

R コマンド : Rcmdr (2)

同志社大学文化情報学部教授

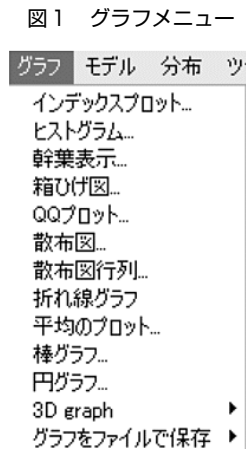
金 明哲 (Jin Mingzhe)

■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. グラフ

R コマンドのグラフメニューを図1に示し、グラフ作成に用いた関数との対応関係を表1に示す。



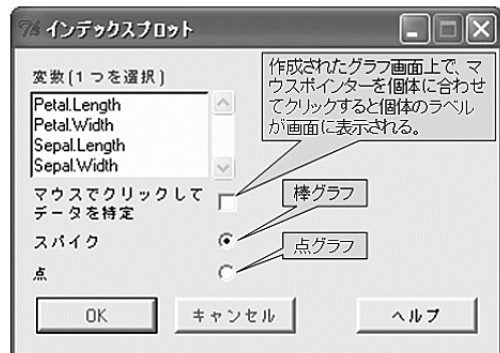
「インデックスプロット」は、1変数の値を縦軸、個体を横軸に並べた棒グラフ、あるいは点グラフを作成する。インデックスグラフのDBox (ダイアログボックス) を図2に示す。DBoxの項目を指定し、[OK]ボタンを押すとグラフがRのウィンドウに返される。

「ヒストグラム」では、パッケージRcmdrの中の関数Histを用いて、引数scale = c("frequency", "percent", "density")の中の1種類のヒストグラムを作成する。

表1 「グラフ」と関数の対応表

項目の名称	対応する関数	パッケージ
インデックスプロット	plot()	graphics
ヒストグラム	Hist()	Rcmdr
幹葉表示	stem.leaf()	Rcmdr
箱ひげ図	boxplot()	graphics
QQプロット	qq.plot()	car
散布図	scatterplot()	car
散布図行列	scatterplot.matrix()	car
折れ線グラフ	matplot()	graphics
平均のプロット	plotMeans()	Rcmdr
棒グラフ	barplot()	graphics
円グラフ	pie()	graphics
3D graph	Scatter3DDialog()	Rcmdr

図2 インデックスグラフのDBox



「幹葉表示」では、パッケージRcmdrの中の関数**stem.leaf**を用いて、Rcmdrの出力ウィンドウに幹葉表示を返す。

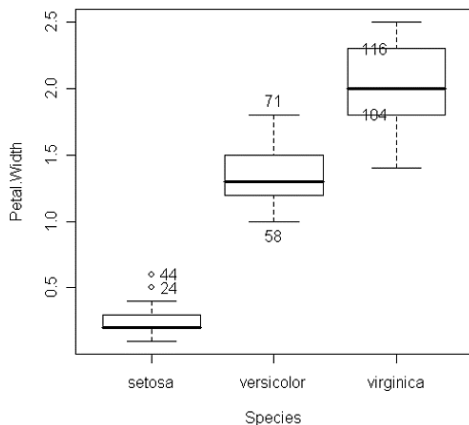
「箱ひげ図」は、パッケージgraphicsの中の関数**boxplot**を用いて作図する。箱ひげ図を作成するには、箱ひげ図における外れ値を含む個体と、箱ひげ図における位置を同定するか否かを指定することができる。同定はマウスのポインターを同定すべき位置に合わせてクリックする。同定操作の終了は、Rのメニューアイコン[STOP]をクリックする。図3にデータirisの箱ひげ図のDBox画面を示し、幾つかの個体を同定した結果を図4に示す。

