R を使った情報教育

奥　村　晴　彦[†1]

C 言語や Java 言語によるプログラミング教育は IT 業界に職を求める学生にとっては適切であるが，これらの言語で有用なプログラムを開発できるようになるまでにはかなりの努力を要し，モチベーションを維持するのは難しい。そこで，対話型のインタープリタ環境で，打ち込めばすぐに結果が得られ，しかも身の回りのデータを簡単にグラフ化・統計的に解析できる R を授業に取り入れることを本稿で提案する。

Informatics Education using R

HARUHIKO OKUMURA[†1]

Although teaching classical programming languages such as C and Java is appropriate for students who seek jobs in IT industry, for many other students it is becoming difficult to maintain motivation before they can write meaningful programs in these languages. In this paper, we propose using R in programming education. With its easy-to-use interactive environment, an R program can be promptly executed when typed in. It is especially suited for drawing statistical graphs and conducting statistical analysis.

1. はじめに

高校の情報科や大学初年次の一般情報教育でもしプログラミングを教えるとすれば，言語は何が適切であろうか。ここでは

- Windows，Mac，Linux 等で動作するフリーな処理系がある
- 型宣言やコンパイル・リンクといったことを意識せず簡単に使える
- グラフィック機能をサポートし，データの視覚化が簡単にできる
- メニューやエラーメッセージが日本語である

といった条件を満たすものとして，R[1] を提案する。

R の元となった S は，1975–1976 年ごろベル研究所で開発された[2]。S は統計（statistics）の頭文字で，同じ研究所で作られた C 言語の伝統に則り，単純な名前が付けられた。R は S の言語仕様にほぼ基づいて 1993 年に作られたオープンソースの処理系であり，現在も活発な開発が続いている。なお，S の現代的な実装としては S-PLUS もある（これはオープンソースではない）。

R は，C 言語などの手続き型の構造化言語と似た文法に従う一種のオブジェクト指向言語で，対話型の処理系を持ち，プログラムを打ち込めば即実行される。GUI のフロントエンド Rcmdr（R コマンダー）や，Excel アドインの RExcel もあるが，本稿では素の R について扱う。

R は，Unicode 文字列，正規表現，複素数，ベクトル，行列が扱え，Excel と違って[3)–5)] プロの統計学者が使えるものである。大学の統計教育の現場では広く用いられるようになりつつあるが，本稿の想定する大学初年次の一般情報教育あるいは高校「情報」では，あまり利用例を聞かない（後述の RjpWiki サイト[10] の「R 本リスト」に列挙されている和書 66 冊にも，このような利用を想定した本は見当たらない）。本稿ではこのような場面を想定した R の利用を提案する。

なお，ソフトを自由にインストールできない PC 教室が多い状況を考えれば，Windows で展開するだけで一般ユーザでも使える点も R の強みである。

2. 関数電卓・グラフ電卓としての利用

数学では R は，まず数表代わりに使える。例えば $\sqrt{2}$ の概数は

```
sqrt(2)
```

と打ち込めば 1.414214 と得られる（表示精度はオプションで加減できる）。0 から 10 まで 1 刻みで $\sqrt{x}$ を求めて表にするには

```
for (x in 0:10) cat(x, sqrt(x), "\n")
```

と打ち込めばよい。

グラフ描画も R の得意とするところである。例えば範囲 $[-1,5]$ で $y=x^2-4x+1$ のグラフを描くには

```
curve(x^2-4*x+1, xlim=c(-1,5))
```

と打ち込めばよい（図 1）。さらに

†1 三重大学教育学部
Faculty of Education, Mie University

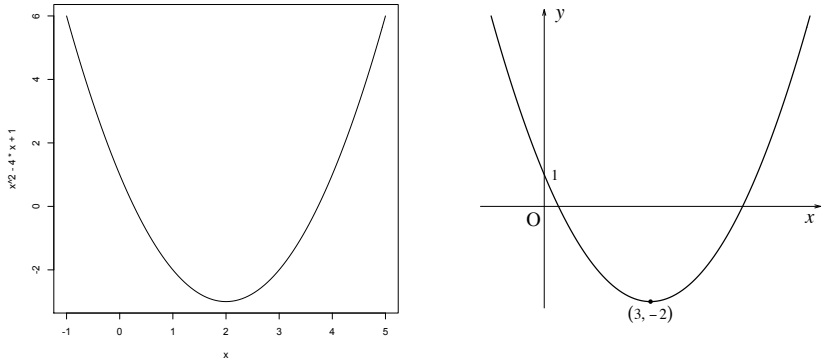


図 1　左：`curve(x^2-4*x+1, xlim=c(-1,5))` の出力。右のような高校数学の教科書風のグラフにするためには少し工夫が必要である（「統計・データ解析」サイト[6] の「グラフの描き方」参照）。

```
locator(1)
```

