多次元尺度構成法
「MDSを使って使って使い倒す！
MDS入門から非対称MDS実習まで」

2010年3月27日～28日
日本行動計量学会 第13回春の合宿セミナー A2コース

中山 厚穂（長崎大学）
横山 暁（慶應義塾大学）

3月28日（日）非対称多次元尺度構成法

- 午前　10:00～12:00　非対称MDSとは何か
  - 非対称MDSについての説明
    - 個人差モデルについても簡単に説明
    - 時間があれば，「R」による分析方法の紹介

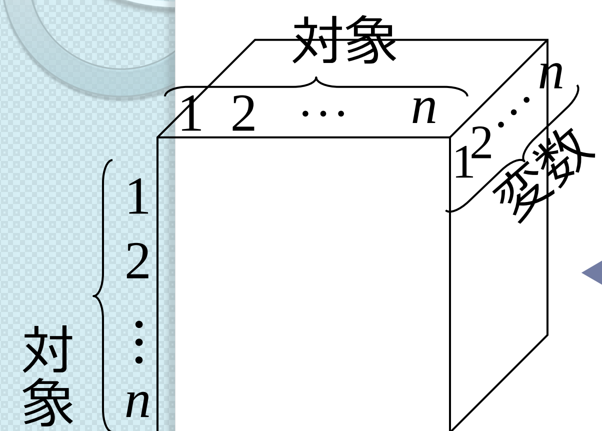
- 午後　13:00～15:00　非対称MDS　事例による実習
  - 非対称MDSの「R」による分析方法の紹介
  - 2日間のQ&A

3 月28日（日）午前(10:00～12:00)

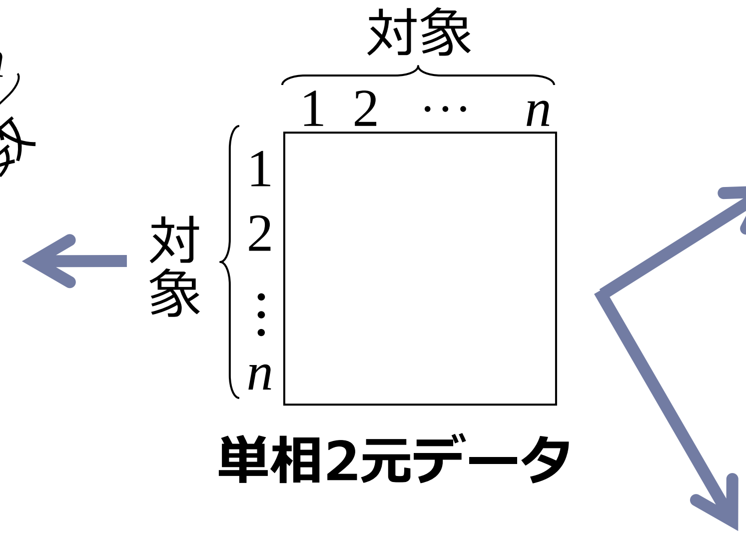
非対称MDSとは

岡太彬訓・守口剛『マーケティングのデータ分析』朝倉書店(近日刊行予定)

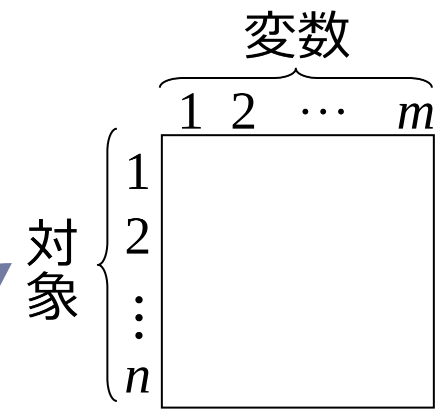
MDSにおいて用いられるデータ
相(mode)と元(way)という概念により整理



