

ESCALONAMENTO MULTIDIMENSIONAL

MONIQUE SARAH LEVY MALZYNER

DISSERTAÇÃO APRESENTADA

AO

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

DA

UNIVERSIDADE DE SÃO PAULO

PARA OBTENÇÃO DO GRAU DE MESTRE

EM

ESTATÍSTICA

ORIENTADOR:

PROF. DR. WILTON DE OLIVEIRA BUSSAB

- São Paulo, Junho de 1981 -

Ao meu marido

Artur

E aos meus pais

Mirella e Victor

Agradecimentos:

Queremos, através desta página, agradecer a todos que colaboraram na realização deste trabalho, mas em especial a:

Professor Doutor Wilton de Oliveira Bussab, por sua profícua orientação ao trabalho;

Professor Doutor Adolpho Walter Pimazoni Canton, pelo interesse demonstrado na execução do nosso trabalho;

Professor Doutor Antonio Ribeiro de Almeida, por ter cedido o programa ALSCAL;

Professor Doutor Arno Engelmann, pela cessão dos dados usados na nossa aplicação;

Professora Graça Bressan, por ter implantado o programa ALSCAL no Instituto de Física da Universidade de São Paulo;

Regina Helena da Silva, por seus préstimos datilográficos;

Todos os amigos do Instituto de Matemática e Estatística da Universidade de São Paulo, por seu apoio e incentivo inestimáveis.

Monique Sarah Levy Malzyner

ÍNDICE

	Pg.
1. CAPÍTULO 1 - INTRODUÇÃO	1
1.1 - Apresentação do Problema	1
1.2 - Aplicabilidade.	6
1.3 - Descrição do Trabalho.	8
2. CAPÍTULO 2 - MEDIDAS DE PROXIMIDADE	11
2.1 - Introdução.	11
2.2 - Considerações Matemáticas	13
2.3 - Medidas de Proximidade	15
2.3.1 - Introdução.	15
2.3.2 - Alguns procedimentos para obtenção de medidas de proximidade.	19
3. CAPÍTULO 3 - ALGUNS MÉTODOS DE E.M.D. A "2 FA TORES".	31
3.1 - Introdução.	31
3.2 - Método Clássico.	33
3.2.1 - Definições e Teoremas	33
3.2.2 - Solução do E.M.D. Métrico.	37
3.2.3 Propriedades ótimas da Solução Clássica e Adequação do Ajuste.	42
3.3 - Métodos Não Métricos.	46
3.3.1 - Método de Coombs.	46
3.3.2 - Método de Shepard: Análise de Proximidades.	46

	pg.
3.3.3 - Método de Kruskal.	50
3.4 - Comparação de Configurações.	51
3.4.1 - Medida de Adequação de Ajuste.	51
3.4.2 - Rotação Procrusteaana.	53
3.4.3 - Fator de Escala.	53
4. CAPÍTULO 4 - MÉTODO DE KRUSKAL.	58
4.1 - Adequação de Ajuste.	58
4.2 - Obtenção da Solução	62
4.2.1 - Nível intuitivo.	62
4.2.2 - Nível teórico.	62
4.2.3 - Nível prático.	63
4.3 - Diagrama de Dispersão.	72
5. CAPÍTULO 5 - ESCALONAMENTO MULTIDIMENSIONAL :	
"3 FATORES"	80
5.1 - Introdução	80
5.2 - Técnicas de E.M.D. a "3 Fatores"	85
6. CAPÍTULO 6 - ALSCAL	96
6.1 - Introdução	96
6.2 - Dados - Mensuração	99
6.2.1 - Processos de Mensuração.	100
6.2.2 - Níveis de Mensuração.	101
6.2.3 - Condicionalidade.	103

	Pg.
6.3 - Medida de Ajuste.	105
6.4 - Descrição da Técnica ALSCAL	109
7. CAPÍTULO 7 - ANÁLISE DA SOLUÇÃO DO E.M.D.	113
7.1 - Escolha da dimensão do espaço.	113
7.1.1 - Introdução.	113
7.1.2 - Ajuste	115
7.1.3 - Interpretabilidade da solução.	120
7.1.4 - Facilidade de uso da solução	122
7.1.5 - Estabilidade da solução.	123
7.2 - Interpretação da Configuração.	125
7.2.1 - Introdução.	125
7.2.2 - Rotação da configuração	126
7.2.3 - Regressão linear.	129
7.2.4 - Análise por vizinhanças	137
7.2.5 - Outras formas de interpretação	144
7.2.6 - Auxílio na análise da solução	150
7.2.7 - Representação de uma configuração num espaço tridimensional	152
8. CAPÍTULO 8 - APLICAÇÃO	154
9. CAPÍTULO 9 - COMENTÁRIOS	173
9.1 - Distâncias Não Euclidianas	173
9.2 - Dados repetidos.	175
9.3 - Medidas de Ajuste.	176
9.4 - Considerações finais	177

	pg.
A. APÊNDICE A - DEMONSTRAÇÕES DOS TEOREMAS DO CA- PÍTULO 3	179
A.1 - Demosntração do teorema 3.1. : :	180
A.2 - Demonstração do teorema 3.2.	183
A.3 - Demonstração do teorema 3.3.	184
A.4 - Demonstração do teorema 3.4.	186
A.5 - Demonstração do teorema 3.5.	187
B. APÊNDICE B - REGRESSÃO MONOTÔNICA.	191
C. APÊNDICE C - "STEEPEST DESCENT".	194
C.1 - Métodos de Gradiente.	195
C.2 - Método do "Steepest Descent".	196
D. APÊNDICE D - FASE INICIAL: ALSCAL.	197
E. APÊNDICE E - FASE DE ESCALONAMENTO ÓTIMO:ALSCAL	205
E.1 - Obtenção de D*.	206
E.1.1 - Constante aditiva	210
E.2 - Normalização da solução do E.M.D.	213
F. APÊNDICE F - FASE DE ESTIMAÇÃO DO MODÉLO:ALSCAL	219
F.1 - Obtenção dos estimadores dos pesos.	220
F.2 - Restrições de não-negatividade.	221
F.3 - Obtenção dos estimadores de mínimos qua- drados da configuração.	223
G. APÊNDICE G - SAÍDA DO PROGRAMA ALSCAL.	231
BIBLIOGRAFIA	244

CAPÍTULO 1

INTRODUÇÃO

1.1 - APRESENTAÇÃO DO PROBLEMA

Supondo que nos seja dado um mapa apresentando a localização de algumas capitais brasileiras, e que se deseje a partir do mesmo a construção de uma tabela das distâncias entre essas capitais, a tarefa é extremamente simples. Basta conhecer a escala na qual o mapa foi construído e com o auxílio de uma régua obtém-se as distâncias pedidas (Figura 1.1).

Agora consideremos o problema inverso, no qual nos é dada uma tabela apresentando todas as distâncias entre algumas capitais brasileiras e nos é pedida a construção do mapa. A solução já não é tão imediata.



Figura 1.1 - Mapa do Brasil

Em essência o Escalonamento Multidimensional, ou E.M.D. é um método que permite a resolução desse tipo de problema. Na figura 1.2 pode-se observar o "mapa" do Brasil obtido através do E.M.D. a partir das distâncias entre algumas capitais brasileiras apresentadas na tabela 1.1. A aplicação típica do E.M.D., no entanto, é bem mais complicada do que este simples exemplo possa sugerir. Por um lado os dados geralmente contêm algum erro. Por outro lado, raramente sabe-se com antecedência se um simples mapa bi-dimensional é adequado, ou se é necessário um "mapa" que use três, quatro, ou mesmo mais dimensões.

O E.M.D. engloba uma classe de técnicas. Essas técnicas usam proximidades entre objetos como entrada. Uma proximidade é um número que indica quão similares ou quão diferentes são dois objetos, ou como são percebidos dois objetos. Ou ainda, é alguma outra medida que interpreta a relação entre dois objetos. O objetivo principal é uma representação espacial que consista em uma configuração geométrica de pontos, semelhante a um mapa. Cada ponto na configuração corresponde a um dos objetos. Espera-se que essa configuração reflita a "estrutura escondida" nos dados e torne os dados mais compreensíveis. Por refletir a estrutura dos dados, entende-se que a maior dissimilaridade (ou a menor similaridade) entre dois objetos, representada pelo valor da medida de proximidade, corresponda o maior afastamento

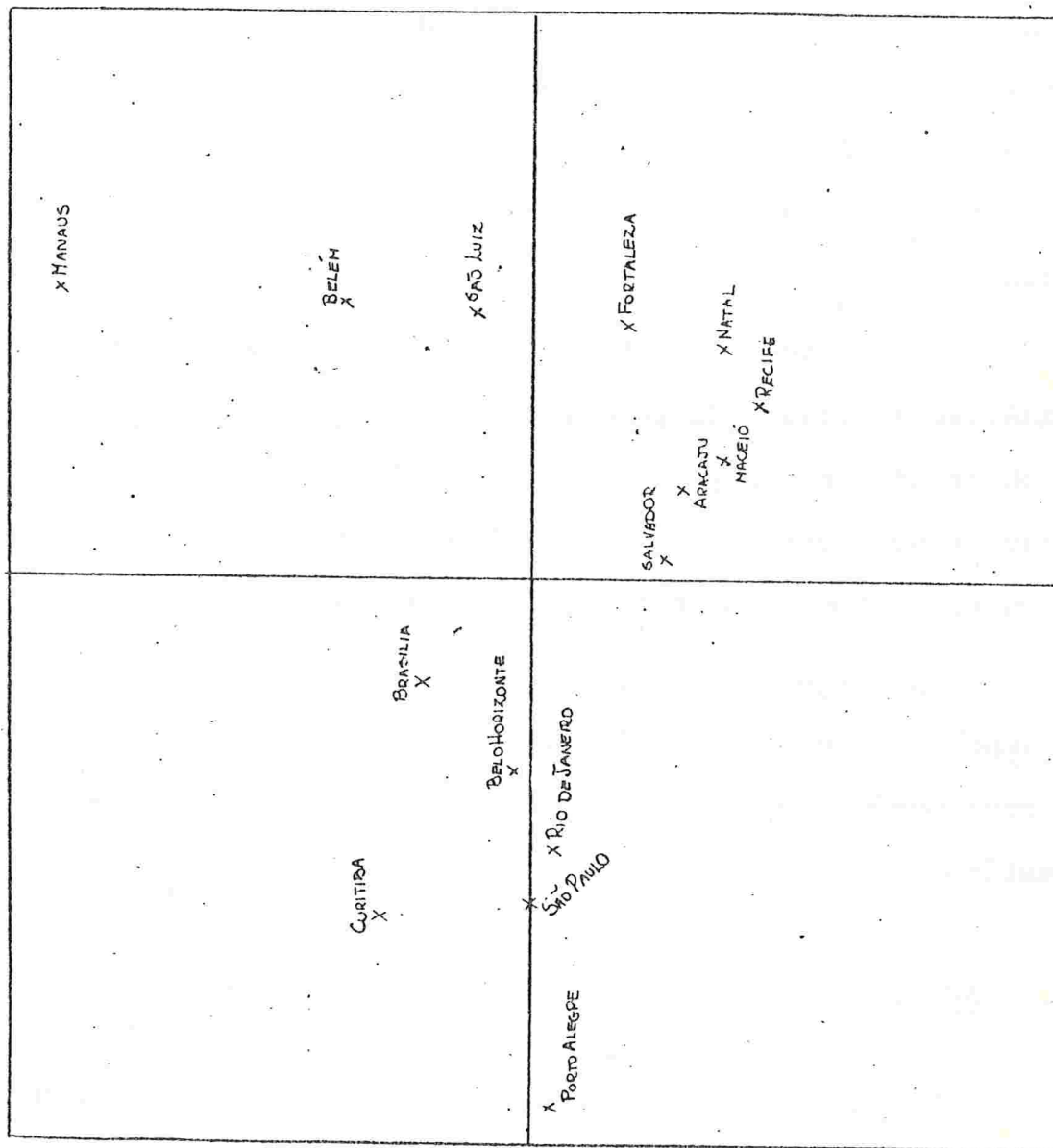


Figura 1.2 - Mapa do Brasil obtido através do E.M.D. (programa ALSCAL, versão métrica)

na configuração espacial.

Uma das formas mais importantes de análise dos resultados de um E.M.D. é feita observando simplesmente os pontos, sendo que cada um deles deve indicar o objeto que ele representa. Se forem produzidos resultados interessantes, estes serão geralmente verificados examinando a configuração obtida.

As técnicas de E.M.D. diferem de outras técnicas de Análise Multivariada pelo fato dos dados não serem n pontos observados num espaço r -dimensional e sim apenas uma função desses pontos, ou seja, as $\frac{1}{2} n(n-1)$ distâncias entre esses pontos. Usualmente são distâncias empíricas.

As origens das aplicações do E.M.D. encontram-se no campo da Psicologia, daí advém o fato dos objetos a serem escalonados constituírem o que habitualmente chamamos de estímulos.

1.2 - APLICABILIDADE

No entanto o E.M.D. não se prendeu apenas às suas origens e atualmente é aplicado em várias áreas de pesquisa.

Na Psicologia, o E.M.D. é usado para entender a avaliação e a percepção de estímulos auditivos (tais como

palavras e tons musicais), estímulos visuais (tais como cores e faces) e entidades sociais (tais como traços de personalidade e situações sociais) (Torgerson 1958, Shepard 1962, Rosenberg 1968, Young 1970).

Sociólogos usam técnicas de E.M.D. para determinar a estrutura de grupos e organizações, baseados nas percepções que os membros tem uns dos outros, e nas interações de comportamento (Shepard 1972, Funk e outros, 1976).

Antropólogos usam essas técnicas para comparar diversos grupos culturais baseados em suas crenças, linguagens e artefatos (Shepard 1972).

Economistas e pesquisadores em mercadologia utilizam-se de técnicas de E.M.D. para investigar as reações dos consumidores com relação a uma variada classe de produtos. Comparando-se inclusive produtos novos com produtos já existentes no mercado. Conhecem-se pesquisas com produtos farmaceuticos (Neidell 1969), classificação de revistas, comparação de grandes firmas (Green e Maheshwari 1969), programas de T.V. (Green e Carmone 1969) e muitas outras.

Pesquisadores Educacionais usam essas técnicas para estudar a estrutura da inteligência baseados em várias baterias de testes e na ambientação em salas de aula (Shepard 1972).

Pode-se encontrar aplicações de E.M.D. em outras

aéreas não citadas aqui (Shepard 1972, Hodson 1966).

O E.M.D. pode ser usado para adotar-se uma estrutura unidimensional num conjunto de dados, isto é, quando esperamos que os dados sejam parametrizados por um único eixo. O exemplo mais comum é a seriação onde queremos descobrir a ordem cronológica dos dados. Observa-se que embora o E.M.D. possa ser usado para ordenar os dados no tempo, a direção do tempo deve ser determinada independentemente. (Mardia 1979)

1.3 - DESCRIÇÃO DO TRABALHO

O presente trabalho se propõe a ser um guia que permita a um pesquisador aplicar o E.M.D. quando julgar necessário. Define e descreve o E.M.D., classifica os diversos tipos de E.M.D., apresenta algumas técnicas para a resolução do problema e finalmente sugere possíveis análises dos resultados.

Estão descritas neste trabalho as várias etapas que devem ser percorridas para se aplicar uma técnica de E.M.D., desde a coleta dos dados até a análise dos resultados.

A nossa aplicação foi de E.M.D. a "2 fatores", e a idéia inicial era se ater a esse tipo de Escalonamento, no entanto, a tendência de todos os autores atualmente é a de

lidar com o E.M.D. a "3 fatores", portanto resolvemos descreverlo também; além de que o programa utilizado na aplicação está habilitado para o E.M.D. com diferenças individuais.

A escolha da medida de proximidade define o E.M.D. a ser aplicado, assim sendo, no Capítulo 2 apresentamos essas medidas e fizemos uma coletânea de algumas delas, utilizadas pelos diversos pesquisadores em aplicações do E.M.D.

No Capítulo 3 há um breve histórico do E.M.D. a "2 fatores", dando um maior destaque ao método clássico de E.M.D. que é o que lança as bases para as técnicas atuais.

No Capítulo 4 foi descrita a técnica de Kruskal para E.M.D. a "2 fatores", pois através dela pode-se entender o procedimento todo de E.M.D.. Além disso, ela é de grande importância por introduzir uma medida de ajuste nos métodos de E.M.D.

No Capítulo 5 há um relato das técnicas de E.M.D. a "3 fatores", salientando aquelas mais divulgadas.

O Capítulo 6 descreve a técnica ALSCAL (Takane, Young e de Leeuw, 1977) de E.M.D. a "3 fatores" utilizada por nós na aplicação do E.M.D..

O Capítulo 7 se preocupa com a análise dos resultados de um E.M.D., prendendo-se novamente ao E.M.D. a "2 fatores".

No Capítulo 8 está relatada a nossa aplicação, de

E.M.D. realizada a pedido do Prof. Arno Engelmann com dados estudados por ele através de outras técnicas estatísticas (Engelmann, 1978).

Finalmente no Capítulo 9 são tecidos alguns comentários a respeito do trabalho.

Foi deixada em apêndice a parte matemática evitando assim um truncamento numa primeira leitura do texto.

CAPÍTULO 2

MEDIDAS DE PROXIMIDADE

2.1 - INTRODUÇÃO

Neste capítulo vamos tratar da primeira etapa do E.M.D., a coleta das medidas de proximidade, ou seja, a coleta dos dados a serem escalonados. O assunto será tratado sob duas formas. A primeira delas mais formal para estabelecer os conceitos, e a outra prática, de modo a servir como guia numa aplicação de E.M.D.

As medidas de proximidade podem se apresentar em quatro níveis de mensuração, sendo que para cada nível há técnicas adequadas de E.M.D.

O nível considerado mais baixo (Roscoe, 1975) é o nominal que consiste simplesmente na classificação das obser

vações em categorias.

O nível de mensuração ordinal, chamado também de escala ordinal distingue-se do nível nominal por existir uma ordem entre as categorias, isto é uma categoria pode ser posta em posição mais alta ou mais baixa com relação a outra. Quando medidas de proximidade estiverem numa escala ordinal a sua magnitude não tem importância, o fundamental é a sua ordenação. Numa escala ordinal não há especificação alguma com relação à magnitude do intervalo entre duas medidas.

O nível intervalar de mensuração distingue-se do ordinal por haver intervalos iguais entre as unidades de medida. Isto significa que a medida 50 pontos está na metade do intervalo de 40 a 60 pontos. Uma vantagem desta escala é que a estas medidas podem ser aplicadas as quatro operações aritméticas (adição, subtração, multiplicação e divisão). A escala intervalar é uma escala realmente quantitativa, porém ela não tem um zero real. Isto significa que não se pode interpretar a medida de 50 pontos como indicando duas vezes mais que a medida de 25 pontos.

A escala razão tem todas as propriedades da escala intervalar acrescidas da existência de um zero real. Numa escala razão pode-se afirmar que a medida zero significa ausência de um determinado atributo e além disso pode-se afirmar que uma medida representa três vezes mais que outra.

As medidas de proximidade quando se apresentam numa escala intervalar podem ser transformadas para uma escala razão mediante o acréscimo de uma constante (constante aditiva) (Torgerson 1962, Young 1977).

2.2 - CONSIDERAÇÕES MATEMÁTICAS

Definição 2.2.1: Medida de Similaridade: Uma medida de similaridade razoável δ_{ij} (similaridade entre o estímulo i e o estímulo j) deve obedecer às seguintes propriedades:

- (i) $\delta_{ij} = \delta_{ji}$ e $\delta_{ii} = \delta_{jj}$
- (ii) $\delta_{ij} > 0$
- (iii) δ_{ij} cresce quando a similaridade entre i e j aumenta.

Definição 2.2.2: Matriz de Similaridades: Uma matriz Δ é chamada matriz de similaridades se $\delta_{ij} = \delta_{ji}$ e se $\delta_{ij} \leq \delta_{ii}$ $\forall i, j$.

Definição 2.2.3: Medida de Dissimilaridade: Definindo-se a medida de dissimilaridade entre o estímulo i e o estímulo j como: $\delta^*_{ij} = \delta_{ii} - \delta_{ij}$ ela deve obedecer às seguintes propriedades:

- (i) $\delta^*_{ij} > \delta^*_{ii} = 0$

$$(ii) \quad \delta^*_{ij} = \delta^*_{ji}$$

(iii) δ^*_{ij} decresce quando a similaridade entre i e j aumenta.

Definição 2.2.4: Matriz de Distâncias: Uma matriz D de ordem $(n \times n)$ é chamada matriz de distâncias se for simétrica, $d_{ii} = 0$ e $d_{ij} \geq 0$, $i \neq j$.

As distâncias definidas segundo 2.2.4 constituem uma semi-métrica, para transformá-las numa métrica (portanto para tornar métrico o espaço dos estímulos) elas devem obedecer a desigualdade triangular.

$$d_{ij} \leq d_{ik} + d_{kj} \quad \forall i, j, k$$



Se as medidas de proximidade (similaridades ou dissimilaridades) estiverem apenas no nível de mensuração ordinal é razoável assumir que na classe das funções monotônicas exista pelo menos uma que possa transformá-las em distâncias.

De fato, para um conjunto finito de pontos isto é trivial, basta definir a função monotônica como sendo a adição de uma constante suficientemente grande às medidas de dissimilaridade (δ^*_{ij}) e deixando-se $\delta^*_{ii} = 0$. A menor constante que funciona é :

$$C_{\min} = \max_{i,j,k} (\delta^*_{ij} - \delta^*_{ik} - \delta^*_{kj})$$

a partir daí pode-se definir d_{ij} por :

$$\begin{cases} d_{ij} = \delta^*_{ij} + C_{\min} & p/ i \neq j \\ d_{ii} = 0 & p/ i = j \end{cases}$$

Demonstração: $d_{ij} \leq d_{ik} + d_{kj}$

$$\delta^*_{ij} + C \leq \delta^*_{ik} + C + \delta^*_{kj} + C$$

$$\delta^*_{ik} + C + \delta^*_{kj} \geq \delta^*_{ij}$$

$$C \geq \delta^*_{ij} - \delta^*_{ik} - \delta^*_{kj}$$

Portanto $C_{\min} = \max_{i,j,k} (\delta^*_{ij} - \delta^*_{ik} - \delta^*_{kj})$

Certamente alguma constante $C > C_{\min}$ pode também converter as medidas de dissimilaridade δ^*_{ij} em distâncias métricas, mas C_{\min} é o menor valor que pode executar a transformação. De fato C_{\min} é uma das estimativas da constante aditiva adotada para converter distâncias comparativas (escala intervalar) em distâncias (escala razão) no método clássico de E.M. métrico (Torgerson 1962)

2.3 - MEDIDAS DE PROXIMIDADE

2.3.1: Introdução

Ao se realizar um trabalho que necessite a aplicação de alguma técnica de E.M.D. a primeira etapa a ser considerada é a definição da medida de proximidade entre um conjunto de estímulos.

Na prática, uma "medida de proximidade" deve ser um valor que indique o grau de semelhança ou disparidade real entre dois estímulos, ou então, o grau no qual eles são percebidos semelhantes ou distintos por determinados sujeitos.

Os estímulos já usados por experimentadores são os mais variados, cores, sinais de Código Morse, revistas, produtos farmacêuticos, nações, jornais, candidatos políticos, descobertas arqueológicas, etc...

Há vários procedimentos para se obter as medidas de proximidade, um procedimento comum consiste em pedir a determinados sujeitos que julguem diretamente a "distância psicológica" (ou a aproximação) entre os estímulos. Por exemplo, o sujeito pode atribuir uma nota numa escala de 1 a 9 para o grau de semelhança entre duas personalidades políticas.

Além dos termos "similaridades" e "dissimilaridades" geralmente usados para designar as "medidas de proximidade" pode-se usar palavras alternativas tais como: relação, dependência, associação, complementação, substituição, etc...

Os atributos segundo os quais os sujeitos devem julgar os estímulos não lhe costumam ser revelados, sendo que

muitas vezes a descoberta desses atributos é um dos objetivos do experimentador.

Em certos tipos de experimentos psicológicos, como por exemplo percepção sonora ou visual, obtém-se como medida de proximidade, "confundimento entre estímulos". Esta medida é obtida quando a tarefa do sujeito consiste em julgar se dois estímulos apresentados a ele são iguais ou diferentes. A percentagem de respostas "iguais" para um par de estímulos fisicamente distintos é a "medida de confundimento" (Shepard 1963, Wish 1967).

Outra medida de proximidade é a chamada "confusão de resposta" é obtida quando o sujeito deve identificar um estímulo apresentado a ele. A percentagem de vezes que o sujeito responde "j" quando lhe é apresentado o estímulo "i" é a medida de "confusão de resposta". Um exemplo dessa medida pode ser obtido em experimentos onde são apresentados aos sujeitos os símbolos do código Morse (Shepard 1974, Wish e Carroll 1974).

A quantidade de comunicação e interação entre indivíduos, grupos, ou outras entidades pode ser considerada como uma medida de sua proximidade. Por exemplo, pelo número de conversas ou encontros entre esses indivíduos. A aplicação de alguma técnica de E.M.D. a esses dados fornece um mapa sociométrico, no qual, grandes distâncias refletem uma me-

nor associação entre as entidades. Da mesma forma, pode-se obter por técnicas de E.M.D aplicadas ao tráfego telefônico, volume de viagens ou quantidade de outras transações de uma cidade para outras, um mapa de comunicações ou então de corrente de informações entre cidades (Slater 1976, Coombs 1964). Estes dados devem ser "processados" antes da aplicação da técnica de E.M.D para remover o efeito do tamanho das cidades sobre os dados.

Uma forma muito comum de se obter "medidas de proximidade" é através de tabelas que forneçam dados quantitativos relativos aos estímulos, sendo que a medida de proximidade é derivada desses dados. Por exemplo, as linhas da tabela original podem corresponder a várias nações e as colunas podem ser medidas do produto nacional bruto, energia consumida, nº de partidos políticos, etc... A forma mais comum de se derivar uma medida de proximidade de perfil é se computar as correlações entre as variáveis ou as distâncias quadráticas entre os estímulos (Wish e Carroll 1974). Apesar das correlações serem normalmente analisadas por técnicas de Análise Fatorial e procedimentos análogos (Banks e Gregg 1965, Rummel 1969, Sawyer, 1967) elas também podem ser consideradas "medidas de proximidade" e ser analisadas por técnicas de E.M.D. (Guttman 1966, Weisberg e Rusk 1970).

