

心理データ解析演習 (2011.5.11)

R入門

PD 井関龍太

Rとは

- オープンソースの統計解析向けプログラミング言語およびその開発実行環境
 - プログラム言語：S言語をもとにしたベクトルや行列の操作がしやすい計算機言語
 - ソフトウェア：R Development Core Teamによってメンテナンス・アップデートされているコマンドライン方式のソフト
- CRAN (The Comprehensive R Archive Network)
 - <http://cran.r-project.org/>

Rの利点と欠点

□利点

- 無料**で使える（いつでもどこでも）
- 豊富な**パッケージ**（SPSSにはない分析も）
- カスタマイズ可能性**が高い（自分の使用法に合った関数が組める）

□欠点

- 使用に際しての**公式サポート**はない（自助努力と情報交換が重要）
- 関数が**多すぎる**（同じことをやるための方法が複数存在する）
- コマンドライン**への抵抗感（慣れましょう）

発表の流れ

□Rの基本

- 計算してみよう
- 代入と関数
- データの型

□統計解析：重回帰分析

- 基本統計量，相関，対散布図，非標準偏回帰係数，標準偏回帰係数

□統計解析：因子分析

- 固有値の計算，因子分析の実行，出力

□パッケージを使おう

Rの基本：計算してみよう

□四則演算（+ - * /）

- 数字と数学記号を入力する

```
> 3 * 5
```

- ✓「>」の後にユーザーのキー入力
- ✓文字間のスペースの有無は任意

- “Enter”を押すと解が得られる

```
[1] 15
```

- ✓出力については「[1], [2]」のような形で行番号がつく

基本的な流れ

Rの基本：代入と関数（1）

□ “<-”を作って，変数に値を代入する

- 適当な文字列に数値を代入

```
> x1 <- 7
```

✓ Enterを押しても何も出力はされない

- 変数名を入力してEnterを押す

```
[1] 7
```

✓ 格納された値が出力される

✓ Rでは大文字と小文字を区別する
(x1 と X1 は別の変数)

Rの基本：代入と関数（2）

□代入した変数は，元の値と同じように使える

- 変数に対して計算式を入力

```
> x1 * 2  
[1] 14
```

□関数を適用する

- 関数(変数)を入力

```
> log(x1)  
[1] 1.94591
```

- ✓対数に変換するための関数logを使用
- ✓変数に対しても数値に対しても適用できる

Rの基本：データの型（1）

- もっと複雑な関数を使うには.....
 - 平均値 複数の数値に関数を適用
 - 回帰分析 複数の変数群に関数を適用
数値を意味のあるセットにする
- データ型：Rのデータは必ず何らかの型を持つ（基本的なもののみ紹介）
 - 数値型：数字
 - ✓ 半角英数で入力する
 - 文字型：文字もデータとして扱える
 - ✓ 半角英数の「"”」で囲む（日本語も可）

Rの基本：データの型（2）

□ベクトル：複数のデータを一行にまとめたもの

- 関数cを使って作る

```
> c(5, 8, 3, 7)
```

- ✓変数に代入してみよう
- ✓関数meanを適用してみよう

□行列： $n \times m$ の形にデータをまとめたもの（3次元以上の行列も可）

- 関数matrixを使って作る

```
> matrix(c(5, 2, 6, 8, 3, 4), 2, 3)
```

- ✓matrix（ベクトル，行数，列数）

Rの基本：データの型（3）

□ データフレーム：様々な型のデータをまとめて扱える（複雑な統計解析に必須）

● 関数 `data.frame` を使って作る

```
> dat <- data.frame(y = c(5, 9, 6), x1 =  
c(2, 4, 1), x2 = c(3, 3, 5))
```

