

R コマンドー入門*

John Fox

2006 年 8 月 26 日

日本語訳†

2008 年 8 月 28 日

1 R コマンドーを起動する

R が起動しているとき、*R Console* にコマンド `library(Rcmdr)` を入力して **Rcmdr** をロードすることにより、R コマンドーのグラフィカル・ユーザー・インターフェース (GUI) を起動することができる。Windows のもとで R コマンドーが適切に機能するには、R がシングル・ドキュメント・インターフェース (SDI) の形で設定されている必要がある。^{*1}パッケージをロードした後、*R Console* と R コマンドーのウィンドウは図 1 と図 2 のようになる。本稿におけるスクリーンイメージは、Windows XP の元で作成した。Windows の他のバージョン (当然、他のプラットフォームも) を利用している場合は、スクリーンの様子は異なる。^{*2}

R コマンドーと *R Console* のウィンドウは、デスクトップ上で自由に移動させることができる。R コマンドーのメニューとダイアログボックスを利用して、データを読み込み、処理、分析を行うことができる。

- R コマンドーの GUI で利用された R のコマンドは、R コマンドーの主ウィンドウの上部にあるテキストウィンドウ (スクリプトウィンドウという) に表示される。スクリプトウィンドウまたは *R Console* のプロンプト (`>`) の位置に直接キー入力してもよい。
- 出力内容は、出力ウィンドウという 2 番目のテキストウィンドウに表示される。
- 下部にあるグレーのウィンドウ (メッセージ というラベルがついている) は、エラーメッセージや警告、

* 本マニュアルは、Fox(2005) に基づく。連絡は jfox@mcmaster.ca まで。

† 翻訳版における画像等は、R-2.6.2, Rcmdr.1.3-5 に基づく。本翻訳に関する問い合わせ等は、荒木孝治 (arakit@kansai-u.ac.jp) まで。

^{*1} R の Windows バージョンは、通常、マルチ・ドキュメント・インターフェース (MDI) で動く。それは、*R Console* ウィンドウ、セッション中に作られるグラフィカル・デバイス・ウィンドウ、他の R のプロセスに関連したウィンドウを含む。これに対してシングル・ドキュメント・インターフェース (SDI) では、*R Console* とグラフィカル・デバイスは主ウィンドウ内に含まれない。R を SDI モードで動かすには、いくつかの方法がある。例えば、R の `etc` サブディレクトリにある `Rconsole` ファイルを編集するか、R のデスクトップアイコンのプロパティのショートカットタブのリンク先に `--sdi` を追加すればよい。**Rcmdr** パッケージの限界は、`tcltk` パッケージの限界による。

^{*2} **Rcmdr** は、R とともに配布されている推奨パッケージに加えていくつかのパッケージを必要とし、**Rcmdr** の起動とともにこれらをロードする。**Rcmdr** および必要なパッケージ、他の多くの追加パッケージは、Comprehensive R Archive Network (CRAN) at (<http://cran.r-project.org/>) よりダウンロードできる。

これらのパッケージがインストールされていないと、**Rcmdr** はインターネットまたはローカルファイル (例えば、CD-ROM) からインストールするかどうか聞いてくる。Windows の “R GUI” 経由で **Rcmdr** をインストールする場合、**Rcmdr** が必要とするパッケージも自動的にインストールされる。より一般的には、パッケージ **Rcmdr** と他の必要なパッケージは、関数 `install.packages` を用いて、引数を `dependencies = TRUE` と設定しておくことによりインストールすることが可能である。

Dirk Eddelbuettel のおかげで、Debian Linux のユーザーはコマンド `$ apt-get install r-cran-rcmdr` を用いるだけで、**Rcmdr** と他の必要なパッケージ全てをインストールすることができる。いずれにしても、Linux システム上でパッケージ **Rcmdr** をビルドし、インストールする方が一般に簡単である。マッキントッシュシステムの OS/X 上での作業はより面倒である。それは、**Rcmdr** が依存するパッケージ `tcltk` が Tcl/Tk のインストールを要求し、R が X-Window 上で作動するからである。

Rcmdr で 3 次元グラフィックスを利用するには、`rgl` パッケージをインストールする必要がある。

インストールに関するその他の情報は、R コマンドーのウェブページ (<http://socserv.socsci.mcmaster.ca/jfox/Misc/Rcmdr/index.html>) にある。

あるいは、図 2 のスタートアップメッセージのような情報を表示する。

- グラフを作成すると、別のグラフィックスデバイスのウィンドウに表示される。

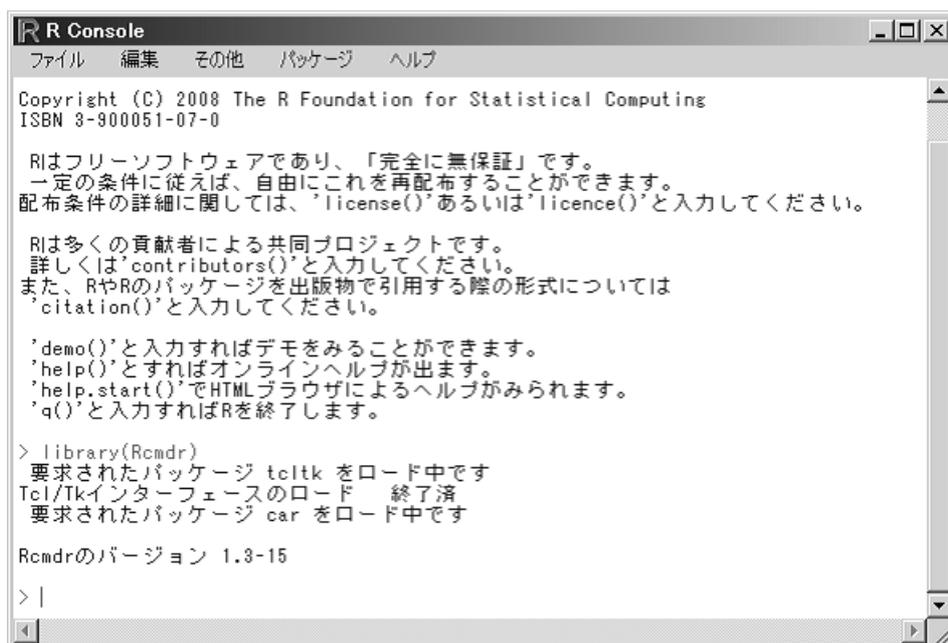


図 1 Rcmdr ロード後の R Console ウィンドウ

R コマンダーの上部には次のようなメニューがある。

- ファイル スクリプトファイルを読み込んだり保存したりする。出力、R ワークスペースの保存、終了の機能を持つ。
- 編集 スクリプトウィンドウと出力ウィンドウの内容を編集（切り取り、コピー、ペースト他）するためのメニュー。スクリプトウィンドウまたは出力ウィンドウで右クリックすると、編集のコンテキストメニューが表示される。
- データ データの読み込み、処理のためのメニュー。
- 統計量 基本的な統計分析を行うためのメニュー。
- グラフ 簡単な統計グラフを作成するためのメニュー。
- モデル 統計モデルに対する数値による要約、信頼区間、仮説検定、診断、グラフのためのメニュー。残差といった診断の統計量をデータセットに追加する機能を持つ。
- 分布 標準的な分布の確率、分位点を求め（数値表の代わりに用いることができる）、グラフを作成する。分布からのサンプルを得ることも可能。
- ツール パッケージ Rcmdr とは関係のないパッケージをロードしたり、オプションを設定したりするためのメニュー。
- ヘルプ R コマンダーの情報（このマニュアルを含む）を得るためのメニュー。なお、R コマンダーの各ダイアログボックスは Help ボタンを持つ（下記参照）。

R コマンダー（バージョン 1.2-0）の完全なメニューツリーを以下に示す。ほとんどのメニュー項目では、本稿で例示するように、ダイアログボックスを表示する。現在の状況において利用できないものはグレイで表示され、選択できないようになっている。

