

R 入門 (3)

公衆衛生学教室セミナー 2004年7月1日 中澤 港

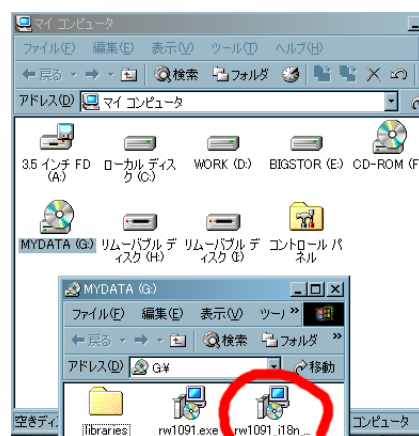
今回は、R をメニューで操作可能にするパッケージ R Commander の使い方を中心に説明する。

(注) この文書において、\ は、半角の ¥ 記号を意味する。

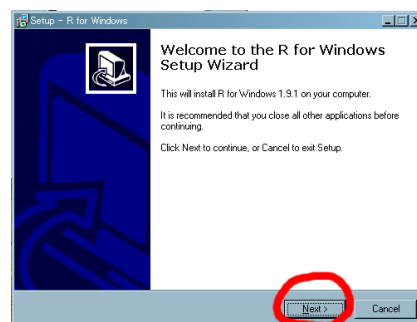
1 R-1.9.1 日本語対応 Windows 版のインストール

第1回に説明したが、新バージョンである 1.9.1 が 2004 年 6 月下旬に公開された^{*1} こともあり、図解しておく。

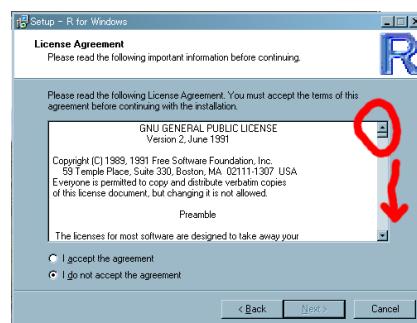
CD-R (予め焼いたもの。中身はすべてダウンロードできる) をドライブに入れ、マイコンピュータから開くと、右図に線で囲って示すように、rw1091-i18... というファイルがあるので、それをダブルクリックする。



ウィンドウが開いてインストールを開始するかどうか聞いてくるので OK をクリックする。

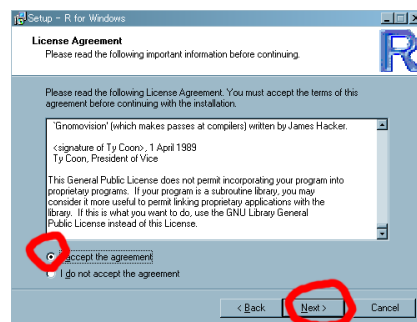


ライセンスを読んで確認するように促されるのでスライダーをドラッグして最後まで読む。

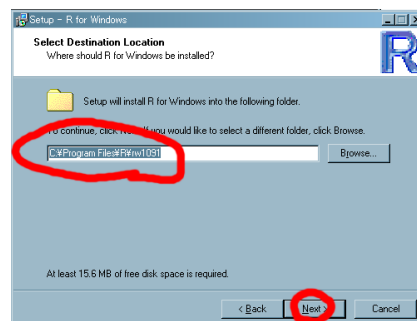


^{*1} <http://r.nakama.ne.jp/R-1.9.1/binary/win32/> から日本語対応版がダウンロードできる。ただし MacOS については日本語対応版バイナリは未公開なので公式版、つまり英語版を使うしかない。しかし、MacOS では、Java から R を呼び出す JGR (<http://stats.math.uni-augsburg.de/JGR/>) という環境を使えば日本語も通るという報告がある。

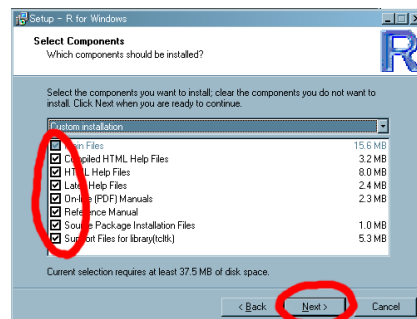
I accept の方にチェックを入れて Next をクリックする。



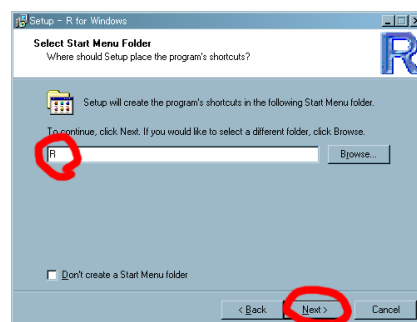
枠内にインストール先ディレクトリを指定して Next をクリックする。通常はデフォルトのままでよい。



