

R コマンダー : Rcmdr (1)

同志社大学文化情報学部教授

金 明哲 (Jin Mingzhe)

■中国生まれ。総合研究大学院大学数物研究科統計科学専攻博士後期課程修了。博士(学術)。1995年札幌学院大学社会情報学部、助教授、教授を経て、2005年4月より現職。E-mail: mjin@mail.doshisha.ac.jp



1. R コマンダーとは

データ解析の商業ソフト (SAS、SPSS、S-PLUSなど) は、使用しやすくするため GUI (Graphical User Interface) に多くの工夫が施されている。R は商業ソフトに比べ GUI が貧弱で、コマンドライン (CUI: Character User Interface) を用いるのが普通である。コマンドラインによるデータ解析は、GUI に慣れた方にとっては馴染みやすいものではない。R コマンダー (R Commander) は、R の基本的な統計関数を使いやすくするための GUI パッケージ (Rcmdr) である。パッケージ Rcmdr は、McMaster University の John Fox が作成・公開し、関西大学の荒木孝治氏により日本語化された。Rcmdr は CRAN ミラーサイト、あるいは John Fox、荒木孝治氏の HP からダウンロードすることができる。

<http://socserv.socsci.mcmaster.ca/jfox/Misc/Rcmdr/>

<http://www.ec.kansai-u.ac.jp/user/arakit/R.html>

2. R コマンダーの起動と終了

CRAN ミラーサイトからインストールを行うと、必要とするパッケージが自動的にインストールされる。

(1) R コマンダーの起動

パッケージ Rcmdr をインストールし、コマンド library (Rcmdr) を実行すると R コマンダーが起動し、図 1 のようなウィンドウが開かれる。

図 1 R コマンダーの起動画面



R コマンダーのウィンドウは、メニュー、データセットの操作に関するアイコン、スクリプトウィンドウ、結果の出力ウィンドウ、メッセージのウィンドウより構成されている。

R コマンダーは、メニューとダイアログボックス (dialog box、以後略してDBoxと呼ぶことにする) の操作でデータ解析・処理を行う。

R コマンダーで必要とするパッケージがインストールされていない場合は、R コマンダーを起動すると、メッセージウィンドウにメッセージが返される。そのときには、必要とするパッケージをインストールし、R コマンダーを再起動する作業が必要である。

(2) R コマンダーの終了

R コマンダーの終了には2つの方法がある。

① R コマンダーのみを終了

メニュー「ファイル」⇒「終了」⇒「R コマンダー」を実行すると確認DBoxが返される。終了したいときには [OK] ボタンを、そうではない場合は [キャンセル] ボタンを押す。1 回 R コマンダーを終了し、再起動する場合には、次のコマンドを実行する。

>Commander()

② R コマンダーと R を同時に終了

「ファイル」⇒「終了」⇒「R コマンダーと R」を実行する。

3. R コマンダーの基本機能

R コマンダーは、その機能を8項目に分けてメニューを設けている。メニューごとの項目リストを表1に示す。より詳細な日本語によるメニューツリーは、荒木氏のHPに公開されている。

http://www.ec.kansai-u.ac.jp/user/arakit/Rcmdr_menu_tree.txt

表1 R コマンダーの基本機能

	機能の項目
ファイル	スクリプトファイルを開く・保存、出力結果の保存、ワークスペースの保存、R コマンダーの終了
編集	ウィンドウをクリア、切り取り、コピー、貼り付け、削除、見つける、すべてを選択
データ	新しいデータセットの作成、データのインポート、パッケージ内のデータセット、アクティブデータセット内の変数管理
統計量	要約、分割表、平均、比率、分散、ノンパラメトリック検定、次元解析、モデルへの適合
グラフ	インデックスプロット、ヒストグラム、幹葉図、箱ひげ図、QQプロット、散布図、散布図行列、折れ線グラフ、平均のプロット、棒グラフ、円グラフ、3Dグラフ、グラフをファイルで保存
モデル	アクティブモデルの選択、モデルの要約、計算結果の保存、信頼区間、仮説検定、数値による診断、モデル関連のグラフ
分布	正規分布、t分布、カイ2乗分布、F分布、2項分布、ポアソン分布
ツール	パッケージロード、オプション

誌面上の制約により、すべての機能について詳しく説明することは不可能である。本稿では、使用例を示しながら主な機能と使用法を簡潔に説明する。

4. データの操作

R コマンダーでは、新しいデータセットの作成、データのインポート、パッケージ内のデータセットの読み込み、アクティブデータセット内の変数管理などを行うことが可能である。