Uma vez obtidas as medidas de proximidade elas são dispostas em matrizes, nas quais cada elemento δ_{ij} represen

ta a proximidade entre o estímulo i e o estímulo j .

Para aplicação de técnicas de E.M.D. a "3 fatores" (vide Cap. 5) são construídas matrizes individuais Δ_k onde cada elemento δ_{kij} representa a medida de proximidade entre os estímulos i e j , julgada pelo sujeito k . No caso de aplicação de E.M.D. a "2 fatores" (vide Cap. 3) é construída uma única matriz Δ , onde cada elemento δ_{ij} representa a medida de proximidade entre o estímulo i e o estímulo j , julgada pelo grupo de sujeitos.

2.3.2: Alguns Procedimentos Para Obtenção de Medidas de Proximidade

2.3.2.1: Tríades

A: Método das combinações triádicas (Torgerson, 1962)

Sejam dados n estímulos, eles são apresentados aos sujeitos em tríades, num total de $C_{n,3} = \frac{n(n-1)(n-2)}{6}$ tríades (grupos de 3 estímulos)

Ao lhe ser apresentada a tríade (i, j, k) o sujeito deverá indicar dentre os três estímulos, quais são os dois mais semelhantes e quais são os dois mais diferentes.

A partir de cada resposta do sujeito são inferidos três julgamentos. Por exemplo, se é apresentada a tríade (i, j, k) e o sujeito responde: i e j são os mais parecidos e

i e k são os mais diferentes, podemos concluir que

- (1) i é mais parecido com j do que com k
- (2) j é mais parecido com i do que com k
- (3) k é mais parecido com j do que com i

Depois de apresentadas todas as tríades a todos os sujeitos, construímos para cada estímulo k a matriz $[k^{p_{ij}}]$ onde $k^{p_{ij}}$ é a proporção de vezes que o estímulo k é julgado mais parecido com o estímulo i do que com o estímulo j. Os elementos dessas matrizes são então processados (Torgerson 1962) até que sejam obtidas medidas de distância entre os estímulos numa escala razão.

B: Método Completo das tríades

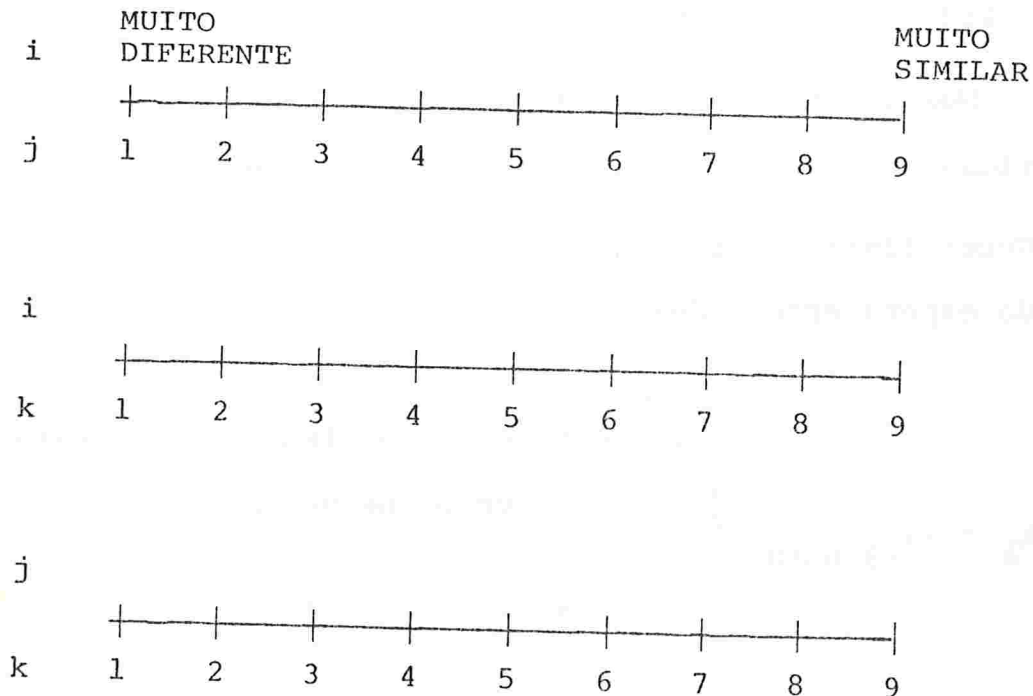
Neste método cada tríade (i, j, k) é apresentada três vezes ao sujeito, em cada apresentação obtém-se diretamente do sujeito uma das conclusões acima. Por exemplo, considere-se uma dada tríade (i, j, k), na primeira vez em que ela lhe é apresentada o sujeito deverá julgar se i é mais parecido com j ou com k; na 2ª vez, deverá julgar se j é mais parecido com i ou com k; e na 3ª vez se k é mais parecido com i ou com j. A partir destes julgamentos realizados por todos os sujeitos com relação a todas as possíveis tríades são construídas as matrizes $[k^{p_{ij}}]$ analogamente ao método anterior.

2.3.2.2 Método dos pares (Wish e Carroll 1973)

Os n estímulos são agrupados em pares num total de $C_{n,2} = \frac{n(n-1)}{2}$ pares.

Os pares são apresentados aos sujeitos sob forma de questionário ou em entrevistas individuais. Cada sujeito deve atribuir uma nota de similaridade a cada par numa escala que costuma ser de 9 ou 10 pontos.

EXEMPLO 2.1: Sejam os estímulos i , j e k . Obtêm-se os pares (i,j) ; (i,k) e (j,k) . A cada par associa-se uma escala na qual o sujeito assinala (X) sobre a nota que resolve atribuir ao par.



A medida de similaridade δ_{ij} pode ser obtida através da média ou da mediana das notas atribuídas por todos os sujeitos ao par de estímulos (i, j) .

2.3.2.3: Agrupamento em categorias

Os n estímulos são apresentados a cada sujeito, e a tarefa dele consiste em dividir esses estímulos em categorias mutuamente exclusivas, de modo que os estímulos pertencentes a cada categoria sejam similares entre si. Pode-se em caso de necessidade, estabelecer limites máximo ou mínimo para o número de categorias.

Este método é recomendado quando o experimentador estiver lidando com um número grande de estímulos.

Neste método as medidas de similaridade δ_{ij} podem ser obtidas de diversas formas, tem-se abaixo algumas:

(i) Green (1969) constrói uma matriz $A(n \times n)$ para cada sujeito do experimento onde:

$$A_k = (a_{ij})_{n \times n} \left\{ \begin{array}{l} a_{ij} = 1 \quad \text{se os estímulos } i \text{ e } j \text{ estiverem na mesma categoria} \\ a_{ij} = 0 \quad \text{caso contrário} \end{array} \right.$$

Em seguida efetua a soma de todas as matrizes A_k obtendo assim a matriz $\Delta = (\delta_{ij})_{n \times n}$ onde δ_{ij} é considerada a medida de similaridade entre o estímulo i e o estímulo j .

(ii) Burton (1972) usa princípios da Teoria da Informação para chegar à matriz de similaridades. Em seu experimento trabalha com sessenta palavras que discriminam ocupações, ele pretende um estudo semântico desses termos. O método usado por ele está descrito abaixo:

Cada categoria de estímulos identificada por um sujeito recebe o nome de "célula" da partição.

Para cada sujeito define-se B_r como sendo o n^o de células obtidas e $C_{r,\alpha}$ o n^o de estímulos da célula α para o sujeito r .

Escolhendo-se dois estímulos ao acaso, a probabilidade deles estarem numa determinada célula α é :

$$P_{r,\alpha} = \frac{C_{r,\alpha} (C_{r,\alpha} - 1)}{60.59} = \frac{\text{n}^\circ \text{ de pares de estímulos da célula } \alpha}{\text{n}^\circ \text{ total de pares de estímulos}}$$

Considera-se $-\log_2 P_{r,\alpha}$ a informação contida no evento: "Estímulos i e j estão ambos na célula α " e $-\log_2 Q_r$ a informação contida no evento: "Estímulos i e j estão em células diferentes". Sendo

$$Q_r = 1 - \sum_{\alpha=1}^{\Delta_r} P_{r,\alpha}$$

A medida de similaridade proposta considera o evento "i e j estão na célula α " como um incremento positivo para o grau de similaridade entre os estímulos i e j , e considera o evento "i e j estão em células diferentes" como um decréscimo do grau de similaridade entre eles.

Define-se inicialmente uma medida de similaridade entre os estímulos i e j para o sujeito r : $\delta_{i,j,r}$ onde :

$$\delta_{i,j,r} = \begin{cases} - \log_2 P_{r,\alpha} & : \text{quando os estímulos } i \text{ e } j \text{ estão na} \\ & \text{mesma célula} \\ + \log_2 Q_r & : \text{quando os estímulos } i \text{ e } j \text{ estão em} \\ & \text{células diferentes} \end{cases}$$

Esta medida permite compensar as diferenças de tamanhos de células nos diversos sujeitos, pois quando um sujeito coloca dois estímulos numa célula muito pequena o evento tem uma alta informação (provavelmente ele os considera extremamente similares) e há relativamente um grande incremento na medida de similaridade destes dois estímulos; da mesma forma quando o sujeito localiza os dois estímulos em células diferentes porém Q_r é pequena (baixa probabilidade dos 2 estímulos estarem em células diferentes) há então um grande decréscimo na medida de similaridade (certamente para este sujeito os estímulos parecem bastante diferentes).

Para dar pesos iguais a todos os sujeitos é neces

sário normalizar os acréscimos e decréscimos propostos.

Define-se para cada sujeito r :

$$e_r = - \sum_{\alpha=1}^{\Delta_r} P_{r,\alpha} + Q_r \log_2 Q_r$$

e_r = (média da informação do evento "i e j juntos") - (média da informação do evento "i e j separados")

Define-se também a variância :

$$\tau_r^2 = \sum_{\alpha=1}^{\Delta_r} P_{r,\alpha} \left[\log_2 P_{r,\alpha} \right]^2 + Q_r \left[\log_2 Q_r \right]^2$$

Obtêm-se a seguir a medida final de similaridade entre os estímulos i e j

$$\delta_{ij} = \sum_{r=1}^R \frac{(\delta_{i,j,r} - e_r)}{\left(\sum_{r=1}^R \tau_r^2 \right)^{1/2}}$$

R: nº de sujeitos que participam do experimento

Pode-se notar que uma modificação desta medida da forma $\delta_{ij} = \sum_{r=1}^R \frac{\delta_{ijr} - e_r}{\tau_r}$ estaria mais de acordo com o espírito do procedimento de normalização. No entanto esta versão resulta de uma dedução errônea que acredita produzir uma medida assintoticamente normal. Como a medida não é uma soma de eventos independentes (todos os julgamentos são oriundos dos mesmos dois estímulos e portanto não são independentes entre si) esta dedução não é válida. De fato a distribui

ção das medidas de similaridade entre dois estímulos é altamente viesada com uma alta concentração no valor mínimo. Assim sendo, a distribuição reflete o fato de haver vários pares de estímulos que são tão diferentes que aparecem em células distintas na partição de cada sujeito.

(iii) Wish e Carroll (1973) constroem a matriz $\Delta = (\delta_{ij})_{n \times n}$ medindo a proporção de vezes que os estímulos i e j se apresentam na mesma categoria.

$$\delta_{ij} = \frac{S}{R} = \text{medida de similaridade entre o estímulo } i \text{ e o estímulo } j$$

R: nº total de sujeitos do experimento

S: nº de sujeitos que classificaram os estímulos i e j na mesma categoria.

2.3.2.4: Método do Ponto de Âncora (Green e Carmone 1972)

Os n estímulos são apresentados n vezes a cada um dos sujeitos do experimento, sendo que em cada vez o experimentador escolhe um dos estímulos como "ponto de âncora" isto é, o sujeito deve ordenar os demais $(n-1)$ estímulos de acordo com o seu grau de semelhança em relação ao estímulo "ponto de âncora". Cada um dos n estímulos será sucessivamente escolhido para desempenhar o papel de "ponto de âncora".

EXEMPLO 2.2: São apresentados ao sujeito os estímulos i, j, k, ℓ . Escolhe-se como primeiro "ponto de âncora" o estímulo i , a partir dele o sujeito dá a ordenação (j, k, ℓ) isto significa que na opinião deste sujeito o estímulo j é o mais parecido com o estímulo i e o estímulo ℓ é o mais diferente. Em seguida é escolhido outro "ponto de âncora" e o sujeito fornece a respectiva ordenação. As ordenações de um determinado sujeito estão exemplificadas abaixo:

<u>"ponto de âncora"</u>	<u>ordenação</u>
i	(j, k, ℓ)
j	(i, k, ℓ)
k	(i, j, ℓ)
ℓ	(k, j, i)

Há um programa para computador, o Tricon, desenvolvido por Green, Carmone e Robinson (1968) que a partir das ordenações obtém as medidas de dissimilaridades, para cada sujeito do experimento.

2.3.2.5: Outros Métodos

Estão selecionados abaixo alguns dentre os vários outros métodos de coleta de dados adotados pelos diversos pesquisadores.

(i) Frequências de semelhanças: Apresenta-se a cada su-

jeito uma lista com os n estímulos. Escolhe-se um estímulo i dentre os n apresentados e pede-se ao sujeito que identifique entre os $(n-1)$ estímulos restantes algum que se assemelhe ao estímulo i segundo alguma característica. O sujeito dará como resposta um estímulo j . Cada um dos n estímulos fará sucessivamente o papel de estímulo i .

A partir das respostas dos sujeitos constrói-se uma matriz $[f_{ij}]$ de frequências onde cada elemento f_{ij} corresponde à frequência na qual o estímulo i foi considerado similar ao estímulo j . A partir dos f_{ij} calcula-se s_{ij} , índice de similaridade dado por :

$$s_{ij} = \frac{f_{ij} + f_{ji}}{f_{i.} + f_{j.} + f_{.i} + f_{.j}}$$

Estes índices s_{ij} são então normalizados para se obter as medidas de similaridade (δ_{ij}) (Mauser, 1972)

(ii) Caracteres concordantes: Este método foi usado por Sueath e Doran (1966) num experimento com materiais arqueológicos.

Os n estímulos são estudados aos pares, a luz de alguns caracteres que interessam ao experimentador. O que interessa é se o caracter está ou não presente no estímulo, a partir daí é calculada a medida de similaridade δ_{ij} .

Tomando-se como exemplo os estímulos i e j e os ca

racteres A, B, C e D. Atribui-se um sinal (+) ao estímulo se ele tiver o caracter e o sinal (-) caso contrário. Mede-se a proporção de sinais concordantes, como na tabela abaixo:

<u>caracteres</u> <u>estímulos</u>	A	B	C	D
i	+	-	+	-
j	+	+	+	-

Neste exemplo $\delta_{ij} = \frac{3}{4}$

(iii) Em experimentos com estímulos visuais, auditivos ou outros estímulos sensoriais, estes podem ser apresentados aos sujeitos em pares, sendo tarefa de cada sujeito, classificá-los em iguais ou diferentes. A medida de similaridade (δ_{ij}) é obtida através da proporção de vezes que os estímulos "i e j foram considerados iguais pelos sujeitos do experimento (Rothkopf, 1957)

Outra forma de se obter as medidas de similaridade ao lidar com este tipo de estímulo consiste em apresentar a cada sujeito o estímulo i e pedir a ele que o identifique. A medida de similaridade entre o estímulo i e o estímulo j é obtida pelo nº de vezes que os sujeitos responderem j ao observar o estímulo i. Cada sujeito deve observar os n estímulos sucessivamente (Shepard 1974, Wish e Carroll, 1974)

O livro de Shepard (1972) apresenta uma série de aplicações de E.M.D. e alí encontram-se várias outras medi-das de proximidade.

CAPÍTULO 3

ALGUNS MÉTODOS DE E.M.D. A 2 FATORES

3.1 - INTRODUÇÃO

Nós vamos desenvolver neste capítulo alguns métodos de E.M.D. a 2 fatores.

Dentre estes E.M.D. há uma classificação em E.M.D. métricos e E.M.D. não métricos.

Esta classificação foi introduzida por Coombs (1958) no entanto foi elaborada por Kruskal (1964 a) que chamou de E.M.D. não métrico o que Shepard (1962 a.b) havia chamado "análise de proximidades".

Na conceituação atual há dois aspectos a serem considerados para classificar o E.M.D. em métrico ou não-métrico.

O 1º deles é o nível de mensuração dos dados. Se eles estiverem em escala razão portanto com propriedades de distâncias ou na escala intervalar podendo ser transformadas em distância mediante a adição de uma constante, o E.M.D. é métrico. Com dados em outros níveis de mensuração o E.M.D. é não métrico, e a matriz de dados não é de distâncias e sim de dissimilaridades ou similaridades.

O 2º aspecto diz respeito ao procedimento utilizado para a determinação da configuração. O E.M.D. métrico admite a existência de uma "verdadeira" configuração em k dimensões cujas distâncias entre os pontos são v_{rs} . Pretende-se reconstruir esta configuração usando a matriz de distâncias observadas $\Delta = (\delta_{rs})$ cujos elementos são da forma

$$\delta_{rs} = v_{rs} + e_{rs}$$

onde e_{rs} representam erros de medida acrescidos de erros de distorções pelo fato das distâncias observadas não corresponderem exatamente às distâncias de uma configuração em R^k . O E.M.D. não métrico numa atitude mais realista, toma como hipótese uma relação menos rígida entre δ_{rs} e v_{rs} , como por exemplo, supondo

$$\delta_{rs} = f(v_{rs} + e_{rs})$$

sendo f uma função desconhecida monotônica crescente. Neste caso a única informação a ser usada para a reconstrução da configuração é a ordenação das δ_{rs} .

As bases teóricas do E.M.D. foram lançadas por Young e Householder (1938) que demonstraram teoremas que fornecem as condições para se acomodar distâncias em um Espaço Euclidiano real, e determinam a dimensão mínima desse espaço.

Richardson (1938) e Klingberg (1941) usaram estes conhecimentos nas primeiras aplicações conhecidas de E.M.D., no entanto a metodologia só começou a se desenvolver a partir dos anos 50 juntamente com o surgimento dos computadores que permitiram a Torgerson (1958), Messick e Abelson (1956) e outros lidarem com um grande número de dados e com espaços de dimensão alta.

Neste capítulo descreveremos o método de E.M.D. não métrico de Torgerson conhecido por método clássico, e também comentaremos os métodos de E.M.D. não métrico de Coombs, Shepard e Kruskal.

3.2 - MÉTODO CLÁSSICO

3.2.1: Definições e Teoremas

Definição 3.1: E.M.D. Métrico: Partindo de uma matriz $D = [d_{ij}]$ de distâncias, o objetivo do E.M.D. métrico é encontrar n pontos x_1, x_2, \dots, x_n num espaço k dimensional de modo que se $\hat{d}_{ij} = d(x_i, x_j)$ então a matriz $\hat{D} = [\hat{d}_{ij}]$

deva estar o "mais próxima possível" da matriz D.

Os pontos x_i são desconhecidos e geralmente a dimensão k também não se conhece.

Definição 3.2: Matriz Euclidiana: Uma matriz de distâncias D é dita Euclidiana se existir uma configuração de pontos num Espaço Euclidiano, cujas distâncias entre os pontos sejam dadas por D, isto é, se para um dado k , existirem pontos $x_1, x_2, \dots, x_n \in \mathbb{R}^k$ tal que

$$d_{rs}^2 = (x_r - x_s)' (x_r - x_s)$$

Os teoremas abaixo permitem verificar se a matriz D é Euclidiana e se for permitem descobrir a configuração de pontos correspondente.

Vamos inicialmente introduzir a notação.

Para uma matriz de distâncias D seja:

$$A = [a_{rs}] \quad \text{onde} \quad a_{rs} = -\frac{1}{2} d_{rs}^2$$

$$B = H.A.H \quad \text{onde} \quad H = I - n^{-1} \underline{1}.\underline{1}'$$

onde I é a matriz identidade e $\underline{1}$ o vetor coluna de 1, H de ordem $(n \times n)$: matriz "centrada"

$$b_{rs} = a_{rs} - \frac{1}{n} \sum_{s=1}^n a_{rs} - \frac{1}{n} \sum_{r=1}^n a_{rs} + \frac{1}{n^2} \sum_{r=1}^n \sum_{s=1}^n a_{rs}$$

$$b_{rs} = a_{rs} - \bar{a}_{r.} - \bar{a}_{.s} + \bar{a}_{..}$$

Teorema 3.1: Seja D a matriz de distâncias, define-se $B = H \cdot A \cdot H$. Nestas condições D é Euclidiana se e somente se B for positiva semi-definida.

Em particular verificam-se as afirmações:

(a) Se D é a matriz das distâncias Euclidianas entre os pontos da configuração $Z = (z_1, \dots, z_n)'$, então $b_{rs} = (z_r - \bar{z})' \cdot (z_s - \bar{z})$ $r, s = 1, \dots, n$. Na forma matricial ter-se-ia $B = (HZ) \cdot (HZ)^1$ sendo assim $B \geq 0$. Nota-se que B pode ser interpretada como "matriz dos produtos internos centrados" para a configuração Z .

(b) Se B for positiva semi-definida de posto k , a configuração correspondente a B pode ser construída da seguinte forma. Sejam $\lambda_1 > \dots > \lambda_k$ as raízes características positivas de B e $X = (x_{(1)}, \dots, x_{(k)})$ onde $x'_{(i)} = (x_{1i}, x_{2i}, \dots, x_{ni})$ seus vetores característicos normalizados por:

$$x'_{(p)} x_{(p)} = \lambda_p \quad p = 1, 2, \dots, k$$

Então os pontos x_i em \mathbb{R}^k com coordenadas $x_i = (x_{i1}, \dots, x_{ik})'$ (sendo x_i a i -ésima linha de X) tem as distâncias entre eles dadas pela matriz D . Além disto esta configuração tem centro de gravidade $\bar{x} = 0$ e B representa a matriz do produto interno para esta configuração. (Demonstração - vide apêndice A.1).

Observações:

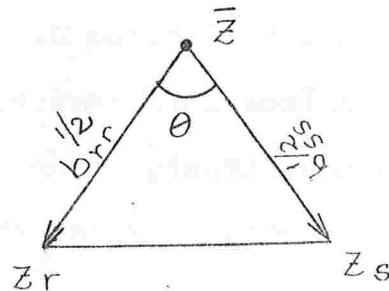
1) A matriz configuração X pode ser visualizada da seguinte forma em função dos vetores características de B e os pontos correspondentes

	raízes características			
	λ_1	λ_2	\dots	λ_k
pontos x' (1)	x_{11}	x_{12}	\dots	x_{1k}
x' (2)	x_{21}	x_{22}	\dots	x_{2k}
	.	.	\dots	.
	.	.	\dots	.
	.	.	\dots	.
x' (n)	x_{n1}	x_{n2}	\dots	x_{nk}

Centro de gravidade $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{0}$

Resumindo a r-ésima linha da matriz X contém as coordenadas do r-ésimo ponto da configuração e a i-ésima coluna de X contém o vetor característico correspondente à raiz característica λ_i .

2) Geometricamente se B é a matriz dos produtos internos centrados para a configuração Z, então $b_{rr}^{1/2}$ é igual à distância entre z_r e \bar{z} e $\frac{b_{rs}}{(b_{rr} b_{ss})^{1/2}}$ é igual ao cosseno do ângulo formado em \bar{z} entre z_r e z_s



Justificativa: $(z_r - z_s)' (z_r - z_s) = b_{rr} + b_{ss} - 2b_{rr}^{1/2} b_{ss}^{1/2} \cos \theta$

3) Observe-se que o vetor l é vetor característico de B seja ou não D uma matriz Euclidiana.

4) O teorema não é válido se B tiver raízes características negativas ($x'(i) x_{(i)} = \lambda_i$), é impossível normalizar-se um vetor de modo a obter uma norma quadrática negativa (Mardia, Kent - Bibby 1979)

3.2.2: Solução do E.M.D. Métrico

3.2.2.1: Solução a partir de uma matriz de distâncias

Uma matriz D que represente aproximadamente as distâncias entre os pontos de uma configuração num Espaço Euclidiano, com uma dada dimensão, pode ou não ser Euclidiana, no entanto, mesmo que D seja Euclidiana, a dimensão do espaço no qual ela pode ser representada é geralmente muito alta para interesses práticos. Uma escolha possível para configuração em k dimensões é sugerida pelo Teorema (3.1). "Escolhe-se a configuração em R^k cujas coordenadas são determi

nadas pelos k primeiros vetores característicos de B ". Se as k primeiras raízes características de B forem "grandes" positivas, e as demais raízes características forem próximas de 0 (positivas ou negativas), então felizmente, as distâncias entre os pontos da configuração obtida estarão bem próximas da matriz D .

Esta configuração é chamada "solução clássica do problema de E.M.D. em k dimensões" e se encaixa na definição de solução métrica.

Para efeitos de programação de computador tem-se abaixo um sumário dos cálculos envolvidos.

1. A partir de D constrói-se a matriz $A = (-\frac{1}{2} d_{rs}^2)$

2. Obtém-se a matriz B cujos elementos são:

$$b_{rs} = a_{rs} - \bar{a}_{r.} - \bar{a}_{.s} + \bar{a}_{..}$$

3. Encontram-se as k maiores raízes características $\lambda_1 > \dots > \lambda_k$ de B (k previamente escolhido) e seus vetores característicos correspondentes $X = (x_{(1)}, \dots, x_{(k)})$ que são então normalizados por $x'_{(i)} \cdot x_{(i)} = \lambda_i$, $i = 1, \dots, k$ (Supõe-se aqui que as k primeiras raízes características sejam todas positivas)

4. As coordenadas dos pontos procurados são $x_r = (x_{r1}, \dots, x_{rp})'$, $r = 1, 2, \dots, k$ dadas pelas linhas da matriz X .

