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27日（土）午後2(15:30～17:30)

MDS事例による実習 -個人差MDSと多次元展開法の説明と「R」による分析方法の紹介-

個人差MDS(INDSCAL)とは

個人差MDS(INDSCAL)

- 多次元尺度構成法（MDS）
  - 布置により対象間の関係を表現
- 個人差MDS(INDSCAL)
  - 個人ごとの差を取り込み，対象間の関係と各条件による差を表現することができる
    - 共通対象布置：対象間に共通の関係を表現
      - すべての個人の布置の原型
    - 重み布置：個人ごと差を表現
      - 必ずしも一人一人の差異という意味に限定されない
      - 実験条件やセグメント，期間毎の特徴の差異等についても分析可能

共通対象布置での点間距離

- 共通対象布置における対象jを表現する点と対象kを表現する点間の距離d_{jk}は以下のように定義される

$$d_{jk} = \left[\sum_{t=1}^{P} (x_{jt} - x_{kt})^2 \right]^{1/2}$$

ただし，x_{jt}は共通対象布置における対象jの次元tの座標で，x_{jt}を(j，t)要素にもつ$n \times p$行列を$\boldsymbol{X}$とする

個人iの対象布置(個人iの専用布置)

- 個人iの対象布置(個人iの専用布置)
  - 共通対象布置の次元を伸縮させることで得られる
  - 次元の伸縮は各次元に重みをかけて行う
  - 個人iの次元tに対する重みをw_{it}(非負)とし，w_{it}を(i，t)要素にもつ$N \times p$行列を$\boldsymbol{W}$とする

  - 個人iの専用布置における，対象jと対象kとの間の距離d_{jki}は以下のように定義できる

$$d_{jki} = \left[\sum_{t=1}^{P} w_{it} (x_{jt} - x_{kt})^2 \right]^{1/2}$$

共通対象布置での点の位置と重みの決定

- 対象jと対象kの条件iでの類似度δ_{jki}を，nを対象数，Nを条件数として，N個の$n \times n$の対称類似度行列で表すと，d_{jki}は，δ_{jki}に対応するように決定
- 共通対象布置と重みに基づく内積と，類似度データから求めた内積とが一致するように，共通対象布置における点の位置と重みは決定

共通対象布置での点の位置と重みの決定

- 個人iの専用布置では対象jと対象kとを表現する点の座標はそれぞれ以下のように表現できる

$$\left(\sqrt{w_{i1}}x_{j1}, \sqrt{w_{i2}}x_{j2}, \cdots, \sqrt{w_{ip}}x_{jp}\right), \quad \left(\sqrt{w_{i1}}x_{k1}, \sqrt{w_{i2}}x_{k2}, \cdots, \sqrt{w_{ip}}x_{kp}\right)$$

- 原点からこれら2つの点へ向かうベクトルの内積 は

$$\hat{b}_{jki} = \sum_{t=1}^{p} w_{it} x_{jt} x_{kt}$$

- 個人iの類似度データから点間距離を計算して類似度データから求めた内積b_{jki}とこの$\hat{b}_{jki}$の差が最小となるように共通対象布置と重み布置の座標を決定

共通対象布置での点の位置と重みの決定

- 一定の次元pのもとで，共通対象布置の初期布置を求め，この初期共通対象布置を少しずつ改善
- 適合度指標であるVAF比を最大化する共通対象布置と重み布置を反復計算により求める

$$\text{VAF比} = 1 - \frac{\sum_{i=1}^{N}\sum_{j=1}^{n}\sum_{k=1}^{n}(b_{jki} - \hat{b}_{jki})^2}{\sum_{i=1}^{N}\sum_{j=1}^{n}\sum_{k=1}^{n}(b_{jki} - \bar{b}_i)^2}$$

ただし，$\bar{b}_i$ は個人iの内積の平均値で，データの基準化により通常は0である

- 個人が分析に対する影響を平準化するために，各個人の内積行列の要素の平均値が0，2乗和が1になるよう基準化

共通対象布置での点の位置と重みの決定

- VAF比
  - 類似度データに対する布置の適合度
- 類似度データδ_{jki}と距離d_{jki}が単調関係を満たすように，VAF比を最大化する布置を求める
  - 共通対象布置の初期布置を求める
  - この初期共通対象布置をから重み布置を求める
  - この重み布置から共通対象布置を求める
  - このように共通対象布置と重み布置を交互に求める手順を反復する

