

R コマンダー入門 *

John Fox and Milan Bouchet-Valat
Version 2.0-0 (最終改訂 : 2013 年 7 月 29 日)

日本語訳 †

2013 年 8 月 23 日

1 はじめに

R コマンダー (Fox, 2005) は、オープンソースの統計解析環境 **R** (R Core Team, 2013) に対してグラフィカル・ユーザー・インターフェース (“GUI”) を提供する。 **R** はコマンド駆動型システムで、 **R** の新しい利用者は、その学習を困難と感ずることが多い。これは特に、基本的な統計コースの学生といった統計手法の入門者に対してあてはまる。 **R コマンダー** は、ポイント・アンド・クリック方式のインターフェースを **R** に与えることにより、 **R** のコマンドを覚えたり生成したりといった作業から学生を解放し、統計手法に集中させることができる。さらに、生成されたコマンドをユーザーに可視化することにより、 **R コマンダー** は、少なくとも一部のユーザーに対して、 **R** のコマンドを書くという作業への移行を容易にする可能性を持つ。しかし、 **R コマンダー** では、 **R** の機能や the Comprehensive R Archive Network (CRAN) に提供されている文字通り数千のパッケージの機能の一部しか利用できない。 **R コマンダー** 自体は、プラグインパッケージにより機能を拡張することができ、これらは CRAN より利用可能である (6.4 節参照)。

本稿は、 **R** の **Windows** バージョンでの **R コマンダー (Rcmdr)** の利用法を説明する。 **Mac OS X** や **Linux**, **Unix** での **R コマンダー** の見かけと利用法には少し異なる点がある。これらの OS における **R コマンダー** のインストール方法については、 **R コマンダー** のウェブページ <<http://socserv.socsci.mcmaster.ca/jfox/Misc/Rcmdr/index.html>>, または、直接 <tinyurl.com/Rcmdr> にあるインストールノートへのリンクをたどることにより利用可能である。

本稿では、次の記法を利用する。 **Windows**, **R**, **Rcmdr** パッケージといったソフトウェア名に対しては太字を用いる。メニューやメニューの要素、ウィンドウ、ダイアログボックスといった GUI の要素については、イタリックとする。変数名やデータセット名、 **R** のコマンドはイタリックとする。*¹

2 R コマンダーの起動

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 コマンダー** 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 コマンダー** 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。*² パッケージをロードした後、 *R Console* と **R コマンダー** のウィンドウは図 1 と

* 本マニュアルの一部は、Fox(2005) の改訂版である。連絡は jfox@mcmaster.ca まで。

† 翻訳版における画像等は、R-3.0.1, Rcmdr.2.0-0 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 (訳注) 翻訳において、これらを必ずしも忠実に守っていない。

*² **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 1 つの主ウィンドウに含む。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

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 7** の元で作成している*3。 **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可能性がある。*4

R コマンダーと *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 コマンダー**のメニューとダイアログボックスを利用して、データの読み込み、処理、分析を行うことができるので、*R Console* を最小化しておいてもよい..

- **R コマンダー**の GUI で利用された **R** のコマンドは、**R コマンダー**の主ウィンドウの上部にあるスクリプトタブに表示される。スクリプトウィンドウに **R** のコマンドを直接入力してもよい。*5しかし、**R コマンダー**の開発目的は、コマンドを入力しないで済むようにすることにある。上部のウィンドウの 2 つめのタブ (**R マークダウン**という) は、**R コマンダー**で生成されたコマンドを蓄積し、レポートを表示するのに利用することができる。**R マークダウン**タブについては、6.1 節で説明する。
- 計算等の結果は、「出力」という 2 番目の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、あるいは、図 2 に示すスタートアップメッセージといった情報 (“メモ”) を表示する。
- グラフを作成すると、それらは別の *Graphics Device* ウィンドウに表示される。

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。出力や **R** ワークスペースの保存、終了の機能を持つ。

編集 様々なウィンドウとタブの内容を編集（切り取り、コピー、ペースト他）するためのメニュー。ウィンドウまたはタブ内でマウスの右クリックをすると、内容を編集するメニューが表示される。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、データ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 様々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 標準的な分布の累積確率、確率密度、分位点を求め、グラフを作成するサブメニュー（数値表の代わりに用いたり、これらの分布からのサンプルを得たりすることも可能）。

ツール **Rcmdr** とは関係のないパッケージをロードしたり（例えば、他のパッケージに付属するデータセット

ンドウとは独立して表示される。**R** を SDI モードで動かすにはいくつか方法がある。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。**R コマンダー**を MDI で利用してもよいが、それは **R** の主ウィンドウ内には表れず、スクリーンの配置が面倒になる。

*3 (訳注) 翻訳は、**Windows 8** の元で作成している。

*4 **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) <<http://cran.r-project.org/>> よりダウンロードできる。

もし、これらのパッケージがインストールされていないと、**R コマンダー**は、それらをインターネットやローカルファイル（例えば、CD-ROM）からインストールするかどうかを聞いてくる。**Windows** の “**R GUI**” から **Rcmdr** をインストールする場合、**Rcmdr** が依存するパッケージ全てがインストールされるとは限らない。関数 `install.packages` を用いて **Rcmdr** をインストールするときに、引数を `dependencies = TRUE` と設定しておくことにより、依存するパッケージ全てをインストールすることもできる。しかし、この方法は、再帰的な依存性のため、必要以上のパッケージをインストールする可能性がある。

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。**Mac OS X** 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 **tk** が、**X-Windows** を要求し、**Tcl/Tk** の **X-Windows** バージョンがインストールされることを要求するからである。これについては、**R コマンダー**のインストールノートを参照。

*5 *R Console* のプロンプト (`>`) の位置に直接キー入力してもよい。しかし、これらのコマンドで生成された結果は **R コマンダー**の出力ウィンドウには表示されず、エラーや警告メッセージは通常見えない