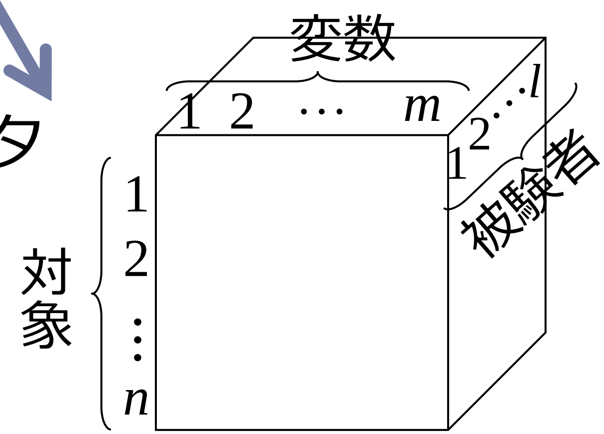
2相3元データ



単相2元データ



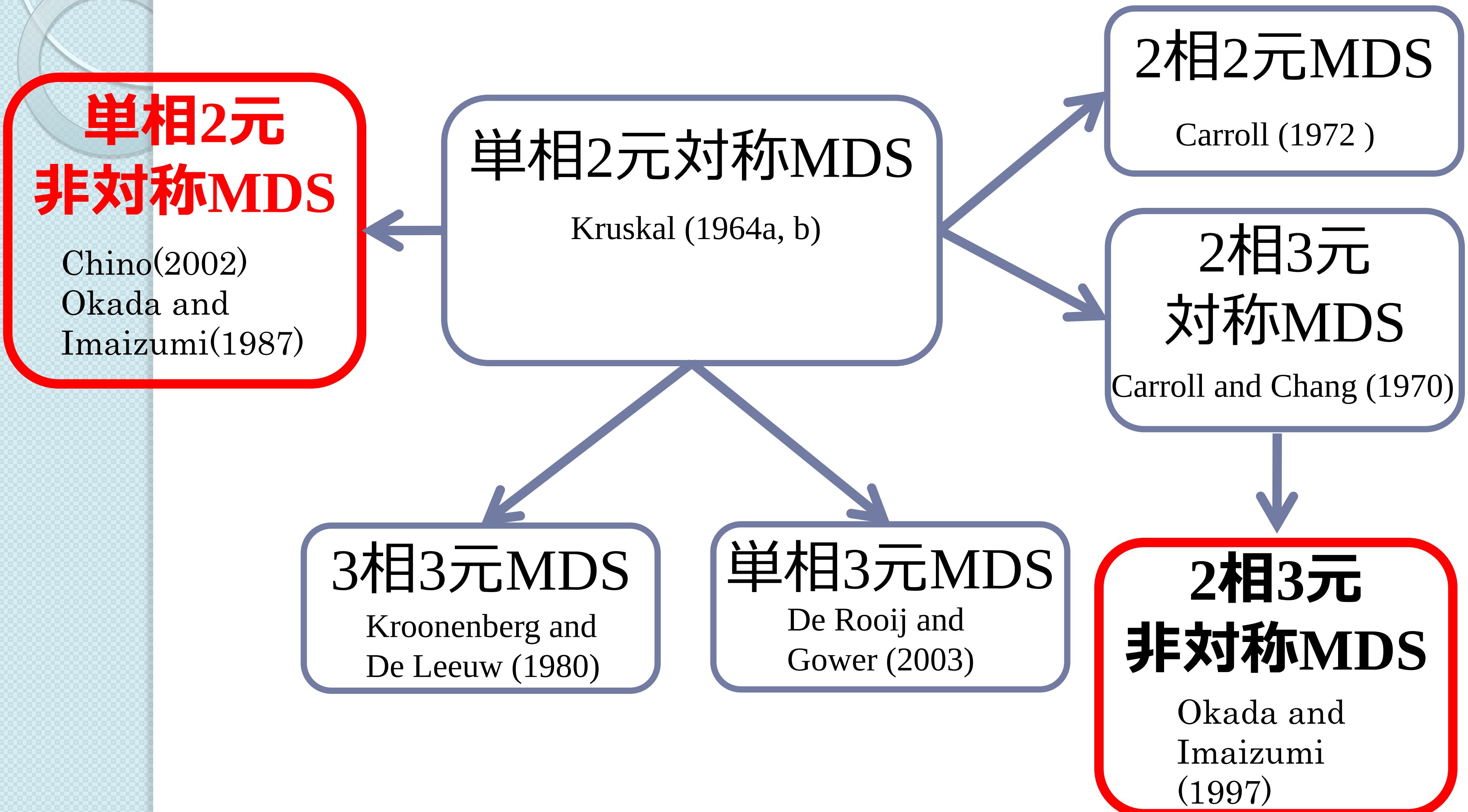
2相2元データ



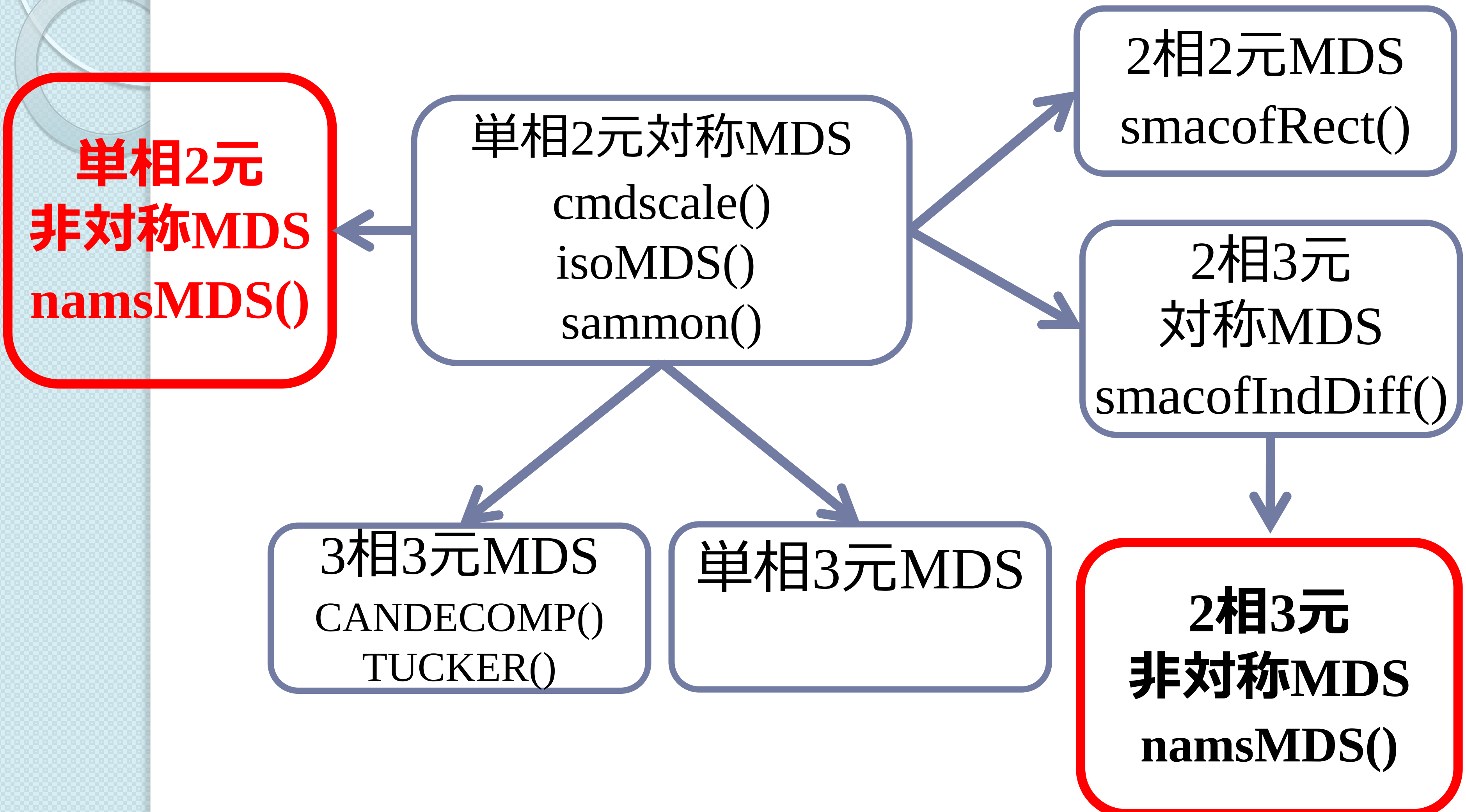
3相3元データ

- M個の相とN個の元をもつデータ
  - M相N元データ (M≦N)
- データの相
  - 1組の対象を意味
- 元の数
  - 相がいくつ組合されているかで決定

データとの対応関係


単相2元
非対称MDS
Chino(2002)
Okada and
Imaizumi(1987)
単相2元対称MDS
Kruskal (1964a, b)
2相2元MDS
Carroll (1972)
2相3元
対称MDS
Carroll and Chang (1970)
3相3元MDS
Kroonenberg and
De Leeuw (1980)
単相3元MDS
De Rooij and
Gower (2003)
2相3元
非対称MDS
Okada and
Imaizumi
(1997)

Rでの分析可能モデル


単相2元
非対称MDS
namsMDS()
単相2元対称MDS
cmdscale()
isoMDS()
sammon()
2相2元MDS
smacofRect()
2相3元
対称MDS
smacofIndDiff()
3相3元MDS
CANDECOMP()
TUCKER()
単相3元MDS
2相3元
非対称MDS
namsMDS()

非対称データ

- 非対称データ
  - データの各行・列の総和が外的影響で異なる(データの上三角部分と下三角部分の値が異なる)
  - マーケティングでは分析のニーズが高い
    - ブランドスイッチングデータや複数商品の同時購買確率
- 非対称データの分析：多元・多相データへの拡張
  - 解析前に何らかの基準化(行列の再構成)を行う
    - Harshman,Green, Wind, and Lundy(1982)
  - 非対称モデルを活用
    - Okada and Imaizumi(1987,1997)
    - De Rooij & Heiser(2003)

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & \ddots & & x_{2n} \\ \vdots & & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{bmatrix}$$