「QQプロット」では、パッケージcarの中の関数**qq.plot**を用いて、正規分布、t分布、カ

図3 箱ひげ図のDBox



図4 データirisの品種別の箱ひげ図



イ2乗分布、F分布などのQQプロットを作成することができる。また、個体の同定操作のオプションもある。

「散布図」は、パッケージcarの中の関数**scatterplot**を用いる。散布図では、回帰直線と平滑化曲線を加えたり、箱ひげ図を対応する軸の外側に加えたりすることができる。また、層別と個体の同定を指定するオプションもある。図5に横軸をirisのPetal.Length、縦軸をPetal.Widthにした散布図DBoxの画面を示し、その結果を図6に示す。

図5 散布図DBox画面

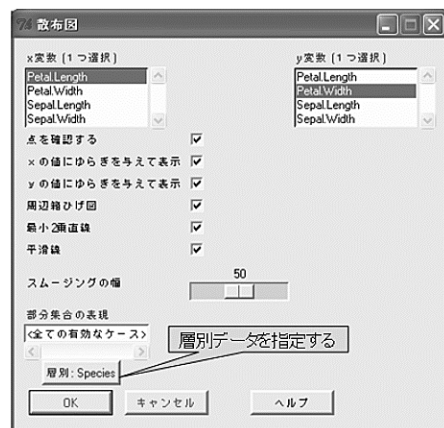
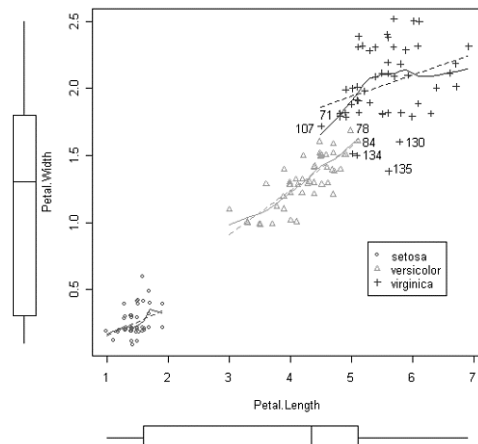
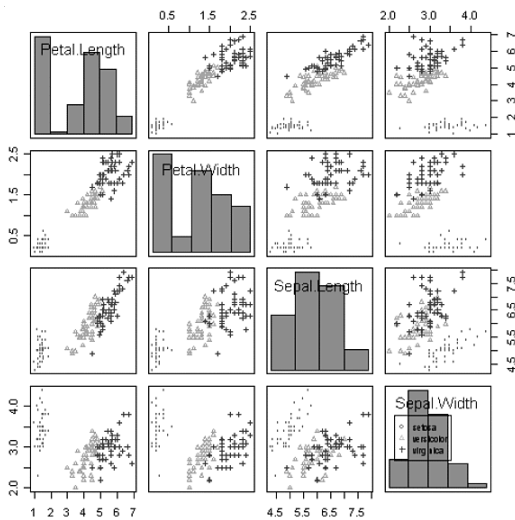


図6 irisの散布図



「散布図行列」は、パッケージcarの中の関数**scatterplot.matrix**を用いる。散布図行列では、データを層別して示すことができ、対角線には、密度プロット（曲線）、ヒストグラム、箱ひげ図、QQプロットのいずれかを加えることができる。また、回帰直線、平滑化曲線を加えることもできる。図7にirisを層別し、対角線にヒストグラムを加えた散布図行列を示す。

図7 irisの散布図行列



「折れ線グラフ」では、パッケージgraphicsの中の関数**matplot**を用いる。横軸には1つの変数、縦軸には複数の変数を指定することができる。図8に横軸をデータirisのPetal.Lengthにし、縦軸をirisの残りの3つの変数にした折れ線グラフを示す。

「平均のプロット」は、パッケージRcmdrの中の関数**plotMeans**を用いる。平均のプロットでは、エラーバーとして標準誤差、標準偏差、信頼区間の中から1つ選択して指定することができる。信頼水準は自由に指定することができる。デフォルトの値は0.95になっている。図9にデータirisのPetal.Lengthの品種別の平均に標