EXEMPLO 3.1: Para ilustrar o algoritmo considere a matriz de distâncias D abaixo

$$D = \begin{bmatrix} 0 & 1 & \sqrt{3} & 2 & \sqrt{3} & 1 & 1 \\ & 0 & 1 & \sqrt{3} & 2 & \sqrt{3} & 1 \\ & & 0 & 1 & \sqrt{3} & 2 & 1 \\ & & & 0 & 1 & \sqrt{3} & 1 \\ & & & & 0 & 1 & 1 \\ & & & & & 0 & 1 \\ & & & & & & 0 \end{bmatrix}$$

Construindo-se a matriz A encontra-se:

$$\bar{a}_{r.} = -\frac{13}{14} \quad r = 1, \dots, 6 \quad \bar{a}_{7.} = -\frac{3}{7}$$

$$\bar{a}_{r.} = \bar{a}_{.r} \quad \bar{a}_{..} = -\frac{6}{7}$$

De $b_{rs} = a_{rs} - \bar{a}_{r.} - \bar{a}_{.s} + \bar{a}_{..}$ Obtém-se:

$$B = \begin{bmatrix} 2 & 1 & -1 & -2 & -1 & 1 & 0 \\ & 2 & 1 & -1 & -2 & -1 & 0 \\ & & 2 & 1 & -1 & -2 & 0 \\ & & & 2 & 1 & -1 & 0 \\ & & & & 2 & 1 & 0 \\ & & & & & 2 & 0 \\ & & & & & & 0 \end{bmatrix}$$

Pode-se verificar que as colunas de B são linear-

mente dependentes.

$$b_{(3)} = b_{(2)} - b_{(1)} \quad b_{(4)} = -b_{(1)} \quad b_{(5)} = -b_{(2)}$$

$$b_{(6)} = b_{(1)} - b_{(2)} \quad b_{(7)} = 0$$

A análise dessa matriz mostra que ela é de posto 2. Portanto a configuração que se ajusta à matriz de distâncias pode ser construída num espaço bi-dimensional.

As raízes características de B encontradas são:

$$\lambda_1 = 3, \quad \lambda_2 = 3, \quad \lambda_3 = \dots = \lambda_7 = 0$$

A configuração pode ser construída usando dois vetores ortogonais do espaço característico correspondente a $\lambda = 3$, tais como:

$$x'_{(1)} = (a, a, 0, -a, -a, 0, 0) \quad a = \frac{1}{2} \sqrt{3}$$

$$x'_{(2)} = (b, -b, -2b, -b, b, 2b, 0) \quad b = \frac{1}{2}$$

A partir daí as coordenadas dos 7 pontos são:

$$A\left(\frac{1}{2} \sqrt{3}, 1/2\right) \quad B\left(\frac{1}{2} \sqrt{3}, -1/2\right) \quad C(0, -1)$$

$$D\left(-\frac{1}{2} \sqrt{3}, -1/2\right) \quad E\left(-\frac{1}{2} \sqrt{3}, 1/2\right) \quad F(0, 1) \quad G(0, 0)$$

(figura 3.1)

O centro de gravidade desses pontos é certamente $(0, 0)$ e pode-se verificar que a matriz de distâncias para esses pontos é D. De fato, os pontos de A a F são vértices

de um hexágono cujos lados tem comprimento 1 e a linha \overline{FC} é o eixo y. Seu centro é G (na verdade a matriz D foi construída através desses pontos)

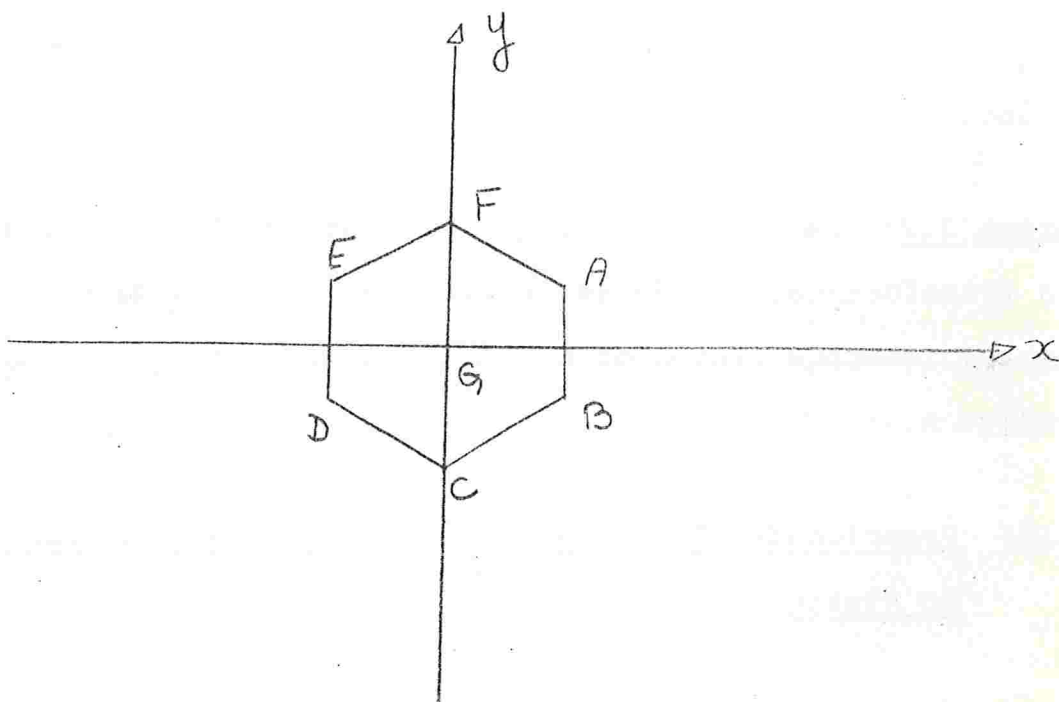


Figura 3.1

3.2.2.2: Solução a partir da matriz de similaridades

Pode-se usar a técnica anteriormente descrita de E.M.D. se as observações constituírem uma matriz de similaridades, para tanto é necessário transformar a matriz de similaridades em matriz de distâncias, através da transformação definida abaixo:

Definição 3.3: A transformação padrão da matriz Δ de simi-

laridades em matriz D de distâncias é definida por

$$d_{rs} = (\delta_{rr} - 2\delta_{rs} + \delta_{ss})^{1/2} \quad (4)$$

Observa-se que sendo $\delta_{rs} \leq \delta_{rr}$ tem-se $\delta_{rr} - 2\delta_{rs} + \delta_{ss} \geq 0$ e $d_{rr} = 0$. Portanto a matriz D é uma matriz de distâncias.

Teorema 3.2: Se $\Delta \geq 0$, a matriz de distâncias D definida pela transformação padrão (4) é Euclidiana, e tem matriz de produtos internos centrados $B = H \Delta H$ (Demonstração - vide apêndice A.2).

3.2.3: Propriedades Ótimas da Solução Clássica e Adequação do Ajuste

Dada uma matriz D, o objetivo do E.M.D. é encontrar uma configuração \hat{X} num espaço Euclidiano de baixa dimensão k , de modo que as distâncias entre os pontos $\hat{d}_{rs}^2 = (\hat{x}_r - \hat{x}_s)' (\hat{x}_r - \hat{x}_s)$ se aproximem ao máximo de D. O circunflexo é usado para indicar que as distâncias \hat{D} entre os pontos da configuração \hat{X} estão ajustadas às distâncias originais D. Da mesma forma, \hat{B} representa a matriz dos produtos internos centrados ajustados.

Seja X a configuração em R^P e seja $L = (L_1, L_2)$ uma matriz ortogonal ($p \times p$) onde L_1 é ($p \times k$). Então XL_1 representa a projeção da configuração X sobre o sub-espaço de R^P

gerado pelas colunas de L_1 . Pode-se considerar $\hat{X} = X \cdot L_1$ como sendo a configuração ajustada em k dimensões.

Sendo L ortogonal, as distâncias entre as linhas de X são iguais às distâncias entre as linhas de $X \cdot L$.

$$d_{rs}^2 = \sum_{i=1}^p (x_{ri} - x_{si})^2 = \sum_{i=1}^p (x'_{r \ell(i)} - x'_{s \ell(i)})^2$$

$\ell(i)$: i -ésima linha da matriz L

Notando as distâncias entre as linhas de $X L_1$ por \hat{D} ,

tem-se:

$$\hat{d}_{rs}^2 = \sum_{i=1}^k (x'_{r \ell(i)} - x'_{s \ell(i)})^2 \rightarrow$$

$$\hat{d}_{rs}^2 - d_{rs}^2 = \sum_{k+1}^p (x'_{r \ell(i)} - x'_{s \ell(i)})^2$$

Portanto, $\hat{d}_{rs} \leq d_{rs}$; isto é, projetando a configuração, a distância entre os pontos fica reduzida. Logo, a medida da discrepância entre a configuração original X e a configuração projetada \hat{X} é dada por:

$$\theta = \sum_{r,s=1}^n (d_{rs}^2 - \hat{d}_{rs}^2)$$

Então a solução clássica do E.M.D. em k dimensões tem as seguintes propriedades ótimas:

Teorema 3.3: Seja D uma matriz de distâncias Euclidianas correspondente à configuração X em R^p , e fixando $k(1 \leq k < p)$. Entre todas as projeções $X L_1$ de X em sub-espacos k dimensionais de R^p , θ é minimizada quando X é projetada sobre suas

coordenadas principais em k dimensões (Demonstração - vide apêndice A.3).

Quando a matriz D não for necessariamente Euclidiana é mais conveniente trabalhar com a matriz $B = H A H$. Se \hat{X} for a configuração ajustada com matriz de produtos internos centrados \hat{B} , então a medida da discrepância entre B e \hat{B} é dada (Mardia, 1978) por:

$$\psi = \sum_{r,s=1}^u (b_{rs} - \hat{b}_{rs})^2 = \text{tr} (B - \hat{B})^2$$

Pode-se provar que a solução clássica do problema de E.M.D. é ótima para esta medida também.

Teorema 3.4: Se D é uma matriz de distâncias (não necessariamente Euclidiana), para um k fixo, ψ atinge seu mínimo para todas as configurações \hat{X} em k dimensões quando \hat{X} for a solução clássica do problema de E.M.D. (Demonstração - vide apêndice A.4).

Os dois últimos teoremas sugerem uma possível "medida de ajuste" para a "proporção da matriz de distâncias que é explicada" pela solução clássica do problema de E.M.D. em k dimensões.

Supondo $\lambda_k > 0$ essas medidas são: (Mardia 1978)

$$\alpha_{1,k} = \frac{\sum_{i=1}^k \lambda_i}{n \sum_{i=1}^k |\lambda_i|} \cdot 100\%$$

$$\alpha_{2,k} = \frac{\sum_{i=1}^k \lambda_i^2}{n \sum_{i=1}^k \lambda_i^2} \cdot 100\%$$

Deve-se usar os módulos em $\alpha_{1,k}$ pois alguns dos valores pequenos das raízes características podem ser negativos.

3.3 - MÉTODOS NÃO MÉTRICOS

Diversos autores pesquisaram o E.M.D. não métrico, no entanto neste capítulo vamos destacar apenas alguns que representaram avanços no desenvolvimento das técnicas de E.M.D.

3.3.1: Método de Coombs

Coombs (1958) desenvolveu o seu método de E.M.D. não métrico baseado em dados de proximidade ordinais, isto é, ele partia de apenas uma ordenação das medidas de proximidade. Obtinha também um espaço não métrico, isto é, a solução consistia na ordenação dos estímulos em cada uma das dimensões do Espaço obtido, sendo impossível calcular-se a distância entre esses pontos, pois para isto seria necessário conhecer pelo menos uma função monotônica para cada dimensão, que transformasse as coordenadas de ordenação em coordenadas de escala intervalar. O procedimento de Coombs é conhecido como "duplamente não métrico" visto que os dados e a solução são não - métricos.

3.3.2: Método de Shepard: Análise de Proximidades

Shepard (1962, a, b) elaborou um algoritmo para computador que permitia obter uma solução métrica a partir de

dados não métricos (ordinais) chamou o seu método de Análise de Proximidades.

Shepard tinha por objetivo uma configuração num espaço de dimensão mínima de modo que as distâncias entre os pontos da configuração estivessem monotonicamente relacionadas com as medidas de proximidade entre os estímulos.

Isto é, Shepard procurava uma configuração de pontos P_1, P_2, \dots, P_n num espaço \mathbb{R}^k , tal que k fosse mínimo e de modo que $d(P_i, P_j) = f(\delta_{ij})$ $i, j = 1, \dots, n$ onde f seria uma função monotônica.

Uma vez colhidas as medidas de proximidade (δ_{ij}) Shepard partia para um método iterativo visando obter a configuração de pontos (P_1, P_2, \dots, P_n) que haveria de representar os estímulos num espaço Euclidiano \mathbb{R}^k .

3.3.2.1: Técnica de Shepard

Vamos descrever a técnica de Shepard, sem nos preocupar com os algoritmos numéricos que ela envolve e sim com as suas idéias que serviram de base para outros procedimentos. Vamos dividi-la em duas etapas:

1ª Etapa:

Escolhia uma configuração inicial, que considerava ter "entropia máxima", um simplex regular em um espaço de

dimensão $(n - 1)$, o equivalente ao triângulo equilátero no espaço bi-dimensional ou ao tetraedro regular no espaço tri-dimensional.

2ª Etapa:

Movia os pontos do simplex respeitando dois objetivos, ou ainda como ele considerava, dois processos: um deles visando diminuir o número de dimensões do espaço chamado processo β , e o outro procurando aumentar o ajuste dos pontos da configuração, à ordenação das medidas de proximidade observadas, chamado processo α .

Iniciava no espaço de maior dimensão possível, baixava a dimensão do espaço pelo processo β e uma vez no espaço de dimensão menor procurava neste espaço a melhor configuração pelo processo α .

Processo β : Sua idéia básica era aumentar as distâncias grandes e diminuir as distâncias pequenas. Para visualizá-la pode-se imaginar um conjunto de pontos em forma de arco portanto localizado num espaço bi-dimensional, para localizar esses pontos num espaço uni-dimensional deve-se retificar o arco, nesta retificação os pontos extremos do arco tenderão a afastar-se e os pontos centrais que já estavam próximos tenderão a unir-se.

O processo cessava quando diminuía ao máximo a di

mensão do espaço e obtinha os pontos no espaço de dimensão "correta" .

Processo α : Este processo era usado numa 1^a fase junto com o processo β e na 2^a fase sozinho quando o processo β atingia o seu objetivo (a dimensão "correta").

Sua idéia básica era procurar uma relação monotônica entre as distâncias e as proximidades.

Isto é aproximava pontos cujas medidas de proximidade estavam próximas na escala ordinal e afastava os pontos correspondentes a medidas de proximidade distantes na escala ordinal.

Esta movimentação dos pontos era feita usando a nação de forças vetoriais. Numa determinada configuração eram construídos com origem em cada um dos n pontos, $(n - 1)$ vetores dirigidos aos demais $(n - 1)$ pontos.

Para se determinar o sentido do vetor com origem no ponto P_i na direção de P_j , observava-se que se ponto P_i tivesse que ser aproximado de P_j o sentido do vetor seria $\overrightarrow{P_i P_j}$, se este devesse ser afastado o vetor teria sentido contrário.

O módulo de cada vetor era determinado medindo-se a discrepância entre as posições da distância $d(P_i P_j)$ e da medida de proximidade δ_{ij} , nas respectivas escalas de orde-

nação. Se as posições fossem concordantes o vetor teria módulo nulo.

Sendo assim cada ponto, como origem de $(n - 1)$ vetores, considerados $(n - 1)$ forças, se moveria na direção da resultante desses vetores.

Sendo o deslocamento de cada ponto devido a $(n - 1)$ tendências distintas, não se esperava que o deslocamento simultâneo dos n pontos levasse imediatamente à condição de monotonicidade desejada entre d_{ij} e δ_{ij} .

Sendo assim o procedimento para obtenção da solução era iterativo, procurando em cada passo aproveitar a parcela de monotonicidade obtida no passo anterior.

Este processo deveria continuar até que o sistema atingisse o equilíbrio perfeito, isto é, até que todos os vetores passassem a ter módulo nulo.

Uma vez que esse estado de perfeito equilíbrio era raramente obtido, Shepard estabeleceu um momento de parada em seu algoritmo.

3.3.3: Método de Kruskal

Kruskal (1964 a, b) construiu a sua técnica de E.M.D. a partir da "Análise de proximidades" de Shepard. A inovação de Kruskal foi a centralização da técnica numa medida de "Adequação de ajuste", ele assim pretendeu obter a configu-

ração que melhor se ajustasse aos dados.

Kruskal basicamente exigiu uma relação monotônica crescente, ou decrescente, entre as medidas de proximidade e as distâncias na configuração procurada. Supos que há uma configuração de pontos num Espaço Euclidiano k - dimensional que é verdadeira, da qual ele poderia descobrir apenas a ordenação linear das distâncias entre os pontos e devia a partir desta informação não métrica recuperar a configuração.

Devido à sua grande importância como introdutor do ajuste e pela grande aplicabilidade de sua técnica, o procedimento de Kruskal será estudado detalhadamente, por nós, no capítulo 4.

3.4 - COMPARAÇÃO DE CONFIGURAÇÕES

3.4.1: Medida de Adequação de Ajuste

Será descrita abaixo uma medida de adequação de ajuste apresentada por Mardia (1979), e usada para comparar duas configurações. Seja X de ordem $(n \times p)$ a matriz das coordenadas dos n pontos obtidas a partir de D por alguma técnica. Supondo-se Y de ordem $(n \times q)$ a matriz de coordenadas de outro conjunto de pontos obtida por outra técnica ou usando outra medida de distância. Seja $q \leq p$, acrescentando-

-se colunas de zeros a Y , esta torna-se de ordem $(n \times p)$ também.

A medida de adequação de ajuste adotada é obtida movendo-se os pontos y_r com relação aos pontos x_r , até que a soma dos quadrados "residual" $\sum_{r=1}^n (x_r - y_r)' (x_r - y_r)$ (7) seja mínima.

Pode-se mover y_r com relação a x_r através de rotação, reflexão e translação isto é pela transformação $A'y_r + b$ $r=1, \dots, n$ onde A' é uma matriz ortogonal de ordem $(p \times p)$.

Deve-se então procurar:

$$R^2 = \min_{A, b} \sum_{r=1}^n (x_r - A'y_r - b)' (x_r - A'y_r - b)$$

para A e b . Note-se que A e b são obtidos por mínimos quadrados baseando-se no teorema abaixo

Teorema 3.5: Sejam $X(n \times p)$ e $Y(n \times p)$ duas configurações de n pontos, por conveniência, com centro na origem tal que $\bar{x} = \bar{y} = 0$. Seja $Z = Y'X$ e usando o teorema da decomposição (Mardia, 1979) escreve-se $Z = V \Gamma U'$ onde V e U são matrizes ortogonais $(p \times p)$ e Γ é uma matriz diagonal de elementos não negativos. Então os valores de A e b que minimizam R^2 são dados por $\hat{b} = 0$ e $\hat{A}A = V U'$ e

$$R^2 = \text{tr } XX' + \text{tr } YY' - 2\text{tr } \Gamma \quad (8)$$

(Demonstração - vide apêndice A.5).

3.4.2: Rotação Procrustea

Supondo que as médias das colunas de X e Y são nulas, a "melhor" rotação de Y com relação a X é $Y \hat{A}$, onde $\hat{A} = V U'$ e \hat{A} é chamada rotação procrustea de Y com relação a X.

Notar que a partir de $Z = V \Gamma U'$ tem-se:

$$Y'X = V \Gamma U'$$

$$X'Y Y'X = U \Gamma V'V \Gamma U'$$

$$X'Y Y'X = U \Gamma^2 U'$$

então o valor mínimo de R^2 pode ser escrito

$$R^2 = \text{tr } XX' + \text{tr } YY' - 2\text{tr}(X'Y Y'X)^{1/2}$$

Observa-se em (7) que R^2 é zero se e somente se y_r puder ser rotacionado exatamente sobre x_r .

3.4.3: Fator de Escala

Se as escalas das duas configurações forem diferentes, a transformação $A'Y_r + b$ deve ser da forma $c A'Y_r + b$ onde $c > 0$. Através do procedimento anterior observa-se que:

$$\hat{c} = \frac{\text{tr } \Gamma}{\text{tr } YY'} \quad (9)$$

e os demais estimadores permanecem os mesmos. Esta transfor

mação é chamada rotação de Procruste com escalonamento de Y com relação a X. Então a nova soma dos quadrados dos resíduos mínima é dada por:

$$R^2 = \text{tr}(XX') + \hat{c}^2 \text{tr}(YY') - 2 \hat{c} \text{tr}(X'Y Y'X)^{1/2} \quad (10)$$

Notar que este procedimento não é simétrico com relação a X e Y. A simetria pode ser obtida selecionando o escalonamento de modo que $\text{tr}(XX') = \text{tr}(YY')$ (ver Sibson-1978)

Se forem dadas duas matrizes de distâncias D_1 e D_2 mas não forem dados os pontos correspondentes, R^2 (8) não pode ser calculado sem usar algum método para determinar os "pontos". As duas primeiras parcelas de R^2 podem ser expressar em função de D_1 e D_2 o $\text{tr} \Gamma$ não pode.

EXEMPLO 3.2: Neste exemplo pretendemos comparar através de R^2 a configuração para o mapa do Brasil obtida pelo método de Torgerson Y_1 (Figura 1.1 e Tabela 3.2) com aquela obtida pelo método de Kruskal Y_2 (Figura 3.2 e Tabela 3.3). Para isto estabelecemos um sistema de coordenadas cartesianas sobre um mapa do Brasil (Figura 3.2) e construímos assim uma matriz X (Tabela 3.1) das coordenadas reais das capitais brasileiras escolhidas.

Sendo que as configurações X e Y_1 ou X e Y_2 não estão na mesma escala adotamos. (ver fórmula 8 e fórmula 10)

$$R^2 = \text{tr}(XX') + \hat{c}^2 \text{tr}(YY') - 2 \hat{c} \text{tr} \Gamma$$

substituindo o valor de \hat{c} (Fórmula 9) obtêm-se:

$$R^2 = \frac{\text{tr}(XX') - (\text{tr } \Gamma)^2}{\text{tr}(YY')} \quad (11)$$

O R^2 obtido por (11) entre X e Y_1 foi 283,11 e aquele obtido entre X e Y_2 foi 333,99 o que nos leva a concluir que a solução clássica de Torgerson se ajusta melhor ao mapa do Brasil do que a solução não - métrica de Kruskal.

Tabela 3.1: Matriz da configuração X obtida para 15 capitais brasileiras

X =	3,25	-9,5
	11,2	-0,7
	-3,8	-4,0
	-0,5	-1,0
	-8,2	0,2
	9,0	-8,7
	4,3	-10,6
	9,9	8,9
	7,4	-11,1
	-12,0	1,5
	5,5	-11,5
	-6,3	-4,5
	-10,2	-4,2
	-6,7	-1,2
	1,9	-8,5

Tabela 3.2: Configuração y_1 , obtida através do programa ALSCAL, para o mapa do Brasil (Método de Torgerson)

$$y_1 = \begin{bmatrix} 0,3841 & -0,6757 \\ 1,2604 & 0,8535 \\ -0,8393 & 0,0477 \\ -0,4537 & 0,5281 \\ -1,4979 & 0,6360 \\ 1,1657 & -0,4573 \\ 0,5603 & -0,8286 \\ 1,3108 & 2,1248 \\ 1,0440 & -0,8627 \\ -2,3595 & -0,0620 \\ 0,7971 & -0,9687 \\ -1,2175 & -0,0455 \\ 1,2143 & 0,2493 \\ -1,4466 & 0,0411 \\ 0,0779 & -0,5801 \end{bmatrix}$$

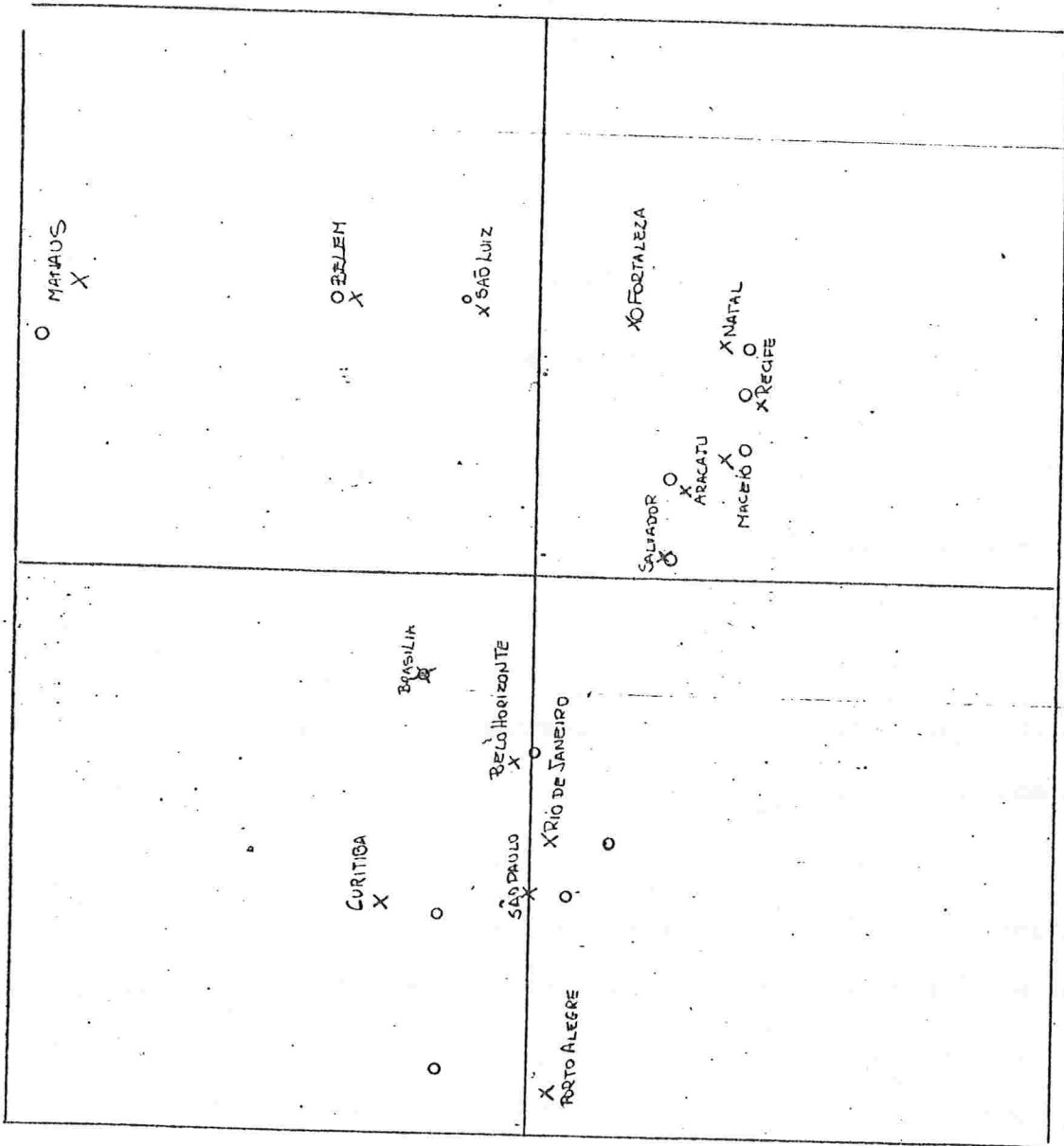
Tabela 3.3: Configuração y_2 , obtida através do programa ALSCAL, para o mapa do Brasil (Método de Kruskal)

$$y_2 = \begin{bmatrix} 0,4338 & -0,6155 \\ 1,2450 & 0,8928 \\ -0,8150 & 0,0176 \\ -0,4564 & 0,4878 \\ -1,5703 & 0,4347 \\ 1,1813 & -0,4079 \\ 0,6149 & -0,8848 \\ 1,0720 & 2,2431 \\ 1,0292 & -0,8881 \\ -2,2635 & 0,3926 \\ 0,8608 & -0,9227 \\ -1,2117 & -0,3257 \\ 1,2392 & 0,2966 \\ -1,4582 & -0,1304 \\ 0,0988 & -0,5901 \end{bmatrix}$$

Figura 3.2 - Mapa do Brasil obtido pelo E.M.D.