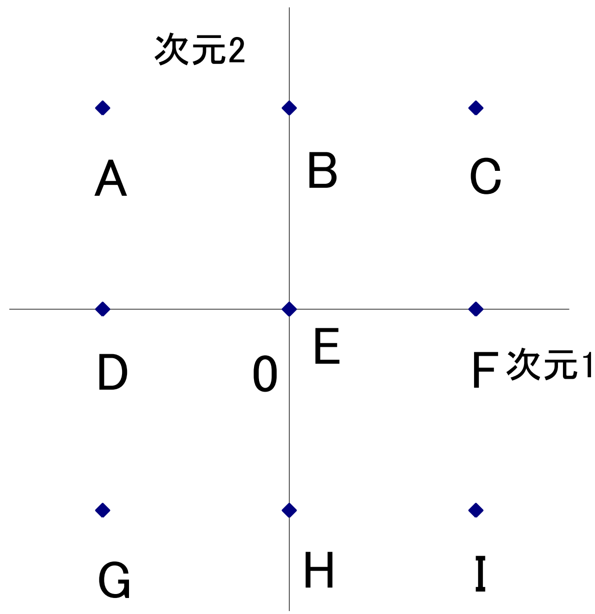
重み布置

- 重み布置
  - w_{it}を個人iの次元tの座標とみなとき，N人の個人のp次元空間における布置が得られる
  - 重み布置は共通対象布置と同じ次元をもつが，共通対象布置とは別々に表現

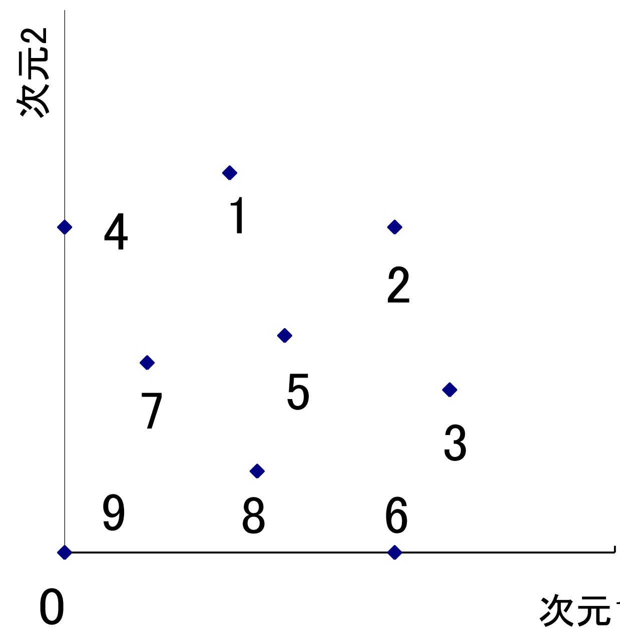
- 重みw_{it}は個人iの類似度判断に関する次元tの重要度を示す
  - 各次元での対象jと対象kの座標の差の2乗$(x_{jt}-x_{kt})^2$がw_{it}により重み付けられる