✓ `data.frame(列名1 = ベクトル1, 列名2 = ベクトル2, ...)`

● 列名，行番号でデータにアクセスできる

	y	x1	x2
1	5	2	3
2	9	4	3
3	6	1	5

✓ 変数名\$列名で列ごとのデータにアクセス

✓ 変数名[n, m]で指定した行・列番号のデータにアクセス

ワンポイント： クリップボードからのデータのコピー

□ Excelなどで作ったデータをRに読み込む

- Excel等を開いて、コピーしたいデータを選択する
- 右クリック コピー
- read.table関数を使う

```
> dat <- read.table("clipboard", T)
```

- ✓ read.table("clipbord", ヘッダの有無)
- ✓ ヘッダがある場合はT (TRUE), ない場合はF (FALSE) を入力 (省略すると, F)
- ✓ この関数を使うと、取り込んだデータはデータフレームになっている (ヘッダは列名になっている)

統計解析：重回帰分析（1）

□データ例：青年期女子の減量希望量

- 独立変数：身長・体重・年齢・体型不満度
- 従属変数：減量希望量（kg）

✓服部環・海保博之（1996）. Q&A心理データ解析
福村出版 表6-1（p. 129）

□基本統計量の確認

- 関数`sapply`を使って各変数の平均値，標準偏差を一挙に計算する

```
> sapply(dat, function(x) mean(x))
```

✓`sapply(データフレーム, function(x) 適用したい関数(x))`

統計解析：重回帰分析（２）

□ 相関関係の確認

- 関数 `cor` を使って相関行列を作る

```
> cor(dat)
```

- ✓ `cor` (データフレーム)
- ✓ `round(cor(データフレーム), 小数点以下の桁数)` で数値をまるめることができる

□ 対散布図を検討する

- 関数 `pairs` を使って変数の各対の散布図を作る

```
> pairs(dat)
```

- ✓ `pairs`(データフレーム)

統計解析：重回帰分析（3）

□ 重回帰分析の実行

- 関数 `lm` を使って式を書く（一般線形モデル）

```
> ans <- lm(expectancy ~ height + weight +  
age + complaint, dat)
```

- ✓ `lm(y ~ x1 + x2 +, データフレーム)`

- ✓ 結果を変数に保存すること

- 関数 `summary` を使って解にアクセス

```
> summary(ans)
```

- ✓ `summary(lm関数の返り値)`

- ✓ 残差，非標準偏回帰係数， R^2 ，F値

統計解析：重回帰分析（４）

□標準偏回帰係数を算出する

- データをあらかじめ標準化する（関数scale）

```
> sdat <- scale(dat)
```

✓ scale(データ)

✓ ただし，この関数を適用すると，データの型がデータフレームでなくなってしまう

- 標準化したデータをデータフレームに変換

```
> sdatf <- data.frame(sdat)
```

✓ 新しく作ったデータフレームを使って，lm関数で重回帰分析を行う

ワンポイント：ヘルプを使おう

□関数の使い方を調べる

- help(関数名), または, ?関数名

□ヘルプをよく読もう

- Description : 関数の基本的な説明
- Usage : 関数の記述の仕方
- Arguments : 引数 (**データの型**に注意!)
- Value : 返り値
 - ✓ 表示に **summary** や **print** を要するものも.....
- Author : 関数の作者
- See Also : 関連する他の関数の参照
- Examples : 実際に **貼り付けて動かせる**

統計解析：因子分析（1）

□ データ例：5教科と国家試験の成績

- テストの得点：国語・社会・数学・理科・英語・国家試験

✓ 服部環・海保博之 (1996). Q&A心理データ解析
福村出版 練習用の基本データ (p. 213)

□ 固有値を計算する

- 関数 `eigen` を使う

```
> eigen(cor(dat))
```

- ✓ `eigen`(相関行列)；最初から相関行列になっていれば，`cor()`は必要ない
- ✓ 固有値と固有ベクトルが出力される

統計解析：因子分析（2）

□ 因子分析を行う

● 関数 `factanal` を使う（最尤法）

```
> factanal(dat, 2)
```

- ✓ `factanal`(データフレーム, 因子数, rotation = “回転の方法”)
- ✓ 因子数は必須；指定しないと動作しない
- ✓ 回転の方法は, none, varimax, promax から選べる（省略すると, varimaxになる）
- ✓ 青木先生の関数 `factanal2` を使えば, 因子数の指定を省略できる (<http://aoki2.si.gunma-u.ac.jp/R/factanal2.html>)