O: configuração obtida pelo procedimento não métrico (Kruskal).

X: configuração obtida pelo procedimento mé-



CAPÍTULO 4

MÉTODO DE KRUSKAL

4.1 - ADEQUAÇÃO DE AJUSTE

Para se descrever o procedimento de Kruskal para E.M.D. não métrico, deve-se começar pelo seu ponto chave, a medida de ajuste.

Inicia-se pela procura de uma "função objetiva" \emptyset (Kruskal, 1978). Para se definir esta função a condição é que ela traduza o ajuste da configuração $X = \{x_1, x_2, \dots, x_n\}$ de pontos P_1, P_2, \dots, P_n onde $x_i = (x_{i1}, x_{i2}, \dots, x_{iK})$ vetor de coordenadas do ponto P_i , à matriz de observações (medidas de proximidade, δ_{ij}). Isto é, ela deve produzir um n^o que mostre o quanto a solução está ajustada aos dados.

Para definir a "função objetiva" \emptyset , Kruskal impõe

que a distância entre os pontos obtidos seja uma função mo notônica das proximidades observadas, isto é,

$$d_{ij} = d(P_i, P_j) = f(\delta_{ij})$$

onde f deve ser uma função monotônica qualquer.

Pensando em definir ϑ o seu raciocínio é o mais intuitivo possível. Começa analisando as discrepâncias existentes entre d_{ij} e $f(\delta_{ij})$. Estas discrepâncias podem ser observadas representadas por linhas pontilhadas no gráfico abaixo que mostra uma função f qualquer monotônica.

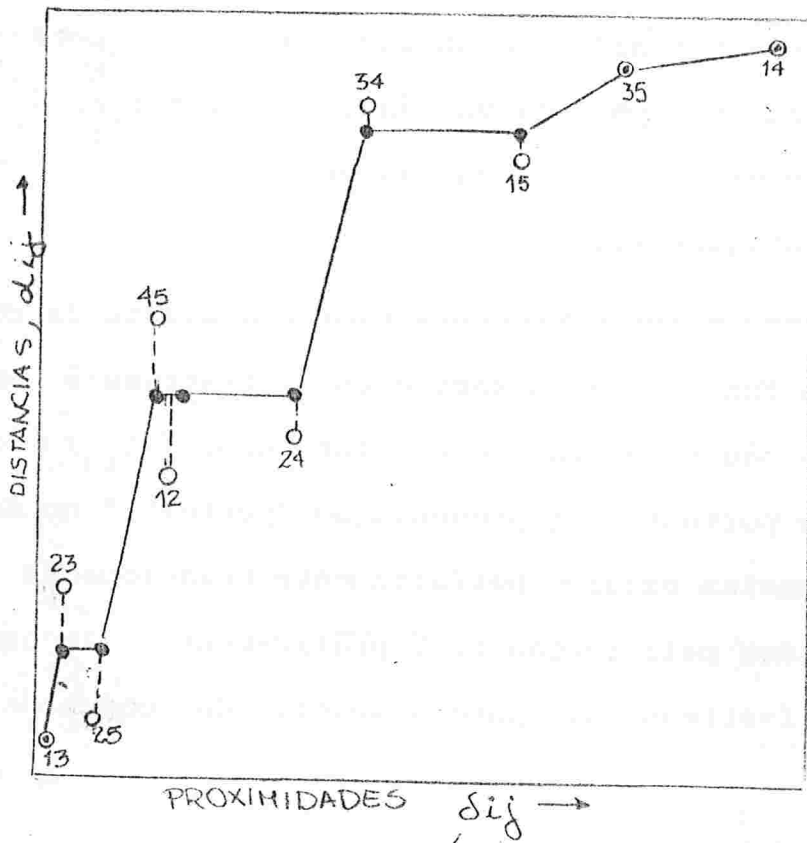


Figura 4.1 - Diagrama de dispersão

As discrepâncias são dadas por $f(\delta_{ij}) - d_{ij}$, não interessando o seu sinal, tanto as positivas como as negativas são interpretadas igualmente e consideradas indesejáveis.

Define então uma função objetiva o f-stress em função dessas discrepâncias quadráticas, dado pela fórmula abaixo

$$\text{f-stress} = \frac{\sum_i \sum_j (f(\delta_{ij}) - d_{ij})^2}{\sum_i \sum_j d_{ij}^2}$$

a raiz quadrada é tomada por analogia ao desvio padrão muitas vezes usado no lugar da variância e o fator $\sum_i \sum_j d_{ij}^2$ no denominador padroniza a medida uma vez que está na mesma escala da configuração.

Quanto maior o f-stress pior é o ajuste da configuração e da função f às observações. O f-stress é obviamente sempre não negativo. Se ele for nulo, $f(\delta_{ij}) = d_{ij}$, $\forall i, j$, tem-se portanto a representação "perfeita" no sentido das distâncias estarem perfeitamente relacionadas com as proximidades pela função f . É praticamente impossível obter-se um f-stress nulo para a maioria dos conjuntos de dados.

Em seguida, procura-se uma medida de ajuste da configuração independente da função f . Define-se então:

stress $(\Delta, X) = \min f\text{-stress } (\Delta, X)$

para todo f

Δ : matriz de proximidade observada

X : configuração obtida

significando que está sendo usada a melhor função f , para a dada configuração.

Uma vez que a monotonicidade exigida pelo modelo, nem sempre é obtida na configuração define-se $d_{ij}^* = f(\delta_{ij})$ onde d_{ij}^* são números monotonicamente relacionados com as proximidades δ_{ij} , e minimizam o stress dado por

$$\varnothing = \sqrt{\frac{\sum_{ij} \sum_{ij} (d_{ij} - d_{ij}^*)^2}{\sum_{ij} d_{ij}^2}} = \text{stress } (\Delta, X)$$

Uma vez definido o stress, define-se a configuração X a partir das observações Δ .

A obtenção da melhor configuração consiste na obtenção daquela que produz o menor valor possível para o stress. Esta configuração \hat{X} (estimada) é aquela que tem o menor \varnothing para uma dada matriz de proximidades observadas Δ . Ou seja, a configuração \hat{X} , solução do E.M.D. é aquela para a qual é válida a igualdade

$$\text{stress } (\Delta, \hat{X}) = \min \text{ stress } (\Delta, X)$$

para todo X

4.2 - OBTENÇÃO DA SOLUÇÃO

Uma vez adotada a idéia básica de Kruskal vamos descrever o procedimento usado por ele para obter a solução do E.M.D.

Kruskal responde em 3 níveis, à pergunta: como obter a configuração de stress mínimo?

4.2.1 - Nível intuitivo

No nível intuitivo ele descreve o procedimento de aproximações sucessivas. Ele começa com uma configuração arbitrária, move um pouco os pontos desta configuração para haver uma ligeira melhora no ajuste. Em seguida repete o processo até obter a configuração onde não haja melhora possível.

O movimento dos pontos é feito obedecendo a seguinte condição: Os pontos P_i e P_j da configuração obtida são aproximados se $d_{ij}^* < d_{ij}$ e afastados caso $d_{ij}^* > d_{ij}$, com o objetivo de se obter os d_{ij} o mais próximos possível dos d_{ij}^* .

4.2.2 - Nível teórico

A nível teórico o problema consiste na minimização

de uma função de várias variáveis, o stress \varnothing

$$\varnothing = \sqrt{\frac{\sum_{ij} \sum_{ij} (d_{ij} - d_{ij}^*)^2}{\sum_{ij} d_{ij}^2}}$$

$$\varnothing = f(x_1, x_2, \dots, x_n) \quad \text{onde}$$

$x_i = (x_{i1}, x_{i2}, \dots, x_{ik})$ vetor de coordenadas \mathbb{R}^k do ponto P_i num espaço

Sendo n pontos P_i num espaço de dimensão k \varnothing é uma função de (nxk) variáveis.

A solução do E.M.D. é dada pela configuração $\hat{X} = (\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n)$ que minimiza a função \varnothing .

Esta solução pode ser obtida resolvendo um problema de análise numérica que usa uma técnica conhecida como "método dos gradientes" ou "método do Descent" (vide Apêndice A.2).

4.2.3 - Nível prático

A nível prático Kruskal descreve as etapas de um programa que permitem a um computador obter a solução do E.M.D.

1ª etapa: Escolhe uma configuração inicial X arbitrária num espaço K -dimensional \mathbb{R}^k .

2ª etapa: Mede as distâncias d_{ij} entre os pontos desta configuração X. A partir das medidas d_{ij} estima os d_{ij}^* , ou seja, obtêm os números que estão monotonicamente relacionados com as medidas de proximidade δ_{ij} .

3ª etapa: Através dos d_{ij}^* é obtido o stress \emptyset para esta configuração X. Se o valor do stress satisfizer as condições do experimentador, esta configuração é a solução do E.M.D.. Caso contrário, passa-se à etapa seguinte.

4ª etapa: Procura-se uma nova configuração X que minimize o stress, para isto é usada a técnica do "steepest descent"; retorna-se em seguida à 2ª etapa do procedimento.

Este é portanto um procedimento iterativo que visa obter a configuração \hat{X} com o menor valor de \emptyset possível.

4.2.3.1 - 1ª etapa: Configuração inicial

No que toca à configuração inicial, uma escolha bem feita economiza várias iterações no processo de obtenção da solução do E.M.D.

Pode-se escolher uma configuração X arbitrária, porém essa configuração deve obedecer a duas condições:

- 1) $P_i \neq P_j$, $\forall i \neq j$, P_i e P_j pontos da configuração.

- 2) X não pode estar contida em \mathbb{R}^t tal que $t < k$, onde k é a dimensão do espaço da configuração procurada.

Uma sugestão de Kruskal seria a escolha de P_1, P_2, \dots, P_n os n primeiros pontos do conjunto abaixo:

(1,0,0,...,0,0)
(0,1,0,...,0,0)
(0,0,1,...,0,0)
⋮
(0,0,0,...,0,1)
(2,0,0,...,0,0)
(0,2,0,...,0,0) etc...

Outra sugestão seria gerar os pontos P_1, P_2, \dots, P_n a partir de um gerador de números aleatórios:

Qualquer que seja a escolha da configuração X , esta deve ser normalizada, isto é:

- 1) O seu centro de gravidade deve se localizar na origem do sistema.
- 2) A soma das distâncias de seus pontos à origem do sistema deve ser igual a 1

$$\sum_{i=1}^n \sum_{t=1}^k x_{it}^2 = 1$$

x_{it} → coordenada de P_i na dimensão t

4.2.3.2 - 2ª etapa: Obtenção dos d_{ij}^*

Ac se estimar os d_{ij}^* , deve-se lembrar que estes são apenas números reais, não são distâncias. Não há uma configuração cujas distâncias entre os pontos sejam d_{ij}^* . Os d_{ij}^* são apenas uma sequência monotônica de números, escolhidos o mais próximo possível dos d_{ij} , que são usados como referência para o grau de monotonicidade das medidas d_{ij} , sendo assim a estimativa dos d_{ij} é um problema de Regressão Monotônica (ver Apêndice C.3) que consiste em procurar os d_{ij}^* sujeitos à restrição de monotonicidade e que tornam mínima a soma de quadrados do numerador do stress

$$\phi = \sqrt{\frac{\sum_{ij} (d_{ij} - d_{ij}^*)^2}{\sum_{ij} d_{ij}^2}}$$

uma vez que para uma dada configuração X , a $\sum_{i=1}^u \sum_{j=1}^n d_{ij}^2$ é fixa.

EXEMPLO 4.1 - Considere-se a matriz $\Delta = (\delta_{ij})_{5 \times 5}$ simétrica, medidas de proximidade observadas.

$$\Delta = \begin{bmatrix} 0 & & & & \\ 1 & 0 & & & \\ 8 & 7 & 0 & & \\ 3 & 2 & 5 & 0 & \\ 4 & 10 & 9 & 6 & 0 \end{bmatrix}$$

Ordenando-se as medidas de proximidade δ_{ij} obtêm-se a sequência

$$\begin{array}{cccccccccc} \delta_{21} < & \delta_{42} < & \delta_{41} < & \delta_{51} < & \delta_{43} < & \delta_{54} < & \delta_{32} < & \delta_{31} < & \delta_{53} < & \delta_{52} \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{array}$$

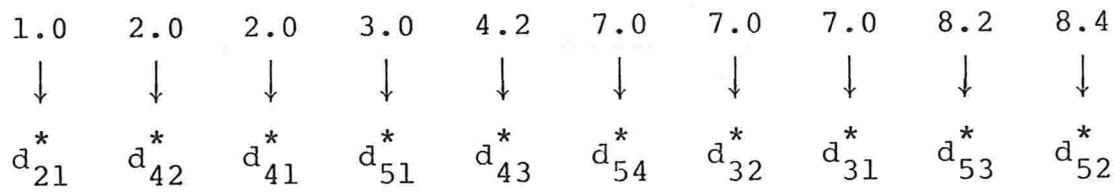
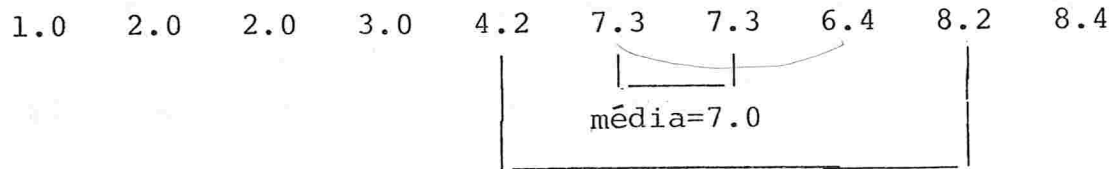
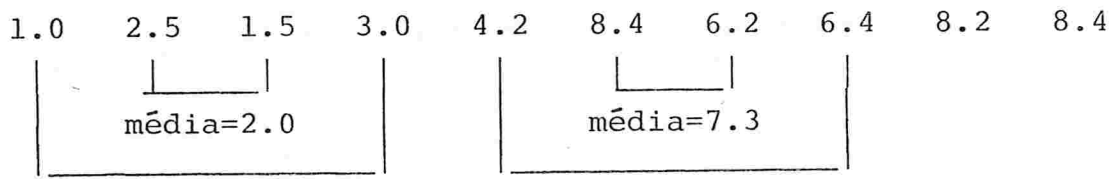
Seja $D = (d_{ij})_{5 \times 5}$ uma matriz de distâncias obtida a partir de uma certa configuração inicial X em \mathbb{R}^2

$$D = \begin{bmatrix} 0.0 & & & & & & & & & & \\ 1.0 & 0.0 & & & & & & & & & \\ 6.4 & 6.2 & 0.0 & & & & & & & & \\ 1.5 & 2.5 & 4.2 & 0.0 & & & & & & & \\ 3.0 & 8.4 & 8.2 & 8.4 & 0.0 & & & & & & \end{bmatrix}$$

Arranjando-se as distâncias na ordem das proximidades obtêm-se

$$\begin{array}{cccccccccc} d_{21} & d_{42} & d_{41} & d_{51} & d_{43} & d_{54} & d_{32} & d_{31} & d_{53} & d_{52} \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ 1.0 & 2.5 & 1.5 & 3.0 & 4.2 & 8.4 & 6.2 & 6.4 & 8.2 & 8.4 \end{array}$$

Observando-se a sequência, nota-se que ela não é monotônica. Em cada sub-sequência que quebra a monotonicidade, seus termos são substituídos pela sua respectiva média



Tem-se portanto os d_{ij}^* monotonicamente relacionados com os δ_{ij} e a partir deles é montada a matriz $D^* = (d_{ij}^*)_{5 \times 5}$

$$D^* = \begin{bmatrix} 0.0 & & & & \\ 1.0 & 0.0 & & & \\ 7.0 & 7.0 & 0.0 & & \\ 2.0 & 2.0 & 4.2 & 0.0 & \\ 3.0 & 8.4 & 8.2 & 7.0 & 0.0 \end{bmatrix}$$

Este algoritmo é ligeiramente alterado quando houver observações repetidas (ver Capítulo 9).

4.2.3.3 - 3ª etapa: Cálculo do stress

Esta etapa consiste apenas de um cálculo do stress e da análise do seu resultado.

Tem-se abaixo o cálculo do stress para o exemplo

4.1 :

$$\sigma = \sqrt{\frac{\sum_{i=1}^5 \sum_{j=1}^5 (d_{ij} - d_{ij}^*)^2}{\sum_{i=1}^5 \sum_{j=1}^5 d_{ij}^2}}$$

$$\sum_{i=1}^5 \sum_{j=1}^5 (d_{ij}^*)^2 = 6,92$$

$$\sum_{i=1}^5 \sum_{j=1}^5 d_{ij} = 651,72$$

$$\sigma = 0,1030439$$

Este valor de stress é considerado alto e portanto seria necessária uma nova iteração no caso deste exemplo para procurar uma outra configuração que dê menor stress.

4.2.3.4 - 4ª etapa: Procura da configuração de stress mínimo

Kruskal (1978) faz uma analogia que permite entender os objetivos e o funcionamento da técnica do "Steepest Descent" (ver Apêndice B.) . Esta analogia esta descrita a

baixo:

Imagine um terreno ondulado, com montanhas e vales. A cada ponto do terreno corresponde uma configuração X (não um ponto da configuração). Cada ponto do terreno pode ser descrito por três coordenadas, a altitude e as duas coordenadas de localização, Norte-Sul e Leste-Oeste. As coordenadas de localização são as mesmas para todos os pontos da configuração. (Certamente, uma configuração de n pontos em um espaço k -dimensional tem $n \cdot k$ coordenadas e $n \cdot k$ é bem maior que 2, portanto a analogia não transfere toda a riqueza e a dificuldade da situação). A altitude é análoga à função objetiva, ou seja, é o stress.

Procurar a configuração com stress mínimo corresponde a procurar o ponto de altitude mínima, o ponto mais baixo do terreno. Como não se conhece a função stress, é impossível ter-se uma visão dela, isto equivale a procurar o ponto mais baixo do terreno de olhos vendados.

Precisa-se de um ponto inicial para a procura. Algumas vezes escolhe-se um ponto aleatório do terreno para este fim; isto é semelhante a se lançar de um avião, voando sobre o terreno, um para-quedista de olhos vendados numa noite escura. Este para-quedista marcha montanha abaixo, a partir do ponto inicial, por passos, imaginando-se que este já procurando o ponto mais baixo do terreno.

Ele se mantém sobre um pé só, tateia ao seu redor com o outro pé, toma a direção que lhe parece mais sensivelmente descendente (matematicamente na direção do menor gradiente) e dá um passo nessa direção. Ele então repete o processo na nova direção e assim por diante. Eventualmente, chega a um local onde nenhuma direção leva a pontos mais baixos, aí então, ele para.

Se o terreno não possui ondas e irregularidades, então este local é provavelmente o ponto mais baixo do vale e ele estará feliz em tê-lo alcançado. Se o terreno todo tiver o formato de uma enorme tijela, tendo portanto um único vale, o para-quedista estará no seu ponto mais baixo.

Se o terreno possuir somente um vale central mas várias ligeiras irregularidades, há uma grande chance que ele tenha alcançado o ponto mais baixo, no entanto ele pode também estar no fundo de uma das irregularidades.

Qualquer lugar onde o para-quedista possa parar é chamado um "mínimo local". Se for o ponto mais baixo do terreno é chamado "mínimo global".

Na prática a possibilidade de encontrar-se um mínimo local ao invés de um mínimo global existe. Felizmente, geralmente ele difere ligeiramente do mínimo global, no entanto se o mínimo local diferir drasticamente do mínimo global, a configuração não pode ser interpretada.

O stress do mínimo local costuma ser maior que o do mínimo global. Comparando-se o stress da solução obtida com aqueles obtidos por soluções em espaços de dimensão mais alta e mais baixa, pode-se perceber que foi obtido um mínimo local.

Quando houver suspeita, pode-se investigar recomeçando o processo com configurações iniciais distintas, o que na analogia equivale a lançar vários para-quedistas em locais distintos e deixar que cada um separadamente procure o ponto mais baixo. É pouco provável que todos os cálculos convirjam para o mesmo mínimo local que não seja o mínimo global.

Finalmente, a experiência sugere que em se suspeitando de um mínimo local seja feita uma nova análise dos dados.

4.3 - DIAGRAMA DE DISPERSÃO

O diagrama de dispersão para uma determinada configuração é um gráfico das medidas d_{ij} em função das medidas δ_{ij} . Portanto este diagrama tem assinalado um ponto para cada par de estímulos (Fig. 4.1).

Este diagrama foi utilizado na definição da função de ajuste e pode também ser um auxiliar do próprio stress

na avaliação do ajuste da configuração aos dados.

Uma vez obtida a solução do E.M.D pode-se construir um diagrama de dispersão com duas finalidades. Primeiramente verificar se a relação exigida entre distâncias (d_{ij} e medidas de proximidade (δ_{ij}) foi obedecida nessa solução. Secundariamente, procurando obter uma impressão visual do ajuste da configuração aos dados, uma vez que já foi obtido o valor numérico do ajuste através do stress.

Kruskal supos $d_{ij} = f(\delta_{ij})$ onde f é uma função monotônica, porém o E.M.D. é muito robusto no sentido de que configurações obtidas com suposições diferentes com relação a f não diferem muito entre si (Kruskal, 1978). (Quando o nº de estímulos é pequeno esta robustez desaparece).

Qualquer que seja a suposição feita com relação a f , o diagrama de dispersão pode mostrar que os dados pedem ou sugerem algum outro tipo de função. Ele mostra isto localizando os pontos nitidamente em torno de uma curva diferente daquela correspondente à função f .

Por exemplo, se for suposto f linear ao invés de simplesmente monotônico para os dados que forneceram o diagrama de dispersão da figura 4.2, poder-se-ia até eventualmente chegar a mesma configuração e mesmo a um diagrama de dispersão bastante semelhante ao da figura junto com a conclusão que uma função linear seria a melhor aproximação pa

ra a curva claramente mostrada na figura.

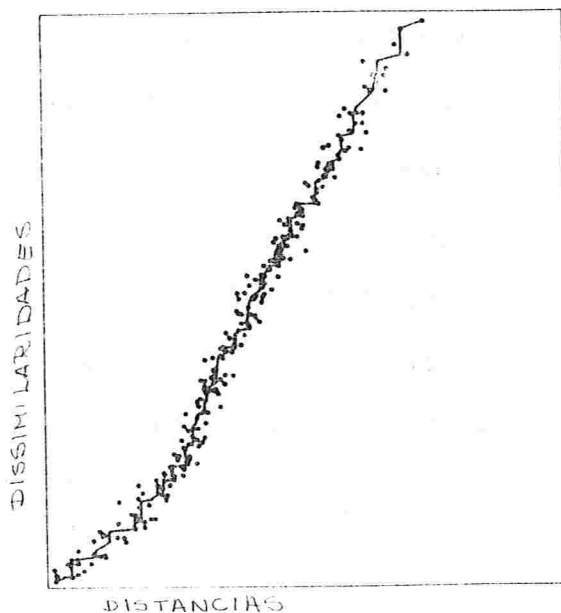


Figura 4.2 - Diagrama de dispersão.

Se os pontos do diagrama de dispersão mostrarem claramente uma curva diferente daquela referente à função f suposta, o valor do stress poderá estar indevidamente aumentado pelo suposição inadequada com relação a f . É melhor analisar novamente os dados usando uma suposição mais adequada.

Quando se assume f monotônica, o diagrama de dispersão tem quase sempre a mesma aparência característica irregular. (Figs. 4.2, 4.3, 4.4, 4.1)

Os zig-zags individuais raramente tem algum significado real, uma vez que podem aparecer zig-zags diferentes

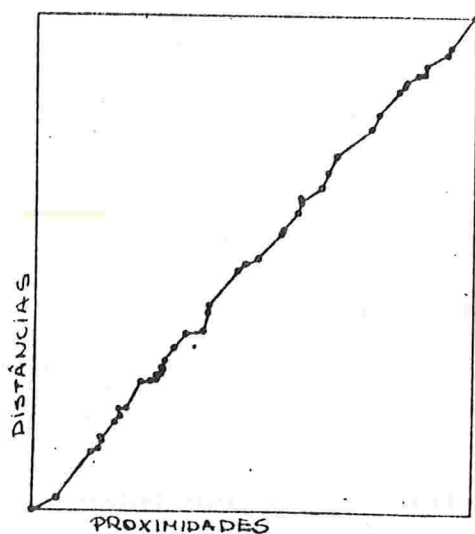


Figura 4.3 - Diagrama de dispersão.