非対称MDSとは

- 類似度の非対称性を無視せずに分析し，非対称性をわかりやすく表現
  - 類似度が非対称

    ×対象j対象とkの間の類似度という表現

    ○「対象iから対象jへの類似度s_{ij}」と「対象jから対象iへの類似度s_{ji}」
  - 類似度が非対称であるというのは

    類似度s_{ij}≠の類似度s_{ji}

非対称MDSとは
例えば・・・

- ブランドスイッチ頻度をブランド間の類似度と考える
  - 2つのブランドiとjの間の類似度

    ブランドiからブランドjへのブランド変更

    ブランドjからブランドiへのブランド変更
    - 2つの変更の頻度の値は(通常は)異なり非対称となる
      - ブランド変更の非対称性はブランド間競争力の優劣や魅力の大小を表現

  - ブランドiがブランドjよりも好まれるならば，以下のような関係が成立

    ブランドiからjへのブランド変更の頻度

    < ブランドjからiへの変の更頻度

  - 類似度の非対称性を無視せず分析し非対称性を表現

非対称MDS

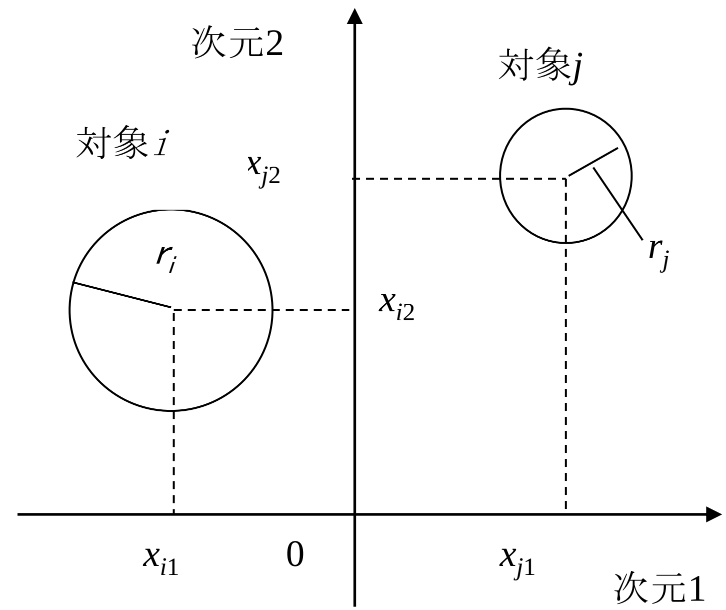
- 類似度のもつ非対称性を直接的に表現するための多次元尺度構成法
  - 1970年代後半から研究が開始され多くのモデルが開発
    - Borg and Groenen (2005) pp. 495-581
    - Bove and Rocci (1999)
    - 千野(1997)
    - 千野・岡太(1996)
    - Cox and Cox (2001), pp. 116-121, 243-244
    - Zielamn and Heiser(1996)
- 今回は，Okada and Imaizumi (1987) の非対称多次元尺度構成法に沿って解説
  - 多次元空間における距離によって非対称な類似度関を表現
  - したがって，布置の解釈が容易であると考えられる(Okada, 1988; 岡太, 1989; 岡太・元治, 1995; 岡太・今泉, 1996; Okada and Imaizumi, 2003)

Okada and Imaizumi (1987)のモデル

- 単相2元類似度データを分析
  - ブランドスイッチングや複数商品同時購買確率のデータ
  - 非対称なブランド変更行列などは，2相2元類似度と考え多次元展開法で分析も可能
- 非対称類似度データでは，「対象iから対象jへの類似度s_{ij}」と「対象jから対象iへの類似度s_{ji}」は必ずしも一致しない
  - 対象が同一(ブランド変更前後でブランドが変わらない)の場合の類似度s_{ii}(i=1, 2, …, M)は分析の対象としない
  - M個の対象からなる非類似度データの主対角要素(M個)を除いたM(M-1)個を分析

Okada and Imaizumi (1987)のモデル

- 類似度の大小だけを用いるKruskalの方法を発展
  - 各対象は多次元空間に点とその点に中心をもつ円(2次元空間)，球(3次元空間)，超球(4次元以上の空間)で表現
  - R次元空間で対象i
    - 点 $(x_{i1}, x_{i2}, \ldots, x_{it}, \ldots, x_{iR})$ と半径r_iの円(球，超球)により表現
    - x_{it}は対象iを表現する点の次元tの座標



Okada and Imaizumi (1987)のモデル

- Kruskalの方法
  - 対象iとjの間の類似度s_{ij}に距離d_{ij}を対応
- Okada and Imaizumi (1987)のモデル
  - 類似度の非対称性を以下のように対応付けることで表現
  - 対象iとjの間の類似度s_{ij}にm_{ij}を対応

$$m_{ij} = d_{ij} - r_i + r_j$$

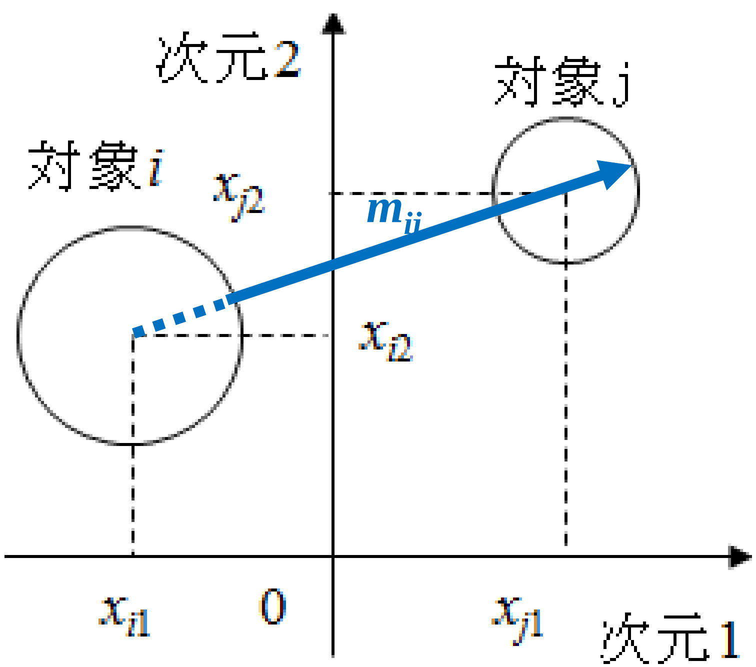
  - 対象jとiの間の類似度s_{ji}にm_{ji}を対応

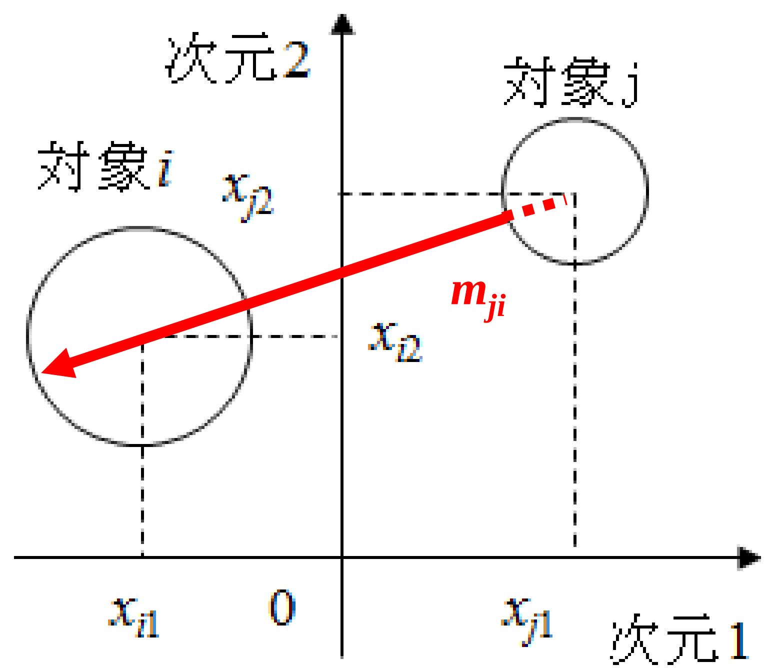
$$m_{ji} = d_{ij} - r_j + r_i$$

  - d_{ij}は対象iと対象jを表す点間の(ユークリッド)距離
  - m_{ij}とm_{ji}ではr_iとr_jが入れ替わる(一般的には$m_{ij} \neq m_{ji}$)