共通対象布置と重み布置の例

- 6個の対象からなる2次元の共通対象布置と9人の個人からなる重み布置



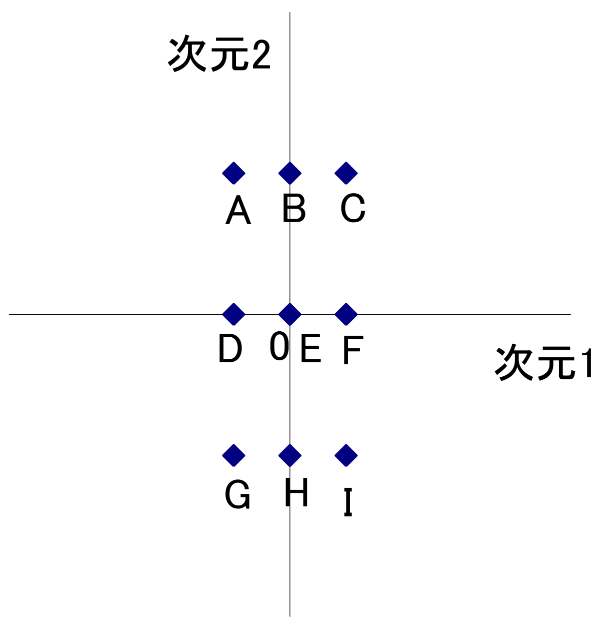
共通対象布置



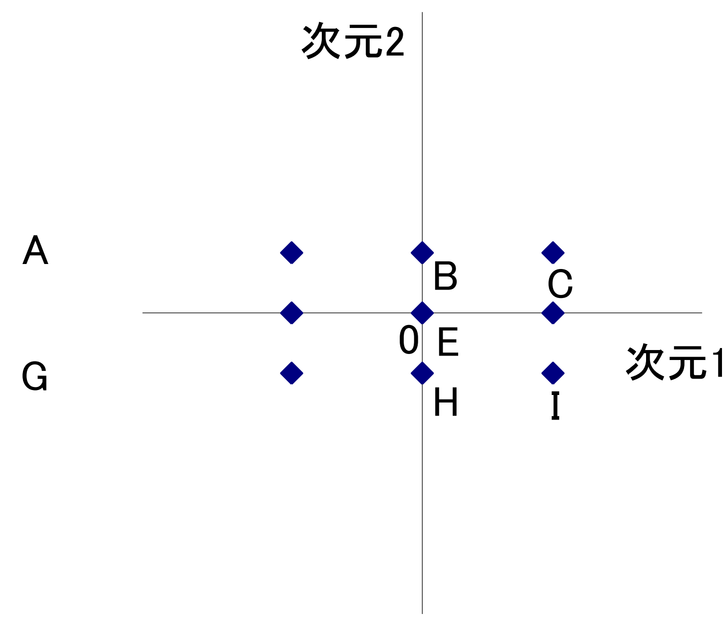
重み布置

専用布置：個人1と個人3

- 個人1では次元1の重みが次元2の重みよりも小さくく($w_{11} < w_{12}$)，個人1の専用布置は，次元1の方向が圧縮され，共通対象布置と比較すると縦長となる
- 個人3では次元1の重みが次元2の重みよりも大きく($w_{31} > w_{32}$)，個人1の専用布置は，次元2の方向が圧縮され，共通対象布置と比較すると横長となる



個人1の専用布置



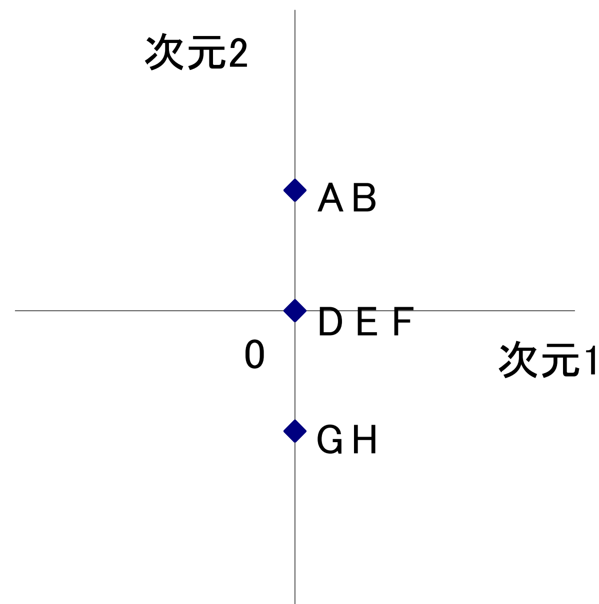
個人3の専用布置

専用布置：個人1と個人3

- 個人1の専用布置が横長であるということ
  - 共通池沼布置の次元1と次元2の座標の同じ長さの差が，専用布置では次元1の差の方が次元2の差の方が大きくなる
- 個人1が対象間の類似度判断を行うときに次元1を次元2よりも重視する
- 個人3の専用布置が縦長であるということ
  - 個人2が次元2を次元1よりも重視することを意味