図8 縦軸を複数の変数とした折れ線グラフ

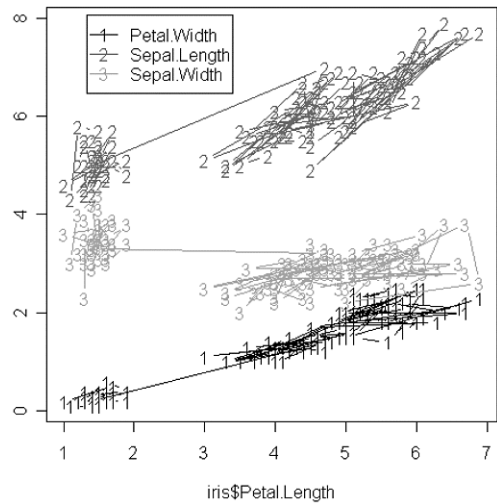
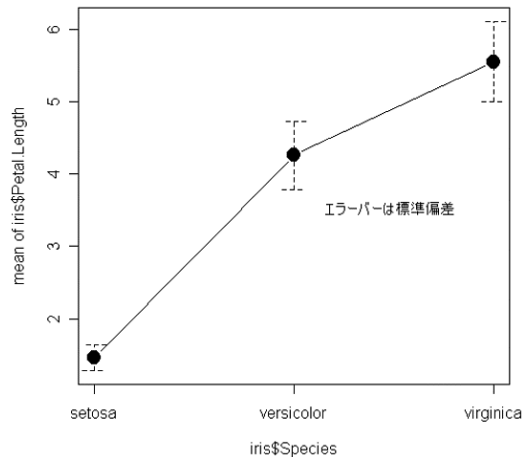


図9 平均プロット



準偏差をエラーバーとした平均プロットを示す。

「3D graph」（3次元グラフ）では、パッケージRcmdrの中の関数**Scatter3DDialog**を用いて作図する。3次元グラフでは、曲面とグリッド線を層別に表示することができる。曲面とグリッド線のオプションとしては、線形最小2乗法、2次最小2乗法、スムーズ回帰、加法回帰

の方法がある。作成されたグラフは、マウスで角度を自由に変えることができる。データirisの3変数（Petal.Width、Sepal.Length、Sepal.Width）を用いたDBoxを図10に、その結果を図11に示す。

作成したグラフは、ビットマップ形式、PDF/Postscript/EPS/png形式で保存することができる。

図10 3次元グラフの設定DBox

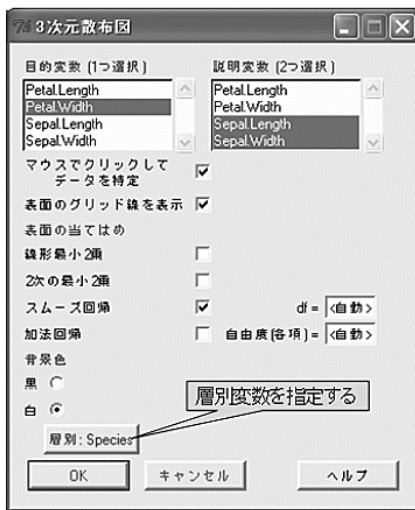
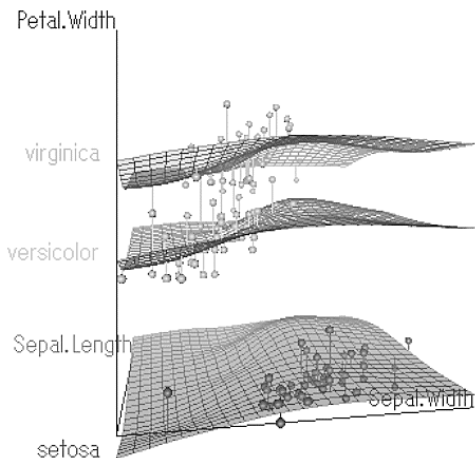