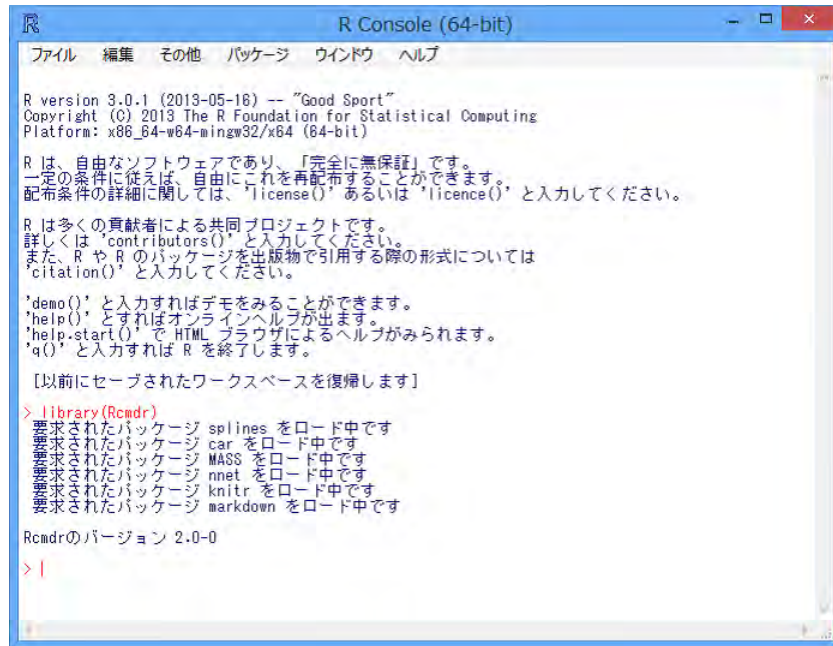


図 1 パッケージ **Rcmdr** ロード後の *R Console* ウィンドウ

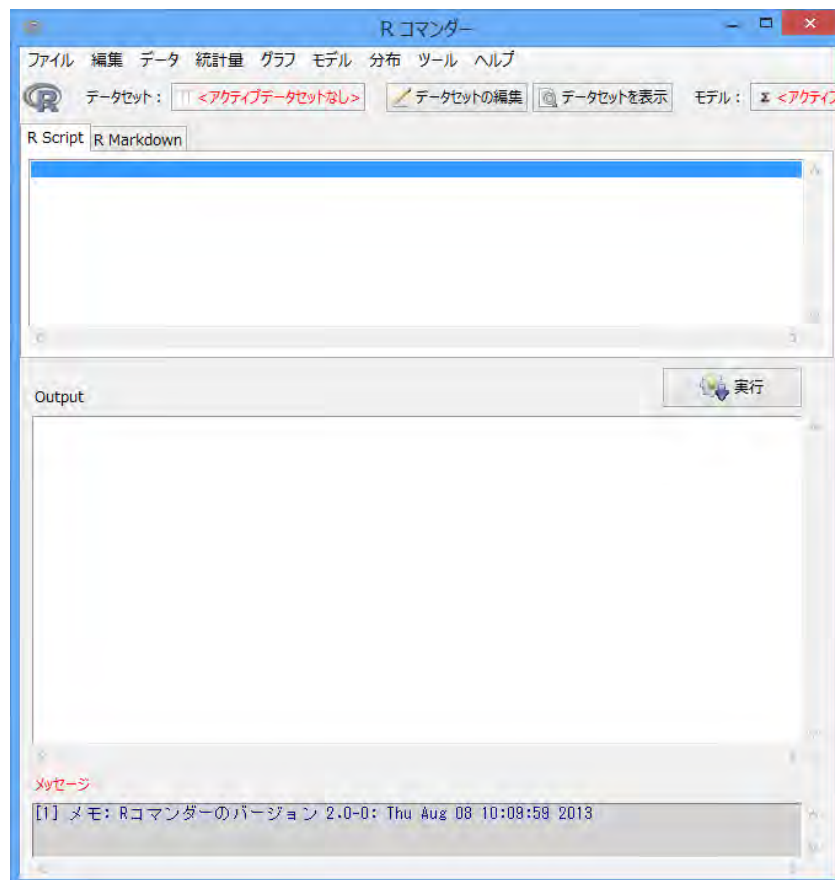


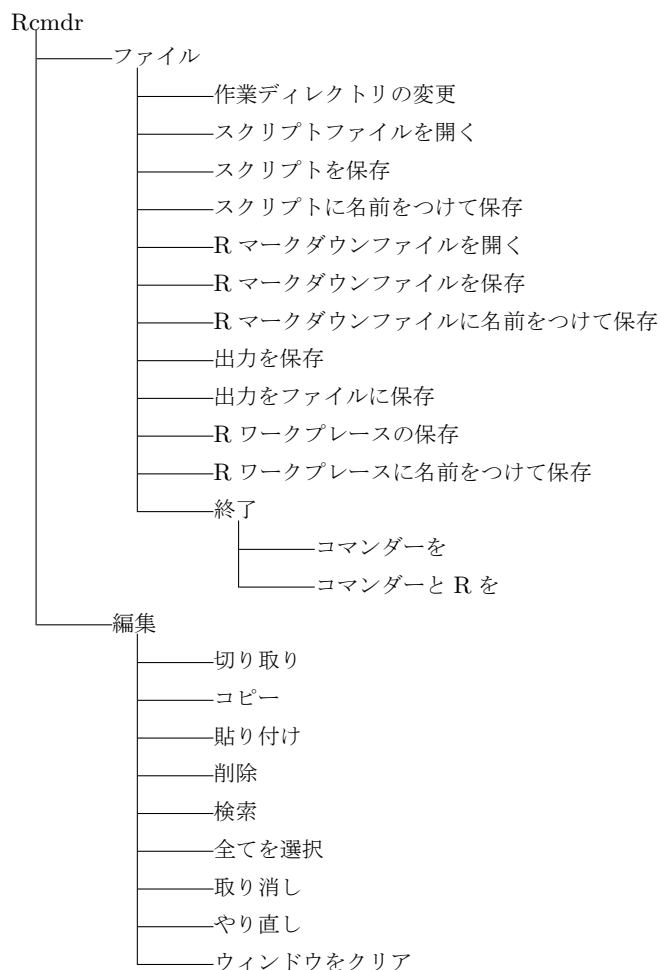
図 2 起動直後の **R コマンドー**のウィンドウ

にアクセスする), **Rcmdr** のプラグインパッケージを起動したり (Fox, 2007 および Fox and Carvalho, 2012, 6.4 節 参照), オプションを設定したり, オプションを保存したりするためのメニュー. こうしたオプションは, 後続のセッションで利用できる.

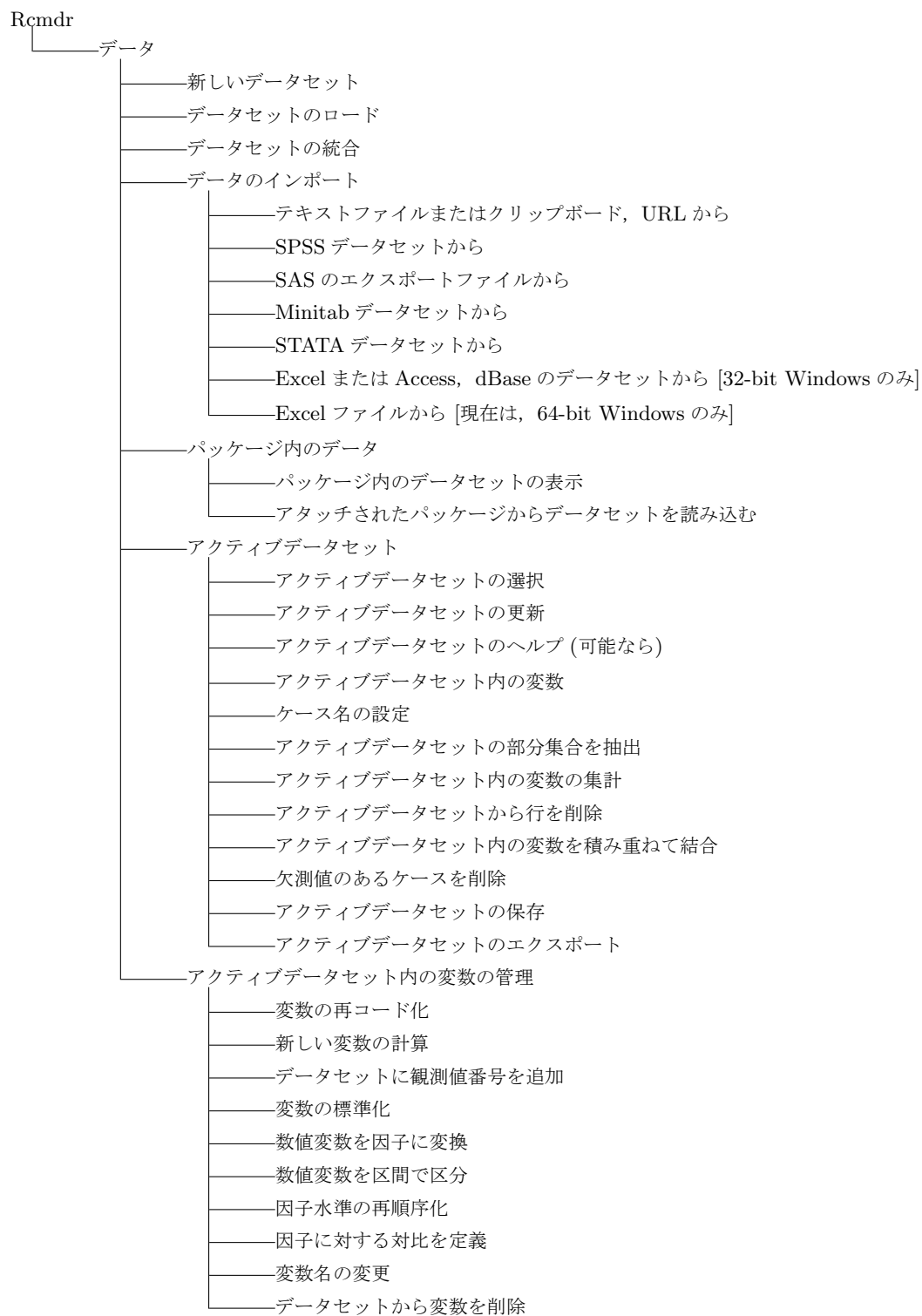
ヘルプ R コマンダーの情報 (このマニュアルを含む) や関連するソフトウェアの情報を得るためのメニュー.