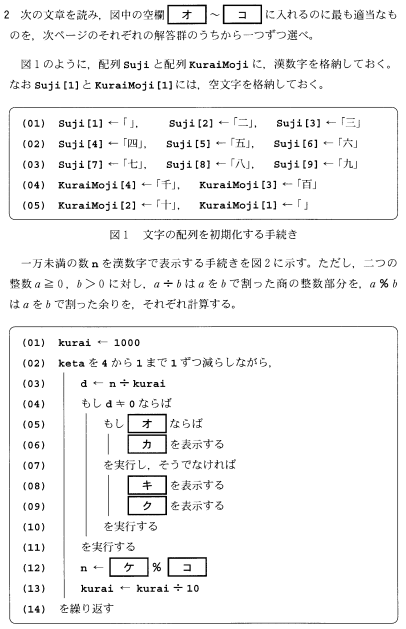
と打ち込んで図上をクリックすればその点の座標値が返される。

3. プログラミング教育への利用

R は汎用プログラミング言語であり，C や Java などと同様にアルゴリズムの学習に使える。ここでは例として，大学入試センター試験「情報関係基礎」の問題を解いてみる。この試験では，DNCL と呼ばれる手続き型言語を使ったプログラミングの問題が毎年出題されている。2010 年の問題を図 2 に示す。

変数は数値型と文字列型が混在するが，DNCL 同様，R は型を意識せずにプログラムを書くことができる。配列の添字が 1 から始まることも DNCL と同様である。

```
Suji = c("","二","三","四","五","六","七","八","九")
KuraiMoji = c("","十","百","千")
n = 1517            # 例えば
kurai = 1000
for (keta in 4:1) {
  d = n %/% kurai   # %/% は整数商
  if (d != 0) {
    if (d == 1 && keta == 1) {
```



問 2　次の文章を読み，図中の空欄 オ ～ コ に入れるのに最も適当なものを，次ページのそれぞれの解答群のうちから一つずつ選べ。

図 1 のように，配列 `Suji` と配列 `KuraiMoji` に，漢数字を格納しておく。なお `Suji[1]` と `KuraiMoji[1]` には，空文字を格納しておく。

```
(01)  Suji[1] ← 「」,    Suji[2] ← 「二」,   Suji[3] ← 「三」
(02)  Suji[4] ← 「四」,  Suji[5] ← 「五」,   Suji[6] ← 「六」
(03)  Suji[7] ← 「七」,  Suji[8] ← 「八」,   Suji[9] ← 「九」
(04)  KuraiMoji[4] ← 「千」,  KuraiMoji[3] ← 「百」
(05)  KuraiMoji[2] ← 「十」,  KuraiMoji[1] ← 「」
```

図 1　文字の配列を初期化する手続き

一万未満の数 n を漢数字で表示する手続きを図 2 に示す。ただし，二つの整数 $a \geqq 0$，$b > 0$ に対し，$a \div b$ は a を b で割った商の整数部分を，a % b は a を b で割った余りを，それぞれ計算する。

```
(01)  kurai ← 1000
(02)  keta を 4 から 1 まで 1 ずつ減らしながら，
(03)      d ← n ÷ kurai
(04)      もし d ≠ 0 ならば
(05)          もし [ オ ] ならば
(06)              [ カ ] を表示する
(07)          を実行し，そうでなければ
(08)              [ キ ] を表示する
(09)              [ ク ] を表示する
(10)          を実行する
(11)      を実行する
(12)      n ← [ ケ ] % [ コ ]
(13)      kurai ← kurai ÷ 10
(14)  を繰り返す
```

図 2　一万未満の数 n を漢数字表示する手続き

図 2　情報関係基礎，2010 年，第 3 問

```
      cat("一")      # cat は出力
    } else {
      cat(Suji[d])
      cat(KuraiMoji[keta])
    }
  }
  n = n %% kurai          # %% は剰余
  kurai = kurai %/% 10    # %/% は整数商
}
```

これくらいの長さになれば，テキストエディタに打ち込んでから，R にコピー＆ペーストするほうが楽である。もっと長いプログラムであれば，R からファイルを読み込むとよい。いずれにしても，上のプログラムを実行すると「千五百十七」と表示する。あるいは

```
kansuji = function(n) {
  kurai = 1000
```

```
  for (keta in 4:1) {
    ...
  }
}
```