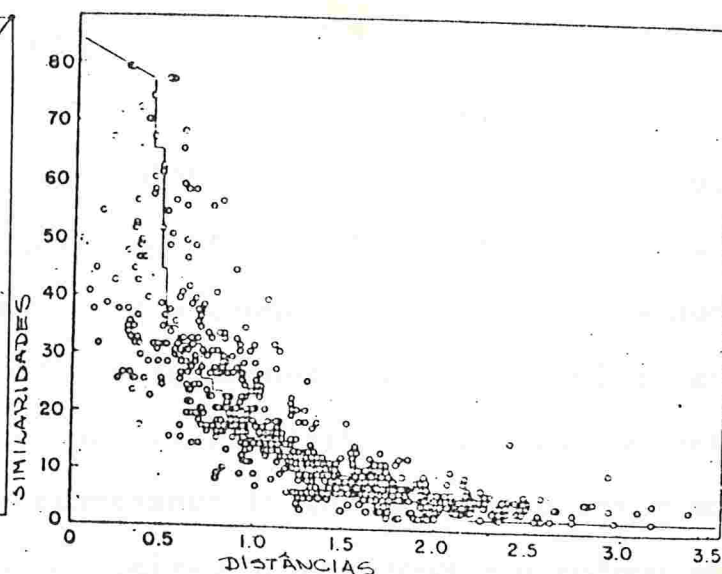


Figura 4.4 - Diagrama de dispersão.

se foram coletados dados novos. No entanto a forma geral da curva que passa pelos pontos do diagrama é frequentemente significativa. Por exemplo, a curvatura observada na parte inferior esquerda da figura 4.2, revela que para pequenos valores de d_{ij} o comportamento de f difere do seu comportamento geral podendo até ser devido a erros.

A figura 4.4 mostra um diagrama de dispersão relativo a um experimento em que se fosse assumido f como função linear estar-se-ia errando totalmente.

O diagrama de dispersão pode mostrar também um fenômeno muito importante chamado "degeneração". Este fenômeno significa que os pontos da configuração estão fortemente

agrupados, isto é, a maioria dos pontos estão sobre ou estão próximos de um número muito pequeno de posições.

A "degeneração" pode ocorrer principalmente sob as seguintes condições: (1) sendo usado E.M.D. não métrico . (2) se os estímulos tiverem um agrupamento natural, geralmente três ou menos grupos, e se as dissimilaridades entre os estímulos dos diferentes grupos forem todas ou quase todas maiores que as dissimilaridades entre os estímulos de um mesmo grupo. Nestas circunstâncias todos ou quase todos os pontos que representam estímulos de um mesmo grupo convergem para a mesma posição. Além disso o stress converge para valores próximos de zero, apesar de não atingir o zero na prática.

Se houver três ou quatro agrupamentos as suas respectivas posições frequentemente formam um triângulo equilátero ou um tetraedro.

Se ocorrer "degeneração" os agrupamentos decorrentes dela devem ser notados e considerados, mas não devem ser tiradas outras conclusões. Em particular o stress muito pequeno não deve ser tomado como indicador de um bom ajuste, uma vez que foi obtido violando duas suposições tácitas: (1) que a verdadeira relação entre distâncias e proximidades é monotônica (2) que os pontos da configuração só estariam na mesma posição se representassem estímulos virtualmente idênticos. Uma análise métrica ou uma nova a-

nálise para cada grupo separadamente é desejável nestas circunstâncias.

O diagrama de dispersão de uma solução degenerada tem uma aparência característica, parece uma escada constituída de vários degraus muito longos. A grande maioria dos pontos localiza-se sobre segmentos horizontais. Esta aparência resulta do fato dos d_{ij} variarem pouco, uma vez que praticamente todas as distâncias entre estímulos de um mesmo conglomerado estão bem próximas de zero.

EXEMPLO 4.2 - No experimento realizado por nós (ver Capítulo 8) foram observadas algumas soluções degeneradas, destacamos abaixo uma delas relativa à seguinte categoria de palavras.

- | | |
|--------------------------------------|--|
| A - sentir-se com apetite | E - sentir-se com prudência |
| B - sentir-se com ambição | F - sentir que está tomando
precaução |
| C - sentir-se com vontade de
algo | G - sentir-se tomando cuidado |
| D - sentir-se com impulso | H - estar atento |
| | I - sentir-se com uma pretensão |

A figura 4.5 apresenta a configuração obtida e a figura 4.6 seu respectivo diagrama de dispersão.

Figura 4.5 - Configuração relativa à solução degenerada do exemplo 4.2.

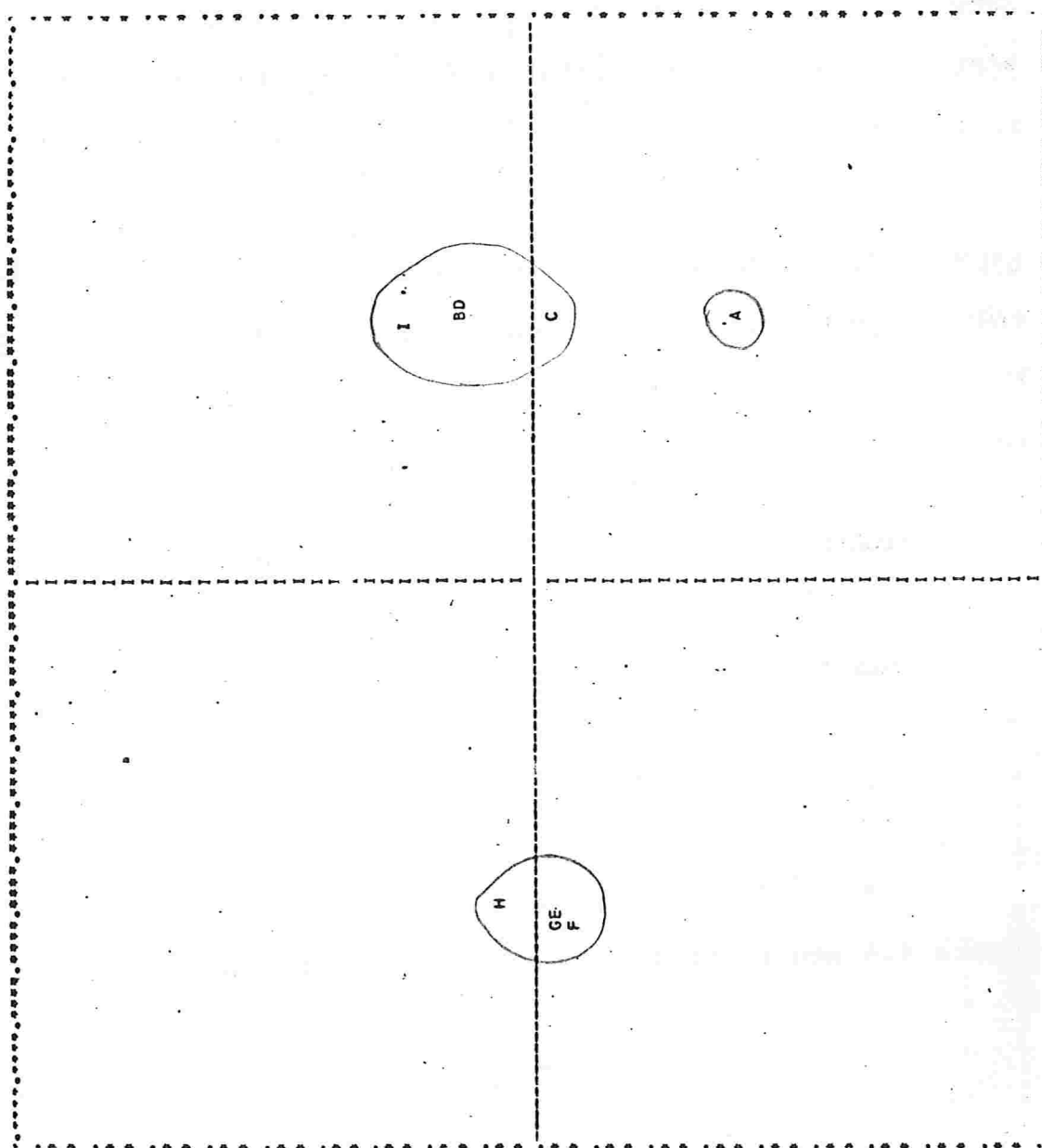
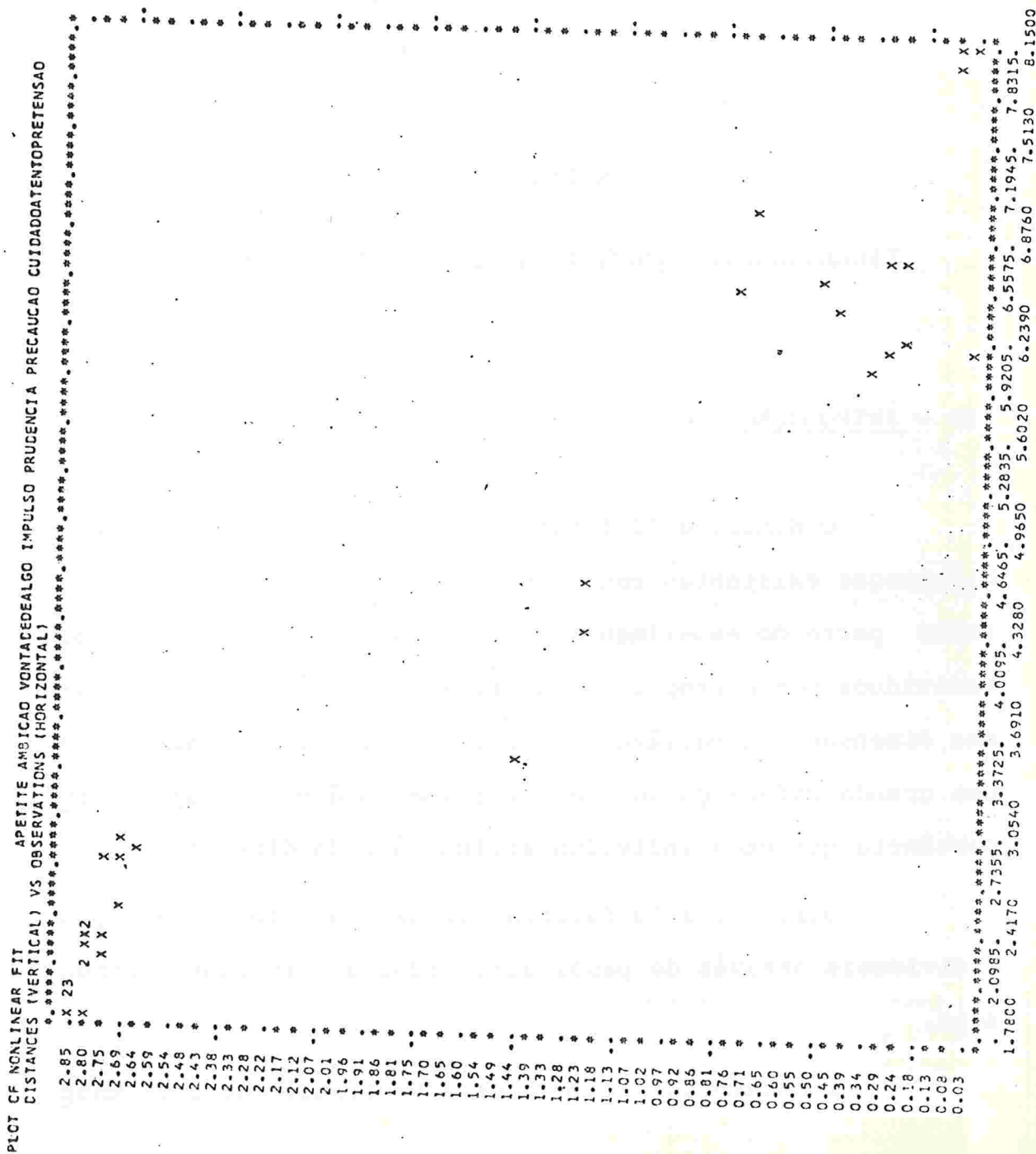


Figura 4.6 - Diagrama de dispersão relativo à solução degerada do exemplo 4.2.



CAPÍTULO 5

ESCALONAMENTO MULTIDIMENSIONAL : " 3 FATORES "

5.1 - INTRODUÇÃO

O E.M.D. a "2 fatores" não considera as possíveis diferenças existentes entre os julgamentos dos sujeitos que tomam parte do experimento. Apesar da grande maioria dos indivíduos perceberem o mundo praticamente segundo as mesmas dimensões ou variáveis de percepção, existe certamente uma grande diferença no que diz respeito à valoração e a importância que cada indivíduo atribui à cada dimensão.

O E.M.D. a "3 fatores" incorpora estas diferenças individuais através de pesos atribuídos às diversas dimensões.

O modelo utilizado para esta finalidade é o mode

lo Euclidiano Ponderado onde:

$$d_{kij}^2 = \sum_{r=1}^R w_{kr} (x_{ir} - x_{jr})^2 \quad w_{kr} \geq 0$$

sendo: d_{kij} : distância entre o estímulo i e o estímulo j para o indivíduo k

w_{kr} : peso atribuído pelo indivíduo k à dimensão r

x_{ir} : coordenada na configuração X do estímulo i na dimensão r

Para o E.M.D. a 3 fatores tem-se uma matriz de proximidades para cada indivíduo do experimento, sendo permitidas diferenças sistemáticas grandes entre essas matrizes. Os dados para este E.M.D. são N matrizes Δ indexadas de 1 a N , de modo que os elementos da matriz Δ_k representem as medidas de proximidade δ_{kij} julgadas pelo k -ésimo sujeito. As medidas de proximidade (simularidades ou dissimularidades) apropriadas para E.M.D. a "3 fatores" são iguais àquelas do E.M.D. a "2 fatores" (vide Cap. 2).

Da mesma forma que o E.M.D. a "2 fatores" o E.M.D. a "3 fatores" produz uma configuração $X = (x_{(1)}, x_{(2)}, \dots, x_{(n)})$ chamada "espaço dos estímulos do grupo". Além disso produz um conjunto de N pontos w_1, w_2, \dots, w_N em outro espaço chamado "espaço ponderado" ou "espaço dos sujeitos" correspondendo um ponto a cada sujeito do experimento, sendo este espaço inexistente no E.M.D. a "2 fatores". As coordenadas dos pontos no "espaço ponderado" devem ser sempre positivas ou

nulas ($w_{kr} \geq 0$).

As distâncias entre os pontos no "espaço do grupo" não são usadas no E.M.D. a "3 fatores". Ao invés, cria-se uma configuração nova para cada sujeito k e são usadas as distâncias nessas configurações.

A configuração para o indivíduo k é construída alterando-se a configuração do "espaço grupal" de acordo com os pesos do vetor de pesos $w_{(k)}$. Isto é cada coordenada dos pontos deste espaço é multiplicada por um fator que leva em conta o peso que o indivíduo k atribui a cada uma das dimensões.

Assim sendo as coordenadas da configuração X^k do indivíduo k são dadas por:

$$x_{ir}^{(k)} = \sqrt{w_{kr}} x_{ir}$$

x_{ir}^k : coordenada na dimensão r do estímulo i no espaço correspondente ao sujeito k .

A distância entre os pontos P^i e P^j na configuração X^k indicada por d_{kij} é dada por:

$$\begin{aligned} d_{kij}^2 &= \sum_{r=1}^R (\sqrt{w_{kr}} x_{ir} - \sqrt{w_{kr}} x_{jr})^2 \\ &= \sum_{r=1}^R w_{kr} (x_{ir} - x_{jr})^2 \end{aligned}$$

de acordo com o modelo Euclidiano Ponderado.

Segundo esta interpretação o k-ésimo sujeito dá à diferença quadrática $(x_{ir} - x_{jr})^2$ a ponderação w_{kr} .

EXEMPLO 5.1 - Esta descrito abaixo um exemplo artificial a apresentado por Kruskal (1978) na sua exposição sobre o INDS CAL (uma das técnicas de E.M.D. a "3 fatores").

Consideram-se $n = 4$ pontos (4 estímulos) e $N = 3$ sujeitos.

Na figura (5.1) tem-se o "espaço grupal" bi-dimen sional onde estão representados os pontos P_1, P_2, P_3, P_4 que constituem a configuração. Ao lado está construído o "espaço ponderado" onde pode-se observar que o sujeito 2 atribuiu peso 4 à dimensão I e peso 1 à dimensão II ($w_2 = 4, 1$).

Abaixo estão os 3 espaços individuais; observando se o espaço do indivíduo 2, verifica-se por exemplo que x_{31} e x_{32} (coordenadas de P_3 neste espaço) são obtidas por:

$$x_{31}^{(2)} = \sqrt{w_{21}} x_{31} = \sqrt{4} \cdot (-0,5) = -1,0$$

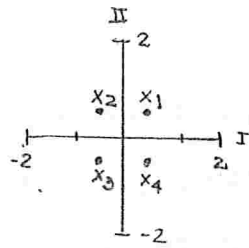
$$x_{32}^{(2)} = \sqrt{w_{22}} x_{32} = \sqrt{1} \cdot (-0,5) = -0,5$$

$$x_3^{(2)} = (-1,0; -0,5)$$

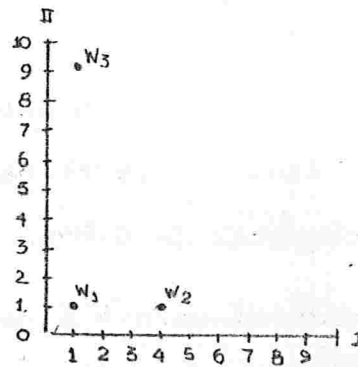
Ainda na figura (5.1) pode-se observar as tabelas que fornecem as distâncias entre os pontos em cada uma das três configurações.

ESPAÇO GRUPAL

ESPAÇO DOS SUJEITOS



X	I	II
1	.5	.5
2	-.5	.5
3	-.5	-.5
4	.5	-.5

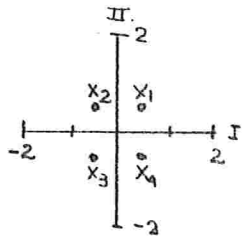


W	I	II
1	1	1
2	4	1
3	1	9

ESPAÇO INDIVIDUO-1

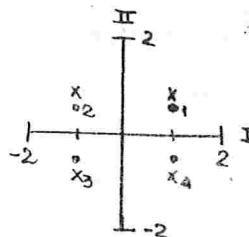
ESPAÇO INDIVIDUO-2

ESPAÇO INDIVIDUO-3



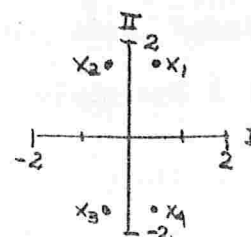
	1	2	3	4
1				
2	1,0			
3	1,4	1,0		
4	1,0	1,4	1,0	

$d_{ij,1}$



	1	2	3	4
1				
2	2,0			
3	2,2	1,0		
4	1,0	2,2	2,0	

$d_{ij,2}$



	1	2	3	4
1				
2	1,0			
3	3,2	3,0		
4	3,0	3,2	1,0	

$d_{ij,3}$

Figura 5.1 - Conceitos básicos do E.M.D a "3 fatores"

Devemos tecer ainda alguns comentários a respeito do E.M.D. a "3 fatores". (1) As N matrizes Δ_k de medidas de proximidade podem eventualmente representar ocasiões diferentes nas quais as medidas de proximidade tenham sido colhidas, ou condições diferentes sob as quais foram obtidas

essas medidas, ou ainda podem ser medidas de proximidade com relação a características distintas. (2) Na prática as estimativas dos pesos w_{kr} podem levar a valores ligeiramente negativos. Se estes valores forem próximos de zero, refletem apenas uma flutuação estatística, se não forem refletem algum desvio entre o modelo e os dados. Frequentemente significa que o nº de dimensões está muito alto para os dados ou que foi usada uma opção inapropriada na entrada (Kruskal, 1978). (3) Segundo vários autores uma vantagem do E.M.D. a "3 fatores" reside no fato das dimensões do espaço obtido por esse escalonamento serem não rotacionáveis, (Takane, Young 1977).

5.2 - TÉCNICAS DE E.M.D. A "3 FATORES"

Nesta secção vamos descrever brevemente como se desenvolveu o E.M.D. a "3 fatores". O primeiro método foi o E.M.D. "tri-modal" (Tucker 1964, 1972). No entanto a complementação de maior sucesso do modelo é devida a Carroll e Chang (1970). O modelo mais recente de E.M.D. a "3 fatores" é o ALSCAL (Takane e Young 1977). O modelo Euclidiano ponderado e os procedimentos associados a ele para ajustar o modelo a dados empíricos foram propostos por várias pessoas aproximadamente ao mesmo tempo (Horan 1969, Bloxom 1968, Carroll e Chang 1970).

O procedimento de Carroll e Chang, chamado

INDSCAL (escalonamento com diferenças individuais) é formalmente uma generalização em n-fatores da decomposição canônica para 2 fatores de Eckart e Young. Esta decomposição é o procedimento por eles chamado de CANDECOMP.

Este procedimento é realizado após uma conversão inicial das dissimilaridades observadas em produtos escalares, obtendo-se alternadamente as estimativas de mínimos quadrados dos pesos (w) (para uma estimativa fixa da configuração X) e depois da configuração X supondo w dado.

Há duas consequências do procedimento CANDECOMP a serem destacadas:

- (1) O critério de minimização (chamado STRAIN, por Carroll) é definido em termos dos produtos escalares computados a partir dos dados. Portanto o INDSCAL não otimiza o ajuste entre o modelo Euclidiano Ponderado e os dados e sim trata do ajuste entre o modelo e uma transformação dos dados.
- (2) Devido à operação que converte dissimilaridades em produtos escalares (que envolve adição, etc...) o procedimento é métrico.

Bloxom (1974) procurou melhorar o método de Carroll e Chang e os resultados de seus trabalhos ainda estão sendo investigados.

Todos os procedimentos citados até aqui, restringem muito os dados aos quais eles se prestam. Especifica

mente exigem que as matrizes de dados sejam simétricas, que não se percam observações, que não haja réplicas, que os dados sejam não condicionados (vide Cap. 6) e ainda que os dados estejam definidos pelo menos no nível de mensuração intervalar.

Alguns procedimentos relaxando algumas dessas exigências foram sendo propostos com graus de sucesso variável.

O primeiro deles foi mencionado brevemente por Carroll e Chang em seu artigo original (1970) chamado NINDSCAL (INDSCAL não métrico) é um procedimento em duas fases que usa o procedimento métrico CANDECOMP na 1^a fase (iterativamente até a convergência) e a regressão monotônica de Kruskal (1964) na segunda fase. As duas fases são aplicadas iterativamente. A 1^a fase minimiza o STRAIN (definido em função dos produtos escalares) e a 2^a fase minimiza o STRESS de Kruskal que é definido em função dos próprios dados.

Envolvendo duas funções a serem minimizadas o NINDSCAL não assegura a convergência num ponto fixo e eventualmente oscila ou diverge após algumas iterações. Além disso o procedimento é ineficiente e dentre as restrições sobre os dados abole apenas a do nível de mensuração.

Pelos motivos acima, Carroll e Chang (1974) propuseram um outro procedimento não métrico para minimizar o STRAIN usando o método de Mínimos Quadrados Alternados (vide Cap. 6) após obter estimativas iniciais de X e w através

de um procedimento CANDECOMP melhorado. Este procedimento que envolve o STRAIN em todas as fases de cada iteração é o primeiro procedimento estável para E.M.D. não métrico envolvendo o modelo Euclidiano Ponderado. Este procedimento tem a propriedade altamente vantajosa que é o relaxamento das restrições sobre os dados. No entanto a desvantagem deste procedimento reside no fato dele estar relacionado com o STRAIN e portanto não otimizar o ajuste entre o modelo de distância e os dados mas sim entre os produtos escalares computados por uma transformação monotônica ótima dos dados e os produtos escalares computados pelas coordenadas da configuração. Dos vários procedimentos revistos. (Young e Takane 1977) este é o mais seguro (pelo menos teoricamente) apesar da sua eficiência ainda ter que ser verificada.

Um terceiro procedimento não métrico para ajustar o modelo Euclidiano Ponderado foi tentado por Young (Takane e Young, 1977). Este procedimento usa uma técnica de gradiente para melhorar simultaneamente as estimativas de X (configuração) e W (pesos) usando as derivadas da função de perda da STRESS.

Apesar deste procedimento usar uma única função de perda e otimizar o ajuste diretamente com relação aos dados, mostrou-se extremamente dependente da configuração inicial. É necessária uma escolha cuidadosa da orientação inicial de X . Apesar desta dificuldade poder ser contornada

usando um procedimento de rotação descrito por de Leeuw (1974), parece que o procedimento ainda padece do fato de usar o gradiente em seu desenvolvimento.

Finalmente Yates (1972) propôs um procedimento com gradiente para ajuste não métrico do modelo Euclidiano Ponderado. Este procedimento não considera nem o STRAIN nem o STRESS, ele minimiza a proporção de variância no modelo que é devida à ordenação não correta de certos pares de distâncias (com relação à ordem das dissimilaridades).

Este objetivo tem sido adotado por vários autores no contexto do modelo Euclidiano não ponderado (Guttman, 1969), de Leeuw 1970, Johnson 1973) e foi muito discutido por de Leeuw (1975) e Young (1975). Apesar desse procedimento ter as vantagens de otimizar uma relação definida diretamente a partir dos dados, e ainda de não restringir os dados, padece do fato de misturar duas funções diferentes a serem otimizadas como foi mostrado por de Leeuw (1975) e discutido por Young (1975).

O procedimento mais recente não métrico que ajusta o modelo Euclidiano é o ALSCAL desenvolvido por Takane, Young e de Leeuw (1977), ele trabalha com uma medida de ajuste semelhante do STRESS chamada SSTRESS, usa uma técnica de Mínimos Quadrados Alternados e remove todas as restrições sobre os dados que forem mencionados anteriormente.

O ALSCAL por ter sido aplicado por nós em alguns exemplos e pela viabilidade do programa será descrito separadamente no Capítulo 6.

No artigo de Carroll e Wish (1974) encontram-se exemplos de aplicação de E.M.D. a "3 fatores" mostrando e comentando possíveis interpretações dos resultados. Nós apresentaremos abaixo um exemplo descrito por Kruskal (1978) apenas para ilustrar o E.M.D. a "3 fatores".

EXEMPLO 5.2 - A técnica INDSCAL foi aplicada a dados de percepção de similaridades entre nações (Wish, Deutsch e Bienen, 1970).