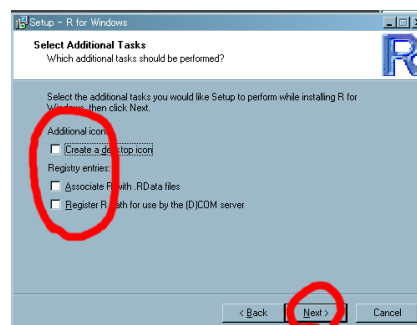
必要なコンポーネントにチェックを入れてから Next をクリックする。通常は全部入れればよい。



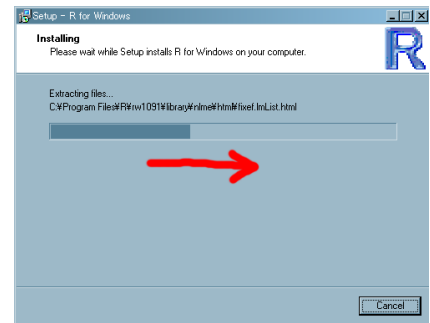
スタートメニューフォルダを指定してから Next をクリックする。通常はデフォルトのままでよい。



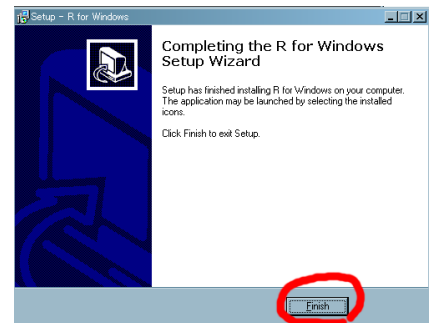
追加インストールオプションにチェックを入れてから Next をクリックする。オプションは上から順に、「デスクトップに起動用アイコンを作成」「.Rdata という拡張子を R に関連付け」「R のパスを (D)COM サーバで使うためにレジストりに登録」であるが、必要を感じなければチェックしなくてもよい。



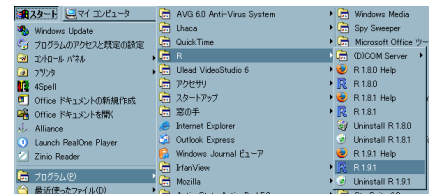
インストール実行中。



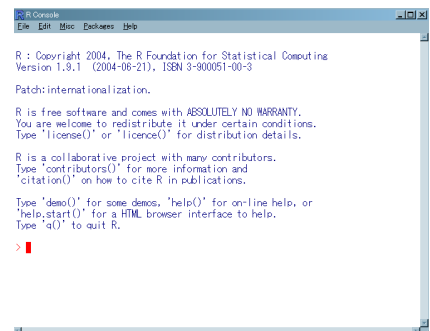
暫く待つて右の画面が出たら、無事に R のインストールは完了なので Finish をクリックする。



R を起動するには、スタートメニューからプログラム→R → R 1.9.1 を選べばよい。面倒ならこれをデスクトップなどにコピーしておいてもよい。



無事に起動すると、このウィンドウが表示され、赤い > マークの右でカーソルが点滅して入力待ちとなる。これをプロンプトと呼ぶ。



2 R Commander パッケージのインストール

インターネットに接続された環境であれば R Commander パッケージのインストールは、プロンプトに対して

```
install.packages("Rcmdr")
```

とすればよいが*2、ここでは CD-R からインストールする方法を紹介する。

*2 R-1.9.1 に Rcmdr-0.9-9 をインストールするなら、予め car パッケージを入れる必要はなくなった。自動的に必要なパッケージもダウンロードしてインストールしてくれる。

R のプロンプトに対して、まず

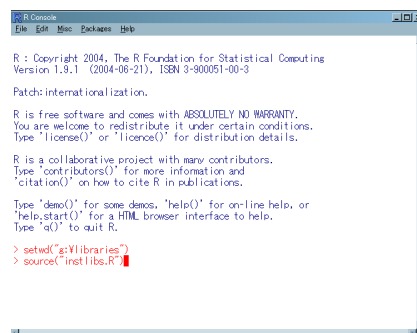
```
setwd("g:/libraries")
```

と入力し、作業ディレクトリを CD-R 内の各種ライブラリが圧縮保存されている場所に指定する。もちろん CD-R を入れたドライブが E: ドライブであれば、g: の代わりに e: としなければならない。続けて、

```
source("instlibs.R")
```

