



1. Rをいじる前に…

まず R をいじって解析をする前に、パソコンの設定を拡張子が表示されるように変えておきましょう。

Windows : コントロールパネル→フォルダオプション→表示→詳細設定の「登録されている拡張子は表示しない」のチェックを外す。
Mac OS X : Finder→環境設定…→詳細の「すべてのファイル拡張子を表示」をチェック。

2. データファイルの作成

今回使用する EXCEL データ“r_data.xls”を開いてください。

ファイル→別名で保存→CSV（カンマ区切り）形式を選択して保存。¹

今回はマイドキュメントに Rlect というフォルダを作ってそこにデータファイルを保存しましょう。

3. ディレクトリの変更

では、R を起動しましょう。

R を起動したら、データファイルの保存してあるフォルダ（今回は先ほど作成した Rlect）にディレクトリを変更しましょう。

ファイル→ディレクトリの変更で、Rlect のフォルダを選択。²

4. データファイルの読み込み

ディレクトリが変更できたら、R にデータを読み込んでみましょう。データファイルを読み込むには read.table()³を使います。

```
> d <- read.table("r_data.csv", header=T, sep=",") #r_data.csv を読み込んで d に挿入。  
>
```

これで読み込んだデータは“d”の中に入りました。

ちなみに、<-は矢印（←）、header=T はデータの先頭行は列の名前として扱う、sep=","はデータのセルの区切りがカンマ（,）であることを表わしています。

¹ EXCEL ファイル（.xls、.xlsx）のままでは R に読み込めないで、CSV 形式（.csv）がテキスト形式（.txt）に変換する必要がある。

² ディレクトリの変更は setwd()でも可能。例えば setwd("C:/Users/ユーザ名/Documents/Rlect")とか。今のディレクトリの場所を知りたいければ、getwd()を使う。

³ csv ファイルの場合、read.csv()でも読み込める。今回の場合は read.csv("r_data.csv", header=T)で読み込める。txt ファイルを読み込むときには、read.table("ファイル名", header=T)で読み込める。

⁴ R でコマンドを打つ時、基本的にスペースはしかとされるので気にしなくていい。また#の後ろも読み込まれない。

5. データの確認

上で読み込んだデータが本当に読み込めたか確認してみましょう。

```
> d #データの表示。
  ID sex length biomass
1  1 male 93.04457 43.61900
2  2 male 88.31661 44.50584
3  3 female 57.92402 55.98809
4  4 male 55.11211 25.80724
以下略
```

str()⁵や summary()⁶で要約した形で情報を確認することもできます。⁷

```
> str(d) #列ごとの情報を表示。
'data.frame': 200 obs. of 4 variables:
 $ ID      : int  1 2 3 4 5 6 7 8 9 10 ...
 $ sex     : Factor w/ 2 levels "female","male": 2 2 1 2 2 2 1 1 2 1 ...
 $ length  : num  93 88.3 57.9 55.1 72.2 ...
 $ biomass: num  43.6 44.5 56 25.8 36.5 ...
>
> summary(d) #列ごとに要約したデータを表示。
      ID          sex          length          biomass
Min.   : 1.00   female:100   Min.   : 23.45   Min.   : 12.82
1st Qu.: 50.75   male :100   1st Qu.: 52.36   1st Qu.: 30.10
Median :100.50                Median : 61.51   Median : 42.71
Mean   :100.50                Mean   : 61.95   Mean   : 45.34
3rd Qu.:150.25                3rd Qu.: 69.96   3rd Qu.: 58.67
Max.   :200.00                Max.   :105.83   Max.   :101.45
>
```

6. データの整理

今回のデータから、オス (male) のデータだけ取り出してみましょう。

⁵ str()で表示される int は整数列 integer、num は実数列 number、Factor は文字列を表わす。

⁶ summary()では、データが分位数の形や各データの個数で表示される。分位数での表示された場合、Min (最小値)、1st Qu. (第1四分位数)、Median (中央値)、Mean (平均)、3rd Qu. (第3四分位数)、Max (最大値) が表示される。

