

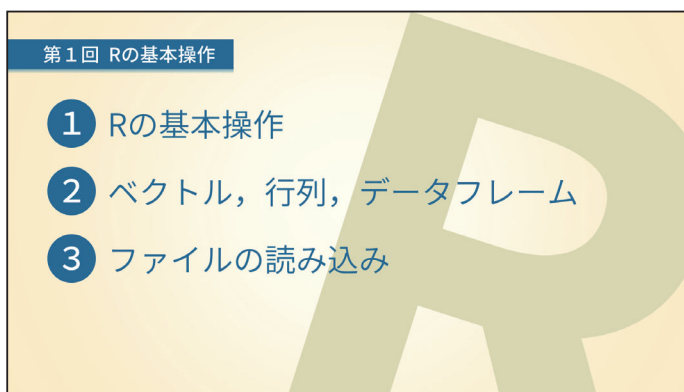
Rで学ぶ確率統計

第1回 Rの基本操作

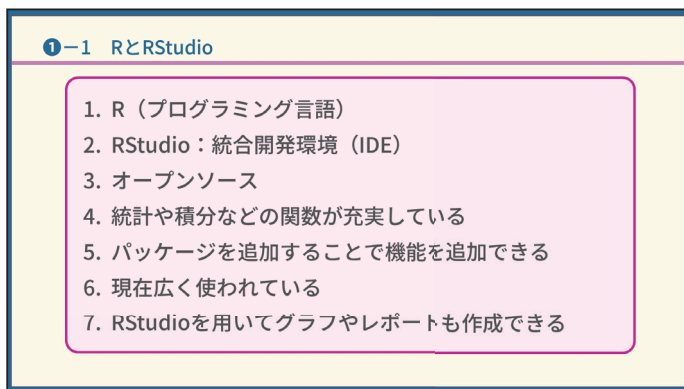
1



2



3



4



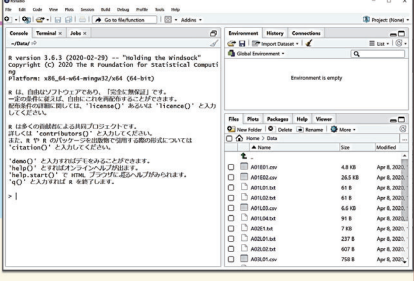
R で学ぶ確率統計

第 1 回 R の基本操作

5

①-3 Rの基本操作

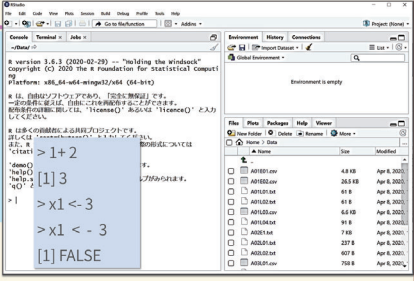
1. 電卓として利用する
2. 変数を代入する
3. 関数を利用する
4. ファイルからデータを読み込む
5. 結果を変数に代入する
6. 結果を理解する
7. 表やグラフを作成する
8. レポートを作成する



6

①-3 Rの基本操作

1. 電卓として利用する
2. 変数を代入する
3. 関数を利用する
4. ファイルからデータを読み込む
5. 結果を変数に代入する
6. 結果を理解する
7. 表やグラフを作成する
8. レポートを作成する



塊ごとでスペース
2文字で1セットも

7

①-4 関数を利用する(1)

```
> x1 <- c(118,119,121,122,170)
> mean(x1)
[1] 130
> mean(c(118,119,121,122,170))
[1] 130
> var(x1)
[1] 502.5
```

c () はベクトルを作る関数
mean、varなどの統計関数が豊富
function(引数)
様々な関数があり、
引数やオプション、出力を覚える
?関数 でヘルプが表示される

8

①-4 関数を利用する(2)

演算子			関数		
演算子	説明	例と意味	関数	説明	例と意味
+	足し算	a + b	c()	連結	c(1,2,3) concatenate()
-	引き算	a - b	count()	個数	ベクトルの要素数
*	掛け算	a * b	abs()	絶対値	abs(a)
/	割り算	a / b	sqrt()	平方根	sqrt(a) は $\sqrt{a}$
%%	余り	a %% b	exp()	指数関数	e のべき乗 e は exp(1)
^	べき	a ^ b (a ^b)	log()	自然対数	log(a) 常用対数は log10()