図11 データirisの3次元グラフ



2. モデル

「モデル」メニューでは、「統計量」⇒「モデルへの適応」で作成したモデルについて操作を行う。図12にモデルのメニュー画面を、表2にメニューの項目と用いた関数との対応表を示す。

「アクティブモデルを選択」は、複数のモデルを作成した際、操作対象とするモデルをアクティブ化するのに用いる。

図12 モデルのメニュー画面

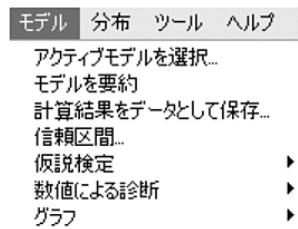


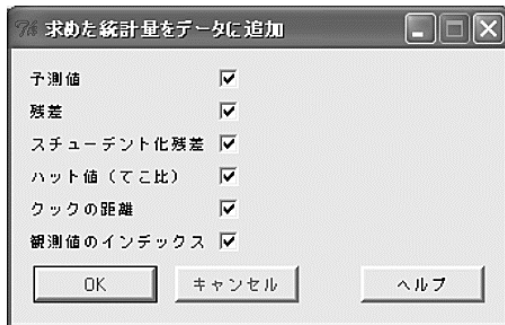
表2 「モデル」に用いる関数の対応表

項目名称	関数	パッケージ
モデルを要約	summary()	stats
計算結果をデータとして保存	influence.measures()	stats
信頼区間	Confint()	Rcmdr
仮説検定	Anova()	car
	anova()	stats
	linear.hypothesis()	car
数値による診断		
分散拡大要因	vif()	car
ブルーシュ・ペーガンの分散不均一性の検定	bptest()	lmtest
自己相関のダービン・ワトソン検定	dwtest()	lmtest
非線形性のRESET検定	resettest()	lmtest
ボンフェローニの外れ値検定	outlier.test()	car
グラフ		
基本的診断プロット	plot.lm()	stats
残差QQプロット	qq.plot()	car
偏残差プロット	cr.plots()	car
偏残差プロット	av.plots()	car
影響プロット	influence.plot()	car
効果プロット	trellis.device()	lattice

「モデルを要約」では、パッケージstatsの中の関数**summary**を用いて、モデルへの適応の結果の要約を呼び出す。

「計算結果をデータとして保存」では、作成したモデルの結果に、統計量を追加する。図13に追加オプションDBoxを示す。

図13 統計量を追加するオプションDBox



「信頼区間」は、パッケージRcmdrの中の関数**Confint**を用いて、モデルの係数について信頼係数の尤度比統計量とワルド統計量を返す。

「仮説検定」では、分散分析表、2つのモデルの比較、線形仮説の検定を行うことが可能である。「分散分析表」は、パッケージcarの中の関数**Anova**を用いて分散分析表を返す。「2つのモデルを比較」は、パッケージstatsの中の関数**anova**を用いて、2つのモデルの残差の分散分析の結果を返す。「線形仮説」は、モデルの係数の帰無仮説結果を返す。

「数値による診断」では、分散拡大要因の計算とブルーシュ・ペーガン (Breusch-Pagan) の分散不均一性の検定、自己相関のダービン・ワトソン (Durbin-Watson) 検定、非線形性のRESET (Ramsey's RESET) 検定、ボンフェローニ (Bonferroni) の外れ値検定を行う。

「グラフ」では、基本的診断プロット、残差QQプロット、要素+偏残差プロット、偏回帰

プロットなどを作成する。

基本的診断プロットでは、パッケージstatsの中の関数**plot.lm**を用いて "Residuals vs Fitted"、"Normal Q-Q"、"Scale-Location"、"Cook's distance" グラフを返す。