のように関数として定義すれば，`kansuji(1517)` で「千五百十七」が得られる。

変数名や関数名にも日本語などの Unicode 文字が使えるので，`KuraiMoji[keta]` の代わりに `位文字[桁]` などと書くこともできる。

4. モデル化とシミュレーションでの利用

高校「情報 B」や後継の「情報の科学」の「モデル化とシミュレーション」では，何らかの数理モデルを設定して模擬実験を行う。

例えば，単純な（hit-or-miss）モンテカルロ計算で円周率 π を求めることができる。R では

```
x = runif(10000)      # 0〜1の一様乱数10000個からなるベクトル
y = runif(10000)
mean(x^2 + y^2 < 1)  # 単位円内に入る割合
```

で $\pi/4$ の概数が求められる。これは概念的にはおもしろいが，収束は非常に遅く，実用的ではない。

あるいは，硬貨を 100 回投げる実験の代わりに使うことができる。

```
sample(0:1, 100, replace=TRUE)
```

と打ち込めば，`0:1` つまり 0 から 1 までの整数から復元抽出（`replace=TRUE`）で 100 個取り出すことができる。この合計

```
sum(sample(0:1, 100, replace=TRUE))
```

が表（おもて）の枚数である。

より実用的な利用法として，平均値や中央値といった統計量の誤差をシミュレーションで求めることができる。ある母集団から n 個を取り出した標本が与えられたとき，元の母集団は未知であるので，標本で母集団を近似し，この標本から重複を許して n 個を再抽出（リサンプル）し，統計量を計算する。これを何度も繰り返せば，統計量のばらつき方がわかる。例えばサンプルが

```
x = c(85,43,72,90,55)
```

であれば，

```
sample(x, replace=TRUE)
```

で復元抽出が行える。その平均値は

```
mean(sample(x, replace=TRUE))
```

で求められる。これをクラス全員で行って結果を持ち寄れば，平均値の分布がわかり，したがって平均値に含まれる誤差が見積もれる。ここまでは数値を書き込んだカードをシャフルして抜き出してもできる。R で自動化するには，例えば

```
a = replicate(10000, mean(sample(x,replace=TRUE)))
```

とすれば 10000 回の結果が配列 `a` に代入される。その度数分布図は

```
hist(a)
```

で描くことができる。同じことを表計算ソフトでしようとすればたいへん手間がかかる。

このように標本から標本を再抽出することをブートストラップ（bootstrap）という[7]。標本だけから統計量の誤差分布を求める方法として近年広く用いられているが，原理は高校生でも十分理解できる。

5. データの視覚化での利用

数学科で復活した統計分野と並んで，高校「情報」でもデータの視覚化を実習に取り込むことが多い[8]。

このいわゆるデータ・リテラシーに関する内容では，まずグラフを正しく描き，正しく読み取る能力を養うことが望まれる。図 3 のような OECD 生徒の学習到達度調査（PISA）問題の影響もあって，こういった能力はますます重視されている。

R ではベクトル `x`（や `y`）に入ったデータの度数分布図は `hist(x)`，散布図は `plot(x,y)` という簡単なコマンドで描くことができる。また，平均値 `mean(x)`，中央値 `median(x)`，四分位偏差 `IQR(x)`，分散 `var(x)`，標準偏差 `sd(x)`，相関係数 `cor(x,y)` およびその検定 `cor.test(x,y)` が簡単に計算できる。ただし分散・標準偏差は統計学で用いる $n-1$ で割る方式であり，学校数学で用いる n で割る方式にするには，例えば

```
varp = function(x) { var(x) * (length(x) - 1) / length(x) }
stdevp = function(x) { sqrt(varp(x)) }
```

のようにして新たな関数 `varp`，`stdevp` を定義する必要がある。

6. 大学での情報教育への利用

大学で R を教えている事例は無数にある。RjpWiki サイト[10] には R を使った大学の授業一覧を列挙した「R シラバス」ページがあるが，実際にはもっと多いと思われる。