R で学ぶ確率統計

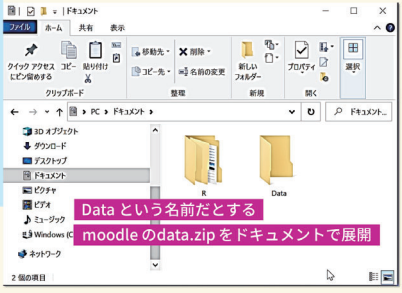
第 1 回 R の基本操作

9

①-5 RStudioの基本設定 (1)

自分のコンピュータ上で作業用のフォルダを作成

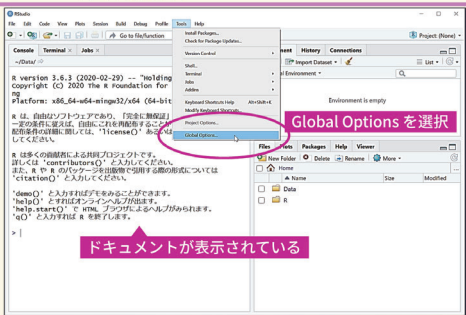
1. 電卓として利用する
2. 変数を代入する
3. 関数を利用する
4. ファイルからデータを読み込む
5. 結果を変数に代入する
6. 結果を理解する
7. 表やグラフを作成する
8. レポートを作成する



Data という名前だとする
moodle の data.zip をドキュメントで展開

10

①-5 RStudioの基本設定 (2)

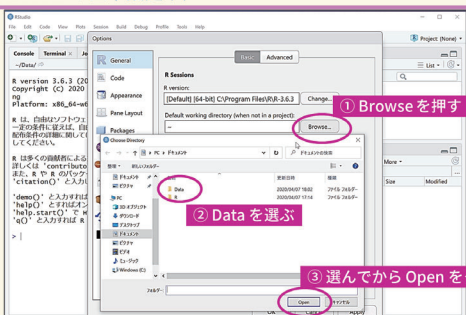


Global Options を選択

ドキュメントが表示されている

11

①-5 RStudioの基本設定 (3)



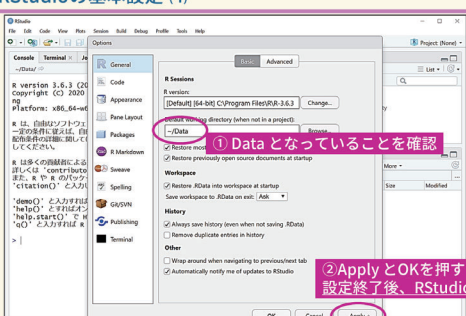
① Browse を押す

② Data を選ぶ

③ 選んでから Open をクリック

12

①-5 RStudioの基本設定 (4)



① Data になっていることを確認

② Apply と OK を押す
設定終了後、RStudio を再起動

R で学ぶ確率統計

第 1 回 R の基本操作

13

①-5 RStudioの基本設定 (5)

Data となっていることを確認

作業フォルダ
ファイルを探しに行くときの起点

14

②-1 Rにおけるデータの型と形式

Rにおける主なデータの型

- 論理値 TRUE, FALSE
- 整数 0 1 2 ...
- 高精度浮動小数点数
- 数値 区切り文字
- 文字列 a b c ...
- 日付
- 日付時刻
- ファクター
(カテゴリーデータ)

データ形式

- スカラー (1つの値)
- ベクトル (複数)
- 行列
- リスト
- データ・フレーム
- tibble

15

②-2 リスト

```
> t1 <- list("英語", 35, c(70, 60, 40, 30))
> names(t1) <- c("科目", "年齢", "点数")
> t1
$科目
[1] 英語
$年齢
[1] 35
$点数
[1] 70 60 40 30
```

リスト

list(要素1, 要素2, 要素3)