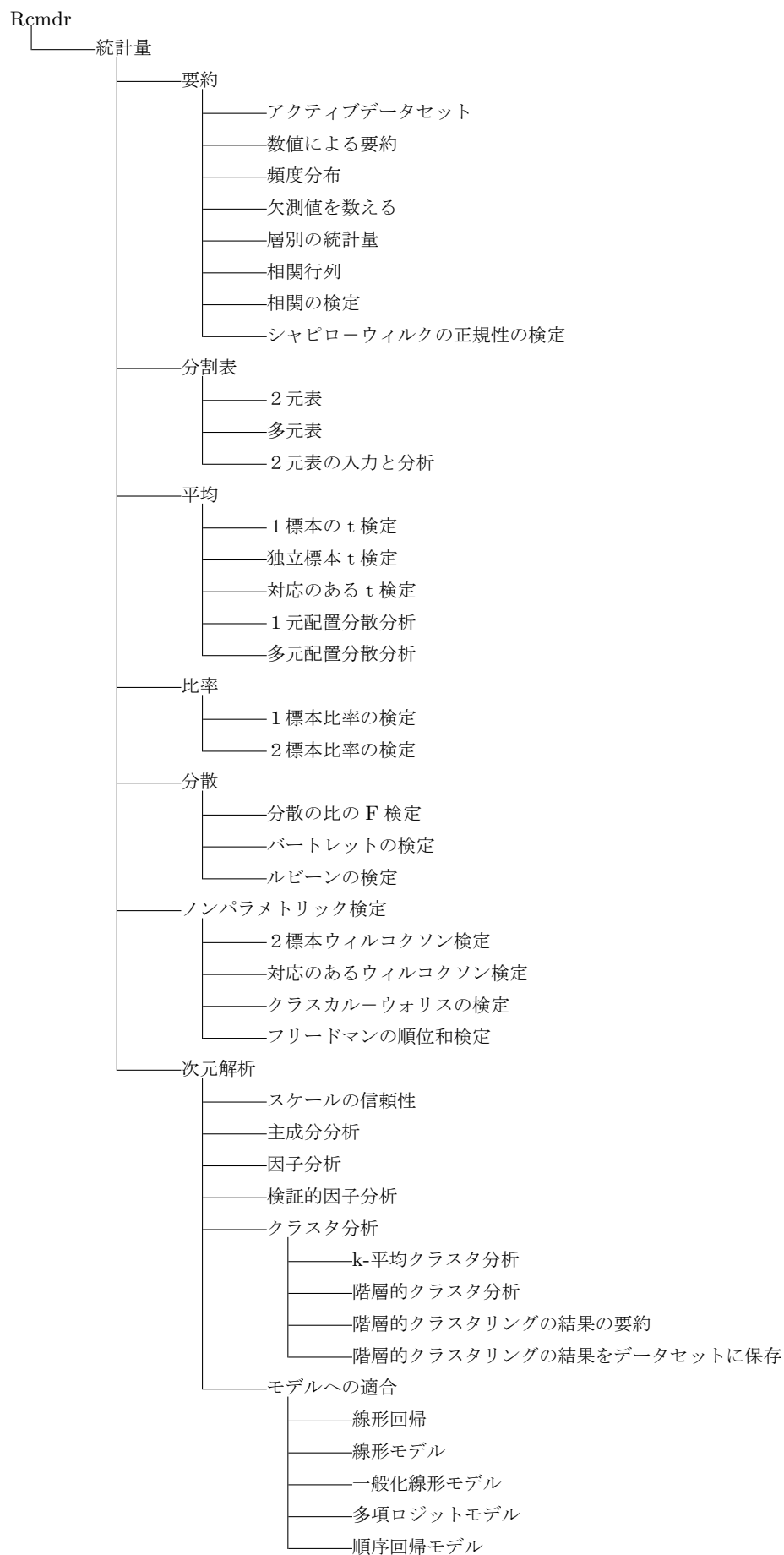
R コマンダーの各ダイアログボックスは**ヘルプ**ボタンを持つ (下記参照).

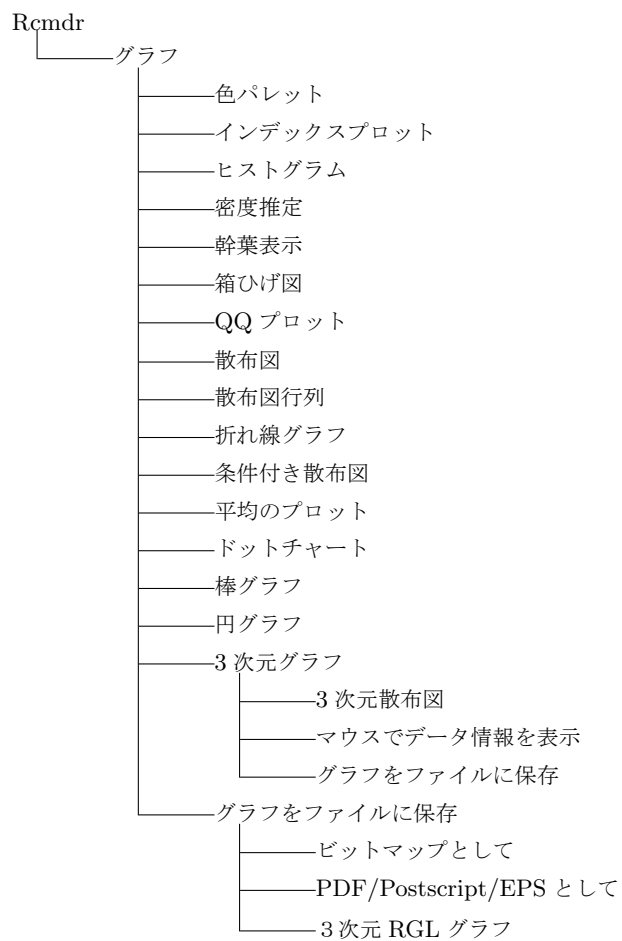
R コマンダー (バージョン 2.0-0) の完全なメニューツリーを以下に示す. ほとんどのメニュー項目では, 後で例示するように, ダイアログボックスが表示される. 表示した状況において利用できないメニューはグレイ表示され, 選択できないようになっている. 例えば, データセットが因子 (質的変数) を含まない場合, 分割表に対するメニュー項目は選択できない.*⁶

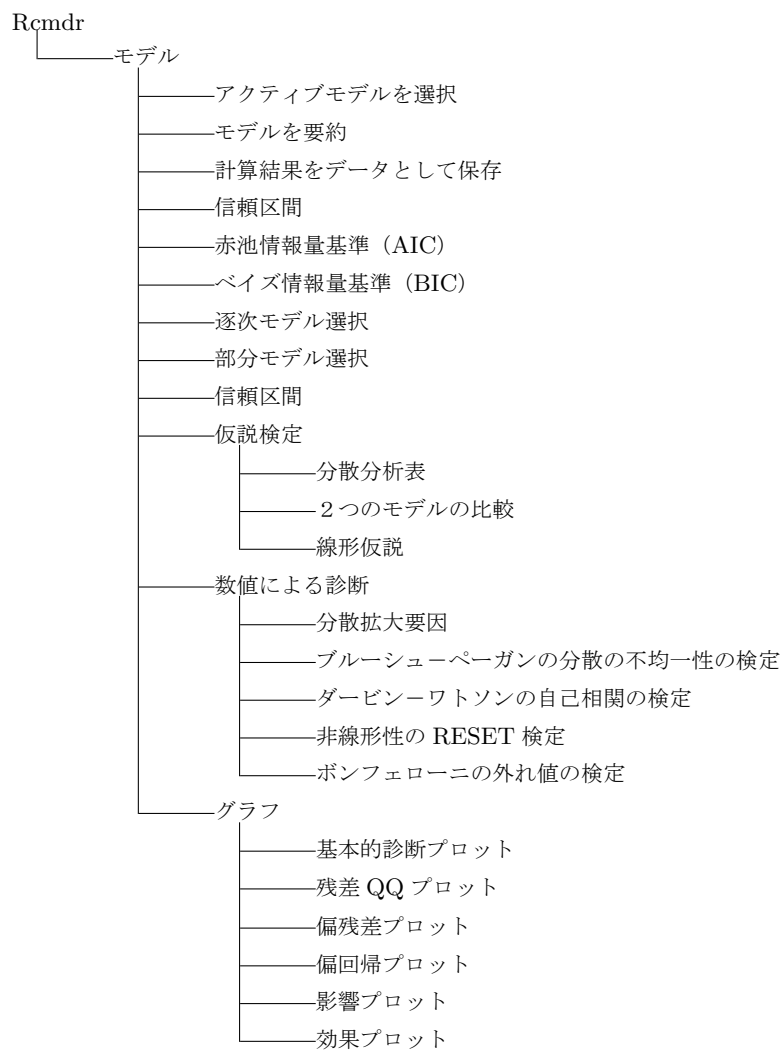


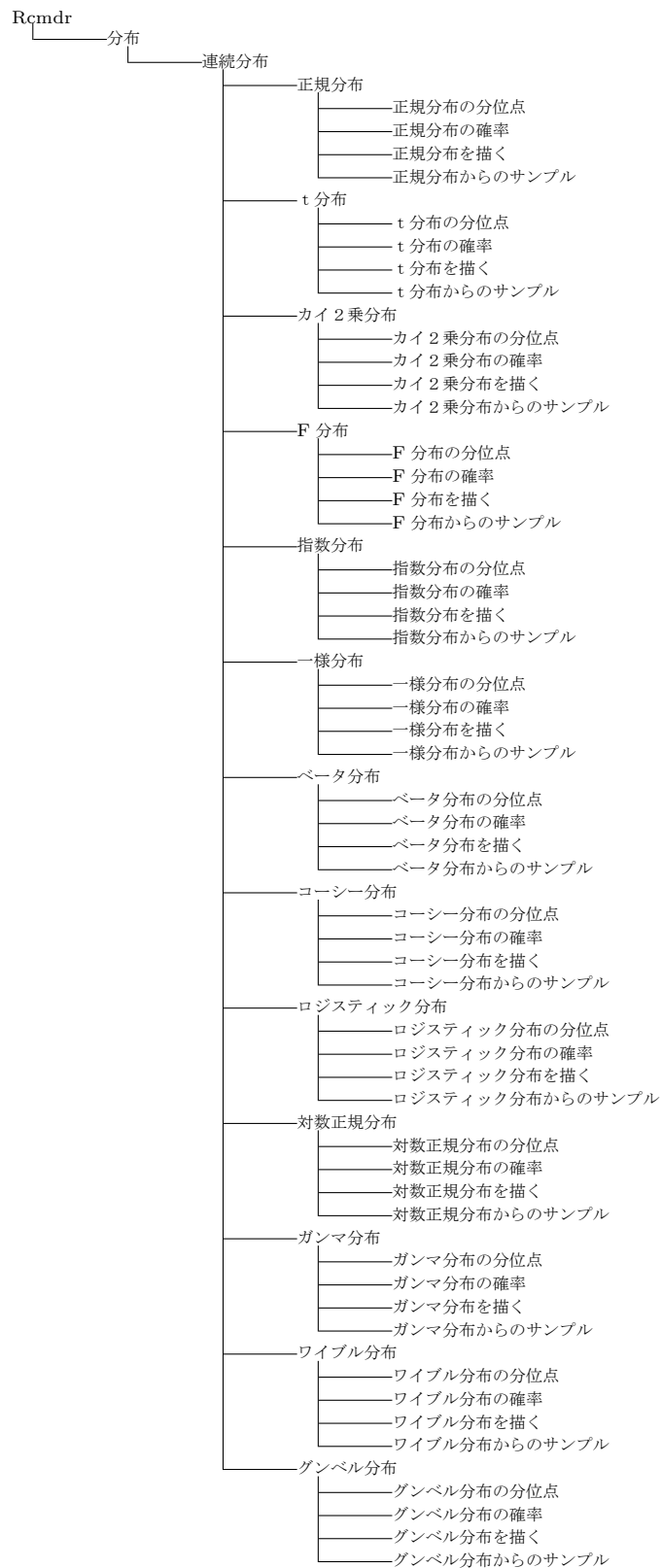
*⁶ ある状況においては, 表示されないメニュー項目がある. 例えば, ファイルメニューの **R マークダウン**メニュー項目は, **R マークダウン**タブがアクティブになっているときのみ, 表示される. ここで示すメニュー項目は, **Windows** で **R コマンダー**を用いているときにデフォルトで利用可能なもののみである. メニューはまた, `dividers` を含んでいるが, これは表示していない. また, ダイアログボックスを従えるメニューは, 慣習的に... をつけているが, これらも表示していない.