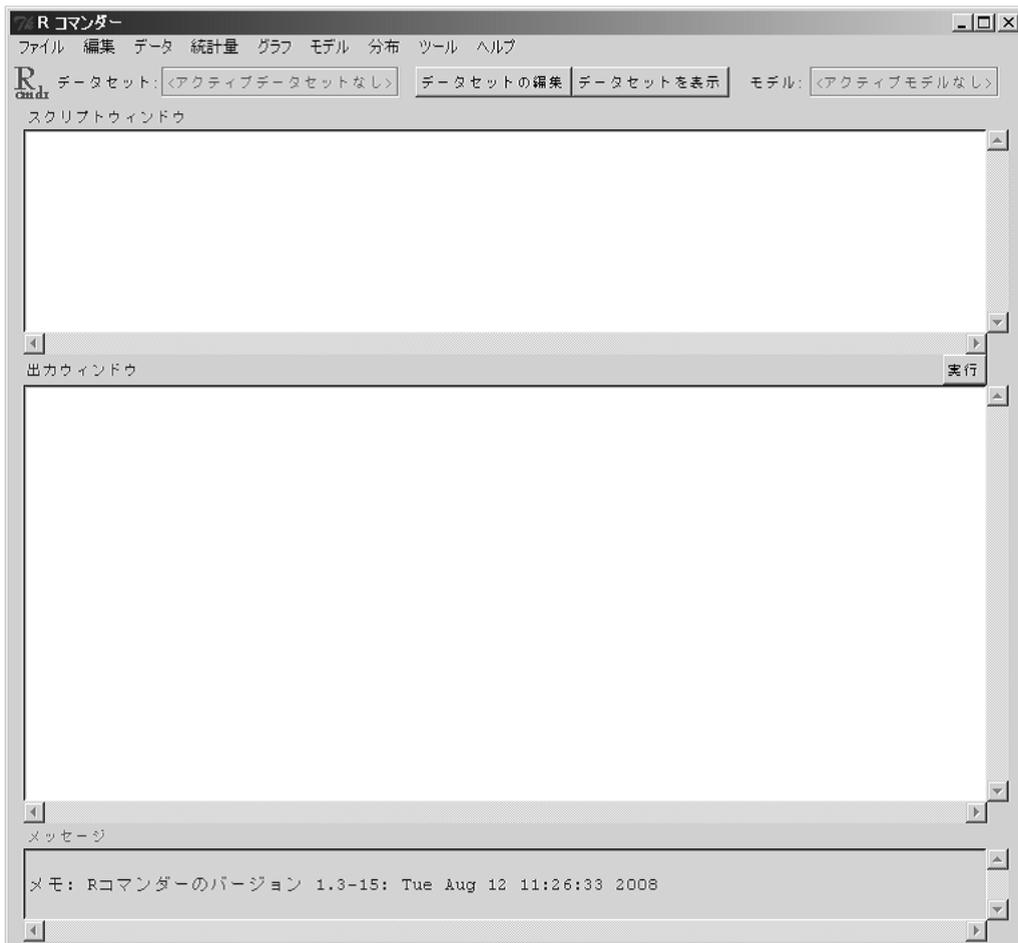
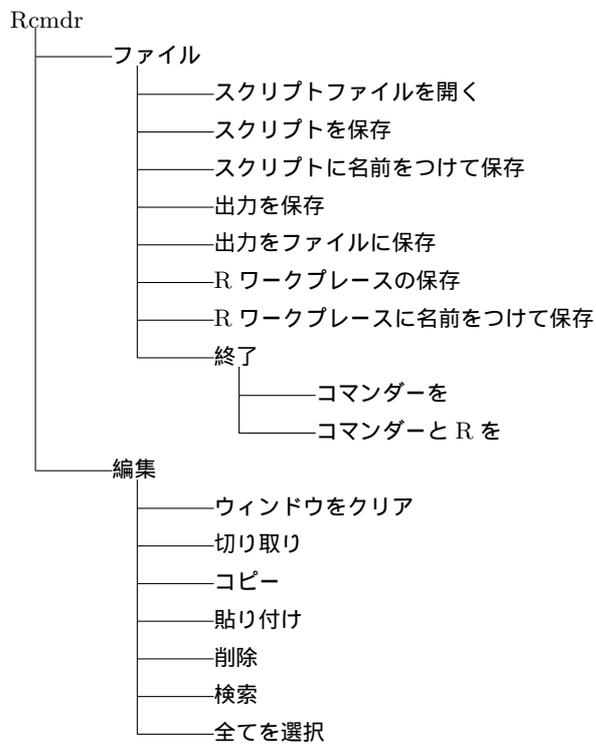
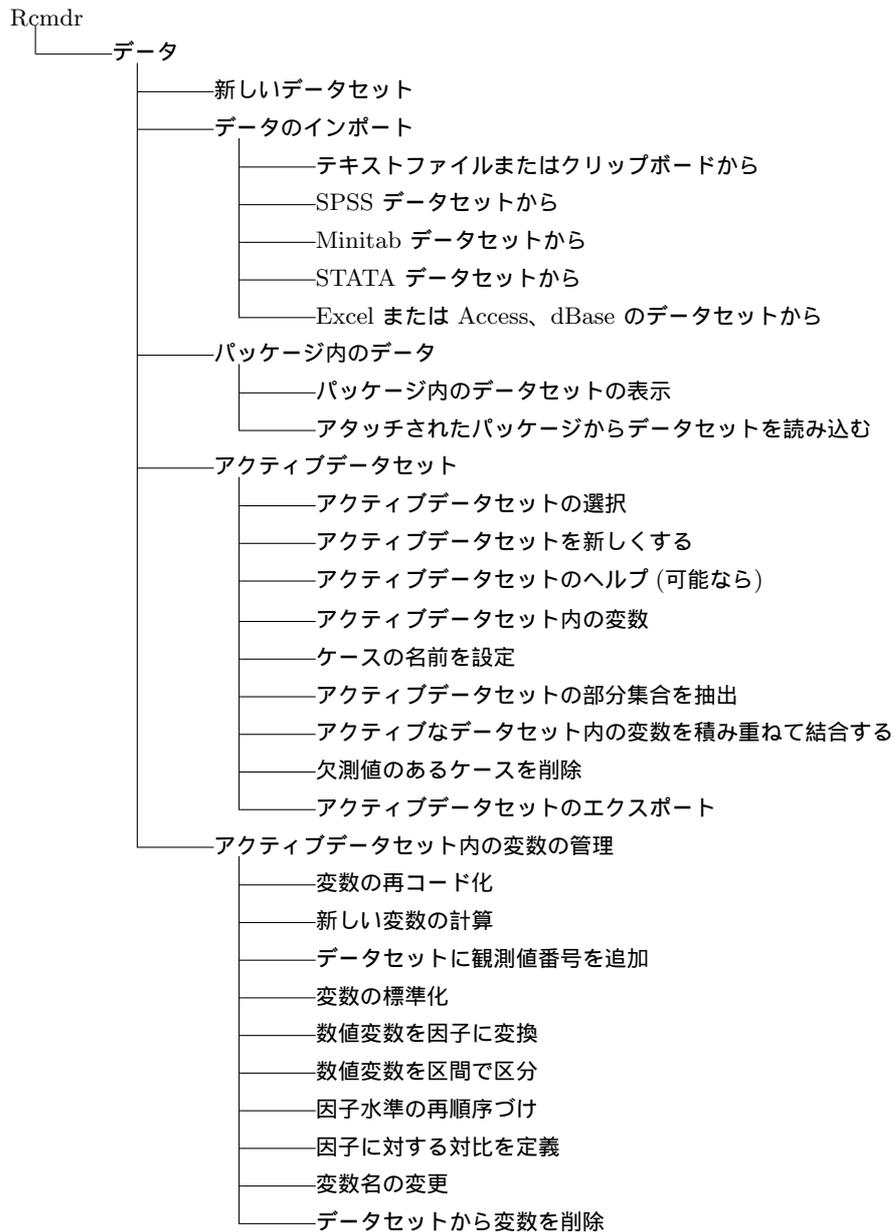


