

Rの基礎知識

資料: Rによる統計的検定と推定

著者: 内田 治

西澤 英子

範囲: 第1章 (P.1～P.19)

60191008 大石 聖也

R(R言語、R環境)

- ・コンピュータ言語として開発されたフリーのソフトウェア
- ・データ操作、データ解析、グラフィックス機能に優れ、統計計算を行うための
さまざまな手法を実装
- ・オープンソースであるため、世界中のユーザがプログラム開発や検証に参加できる
- ・最大の魅力は、無料で使うことができること

CRAN (Comprehensive R Archive Network)

- CRANには、Rに関するさまざまな情報が掲載されている
- Rのプログラムを無料でダウンロードできる
- 通信時間を考慮すると、近場のミラーサイト(CRAN上の情報と同じ情報を持つサイト)を利用したほうがダウンロードの際に便利
- CRANのURL・・・<http://cran.r-project.org/>

Rのインストール

手順1: 画面左側の「Mirrors」を選択

手順2: 画面右側から近場のミラーサイト選択

手順3: インストールするコンピュータのOSに合わせ、ダウンロードのリンクを選択

手順4: 「install R for the first time」を選択

手順5: 「Download R 2.14.1 for Windows」を選択

手順6: 「実行」ボタンをクリック

パッケージ

- ・特定の統計計算やグラフィックス操作を実行するための関数やデータ、リファレンス情報などまとめられている単位
- ・初回時のインストールで、Rの基本システムに加え、さまざまなパッケージがインストールされる
- ・インストール済みのパッケージは、コンソール(R起動後の画面で「R Console」とタイトルがついた領域)で、「library()」と入力、実行して確認できる
- ・ただし、特定のパッケージを追加するには、CRANから追加する必要がある
- ・パッケージの追加は、R起動後の画面上からでも指示できる

パッケージの追加(手順1～3)

Excelファイルからの読み込みやExcelファイルへの書き出しを行うことができるxlsxパッケージの追加

手順1:「パッケージ」メニューの「CRANミラーサイトの設定」を選択

手順2: 最寄りのミラーサイトを選択

手順3:「OK」ボタンをクリック

手順1からの追加手順

手順1の際に次のメッセージが表示される場合は、Rを終了してから以下の追加手順を実施する

警告メッセージ:

In open.connection(con, "r"):

'cran.r-project.org' をポート 80 でコネクトできません

追加手順1: コンソールに「q()」を入力しRを終了
作業スペースの保存は「いいえ」を選択

追加手順2: デスクトップ上のRのアイコンを右クリックし、「プロパティ」を選択

追加手順3: プロパティ画面で「リンク先」に「--internet2」を追加入力し、「OK」ボタンをクリック

※手順4以降に進む前に、手順3(ミラーサイトの設定)までを実施する

パッケージの追加(手順4～5)

手順4:「パッケージ」メニューの「パッケージのインストール」を選択

手順5:ダウンロードするパッケージとして「xlsx」を選択し、「OK」ボタンをクリック

※この手順ではxlsxパッケージのみをインストールしたが、関連パッケージであるxlsxjarsとrJavaも自動的にインストールされる

計算式の代入

変数への代入

例1: 計算して変数xに代入

- ・1+1を変数xへ代入する
- ・その結果を表示する

```
> x <- 1+1  
> x  
[1] 2
```

例2: 数式を変数yに代入

- ・x+100を変数yへ代入する

```
> y <- x+100  
> y  
[1] 102
```

※例1の結果(x)を使い、さらに別の変数(y)に値を代入することができる

データの代入 ~関数c()を使ったベクトルの作成~

データは関数「c()」を使い、ベクトル形式で入力する

例3: データの入力方法

- ・5人の学生の英語の試験結果、50, 60, 80, 30, 70であったとする。このデータを変数x1へ代入する。

```
> x1 <- c(50, 60, 80, 30, 70)
```

- ・同じ5人の学生の数学の試験結果が、40, 70, 50, 40, 80であったとする。このデータを変数x2へ代入する。

```
> x2 <- c(40, 70, 50, 40, 80)
```

データの代入 ～基本的な統計量を計算する～

例4: 要約統計量の算出

- 関数summary()を使用して変数x1の基本的な統計量を求める

```
> summary(x1)
```

```
Min. 1st Qu. Median Mean 3rd Qu. Max.
 30    50    60    58    70    80
```

※関数summary()で算出される統計量は以下である
基本的な統計量

統計量	説明
Min.	最小値
1st Qu.	25パーセントタイル値
Median	中央値
Mean	平均値
3rd Qu.	75パーセントタイル値
Max.	最大値