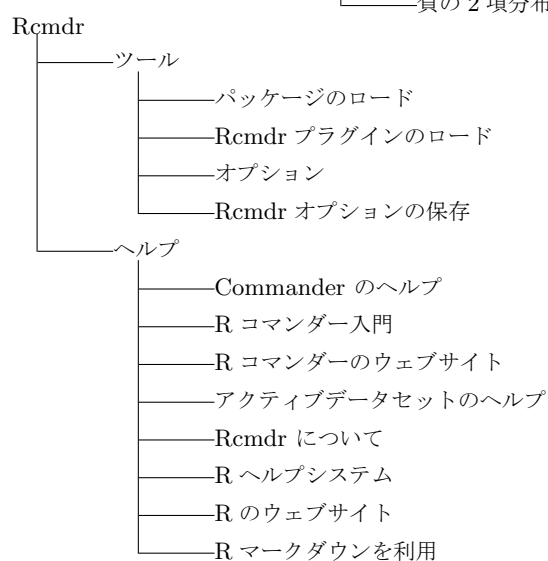
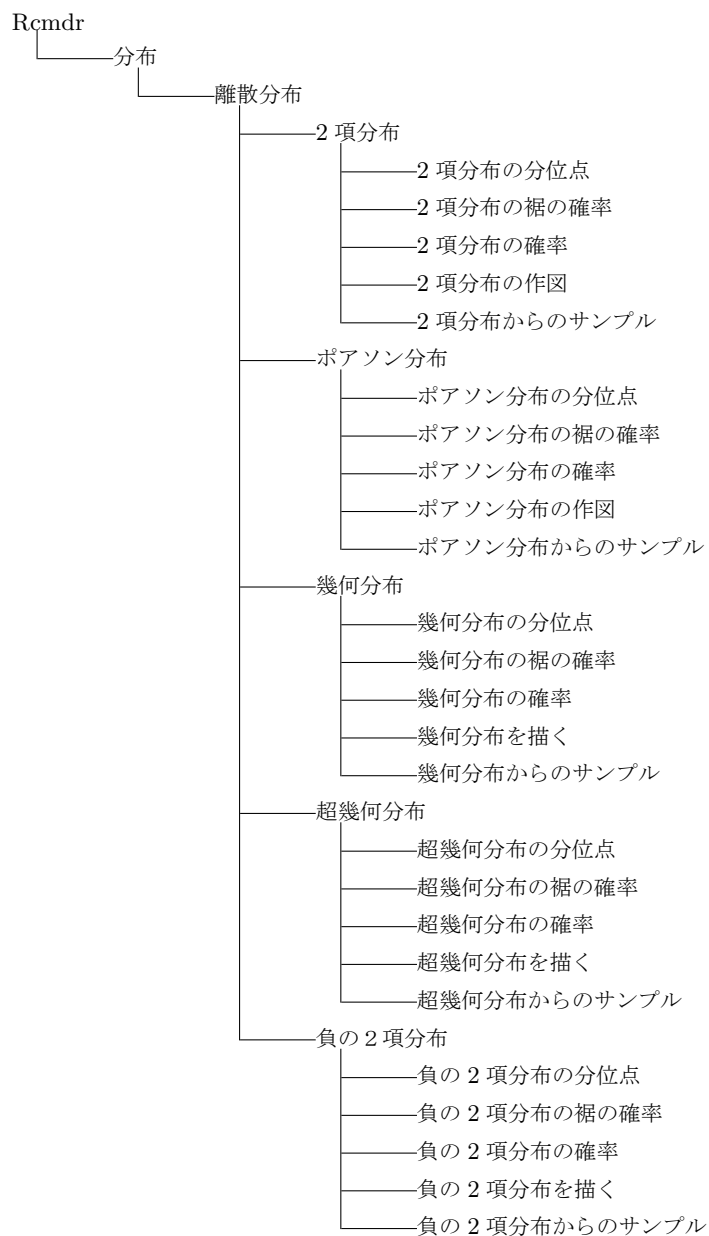












R コマンダーのインターフェースは、メニューとダイアログに加えて、他にいくつかの要素を持つ。

- メニューの下にボタンが並んだツールバーがある。
 - － 最も左にある（フラット）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。このボタンを押すことにより、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を対象とする（**ファイル**、**編集**、**分布メニュー**を除く）。
 - － 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しておくことができる。^{*7}
 - － 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、通常の線型モデル — の名前が表示される。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モデルを選択することができる。 **R コマンダー**はモデルと、当てはめたいデータとをシンクロナイズする。
- 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 **R スクリプトタブ**を含むウィンドウがある。既に述べたように、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。スクリプトタブの右下にある**実行**ボタンをクリックする（*Ctrl-r*^{*8}または *Ctrl-Tab* 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。複数行を選択した場合（それらの上で左クリックし、ドラッグすることにより）、**実行**ボタンによりそれら全てを実行することができる。 **R** スクリプトタブにコマンドを複数行に渡って入力してもよいが、そのときは、全ての行を同時に実行する必要がある。キーの組合せ入力 *Ctrl-a* により、スクリプトタブ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。
- **R** スクリプトと **R マークダウンタブ**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（標準の *R Console* と同じ）。
- 一番下には、メッセージを表示するための小さなグレーの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。

既に述べたように、パッケージ **Rcmdr** を起動した後、*R Console* を最小化してもよい。 **R コマンダー**のウィンドウの大きさは、通常の方法で変更したり最小化したりできる。 **R コマンダー**の大きさを調整すると、以降の出力は出力ウィンドウの大きさに自動的に調整される。

R コマンダ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** のオプションで設定する。^{*9} 詳細については、**Rcmdr** のヘルプメニューより、**R コマンダー**のヘルプを参照のこと。

3 データの入力

R コマンダーのほとんどの手続きは、アクティブデータセット^{*10}があるという前提で実行される。メモリに複数のデータセットがある場合、1つのみを選択することができる。 **R コマンダー**を起動した直後には、アクティブデータセットはない。

^{*7} David Firth のパッケージ **relimp** にある関数 **showData** によるデータビューアは、多くの変数を持つデータセットを表示するとき遅くなることがある。変数の数の閾値（初期設定は 100）を越える場合、データセットを表示するかわりにすこし不細工であるが、**R** の **View** コマンドが用いられる。変数の数に関係なく **R** の **View** コマンドを利用するには、閾値を 0 に設定する。詳細については、**R コマンダー**のヘルプを参照のこと。

^{*8} Ctrl (Control) キーを押したまま、同時にキー **r** を押す。

^{*9} 省略記号 (...) 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 GUI の標準的な仕様である。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。

^{*10} 分布メニューで選択された手続き、および、**統計量** → **分割表**メニューにおける分割表の入力および分析の手続きに関しては、この限りではない。

R コマンダーによるデータの入力方法にはいくつかある。^{*11}

- **Mac OS X** 以外のプラットフォームでは、**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。データセット数が非常に少ないときは、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、URL によるインターネット、他の統計パッケージ（**Minitab** や **SPSS**, **Stata**）から、あるいは（**Windows**）では、**Excel** または **Access**, **dBase** のデータをインポートすることができる。
- **R** のパッケージに含まれるデータセットを読み込むことができる。その名前を知っている場合はそれをキー入力するか、ダイアログボックスから選択する。

3.1 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** ^{*12}を取り上げる。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**（出生児 1000 人当たりの乳児死亡率）、**GDP**（1 人当たり国民総生産、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、各行にはデータ値が 6 つあり、最初の値はアルファベットである。このような場合、**R** の **read.table** コマンド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** には地域名が入力されている。これが読み込まれると **R** は **region** を因子、つまり質的変数として取り扱う。たいていの場合、**R コマンダー**は、数値変数と因子とを区別し、因子の平均を計算するといった間違いを行うことを防ぐ。