として、そのディレクトリに入っている、すべての圧縮パッケージをインストールするために予め用意した R の命令文を書き連ねたファイルを読み込んで実行させる。

ディレクトリ内のパッケージが順次無事にインストールされたことが表示され、最後にアップデート中と表示されて止まる。これは作業ディレクトリが CD-R 上にあって書き込めないためなので、Esc キーを押して中断すればよい。これを避けるためには、CD-R の中身を予めハードディスクなど、書き込み可能なメディアにコピーしてからインストール作業をすればよい。



```
R Console
File Edit Misc Packages Help

R: Copyright 2004, The R Foundation for Statistical Computing
Version 1.9.1 (2004-06-21), ISBN 3-900051-00-3

Patch: internationalization.

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for a HTML browser interface to help.
Type 'q()' to quit R.

> setwd("g:/libraries")
> source("instlibs.R")
```



```
R Console
File Edit Misc Packages Help

package 'mev' successfully unpacked and MD5 sums checked
package 'mev' successfully unpacked and MD5 sums checked
package 'multcomp' successfully unpacked and MD5 sums checked
package 'multcomp' successfully unpacked and MD5 sums checked
package 'mvtnorm' successfully unpacked and MD5 sums checked
package 'mvtnorm' successfully unpacked and MD5 sums checked
package 'R2HTML' successfully unpacked and MD5 sums checked
package 'R2HTML' successfully unpacked and MD5 sums checked
package 'randomForest' successfully unpacked and MD5 sums checked
package 'randomForest' successfully unpacked and MD5 sums checked
package 'Rcmdr' successfully unpacked and MD5 sums checked
package 'Rcmdr' successfully unpacked and MD5 sums checked
package 'Rcmdr' successfully unpacked and MD5 sums checked
package 'replots' successfully unpacked and MD5 sums checked
package 'replots' successfully unpacked and MD5 sums checked
package 'rimage' successfully unpacked and MD5 sums checked
package 'rimage' successfully unpacked and MD5 sums checked
package 'rpart' successfully unpacked and MD5 sums checked
package 'rpart' successfully unpacked and MD5 sums checked
package 'scatterplot3d' successfully unpacked and MD5 sums checked
package 'scatterplot3d' successfully unpacked and MD5 sums checked
package 'sem' successfully unpacked and MD5 sums checked
package 'sem' successfully unpacked and MD5 sums checked
package 'survival' successfully unpacked and MD5 sums checked
package 'survival' successfully unpacked and MD5 sums checked
package 'survival' successfully unpacked and MD5 sums checked
package 'vcd' successfully unpacked and MD5 sums checked
package 'vcd' successfully unpacked and MD5 sums checked
bundle 'VR' successfully unpacked and MD5 sums checked
bundle 'VR' successfully unpacked and MD5 sums checked
package 'XML' successfully unpacked and MD5 sums checked
package 'XML' successfully unpacked and MD5 sums checked
package 'xtable' successfully unpacked and MD5 sums checked
package 'xtable' successfully unpacked and MD5 sums checked
updating HTML package descriptions
trying URL 'http://cran.r-project.org/bin/windows/contrib/1.9/PACKAGES'
```

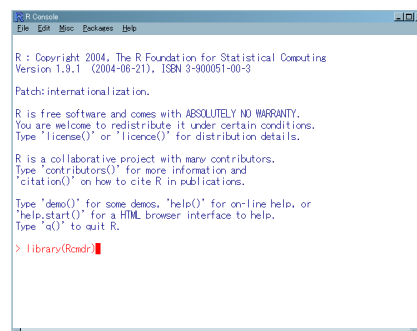
3 R Commander を起動する

ここではオリジナルの R Commander を使ったが、メニューはテキストファイルで構成されているので、中間栄治さんが公開されているようなテキストファイルと入れ替えれば（または自分で書き換えれば）メニューを日本語化することは簡単である*3。また、R Commander を使うには、R Console のメニューの GUI Preference で、予め SDI にしておく方がよい。

他のライブラリを使えるようにするときも基本的に同じだが、R Commander ライブラリを呼び出すには、

```
library(Rcmdr)
```

と入力する。エラーが 1 つでるが無視してよい。



```
R Console
File Edit Misc Packages Help

R: Copyright 2004, The R Foundation for Statistical Computing
Version 1.9.1 (2004-06-21), ISBN 3-900051-00-3

Patch: internationalization.

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

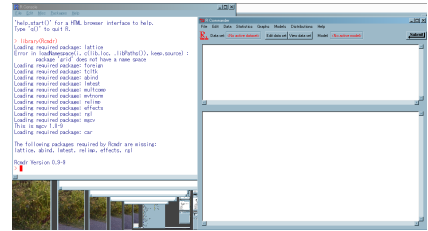
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for a HTML browser interface to help.
Type 'q()' to quit R.

> library(Rcmdr)
```