Os 18 sujeitos do estudo piloto do qual os dados foram colhidos atribuíram notas de similaridades numa escala de 9 pontos aos pares obtidos dentre 12 nações. Após ter feito estes julgamentos e alguns outros, os sujeitos indicaram a sua posição com relação à guerra do Vietnã. Foram então classificados em "pacifistas", "moderados" ou "radicais" de acordo com essa sua posição.

A figura 5.2 mostra o "espaço grupal" obtido através de uma análise INDSCAL bi-dimensional dos dados, ou seja das 18 matrizes de medidas de proximidade obtidas. As dimensões foram interpretadas sem rotação como sendo "alinhamento político" e "desenvolvimento econômico". Ao ser feita uma análise em três dimensões, a terceira dimensão pode ser

interpretada como "geográfica ou cultural".

A figura 5.3 mostra os pesos atribuídos pelos sujeitos às duas dimensões no "espaço dos sujeitos"

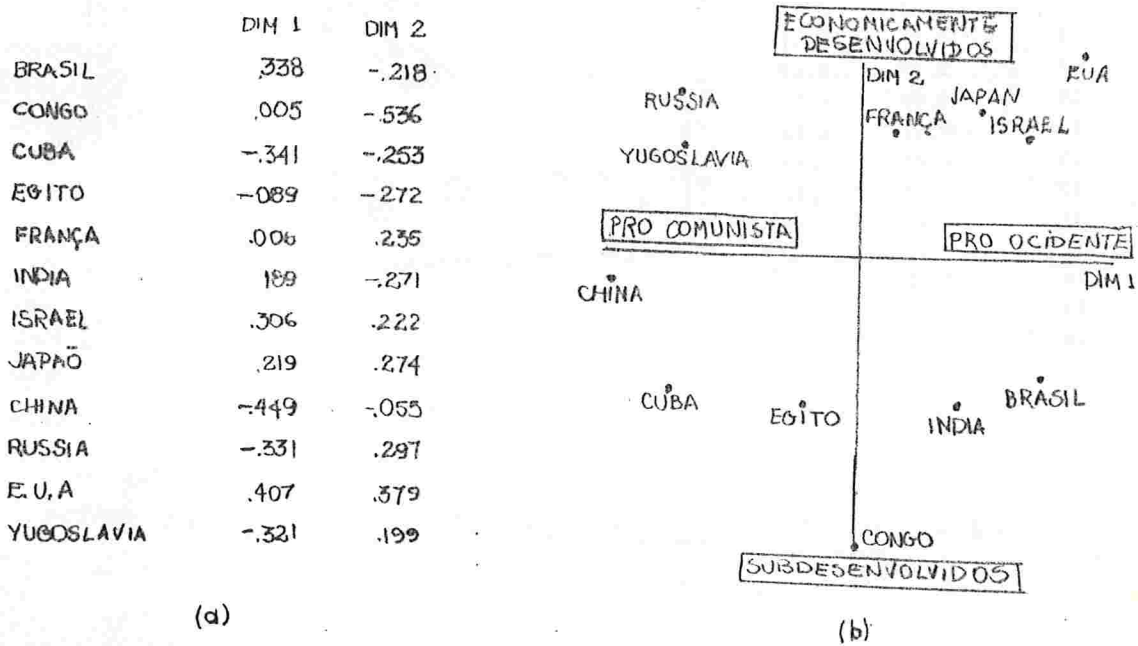


Figura 5.2 - Espaço grupal obtido através da técnica INDSCAL para dados do exemplo 5.2.

SUJEITOS	DIM 1	DIM 2
1	0,23	0,09
2	0,22	0,52
3	0,26	0,50
4	0,26	0,59
5	0,60	0,22
6	0,36	0,56
7	0,62	0,38
8	0,50	0,42
9	0,38	0,57
10	0,33	0,70
11	0,63	0,10
12	0,47	0,43
13	0,45	0,31
14	0,41	0,08
15	0,25	0,46
16	0,52	0,34
17	0,55	0,05
18	0,53	0,39

(a)

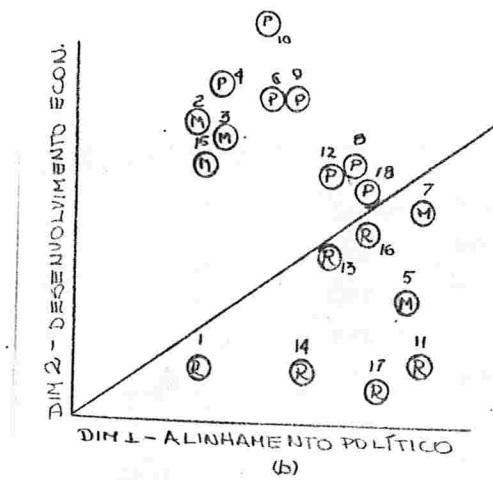


Figura 5.3 - Espaço dos sujeitos relativo aos dados do exemplo 5.2.

A real diferença existente nos julgamentos dos sujeitos pode ser percebida observando-se as matrizes de proximidade dos sujeitos 17 e 4, (figura 5.4), o primeiro considera todos os países comunistas semelhantes entre si e muito diferentes dos não-comunistas enquanto que o outro tem uma "visão de mundo" totalmente distinta, confirmando a observação das matrizes tem-se as suas posições no "espaço dos sujeitos".

	BR	CON	CUB	EGT	FR	IND	ISR	JP	CH	RUS	USA	YUG
Brasil												
Congo	7											
Cuba	8	5										
Egito	2	7	6									
França	3	3	2	6								
India	2	7	2	6	2							
Israel	3	2	3	8	5	7						
Japão	2	1	1	4	3	7	2					
China (Mainland)	2	6	8	7	4	5	3	7				
Russia	2	3	7	7	3	6	3	6	8			
U.S.A.	5	3	3	4	8	5	7	8	1	2		
Yugoslavia	2	2	8	3	4	5	4	6	8	8	2	

(A) SUJEITO 17

	BR	CON	CUB	EGT	FR	IND	ISR	JP	CH	RUS	USA	YUG
Brasil												
Congo	6											
Cuba	8	6										
Egito	5	7	4									
França	3	3	2	2								
India	4	5	3	3	3							
Israel	3	4	2	5	4	3						
Japão	3	3	3	2	3	5	3					
China (Mainland)	2	2	3	2	2	5	3	7				
Russia	2	2	2	2	4	4	4	6	8			
U.S.A.	5	2	3	2	8	5	6	7	3	5		
Yugoslavia	2	2	2	4	5	4	7	4	4	6	3	

(B) SUJEITO 4

Figura 5.4 - Matrizes de Similaridades entre 12 Nações para dois sujeitos

No "espaço dos sujeitos" observa-se também que os indivíduos classificados como "radicais" deram mais peso ao "alinhamento político" enquanto que os "pacifistas" deram peso maior ao "desenvolvimento econômico".

Nas figuras 5.5 (a) e 5.5 (b) tem-se os espaços individuais referentes a um "radical" (sujeito 17) e a um "pacifista" (sujeito 4) respectivamente; a diferença entre eles é dramática. O "radical" localiza as nações ao longo da dimensão I, desconsiderando praticamente o desenvolvimento econômico e o "pacifista" tem uma configuração mais vertical devido ao alto peso por ele atribuído à dimensão II.

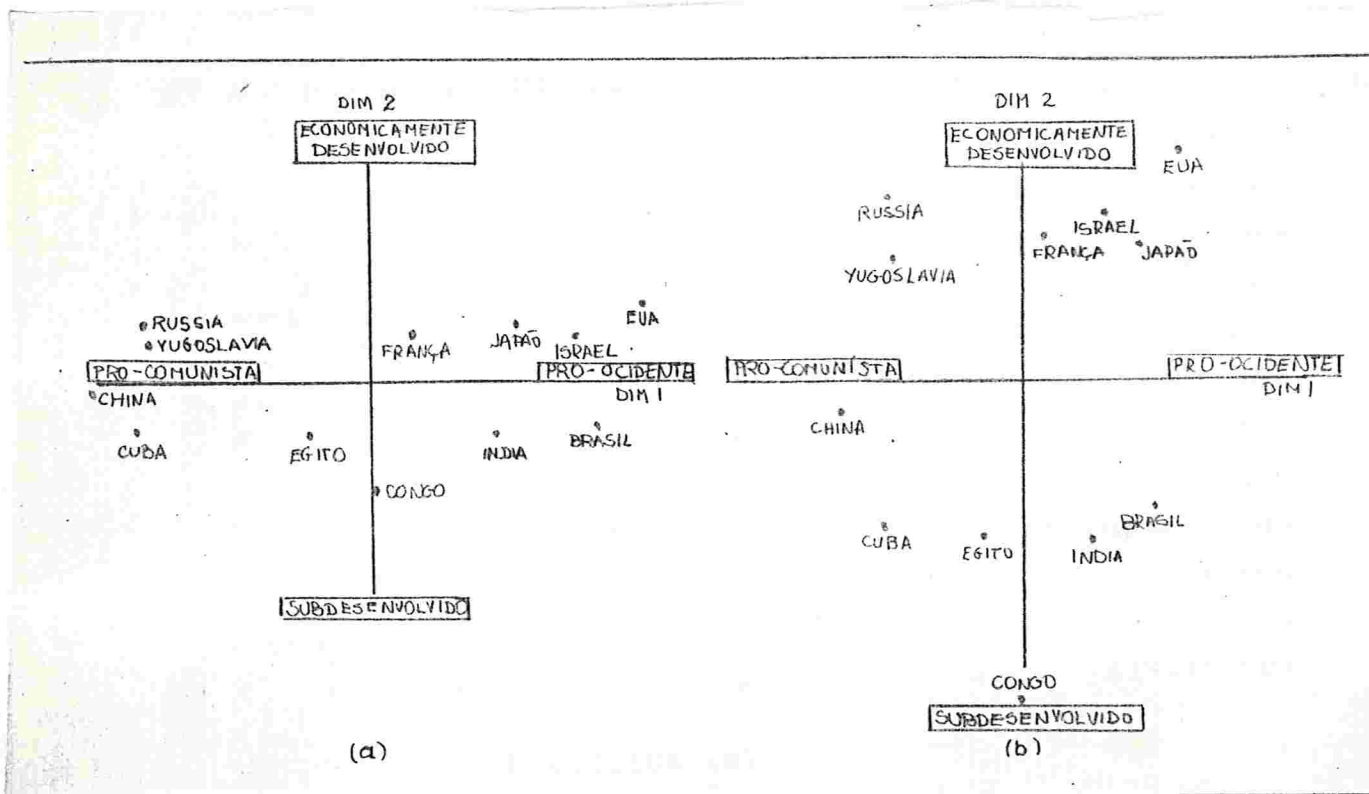


Figura 5.5 - Espaços individuais dos sujeitos 17 e 4 mostrados respectivamente em (a) e (b); dados do exemplo 5.2

O exemplo 5.2 demonstra como a técnica INDSCAL assim como outras técnicas de E.M.D. a "3 fatores" podem a comodar grandes diferenças entre indivíduos. Os dados de do is sujeitos, tão diferentes como aqueles discutidos no exem plo, podem se ajustar no mesmo espaço apesar das matrizes serem totalmente correlacionadas (Kruskal, 1978).

Kruskal (1978) apresenta algumas outras formas de aplicação da própria técnica de E.M.D. a "3 fatores" IND SCAL na análise de dados, uma das aplicações se ajusta ao caso de haverem muitos sujeitos no experimento o que difi - cultaria a inclusão de todos eles numa única análise, a ou tra se aplica se já forem conhecidas as dimensões do espaço, através de pesquisas anteriores, e deseja-se conhecer os pe sos que um novo grupo de sujeitos atribuiria a essas dimen - sões.

CAPÍTULO 6

ALSCAL

6.1 - INTRODUÇÃO

O ALSCAL é um procedimento para E.M.D. a "3 fatores", que apresenta certas vantagens com relação aos métodos anteriormente citados. Foi elaborado por Young, De Leeuw e Takane (1977).

Ele se aplica a dados que tenham observações perdidas, que estejam definidos em qualquer um dos quatro níveis de mensuração (normal, ordinal, intervalar, razão), que sejam discretos ou contínuos, e que sejam simétricos ou as simétricos, condicionais ou com réplicas.

Além disso, a técnica ALSCAL está habilitada, sem maiores complicações, a ajustar o modelo Euclidiano não pon

derado (E.M.D. a "2 fatores") e outros modelos que incluem diferenças individuais como os de Carroll e Chang (1970) , McGee (1968) e Young (1975), dentro de um mesmo procedimento.

Esta técnica baseia-se no conceito de Fisher (1946) de escalonamento ótimo, isto é, as observações devem ser escalonadas obedecendo a dois objetivos: (a) elas devem se ajustar o máximo possível ao modelo segundo um critério de mínimos quadrados; (b) as suas características de mensuração devem ser estritamente mantidas.

Para a aplicação da técnica ALSCAL define-se:

Δ : coleção de matrizes Δ_k das observações quadráticas do indivíduo $\Delta_k = (\delta_{kij}^2)_{n \times n}$

D^* : coleção de matrizes D_k^* das observações quadráticas otimamente escalonadas do indivíduo k $D_k^* = (d_{kij}^{*2})_{n \times n}$

D : coleção de matrizes D_k das distâncias quadráticas onde $D_k = (d_{kij}^2)_{n \times n}$

$$d_{kij}^2 = \sum_{r=1}^R \omega_{kr} (x_{ir} - x_{jr})^2 \quad \omega_{kr} \geq 0$$

As observações quadráticas otimamente escalonadas são comumente chamadas de disparidades para efeito de E.M.D.

ou ainda pode-se referir a elas como sendo estimativas, uma vez que elas são os estimadores de mínimos quadrados das distâncias quadráticas.

Com essas definições pode-se representar formalmente o problema de escalonamento ótimo como um problema de transformação, da seguinte forma: Procura-se uma transformação t dos dados brutos que gere os dados otimamente escalonados d_{kij}^* , tal que:

$$t [\delta_{kij}] = [d_{kij}^*]$$

Existem várias transformações t possíveis; o ALSCAL limita a escolha da transformação t através de três critérios relativos à mensuração dos dados. Esses critérios são o processo de mensuração, o nível de mensuração e a condicionalidade dos dados. Adiante nós estudaremos cada um desses critérios em separado. O processo de mensuração pode ser discreto ou contínuo; os níveis de mensuração são normal, ordinal, intervalar e razão; e quanto a condicionalidade os dados podem ser matriz condicional, linha condicional ou não condicionados.

Além disso, a transformação t tem que ser tal que exista uma relação de mínimos quadrados entre d_{kij}^* e d_{kij} , dado que as características de mensuração sejam estritamente mantidas.

condicionalidade) dizem respeito a três aspectos diferentes das categorias de observações. O processo de mensuração diz respeito a todas as observações dentro de uma particular categoria. O nível de mensuração diz respeito às relações entre todas as observações entre categorias diferentes. E a condicionalidade dos dados, está relacionada com a possibilidade de ajuntamento de categorias.

6.2.1 - Processos de Mensuração

Nesta secção vamos descrever como a escolha da transformação t é limitada em função do processo de mensuração.

Há dois tipos de restrições de processo, uma é considerada quando se assume que o processo gerador é discreto e a outra quando é contínuo. Deve-se assumir sempre uma das duas hipóteses.

Acreditando-se que o processo é discreto, então todas as observações que estejam numa mesma categoria deverão ser representadas pelo mesmo nº real após ter sido feita a transformação t ($t [\delta_{kij}] = [d_{kij}^*]$) isto é:

$$t^d : (\delta_{kij} \sim \delta_{mno}) \rightarrow (d_{kij}^* = d_{mno}^*)$$

\sim : indica equivalência empírica; pertencem à mesma categoria.

6.2 - DADOS - MENSURAÇÃO

Antes de analisar a escolha da transformação t segundo os três critérios relativos à mensuração, devemos estabelecer o conceito básico no desenvolvimento do ALSCAL, que é do ponto de vista de seus autores a consideração de todas as observações como sendo categóricas. Isto é, eles enxergam uma variável-observação consistindo de observações que caem numa variedade de categorias, tal que todas as observações numa particular categoria são equivalente.

Estendem esse ponto de vista "categórico" com relação ao nível de mensuração da variável e com relação à natureza do processo que gera as observações. Resumidamente, seu ponto de vista é que o processo de observação fornece observações categóricas devido às condições de mensuração e de observação.

Por exemplo, ao ser feita uma leitura de 40° num termômetro comum, o que equivale a uma mensuração de nível intervalar e processo contínuo, sabe-se que a temperatura real deve estar entre $39,5^{\circ}$ e $40,5^{\circ}$, sendo assim há um grande número de temperaturas que são relatadas pelo mesmo valor 40° , constituindo assim uma categoria de observações equivalentes à observação 40° .

Os três critérios de mensuração (processo, nível e

vação limita a escolha da transformação t .

Para variáveis nominais, o nível de mensuração não introduz para t outros limites além daqueles já impostos pelo processo de mensuração.

Para variáveis ordinais, exige-se, além das restrições devidas ao processo de mensuração que os números reais associados às observações nas diferentes categorias representem a ordem das observações empíricas, isto é:

$$t^o : (\delta_{kij} \{ \delta_{mno}) \rightarrow (d_{kij}^* \leq d_{mno}^*)$$

t^o : restrição de ordem

$\{$: ordem empírica

Observe-se que se permite que os números sejam iguais mesmo que as observações não o sejam.

Quanto ao problema das observações, ele pode ser tratado pelas restrições de processo. Se a variável for discreta-ordinal (t^{do}) observações iguais permanecem iguais após a transformação, no entanto para uma variável contínua-ordinal (t^{co}) observações iguais podem ser diferentes após a transformação.

Para variáveis quantitativas (intervalar ou razão) de um modo geral, os números reais relacionam-se com as observações através de um polinômio de grau conhecido:

Por outro lado, adotando-se a hipótese de processo contínuo, cada uma das observações de uma determinada categoria deve ser representada por um número real selecionado num intervalo fechado de números reais, isto é:

$$t^c : (\delta_{kij} \sim \delta_{mno}) \rightarrow \begin{cases} d_{kij}^- = d_{mno}^- \leq d_{kij} \leq d_{kij}^+ = d_{mno}^+ \\ d_{kij}^- = d_{mno}^- \leq d_{mno}^* \leq d_{kij} = d_{mno}^+ \end{cases}$$

d_{kij}^- : limite inferior do intervalo de números reais

d_{kij}^+ : limite superior do intervalo de números reais

Observe-se que uma das implicações da equivalência empírica é que os limite superior e inferior de todas as observações em uma determinada categoria são os mesmos para as observações. Portanto os limites se aplicam mais às categorias do que às próprias observações. Note também que para todas as observações numa determinada categoria, as respectivas observações reescaloadas devem cair no intervalo mas não precisam ser iguais.

6.2.2 - Níveis de Mensuração

Veremos agora como o nível de mensuração da obser

$$t^p : d_{kij}^* = \sum_{q=0}^p a_q \delta_{kij}^q$$

(a somatória começa com $q = 1$ para variáveis razão)

No entanto pode-se exigir que os números reais sejam linearmente relacionados com as observações, isto é:

$$t : d_{kij}^* = a_0 + a_1 \delta_{kij}$$

onde $a_0 = 0$ para variáveis em escala razão.

Quando necessário denota-se a transformação intervalar por t^i e a transformação de razão por t^r .

6.2.3 - Condicionalidade

Agora será analisado o terceiro critério relativo à mensuração, o que diz respeito à condicionalidade dos dados.

Como foi enfatizado por Coombs (1964) é possível que as características de mensuração das observações sejam condicionais em algum aspecto da situação experimental, ou seja, que algumas observações não possam ser comparadas com outras. Por exemplo, se alguns sujeitos num experimento tiverem que julgar as simularidades entre pares de estímulos numa escala de 1 a 9, cria uma certa indisposição afirmar

que o sujeito que atribui uma nota 7 a um par considera o par mais semelhante do que o sujeito que lhe atribui nota 6. Não se tem certeza que os sujeitos estejam usando a escala da mesma forma. De fato, pode-se quase ter certeza que eles não a usam da mesma maneira, portanto diz-se que as mensurações são condicionadas aos sujeitos. Geralmente este tipo de condicionalidade é referido como matriz-condicional, uma vez que todas as observações dentro de uma matriz Δ_k são comparáveis, mas não são comparáveis as observações entre matrizes.

As observações podem ainda ser linha-condicionadas ou não condicionadas como são geralmente consideradas no E.M.D. a "2 fatores" (Coombs, 1964).

Formalmente estabelece-se que o domínio da transformação t depende do tipo de condicionalidade. Para dados não condicionados o domínio da transformação é todo o conjunto de observações, as N matrizes Δ_k , e a transformação é denotada por t , uma vez que todas as observações são comparáveis entre si. Para dados matriz-condicional o domínio é uma única matriz de dados Δ_k e a transformação é denotada por t_k , uma vez que as observações só são comparáveis dentro da própria matriz. Finalmente para dados linhas-condicionados o domínio é uma única linha de uma matriz Δ_k que é o que constitui um conjunto de dados comparáveis, e a transformação é denotada por t_{kj} .

A discussão anterior de nível e processo de mensuração foi feita em termos de dados não condicionados e as definições devem ser modificadas adequadamente para outras condicionalidades. Estas modificações e outros comportamentos de condicionalidade foram discutidos por Young (1973) e Kruskal, Young e Seery (1973).

Concluída ao ser feita a escolha da transformação t deverão ser combinados os limites impostos pelos três critérios relativos à mensuração.

6.3 - MEDIDA DE AJUSTE

A técnica ALSCAL procura otimizar uma única medida de ajuste bastante semelhante ao STRESS, o SSTRESS (ϕ^2)

$$\phi^2(X, W, D^*) = \sum_k \sum_j \sum_i^{j-1} (d_{kij}^* - d_{kij})^2$$

O SSTRESS difere do STRESS por estar em função das distâncias quadráticas d_{kij}^2 e de seus estimadores d_{kij}^{*2} . Deve-se observar que os d_{kij}^{*2} são os estimadores de mínimos quadrados de d_{kij}^2 e não os quadrados dos estimadores de mínimos quadrados de d_{kij} .

O STRESS e o SSTRESS não são equivalentes, no entanto, apesar disso, nota-se que as transformações

$$t [\delta_{kij}] = [d_{kij}^*] \quad e \quad [\delta_{kij}^2] = [d_{kij}^{*2}]$$

são equivalentes para os níveis de mensuração normal, ordinal e razão, mas não o são para o nível intervalar, porém Takane, Young e De Leeuw (1977) afirmaram que elas podem ser consideradas equivalentes para os quatro níveis de mensuração, e se propõe a estudar o problema mais a fundo.

Pode-se notar uma outra diferença entre o STRESS e o SSTRESS no comportamento relativo a valores altos de d_{kij} e d_{kij}^* , estes tem mais ênfase com o SSTRESS do que com o STRESS.

Uma simples manipulação algébrica pode explicar essa diferença. Define-se

$$d_{kij}^* = d_{kij} + e_{kij}$$

onde e_{kij} pode ser positivo ou negativo.

Com o STRESS, a contribuição da discrepância entre d_{kij} e d_{kij}^* é de apenas e_{kij}^2

$$(d_{kij} - d_{kij}^*)^2 = |d_{kij} - (d_{kij} + e_{kij})|^2 = e_{kij}^2$$

No entanto com o STRESS tem-se:

$$\begin{aligned} |d_{kij}^2 - d_{kij}^{*2}|^2 &= |d_{kij}^2 - (d_{kij} - e_{kij})^2|^2 = \\ &= |d_{kij}^2 - d_{kij}^2 - 2d_{kij}e_{kij} - e_{kij}^2|^2 \\ &= e_{kij}^2 |e_{kij} - 2d_{kij}|^2 \end{aligned}$$

Portanto não s3o o m3dulo de e_{kij} mas seu sinal tam-
b3m (d_{kij} 3 sempre n3o negativo) e o m3dulo de d_{kij} est3o
relacionados na avalia33o do ajuste.

Um exemplo torna isto claro:

EXEMPLO 6.1 - Supondo que se tenha os dois casos abaixo:

$$(A) \quad d_{kij} = 2 \quad d_{kij}^* = 1 \quad \left| d_{kij} - d_{kij}^* \right| = 1$$

$$(B) \quad d_{kij} = 5 \quad d_{kij}^* = 6 \quad \left| d_{kij} - d_{kij}^* \right| = 1$$

Se for usado o STRESS a contribui33o relativa des-
sas discrep3ncias 3 igual, mas se for usado o SSTRESS tem-
-se uma raz3o de 3 para 11 o que 3 bem diferente da igual-
dade.

$$(A) \quad d_{kij}^2 = 4 \quad d_{kij}^{*2} = 1 \quad \rightarrow \left| d_{kij}^2 - d_{kij}^{*2} \right| = 3$$

$$(B) \quad d_{kij}^2 = 25 \quad d_{kij}^{*2} = 36 \quad \rightarrow \left| d_{kij}^2 - d_{kij}^{*2} \right| = +11$$

Este efeito 3 mais marcante quando se compara o ca-
so (C) com o caso (B)

$$(C) \quad \left\{ \begin{array}{l} d_{kij} = 5 \quad d_{kij}^* = 4 \\ d_{kij}^2 = 25 \quad d_{kij}^{*2} = 16 \end{array} \right. \left\{ \begin{array}{l} \left| d_{kij} - d_{kij}^* \right| = 1 \\ \left| d_{kij}^2 - d_{kij}^{*2} \right| = 9 \end{array} \right.$$

No caso (C) a discrepância é de 9 se avaliada pelo SSTRESS. Assim sendo, mesmo que se tenha a mesma d_{kij} e que a diferença seja igual usando o STRESS, o sentido da diferença contribui diferentemente no SSTRESS.

A relação entre o STRESS e o SSTRESS apesar de ser algebricamente tratável não é direta e não torna o problema muito mais claro, uma vez que não se pode comparar a magnitude do ajuste pois os fatores de normalização das fórmulas do STRESS e do SSTRESS podem ser diferentes.