データファイル **Nations.txt** を **R** に読み込むには、**R コマンダー**のメニューから、**データ** → **データのインポート** → **テキストファイルまたはクリップボード、URL から...** を選択する。この操作により、図 3 に

^{*11} これらのデータソースは、全てのプラットフォームで利用可能というわけではない。

^{*12} 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。

示すテキストファイルまたはクリップボード、URL 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。データセット名を Nations に変更していることに注意。

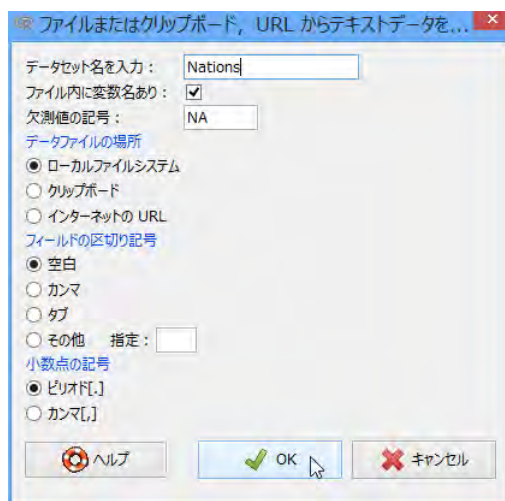


図 3 テキストファイルからのデータの読み込み

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、全てのアルファベット、アンダースコア（_）、数字（0 – 9）で構成される必要がある。空白を用いることはできない。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、異なるデータセット名となる。

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 ファイルを探し、選択した状況を示している。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。

次に、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、R スクリプトタブ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メモが表示される。R コマンダーはまた、`relimp` パッケージをロードするために `library` コマンドを実行するので、データセットが表示される。このように、一般に、パッケージは必要なときに自動的にロードされる。

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、R のデータフレームである。

3.2 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。^{*13} 例として、Moore (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

^{*13} R データエディタの制約により、Mac OS X のもとで直接データセットを入力することはできないので、対応するメニュー項目を以下表示しない。データをテキストファイルとして保存し（ワープロでデータファイルを作成し、それをプレーンテキストファイルまたはアスキーファイルとして保存しようとしている場合は注意）、前節で説明したデータ → データのインポート → テキストファイルまたはクリップボード、URL から... により読み込むのが最も簡単である。

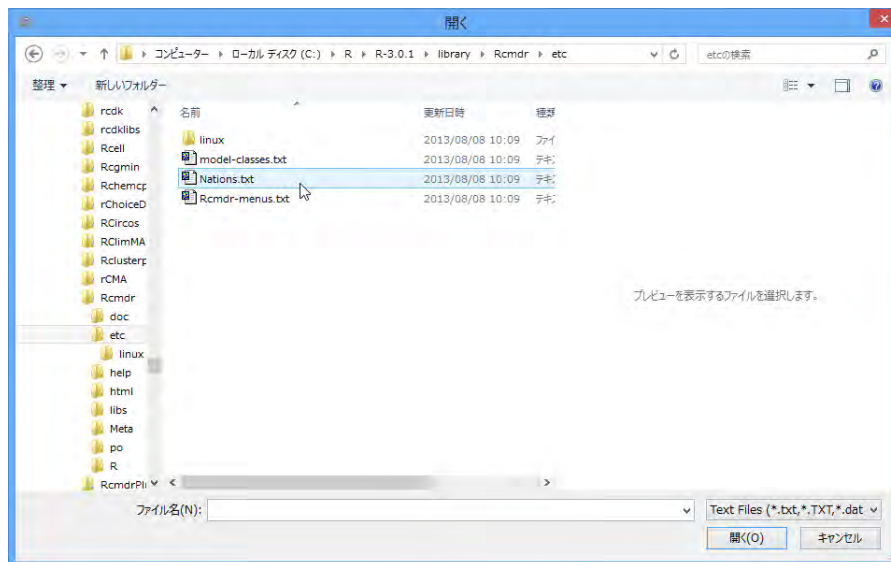


図 4 テキスト形式データファイルを読み込むためのファイルを開くダイアログボックス

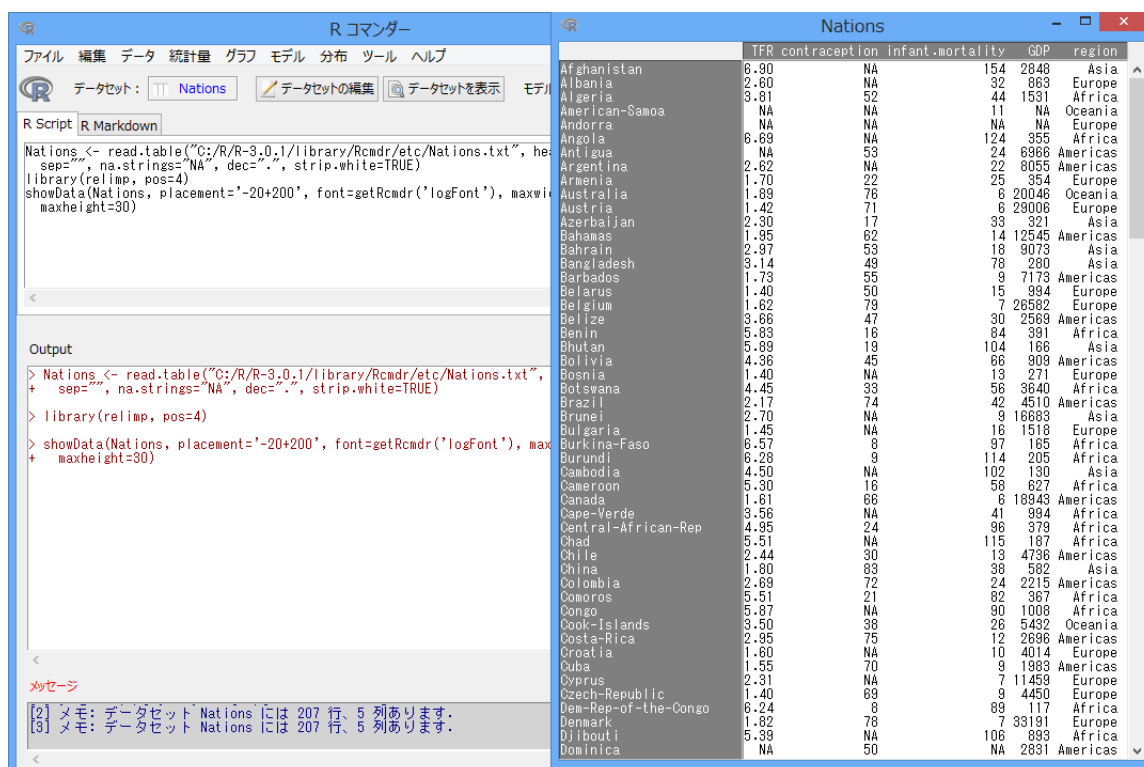


図 5 アクティブデータセットの表示

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	38	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						
19						

図 6 データ入力後のデータエディタ

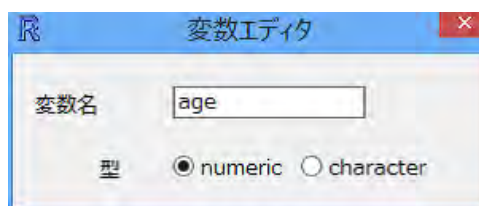


図 7 データエディタの変数名の変更のためのダイアログボックス

3.3 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含んでいる。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リストとして表示でき、**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。^{*14} 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名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、パッケージ名をダブルクリックすると右のリストボックスにデータセット名のリストが表示される。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**データセットを入力**欄にコピーされる。^{*15} データセットのスクロールボックス内の記号キーをクリックすると、その記号で始まる次のデータセットに移動する。**R** の他のパッケージにアクセスするには、**ツール** → **パッケージのロード** によって行う。

^{*14} 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。

^{*15} **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。

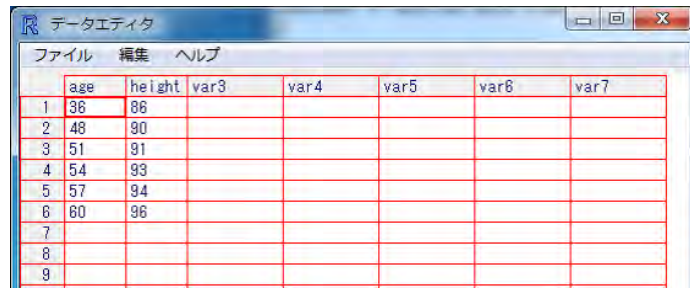


図 8 2 つの変数名を変更した後のデータエディタウィンドウ

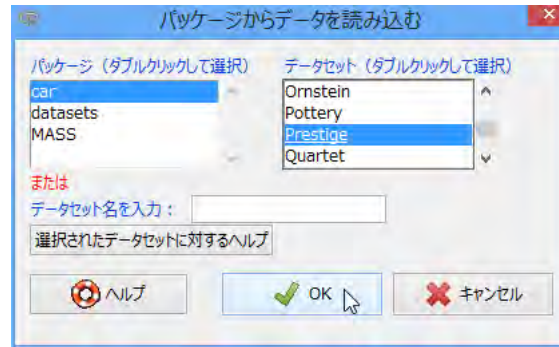


図 9 アタッチされたパッケージからのデータセットの読み込み—car パッケージの Prestige データセットの例.