*3 C:\Program Files\R\rw1091\library\Rcmdr\etc\Rcmdr-menus.txt が標準状態でのメニューファイルのパスである。日本語対応版の R ならば、このファイルをメモ帳などで書き換えれば、それだけで R Commander のメニューを書き換えることが可能である。

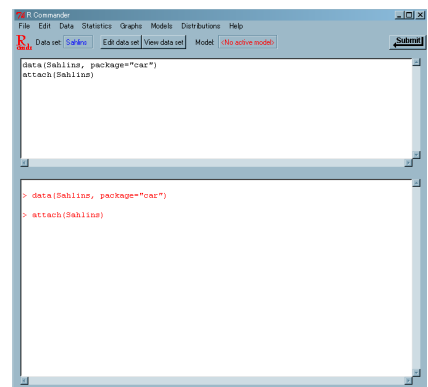
他のライブラリとは異なり、R Commander ライブラリは、呼び出すだけで R Commander のウィンドウが起動する。



4 R Commander を使ってみる

使い方の例として、R Commander に入っているデータ、**Sahllins** を使って、散布図を書き、回帰分析をしてみる。**Sahllins** は、文化人類学者マーシャル・サーリンズが、有名な『石器時代の経済学』*4の中で中央アフリカの Mazulu 村 20 世帯の農業生産について報告しているデータである。

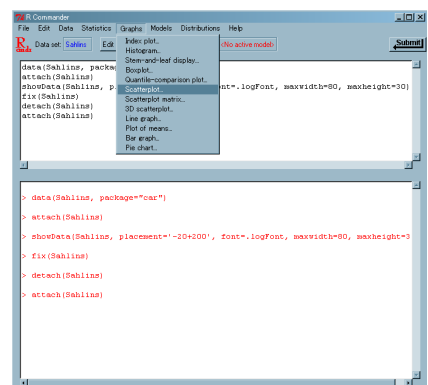
Sahllins というデータフレームが使えるようにするには、プルダウンメニューの **Data** から **Data in packages** → **Read data set from an attached package...** と選んで出てくるパッケージリストの中から **car** を選んで **OK** をクリックし、表示されるウィンドウで **Sahllins** を選んでクリックすればよい。



ここで **Edit data set** をクリックすると、データエディタが起動する（編集可）。このデータフレームには **consumers** と **acres** という2つの変数があることがわかる。**consumers** は各世帯の農業従事者数に対する消費者の人数比を意味し、**acres** は各世帯の農業従事者 1 人あたり耕作面積を意味する変数である。エディタを閉じるにはプルダウンメニューの **File** から **close** を選ぶ。

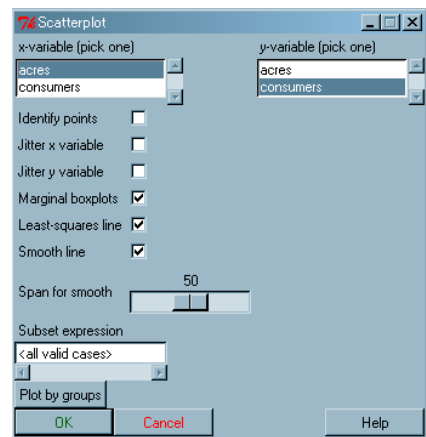
	consumers	acres	var2	var4	var5	var6	var7
1	1	1.71					
2	1.08	1.52					
3	1.15	1.29					
4	1.15	3.09					
5	1.2	2.21					
6	1.3	2.26					
7	1.37	2.4					
8	1.37	2.1					
9	1.43	1.96					
10	1.46	2.09					
11	1.52	2.02					
12	1.57	1.91					
13	1.65	2.17					
14	1.65	2.28					
15	1.65	2.41					
16	1.66	2.23					
17	1.67	3.04					
18	2.03	2.06					
19	2.05	2.73					

次に **consumers** と **acres** の関係をグラフで見してみる。プルダウンメニューの **Graphs** から **Scatterplot** を選ぶ。

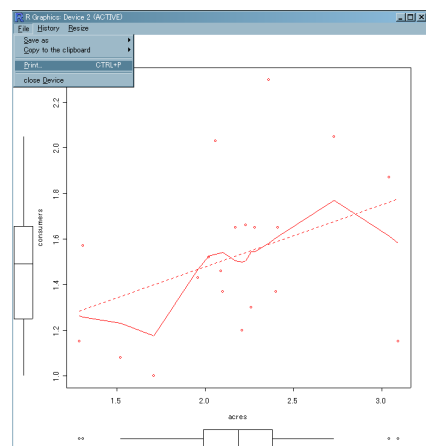


*4 文明の進歩とともに余剰が増えて余暇ができてきたはず、という迷信に反するデータを出し、実は狩猟採集経済に生きる人々の方が労働時間は短くて済むことを示した、経済人類学の先駆的な本である。

