

# 第1章 データマイニングとは

## 1.7 Rことはじめ

06T4029T 齊藤優

# Rの紹介

- Rはフリーの統計解析向けの言語及び環境
- 統計パッケージを作成できるなど拡張性・汎用性が高い

# Rにおけるデータの入出力

- csv形式のファイルをデータ入出力に使用
- データ入力関数 `read.csv()` を利用  
「陸上データ.csv」のデータを入力

```
陸上<-read.csv("陸上データ.csv")
```

- データ出力関数 `write.csv()` を利用  
「陸上」というオブジェクトを「出力.csv」というファイルに出力

```
write.csv(陸上,"出力.csv")
```

# データフレーム

- 先ほどの「陸上」というオブジェクト データフレーム
- データフレームは変数の種類が識別されたデータセット
- データフレームを利用 R内部でのデータの扱いが簡単

# 基礎統計量

- 基礎統計量は標本の分布の特徴を代表的に表す統計学上の値
- 基礎統計量の算出は関数 `summary()` を利用  
「陸上」オブジェクトの基礎統計量を算出

```
summary(陸上)
```

- 群ごとに要約統計量を算出するためには関数 `by()` を利用  
「区分」という変数の値で分けて要約統計量を算出

```
by(陸上, 陸上$区分, summary)
```

# グラフィックス

- Rはグラフィックス機能も充実  
記録の時系列プロットとヒストグラムの描画を実行

```
plot(陸上$開催年, 陸上$高跳び)  
hist(陸上$走り幅跳び)
```

# 組み込みデータセット

- 組み込みデータセット

- すでに内部で定義されたデータセット  
外部から読み込む必要なしで使える
- Rはこの組み込みデータセットが充実
- 統計関数の性能評価や比較・改良に使える

- 利用法

- データフレームとしてすでに定義されているので、データセット名をそのまま関数などに利用可能

# パッケージのインストール

- パッケージとは関数やデータセットなどをひとまとめにした関数ライブラリ・データベース
- R内部で定義されていない関数を使う場合  
関数が含まれるパッケージをダウンロードする必要がある
- ダウンロードしてRのライブラリに登録されたパッケージは  
`library(パッケージ名)`でRに読み込んで利用可能となる

# ヘルプの参照法

- 関数 `help()` でパッケージや関数に付属しているドキュメントを確認できる

```
help(package="パッケージ名")  
help(関数名)
```

- `help.search("キーワード")` でキーワードが記載されているヘルプの一覧が表示