図 2 起動直後の R コマンダーのウィンドウ





Rcmdr

統計量

要約

- アクティブデータセット
- 数値による要約
- 頻度分布
- 統計量の表
- 相関行列
- 相関の検定

分割表

- 2 元表
- 多元分割表
- 2 元表の入力と分析

平均

- 1 標本の t 検定
- 独立サンプルの t 検定
- 対応のある t 検定
- 1 元配置分散分析
- 多元配置分散分析

比率

- 1 標本の比率の検定 (母不良率の検定)
- 2 標本の比率の検定

分散

- 分散の比の F 検定
- バートレットの検定
- ルビーンの検定

ノンパラメトリック検定

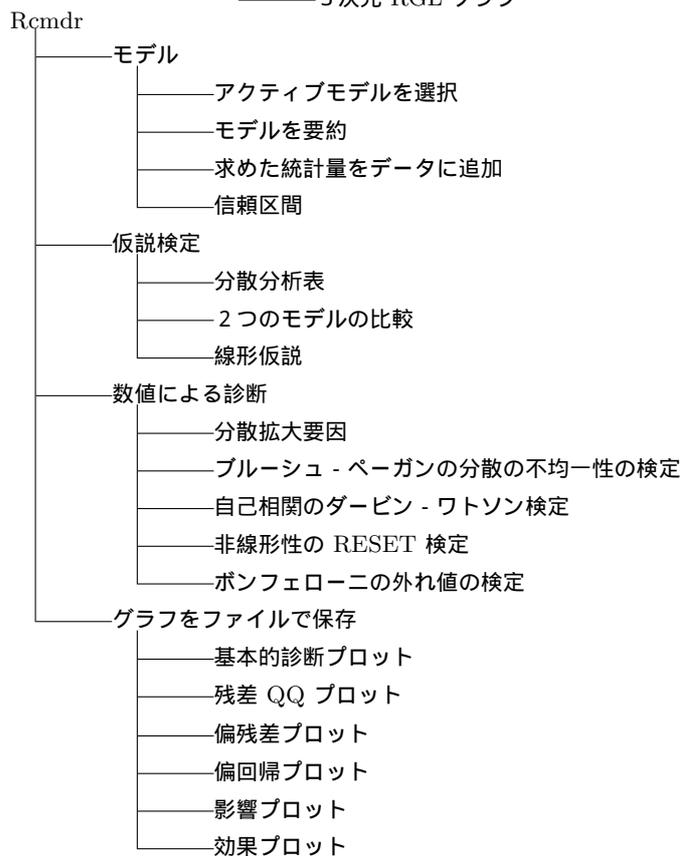
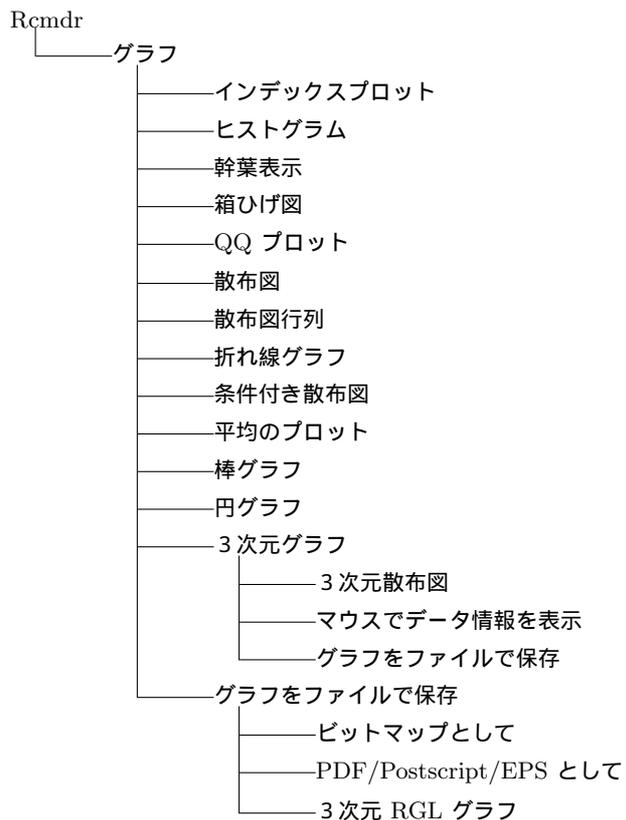
- 2 標本ウィルコクソン検定
- 対応のあるウィルコクソン検定
- クラスカル - ウォリスの検定

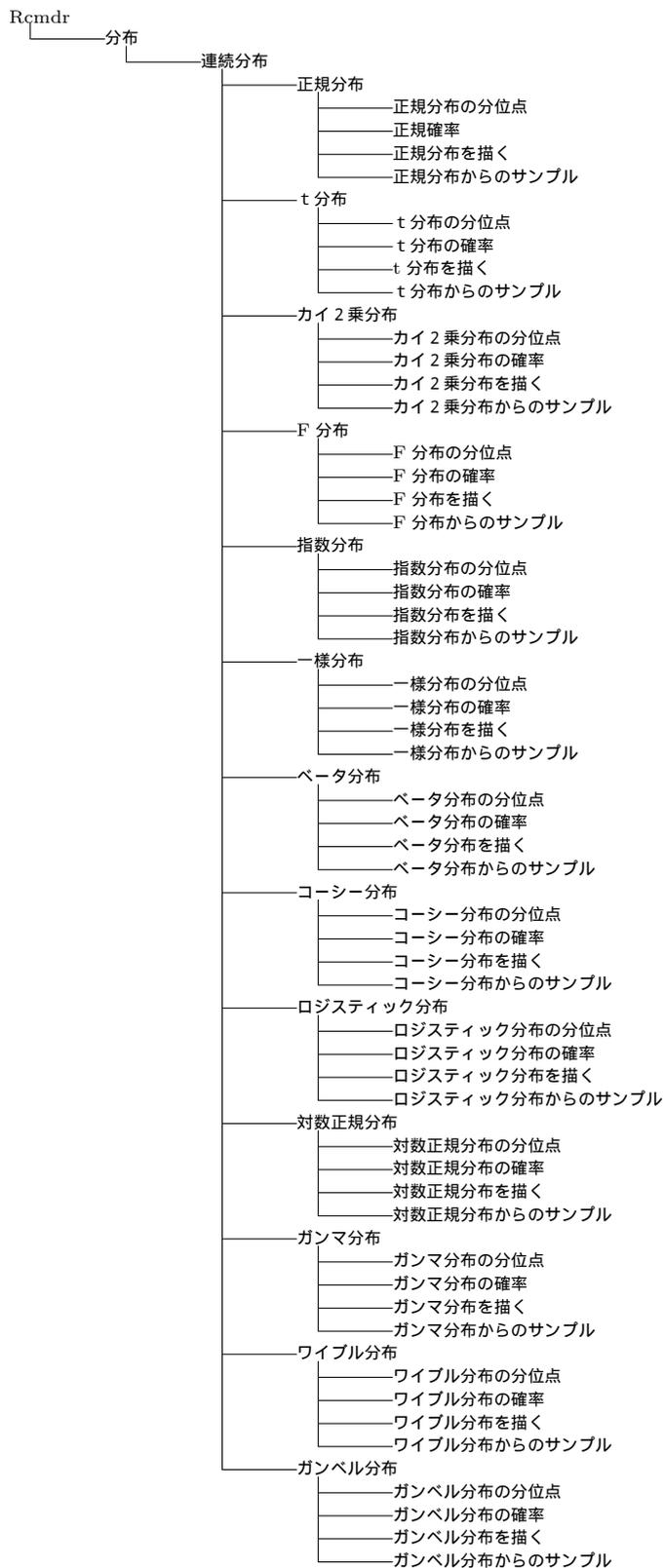
次元解析

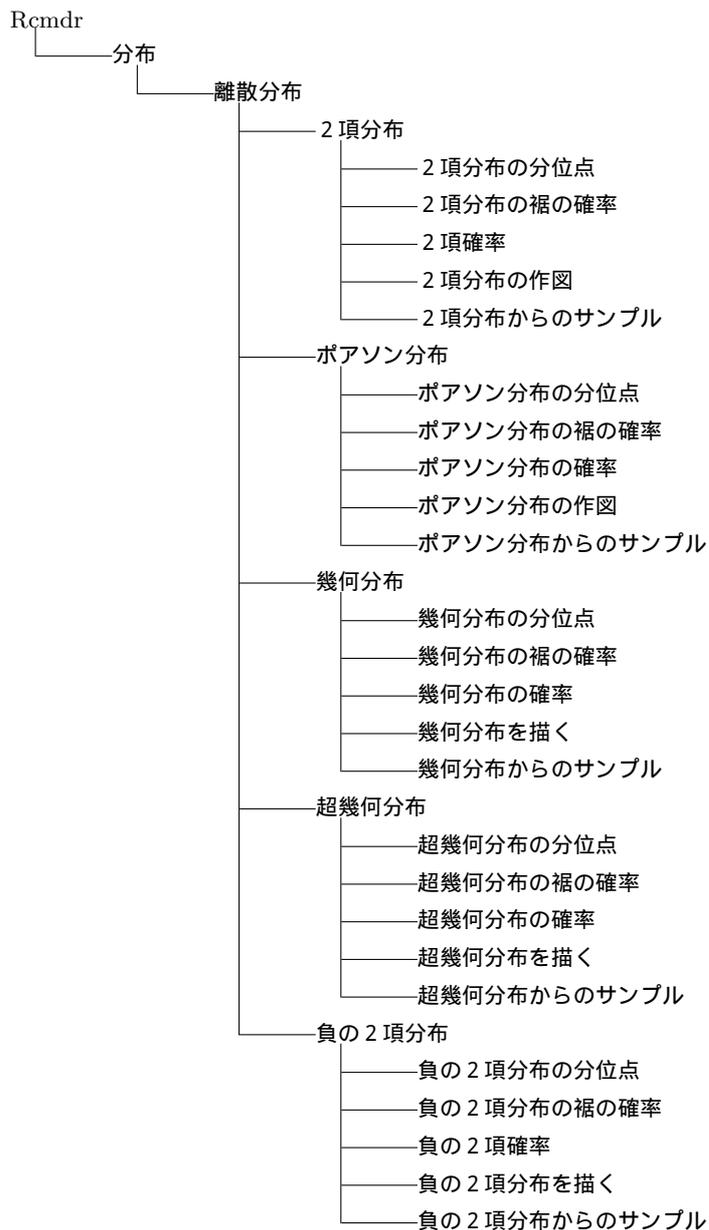
- 尺度の信頼性
- 主成分分析
- 因子分析
- クラスタ分析
 - k-平均クラスタ分析
 - 階層的クラスタ分析
 - 階層的クラスタリングの要約
 - 階層的クラスタリングの結果をデータセットに保存