Okada and Imaizumi (1987)のモデル 2次元布置に表現された2つの対象*i*と対象*j*

- (a)対象*i*から*j*への類似度に対応するm_{ij}
- (b)対象*j*から*i*への類似度に対応するm_{ji}

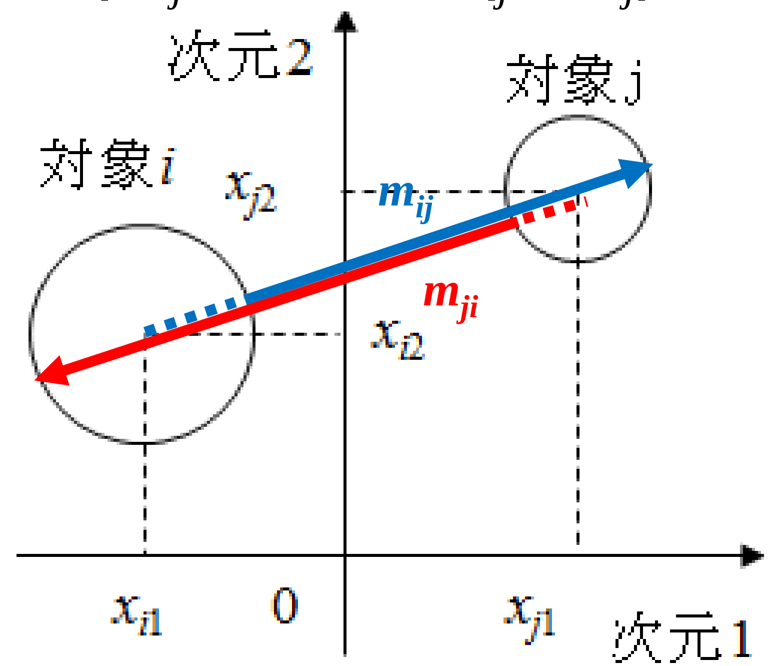




Okada and Imaizumi (1987)のモデル

- m_{ij}とm_{ji}の大小はr_jとr_kの相対的な大きさから決定
  - $r_i<r_j$であれば$m_{ij}>m_{ji}$
  - $r_i=r_j$であれば$m_{ij}=m_{ji}$
  - $r_i>r_j$であれば$m_{ij}<m_{ji}$

$r_i>r_j$であり，$m_{ij} < m_{ji}$



より小さい半径をもつ対象からより大きい半径をもつ対象への類似度に対応するm_{ij}
＞より大きい半径をもつ対象からより小さい半径をもつ対象への類似度に対応するm_{ji}

Okada and Imaizumi (1987)のモデル

- 例えば，ブランドスイッチを例に考えると

  「半径がより小さいブランドからより大きいブランドへの変更」は

  「半径がより大きいブランドからより小さいブランドへの変更」より起きにくい

  - 半径r_iが小さいほどブランドiはブランド変更において優位
  - 半径がブランド間の競争力の優劣や魅力の大小を表す

Okada and Imaizumi (1987)のモデル

- Kruskalの多次元尺度構成法
  - 対象を多次元空間に点として表現
  - 布置での点間距離が対象間間の類似度と単調減少になるように布置を求める

- Okada and Imaizumi (1987)のモデル
  - 点間距離d_{ij}ではなくm_{ij}を用いる
  - 対象iからjへの類似度s_{ij}と対象rからsへの類似度s_{rs}に対し

$$s_{ij} > s_{rs} \quad \Rightarrow \quad m_{ij} \le m_{rs}$$

となるように，これらの点の座標と半径を求め，布置を得る

Okada and Imaizumi (1987)のモデル

- 散布図やストレスはKruskalの方法と同様に考えられる
- Kruskalの方法
  - $M(M\text{-}1)/2$個の類似度と対応する点間距離を考える
- Okada and Imaizumi (1987)のモデル
  - $M(M\text{-}1)$個の類似度と対応するm_{ij}を考える
  - 散布図は，(m_{ij}, s_{ij})で表される$M(M\text{-}1)$個の点からなる

Okada and Imaizumi (1987)のモデル

- 距離との単調減少関係を満たすディスパリティーに対応する $\hat{m}_{ij}$ は，Kruskalの方法と同じ考え方と手順で定義され

$$s_{ij} > s_{rs} \quad \Rightarrow \quad \hat{m}_{ij} \le \hat{m}_{rs}$$

満たす

- ストレス*S*は以下のように*M*(*M*-1)個の組合わせについて定義

$$S = \sqrt{\sum_{j=1}^{M} \sum_{\substack{k=1 \\ j \ne k}}^{M} (m_{jk} - \hat{m}_{jk})^2 \Bigg/ \sum_{j=1}^{M} \sum_{\substack{k=1 \\ j \ne k}}^{M} (m_{jk} - \overline{m})^2}$$

- ただし， $\overline{m} = \sum_{j=1}^{M} \sum_{\substack{k=1 \\ j \ne k}}^{M} m_{jk} \Big/ M(M-1)$
- ストレスを最小化するためのアルゴリズムはKruskalの方法と同様であり，ストレスを最小化する点の座標と半径を求める

Okada and Imaizumi (1987)のモデル

- 分析の手順は，Kruskalの方法の手順と同様
  - 退化については対象を表現する点だけではなく半径もあるため，点間距離と半径を併せて検討
    - 散布図などを吟味して，結果が退化しているかどうかを判断
  - 点間距離や半径は，次元の回転では変化しないため，得られた布置の次元は回転(直交回転)できる
    - 対象を表現する点の位置や半径を考え，布置を解釈しやすい方向に次元を回転する
- このあと「R」による分析方法の紹介について詳しく説明

非対称モデルを用いないで分析するには…

非対称モデルを用いないで分析

- 非対称性の情報は失われる
  - 対応する2つの類似度の平均をとる
  - 2つの類似度に1つの距離を対応させる
- 非対称性を表現可能
  - 行間類似度と列間類似度を別個に分析
  - 2相2元類似度として分析