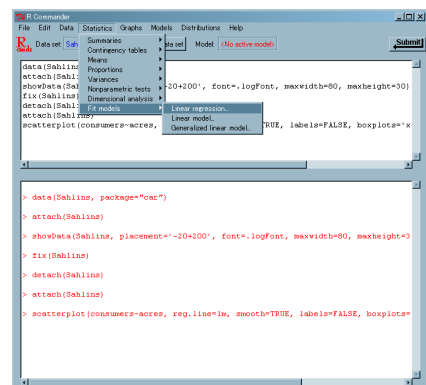
現れるウィンドウで、x-variable として `acres` を、y-variable として `consumers` を選び、OK をクリックする。



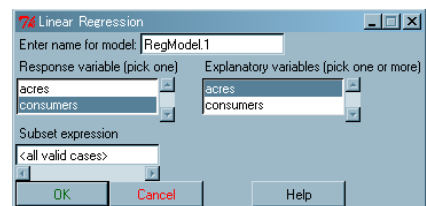
横軸に `acres`、縦軸に `consumers` が取られた散布図が描かれたウィンドウが開く。軸の外側に、各々の変数についての箱ヒゲ図も描かれ、赤の点線で回帰直線が、実線でスムージングした近似曲線が重ね描きされる。プルダウンメニューの `File` から `Print` を選べば、グラフをそのまま印刷することもできる。`Save as` でグラフィックファイルとして保存できるし、`Copy to the clipboard` でビットマップ画像またはメタファイルとしてクリップボードにグラフを記憶させられる。`Close Device` を選べばこのウィンドウが閉じる。



次に回帰分析をしてみる。R Commander のプルダウンメニューの `Statistics` から `Fit models` の `Linear Regression` を選ぶ。



散布図を描くときと同様に設定ウィンドウが開くので、Response variable として `consumers` を選び、Explanatory variables として `acres` を選んで OK をクリックする。なお、ここでは単回帰なので説明変数は 1 つしか選ばないが、重回帰分析をする場合は、ここで複数の説明変数を選べばよい。



R Commander のウィンドウの下側に単回帰分析の結果が表示される。回帰式が

$$\text{consumers} = 0.9321 + 0.2733 * \text{acres}$$

と推定されるが、回帰係数がゼロであるという帰無仮説の下でこの値が偶然得られる確率が 0.1026 であり、自由度調整済み相関係数の 2 乗の値から acres のバラツキが consumers のバラツキの 9.34%しか説明しないことがわかる。

```

R Commander
File Edit Data Statistics Graphs Models Distributions Help
Data set: [Save] [Edit data set] [View data set] [Model] [RegModel] [Submit]

data(Bahline, package="cac")
attach(Bahline)
showData(Bahline, placement="+204200", font=.logFont, maxHeight=80, maxHeight=30)
fix(Bahline)
detach(Bahline)
attach(Bahline)
postscriptplot(consumers~acres, reg.line=T, smooth=TRUE, labels=FALSE, boxplot=T)
RegModel.1 <- lm(consumers~acres, data=Bahline)
summary(RegModel.1)

Call:
lm(formula = consumers ~ acres, data = Bahline)

Residuals:
    Min       1Q   Median       3Q      Max
-0.62664 -0.22598 -0.00099  0.12001  0.72298

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.9321    0.3845    2.426  0.0162 *
acres        0.2733    0.1589    1.720  0.1026 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3305 on 18 degrees of freedom
Multiple R-Squared:  0.1411    Adjusted R-Squared:  0.0934
F-statistic: 2.957 on 1 and 18 DF,  p-value: 0.1026

```

R Commander を終了するには、プルダウンメニューの File から Exit の From Commander を選ぶ。R Commander で実行した命令が記録されている上側ウィンドウの中身をスクリプトと呼び、このメニューの Save script で保存できる（保存したスクリプトは任意のテキストエディタで編集できるし、Load script from file... で呼び出して実行できる）。

```

R Commander
File Edit Data Statistics Graphs Models Distributions Help
Load script from file...
Save script...
Save script as...
Options...
Reset output width...
Exit
Summary & regression...

Exit
  From Commander
  From Commander and R

Call:
lm(formula = consumers ~ acres, data = Bahline)

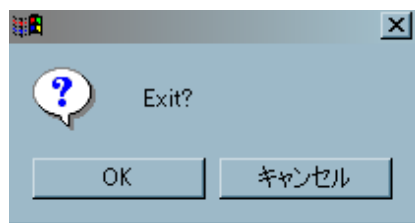
Residuals:
    Min       1Q   Median       3Q      Max
-0.62664 -0.22598 -0.00099  0.12001  0.72298

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.9321    0.3845    2.426  0.0162 *
acres        0.2733    0.1589    1.720  0.1026 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

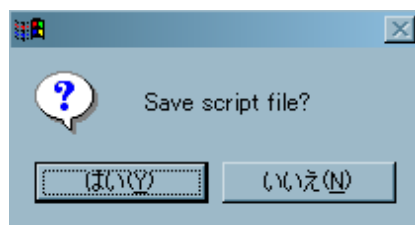
Residual standard error: 0.3305 on 18 degrees of freedom
Multiple R-Squared:  0.1411    Adjusted R-Squared:  0.0934
F-statistic: 2.957 on 1 and 18 DF,  p-value: 0.1026