モデルへの適合

- 線形回帰
- 線形モデル
- 一般化線形モデル
- 多項ロジットモデル
- 比例オッズロジットモデル







R コマンダーのインターフェースは、メニューとダイアログに加えて、他にいくつかの要素を持つ。

- メニューの下にツールバーがある。
 - 最も左にあるボタンはアクティブデータセットの名前を表示する。最初、アクティブデータセットはない。このボタンを押すと、現在メモリにあるデータセットを（2つ以上あれば）選択することができる。R コマンダーのメニューとダイアログのほとんどは、アクティブデータセットに関連している（ファイル、編集、分布メニューは別）。
 - アクティブデータセットを編集したり表示したりするためのボタンが2つある。他のオペレーション

が実行されているとき、データセットビューアを開いたままにすることができる。^{*3}

— フラットボタンには、アクティブな統計モデル — 線形モデル (線形回帰モデル), 一般化線形モデル, 多項ロジットモデル, 比例オッズモデル — の名前が表示される。^{*4} 最初, アクティブモデルはない。メモリにモデルが2つ以上あるとき, このボタンをクリックすることにより, 選択することができる。

- ツールバーの直下に, スクロール可能な大きなテキストウィンドであるスクリプトウィンドウがある。GUI で生成されたコマンドは, このウィンドウに表示される。スクリプトウィンドウのテキストを編集したり, R のコマンドをキー入力したりすることができる。スクリプトウィンドウの右下にある 実行 ボタンをクリックする (*Ctrl-r* でもよい) ことにより, カーソルがある行のスクリプトを実行することができる。数行を選択した場合 (左クリックしながらドラッグして), 実行 ボタンにより全てを実行することができる。スクリプトウィンドウに入力したコマンドは, 複数行に渡ってもよいが, そのときは, 2行目以降を1つ以上の空白またはタブにより字下げする必要がある。キー入力の組合せ *Ctrl-a* により, スクリプトウィンドウ内の全てのテキストを選択することができ, *Ctrl-s* によりウィンドウの内容を保存するためのダイアログボックスを表示することができる。
- スクリプトウィンドウの下に, スクロール可能で編集可能な出力のためのテキストウィンドウがある。このウィンドウに送られたコマンドは赤で, 出力結果はダークブルーで表示される (*R Console* と同じ)。
- 一番下には, 小さなグレーのメッセージを表示するテキストウィンドウがある。エラーメッセージは赤で, 警告は緑で, 他のメッセージはダークブルーで表示される。エラーと警告はベル音によっても知らされる。次のオペレーションに移ると, メッセージは画面上方にスクロールされて消える。

パッケージ *Rcmdr* を起動すると, *R Console* を最小化してもよい。R コマンドーのウィンドウのサイズを通常の方法で変更したり, 最小化したりすることができる。R コマンドーのサイズを調整すると, 以降の出力は出力ウィンドウのサイズに自動的に調整される。

R コマンドーでは設定を柔軟にできる。ここで示したのは標準設定である。設定の変更は, ツール → オプション メニューを用いるか, より徹底的に行うには, R のオプションで行う。^{*5} 詳細については, *Rcmdr* のヘルプファイルを参照のこと。

2 データの入力

R コマンドーのほとんどの手続きでは, アクティブデータセットがあるという前提で実行される。^{*6} R コマンドーが起動した直後には, アクティブデータセットはない。

R コマンドーによるデータの入力方法を次に示す。

- データ → 新しいデータセット... より, 直接入力することができる。データセット数が少ないときは, これでよい。
- プレーンテキスト (“ascii”) ファイルまたはクリップボード, 他の統計パッケージ (Minitab や SPSS, Stata), Excel または Access, dBase からインポートすることができる。
- R パッケージに含まれるデータセットを, その名前をキー入力するか (名前を知っていれば), ダイアロ

^{*3} David Firth のパッケージ *relimp* にある関数 *showData* によって与えられたデータビューアは, 多くの変数を持つデータセットを表示することができる。変数の数の限界 (初期設定は 100) を越えた場合, データセットを表示する代わりに R のデータエディタを利用する。変数の数に関係なくデータエディタを利用するには, 限界値を 0 に設定する。詳細については, R コマンドーのヘルプファイルを参照のこと。データセットを表示するのにデータビューアを利用することの欠点は, 他のオペレーションが実行されている間, エディタウィンドウが表示されないことである。

^{*4} 必要なダイアログボックスとメニューアイテムを準備し, R の *etc* ディレクトリにある *model-classes.txt* を編集することにより, 別の統計モデルを付加することができる。

^{*5} 3つの点 (...) が付いているメニュー項目を選択すると, ダイアログボックスが表示される。これは標準的な GUI の仕様である。本稿では, → はメニュー項目やサブメニューを選択することを意味する。

^{*6} メニュー統計量 → 分割表の2元表を入力して分析... の場合と同様, 分布 メニューで選択された手続きはそうではない。

グボックスでデータセットを選択することにより読み込むことができる。

2.1 テキストファイルからデータを読み込む

例として、データファイル Nations.txt を取り上げる。^{*7} このデータの最初の数行は次のようになっている。

```
TFR contraception infant.mortality GDP region
Afghanistan          6.90   NA   154   2848   Asia
Albania              2.60   NA    32    863   Europe
Algeria              3.81   52    44   1531   Africa
American-Samoa       NA     NA    11    NA    Oceania
Andorra              NA     NA    NA    NA    Europe
Angola               6.69   NA   124    355   Africa
Antigua              NA     53    24   6966   Americas
Argentina            2.62   NA    22   8055   Americas
Armenia              1.70   22    25    354   Europe
Australia            1.89   76     6  20046   Oceania
. . . .
```

- ファイルの第 1 行に変数名がある。これらは、TFR (出生率で、女性 1 人当たりの子供の数)、contraception (既婚女性当たりの避妊具利用率 (%))、infant.mortality (生児 1000 人当たりの乳児死亡率)、GDP (国民総生産、単位は US ドル)、region である。
- 2 行目以下には、国単位でデータ値が入力されている。データ値は余白 (1 つ以上の空白またはタブ) で区切られている。データ値は縦に並んでいると見やすいが、そうである必要はない。データ行が国名で始まっていることに注意。これをデータセットの行名としたいので、国名に対応する変数名を入れていない。すなわち、変数名は 5 つだが、データ値は 6 つある。このような場合、R は各行の最初の値を行名として取り扱う。
- データ値には欠測値がある。R では、欠測値のコードとして NA (not available の意味) を用いるのがよい。
- TFR, contraception, infant.mortality, GDP は数値 (量的) 変数である。これに対して、region は地域名が入力されている。このデータが読み込まれると R は region を 因子、つまり質的変数として取り扱う。R コマンドは、数値変数と因子とを区別する。

データファイルを R に読み込むには、データ → データのインポート → テキストファイルから... を R コマンドのメニューから選択する。この操作により、図 3 に示す テキストファイルからデータを読み込む というダイアログボックスが表示される。データセットのデフォルト名は Dataset であるが、ここでは Nations に変更している。

R において、データセット名は大文字または小文字のアルファベット (または、ピリオド “.” で始まり、以降、全アルファベット、アンダースコア (_), 数字 (0-9) で構成される必要がある。データセット名に空白を含むことはできないことに注意。また、R では大文字と小文字を区別する。そのため、nations, Nations, NATIONS 等は区別されるので、違ったデータセットに対する名前として利用できる。

テキストファイルからデータを読み込む ダイアログで OK ボタンをクリックすると、図 4 に示すようなファイルを開くダイアログが表示される。ここでは、Nations.txt ファイルを読み込む状況を示している。ダイアログの 開く ボタンをクリックすると、データファイルが読み込まれる。データファイルが読み込まれると、そ