対応する2つの類似度の平均をとる

- 主対角線を挟んで対称な位置にある2つの要素の平均を求めて対称化

  対象iとj(対象jとi)間の類似度$=(s_{ij}+s_{ji})/2$

  - 類似度の平均には非対称性の情報は含まれない
  - 非対称性を考慮していない

- 例えば，ブランドスイッチのデータ行列
  - ブランドiからjへのブランド変更の頻度である第(i,j)要素s_{ij}と，ブランドjからiへのブランド変更の頻度である第(j,i)要素s_{ij}の平均を計算
  - この平均値をブランドiとj(jとi)の間の類似度とする

    ブランドiとj(ブランドjとi)間の類似度$=(s_{ij}+s_{ji})/2$

2つの類似度に1つの距離を対応させる

- 対象iとj間の類似度をs_{ij}と対象jとi間の類似度をs_{ji}という必ずしも等しくない2つの類似度s_{ij}とs_{ji}に同一の距離$d_{ij}(=d_{ji})$を対応させる
  - 2つの類似度s_{ij}とs_{ji}のもつ非対称性の情報は失われる
- 2つの相異なる類似度s_{ij}とs_{ji}に同じ距離d_{ij}を対応させるため，2つの類似度の差(非対称性)が大きくなれば，距離の類似度への適合度は低下
  - 小さい適合度が大きい非対称性の可能性を示すが，類似度関係における非対称性は具体的には表されない

行間類似度と列間類似度を別個に分析

- データ行列の行間の類似度と列間の類似度を求め，別個に分析して2つの布置を求める
  - 2つの布置は，行間の類似度と列間の類似度の違いを表し，類似度の非対称性を表現
  - 2つの布置から類似度の非対称性を把握するのは困難
- 例えば，ブランドスイッチのデータ行列
  - 各行はどのブランドから変更されたのか(ブランド変更元)を示している
  - 各列はどのブランドへ変更されたのか(ブランド変更先)を示している
  - ブランド変更行列の行間と列間の類似度をそれぞれ求め，個別に分析して2つの布置を求める
    - 2つの布置によりブランド変更元と変更先の違いが明らかとなる

2相2元類似度として分析

- 行の対応する対象と列の対応する対象を別個の対象と考えて2相2元類似度データとして分析
  - 行の対応する対象を表現する点と列の対応する対象を表現する点とが同じ布置に表現(点の数は2倍)
  - 類似度の非対称性が表現できる
  - 同一の対象が2つの点で表されるため類似度の非対称性を把握するのは必ずしも容易ではない
  - 分析では退化の問題がある
- 例えば，ブランドスイッチのデータ行列
  - 行の対応するブランドと列の対応するブランドを別個のブランドとみなす
  - 2相2元類似度とした分析で得られた布置
  - 行の対応するブランドを表現する点と列の対応するブランドを表現する点とが同じ布置に表現

補足説明

個人差非対称MDS

個人差非対称MDS(2相3元非対称MDS)

- 2相3元非対称類似度がある場合

- 個人ごとの類似度の変動を誤差であると考える
  - 1つの布置を個人ごとの全てのデータに平均的に適合するように求める

- 類似度の非対称性を含めて個人ごとの類似度の変動が構造的な差異を表すと考える
  - 2相3元非対称多次元尺度構成法あるいは個人差非対称多次元尺度構成法(DeSarbo et al., 1992; Okada and Imaizumi, 1997; Zielman, 1991; Zielman and Heise, 1993) を用いて分析

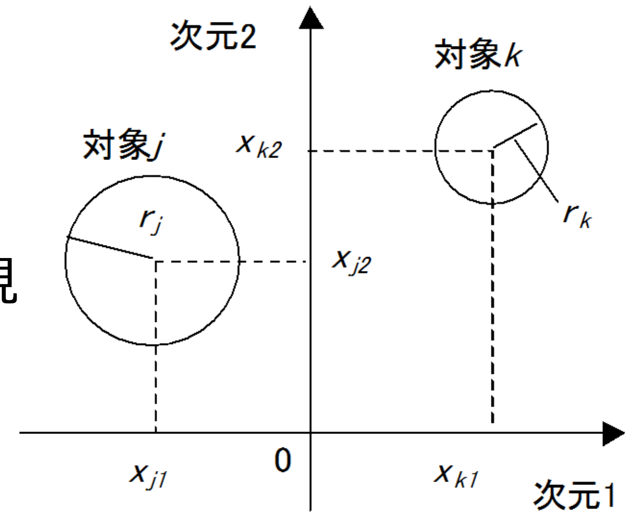
個人差非対称MDS(2相3元非対称MDS)

- 例えば，*N*週分のブランドスイッチのデータが1週間毎に得られている
  - 週ごとの類似度の変動を誤差であると考える
    - 1つの布置を*N*組全てのデータに平均的に適合するように求める
  - 類似度の非対称性を含めて週毎の類似度の変動が構造的な差異を表すと考える
    - 2相3元非対称多次元尺度構成法あるいは個人差非対称多次元尺度構成法で分析

個人差非対称MDS(2相3元非対称MDS)

- 対象間の類似度の非対称性が持つ情報を捨てることなく分析し，条件ごとの差も表現する

- 対象間の関係
  - 点：対象間の対称な関係を表現
  - 円：対象間の非対称な関係を表現



共通対象布置での対象jとk

- 条件ごとの差
  - 対象に対しての重み付けで表現
    - 点：対称重みによって重み付け
    - 円：非対称重みによって各次元に沿って重み付け

個人差非対称MDS(2相3元非対称MDS)

- 条件iでの対象jとkの類似度s_{jki}とm_{jki}が単調関係を満たすようなm_{jki}求め， 対象の位置と半径を決定
- 対象jを中心 とする円の半径の長さをr_j対象jとkの距離をd_{jk}，対称重みw_iでd_{jk}重み付けた条件iでの対象jとkの距離をd_{jki}とするとm_{jki}を以下のように仮定

$$m_{jki} = d_{jki} - v_{jki}r_j + v_{kji}r_k$$

$$v_{jki} = z_{jki}^{1/2}$$

$$z_{jki} = \frac{d_{jki}^2}{\sum_{t=1}^{p}\left(\frac{x_{jt} - x_{kt}}{u_{it}}\right)^2}$$

個人差非対称MDS(2相3元非対称MDS)

- s_{jki}とm_{jki}の単調関係のあてはまりの指標としてストレスを定義し，ストレスが最小となるように反復を繰り返し布置を決定
- ストレスSは以下のように定義できる

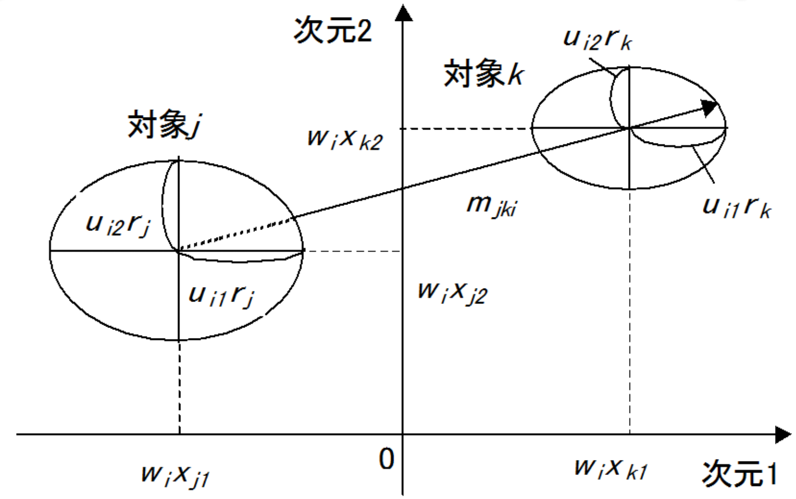
$$S = \left[\frac{1}{N} \sum_{i=1}^{N} S_i^{\,2} \right]^{1/2}$$