専用布置：個人4と個人6

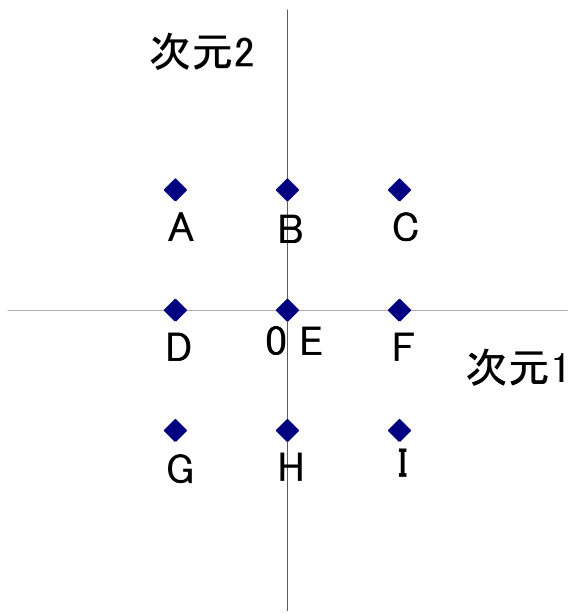
- 個人4は次元1の重みが0(w_{41} = 0)
  - 専用布置は次元2だけで表現される1次元布置
  - 対象間の類似度判断を次元2のみで行うことを表す
- 個人6では次元2の重みが0(w_{62} = 0)
  - 専用布置は次元1だけで表現される1次元布置
  - 対象間の類似度判断を次元1のみで行うことを表す



個人4の専用布置

専用布置：個人2と個人5

- 個人2では次元1の重みと次元2の重みが等しく($w_{21} = w_{22}$)，専用布置は共通対象布置と相似(図？)
  - 個人5についても同様(個人5の専用布置は個人2の専用布置よりも縮小)



個人2の専用布置

専用布置：個人7と個人9

- 個人1と個人7は次元1と次元2の重みの比が等しく($w_{11}/ w_{12}= w_{71}/ w_{72}$)，両者は原点からの1本の直線状
- 両者の専用布置は相似であり，個人7の専用布置は個人1の専用布置よりも縮小
  - 同様の関係は個人3と個人8($w_{31}/ w_{32}= w_{81}/ w_{82}$)にあてはまる
- 個人9は2つの重みがともに0
  - 専用布置は原点に9個の対象を表現する点が集中したものになる

次元の方向の一義性

- INDSCALでは，重み**W**により共通対象布置の次元の方向が一義的に決まる
  - 次元の方向を回転し，回転後の共通対象布置の次元に重み付け
  - 回転前の共通対象布置の次元に重みを与えた場合とは異なる布置
  - 各個人の専用布置における点間距離d_{jki}が共通対象布置の回転により変化し，データに対する適合度が変化
- 次元の反転と互換ではd_{jki}は変化せず，一義的には決定されず，布置の解釈には影響しない

個人差MDS(INDSCAL)を関数smacofIndDiff()で実行

- 関数smacofIndDiffで個人差MDS(INDSCAL) を実行するには

```
> smacofIndDiff(delta, ndim = 2, weightmat = NULL, init = NULL,
  metric = TRUE, ties = "primary", constraint = NULL, verbose =
  FALSE, modulus = 1, itmax = 1000, eps = 1e-6)
```

  - delta : A list of dissimilarity matrices or a list objects of class dist
  - Ndim : Number of dimensions
  - Weightmat : Optional matrix with dissimilarity weights
  - Init : Matrix with starting values for configurations (optional)
  - metric : If FALSE non-metric MDS is performed
  - ties : Tie specification for non-metric MDS
  - constraint : Either NULL, "idioscal", "diagonal", or "identity" (see details)
  - verbose : If TRUE, intermediate stress is printed out
  - modulus : Number of smacof iterations per monotone regression call
  - itmax : Maximum number of iterations
  - eps : Convergence criterion

多次元展開法とは

多次元展開法(Unfolding)

- これまで説明してきたMDS では，行と列に同じ対象が並んだ正方行列（行の数と列の数が等しい行列のこと）を分析の対象としていた

- 多次元展開法
  - 行と列に違う対象が並んだ(正方でない)矩形行列(対象×変数からなるような選好度のデータ)を分析可能
  - 対象×変数からなるデータに対して，次元数を2とした多次元展開法を適用
  - 対象と変数のそれぞれについての座標の値が得られる

多次元展開法(Unfolding)

- ある6人の7つの観光地がどれくらい好きであるかという選好度が10件法で得られているとする
  - 値が大きいほどその観光地が嫌いであることを示す
  - すなわち，値が大きい(小さい)ほどその人がその観光地に対して抱いている心理的距離が遠い(近い)と判断する
- このようなデータに対し，多次元展開法を適用することにより，列成分(観光地)同士や行成分(個人)同士がどの程度似ているか，その特徴を把握することができる

多次元展開法(Unfolding)

- 多次元展開法では，行と列の両方を同じ空間内に表現可能
  - 同じ空間内の距離として評価することができる
- 観光地と各個人の近さ(遠さ)
  - 心理的距離がどれくらい近いか(遠いか)を表現
  - ある個人がその観光地をどれくらい好き(嫌い)と感じているかを表している
  - 観光地についての布置を考察することで，それぞれの座標軸の命名や解釈を行うこともできる