例を用いて説明するため、メニューの「データ」⇒「パッケージ内のデータ」から、パッケージdatasetの中のairqualityを読み込み、データメニューの「統計量」⇒「モデルへの適応」⇒「一般線形化モデル」をクリックし、図14のDBoxのとおり指定し、モデルGLM.1を作成する。

図14 一般線形モデルのDBox



作成したGLM.1モデルの基本的診断プロットは、「モデル」⇒「グラフ」⇒「基本的診断プロット」で作成することができる。その結果を図15に示す。

3. 分布

メニュー「分布」では、正規分布、t分布、カイ2乗分布、F分布、2項分布、ポアソン分布の分位点と確率などを計算し、確率分布のグラフを作成する。図16にメニュー画面を示す。

図15 モデルの基本的な診断プロット

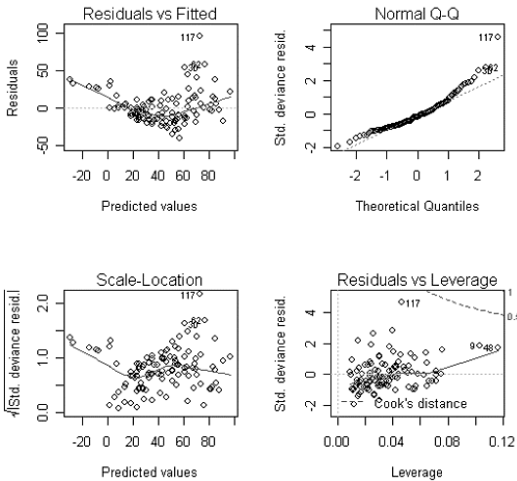
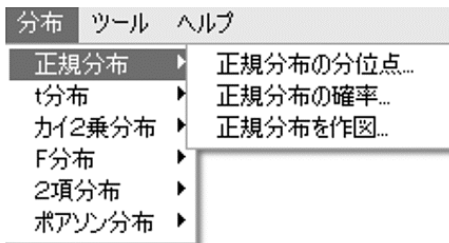


図16 メニュー「分布」の画面



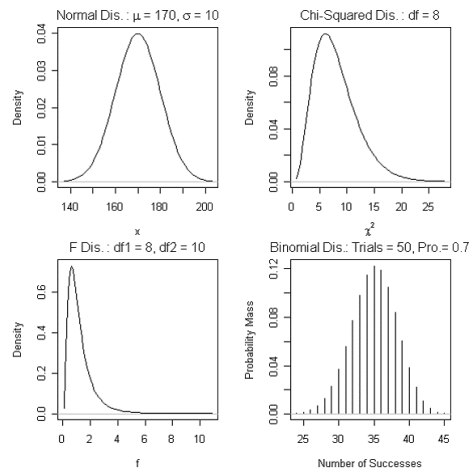
計算はパッケージstatsの中のNormal、TDist、Chisquare、FDist、Binomial、Poisson分布を用いる。これらの文字列は関数ではなく確率分布の分位点、確率、密度、乱数を求める関数の総称である。例えば、正規分布Normalは、確率分布の分位点関数qnorm、分布の確率関数pnorm、分布の密度関数dnorm、乱数関数rnormの総称である。

確率分布の作図について例を示す。「正規分布を作図」をクリックすると図17に示す「正規分布」のDBoxが開かれる。図17のように平均、分散を指定し、[OK]ボタンを押すとR画面にグラフを返す。図18に幾つかのグラフの例を示す。

図17 正規分布DBox



図18 確率分布グラフ



メニュー「ツール」には、「パッケージのロード」と「オプション」がある。「パッケージのロード」では、RにインストールされているパッケージをRコマンドーに読み込み、「オプション」では、Rコマンドーの操作環境などを設定する。

Rコマンドーに関しては、品質管理にポイントを置いた参考書がある[1]。

*参考文献

[1] 荒木孝治(2005): フリーソフトウェアRによる統計的品質管理入門: 日科技連出版社.