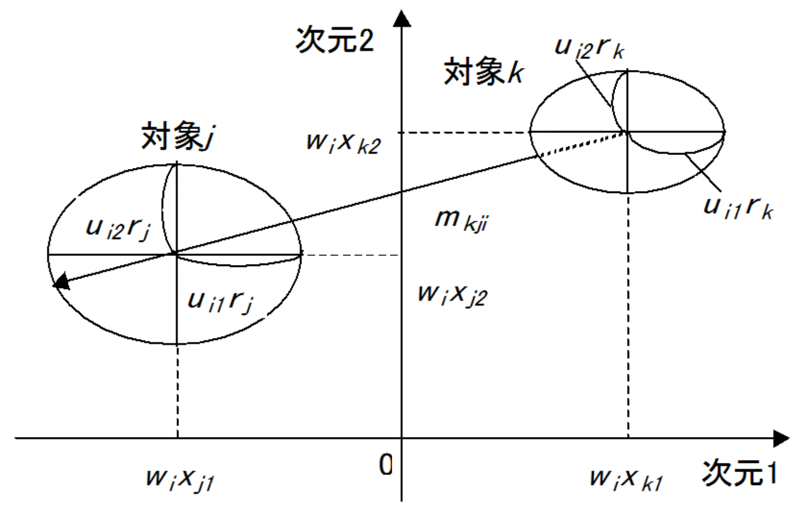
$$S_i = \left[\frac{\sum_{\substack{j=1 \\ j \neq k}}^{n} \sum_{k=1}^{n} (m_{jki} - \hat{m}_{jki})^2}{\sum_{\substack{j=1 \\ j \neq k}}^{n} \sum_{k=1}^{n} (m_{jki} - \overline{m}_i)^2} \right]^{1/2}$$

$$\overline{m}_i = \frac{\sum_{\substack{j=1 \\ j \neq k}}^{n} \sum_{k=1}^{n} m_{jki}}{n(n-1)}$$

個人差非対称MDS(2相3元非対称MDS)

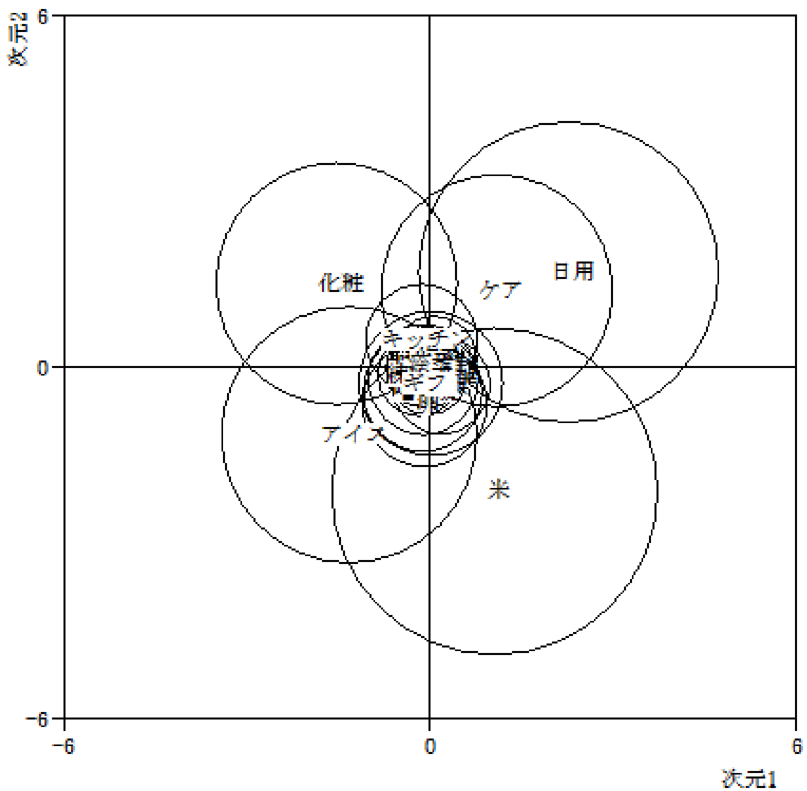
- 対称重みw_i (非負)
  - 条件iにおける布置に重み付け
- 非対称重みu_{it}(非負)
  - 対象jを中心とする円の半径を次元tに沿って重み付け
  - 円は楕円に引き伸ばされるか縮められる



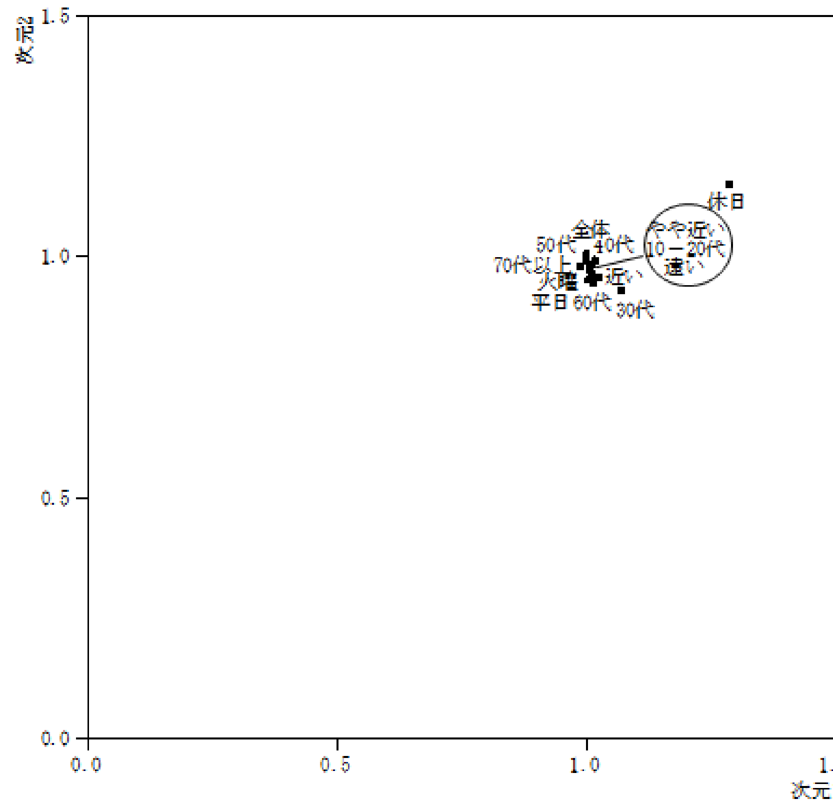


条件iで対象jとkを重み付け

スーパーの取り扱い商品の同時購買比率のデータを属性ごとに分析



2次元共通対象布置



2次元非対称重み布置

非対称MDS
事例による実習

はじめに

- 多摩大学大学院岡太彬訓先生と今泉忠先生から提供していただいた多次元尺度構成法をRで実行するためのプログラムを使用します.
  - このプログラムは現在開発中であり，開発者より今回のセミナーでのみ特別に使用を許可されております.
  - 現在のソースコードが拡散してしまうと，プログラムの使用方法についてのKNOW-HOWが共有できなくなってしまうため，本コース終了後，プログラムをPCから削除していただきますよう，お願いいたします.