^{*7} このファイルは、パッケージ Rcmdr のサブディレクトリ etc にある。

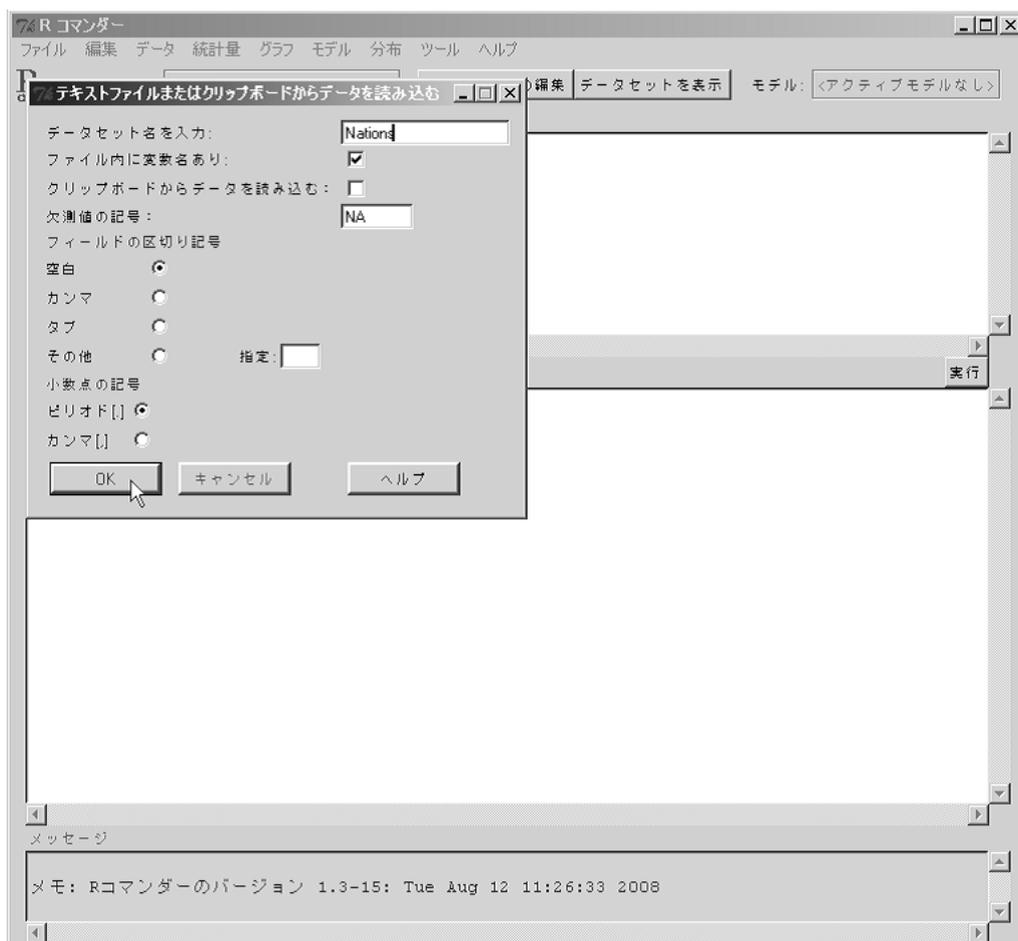


図3 Rcmdr のロード後の R Console ウィンドウ

これは R コマンダーのアクティブデータセットとなる。結果として、図 5 に示すように、読み込まれたデータセット名が R コマンダーのウィンドウの左上部にあるデータセットボタンに表示される。

データセットを表示 ボタンをクリックすると、図 5 に示すようなデータビューウィンドウが表示される。データセット Nations を読み込み、それを表示するコマンド (`read.table` と `showData`) が、スクリプトウィンドウと出力ウィンドウに表示されていることに注意 (データセットの表示により少し隠れてわかりにくい)。データセットが読み込まれてアクティブデータセットになると、メッセージウィンドウにコメントが表示される (さらにコマンド `showData` が実行されると消える)。

コマンド `read.table` は R の “データフレーム” を作る、これは、行をケース、列を変数とする表形式のデータセットのオブジェクトである。行は、ケースまたは観測対象を表し、列は変数である。R コマンダーのデータセットは、データフレームである。

2.2 データを直接入力する

R の表計算に似たデータエディタから直接データを入力するには、次のようにする。例として、Moor(2000) の Problem 2.44 からの非常に小さいデータセットを用いる。

- R コマンダーのメニューから、データ → 新しいデータセット を選択する。オプションとして、データセットの名前を、例えば Problem2.44 をダイアログボックスに入力し、OK をクリックする。(R のデータセット名に空白を使用できないことに注意。) これにより、何も入力されていないデータエディタのウィンドウが表示される。
- データエディタの最初の 2 列にデータを入力する。入力するセルを移動するには、キーボードにある矢印

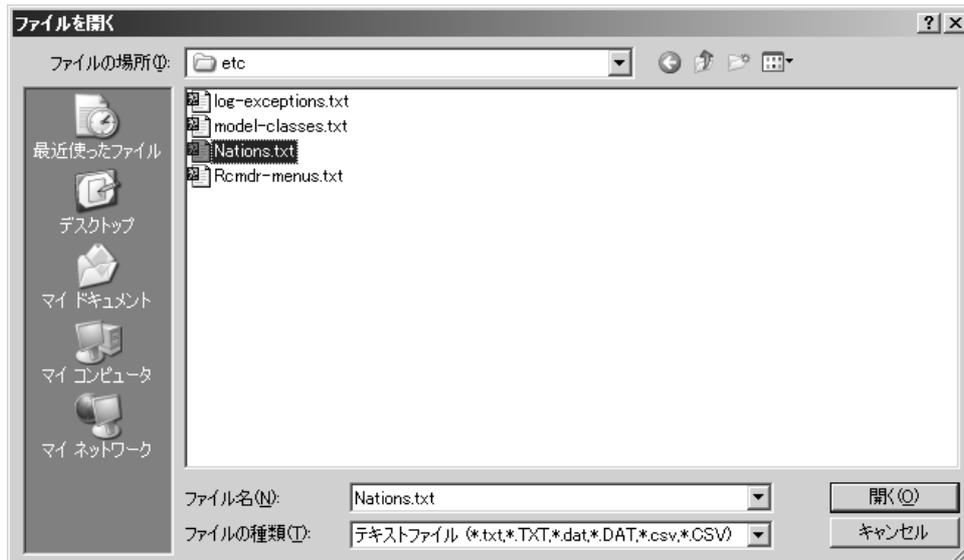


図 4 テキスト形式データファイルを読み込むためのファイルを開くダイアログボックス

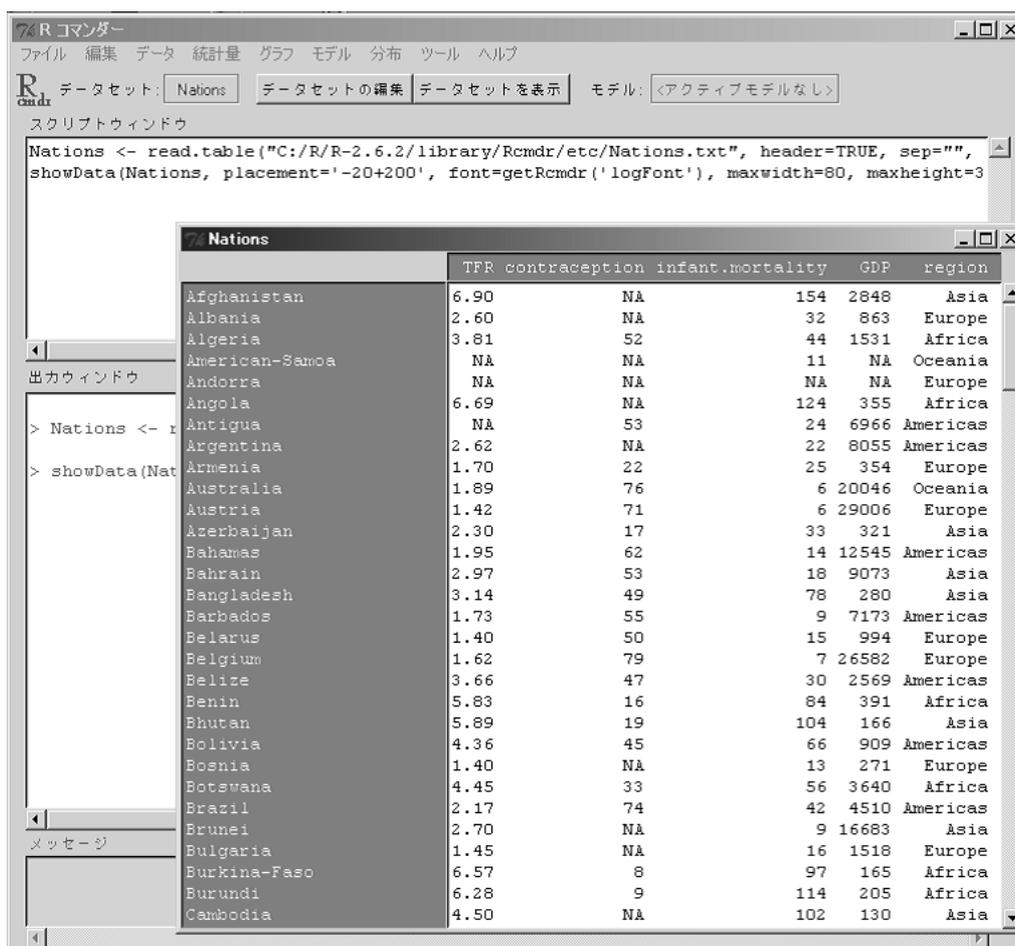
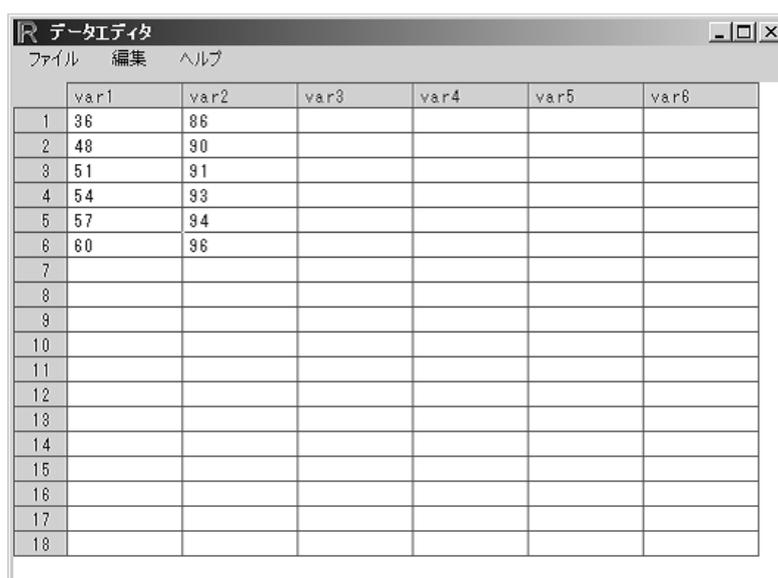