⁷ 列の情報を得たいときは、次の関数で見ることができる。levels(): 列のデータの種類、nlevels(): データの種類の数、table(): 種類ごとのデータ数
また、平均、標準偏差を求めたいときはそれぞれ mean()、sd()を使う。

```

> male <- subset(d, d$sex=="male") #d の sex の列が"male"のものだけ取り出す。
> male
      ID sex  length biomass
1     1 male  93.04457 43.61900
2     2 male  88.31661 44.50584
4     4 male  55.11211 25.80724
以下略
> female<-subset(d, d$sex=="female") #同様にメスのデータも取り出して見ましょう。
>

```

7. ヒストグラムの作成

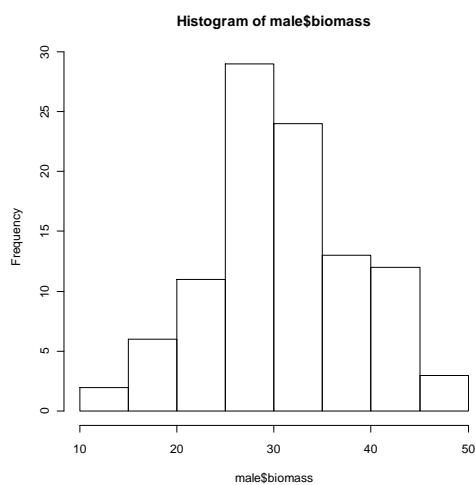
オスのデータ (male) を用いて、バイオマスについてヒストグラムを描いてみましょう。⁸

```

> hist(male$biomass) #male の biomass についてヒストグラムを作成。
>

```

すると R Graphic が開いて、下の図が描かれます。



8. 箱ひげ図の作成

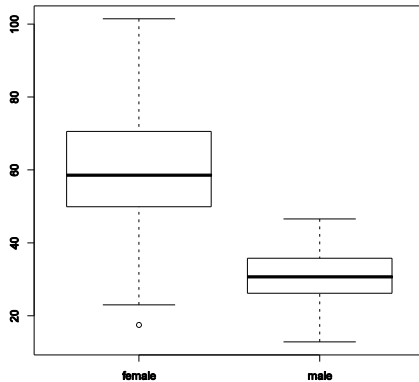
箱ひげ図を描いて、性によるバイオマスの違いを比較してみましょう。

```

> boxplot(d$biomass~d$sex) #d の biomass の箱ひげ図を d の sex ごとに描く。
>

```

⁸ ここでは煩雑になるので軸名などの指定をしなかったが、hist()に関して後述の plot()と同様に指定することができる。



この図から、メスのほうがバイオマスが大きい傾向にあることがわかります。またメスのほうが分布が広く、個体差が大きいこともわかります。

9. 散布図の作成

最後に、散布図を描いてみましょう。

性ごとに、樹幹長とバイオマスの関係を描いてみましょう。

```
> plot(female$length, female$biomass, #まずメスについて図を描く。x, yの順に入力。
+     pch=1, #点の形を指定。
+     col="red", #色をきめる。
+     xlim=c(0, 120), #x軸の範囲。今回は、オスも入る範囲を指定。
+     ylim=c(0, 120), #y軸の範囲。
+     xlab="length (cm)", #x軸の名前。
+     ylab="biomass (g)", #y軸の名前。
+ )
> points(male$length, male$biomass, pch=4, col="blue") #オスのデータを重ねる。
> #ついでに図の左上に凡例を描きましょう。
> legend("topleft", legend=c("female", "male"), pch=c(1, 4), col=c("red", "blue"))
>
```

右のような図は描けましたか？

オスもメスもきれいな正の相関が見られます。

今回の発表は以上です。

ちょっとだけRに興味をもってもらえたら嬉しいです。

あとは、先輩に聞くなり、ネットのサイト（R-Tips や RjpWiki など）を参考にすなりして頑張ってください。