  - なお，今後機能を拡張し，MDSのパッケージとして利用できる環境を提供する予定です.

準備

- パッケージsfsmiscのインストールが必要
- namsmds121_bsj.rdataを読み込む
  - Rの作業スペースとして

実行

- doMDSbatch(data)で実行
  - 内部に
    - BrandswitchingRaw（対象数15）
    - KyusyuODtakecuhi（対象数8）
    - Morsecode（対象数36）
    - SM1（対象数8）

    というデータをもっている
  - 全て非対称の単相2元類似度データ

実行

- result.SM1<-doMDSbatch(SM1)
  - ： DATA S(imilarity) or D(issimilarity)=
    - 類似度なのでS
  - ： FORM S(ymmetric) or A(symmetric) =
    - 非対称データなのでA
  - ： Maximum Dimensionality =
    - 最大次元数はここでは5としてみる

結果

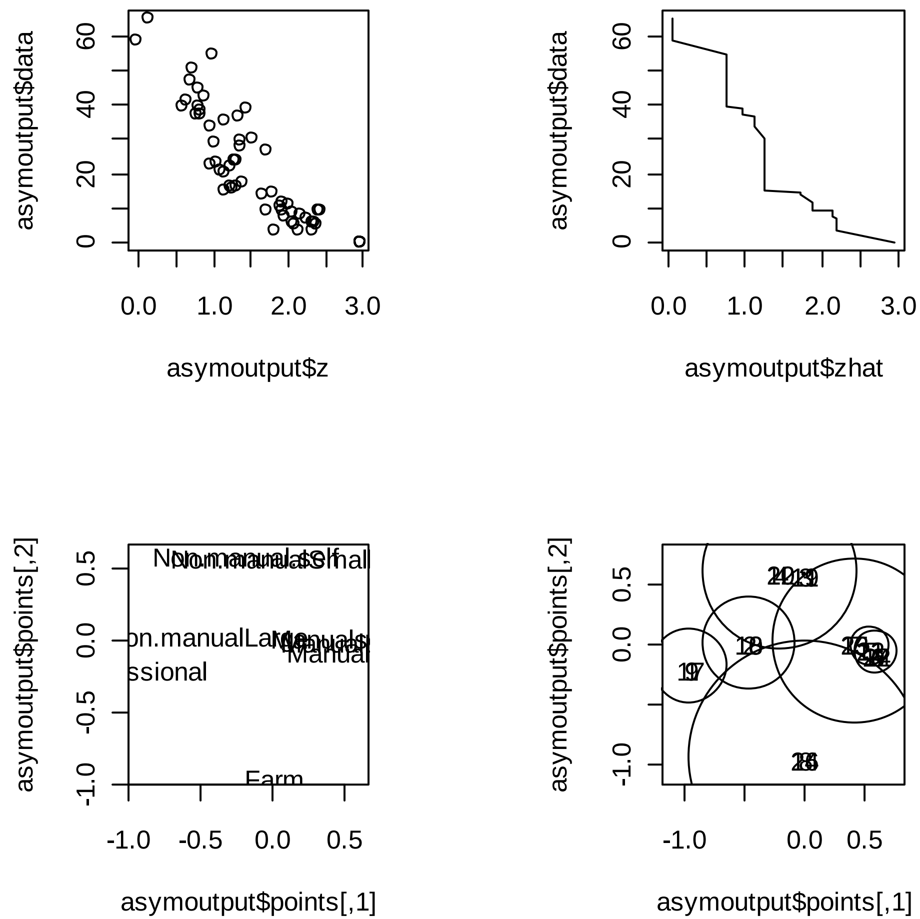
Scaling of 1 mode 2 way Asymmetric Similarity Data by nams
Iter= 0 Stress= 0.7828287
Iter= 100 Stress= 0.2376891
loop= 1 Final Stress 0.2376891 on Dim.= 5
Iter= 0 Stress= 0.3004018
Iter= 100 Stress= 0.2485784
loop= 1 Final Stress 0.2474830 on Dim.= 4
Iter= 0 Stress= 0.3654137
Iter= 100 Stress= 0.2730699
loop= 1 Final Stress 0.2730699 on Dim.= 3
Iter= 0 Stress= 0.3564004
Iter= 100 Stress= 0.3267034
loop= 1 Final Stress 0.3265235 on Dim.= 2
Iter= 0 Stress= 0.4774333
Iter= 100 Stress= 0.4485451
loop= 1 Final Stress 0.4484392 on Dim.= 1
Done
Final Stress is 0.4484392 0.3265235 0.2730699 0.2474830 0.2376891

結果の出力

- 分析が終了すると
  - Summary: plotting Dimensionality=

と聞かれる

  - 3としてみる
  - 右図が得られる



結果の出力

- 結果の概要を見ることができる
  - 左上：点間距離と類似度の散布図
  - 右上：ディスパリティーと類似度の折れ線グラフ
  - 左下：2次元布置
    - 分析した次元に関わらず1次元目と2次元目
  - 右下：非対称MDSの布置
    - 分析した次元に関わらず1次元目と2次元目

結果の出力

- さらに
  - Other Dimensionality Y(es) Or N(o)=

と聞かれる

  - 必要があれば，Yとすることで，再度結果の様子を見ることができる
  - 最終解とする次元が決まればNとする
    - 先に進む
  - Dimensionality you select =
  - 最終解とする次元を入力
    - 2としてみる
  - 結果が出力される