図 5 アクティブデータセットの表示

キーやタブキー，Enter キーを用いたり，マウスでポインターを移動して左クリックしたりする．データの入力が終了すると，図 6 のようになっている．

- 次に，第 1 列の上部にある名前 var1 をクリックする．これにより，図 7 のような変数エディタのダイアログボックスが表示される．
- 変数名 age を入力し，変数エディタのウィンドウの右上隅にある ×（閉じる）ボタンをクリックするか，Enter キーを押してウィンドウを閉じる．同様にして 2 列目の変数名を height に変更する．データエディタは図 8 のようになる．
- データエディタのメニューよりファイル → 閉じる を選択するか，データエディタの右上にある × ボタンをクリックする．これにより，入力したデータセットは R コマンダーのアクティブデータセットとなる．



	var1	var2	var3	var4	var5	var6
1	36	86				
2	48	90				
3	51	91				
4	54	93				
5	57	94				
6	60	96				
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						

図 6 データ入力後のデータエディタ

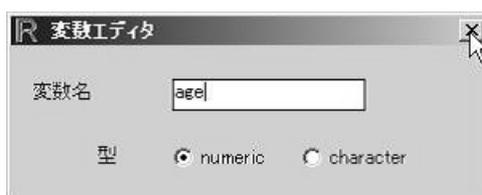


図 7 データエディタの変数名の変更のためのダイアログボックス

2.3 パッケージからデータを読み込む

パッケージの多くはデータを持っている．パッケージ内のデータセットは，データ → パッケージ内のデータ → パッケージ内のデータセットの表示 によりポップアップウィンドウに表示できる．また，データ → パッケージ内のデータ → アタッチされたパッケージからデータセットを読み込む... により R コマンダーに読み込むことができる．*8 結果として表示されるダイアログボックスを図 9 に示す．パッケージ内のデータセットの名前を知っているときは，それを直接入力してもよい．そうでなければ，リストボックスの右に表示されているパッケージ名をダブルクリックする．データセット名をダブルクリックすると，その名前がダイアログ中の

*8 パッケージ内のデータセットは必ずしも全てがデータフレームではなく，データフレームのみが R コマンダーに適していることに注意．データフレームでないデータを読み込もうとすると，メッセージウィンドウにエラーメッセージが表示される．

データセットを入力欄にコピーされる。⁹ R の他のパッケージをアタッチするには、ツール → パッケージのロード によって行う。

	age	height	var3	var4	var5	var6	var7
1	36	86					
2	48	90					
3	51	91					
4	54	93					
5	57	94					
6	60	96					
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							

図 8 変数名を変更した後のデータエディタウィンドウ

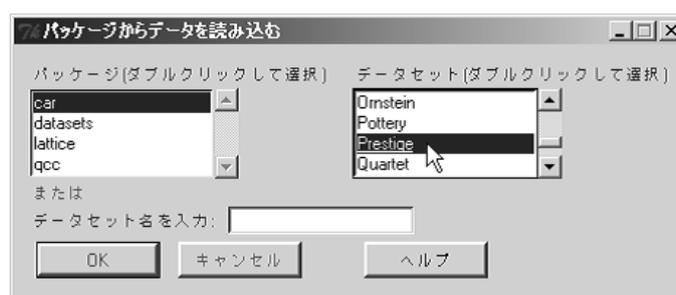


図 9 アタッチされたパッケージからのデータセットの読み込み

3 数値による要約の実行とグラフの作成

アクティブデータセットがあると、数値による要約の実行やグラフの作成を R コマンダーのメニューにより実行できる。基本的な例をいくつか示す。良い GUI というものはだいたい見れば分かるものである。R コマンダーがどのように機能するかを一度知ると、オンラインヘルプファイルを参照すれば、ほとんどトラブル無く利用できるだろう。

下記の例では、アクティブデータセットとして、前節で読み込んだ Nations を取り上げる。もしあなたが、前節で説明したように、Moor(2000) の 5 つのケースのデータセットをキー入力したり、パッケージ car から Prestige を読み込んだりしたならば、これらのどれかがアクティブデータセットとなっている。R コマンダーのウィンドウの上左部にあるアクティブデータセット名を表示するフラットなボタンをクリックし、現在メモリにあるデータセットのリストから選択することにより、アクティブデータセットを切り替えることができる。

統計量 → 要約 → アクティブデータセット により、図 10 に示す結果を求めることができる。データセット内の各数値変数 (TFR, contraception, infant.mortality, GDP) に対して、最小値と最大値、第 1 四分位数、第 3 四分位数、メディアン、平均、欠測値の数を表示する。質的変数である region に対しては、因子

⁹ R コマンダーでは一般に、リストボックス内のアイテムをダイアログの他の箇所にコピーする必要がある場合、ダブルクリックするだけでよい。

の各水準の観測数が表示される。データセット内に変数が 10 個以上ある場合は、計算を進めてよいかどうか問
い合わせるにより、不要な大量の出力を避けるようになっている。

同様に、統計量 → 要約 → 数値による要約 を選択することにより、図 11 に示すダイアログボックスが表
示される。このダイアログには数値変数のみが表示されている。因子 `region` が表示されていないのは、因子に
対して数値による要約を行っても意味がないからである。`infant.mortality` をクリックして選択し、`OK` をク
リックすると、次の結果が出力ウィンドウに表示される。^{*10}

```
> numSummary(Nations[, "infant.mortality"], statistics=c("mean", "sd", "quantiles"))
      mean      sd 0% 25% 50% 75% 100%   n NA
43.47761 38.75604  2  12  30  66 169 201  6
```

デフォルトでは、平均と標準偏差 (sd)、最小値、第 1 四分位数、メディアン、第 3 四分位数、最大値に対応
する分位点 (パーセント点) を表示する。n は有効な観測数で、NA は欠測値数である。

R コマンドのダイアログでは通常そうであるが、数値による要約には `OK`、`Cancel`、`Help` という 3 つの
ボタンがある。`Help` ボタンにより、ダイアログボックス自体のヘルプページまたはダイアログが利用する R の
関数のヘルプページを参照することができる。

数値による要約のダイアログボックスでは、因子の水準によって定義されるグループ内での要約情報、つまり
層別の要約情報を求めることもできる。層別して要約... をクリックすると、図 12 に示す 質的変数 ダイアログ
が表示される。データセット `Nations` には質的変数が 1 つしかないので、変数のリストには `region` のみが表
示されている。これを選択し、`OK` ボタンをクリックすると、層別して要約... ボタンが 層別変数: `region` へ
と変化する (図 13)。`OK` をクリックすると、次の結果が表示される。

```
> numSummary(Nations[, "infant.mortality"],
             groups=Nations$region, statistics=c("mean", "sd", "quantiles"))
      mean      sd 0% 25% 50% 75% 100%   n NA
Africa  85.27273 35.188095  7 61.00 85.0 111.00 169 55  0
Americas 25.60000 17.439713  6 12.00 21.5  36.00  82 40  1
Asia    45.65854 32.980001  5 22.00 37.0  72.00 154 41  0
Europe  11.85366  7.122363  5  6.00  8.0  16.00  32 41  4
Oceania 27.79167 29.622229  2  9.25 20.0  35.75 135 24  1
```

R コマンドの他のダイアログでも、同様にして層別変数を選択することができる。

R コマンドでグラフを作成することも簡単である。例えば、R コマンドのメニューより `グラフ →`
`ヒストグラム...` を選択することにより、図 14 に示すヒストグラムのダイアログボックスが表示される。
`infant.mortality` を選択し、`OK` をクリックすると、図 15 に示すヒストグラムを含むグラフィックスウィ
ンドウが表示される。

1 つのセッションの中でいくつかグラフを作成した場合、グラフィックスデバイス ウィンドウには、最新の
もののみが表示される。キーボードにある `Page Up` または `Page Down` キーにより、以前のグラフを呼び出す
ことができる。^{*11}

^{*10} 変数のリストボックスで 1 つの変数を選択するには、その名前を左クリックするだけでよい。状況によっては 2 つ以上の変数を選択
する必要が生じる。この場合、通常の Windows での方法を適用する。左クリックで変数を選択することができ、再度左クリックす
ると、取り消すことができる。`Shift` キーを押しての左クリックで選択を拡大することができる。`Ctrl` キーを押しての左クリックで
選択を追加することができる。