多次元展開法(Unfolding)

- Coombs (1950, 1964) がスタート
  - 最初は刺激に対する個人の選好順位のデータから，１次元順序尺度を構成する方法として出発
- BenettとHayにより多次元順序尺度構成の方法に発展し(Benett & Hays, 1960; Hays & Bennett, 1961)
  - 現在では古典的展開法とも呼ばれる(斉藤，1980)

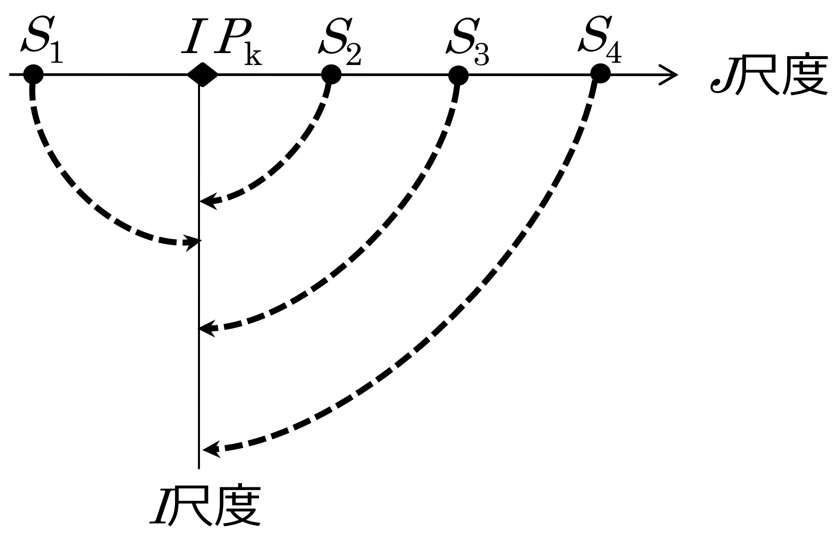
- 非計量的展開法(nonmetric unfolding)
  - 古典的展開法は，選好順位データをもとに刺激布置及び理想点の多次元距離空間上の布置を仮定し，順位データが可能な限り点間距離の単調減少関数となるように両布置を定める非計量的展開法に発展(Lingoes, 1966; Kruskal, 1968; Young & Torgerson, 1967)

多次元展開法(Unfolding)

- 計量的展開法(metric unfolding)
  - Schőnemannらにより，刺激と個人の理想点間の距離の推定値が比率尺度で得られている場合の計量的展開法へ発展(Greenacre & Browne, 1986; Schőnemann, 1972; Sch¨onemann & Wang, 1972; Zinnes & Griggs, 1974)
  - 刺激布置X={x_{ia}}と個人の理想点布置$\boldsymbol{Y}$={y_{ka}}間の距離d_{ik}の推定値が何らかの方法で得られている場合に，刺激布置$\boldsymbol{X}$及び理想点の布置$\boldsymbol{Y}$を求める方法
- ALSCAL (Takane, Young, and De Leeuw, 1977)
  - 計量的と非計量的データのどちらの場合にもデータを距離空間上に位置づける統一的なアルゴリズムを提案し，１つのモデルとして組み込んでいる

古典的展開法： Coombsの１次元展開法

- 刺激と個人は*J*尺度という共通の１次元尺度上の点で表現
- それぞれの個人の刺激に対する選好順位は各個人の理想点と呼ばれる位置から刺激点までの距離の順位に一致すると仮定
- 個人の刺激に対する選好順位は*I*尺度と呼ばれる
  - *J* 尺度上の個々の刺激を個人O_k の理想点IP_k で折りたたむ(fold)ことで得られる
  - 実際には，刺激に対する*I*尺度上の選好順位データから*J*尺度上での刺激の位置と個人の理想点を算出



多次元展開法(Unfolding)を関数smacofRect()で実行

- 関数smacofRectで多次元展開法(Unfolding)を実行するには

  > smacofRect(delta, ndim = 2, weightmat = NULL, init = NULL, verbose = FALSE, itmax = 1000, reg = 1e-6, eps = 1e-6)