Segundo Young, Takane e De Leeuw (1977) não há razão a priori para se escolher uma das duas fórmulas. Uma vantagem que tanto o STRESS quanto o SSTRESS tem sobre o STRAIN (medida de ajuste na técnica INDSCAL) é o fato de suas fórmulas serem perfeitamente compatíveis com as restrições de nível de mensuração. A razão básica da opção pelo SSTRESS é uma conveniência no algoritmo. Como se sabe as diferenças individuais ponderadas W são lineares com relação às distâncias quadráticas, mas não com relação às próprias distâncias

$$d_{kij}^2 = \sum_{r=1}^R w_{kr} (x_{ir} - x_{jr})^2 \quad w_{kr} \geq 0$$

Assim sendo a estimativa de mínimos quadrados quadrados de W pode ser obtida através de uma série de operações matriciais elementares quando o SSTRESS é adotado co-

mo critério de otimização.

6.4 - DESCRIÇÃO DA TÉCNICA ALSCAL

Vamos descrever neste tópico o algoritmo de Mínimos Quadrados Alternados para E.M.D. com diferenças individuais (ALSCAL). Para isto é necessário inicialmente explicar o que é o método de Mínimos Quadrados Alternados (ALS) (Young, De Leeuw e Takane, 1977).

O ALS é um procedimento geral para estimar parâmetros que envolve uma subdivisão dos parâmetros em vários subconjuntos e em seguida obtém o estimador de Mínimos Quadrados de um dos subconjuntos de parâmetros sob a hipótese dos restantes serem de fato conhecidos e constantes. A estimação é então alternadamente repetida para um subconjunto por vez até que todos os subconjuntos sejam estimados. Este processo todo é então iterado até que se obtenha a convergência. Esta convergência é certamente obtida, isto é, jamais num procedimento de ALS obtém-se uma iteração que piore a função que deve ser otimizada (de Leeuw, Young e Takane 1976).

O algoritmo ALSCAL envolve duas fases maiores e duas fases menores.

Na 1^a fase maior obtém-se os estimadores de míni-

mos quadrados das observações otimamente escalonadas (D^*) sob a hipótese de que a configuração (X) e os pesos (W) são constantes. Isto é, resolve-se o problema de mínimos quadrados condicionados que minimiza o SSTRESS sob a hipótese de X e W não variarem.

Notação: $\text{MIN}_{D^*} [\sigma^2 (D^*/X, W)]$

A 2ª fase maior é constituída por duas sub-fases de minimização separadas. A 1ª delas procura os pesos, que minimizam o SSTRESS considerando constantes a configuração e as disparidades.

Notação: $\text{MIN}_W [\sigma^2 (W/X, D^*)]$

E a 2ª fase procura a configuração que minimiza o SSTRESS considerando os pesos e as disparidades constantes.

Notação: $\text{MIN}_X [\sigma^2 (X/W, D^*)]$

As duas fases menores do algoritmo são a fase inicial e a fase final do processo.

1. Fase inicial

Computam-se os valores iniciais de X e W a partir de Δ (coleção das matrizes Δ_k das observações quadráticas)

usando a solução algébrica modificada de Schonemann. (Ver Apêndice D.).

2. Fase de escalonamento ótimo

2.1 - Calculam-se as distâncias Euclidianas ponderadas quadráticas (D) a partir de X e W.

2.2 - Obtêm-se as disparidades otimamente escalonadas D^* (estimadores de Mínimos Quadrados de D) a partir das distâncias D, das observações Δ e das restrições de mensuração usando o método de de Leeuw, Young e Takane (1976). (Vide Apêndice E.1.).

2.3 - Normalizam-se adequadamente as disparidades. (Vide Apêndice E.2.).

3. Fase final

Calcula-se o SSTRESS e verifica-se a taxa de decréscimo do SSTRESS é suficientemente baixa para poder encerrar o processo.

4. Fase de estimação do modelo

4.1 - Calculam-se os novos estimadores de mínimos quadrados dos pesos W a partir da configuração anterior X e das

novas disparidades D^* através de técnicas de regressão. (Ver Apêndice F-1).

4.2 - Impõe-se as restrições de não-negatividade sobre W , se necessário, através de técnicas desenvolvidas por Young (1978). (Ver Apêndice F-2).

4.3 - Calculam-se os novos estimadores de mínimos quadrados da configuração X através dos novos pesos W calculados em 4.1 e 4.2 e das disparidades D^* computadas em 2.3 usando uma técnica numérica (Ver Apêndice F-3).

4.4 - Retorna-se à fase 2.1 para uma nova iteração.

CAPÍTULO 7

ANÁLISE DA SOLUÇÃO DO E.M.D.

Um dos problemas que o pesquisador enfrenta quando aplica uma técnica de E.M.D. é a análise dos seus resultados. Dois aspectos são extremamente importantes nesta análise, um deles a escolha da dimensão do espaço no qual o experimentador deve interpretar os resultados e o outro aspecto é a interpretação da configuração obtida no espaço escolhido.

7.1 - ESCOLHA DA DIMENSÃO DO ESPAÇO

7.1.1 - Introdução

As diversas técnicas de E.M.D. impõem a dimensão

do espaço Euclidiano no qual obter-se-á a configuração desejada.

Na prática, a dimensão do espaço é habitualmente desconhecida e deve ser escolhida segundo alguns critérios. Sendo que o E.M.D. é geralmente usado como um modelo descritivo para representar e entender os dados, esses critérios devem considerar a interpretabilidade da solução, a facilidade de uso dessa solução, a estabilidade do modelo e o ajuste do modelo aos dados.

Neste momento a preocupação maior é a escolha de um espaço de dimensão apropriada aos dados e não a procura de dimensão "correta" do espaço, daquele espaço "hipotético", "verdadeiro" que está sob os dados.

Quanto à dimensão do espaço, esta não precisa necessariamente corresponder ao número de características observáveis dos estímulos envolvidos no experimento. Há características dos estímulos que não se destacam para constituir uma dimensão em separado, podem ser observadas na configuração porém não são identificadas com nenhum dos eixos. Há também soluções que apresentam um número de direções ou características interpretáveis maior que a dimensão do espaço, ou ainda o pesquisador pode não estar apto a interpretar alguns dos eixos apesar dos dados exigirem uma configuração num espaço de dimensão alta. Além disso o

que conceitualmente é uma característica só pode corresponder a duas ou mais dimensões quando a característica não puder ser descrita por uma única variável ordenada.

7.1.2 - Ajuste

Num E.M.D. a "2 fatores" em que seja usado o STRESS como medida de ajuste, este pode servir como guia para a escolha do nº de dimensões do Espaço Euclidiano.

Há dois procedimentos básicos para isto, um deles é estatístico, o outro é baseado em experiência e intuição. O estatístico requer menos pericia e é menos subjetivo, o intuitivo é usável num número maior de situações. As tabelas necessárias para o procedimento estatístico só foram publicadas para certas condições e para um número limitado de valores de n (número de estímulos).

Vamos comparar o procedimento intuitivo com o procedimento estatístico através do exemplo abaixo, apresentado por Kruskal (1978).

EXEMPLO 7.1 - Neste exemplo serão usados dados de três experimentos. No 1^o deles o E.M.D. é aplicado a 12 estímulos que são nações, no 2^o são 12 dialetos BIKOL e no 3^o são 12 dialetos KONSOID.

(1) Procedimento intuitivo: Vamos começar observando para cada um dos experimentos os gráficos do stress em função da dimensão do espaço (fig. 7.1). Os gráficos foram construídos a partir da (Tabela 7.1).

DIMENSÃO DO ESPAÇO (R)	1	2	3	4	5
STRESS: DIALETOS KONSOID	0,56	0,26	0,15	0,09	-
STRESS: DIALETOS BIKOL	0,246	0,069	0,035	0,018	-
STRESS: NAÇÕES	0,388	0,187	0,112	0,049	0,022

TABELA 7.1 - Valores de STRESS obtidos em função da dimensão do espaço.

Observando o gráfico para os dialetos Konsoid não se nota nele nenhum "cotovelo que indique uma queda abrupta no valor do stress e uma posterior estabilização, a impressão que se tem é que ele segue decrescendo continuamente. Kruskal (1978) admitiu para esses dados, a solução unidimensional pelo procedimento intuitivo. Os dados dos dialetos Bikol foram por ele intuitivamente considerados bidimensionais por apresentarem um ligeiro "cotovelo" para $R = 2$, decrescendo mais lentamente para valores de R maio-

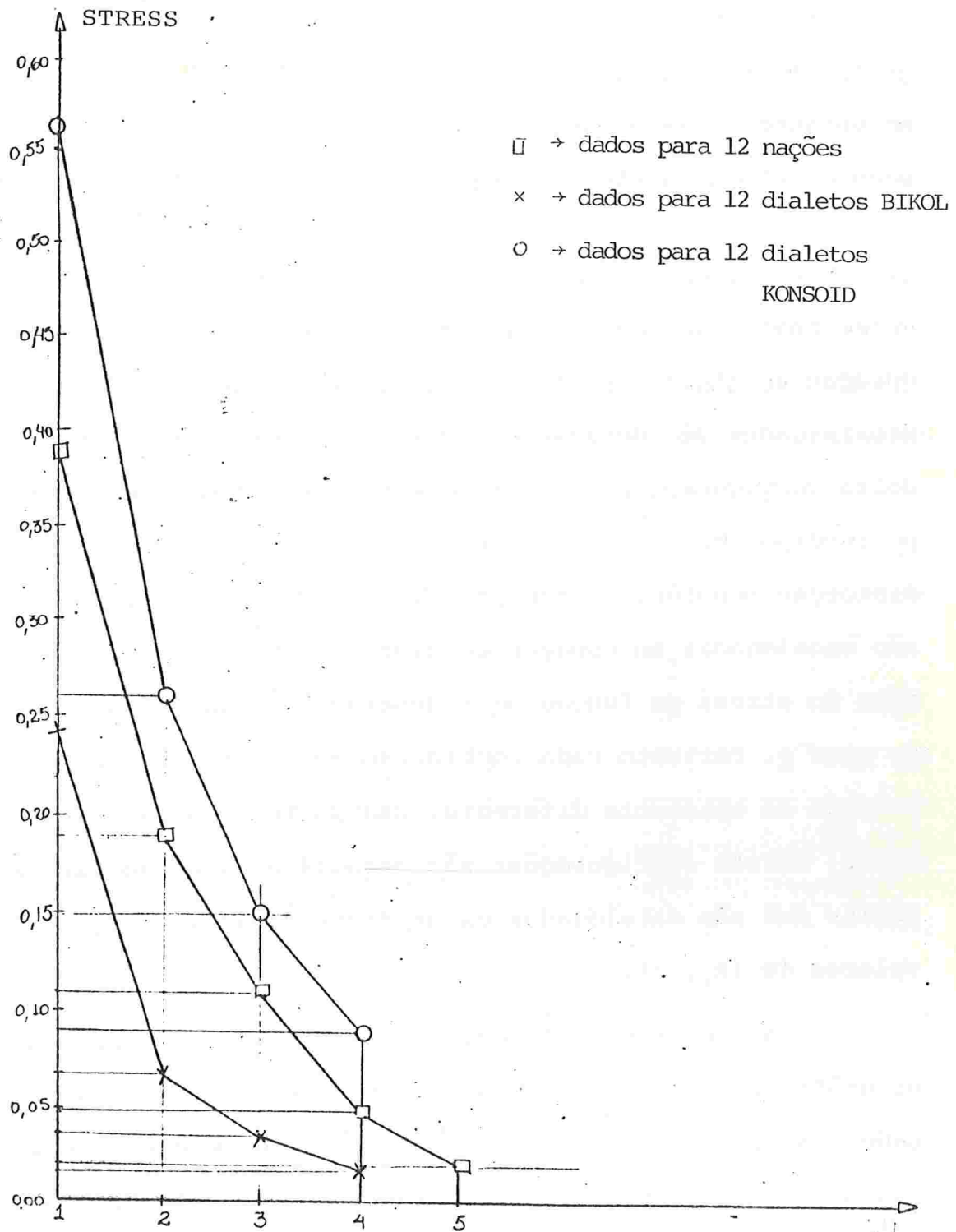


Fig. 7.1 - Gráficos do stress em função da dimensão do espaço para os dados do exemplo 7.1.

res que 2. Quanto aos dados das nações, o gráfico lhe sugeriu uma solução tetra-dimensional ou pentadimensional, no entanto é uma questão a parte usar uma solução num espaço de dimensão tão alta partindo de apenas 12 estímulos.

(2) Procedimento estatístico: O procedimento estatístico se baseia na idéia de que existe uma verdadeira configuração em algum espaço de dimensão R_t . Assume-se que as proximidades são geradas a partir das distâncias na verdadeira configuração, mas com um erro aleatório de algum tipo incorporado num nível controlável e talvez com alguma distorção monotônica incorporada. Quando as proximidades são escalonadas em espaços de dimensões diferentes, o gráfico do stress em função de R depende basicamente de R_t e do erro e . Portanto cada combinação de R_t e e produz um gráfico de aparência diferente. Usando técnicas de Monte Carlo, várias configurações são geradas e escalonadas, a partir daí são catalogados os gráficos de acordo com os valores de (R_t, e) .

Ao se usar o procedimento estatístico compara-se os gráficos obtidos pelos dados com esses gráficos catalogados. Na figura 7.2 pode-se observar os gráficos referentes aos três experimentos comparados com os gráficos catalogados. No 1º deles tem-se $R_t = 1$ e o erro e assumindo os valores $e = 0,06$, $e = 0,12$, $e = 0,25$, $e = \infty$. No 2º e

3^o tem-se respectivamente $R_t = 2$ e $R_t = 3$ para os mesmos valores de e .

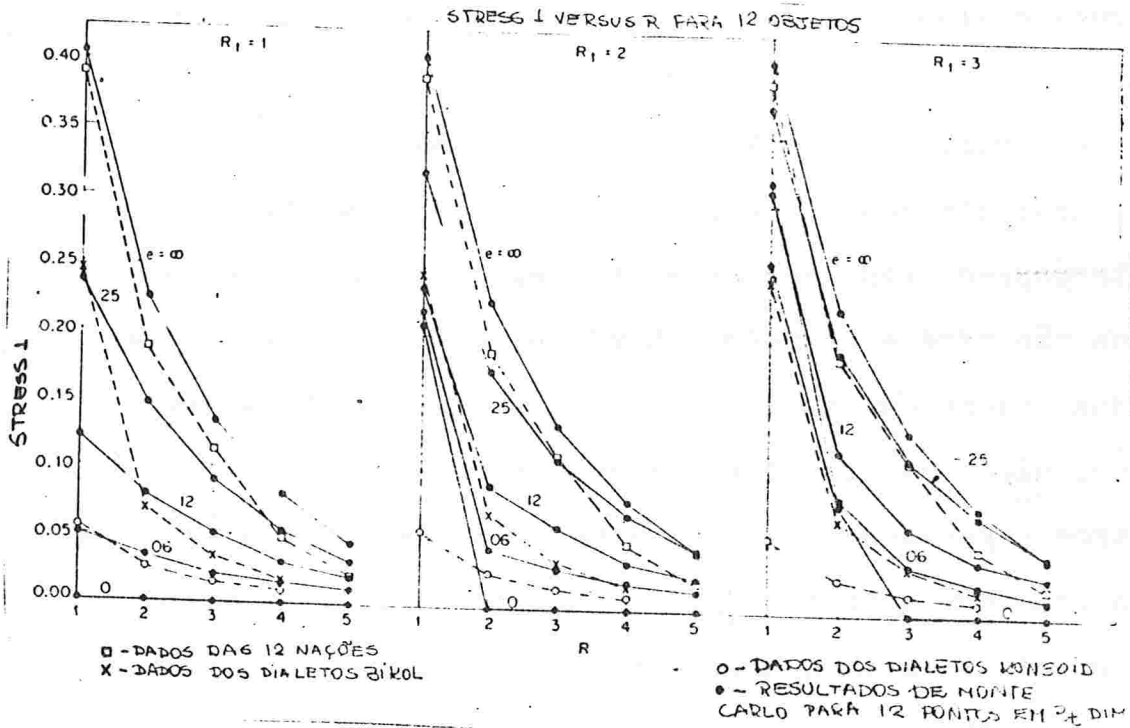


Figura 7.2: Procedimento Estatístico para estimar a di mensão do espaço.

Na comparação infere-se os valores de R_t e e verificando quais são as curvas catalogadas que mais se assemelham com as curvas obtidas. É óbvio que um erro muito alto obscurece qualquer característica dos dados, isto pode ser visto ao se observar que para $e = \infty$ as curvas são praticamente iguais para os três valores de R_t .

Os dados dos dialetos Konsoid tem a curva muito

semelhante àquela catalogada para $R_t = 1$ e $e = 0,06$ estando apenas ligeiramente abaixo dela, o que leva a concluir que estes dados devem se adaptar num espaço unidimensional e ainda que devem ter um nível de erro ligeiramente inferior a 0,06. A curva dos dados dos dialetos Bikol está entre a curva para $e = 0,06$ e aquela para $e = 0,12$ com $R_t = 2$, assemelhando-se bastante a ambas, o que leva a optar pelo espaço bidimensional. A curva para os dados das nações não parece se assemelhar a nenhuma das curvas catalogadas observadas na figura 7.2. Kruskal (1978) afirma que o seu nível de erro é certamente maior do que o dos dois outros experimentos e mais ainda diz que a curva se assemelha com aquela obtida para $R_t = 4$ e $e = 0,25$. Portanto para esses dados o espaço tetradimensional deve ser o que melhor se adapta.

Kruskal (1978) discute com mais detalhes procedimentos estatísticos que usam o stress como guia na escolha da dimensão do espaço apropriado ao E.M.D.

7.1.3 - Interpretabilidade da Solução

Nesta secção vamos comentar como a interpretabilidade da solução influi na escolha da dimensão do espaço.

Começa-se pela facilidade que possa existir em interpretar os eixos. Se uma solução num espaço de R dimen-

sões tiver uma interpretação satisfatória e a solução num espaço de $(R+1)$ dimensões não revelar nenhuma outra característica dos dados, é melhor usar a solução em R dimensões.

Por outro lado, pode-se ter uma configuração em 2 dimensões sem interpretação que passada para 3 dimensões tem uma interpretação total. Isto é possível pois a relação entre configurações em espaços de dimensões diferentes é mais complexa do que em outros métodos.

É necessário atentar para o fato de que se um investigador não consegue interpretar os eixos do espaço isto não significa necessariamente que eles não possam ser interpretados.

Para se interpretar soluções em espaços de dimensão maior ou igual a três muitas vezes procura-se observar as suas projeções em um espaço bi-dimensional de modo a facilitar uma análise visual dos resultados. No entanto esta forma de interpretação não é totalmente segura pois pode haver direções interpretáveis no espaço de dimensão alta que não são identificadas na projeção. Há técnicas estatísticas que podem auxiliar nesses casos (Kruskal, 1978).

Havendo dificuldades em se interpretar uma solução é aconselhável aumentar a dimensão do seu espaço, no entanto, uma dimensão alta apesar de certamente aumentar a adequação do ajuste, pode adaptar erros aleatórios aos dados,

levando a uma interpretação errônea.

Feitas todas essas considerações, se forem rejeitadas todas as possíveis dimensões do espaço, os dados deverão ser manuseados de forma diferente, ou então são inadequados para o E.M.D.

7.1.4 - Facilidade de Uso da Solução

Ao se escolher a dimensão do espaço deve-se considerar a facilidade de uso da solução do E.M.D. nesse espaço.

A solução num espaço bi-dimensional tem uma grande vantagem que é a possibilidade de uma análise visual da configuração obtida.

Quando se tratar de uma configuração num espaço de muitas dimensões, não é útil apenas examiná-la visualmente, pois fica difícil compreendê-la. Configurações em espaços de muitas dimensões são úteis somente quando são usadas técnicas suplementares para encontrar estruturas interessantes e compreensíveis. Estas estruturas são frequentemente direções da configuração, mas podem também ser planos, curvas, vizinhanças e outras estruturas sugeridas por Guttman (1954).

Sendo assim, mesmo quando os dados pedirem uma configuração num espaço de 4 dimensões pode-se querer usar

a configuração em espaços de 2 ou 3 dimensões para que facilite a compreensão da solução no espaço de dimensão mais alta. Além disso se os aspectos mais importantes e mais interessantes do espaço de dimensão mais alta já estiverem expostos na configuração no espaço bi-dimensional (Shepard , 1974) então em algumas situações é desejável usar apenas o espaço bi-dimensional.

Por outro lado, alguns aspectos importantes da estrutura não estão sempre confinados nas duas ou três primeiras dimensões. Em alguns casos pode-se até achar que as dimensões mais altas ajudam a esclarecer a interpretação do resultado bi-dimensional. Quando isto ocorre, a estrutura dos dados pode ser obscurecida tentando representá-los em poucas dimensões.

7.1.5 - Estabilidade da Solução

A solução obtida através do E.M.D, assim como os dados dos quais ela deriva estão virtualmente sujeitos a variações aleatórias.

Frequentemente, apesar dos dados estarem bem colhidos, a técnica utilizada para a obtenção da solução pode levar a mínimos locais que diferem do mínimo global apenas por pequenas modificações das posições dos pontos, porém alteram consideravelmente o aspecto da configuração. Nes

ses casos é aconselhável uma nova análise dos dados.

Uma forma simples de investigar a estabilidade da solução consiste em fazer novas análises nos dados subdivididos em grupos. Esta subdivisão, se houver um número grande de sujeitos participando do experimento, pode ser feita dividindo os sujeitos em sub-grupos aleatórios, e aplicando o E.M.D. aos dados de cada sub-grupo em separado. Ou ainda, pode-se eliminar alguns estímulos da matriz de dados e aplicar o E.M.D. aos estímulos restantes usando a idéia "jackknife" de J. N. Tukey (Mosteller e Tukey, 1977).

Outra forma de investigação consiste em gerar vários conjuntos de dados alternativos, acrescentando variação aleatória aos dados originais. Aplica-se o E.M.D. a todos esses conjuntos de dados e escolhe-se a solução que apresentar as características mais frequentes entre todas as soluções assim obtidas.

A decisão sobre o nº de dimensões do espaço pode se basear nessas considerações de estabilidade. Se o nº R de dimensões do espaço se mantiver para várias soluções, ele é um nº aconselhável mesmo que não diminua consideravelmente a medida de ajuste; a sua consistência é suficiente para aceitá-lo.

É interessante observar que as configurações obtidas através de métodos diferentes de coleta de dados podem

diferir entre si.

É necessário ser cuidadoso ao selecionar os estímulos que são submetidos ao E.M.D., é preferível procurar grupos de estímulos em que todos eles variam ao máximo segundo as dimensões que potencialmente possam aparecer. Isto porque não se deve considerar uma dimensão do espaço na solução do E.M.D. em que haja uma pequena variação dos estímulos ou então apenas a variação de um sub-grupo deles.

Há também uma regra prática que pode ser usada para escolher o nº de dimensões do espaço: o nº de estímulos deve ser pelo menos o dobro do nº de parâmetros (conjunto de coordenadas) a serem estimados, para assegurar um grau adequado de estabilidade estatística. Quando se usa um triângulo da matriz de proximidades, sem diagonal, isto significa que $\frac{n(n-1)}{2} \geq 2 nR$, ou equivalentemente $R \leq \frac{n-1}{4}$. Kruskal (1978) afirma ter verificado empiricamente esta regra prática para $R \leq 3$.

7.2 - INTERPRETAÇÃO DA CONFIGURAÇÃO

7.2.1 - Introdução

Uma etapa importante do E.M.D. consiste na interpretação da configuração obtida. Além da observação visual

são usados alguns recursos que tornam a configuração mais facilmente interpretável. Estes recursos são usados por solicitação do tipo de dados ou da própria configuração.

Através dos exemplos pode-se verificar a adequação dos recursos a cada tipo de situação.

7.2.2 - Rotação da Configuração.

O E.M.D. a "2 fatores" permite uma rotação dos eixos do Espaço Euclidiano obtido, portanto como em Análise Fatorial essa rotação pode permitir uma interpretação melhor da solução. Isso pode ser observado no exemplo abaixo apresentado por Kruskal (1978).

EXEMPLO 7.2 - O experimento consiste de um estudo piloto sobre a percepção das nações; foi realizado em março de 1968 (Wish 1971, Wish, Deutsch e Brener 1970). Foram submetidos ao experimento 18 estudantes de um curso de Medidas Psicológicas ministrado por Wish. Cada um dos estudantes atribui um grau de similaridade numa escala de 1 (muito diferentes) a 9 (muito similares) a cada um dos pares obtidos dentre 12 nações escolhidas. Primeiramente calcularam-se as médias das similaridades para cada um dos 66 ($C_{6,2}$) pares de nações. Estas medidas de similaridades foram arranjadas numa matriz para servirem de entrada para uma análise

lise por alguma técnica de E.M.D.

	BR	CCN	CUB	EGT	FR	IND	ISR	JP	CH	RUS	USA	YUG
Brasil	-											
Congo	4.83	-										
Cuba	5.28	4.56	-									
Egito	3.44	5.00	5.17	-								
França	4.72	4.00	4.11	4.78	-							
Índia	4.50	4.83	4.00	5.83	3.44	-						
Israel	3.83	3.33	3.61	4.67	4.00	4.11	-					
Japão	3.50	3.39	2.94	3.83	4.22	4.50	4.83	-				
China (Mainland)	2.39	4.00	5.50	4.39	3.67	4.11	3.00	4.17	-			
Rússia	3.06	3.39	5.44	4.39	5.06	4.50	4.17	4.61	5.72	-		
U.S.A.	5.39	2.39	3.17	3.35	5.94	4.28	5.94	6.06	2.56	5.00	-	
Yugoslavia	3.17	3.50	5.11	4.28	4.72	4.00	4.44	4.28	5.06	6.66	3.56	-

Figura 7.3: Matriz das similaridades obtidas entre 12 nações.

Através da observação da matriz pode-se perceber que Rússia e Yugoslavia foram percebidos como os mais semelhantes ($\delta_{ij} = 6,67$) por outro lado China e Brasil, E.U.A. e Congo foram julgados como sendo os pares mais diferentes ($\delta_{ij} = 2,39$).

A análise desses dados foi feita pelo programa KYST para E.M.D. num espaço bidimensional.

A saída do computador forneceu a lista das coordenadas das nações e a configuração obtida a partir dessas coordenadas.