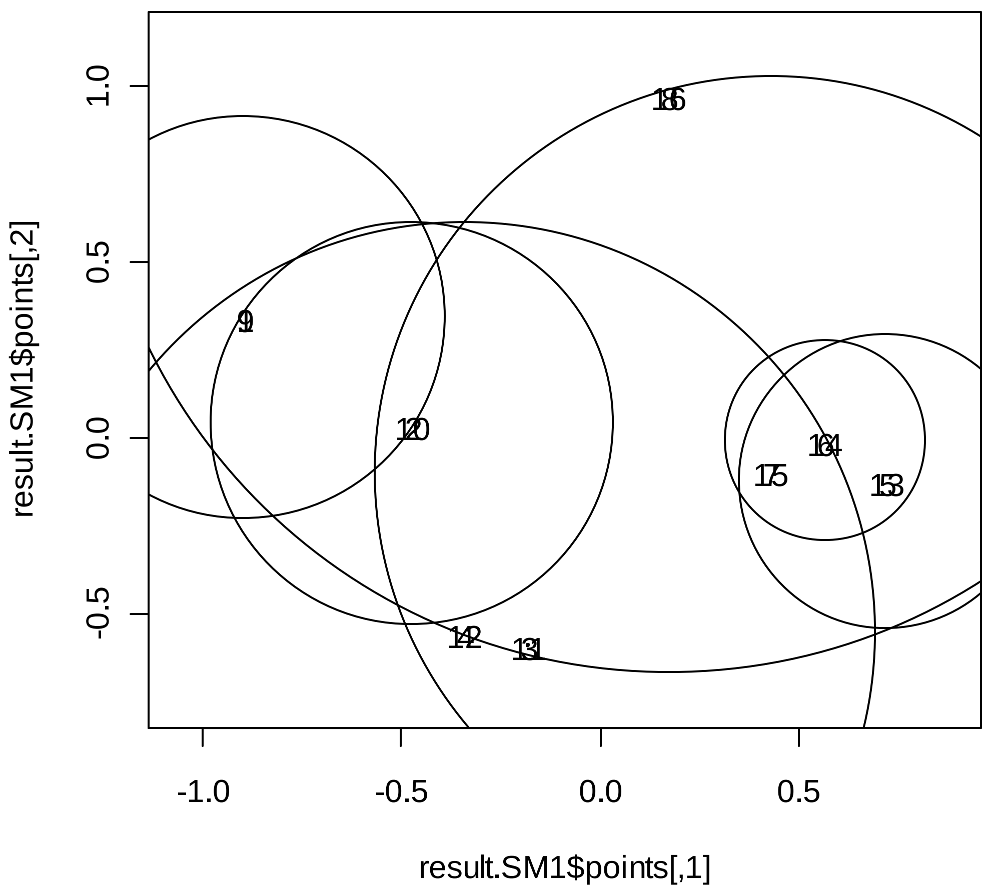
結果の出力

- $Stress, $r, $pointsに着目
  - $Stress : ストレス
  - $r : 円の半径
  - $points : 布置
- 布置を描きたい場合
  - symbols(result.SM1$points,circles=result.SM1$r,inches=FALSE)
  - text(result.SM1$points,rownames(result.SM1))

や

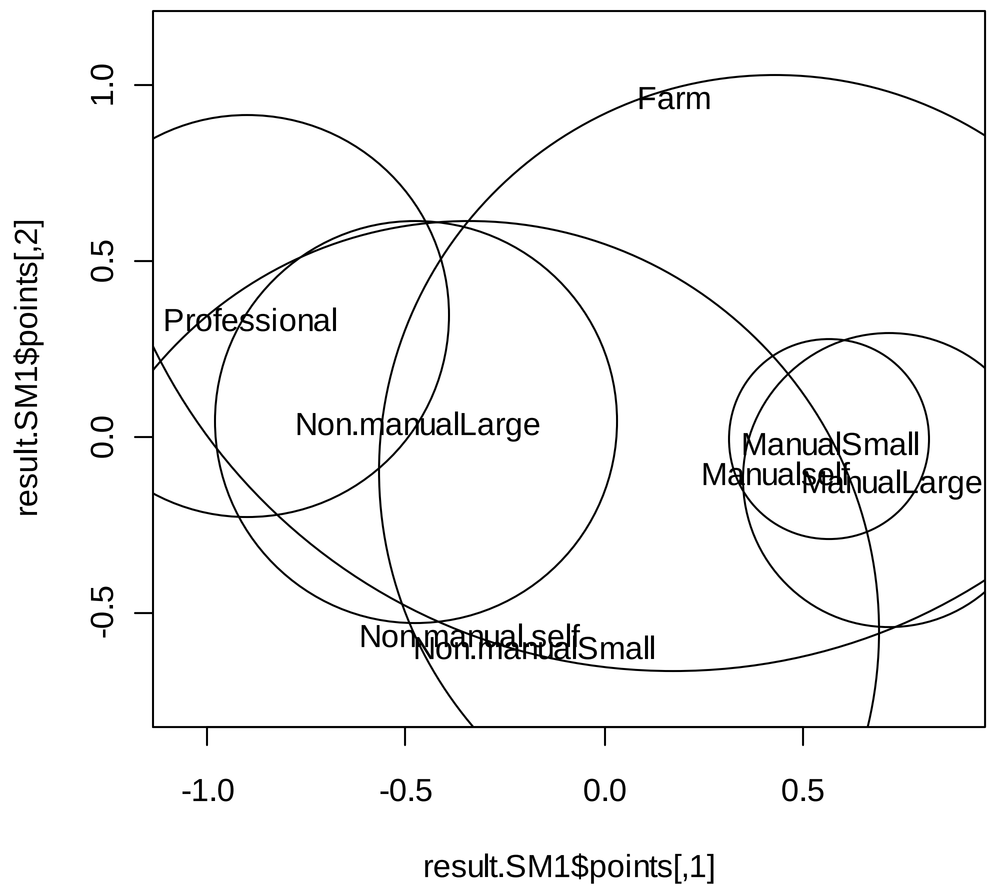
  - symbols(result.SM1$points,circles=result.SM1$r,inches=FALSE)
  - text(result.SM1$points,rownames(result.SM1$points))

結果の出力



18
6
9
1
12
10
16
4
17
5
15
3
14
2
13
11
1.0
0.5
0.0
-0.5
-1.0
-0.5
0.0
0.5
result.SM1$points[,2]
result.SM1$points[,1]

結果の出力



Farm
Professional
Non.manualLarge
ManualSmall
Manualself
ManualLarge
Non.manual.self
Non.manualSmall
1.0
0.5
0.0
-0.5
-1.0
-0.5
0.0
0.5
result.SM1$points[,2]
result.SM1$points[,1]

2日間のQ&A

論文などで活用する際の参考として
社会学におけるクラスター分析とMDS の応用

- 岡太彬訓(2002). 社会学におけるクラスター分析とMDS の応用. 理論と方法(Sociological Theory and Methods), Vol.17, No.2, 167-181.

- 社会学におけるクラスター分析法やMDS（多次元尺度構成法）の応用を概観
  - 手法の適用上で注目すべき点や問題点をとりあげている
  - 手法をより適切に利用し、また、より有意義な結果を導くにはどうすればよいのかについて記述

- 社会学に限らずクラスター分析やMDS の応用で一般的に留意すべき点
- 多くの応用に共通して認められる問題点について指摘

- 社会学における応用に焦点
  - それらの応用の特徴と具体的な問題点を述べ，問題点をどのように解決していけばよいのかを示唆
  - 主として社会学で発達してきたクラスター分析法やMDSの手法やそれらの利用方法を言及
  - クラスター分析法やMDS の社会学における応用について、今後の展開の可能性を述べられている

内容

- クラスター分析とMDS の応用上の一般的な留意点
  - クラスター分析の応用
  - MDSの応用
  - クラスター分析とMDS の併用
- 具体的な問題点
  - 布置の次元数の決定
  - 布置の作図
  - 分析手続の明瞭さ
  - 手法の種類
  - データ形式
  - 樹状図の利用
- 新しい手法の開発

最後に

- 本セミナーAコースで使用した資料やスクリプトはWeb Pageで配布いたします.
  - http://www.ae.keio.ac.jp/~satoru_y/
    から辿れるWeb Page
- 講師連絡先
  - 中山：atsuho@nagasaki-u.ac.jp
  - 横山：satoru_y@ae.keio.ac.jp