- 異なる要素の集まり
- t1[[1]] # 1番目の要素
- t1\$科目

16

②-3 データフレーム

```
> h <- c(148, 152, 154, 158)
> w <- c(52, 54, 56, 58)
> D <- data.frame(h, w)
> D
  h w
1 148 52
2 152 54
3 154 56
4 158 58
> names(D)
[1] "h" "w"
```

データフレーム

data.frame(要素1, 要素2, 要素3)

- 同じ長さ、型の異なるベクトル
- D[1, 1:3]
- D\$h

D[行番号, 列番号]
1:3 は c(1, 2, 3)

R で学ぶ確率統計

第 1 回 R の基本操作

17

③-1 パッケージの利用 (tidyverse)

```
> install.packages("tidyverse")
> install.packages("knitr")
```

```
> library(tidyverse)
```

tidyverse (複数のパッケージ)

- Hadley Wickham
- データ操作 dplyr
- グラフ作成 ggplot2 など

パッケージ

- 便利な機能や解析のための関数など
- インターネットに接続している環境で
- # 利用するときは library() とする

knitr

- Rでレポートを作成するときに利用

18

③-2 tibble (データフレームの改良版: tidyverse)

```
> h <- c(148,152,154,158)
```

```
> w <- c(52,54,56,58)
```

```
> df <- tibble(h, w)
```

```
> df
```

A tibble 4x2

```
h w
<dbl> <dbl>
```

1 148 52

2 152 54

3 154 56

4 158 58

tibble

tibble(要素1,要素2,要素3)

- 列ごとに型が決まっている
- 列名を作らない
- 文字列がファクターにならない
- df\$h, df[["h"]], df[[1]]

19

③-3 データの分析

- ファイルの読み込み
- 表形式のデータ
- グラフの作成
- データの要約
- 数理モデルや多変量解析などによる分析
- 出力

20

③-4 ファイルの読み込み (tidyverse) (1)

```
> library(tidyverse)
```

```
> a01 <- read_table2("A01L01.txt")
```

```
> a02 <- read_csv("A01L02.csv")
```

```
> a03 <- read_tsv("A01L03.tsv")
```

tibble (データフレームの 1 種)

read.csv という関数もある

ファイルの区切り (delim) が何か

table2 スペース区切り

tsv タブ区切り

csv カンマ区切り

1 行目は行の説明

ない場合 col_names=FALSE

name1 A B

A1 148 52

A2 152 54

A3 154 56

A4 158 58

A5 163 60

name1,C,D

B1,148,52

B2,152,54

B3,154,56

B4,158,58

B5,163,60

R で学ぶ確率統計

第 1 回 R の基本操作

21

③-4 ファイルの読み込み (tidyverse) (2)

```
>?read.csv
>?read_csv
```

```
read.table{util}
read_delim{readr}
```

util というパッケージ

```
# read.csv, read.table は R の基本的な関数
# tidyverse 中の readr というパッケージがある
# read_csv, read_table2 などはその中の関数
# read_csv などの方が高速
```

22

③-5 RStudio の画面

⓪ でファイルを新規作成

History 過去の実行記録

Environment 読み込んだ変数

Plots 図の表示

Files ファイル一覧表示 (エクスプローラー)

コマンドの実行タイピング

Run 行を選んで実行

データをインポート

ファイルを編集 (エディター)

23

③-6 Rスクリプト

- 1 つずつ命令を実行
- 結果を確認しながら分析可能

- 不要な変数、サイズの巨大な変数の増加
- やり直して上書きし、必要な命令が不明に
- 分析が複雑になると、命令も増加
- 同一の変数名だと処理手順が不明に

- 分析では再現可能であることが大切
- 一連の手順をスクリプトとして保存
- 履歴から選ぶことで作成可能
- 振り返りの材料になる
- 他人への可読性を高める工夫も
- グラフ含めたレポートのパーツ化

24

この講義を進めるにあたって

- 講義を聞きながらRStudioを動かす
- コマンドを実行して確認しながら学ぶ
- スクリプトにコメントを追記する
- 関数の理解のため英語の意味を調べる
- ファイル名は半角英数字を推奨

```
() や 42.7 など
セットになって用いられる
揃っていないと命令が
終わっていないと解釈され
> read_csv(" a.csv) # "がないと
+
となるが、困ったら Esc を押す
```