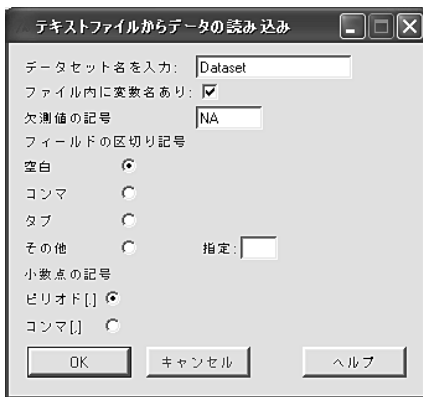
(1) 新しいデータセットの作成

新しいデータセットの作成は、サイズが比較的に小さいデータセットの作成に適している。R コマンダーのメニュー「データ」⇒「新しいデータセット」をクリックするとデータセットの名前を指定するDBoxが開かれる。新しいデータセットのDBoxでデータファイルの名前を指定し、[OK]ボタンを押すとデータ作成シートが開かれる。ここではRの関数editを用いている。

(2) データのインポート

データのインポートでは、テキストファイル、SPSS、Minitab、STATAデータ形式を読み込むことができる。サブメニュー「テキストファイル」をクリックすると図2のようなDBoxが開かれる。DBoxの必要な項目を図2のように指定し、[OK]ボタンを押すとファイルを開く画面が返される。この機能は、関数read.tableを用いている。

図2 テキストファイルを読み込む画面



(3) パッケージ内のデータセット

「パッケージ内のデータセット」には、パッケージ内のデータセットに関する操作を行うサブメニュー「パッケージ内のデータセットの表示」と「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」がある。

「パッケージ内のデータセットの表示」をクリックするとパッケージcar、datasetの中のデータセットのリストをR上に返す。

「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」をクリックすると図3に示すDBoxが開かれる。DBoxからパッケージおよびデータファイルを指定して[OK]ボタンを押すとアタッチされたデータファイルが読み込まれる。これらの操作を行うと、Rコマンダー上の

図3 パッケージからデータを読み込む画面

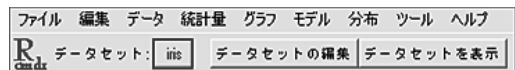


データセットアイコンにデータセットの名前が表示される。[データセットの編集]ボタンを押すと関数fixによるデータシートの画面が返され、[データセットを表示]ボタンを押すとデータセットを表示する画面が返される。

(4) データセットのアクティブ化

上記の操作で読み込んだデータセットが1つの場合は、自動的にそのデータセットがアクティブ化され、処理対象になる。データirisがアクティブ化されている画面を図4に示す。

図4 irisがアクティブ化されている画面



複数のデータセットを読み込んだ場合は、解析対象とするデータセットをアクティブ化させなければならない。データセットのアクティブ化の操作は、メニューの下部のデータセットの名前が表示されているアイコンをクリックして指定する方法と、メニューの「データ」⇒「アクティブデータセット」⇒「アクティブデータセットの選択」をクリックして指定する2つの方法がある。

メニュー「データ」⇒「アクティブデータセット」のリストを表2に示す。

表2 「データ」メニューと対応する関数

メニュー	対応する関数
アクティブデータセットのヘルプ	help()
アクティブデータセット内の変数	names()
ケース名の設定	row.names(),as.character()
アクティブデータセットの部分集合を抽出	subset()
欠測値のあるケースを削除	na.omit()
アクティブデータセットのエクスポート	write.table()

