

データ解析

鈴木 大慈
理学部情報科学科
西八号館 W707 号室
s-taiji@is.titech.ac.jp

この講義について

講義の目的

- フリーの統計解析用言語「R」を用いてデータ解析の仕方を学ぶ.
- 実際に自分で手を動かすことによってデータ解析手法を習得.
- そのため頻繁にレポート提出をしてもらう.

成績評価

- $\min(\text{出席 } 30\% + \text{レポート } 80\%, 100)$

前提知識

- 確率統計の基本的な知識があることが望ましい.
- 前学期の確率統計第二を取っていれば問題なし.

Rとは

- オープンソース・フリーソフトウェア の統計解析向けのプログラミング言語及びその開発実行環境.
- Rの使い方に慣れておけば他の言語にも活用可能.
 - 類似言語 : Matlab, Octave, Python
- 多くの推定・検定方法が実装されていて、誰でも簡単に統計解析ができる.

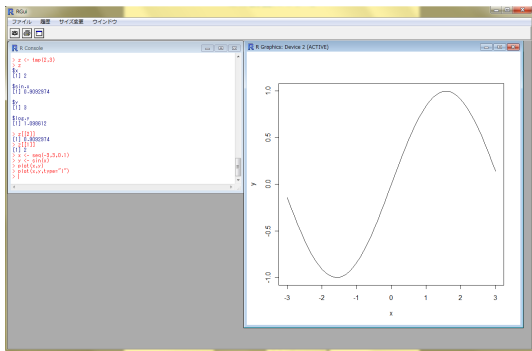


Figure : R の操作画面

講義内容（予定）

- Rの基本的操作
- 確率変数，確率分布の生成
- 回帰分析
 - 単回帰，重回帰
 - ガウスマルコフモデルにおける検定
- 判別分析
 - Fisher 線形判別分析
 - ロジスティック回帰，サポートベクトルマシン
- 検定
 - 適合度検定
 - 独立性検定
 - 分散分析
- 主成分分析
- ノンパラメトリック推定
 - カーネル密度推定
 - カーネル平滑化回帰
- 推薦システム
- クラスタリング
- 時系列解析

Rのインストール

RをWindowsにインストールする方法.

<http://cran.r-project.org/>

へ行き,

「Windows」 → 「base」 → 「Download R (version No.) for Windows」
とクリックしゆく. すると `R-(version No.)-win32.exe` のダウンロードが始まるのでこれを「実行」. 後はデフォルトの設定を利用すれば問題なくインストールできる. 一番最初に言語の選択を尋ねられるので「Japanese」を選べば日本語環境を利用できる.

R for Windows

Subdirectories:

- [base](#) Binaries for base distribution (managed by Duncan Murdoch). This is what you want to [install R for the first time](#).
- [contrib](#) Binaries of contributed packages (managed by Uwe Ligges). There is also information on [third party software](#) available for CRAN Windows services and corresponding environment and make variables.
- [Rtools](#) Tools to build R and R packages (managed by Duncan Murdoch). This is what you want to build your own packages on Windows or to build R itself.

Please do not submit binaries to CRAN. Package developers might want to contact Duncan Murdoch or Uwe Ligges directly in case of questions / suggestions on binaries.

You may also want to read the [R FAQ](#) and [R for Windows FAQ](#).

Note: CRAN does some checks on these binaries for viruses, but cannot give guarantees. Use the normal precautions with downloaded executables.



R-3.0.3 for Windows (32/64 bit)

- [Download R 3.0.3 for Windows](#) (1.4 megabytes, 32/64 bit)
- [Installation and other instructions](#)
- [New features in this version](#)

If you want to double-check that the package you have downloaded exactly matches the package distributed by R, you can compare the [md5sum](#) of the executable. You will need a version of md5sum for windows: both [graphical](#) and [command line versions](#) are available.

