
Rでのデータの操作

データの読み込み（インポート）

(1) データセットの利用

Rには、さまざまな「データセット」とよばれるサンプルデータが提供されている。

【実習1】 ①データセットの読み込み

```
> WorldPhones
```

```
> class(WorldPhones)
```

```
[1] "matrix"
```

```
# データの型 (class())で確認
```

```
> wp <- data.frame(WorldPhones)
```

```
# データフレーム形式に変換
```

【練習1】 Rで提供されている「データセット」を確認しよう。

データの読み込み（インポート）

(2) データの読み込み（※ディレクトリについて）

ここでは、主に**CSV(Comma Separated Values)**ファイルの読み込みについて説明する。

(注) ディレクトリについて

- ・ Windowsではメニュー「ファイル」→「ディレクトリの変更」で指定できる
- ・ Macではメニュー「その他」→「作業ディレクトリの変更」

実際に、それぞれの環境の中でデータを読み込むときは、ディレクトリの理解が不可欠になる。
例えば、**JドライブのRフォルダ**にある**"sheet1.csv"**を読み込むときは次のようにする
（学院のPC教室ではJドライブに資料を格納しておく）。

- | | |
|---------------|--|
| ・ JドライブのRフォルダ | <code>data1 <- read.csv("J:/R/sheet1.csv")</code> |
| ・ 同一階層の場合 | <code>data1 <- read.csv("sheet1.csv")</code> |

データの読み込み（インポート）

【実習1】 ②データの読み込み

```
data1 <- read.csv("J:/R/sheet1.csv", header=TRUE)
```

header=TRUE (最初の行にラベル有), FALSE (ラベル無)

	gender	region	name	Japanese	Math	English
1	1	tokyo	aoki	23	56	65
2	2	tokyo	ehara	43	45	44
3	2	saitama	kato	11	97	76
4	1	chiba	kayama	56	32	23
5	2	chiba	kinoshita	89	41	54

```
> class(data1)
```

```
[1] "data.frame"
```

データの型の確認

データの加工

【実習2】 ①データ構造の確認（変数の種類を確認する）

```
> str(data1)
```

データの構造

```
> dim(data1)
```

データの配列

```
[1] 30 7
```

②抽出

```
> data1[1,2] # 1行2列目の抽出
[1] Tokyo

> data1[1,] data1[1,1:6]と同じ # 1行目の抽出
sex region name Japanese Math English
1 tokyo aoki 23 56 65

> data1[1:3,] data1[1:3,1:6]と同じ # 1~3行目の抽出
sex region name Japanese Math English
1 tokyo aoki 23 56 65
2 tokyo ehara 43 45 44
2 saitama kato 11 97 76

> data1[,5] data1[1:30,5]と同じ # 5列目の抽出
[1] 56 45 97 32 41 78 45 55 (略) 86 76 54 67 76 91 55 69 48 83 81 79 89

> data1$Math # $Math変数へアクセス
[1] 56 45 97 32 41 78 45 55 (略) 86 76 54 67 76 91 55 69 48 83 81 79 89

> data1[,4] + data1[,5] # 演算 (4列目+5列目)
```

③並べ替え

```
> sort(data1$Math)                                # Math列の並べ替え(昇順)
```

```
[1] 12 31 32 34 41 44 45 45 45 48 54 (略) 78 78 79 79 81 83 86 89 91 97
```

```
> sort(data1$Math,decreasing=TRUE)                # Math列の並べ替え(降順)
```

```
[1] 97 91 89 86 83 81 79 79 78 78 76 (略) 48 45 45 45 44 41 34 32 31 12
```

④部分和 **apply**関数：(行・列を一括して計算するには**apply**関数が便利)

```
> sum(data1[1, 4:6]) # 1行目の合計 (個人の三教科合計)
```

```
[1] 144
```

```
> sum(data1$Math) # 各列 (Math列) の合計 (各教科の合計)
```

```
[1] 1861
```

行・列を一括して計算するには**apply**関数が便利である

```
# apply (data, 1:行・2:列, 関数名)
```

```
# data:行列orデータフレームに適用
```

```
> apply(data1[,4:6], 1, sum) # 各行の合計 (個人の三教科合計)
```

```
[1] 144 132 184 111 184 215 125 131 176 108 169 181 253 154 204 187 180 165 182 178 145 188 265 167 188 164  
189 203 223 193
```

```
> apply(data1[,4:6], 2, sum) # 各列の合計 (各教科の合計)
```

```
Japanese Math English
```

```
1681 1861 1746
```

⑤表の再構成（**data2**をつくる）

```
> mark <- apply(data1[,4 : 6], 1, sum)
```

#合計をmarkオブジェクト

```
> data2 <- data.frame(data1, mark)
```

data1とmarkを結合

```
> data2
```

	gender	region	name	Japanese	Math	English	mark
1	1	tokyo	aoki	23	56	65	144
2	2	tokyo	ehara	43	45	44	132
3	2	saitama	kato	11	97	76	184
4	1	chiba	kayama	56	32	23	111