5. 統計量

「統計量」のメニュー画面のコピーを図5に示す。

統計量のメニューには、サブメニューがある。そのリストと関数の対応関係を表3に示す。

図5 メニュー「統計量」の画面

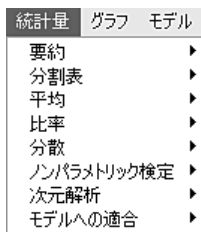


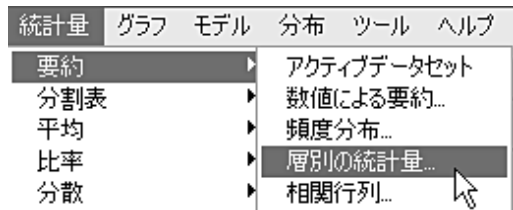
表3 「統計量」メニューと対応する関数

メニュー	対応する関数
要約	
アクティブデータセット	summary(iris)
数値による要約	mean(),sd(),quantile()
頻度分布	table(),chisq.test()
層別の統計量	tapply()
相関行列	cor()
分割表	
2元表	xtabs(),chisq.test()
多元表	xtabs()
2元表の入力と分析	chisq.test()
平均	
1 標本t検定	t.test()
独立サンプルのt検定	t.test()
対応のあるt検定	t.test()
1元配置分散分析	anova(lm())
多元配置分散分析	Anova(lm())
比率	
1 標本比率の検定	prop.test()
2 標本比率の検定	prop.test()
分散	
分散比のF検定	var.test()
バートレット検定	bartlett.test()
ルビーン検定	levene.test()
マンハタメトリック検定	
2 標本ウィルコクソン検定	wilcox.test()
対応のあるウィルコクソン検定	wilcox.test()
クラスカル-ウォリスの検定	kruskal.test()
次元解析	
スケールの信頼性	reliability(cov())
主成分分析	princomp()
因子分析	factanal()
k-平均クラスター分析	KMeans(),biplot(princomp())
階層的クラスター分析	hclust(dist())
モデルへの適合	
線形回帰	lm()
線形モデル	lm()
一般線形モデル	glm()
多項ロジットモデル	multinom()
比例オッズロジットモデル	polr()

(1) 要約

「要約」メニューの画面を図6に示す。

図6 メニュー「要約」の画面



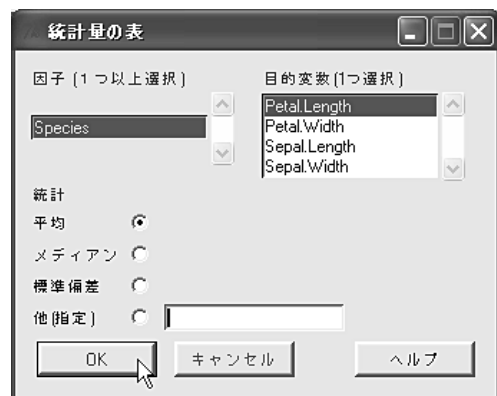
例として「層別の統計量」の使用手順とその結果を示す。

◇ 解析対象とするデータセットをアクティブ化する。

◇ 図6のようにメニューの「統計量」⇒「要約」⇒「層別の統計量」をクリックし、DBoxを開く。

◇ 「統計量の表」のDBoxの設定を行う。層別の要約を行うので、まず層別因子を決める。ここではirisの品種ごとの“Petal. Length”の平均を求めることにする。その設定の画面を図7に示す。DBoxを設定し、[OK]ボタンを押すとirisの品種ごとのPetal. Lengthの平均値がRコマンドーの出力ウィンドウに返される。

図7 「統計量の表」のDBox画面



DBox上の[ヘルプ]ボタンを押すと、用いる関数のマニュアル画面がR上で開かれる。より詳細な引数を指定したいときには、各DBoxに用いる関数を知っておくことがよいであろう。

(2) 分割表

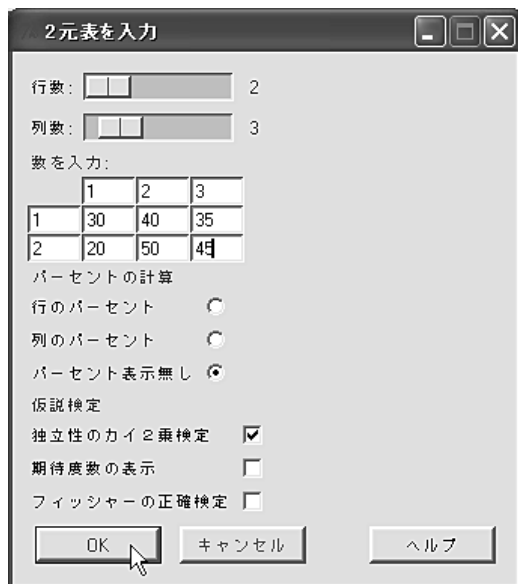
分割表では、分割表の作成（関数table）や分割表の独立性検定の統計量（関数chisq.test、fisher.test）などを求めることができる。分割表の使用例として、分割表のデータを入力してカイ2乗検定を行う手順を示す。