  - delta : Data frame or matrix of preferences, ratings, dissimilarities
  - ndim : Number of dimensions
  - weightmat : Optional matrix with dissimilarity weights
  - init : Matrix with starting values for configurations (optional)
  - verbose : If TRUE, intermediate stress is printed out
  - itmax : Maximum number of iterations
  - reg : Regularization factor, prevents distances from beeing 0
  - eps : Convergence criterion

MDSの事例による実習
個人差MDSと多次元展開法の「R」による分析方法

準備

- INDSCALをする前に
- データの読み込み
  - d1 <- 8-read.table("indscal1.dat", sep="¥t", header=T, row.names=1)
  - d2 <- 8-read.table("indscal2.dat", sep="¥t", header=T, row.names=1)
  - 非類似度データにする必要があるので8から引いている
- データの結合
  - indscaldata <- list(A=d1, B=d2)
    - AさんとBさんのデータだと考える
  - listで結合する
    - 「,」で結合することで複数人分を結合する

準備

- indscaldata

$A

	Greece	Hawaii	West.Coast	Hong.Kong	LondonParis	East.America	Australia
Greece	8	1	4	6	3	2	3
Hawaii	1	8	5	5	6	4	6
West Coast	4	5	8	2	4	3	5
Hong Kong	6	5	2	8	2	1	7
LondonParis	3	6	4	2	8	1	7
East America	2	4	3	1	1	8	4
Australia	3	6	5	7	7	4	8

$B

	Greece	Hawaii	West.Coast	Hong.Kong	LondonParis	East.America	Australia
Greece	8	7	6	7	1	2	4
Hawaii	7	8	5	3	6	5	4
West Coast	6	5	8	2	6	3	2
Hong Kong	7	3	2	8	7	5	3
LondonParis	1	6	6	7	8	1	4
East America	2	5	3	5	1	8	1
Australia	4	4	2	3	4	1	8

smacofIndDiff()

- 個人差MDS(INDSCAL)
  - smacofIndDiff(delta, ndim = 2, weightmat = NULL, init = NULL, metric = TRUE, ties = "primary", constraint = NULL, verbose = FALSE, modulus = 1, itmax = 1000, eps = 1e-6)
    - delta : A list of dissimilarity matrices or a list objects of class dist
    - ndim : Number of dimensions
    - weightmat : Optional matrix with dissimilarity weights
    - Init : Matrix with starting values for configurations (optional)
    - metric : If FALSE non-metric MDS is performed
    - ties : Tie specification for non-metric MDS
    - constraint : Either NULL, "idioscal", "diagonal", or "identity" (see details)
    - verbose : If TRUE, intermediate stress is printed out
    - modulus : Number of smacof iterations per monotone regression call
    - itmax : Maximum number of iterations
    - eps : Convergence criterion

smacofIndDiff()

- ライブラリの読み込み
  - library(smacof)
- 実行
  - indscalout <- smacofIndDiff(delta=indscaldata, constraint=“diagonal”, ndim=2)
- 共通空間布置の表示
  - indscalout$gspace

```
          D1         D2
1 -0.60064182 -0.2438682
2  0.69693572 -0.3029180
3  0.25611294  0.2911673
4  0.50293480  0.4637668
5 -0.57860073  0.3644406
6 -0.26492621  0.1168288
7 -0.01181470 -0.6894173
```

smacofIndDiff()

- 個人空間の座標軸の表示
  - indscalout$conf

```
$A
            D1         D2
1 -0.352003702 -0.3509553
2  0.408436351 -0.4359348
3  0.150093950  0.4190242
4  0.294742902  0.6674153
5 -0.339086613  0.5244732
6 -0.155258933  0.1681305
7 -0.006923955 -0.9921531

$B
           D1          D2
1 -0.79161923 -0.09467626
2  0.91853031 -0.11760097
3  0.33754548  0.11303903
4  0.66284573  0.18004683
5 -0.76257006  0.14148573
6 -0.34916098  0.04535612
7 -0.01557125 -0.26765048
```

smacofIndDiff()