^{*11} R コマンドは、グラフ履歴メカニズムをスタートアップ時にオンにしている。この機能は Windows のみで利用可能である。グラフ
→ 3次元グラフ → 3次元散布図... で作られた動的な 3次元散布図は、特別な `RGL` デバイス中に表示される。同様に、モデル
→ グラフ → 効果プロット で作られる統計モデルに対する効果プロット (Fox, 2003) は、独立したグラフィックスデバイ
ス・ウィンドウに表示される。

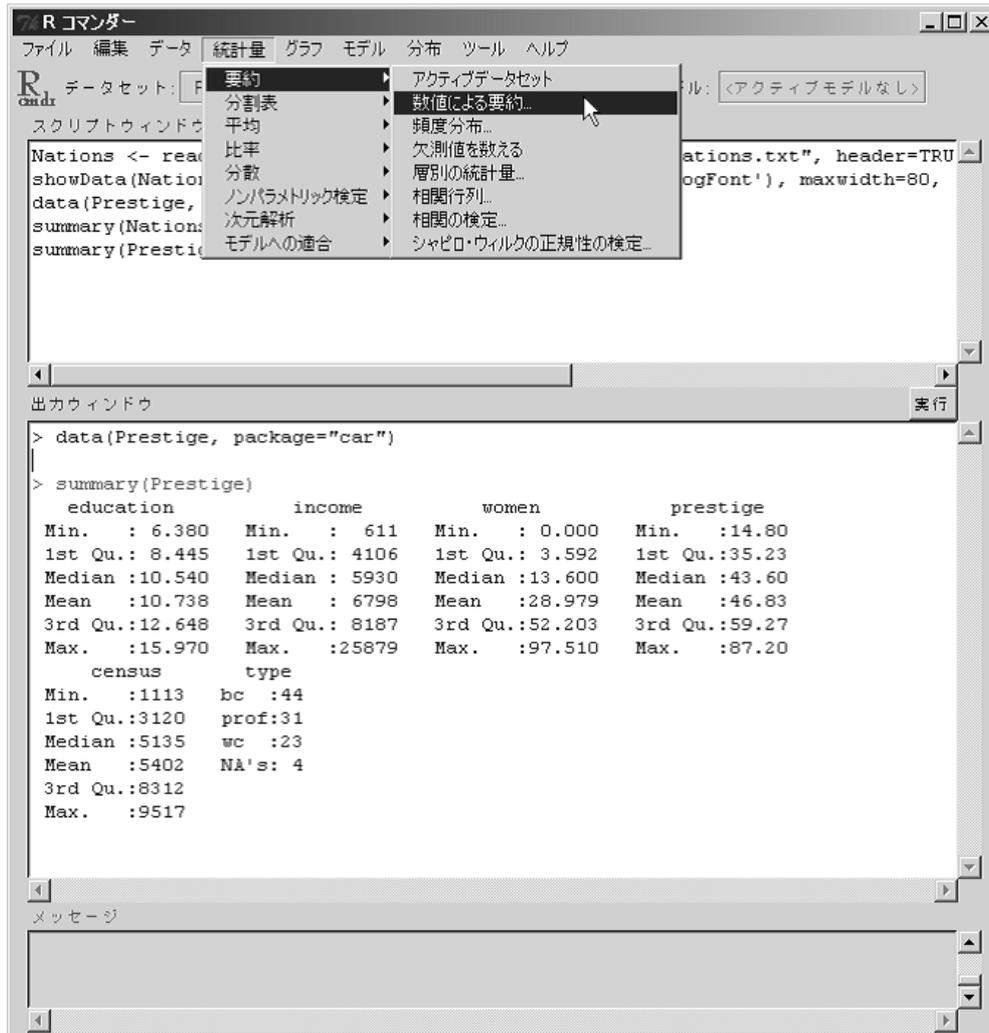


図 10 アクティブデータセットに対する変数の要約情報

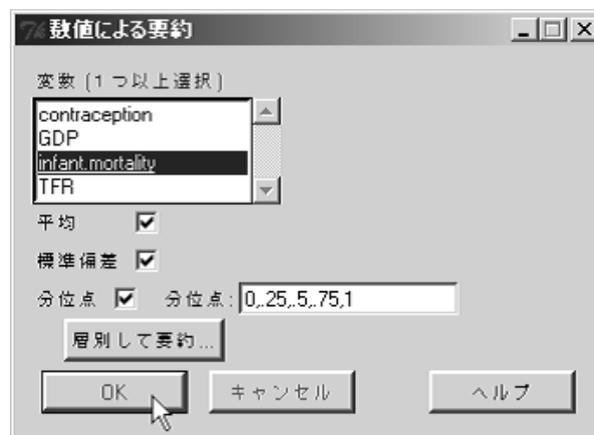


図 11 数値による要約のダイアログボックス

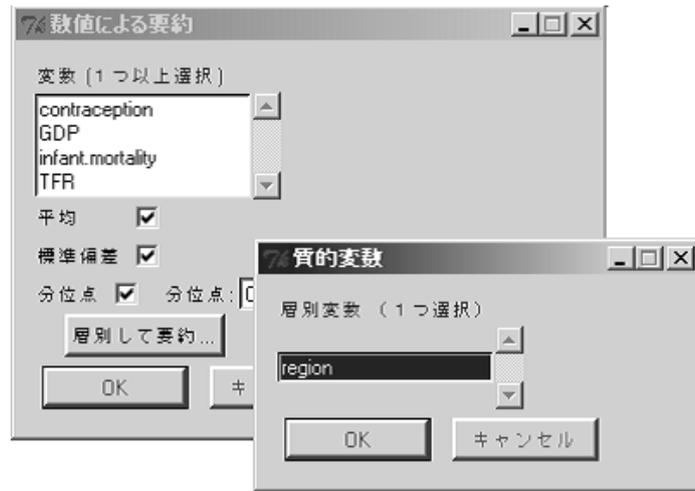


図 12 質的変数ダイアログボックスで層別変数を選択する

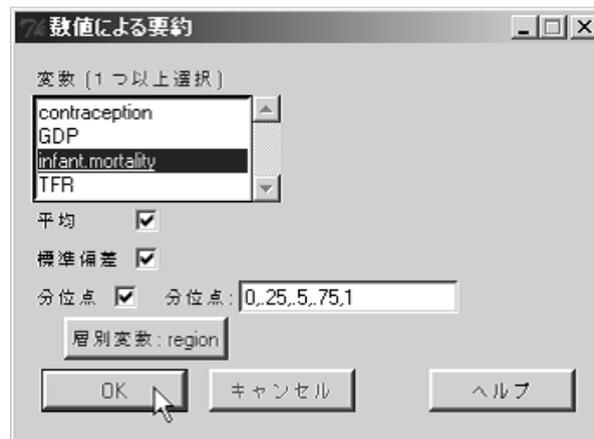


図 13 層別変数を指定した後の数値による要約ダイアログボックス



図 14 ヒストグラムダイアログボックス

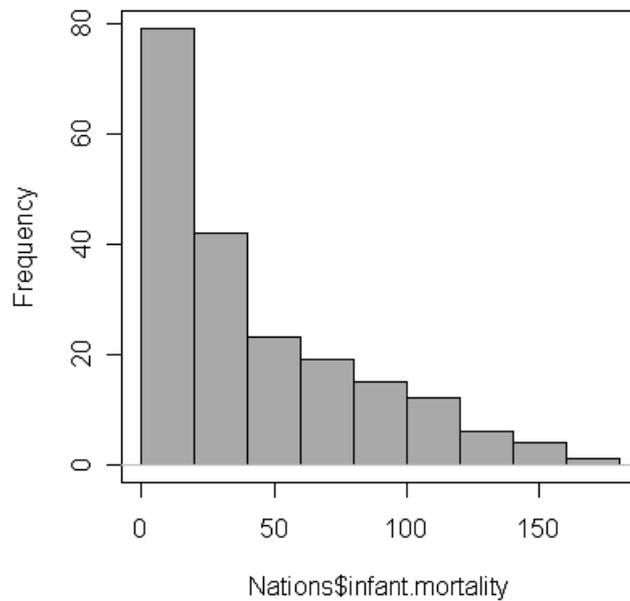


図 15 乳児死亡率のヒストグラムを含むグラフィカルウィンドウ

4 統計モデル

統計量 → モデルへの適合 メニューを用いて、R コマンドでいくつかの統計モデルを作成することができる。線形モデル（線形回帰... および 線形モデル... によって）、一般化線形モデル、多項ロジットモデル、比例オッズモデルである。[最後の2つは、Venables and Ripley(2002)による2つのパッケージ `nnet` と `MASS` による。] ダイアログボックスはモデルによってある程度異なるが（例えば、一般化線形モデルのダイアログには、分布族とこれに対応するリンク関数を選択する機能がある）、図 16 に示す線形モデルのダイアログボックスと共通する一般的な構造を持つ。^{*12}