4 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、**R コマンダー**のメニューにより数値による要約やグラフの作成を実行できる。ここでは、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見ると、必要に応じて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ことを期待する。

下記の最初の例では、前節でテキストファイルから読み込んだ **Nations** をアクティブデータセットとしている。前節で説明したように、Moore (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** データセットを読み込んだりしているときは、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**のウィンドウの左上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、因子の各水準のデータ数が表示される。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うことにより、不要な大量の出力を避けることができるように設定されている。このメニュー項目は特殊である。というのは、**R コマンダー**の典型的なメニューならダイアログボックスを表示するのであるが、これは直接、**R** のコマンドを呼び出すからである。

例えば、**統計量** → **要約** → **数値による要約**を選択すると、図 11 に示すダイアログボックスが表示される。このダイアログの変数のリストには、数値変数のみが表示される。因子 **region** が表示されないのは、因子に対して平均や標準偏差といった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する。^{*16} **数値による要約**のダイアログボックスには、**データ**と**統計量**という 2 つのタブがある。図

^{*16} 変数リストボックスで 1 つの変数のみを選択するには、その名前を左クリックするだけでよい。2 つ以上の変数を選択したい場合は、

12 に示すように、**統計量**タブを左クリックして選択する。今の場合、デフォルトで選択されている項目を取り上げる。**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。

```
> numSummary(Nations[, "infant.mortality"], statistics=c("mean", "sd", "IQR",
+ "quantiles"), quantiles=c(0,.25,.5,.75,1))
      mean      sd IQR 0% 25% 50% 75% 100%   n NA
43.47761 38.75604  54   2  12  30  66  169 201   6
```

デフォルトでは、平均と標準偏差 (sd)、四分位範囲 (IQR)、および、最小値 (0 %)、第 1 四分位数、メディアン、第 3 四分位数、最大値 (100 %) 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 は欠測値の数である。

R コマンダーのダイアログは通常そうであるが、図 11 に示す**数値による要約**ダイアログボックスのように、左からヘルプ、リセット、**OK**、キャンセル、適用というボタンを持つ。^{*17} ヘルプボタンにより、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 (ブラウザーに表示) を参照することができる。リセットボタンは、ほとんどの **R コマンダー**のダイアログにあるが、ダイアログをそのオリジナルな状態に戻す。さもないと、以前の選択からの起動状態を保持する。アクティブデータセットが切り替えられたとき、ダイアログの状態もリセットされる。例示したように、**OK** ボタンは、ダイアログを閉じ、**R コマンド**を呼び出す。適用ボタンもまたコマンドを実行するが、現状のままダイアログを再開する。これは、いくつかの似たオペレーションを実行するときに便利である。ダイアログボックスで間違いを犯すと、例えば、数値による要約のダイアログで変数を選択せずに **OK** をクリックすると、エラーメッセージが表示され、ダイアログが再び開かれる。

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別の要約情報、つまり層別の要約情報を求めることもできる。**数値による要約のデータタブで層別して要約...** 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、あらかじめ選択されている。これを選択し、**OK** ボタンをクリックすると、層別して要約... ボタンの表示が層別変数: region へと変化する (図 14)。今の場合、要約する 2 つの数値変数 GDP と infant.mortality を選択した。**OK** をクリックすると、次の結果が出力パネルに表示される。

```
> numSummary(Nations[,c("GDP", "infant.mortality")], groups=Nations$region,
+ statistics=c("mean", "sd", "IQR", "quantiles"), quantiles=c(0,.25,.5,.75,1))
```

Variable: GDP

	mean	sd	IQR	0%	25%	50%	75%	100%	n	NA
Africa	1196.000	2089.614	795.50	36	209.00	389.5	1004.50	11854	54	1
Americas	5398.000	6083.311	5268.50	386	1749.25	2765.5	7017.75	26037	40	1
Asia	4505.051	6277.738	6062.50	122	345.00	1079.0	6407.50	22898	39	2
Europe	13698.909	13165.412	24582.25	271	1643.75	9222.5	26226.00	42416	44	1
Oceania	8732.600	11328.708	16409.25	654	1102.75	2348.5	17512.00	41718	20	5

Variable: infant.mortality

	mean	sd	IQR	0%	25%	50%	75%	100%	n	NA
Africa	85.27273	35.188095	50.0	7	61.00	85.0	111.00	169	55	0
Americas	25.60000	17.439713	24.0	6	12.00	21.5	36.00	82	40	1

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** キーを押して左クリックすると、選択を拡大することができる。**Ctrl** キーを押したまま左クリックすると、追加選択することができる。

^{*17} これらのボタンの順序はオペレーティングシステムによって異なる。

Asia	45.65854	32.980001	50.0	5	22.00	37.0	72.00	154	41	0
Europe	11.85366	7.122363	10.0	5	6.00	8.0	16.00	32	41	4
Oceania	27.79167	29.622229	26.5	2	9.25	20.0	35.75	135	24	1

R コマンダーの他のダイアログにも、同様にして層別変数を選択することができるものがある。

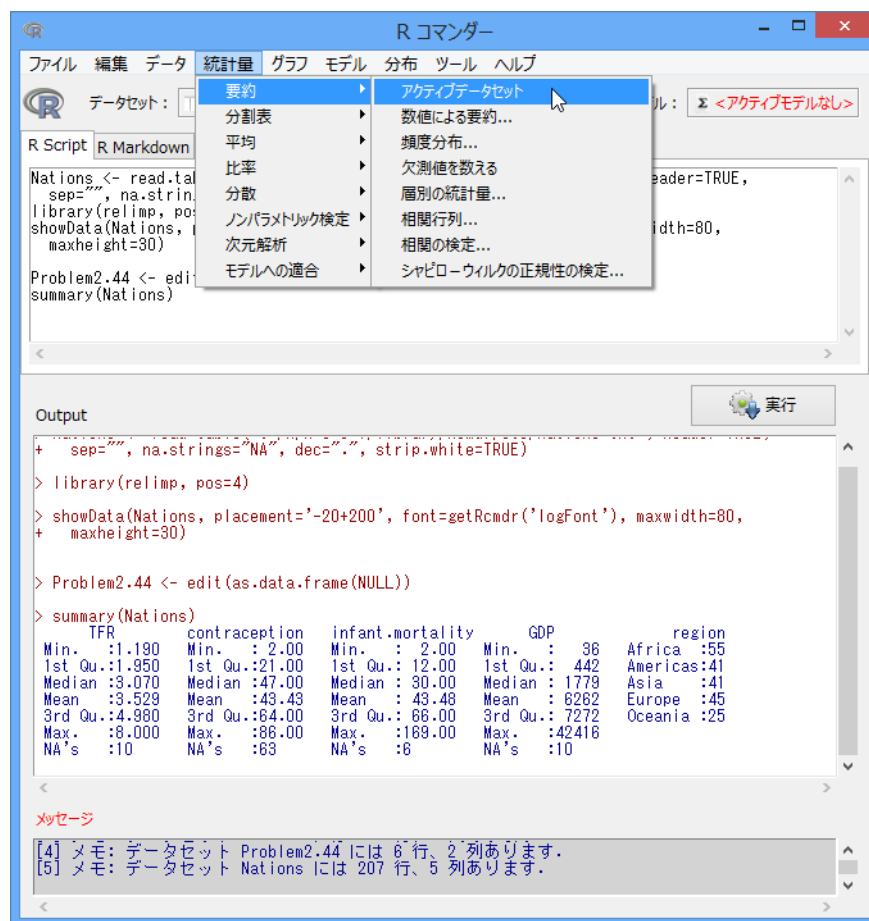


図 10 アクティブデータセットに対する変数の要約情報の取得

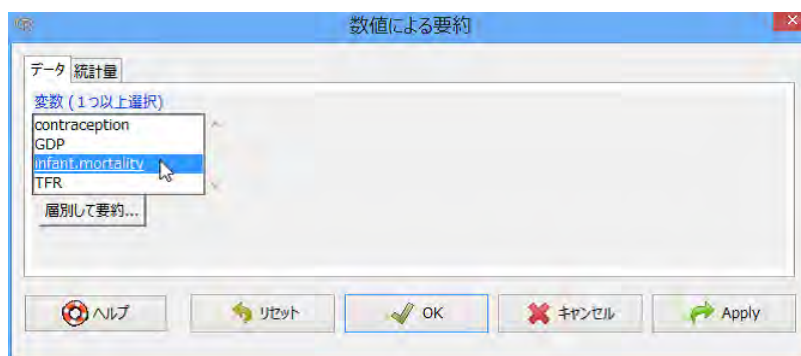


図 11 数値による要約のダイアログボックス

R コマンダーでグラフを作成することも簡単である。例えば、R コマンダーのメニューより **グラフ → ヒストグラム...** を選択すると、図 15 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。このダイアログには、**データタブ** と **オプションタブ** がある。デフォルトのオプションを全て選択し、**infant.mortality** を選択して **OK** をクリックすると、図 16 に示すヒストグラムが表示されたグラフィックスウィンドウが表れる。