- 各次元への重みの表示
  - indscalout$cweights

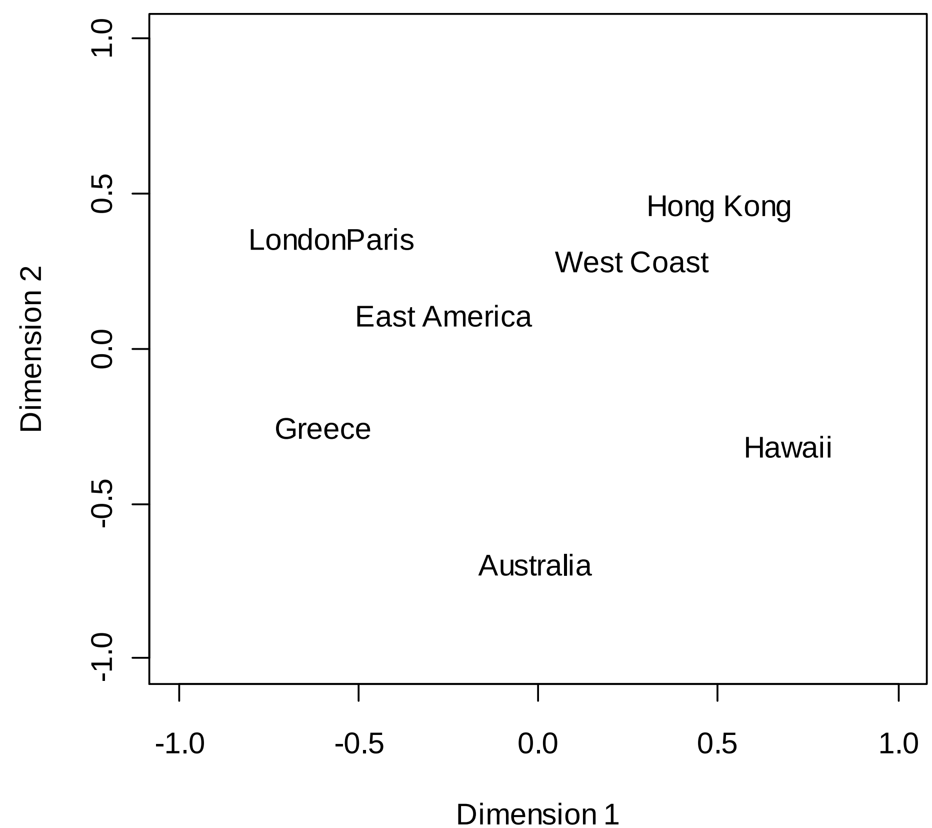
```
$A
        D1       D2
D1 0.586046 0.000000
D2 0.000000 1.439118

$B
        D1        D2
D1 1.317956 0.0000000
D2 0.000000 0.3882271
```

smacofIndDiff()

- 共通布置のプロット
  - plot(indscalout$gspace, type=“n”, xlim=c(-1,1), ylim=c(-1,1), xlab=“Dimension 1”, ylab="Dimension 2")
  - text(indscalout$gspace,labels=row.names(indscaldata[[1]]))
- 個人布置のプロット
  - plot(indscalout$conf$A, type=“n”, xlim=c(-1,1), ylim=c(-1,1), xlab=“Dimension 1”, ylab="Dimension 2")
  - text(indscalout$conf$A,labels=row.names(indscaldata[[1]]))
  - 個人Bは$Bにすればよい