統計解析：因子分析（3）

Call:
factanal(x = dat, factors = 2, rotation = "varimax")

Uniquenesses:
kokugo shakai sugaku rika eigo
0.457 0.749 0.690 0.449 0.659

独自性

Loadings:

	Factor1	Factor2
kokugo		0.731
shakai	0.181	0.467
sugaku	0.525	0.186
rika	0.738	
eigo	0.454	0.368

因子負荷量

0.1未満は空白

	Factor1	Factor2
SS loadings	1.068	0.928
Proportion Var	0.214	0.186
Cumulative Var	0.214	0.399

因子寄与
因子寄与率
累積寄与率

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 0.04 on 1 degree of freedom.
The p-value is 0.836

データとモデルの解離
の検定量

統計解析：因子分析（４）

□補足事項

- 0.1以下の因子負荷も出力させる

```
> print(ans, cutoff = 0)
```

- ✓ print(factanalの出力, cutoff = 数値)

- ✓ カットオフ値より大きい値を表示させる指定

- 指定した変数のみを選択して因子分析にかける

```
> factanal(~kokugo + sugaku + eigo, factors  
= 1, data = dat)
```

- ✓ factanal(~変数1+変数2+....., factors = 因子数, data = データフレーム)

パッケージを使おう（1）

□ インストールした状態のR = すべての機能が有効になっているわけではない

● **パッケージ**：いろいろな関数がセットになったもの

✓ テーマごとに分かれた様々なパッケージがある

✓ ウェブ等を経由して、追加インストールする

□ パッケージの例

● **lme4**：一般化線形混合モデル，マルチレベルモデル

● **psych**：心理学向けの様々な拡張機能

✓ スクリーンプロット，様々な回転による因子分析

● **sem**：構造方程式モデリング

パッケージを使おう（２）

□パッケージのインストール（最初のみ）

- Rのツールバーから“パッケージ”を選択
- “パッケージのインストール”を選択
 - ✓ミラーサイトの選択画面が開く
- ミラーサイトを選ぶ
 - ✓どれでもよいが，“Japan”の方がたぶん速い
 - ✓パッケージの選択画面が開く
- インストールしたいパッケージを選択

□パッケージの読み込み（毎回必要）

- “パッケージの読み込み”を選択
- 使用したいパッケージを選択

参考サイト

- RjpWiki (<http://www.okada.jp.org/RWiki/>)
 - ✓ 巨大データベース , Q&Aも
- R-Tips (<http://cse.naro.affrc.go.jp/takezawa/r-tips/r.html>)
 - ✓ 操作のレファレンスとして
- Rオンラインガイド (http://psycho.edu.yamaguchi-u.ac.jp/?page_id=242)
 - ✓ 動画で学ぶ操作説明
- Rで学ぶデータサイエンス
(<http://www1.doshisha.ac.jp/~mjcin/R/>)
 - ✓ 各種分析手法をじっくり解説
- R による統計処理 (<http://aoki2.si.gunma-u.ac.jp/R/>)
 - ✓ 手法の解説と分析に使える関数群

参考書籍

特に心理学での利用を考慮したもの

- 服部環 (2011). 心理・教育のための R によるデータ解析 福村出版
 - 各種解析法の解説をしながら，それぞれの分析を行うための関数とパッケージを適宜紹介；より高度な内容にもふれている
- 緒賀郷志 (2010). Rによる心理・調査データ解析 東京図書
 - 心理学でよく使用する分析一通りを R と R コマンドで実行する方法を紹介；図版も多い