```

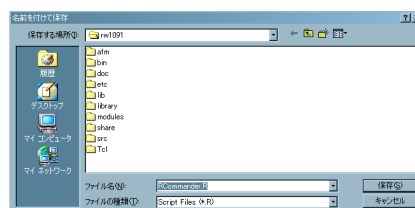
Exit?と確認を求めるウィンドウが現れるので OK をクリックする。



Save script file?と確認を求めるウィンドウが表示される。スクリプトを保存したければ「はい (Y)」をクリックする。



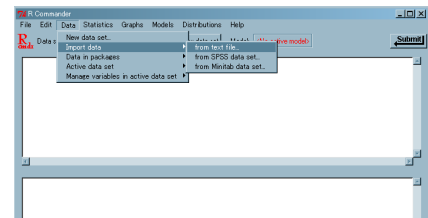
名前を付けて保存するウィンドウが現れるので、ディレクトリとファイル名を適当に選んでから「保存 (S)」をクリックする。続けて Save output file?と確認するウィンドウが現れるので、「はい」と答えれば、グラフ以外の実行結果が表示されている下側ウィンドウの中身を保存できる。



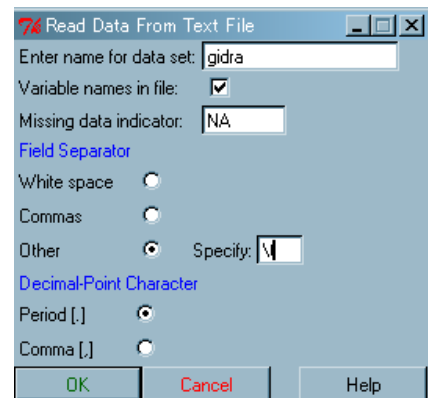
5 もう少し大きなデータなら

Nakazawa M, et al.: Serum biochemical data of the Gidra in lowland Papua New Guinea: Consideration of their normal ranges. J. Nutritional and Environmental Med., 10: 153-162, 2000. において、パプアニューギニア低地に居住するギデラと呼ばれる人々の血清生化学検査値の分析結果を報告した。SAS によって統計解析をしたデータファイルの元になったタブ区切りテキストファイル（変数名が1行目に入っていて欠損データのところには何も入っていないもの）を、エディタの検索置換機能を使って欠損をNAにしたファイルを作成し、これを gidra.txt とした。R Commander によって TABLE 3 と TABLE 4 を求める方法を示す。

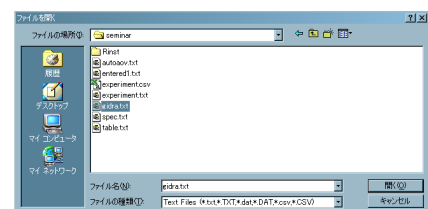
R Commander を呼び出しておき、まずプルダウンメニューの Data から Import Data, from text file を選ぶ。



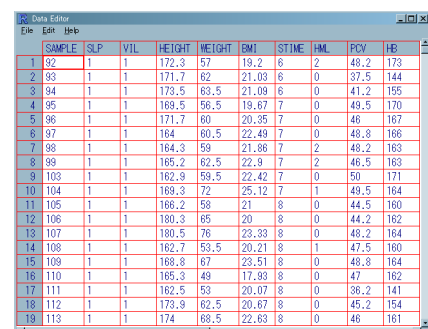
テキストファイルをどのような形式のデータとして読み込むかを指定するウィンドウが開くので設定後 OK をクリックする。タブ区切りテキストの場合は、Field Separator は Other で Specify のところを \t とする。