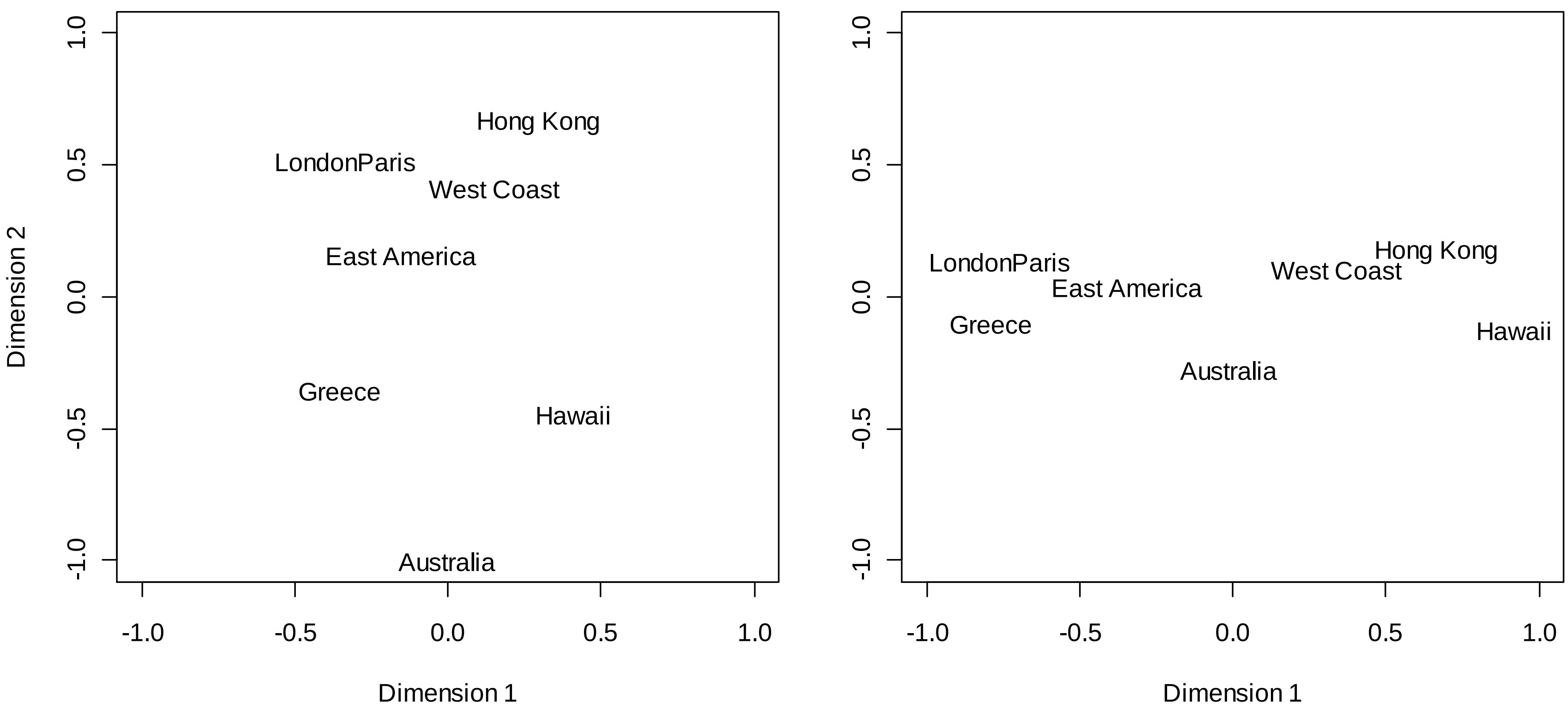
smacofIndDiff()

- 共通布置



smacofIndDiff()

- 個人布置



準備

- Unfoldingをする前に
- データの読み込み
  - unfoldingdata <- 10-read.table("unfolding.dat", sep="¥t", header=T, row.names=1)
  - 非類似度データにする必要があるので10から引いている
  - unfoldingdata

  Greece Hawaii West.Coast Hong.Kong LondonParis East.America Australia

ind1	9	6	7	3	7	5	9
ind2	5	2	4	1	7	4	4
ind3	3	6	2	8	2	1	5
ind4	4	2	1	9	8	6	2
ind5	3	4	1	7	5	5	3
ind6	6	3	1	9	7	2	4

smacofRect()

- 多次元展開法(Unfolding)
  - smacofRect(delta, ndim = 2, weightmat = NULL, init = NULL, verbose = FALSE, itmax = 1000, reg = 1e-6, eps = 1e-6)
    - delta : Data frame or matrix of preferences, ratings, dissimilarities
    - ndim : Number of dimensions
    - weightmat : Optional matrix with dissimilarity weights
    - init : Matrix with starting values for configurations (optional)
    - verbose : If TRUE, intermediate stress is printed out
    - itmax : Maximum number of iterations
    - reg : Regularization factor, prevents distances from beeing 0
    - eps : Convergence criterion

smacofRect()

- ライブラリの読み込み（既に行っている場合は不要）
  - library(smacof)
- 実行（2次元解）
  - unfoldingout <- smacofRect(delta=unfoldingdata,ndim=2)
- ）

smacofRect()

- 結果の表示
  - summary(unfoldingout)

Subjects configurations (rows):

	D1	D2
ind1	5.6282	-0.4433
ind2	2.2371	1.6988
ind3	-0.5589	-2.1464
ind4	-2.9081	1.9410
ind5	-2.5249	0.3816
ind6	-2.5287	0.9743

Objects configurations (columns):

	D1	D2
Greece	-2.5796	-2.9770
Hawaii	0.0462	2.9740
West.Coast	-1.6041	0.4551
Hong.Kong	5.0806	2.4947
LondonParis	0.0852	-4.7708
East.America	0.5670	-1.8074
Australia	-1.5952	3.6314

smacofRect()

- 結果のプロット
  - plot(unfoldingout,plot.type="confplot",joint=T)