Frequently asked questions

- [How do I install R when using Windows Vista?](#)
- [How do I update packages in my previous version of R?](#)
- [Should I run 32-bit or 64-bit R?](#)

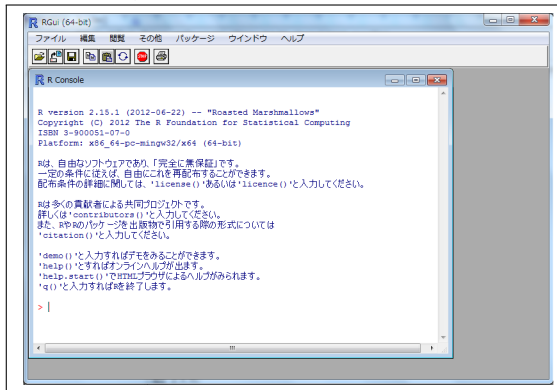
Please see the [R FAQ](#) for general information about R and the [R Windows FAQ](#) for Windows-specific information.

Other builds

- Pre-release test versions are available [here](#).
- Patches to this release are incorporated in the [r-patched snapshot build](#).
- A build of the development version (which will eventually become the next major release of R) is available in the [r-devel snapshot build](#).

Rの起動と終了

- Rの起動画面：



```
R GUI (64-bit)
ファイル 編集 閲覧 その他 パッケージ ウィンドウ ヘルプ

R Console

R version 2.15.1 (2012-06-22) -- "Roasted Marshmallows"
Copyright (C) 2012 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
Platform: x86_64-pc-mingw32/x64 (64-bit)

Rは、自由なソフトウェアであり、「完全に無保証」です。
一定の条件に従えば、自由にこれを再配布することができます。
配布条件の詳細に関しては、'license()'あるいは'licence()'と入力してください。

Rは多くの貢献者による共同プロジェクトです。
詳しくは'contributors()'と入力してください。
また、RやRのパッケージを出版物で引用する際の形式については
'citation()'と入力してください。

'demo()'と入力すればデモをみることができます。
'help()'とすればオンラインヘルプが出ます。
'help.start()'でHTMLブラウザによるヘルプがみられます。
'q()'と入力すればRを終了します。

> |
```

- Rの終了：q() と入力。

基本的なコマンド

- 関数のヘルプ : `help('関数名')`
例 : `help('rnorm')`
- ワーキングディレクトリの表示 : `getwd()`
- ワーキングディレクトリの移動 : `setwd("ディレクトリパス")`
※ バックスラッシュをパス区切りに使う場合は、\\のように二回ずつ入れる。
× `setwd("C:\Users\..")`, ○ `setwd("C:\\Users\\..")`
- パッケージのインストール : `install.packages("パッケージ名")`
例 : `install.packages("Rcmdr")`
※ ミラーサイトの指定を求められたら「Japan (Tokyo)」など近場を選べばよい。
- パッケージ (ライブラリ) の読み込み : `library(ライブラリ名)`
例 : `library(Rcmdr)`
- スクリプトファイルの実行 : `source("hoge.R")`
※ `hoge.R` の中に書いてあるコマンドが実行される。
※ 関数の定義を書いておけば関数を読み込むことができる。

講義資料

授業で用いたスクリプトは以下のリンクに随時掲載する.

<http://www.is.titech.ac.jp/s-taiji/lecture/dataanalysis/dataanalysis.html>

また、レポート問題や講義資料は OCW にもアップロードする.

今日の講義内容

- R ことはじめ
- ベクトル操作
- 行列操作
- データフレーム
- リスト

R 関連参考資料

- CRAN (Complete R Archive Network). R 本体および千を越す膨大な貢献パッケージが入手できる. 日本のミラーサイトは <http://cran.md.tsukuba.ac.jp>
- RjpWiki. 日本の R ユーザーが運営する情報サイト. 各種 Tips や質問コーナーがある. URL は <http://www.okada.jp.org/RWiki/>
- 岡田昌史編「The R Book データ解析環境 R の活用事例集」, 九天社 (2004)
- 間瀬茂著「R プログラミングマニュアル」, 数理工学社 (2007)