分析の例

```
library(cluster) #ライブラリーの呼び出し
data(votes.repub) #大統領選挙における共和党候補者のデータ（サンプルデータ）
cluster <- agnes(votes.repub) #agnes という関数でクラスター解析の実行
plot(cluster) #リターンキーを押すと画面が切り替わっていく
```

## 4.9 分類樹木と回帰樹木 (Classification and Regression Tree)

Classification and Regression Tree(CART) は多変量解析の一種で分類の検索表のような2股の分岐を段階的に作成することで、最終的には均質なグループに分けることを目的とした解析手法である。詳細は大滝ら(1998)を、実際の使用例としては加藤(2003)をそれぞれ参照していただきたい。この解析手法を使用するにはあらかじめtreeパッケージをインストールする必要がある。また、結果の図示も容易である。

#### CARTの解析例

```
library(tree) #ライブラリーの呼び出し（あらかじめインストールが必要）
data(iris) #エドガー・アンダーソンのアヤメのデータ（サンプルデータ）
iris #データファイルの確認
ir.tr <- tree(Species ~., iris) #tree という関数で CART を実行
#「~」の左辺には従属変数、右辺には独立変数を記述する
#Species を他の全ての要因を使用して説明する
ir.tr #データ結果の表示
summary(ir.tr) #解析結果の要約を表示
plot(ir.tr) #結果の図示
text(ir.tr) #結果の文字列を図示
```

## 4.10 一般化線形モデル

一般化線形モデルはt検定・分散分析・共分散分析・回帰分析・重回帰分析・ロジスティック回帰分析などの多くの線形回帰を一般化したモデルのことである。独立変数には名義変数・順序変数・連続変数を使うことができる。従属変数の確率分布にはポワソン分布・二項分布といった離散分布、ガンマ分布・正規分布といった連続分布を使うことができる。これらを統一的なモデルとして扱うのが一般化線形モデルである。

#### 一般化線形モデルの解析例

```
library(MASS) #ライブラリの呼び出し
data(anorexia) #体重変化と拒食症のデータ (サンプルデータ)
anorex.1 <- glm(Postwt ~ Prewt + Treat + offset(Prewt), #次の行に続く
family = gaussian, data = anorexia) #glm 関数で一般化線形モデルによる解析
#「~」の左辺には従属変数、右辺には独立変数を記述する
#family=gaussian で正規分布を仮定
summary(anorex.1) #解析結果の要約を表示
```

この例では、正規分布を仮定しているため「family = gaussian」としている。二項分布の場合は「family = binomial」、ポアソン分布は「family = poisson」、ガンマ分布は「family = gamma」、正規分布は「family = gaussian」とそれぞれ設定する。

#### 4.11 モデル選択

上述の一般線形モデルなどの解析では、変数が本質的な意味を持つものであろうが無意味なものであろうが、モデル内の説明変数を増やせば増やすほど、モデル全体としては当てはまりが良くなり誤差はしだいに小さくなる。一方、モデルとしては本質的な変数だけからなる方が優れており、無意味な変数を除去することが必要である。モデル選択の手法は、上記の考え方にに基づき、本質的な変数だけを選び出そうとする手法である。モデル選択の基準としては赤池情報量基準 (AIC)・自由度調整済み決定係数・Schwarz Criterion などが提案されている。ここでは、最もよく使われている AIC によるモデル選択の実行例を示す。

なお、生態学関係の例題を使った一般化線形モデルとモデル選択の手法の分かりやすい解説が計算生態学 (<http://hosho.ees.hokudai.ac.jp/kubo/ce/index.html>) 内のページにあるので参照していただきたい。

#### モデル選択の例

```
data(swiss) #スイスの 1888 年の社会指標のデータ (サンプルデータ)
summary(lm1 <- lm(Fertility ~ ., data = swiss)) #回帰分析の実行とその結果の要約を表示
#「~」の左辺には従属変数、右辺には独立変数を記述する
s1m1 <- stepAIC(lm1) #AIC によるモデル選択の実行
summary(s1m1) #モデル選択結果の要約を表示
```

#### 4.12 その他

R には本稿で紹介したほかにも様々なパッケージがある。たとえば Rmap や Maptool は地理情報システム (GIS) のパッケージであり、spatstat は点過程 (島谷 2001) のパッケージである。また、R では自分のデータにあわせた独自の解析手法をプログラミングすることができる。詳細は後述の書籍・ウェブページを参照していただきたい。