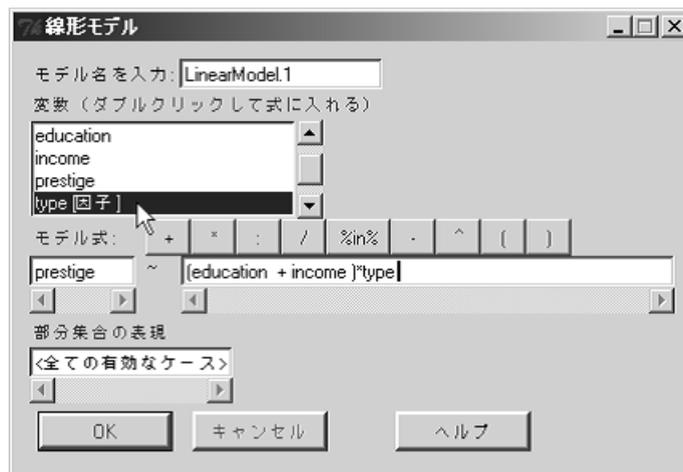


図 16 線形モデルダイアログボックス

- 変数のリストボックスにある変数をダブルクリックすると、モデル式にそれがコピーされる もし式の

^{*12} 例外は線形回帰ダイアログで、分析対象とするデータセットが持つ数値変数のリストから名前を選択することにより、説明変数と目的変数を指定することができる。以下の説明では、R のモデル式についてよく知っていることを仮定している。詳細については、R と一緒にインストールされる *Introduction to R* を参照。これは *R Console* の *Help* メニューからアクセスすることができる。

左辺が空白なら左辺に，そうでないならば右辺に（必要ならば，記号 + が前に添付される）. 変数リストの中で，因子については，名前の後ろに [因子] が追記されている .

- モデル式の上に並んでいるボタンは，式の右辺にオペレータや丸括弧を入力するために利用できる .
- モデル式のフィールドに直接キー入力してもよい . $\log(\text{income})$ といった項を入力するには，このようにする必要がある .
- モデル名は，今は `LinearModel.1` となっている . これは自動的に生成されるが，R の有効な名前の形で変更することもできる .
- 部分集合の表現 というボックスに R の表現を入力することができる . 入力があると，これが関数 `lm` の引数 `subset` に送られ，データセット内の観測値の部分集合に対してモデルが適用される . 部分集合の表現の 1 つの形として，各観測値に対して `TRUE` または `FALSE` を評価する論理的な表現がある . 例えば，`type != "prof"`（これは，データセット `Prestige` で，非専門的職業全てを指定する）である .

OK ボタンをクリックすると，次の結果が出力ウィンドウに表示される . `LinearModel.1` がアクティブモデルとなり，それがモデル ボタンに表示される .

```
> LinearModel.1 <- lm(prestige ~ (education + income)*type , data=Prestige)
```

```
> summary(LinearModel.1)
```

Call:

```
lm(formula = prestige ~ (education + income) * type, data = Prestige)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-13.462  -4.225   1.346   3.826  19.631
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.276e+00	7.057e+00	0.323	0.7478
education	1.713e+00	9.572e-01	1.790	0.0769 .
income	3.522e-03	5.563e-04	6.332	9.62e-09 ***
type[T.prof]	1.535e+01	1.372e+01	1.119	0.2660
type[T.wc]	-3.354e+01	1.765e+01	-1.900	0.0607 .
education:type[T.prof]	1.388e+00	1.289e+00	1.077	0.2844
education:type[T.wc]	4.291e+00	1.757e+00	2.442	0.0166 *
income:type[T.prof]	-2.903e-03	5.989e-04	-4.847	5.28e-06 ***
income:type[T.wc]	-2.072e-03	8.940e-04	-2.318	0.0228 *

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 6.318 on 89 degrees of freedom

(4 observations deleted due to missingness)

Multiple R-squared: 0.8747, Adjusted R-squared: 0.8634

F-statistic: 77.64 on 8 and 89 DF, p-value: < 2.2e-16

アクティブモデルに対するオペレーションは、モデル メニューから選択することができる。例えば、モデル → 仮説検定 → 分散分析表 により、次の結果が表示される。

```
> Anova(LinearModel.1)
Anova Table (Type II tests)

Response: prestige
      Sum Sq Df F value    Pr(>F)
education  1068.0  1 26.7532 1.413e-06 ***
income    1131.9  1 28.3544 7.511e-07 ***
type       591.2  2  7.4044 0.001060 **
education:type 238.4  2  2.9859 0.055574 .
income:type  951.8  2 11.9210 2.588e-05 ***
Residuals  3552.9 89
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

5 その他

5.1 出力の保存と印刷

R コマンダーのファイル メニューより、テキストの出力を直接保存することができる。これは、R のグラフィックスデバイスのウィンドウにおいて、ファイル メニューによりグラフを保存したり、印刷したりできるのと同じである。しかし、保存したい出力テキストやグラフをワープロ文書として保存しておく方が便利である。このようにすると、R の出力に注記や説明をつけて配布することができる。

ワードや Windows のワードパッドといったワープロを起動する。出力ウィンドウからテキストをコピーするには、テキストの範囲をマウスで指定し、編集 メニューから コピー を選択し（あるいは、*Ctrl-c* を押したり、ウィンドウ内で右クリックしてコンテキストメニューから コピー を選択したりする）、編集 → 貼り付け（または、*Ctrl-v*）によりワープロにテキストを貼り付ける。1つ注意すべきことは、R からのテキストの出力に対しては、Courier New といった等幅フォント（タイプライター体）を使うべきである。でないとうまく整列しない。

グラフをコピーするのと同じように、R のグラフィックスデバイスのメニューからファイル → クリップボードにコピー → メタファイルとして を選択する。そして、編集 → 貼り付け（または、*Ctrl-v*）によりワープロにグラフを貼り付ける。別の方法として、*Ctrl-w* により R のグラフィックスデバイスからグラフをコピーするか、グラフ上で右クリックして表示したコンテキストメニューよりメタファイルにコピー を選択してもよい。^{*13} R セッションの最後でも、作成したドキュメントを保存または印刷することができるが、これはあなたの仕事の注釈付きの記録となる。

テキストやグラフを保存するための別の方法として、R コマンダーのファイル メニューと、グラフ → グラフをファイルで保存 により保存することもできる。

^{*13} これらのメニューを調べると分かるように、様々なフォーマットでグラフをファイルのみならずクリップボードに保存することができる。ここで述べた方法は簡単であるが、グラフは高品位である。再度述べるが、ここでの説明は Windows のシステム自体にあてはまるものである。

5.2 R セッションの終了

セッションを終了する方法はいくつかある。例えば、R コマンダーのメニューのファイル → 終了 → コマンダーと R を を選択する。終了してもよいかどうかの確認の後、スクリプトと出力ウィンドウの内容を保存したいかどうかを聞かれる。また、R Console で、ファイル → 終了 を選択してもよい。この場合、R のワークスペース（すなわち、R がメモリに保存しているデータ）を保存したいかどうか聞かれる。通常は No を選択すればよい。

5.3 スクリプトウィンドウにコマンドを入力する

スクリプトウィンドウは、コマンドを編集・入力・実行するための簡単な機能を提供する。R コマンダーが生成したコマンドは、このスクリプトウィンドウに表示され、他のエディタと同じように、コマンドを入力し、編集することができる。しかしながら、R コマンダーは、R に対する真のコンソールを提供しない。スクリプトウィンドウの限界としては次のようなものがある。

- 2 行以上にわたるコマンドでは、2 行目以下の行頭に空白またはタブが必要である。複数行のコマンドの全ては、同時に実行されなければならない。
- 代入の矢印 (<-) を含むコマンドは、通常の R Console では結果を表示するのだが、スクリプトウィンドウからでは結果を表示しない。[例えば、コマンド `print(x<-10)`] 一方、等号 (=) を用いた代入の場合、通常はそうではないのだが、結果を表示する（例えば、`x = 10`）。*14
- 通常、結果を表示しないコマンドが、出力ウィンドウに結果を表示することがある。このふるまいは、R コマンダーの etc ディレクトリにある `log-exceptions.txt` ファイルを編集することにより、変更できる。
- 大括弧 {} によって囲まれたコマンドの集合では、各コマンドの最後にセミコロン (;) をつけておかないと、適切に取り扱われない。これは R の悪いスタイルであり、スクリプトエディタは、プログラミング用エディタとしてはあまり有益ではないことを意味する。R の本格的なプログラミングには、R の Windows 版自体が提供するスクリプトエディタ、または、こちらの方がより望ましいが、プログラミング用のエディタを利用することを勧める。

参考文献

- Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of Statistical Software*, 8(15):1-27.
- Fox, J. (2005). The R Commander: A basic-statistics graphical user interface to R. *Journal of Statistical Software*, 19(9):1-42.
- Moore, D. S. (2000). *The Basic Practice of Statistics, Second Edition*. Freeman, New York.
- Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S, Fourth Edition*. Springer, New York.

*14 訳注：R2.6.2, Remdr.1.3-5 では確認できなかった。