◇ 「統計量」⇒「分割表」⇒「2元表の入」と分析」をクリックし、2元表を入力するDBoxを開く。

◇ 2元分割表のサイズの設定は、DBox上の「行数」「列数」の平行移動ボタンを左右に移動して行う。分割表の数値は直接入力する。その画面を図8に示す。

◇ [OK]ボタンを押すと、Rコマンドーの出力ウィンドウに実行結果が返される。

図8 2元表を入力するDBox



(3) 平均・比率・分散・ノンパラメトリック検定
メニュー「統計量」の中の「平均」、「比率」、「分散」、「ノンパラメトリック検定」ではそれぞれの仮説検定を行う（表3を参照）。例として、「平均」のメニュー画面を図9に示す。使用例としてデータirisを用いた多元配置分散分析のDBoxの画面を図10に示す。

図9 「平均」のメニュー画面

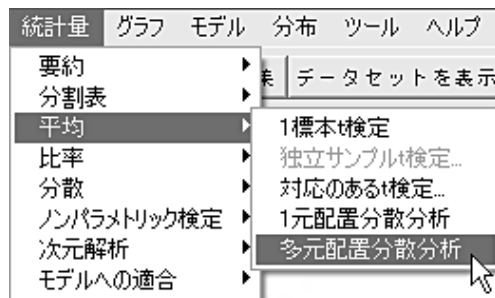


図10 多元配置分散分析DBox



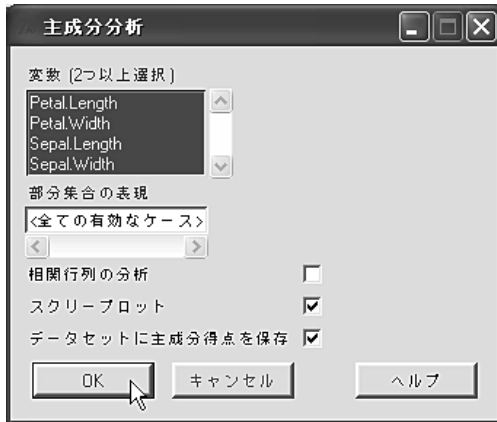
(4) 次元解析

次元解析では、因子分析、主成分分析、クラスター分析などの解析を行う。

例えば、「統計量」⇒「次元解析」⇒「主成分分析」をクリックすると図11のような主成分分析を指定するDBoxが開かれる。用いる変数および関連項目を選択し、[OK]ボタンを押すと主成分の結果はRコマンドーの出力ウィンドウに、寄与率の棒グラフはRに返される。

DBoxにおける「相関行列の分析」項目の右にチェック記号を入れると相関行列を用いた主

図11 主成分分析のDBox画面



成分分析、チェック記号を入れない場合は分散共分散行列を用いた主成分分析を行う。

主成分の散布図を作成するためにはコマンドの操作が必要である。コマンドをスクリプトウィンドウに記述し、Rコマンドの右側の[実行]ボタンを押すとグラフがR画面に返される。このような作業はR画面とRコマンドを入れ替えることになるので若干不便である。Rコンソール上で、Rコマンド上の操作結果のオブジェクトを用いた方が便利であろう。

クラスター分析では、k-平均クラスター分析と階層的クラスター分析を行うことが可能である。k-平均クラスター分析では、R画面の主成分のパイプロットにk-平均クラスター分析の結果を返す。

(5) モデルへの適合

「モデルへの適合」では、線形回帰および線形モデル、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の解析を行うことが可能である。線形回帰と線形モデルの違いは、変数間の交互作用をモデルに入れるか否かである。后者では変数間の交互作用を考慮す

るモデルを作成することができる。

例として、2004年10月号で一般線形モデルを説明する際に用いたデータairqualityを用いて例を示す。

まずデータを読み込み、アクティブ化する。データairqualityはパッケージdatasetの中にあるのでメニューの「データ」⇒「パッケージ内のデータセット」⇒「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」をクリックする。しかし、開かれたDBoxの中のdatasetをクリックしても該当のファイルがリストに現れない場合がある。その際には、DBoxにおける「データセットの名前を入力」の右の窓に直接データセットの名前を入力して、[OK]ボタンを押す。

メニュー「統計量」⇒「モデルへの適合」⇒「一般化線形モデル」をクリックすると「一般化線形モデル」のDBoxが開かれる。データairqualityを用いた一般化線形モデルを作成するDBox画面を図12に示す。作成するモデルの式とリンク関数を指定し、[OK]ボタンを押すと結果が出力ウィンドウに返される。

図12 一般化線形モデルのDBox画面