大学で教科書として使えるような『R による○○』といった本が大量に書かれている。

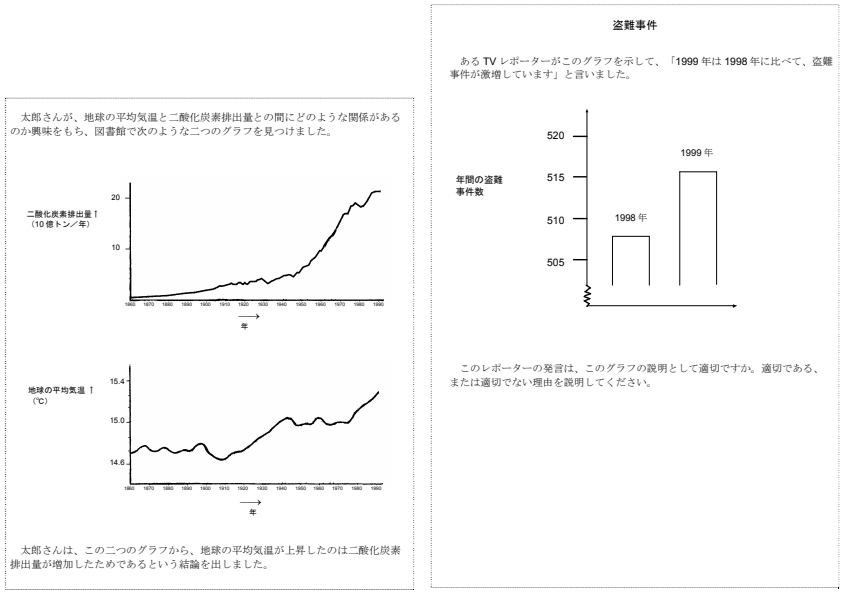


図 3　PISA 問題例[9]。左は相関関係と因果関係の違い，右は省略棒グラフ。

RjpWiki サイト[10] の「R 本リスト」ページには，本稿執筆時点で和書だけで 66 冊も列挙されている。ただ，先に述べたように，このリストを見ても，一般情報処理教育で使われている例は見当たらない。

そこで，実際に初年次の必修の授業で R を利用してみた。例えば，今まで JavaScript で試させていた Collatz の問題[11] [pp. 154–156] を R に切り替えることによって，実習の時間効率が向上した。また，上で述べたセンター試験の問題とその R 版を示し，試験問題を解いたら R で確認するよう指示して実習させたところ，当初の解答を入れて実行してもなかなか正しい結果が得られないのが予想外であったらしく，全員が非常に熱心に取り組んだ。その過程で != が $\neq$ であるといった知識が習得されていく様子が見て取れた。

初年次教育でない例としては，情報理論で音声のサンプル値をマルコフ過程と見たときのエントロピーを求めさせる実習で，簡単な例では全員が Excel を使っていたが，サンプル値が数百万程度になると，一人の学生が統計の授業で学んだ R を使い始めたのをきっかけに全員が R に移行し，しかも数行の入力で簡単に目的が達成できて，学生たちも驚嘆した。

7. 結論と課題

本稿では R を使った情報教育を提案した。従来 C や Java，JavaScript，VBA などが使われてきたプログラミング教育の場では，簡単に R に移行することが可能である。

もちろん R が唯一の解というわけではない。個々の例を見れば，表計算ソフトで十分な場合や，Maxima のほうが適切な場合もあるし，大学入試センター試験「情報関係基礎」に限れば，DNCL の実装である PEN[12] が便利であろう。ここでは，一つ覚えればどの場合にも使えることと，さらにプロの道具として将来も使えることを考えて R を提案している。

まだ R による教育の評価も逸話的な段階であるが，これからも実践を続けて評価を確立していくつもりである。

参　考　文　献

1) R Development Core Team, “R:A Language and Environment for Statistical Computing”, http://www.R-project.org/
2) Richard A. Becker, “A Brief History of S,” http://cm.bell-labs.com/stat/doc/94.11.ps
3) B.D. McCullough and David A. Heiser, “On the accuracy of statistical procedures in Microsoft Excel 2007,” *Computational Statistics and Data Analysis* **52**, 4570–4578 (2008).
4) A. Talha Yalta, “The accuracy of statistical distributions in Microsoft® Excel 2007,” *Computational Statistics and Data Analysis* **52**, 4579–4586 (2008).
5) B.D. McCullough, “Microsoft Excel’s ‘Not The Wichmann-Hill’ random number generators,” *Computational Statistics and Data Analysis* **52**, 4587–4593 (2008).
6) 奥村晴彦「統計・データ解析」 http://oku.edu.mie-u.ac.jp/~okumura/stat/
7) B. Efron, “Bootstrap Methods: Another Look at the Jackknife,” *Annals of Statistics* **7**, 1–26, doi:10.1214/aos/1176344552 (1979).
8) 奥村晴彦「情報教育と統計」，情報処理学会研究報告「コンピュータと教育」2008-CE-97（情処技報 Vol. 2008, No. 128），pp. 81–88, ISSN 0919–6072
9) 文部科学省「国際学力調査」http://www.mext.go.jp/a_menu/shotou/gakuryoku-chousa/sonota/07032813.htm
10) RjpWiki, http://www.okada.jp.org/RWiki/
11) 奥村晴彦＋三重大学学術情報ポータルセンター『基礎からわかる情報リテラシー』技術評論社，2007 年。
12) 初学者向けプログラミング学習環境 PEN http://pen.moe.hm/