1 つのセッションで複数の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** には通常、最新のものの



図 12 数値による要約のダイアログボックスにおける統計量タブ



図 13 質的変数ダイアログボックスで層別変数を選択する

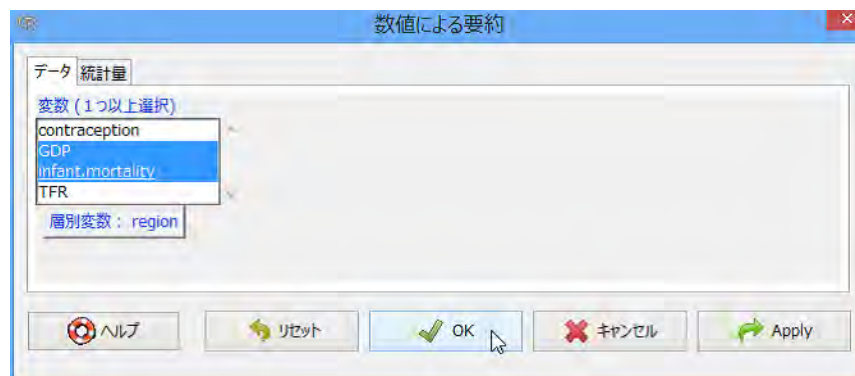


図 14 層別変数として region を、2 つの数値変数を選択した後の数値による要約のダイアログボックス

みが表示される。^{*18}

5 統計モデル

統計量 → モデルへの適合メニューを用いて、R コマンダー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... メニューにより），一般化線形モデル，多項ロジットモデル，順序回帰モデル（比例オッズモデル等）である。[最後の 2 つは，それぞれ Venables and Ripley (2002) に

^{*18} Windows では，Windows の R グラフィックスデバイスのグラフ履歴メカニズムをヒストリ → 記録によりオンにしておくと，キーボードの Page Up または Page Down キーにより，前のグラフを呼び出すことができる。この特徴は，Windows システムにおいてのみ利用可能である。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 ; Fox and Hong, 2009) 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。

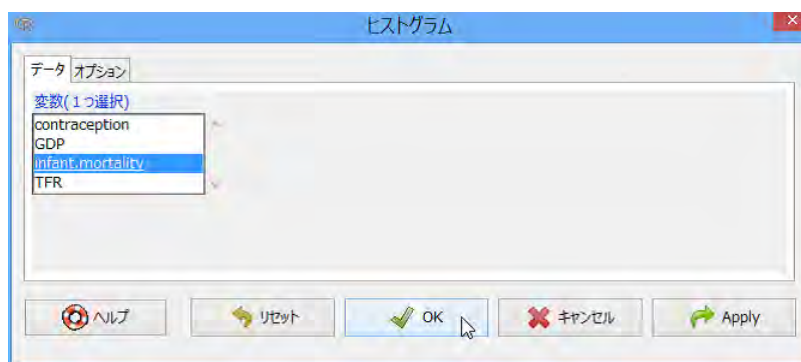


図 15 ヒストグラムのダイアログボックス

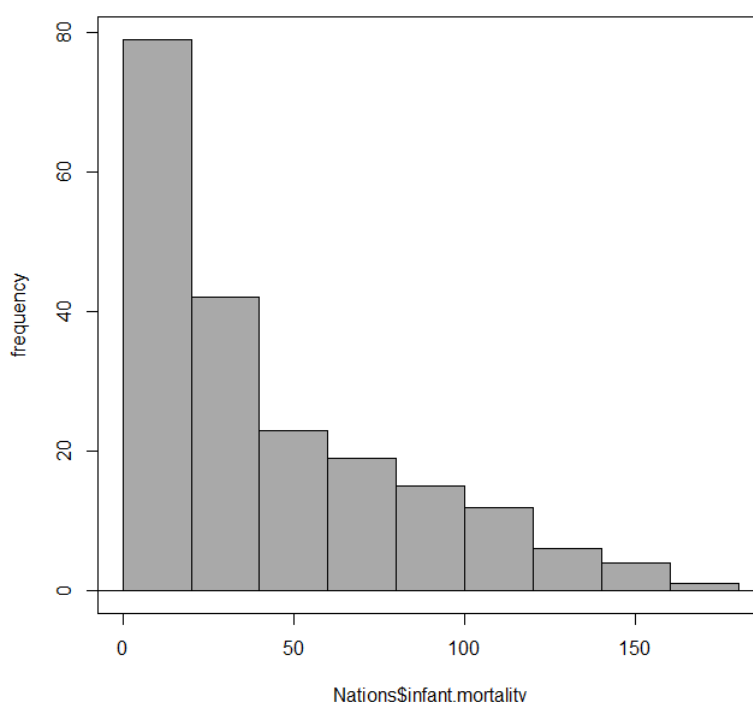


図 16 Nations データセットの `infant.mortality` のヒストグラムを表示するグラフィックスウィンドウ

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少し異なるが (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 17 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.*¹⁹統計量 → モデルへの適合 → 線型モデル... を選択する前に, **Prestige** をアクティブデータセットにする. これには, アクティブデータセットボタンをクリックし, リストから **Prestige** を選択する. なお, **Prestige** データは, 3.3 節で **car** パッケージから読み込んだ.

- 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 —— 式の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 $+$ が前に入力される). 変数リストの中で, 因子については, 変数名の後ろに【因子】と記されている. モデル式の右辺に因子が入力されると, ダミー変数に変換される.

*¹⁹ 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. 以下の説明では, **R** のモデル式についてある程度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* マニュアルを参照. これには, *R Console* のヘルプ → マニュアル (PDF) からアクセス可能.

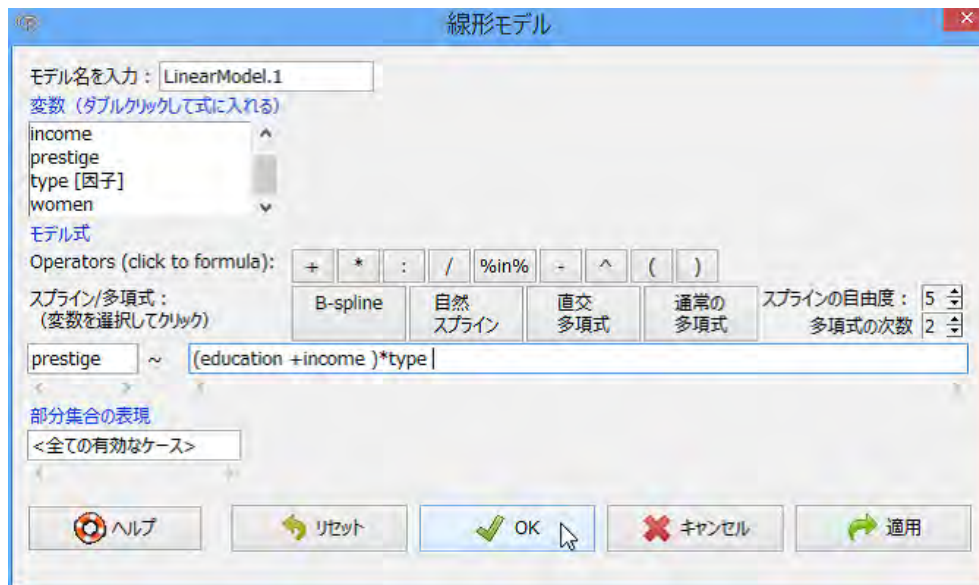


図 17 線形モデルダイアログボックス

- モデル式の上に並んでいるツールバーのボタンは、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。
- ツールバーの最下行のボタンは、モデル式に回帰スプラインや多項式項、およびスピンボックスにより、スプラインの自由度および多項式の次数を入力するのに便利である。
- モデル式のフィールドに直接キー入力してもよい。例えば、`log(income)` といった項を入力するには、直接入力する必要がある。
- 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、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` として送られ、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、各データに対して `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 (`prof` ではない) を指定する) といったものである。

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。また、`LinearModel.1` がアクティブモデルとなり、その名前が**モデル**ボタンに表示される。

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

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

Call:

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

Residuals:

Min	1Q	Median	3Q	Max
-13.970	-4.124	1.206	3.829	18.059

Coefficients:

Estimate	Std. Error	t value	Pr(> t)
----------	------------	---------	----------

```

(Intercept)          -120.0459    20.1576   -5.955 5.07e-08 ***
education             2.3357      0.9277    2.518 0.01360 *
log(income)          15.9825      2.6059    6.133 2.32e-08 ***
type[T.prof]         85.1601     31.1810    2.731 0.00761 **
type[T.wc]           30.2412     37.9788    0.796 0.42800
education:type[T.prof] 0.6974      1.2895    0.541 0.58998
education:type[T.wc]   3.6400      1.7589    2.069 0.04140 *
log(income):type[T.prof] -9.4288     3.7751   -2.498 0.01434 *
log(income):type[T.wc] -8.1556      4.4029   -1.852 0.06730 .