## 5 R の情報源

### 5.1 書籍

「R による統計解析の基礎」(中澤 2003) では統計の基礎と同時に R の使用法を学ぶことができる。

「工学のためのデータサイエンス入門」(間瀬ら 2004) では基礎統計量からシミュレーションまで幅広いの統計手法が R での使用方法とあわせて掲載されている。

「The R Book」(岡田 2004) は様々な専門家がそれぞれの分野での解析手法を R で実現する方法を説明している。

### 5.2 ウェブページ

「CRAN」(<http://cran.r-project.org/>) は R の情報源としては、最大のものである。

「Rjpwiki」(<http://www.okada.jp.org/RWiki/>) は岡田昌史氏が運営されており、R に関する情報がたくさん集まっている。日々情報が増えており、日本語での R の情報源としては最も重要なものの一つである。

「R による統計処理」(<http://aoki2.si.gunma-u.ac.jp/R/>) は群馬大学の青木繁伸氏が作成されたページである。R の有用な関数が多数ある。また、同氏作成の「統計学自習ノート」(<http://aoki2.si.gunma-u.ac.jp/lecture/index.html>) には一般的な統計学に関する情報が豊富である。統計学の勉強にはこのページが役立つ。

「統計解析用フリーソフト・R-Tips」(<http://cse.naro.affrc.go.jp/takezawa/r-tips/r.html>) は大阪大に在籍していた舟尾暢男氏が作成し、現在は中央農業総合研究センターの竹澤邦夫氏のウェブページ内で保存されている。基礎的な R の使用方法から応用までがまとまっている。

「General Rweb-jp interface」(<http://r.nakama.ne.jp/Rweb-jp/>) は中間栄治氏作成のページで、ウェブ上で R が体験できる。また、中間氏のウェブページ (<http://r.nakama.ne.jp/>) からは日本語化された R をダウンロードできる。

「計算生態学」(<http://hosho.ees.hokudai.ac.jp/kubo/ce/index.html>) は日本生態学会での自由集会の一つである。R に特化した説明ではないものの、生態学全般での統計手法についてのいろいろな解説に含まれる形で、R を使用した解析手法の説明がある。一般化線形モデルやモデル選択について、R で解析することを前提とした解説がある。

「R でプログラミング」([http://takenaka-akio.cool.ne.jp/doc/r\\_auto/index.html](http://takenaka-akio.cool.ne.jp/doc/r_auto/index.html)) は執筆途中 (2005-06-11 現在) ですが、国立環境研究所の竹中明夫さんが作成されているページで、R を使ってデータの一括処理と図化をする際に役立ちそうです。

## 6 おわりに

本稿では R の魅力のうちの一部を紹介することはできたものの、実際に使用しなければ本当の R の価値を判断することはできない。松村のウェブページ (<http://kobe.cool.ne.jp/matsut/R.html>) に紹介した使用方法の実行例とデータの例を用意している。これらと上述の書籍・ウェブページ・R のマニュアルを参考にして各自で R を実際に使用して、R の有用性を実感してほしい。

## 参考文献

- [1] 加藤和弘・一ノ瀬友博・高橋俊守 2003. 分類樹木を用いた生物生息場所の分類 - 河川水辺の鳥類を対象とした事例研究 - . 応用生態工学, 5, 189-201.
- [2] 小林四郎. 1995. 生物群集の多変量解析. 蒼樹書房.
- [3] 間瀬茂・神保雅一・鎌倉稔成・金藤浩司. 2004. 工学のためのデータサイエンス入門 - フリーな統計環境 R を用いた解析 - . 数理工学社.
- [4] 中澤港. 2003. R による統計解析の基礎. ピアソン・エデュケーション
- [5] 岡田昌史 (編). 2004. The R Book - データ解析環境「R」の活用事例集 - . 九天社.
- [6] 大滝厚・堀江宥治. 1998. 応用 2 進木解析法 - CART による - . 日科技連出版社.
- [7] 島谷健一郎. 2001. 点過程による樹木分布地図の解析とモデリング. 日本生態学会誌, 51, 87-106.
- [8] Ter Braak, C. J. F. 1986. Canonical Correspondence Analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology*,
- [9] 1986, 67, 1167-1179.
- [10] Palmer, M. W. 1993. Putting things in even better order: The advantages of canonical correspondence analysis. *Ecology*, 74, 2215-2230.
- [11] McCune, B. 1997. Influence of noisy environmental data on canonical correspondence analysis. *Ecology*, 78, 2617-2623.