2016/6/17 Mon.

### 問題解決技法入門

# **Cluster analysis using R**



Rでクラスター分析

 Rを起動, csv ファイルをデータとして読込み –「マイドキュメント(Y:)」の「R」フォルダに保存

		算数	理科	国語	英語	社会
	太郎	90	100	70	90	30
	次郎	80	60	70	70	20
data-seiseki.csv	三郎	100	40	30	70	80
	四郎	60	30	40	80	80
	花子	30	60	80	90	90
	寒子	50	60	40	30	60
	湘子	90	100	90	80	70

※ファイルのフルパス 例)YドライブのRフォルダ内にあるdata-seiseki.csvという名前のファイル

• R Console

※1行目にheaderあり ※各行1列目は名前

> seiseki <- read.csv("Y:/R/data-seiseki.csv", header=T, row.names=1)</p>

※csvファイルを読み込み, 変数seiseki に代入

Rでクラスター分析

> seiseki.d <- dist(seiseki, "manhattan")</pre>

※マンハッタン距離("manhattan")を用いて距離を計算している 他の距離を使いたいとき "euclidean" =ユークリッド距離 は"manhattan"を右に変更 "minkowski", p=3 = p=3のミンコフスキー距離

"maximum" =  $l_{\infty} / \mu \Delta$  ( $z_{\delta}$  tribution of  $\delta z_{\delta}$ )

### 階層クラスター分析をし、結果をseiseki.hcに代入

> (seiseki.hc <- hclust(seiseki.d, "ward"))</pre>

※ウォード法("ward")を用いてクラスター分析を実施している 他の方法を使いたいときは、"ward"を以下に変更

"single"=最短距離法, "complete"=最長距離法 "average"=群平均法, "centroid"=重心法,

"median"=中央值法

#### • 結果をデンドログラム(樹形図)で描画

> plot(seiseki.hc, hang=-1)

## 参考文献

- ◆ 田中豊・脇本和昌『多変量統計解析法』現代数学社(1983)
- 河口至商『多変量解析入門Ⅱ』森北出版(1978,2005)
- ◆ 青木繁伸『Rによる統計解析』オーム社(2009)
- ◆ 荒木孝治 『RとRコマンダーではじめる多変量解析』 日科技連(2007)
- ◆ 金明哲 『Rによるデータサイエンス』 森北出版(2007)
- 新納浩幸『Rで学ぶクラスタ解析』オーム社(2007)

# もっと知りたい人へ

- 関連する経営学科の授業
  - 「統計の見方」(1/2セメ)
  - 「統計の分析と利用」(2セメ)
  - 「統計データの扱い方」(3/4セメ)
  - 「**多変量の統計データ解析**」(4セメ)