```

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

Residual standard error: 6.409 on 89 degrees of freedom

(4 observations deleted due to missingness)

Multiple R-squared: 0.871, Adjusted R-squared: 0.8595

F-statistic: 75.15 on 8 and 89 DF, p-value: < 2.2e-16

アクティブモデルに対する操作は、**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、デフォルトの“Type II”を選択すると、次の結果が表示される。

```
> Anova(LinearModel.1, type="II")
```

Anova Table (Type II tests)

Response: prestige

	Sum Sq	Df	F value	Pr(>F)
education	1209.3	1	29.4446	4.912e-07 ***
log(income)	1690.8	1	41.1670	6.589e-09 ***
type	469.1	2	5.7103	0.004642 **
education:type	178.8	2	2.1762	0.119474
log(income):type	290.3	2	3.5344	0.033338 *
Residuals	3655.4	89		

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

6 その他

6.1 レポートの生成

デフォルトの設定では、**R コマンド**は上部のウィンドウに **R マークダウン**タブを持つ。これは、セッションの間に生成されたコマンドを **R マークダウン**ドキュメントに蓄積する。図 18 に、現在のセッションのトップにスクロールした **R マークダウン**タブを示す。その名前から分かるように、**R マークダウン**は、単純なマークアップ言語で、**R** のコマンドのブロックを含み、その名前から分かるように、HTML（つまり、ウェブ）ページを生成するのに利用できる。**R マークダウン**に関する情報については、**ヘルプ** → **R マークダウンの利用**を参

照.*20

R コマンドにより生成されたコマンドの各集合は、**R マークダウン**ドキュメントに **R** のコマンドのブロックを構成する.*21 こうしたブロックは、`““{r}`で始まり、`““`で終わる。**R マークダウン**タブは編集可能で、タブ内を修正したりテキストを追加したりすることができる。一般に、**R** コードのブロックの間、あるいは、あなた自身のコードブロックに説明的なテキストを入力する方が安全である（下記参照）。しかし、エラーとなったり意図しない出力を生み出すブロック全体を削除してもよい（以降のブロックがそのブロックに依存していないなら）。たとえ **R** のコマンドをよく知らなくても、出力ウィンドウにおけるコマンドと出力結果との関係を理解することができる。最初のコードブロック（`““{r echo=FALSE}`）で始まる）は、**R マークダウン**テキストを処理するために利用される **knitr** パッケージ（Xie, 2013）からのソフトウェアに対するオプションを設定するので、修正しない方がよい。

行を少し省略して（... で示している）、現在のセッションで生成された **R マークダウン**ドキュメントを次に示す。

```
<!-- R Commander Markdown Template -->
```

```
Replace with Main Title
```

```
=====
```

```
### Your Name
```

```
### `r as.character(Sys.Date())`
```

```
““{r echo=FALSE}
```

```
# include this code chunk as-is to set options
```

```
opts_chunk$set(comment=NA, prompt=TRUE, out.width=750, fig.height=8, fig.width=8)
```

```
library(Rcmdr)
```

```
““
```

```
““{r}
```

```
Nations <- read.table("C:/R/R-3.0.1patched/library/Rcmdr/etc/Nations.txt",
```

```
header=TRUE, sep="", na.strings="NA", dec=".", strip.white=TRUE)
```

```
““
```

```
. . .
```

```
““{r}
```

```
data(Prestige, package="car")
```

```
““
```

```
. . .
```

```
Let us regress occupational prestige on the education and income levels of the occupations,
```

*20 ヘルプメニューのこの機能および他のものに関して、インターネットへの接続が必要なものがある。

*21 ユーザーとのインタラクションを直接必要とするコマンド、例えば、グラフにおける点のインタラクティブな識別は、**R マークダウン**ドキュメントでは機能しない。例外は、データエディタを起動するコマンドである。なぜなら、以降のコマンドはこれに依存しているはずだからである。

```
transforming income to linearize its relationship to prestige:
```

```
““{r}
LinearModel.1 <- lm(prestige ~ (education + log(income))*type,
data=Prestige)
summary(LinearModel.1)
““

““

Anova(LinearModel.1, type="II")
““
```

“Your Name”（あなたの名前）をあなたの名前で置き換えたり，“Replace with Main Title”（主タイトルで置き換える）を作成するレポートのタイトルで置き換えるといったことは説明する必要はないだろう。おそらく、あまり明確ではないのは、**R** のコードブロックの間、つまり、ブロックの終わりを示す “”と次のブロックの始まりを示す ““{r}”の間に任意の説明文を入力してもよいことである。そうすることにより、**R マークダウン**により提供される簡潔なマークアップの利点を利用することができる；例えば、アスタリスク*で囲まれたテキストは（例えば、*this is important*）、イタリックで表示される。回帰を行うブロックの直前に、テキスト “Let us regress occupational prestige ...”（prestige を回帰させる...）を追加している。

R マークダウンドキュメントの編集が終了すると、**R マークダウン**タブの下にある **HTML レポートの生成** ボタンを押すことにより、HTML のページを作ることができる。レポートは、ウェブブラウザで開くことができる。**R マークダウン**ドキュメントは、**ファイルメニュー**で保存することができる。

6.2 出力の保存と印刷

R コマンドの**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。**R マークダウン**タブを利用するのが好きでなければ、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。しかしこの方法は、直接的には再現可能ではないという欠点を持つ。しかし **R マークダウン**ドキュメントなら、分析を再現するとともに修正も可能である。

Word や **OpenOffice**, **Writer** といったワープロを起動する。**Windows WordPad** なんてものもよい。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* や、マウスの右クリックで貼り付け）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。でないときれいに整列しない。

同様に、グラフをコピーするには、**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* や、マウスの右クリックで貼り付け）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、表示されたコンテキストメニューより**メタファイル**としてを選択してもよい。^{*22} **R** のセッションの最後に、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、これは注釈付きの記録となる。

^{*22} 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。しかし、ここで説明した方法は簡単であり、一般に高品位のグラフとなる。再度述べるが、ここでの説明は **Windows** のみにあてはまる。

R コマンダーの下で、テキストやグラフを保存するための別の方法として、それぞれ、**R コマンダー**のファイルや、グラフ → グラフをファイルで保存を利用することもできる。**R コマンダー**のスクリプトタブをファイル → スクリプトの保存により保存しておく、将来のセッションにおいて作業を再現する助けとなる。

6.3 スクリプトタブにコマンドを入力する

スクリプトタブは、コマンドを編集・入力・実行するための機能を簡単に提供する。**R コマンダー**が生成したコマンドは、スクリプトタブに表示され、エディタと同じように、コマンドを入力したり編集したりすることができる。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”ではなく、限界がある。例えば、複数行にわたるコマンドは、全て同時に実行されなければならない。**R** で本気でプログラミングするときは、**R** の **Windows** や **Mac OS X** バージョンが提供するスクリプトエディタを利用するか、より望ましいのは、プログラミングエディタやインタラクティブな開発環境、例えば、**RStudio** <www.rstudio.org>を用いる方がよい。^{*23}

6.4 R コマンダーのプラグインを利用する

R コマンダーのプラグインは、**R コマンダー**に機能を追加する **R** のパッケージである。現在、多くのプラグインが CRAN で利用可能であり、通常の方法で、ダウンロードおよびインストールすることができる。プラグインは、**R コマンダー**にメニューやメニュー項目、関連するダイアログボックスを追加する。既存のメニュー項目やダイアログを変更したり削除したりすることもある。適切にプログラムされた **R コマンダー**プラグインは直接 **R コマンダー**の起動と同時に、あるいは、**R コマンダー**のツール → *Rcmdr* プラグインのロードより起動することができる。後者の場合、**R コマンダー**はプラグインをアクティベートするために再起動する。複数のプラグインを同時に利用することも可能であるが、互いにコンフリクトを起こすことがあるので注意。例えば、あるプラグインが削除しようとするメニューに対してメニュー項目を追加しようとするプラグインもある。

6.5 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。終了してもよいかという確認の後、**R** スクリプトと出力、**R** マークダウンウィンドウの内容をを保存したいかどうかを聞かれる。なお、*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を聞かれるが、通常は **No** を選択するとよい。

参考文献

- [1] Firth, D. (2011). relimp: Relative Contribution of Effects in a Regression Model. R package version 1.0-3.
- [2] Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of Statistical Software*, 8(15):1-27.
- [3] Fox, J. (2005). The R Commander: A basic-statistics graphical user interface to R. *Journal of Statistical Software*, 19(9):1-42.
- [4] Fox, J. (2007). Extending the Rcmdr by “plug-in” Packages. *R News*, 7(3):46-52.
- [5] Fox, J. and Sa Carvalho, M. (2012). The RcmdrPlugin.survival package: Extending the R Commander to survival analysis. *Journal of Statistical Software*, 49(7):1-32.

^{*23} **R コマンダー**は、**RStudio** の元で機能する。しかし、デフォルトでは、**R コマンダー**の出力とメッセージは **RStudio** の中で **R** コンソールに送られるが、不安定性や **RStudio** のグラフィックスデバイスとの不適合性といった問題がある。

- [6] Moore, D. S. (2000). *The Basic Practice of Statistics*. Freeman, New York, second edition.
- [7] Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer, New York, fourth edition.
- [8] Xie, Y. (2013). knitr: A general-purpose package for dynamic report generation in R. R package version 1.2.