読み込みたいテキストファイルを指定して、「開く」をクリックする。

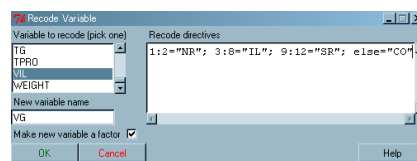


Edit data set をクリックするとデータエディタが開いて、ちゃんと読めていることが確認できる。

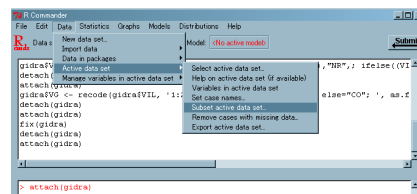


SAMPLE	SLP	WTL	HEIGHT	HEIGHT	BMI	STIME	HM	PCY	HB	
1	92	1	1	172.5	57	19.2	8	2	49.2	173
2	93	1	1	171.7	62	21.03	6	0	37.5	144
3	94	1	1	173.5	63.5	21.09	6	0	41.2	195
4	95	1	1	169.5	56.5	19.67	7	0	49.5	170
5	96	1	1	171.7	60	20.35	7	0	46	167
6	97	1	1	164	60.5	22.49	7	0	48.8	166
7	98	1	1	164.3	59	21.86	7	2	48.2	163
8	99	1	1	165.2	62.5	22.9	7	2	46.5	163
9	103	1	1	162.9	59.5	22.42	7	0	50	171
10	104	1	1	169.3	72	25.12	7	1	49.5	164
11	105	1	1	166.2	58	21	8	0	44.5	160
12	106	1	1	180.3	65	20	8	0	44.2	162
13	107	1	1	180.5	76	23.33	8	0	48.2	164
14	108	1	1	162.7	53.5	20.21	8	1	47.5	160
15	109	1	1	168.8	67	23.51	8	0	48.8	164
16	110	1	1	165.3	49	17.93	8	0	47	162
17	111	1	1	162.5	53	20.07	8	0	38.2	141
18	112	1	1	173.3	62.5	20.67	8	0	45.2	154
19	113	1	1	174	68.5	22.63	8	0	46	161

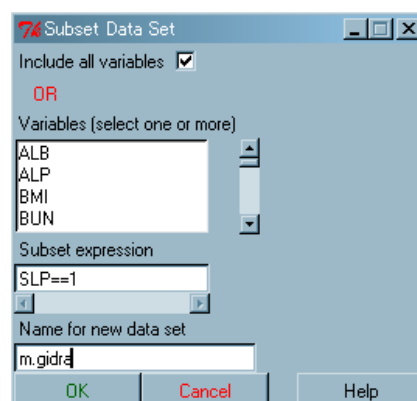
TABLE 3 の分析では男女別に村落群間の比較が必要なので、まず村落番号から村落群を factor として再コーディングする。プルダウンメニューの Data で Manage variables in active data set から recode を選ぶとウィンドウが開くので、再コード化の元にする変数、新しい変数名、再コード化定義式を設定する。Make new variable a factor にチェックが入っていないと Kruskal-Wallis 検定ができないので注意。



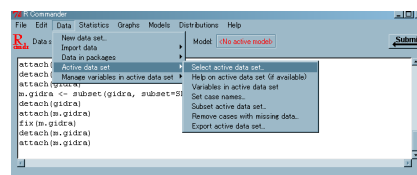
次に男女別にデータのサブセットを作る。Data の Active data set の Subset active data set を使う。元データファイルには、SLP という変数で、1 が男性、2 が授乳中でない女性、3 が授乳中の女性とコードされている。



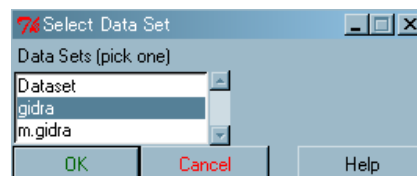
開くウィンドウで、Subset expression のところに SLP==1 とし、Name for new data set のところに m.gidra とした。



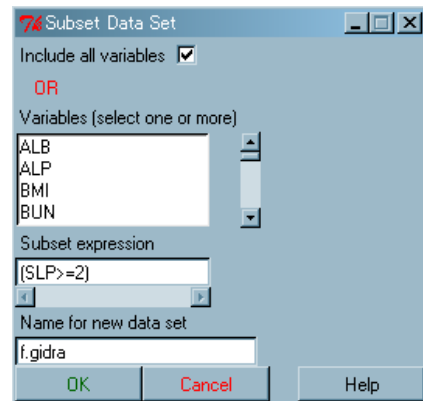
女性のデータセットを作るには、一度全体のデータを active にし直さなくてはならないので、Data の Active data set から Select active data set を選んで、



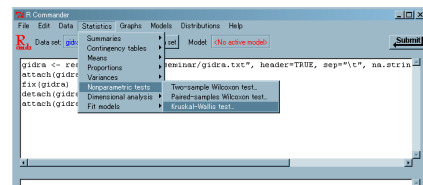
開くウィンドウで gidra を選んで OK をクリックする。



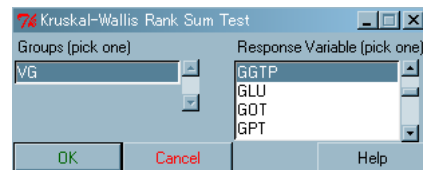
男性のサブセットを作ったときと同様にし、Subset expression で (SLP>=2) とすれば、女性のサブセット f.gidra を作れる。この段階では f.gidra が active になっている