データの代入 ～基本的な統計量を計算する～

例5: 標準偏差と分散の算出

- ・変数x1とx2の標準偏差と分散を求める。標準偏差を求めるには関数sd()を、分散を求めるには関数var()を使用する。

```
> sd(1) ; sd(x2) ; var(x1) ; var(x2)
```

※命令文はセミコロン(;)を使用すると並列させることができる

データフレームの作成

- ・数値データや文字データが混在したデータ表を、Rではデータフレームとして扱う
- ・データフレームの作成には、以下の方法などが考えられる
 - 1、複数のベクトルを作成してから、データフレームとしてまとめる
 - 2、保存済みの外部ファイルを読み込む

データフレームの作成 ~データフレームの作成1~

ベクトルを結合してデータフレームを作成する

- ・例3で作成したx1とx2を結合し、データフレームに「date01」という名前をつける

作成したデータ表のイメージ

X2	x2
50	40
60	70
80	50
30	40
70	80

- ・関数date.frame()を使って、作成済みのベクトルx1とx2からデータフレームを作成する

```
> date01 <- date.frame(x1, x2)
```

データフレームの作成

～データフレームの作成1～

- ・作成したデータフレームdate01を表示する

```
> date01
  x1x2
1  50 40
2  60 70
3  80 50
4  30 40
5  70 80
```

- ・変数名を変更する

```
> date01 <- date.frame(英語=x1, 数学=x2)
> date01
  英語 数学
1  50  40
2  60  70
3  80  50
4  30  40
5  70  80
```

データフレームの作成 ~データフレームの作成2~

Rでは作業ディレクトリという場所を対象に、データファイルの読み込みや書き出しを行う。今回は事前に作成しておいたCドライブ内のRフォルダを作業ディレクトリとして設定する。

- ・作業ディレクトリを設定するには関数setwd()を使う

```
> setwd( "C:/R" )
```

- ・現在の作業ディレクトリの場所を確認するにはgetwd()を使う

```
> getwd()  
[1] "C:/R"
```

データフレームの作成 ～データフレームの作成2～

- ・カンマ区切りファイルを読み込む

例6: 変数名のないデータを読み込む

変数名が記録されていないdate02.csvを読み込む。
カンマ区切りファイルを読み込むには関数read.csv()を使う。

50, 40, 70
60, 70, 60
80, 50, 90

date02.csv

```
> date02<- read.csv("date02.csv", header=F)
> date02
  V1 V2 V3
1 50 40 70
2 60 70 60
3 80 50 90
```

※命令文中の「header=F」は、データ表の1行目に変数名の記録がないことを示している。1行目に変数名の記録がある場合は、「header=T」と入力するか、「header」に関する指示を記載しない。

また、コンソール内に表示されているデータフレームdate02の変数名は、Rが自動的に作成したものである。

データフレームの作成 ~データフレームの作成2~

- xlsxファイルの読み込み

例7: Excelに入力したデータの読み込み

20名の学生に対してアンケートを実施した。その結果を保存しているファイルdate03.xlsx(シート名は「Sheet1」)を読み込む。ファイルの1行目には変数名が記録されている。

```
> library(xlsx)
```

```
> date03<- read.xlsx("date03.xlsx", sheetName="Sheet1")
```

※パッケージ内の関数を使う場合、はじめに命令文library(xlsx)を使ってxlsxパッケージの仕様準備を行う。

その後、関数read.xlsx()を使って、作業ディレクトリ内のdate03.xlsxファイルのSheet1シートを、date03という名前のデータフレームとして取り込むことができる。

※関数read.xlsxを使う場合、シート名、変数名は全て半角英数字を使う。ファイル名とデータフレーム名には全角も使えるが、コンソールでは半角英数字を使うほうが操作しやすい。

データフレームの作成

～データフレームの作成2～

- ・データフレーム内の変数の基本的な統計量を計算する

例8: データフレーム内の平均の算出

データフレームdate03内の変数q3の平均値を求める。
平均値を求めるには関数mean()を使う。

```
> mean(date03$q3)
```

例9: データフレーム内の要約統計量の算出

データフレームdate03内の全ての変数の要約統計量を求める。
要約統計量を求めるには関数summary()を使う。

```
> summary(date03)
```

まとめ

- ・Rとは統計計算を行うために開発された無料のソフトウェア
- ・Rのインストール方法
- ・Rを使った、式の代入方法、データの入力方法、ファイルの読み込み方法