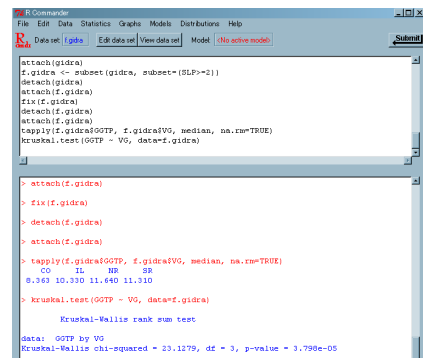
プルダウンメニューの Statistics で Nonparametric tests から Kruskal Wallis test... を選ぶ。



開くウィンドウで Groups として VG, Response variable として GGTP を選んで OK をクリックすれば γ -GTP レベルに村落群間で差がないかの検定ができる。



検定結果の前に村落群ごとの中央値も自動的に計算される。



Response variable は1つしか指定できない。いちいちメニューから選び直すのは面倒なので、R Commander の上側ウィンドウでカットアンドペーストを使って編集後、実行させたい部分を選択して Submit をクリックすると、選んだ部分のコマンドが一気に実行できる。

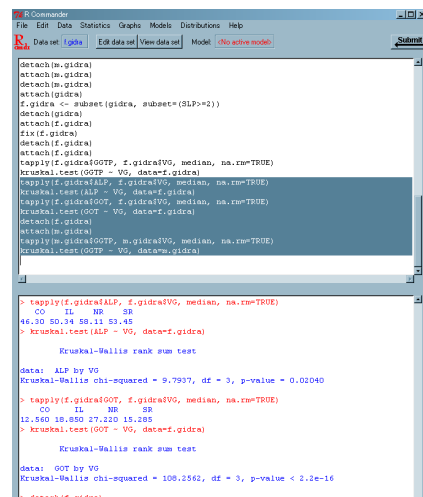
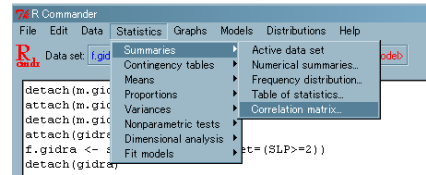
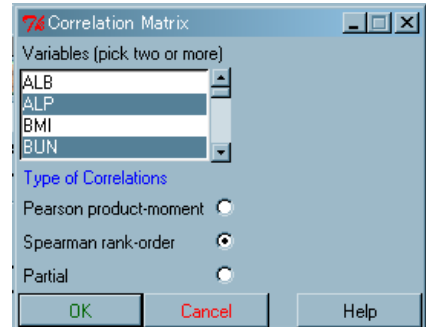


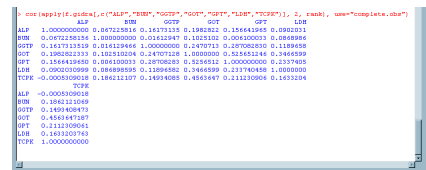
TABLE 4 の Spearman の順位相関係数行列を計算するには、プルダウンメニュー **Statistics** の **Summaries** から **Correlation matrix...** を選ぶ。



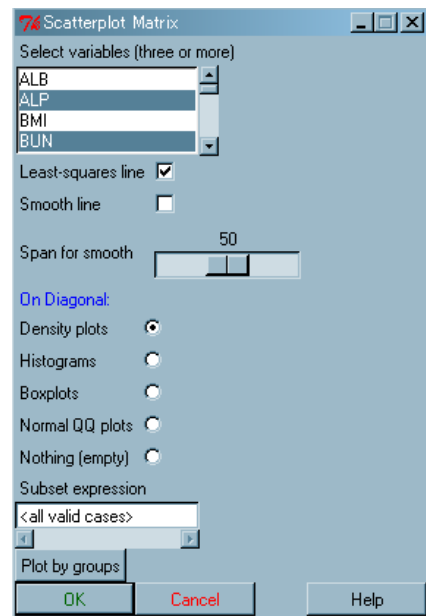
Type of Correlation で Spearman のところにチェックを入れ、計算したい変数を (いくつでも) クリックして選び、OK をクリックする。



相関係数行列が一気に計算され、R Commander の下側に表示される。残念ながら相関係数がゼロという帰無仮説の検定結果はこれでは出てこないなので、一つずつ計算するしかない。



プルダウンメニューの **Graphs** から選べば、**Scatterplot matrix** を作ることもできる。開くウィンドウで必要な情報を設定する。



得られる散布図行列。対角線上には各変数の分布が表示される。