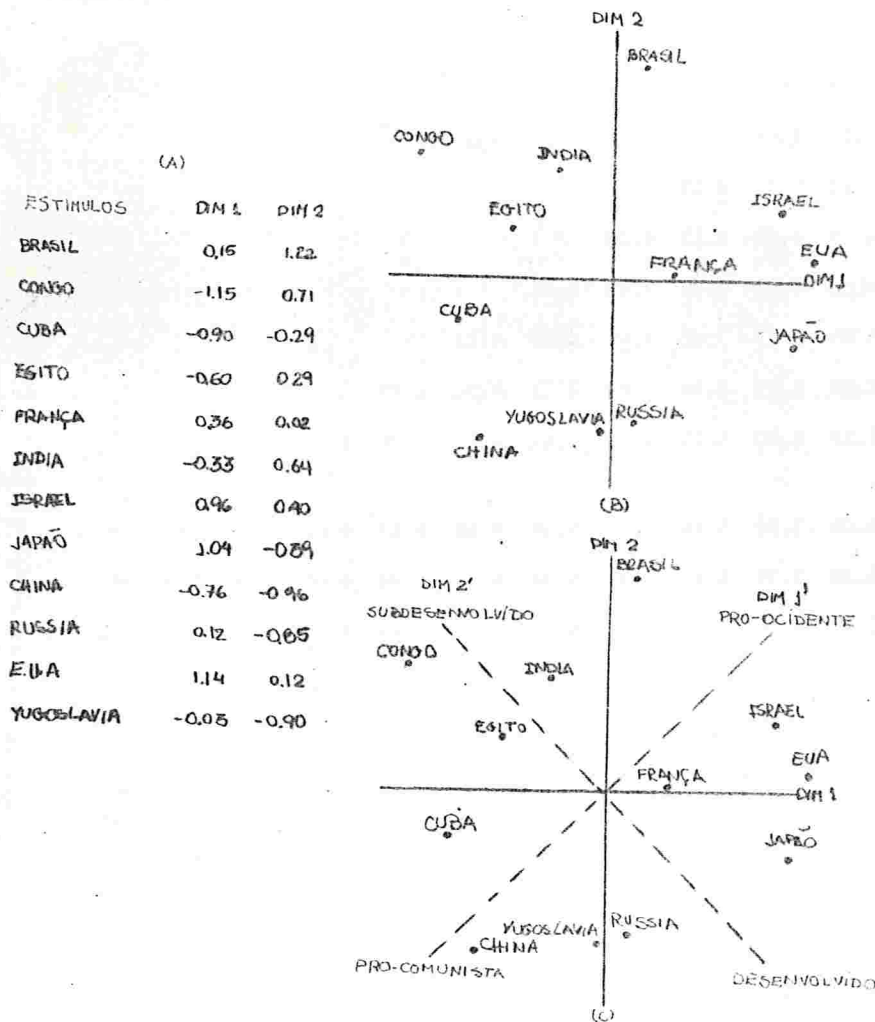


Figura 7.4 - Configuração obtida pelo programa KYST para os dados de similaridades entre nações (exemplo 7.2)

Observa-se que as dimensões I e II ilustradas na figura 7.4 (B) são dificilmente interpretáveis. Neste caso, os experimentadores fizeram o que é muito comum em E.M.D. ;

uma rotação dos eixos.

Os novos eixos estão ilustrados na figura 7.4 (C) , identificados como Dim I' e Dim II'. Observou-se que as nações pró-comunistas ficaram localizadas na parte superior da Dim I' e as pró-ocidente na outra parte. A Dim II' foi identificada como o desenvolvimento econômico das nações , observa-se as nações menos desenvolvidas na região superior da Dim II' e as mais desenvolvidas localizadas mais abaixo. Desta forma, pareceu-lhes razoável interpretar essas dimensões como sendo "Alinhamento Político" e "Desenvolvimento Econômico" respectivamente.

7.2.3 - Regressão Linear

As posições dos pontos na configuração obtida através de um E.M.D. podem estar relacionadas com as características dos estímulos que eles representam. Na verdade, a maior parte das vezes o E.M.D. é usado para descobrir essas características.

Em muitas aplicações úteis e interessantes as características encontradas são as esperadas.

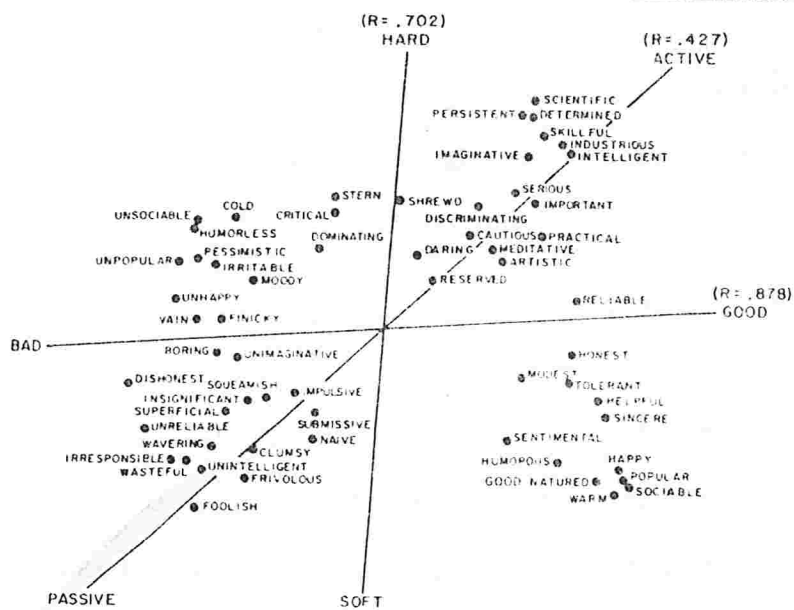
Geralmente essas características fazem parte de um rol plausível para o conjunto de estímulos. A função do E.M.D. é discriminar quais características são realmente importantes e quais são apenas plausíveis.

Uma forma de descobrir essas características é apenas visual, isto é, observando a configuração e procurando relembrar o rol de possíveis características do conjunto de estímulos. Esse método, no entanto, é falho por diversas razões: (1) É difícil observar a configuração em todas as direções e não só naquelas definidas pelos eixos coordenados sendo que muitas vezes numa configuração há mais direções interpretáveis do que aquelas definidas pelos eixos coordenados, isto se observa no exemplo abaixo.

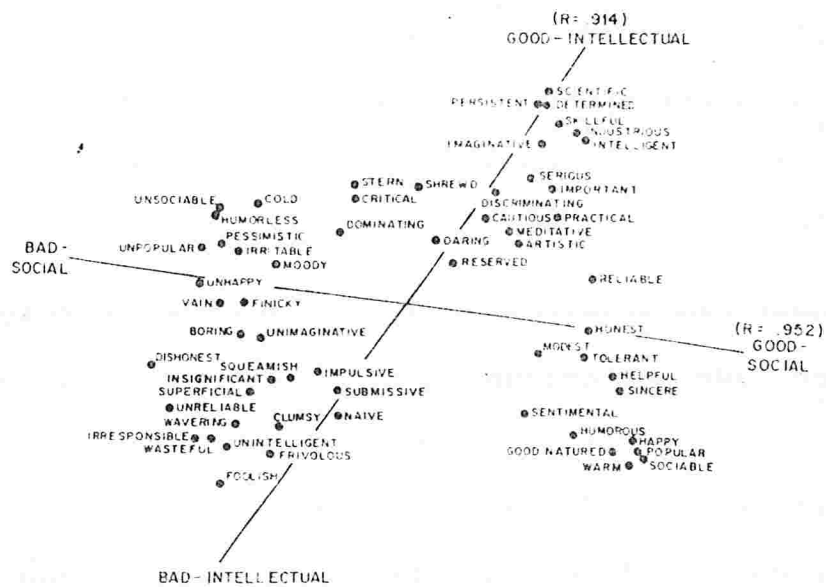
EXEMPLO 7.3 - Foi realizado por Rosenberg (1968) um estudo sobre termos que identificam traços de personalidade, pode-se perceber na figura 7.5 quatro direções interpretáveis e não apenas duas que constituiriam os eixos do espaço bi-dimensional. Essas direções são interpretadas como: " Bom Social versus Mal Social", "Bom Intelectual versus Mal Intelectual", "Duro versus Mole" e "Ativo versus Passivo".

(2) Se a configuração estiver num espaço de três ou mais dimensões a observação visual torna-se mais difícil, acaba-se observando apenas algum plano usado para a representação, geralmente são os planos coordenados. Nestas situações deve-se fazer um grande esforço para se observar outras direções ou então possuir uma excelente imaginação espacial.

(3) Ao ser observada alguma relação entre a configuração



(a)



(b)

Figura 7.5 - Configuração num espaço bidimensional obtida por Rosenberg para a percepção de 60 traços de personalidade. As linhas traçadas na configuração representam as direções interpretáveis (Figura retirada de Kruskal (1978)).

e alguma presumível característica é necessário se tomar cuidado para não incorrer num erro muito comum, gerado pela tendência humana de enxergar comportamentos sistemáticos onde não existem, que é o de super-valorizar o observado .

Exatamente para suprir as falhas da observação pode-se usar um procedimento estatístico de regressão linear quando for conhecido um rol de presumíveis características do conjunto de estímulos. Verifica-se a aplicação desse procedimento no exemplo abaixo apresentado por Kruskal (1978)

EXEMPLO 7.4 - O experimento faz parte de um estudo piloto que lida com a percepção do relacionamento entre vários problemas sociais (por exemplo, guerra, poluição, pobreza, crime e violência).

Quatorze sujeitos taxaram o grau de relação existente entre cada par obtido pela combinação de 22 problemas sociais (os sujeitos faziam parte de uma organização , sendo assim não constituíam uma amostra aleatória da população americana). Em seguida atribuíam notas a cada um dos estímulos isoladamente segundo 15 características desses estímulos. As notas eram atribuídas em escalas bipolares como por exemplo, uma das escalas que variava de 0 (não é de responsabilidade do governo local) a 9 (forte responsabilidade do governo local).

A matriz $\Delta = (\delta_{ij})_{22 \times 22}$ das médias das medidas de relação entre os estímulos foi submetida a um E.M.D.

O E.M.D. produziu os valores de stress de acordo com a tabela abaixo.

dimensão do espaço	1	2	3	5	6
stress	0,41	0,23	0,15	0,10	0,08

Considerou-se o mais apropriado o espaço Euclidiano de dimensão 3.

A partir daí, partiu-se para a interpretação dos eixos do espaço usando-se as escalas bipolares. O primeiro passo foi calcular as médias das notas que os sujeitos atribuíram em cada escala bipolar para cada estímulo. A figura 7.6 mostra essas médias para quatro dessas escalas além das coordenadas nos 3 eixos.

Em seguida usou-se regressão linear múltipla, fazendo uma regressão das médias obtidas em cada escala bipolar, em função das coordenadas resultantes do E.M.D.

Os coeficientes de correlação múltipla entre cada escala bipolar e as coordenadas do espaço tridimensional estão na figura 7.7.

Verificaram-se quatro coeficientes altamente significantes, os das 3^a, 4^a, 10^a e 12^a escalas bipolares, pa

Figura 7.6 - Coordenadas dos 22 problemas sociais nas 3 dimensões e as médias obtidas por eles em 4 escalas bipolares.

ESTÍMULOS	DIM 1	DIM 2	DIM 3	(3)	(4)	(10)	(12)
				AFETA A MIN.	AFETA A MAIORIA DAS PES SOAS	PROBLEMA TECNOLÓ- GICO	RESPONSA BILIDADE GOV. LOCAL
A. Exploração do consumidor	0,78	-0,31	0,71	5,71	8,07	2,54	6,92
B. Crime e violência	-0,67	0,15	0,24	3,50	6,07	2,77	8,46
C. Deterioração de Educação Pública	0,15	-0,30	-0,91	3,00	6,50	5,62	8,08
D. Abuso do Álcool e da Droga	-1,03	0,59	-0,29	2,43	5,29	3,23	7,54
E. Falta de prosperidade	0,86	-0,52	-0,59	2,29	4,71	2,46	7,77
F. Representação política desigual	-0,16	-1,03	0,12	4,50	5,57	2,77	7,85
G. Cuidados de saúde inadequados	-0,26	0,49	-0,90	2,57	6,29	5,00	7,62
H. Ineficiência do governo local	0,25	-0,83	-0,40	4,43	6,07	2,62	8,15
I. Injustiça no sistema judicial	-0,37	-0,71	0,91	2,64	5,57	2,00	8,08
J. Inflação	0,16	0,92	0,66	4,64	7,57	2,62	4,38
K. Discriminação no trabalho	-0,97	-0,10	0,45	2,21	4,43	1,62	7,62
L. Abusos de tecnologia	0,98	0,52	-0,01	4,71	7,29	6,85	5,77
M. Negligência nos transportes públicos	1,07	-0,83	-0,24	5,43	6,50	6,38	8,00
N. Superpopulação	0,18	0,66	-0,78	5,79	7,79	5,08	5,54
O. Poluição do ambiente	1,23	0,16	-0,42	6,86	8,21	7,38	8,38
P. Pobreza	-0,32	0,26	-0,13	3,07	4,64	3,85	7,54
Q. Voracidade privada e pública	0,49	-0,11	0,41	4,71	6,79	2,08	7,38
R. Racismo e fanatismo	-0,64	-0,32	0,40	3,71	5,29	2,08	7,38
S. Habitação abaixo da média	-0,14	-0,00	-0,04	2,43	4,29	6,15	7,54
T. Desemprego	-0,48	0,54	0,10	2,14	3,79	6,31	6,92
U. Decadência Urbana	0,03	0,04	-0,24	4,43	6,43	5,92	8,23
V. Guerra	0,59	0,74	0,87	5,36	7,79	4,46	5,08

POLOS POSITIVOS DAS ESCALAS	PESOS DA REGRESSÃO COSSENO DIRETORES			CORRELAÇÃO MÚLTIPLA
	DIM 1	DIM 2	DIM 3	
1. Muito importante	0,538	0,765	-0,354	0,491
2. Muito interessante	0,909	0,386	0,160	0,364
3. Afeta-me muito	0,994	0,045	0,105	0,841 **
4. Afeta a maioria das pessoas	0,932	0,342	0,049	0,804 **
5. Necessita de ação urgente	0,557	0,424	-0,714	0,301
6. Problema econômico	0,329	0,942	-0,072	0,523
7. Problema moral	-0,170	-0,117	0,979	0,519
8. Problema político	0,364	-0,685	-0,631	0,300
9. Problema organizacional	0,320	-0,474	-0,820	0,638 *
10. Problema tecnológico	0,612	0,416	-0,672	0,767 **
11. Responsabilidade do governo federal	-0,032	0,736	0,677	0,437
12. Responsabilidade do governo local	-0,262	-0,863	-0,432	0,767 **
13. Responsab. das instituições s/ lucro	0,592	0,435	-0,679	0,409
14. Responsab. das instituições c/ lucro	0,555	0,767	0,320	0,599 *
15. Responsab. das pessoas diretamente afetadas	0,302	-0,859	-0,414	0,386

** = $F_{3,18}$ significativa a nível menor que 0,001

* = significativa a nível menor que 0,05

Figura 7.7: Regressão Múltipla das médias nas Escalas Bipolares com relação às dimensões obtidas no E.M.D. dos Problemas Sociais

ra estas escalas pode-se considerar que as dimensões fazem sentido. Para duas escalas (9^a e 14^a) as dimensões são significantes e para as demais não há coeficientes associados às coordenadas que produzam uma correlação significativa entre elas e as três dimensões.

As três primeiras colunas da figura 7.7 apresentam os pesos ótimos correspondentes a cada correlação múltipla; são os valores dos cossenos diretores, isto é, coeficientes de regressão normalizados, sendo assim a soma de seus quadrados é 1 para cada escala bipolar.

Há duas condições necessárias para que as notas atribuídas numa escala bipolar levem a uma interpretação satisfatória de uma determinada dimensão. (1) a correlação múltipla para a escala deve ser alta, indicando assim que a escala pode ser bem ajustada às coordenadas da configuração. (2) a escala deve ter um coeficiente de regressão alto na dimensão, indicando assim que o ângulo entre a dimensão e a direção dela é pequeno.

É desejável portanto que se obtenha coeficientes de correlação múltipla próximos de 0,90 para uma boa interpretação da dimensão; correlações acima de 0,70 podem ser suficientes em determinadas situações. É claro que quanto maior o coeficiente, mais confiança pode-se ter na interpretação da dimensão. Exige-se que a correlação seja estatisticamente significativa pelo menos ao nível de 0,01.

Observando-se a tabela da figura 7.7 a primeira dimensão pode ser imediatamente interpretada. A 3^a escala é altamente correlacionada com as dimensões e tem cosseno diretor com a 1^a dimensão de 0,94 o que corresponde a um ângulo de 6^o; pode-se observar na figura 7.8 (a) que isso torna a direção da escala bastante próxima da direção da 1^a dimensão. O mesmo ocorre com a 4^a escala. Sendo que estas duas escalas se relacionam com os efeitos dos problemas sobre a própria pessoa ou sobre o povo em geral; a 1^a dimensão é interpretada como "grande versus pequeno efeito sobre o povo".

As figuras 7.8(a) e 7.8 (b) mostram respectivamente a configuração nos planos determinados pelas dimensões 1 e 2 e pelas dimensões 2 e 3. Percebe-se alguma significância em se interpretar a dimensão 2 como "Responsabilidade do governo local versus do governo não local" pois a escala 12 tem coeficiente de correlação significativa e o cosseno do ângulo com esta dimensão é grande. O que mais se aproxima da dimensão 3 é uma interpretação de "Problemas tecnológicos versus problemas não tecnológicos" é o que pode ser percebido observando na figura 7.7 a 10^a escala.

Uma interpretação melhor dessas dimensões pode ser obtida pensando-se em outras variáveis que distingam melhor os estímulos que se localizem nos extremos opostos das várias direções dos planos da configuração, em seguida es-

estruturas podem ser observadas no espaço multi-dimensional ao invés da interpretação dos eixos.

Pode haver na configuração vizinhanças ou regiões que tenham algum significado quando associadas com características de alguns dos estímulos. Isto frequentemente ocorre quando se obtém uma solução bidimensional para dados apropriados a um espaço de dimensão maior. Uma razão importante para que a interpretação por vizinhança revele outros comportamentos nos dados é o fato dela focalizar primariamente as distâncias pequenas (grandes similaridades) enquanto que a interpretação das dimensões se baseia fundamentalmente nas distâncias grandes.

Segundo Kruskal (1978), Guttman argumenta que uma análise por vizinhanças ou comportamental é preferível ao procedimento tradicional baseado nas dimensões. Isto se verifica no seu espaço que analisa correlações entre testes mentais, este revela comportamentos importantes dos dados (simplexes, circumplexes e radexes) que haviam sido perdidos em uma análise fatorial anterior. Além disso, o próprio Kruskal acredita na interpretação por vizinhanças como sendo um auxílio na interpretação da configuração que suplementa e torna mais clara as dimensões do espaço e não compete com elas.

Justificando seu ponto de vista apresenta um exem-

plo onde a interpretação por vizinhanças lhe parece mais razoável do que a das dimensões do espaço.

EXEMPLO 7.5 - Este exemplo é parte de um experimento descrito por Wish (1970) e por Wish e Carroll (1974). As medidas de proximidade foram obtidas através de julgamentos da similaridade cultural entre 21 nações. A figura 7.9 (a) apresenta a solução obtida num espaço bidimensional.

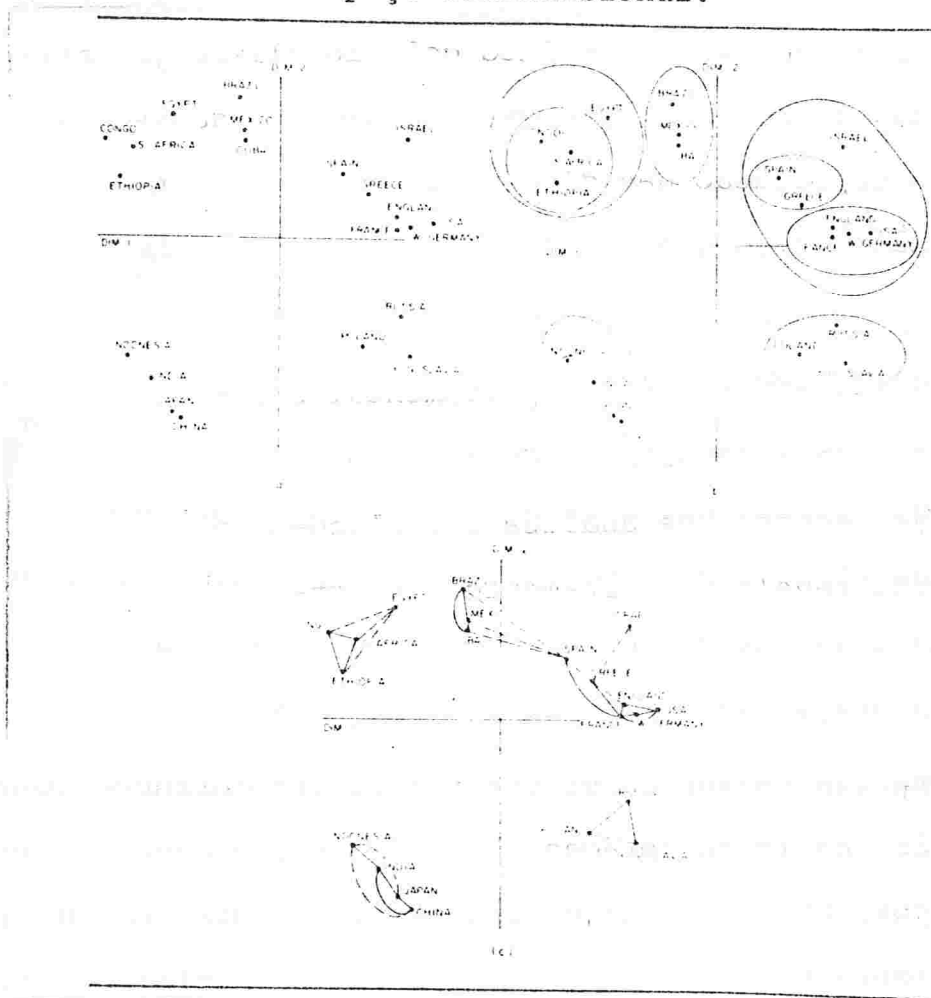


Figura 7.9

Uma interpretação das dimensões deste espaço consideraria que o eixo horizontal separa as culturas Europeias das não-Europeias, ou então culturas brancas das não-brancas ou ainda países do Norte dos países do Sul. Forçando a imaginação poder-se-ia dar uma interpretação de Leste versus Oeste para a dimensão vertical desde que isso confronte os países do Leste da Ásia e o Leste da Europa com os outros.

Parece menos forçado, no entanto, observar apenas que os países de culturas similares tendem a formar conglomerados no espaço, não sendo estritamente no sentido geográfico, por exemplo E.U.A. e México estão bastante distantes o mesmo ocorrendo com Israel e Egito. Os E.U.A. formam um conglomerado com os países da Europa Ocidental dos quais sua cultura é originária, enquanto que o México forma um conglomerado com os países da América Latina. Nota-se que os países da África estão na parte superior esquerda, os do Leste da Ásia na parte inferior esquerda e os da Europa Oriental na porção inferior direita do espaço.

Uma das formas de se interpretar vizinhanças numa configuração envolve a aplicação de técnicas de Análise de Conglomerados. Existem métodos sistemáticos para se encontrar conglomerados baseando-se diretamente na matriz de proximidades (Kruskal 1978). Os conglomerados podem ser determinados através de "loops" em torno dos pontos que cor-

respondem aos estímulos. Uma vez determinados pode-se procurar observar quais as características comuns aos estímulos que integram o mesmo conglomerado. Isto geralmente é feito de forma subjetiva como um ato de interpretação criativa, em seguida pode-se atribuir ao conglomerado um rótulo que explique o significado da sua localização.

Dessa forma pode-se prever em que conglomerado se localizariam outros estímulos, que não foram analisados, em caso de coleta e estudo de mais dados.

A figura 7.9 (b) mostra os conglomerados obtidos através de uma técnica de Análise de Conglomerados Hierárquica (Johnson, 1967) realizada sobre a configuração bidimensional das similaridades culturais entre as nações. O fato dos "loops" serem bastante pequenos mostra uma boa correspondência entre as vizinhanças do E.M.D. e os conglomerados determinados pela Análise de Conglomerados Hierárquica.

Deve-se observar que o fato de dois estímulos estarem próximos na configuração não representa obrigatoriamente uma alta medida de similaridade, pois vizinhanças em espaços de dimensão baixa podem estar representando mal os dados dos quais a configuração deriva (ver em Kruskal (1978) experimento de Rosemberg, Nelson e Vivikananthau (1978)).

É necessário portanto suplementar a proximidade na

configuração com a proximidade baseada diretamente na matriz de dados, além de que o E.M.D. funciona melhor na representação de grandes distâncias (estrutura global) do que na representação de distâncias pequenas (estrutura local). Graef e Spence (1976) mostraram que descartando apenas o terço em que as dissimilaridades são pequenas ou o terço em que elas são médias não há alteração na reconstrução do espaço multidimensional, ao passo que descartando o terço mais dissimilar de estímulos há uma grande alteração na solução do E.M.D.

Uma forma de juntar as informações da matriz de dados ao que se observa na configuração consiste em ligar por uma linha os pares de pontos que representam estímulos cuja medida de proximidade excede um determinado limite fixado, como na figura 7.9 (c). O objetivo é que haja linhas ligando pontos próximos na configuração pois estas representam pequenas dissimilaridades ou grandes similaridades. Geralmente usa-se mais que um limite, representando-se por linhas cheias, tracejadas ou pontilhadas.

Quando há linhas muito longas unindo os pontos da configuração, isso indica uma discrepância entre a proximidade na matriz de dados e na configuração. Se a configuração for bidimensional cada linha longa identifica um resíduo grande que deve ser levado em consideração na interpretação da configuração. Quando as linhas longas esti-

verem traçadas em planos de espaços de dimensão maior, a discordância entre os dados e as vizinhanças pode ser devida às outras dimensões.

Os conglomerados em espaços bidimensionais, provavelmente só são válidos se estiverem de acordo com as linhas traçadas, isto é, dentro do conglomerado, a maioria dos pontos devem estar ligados uns aos outros e poucos dentre eles podem estar ligados a pontos de outros conglomerados.

A figura 7.9 (c) mostra as linhas traçadas no espaço de acordo com os dados de similaridade cultural entre as nações. As linhas cheias mostram os pares que os dados indicam ser mais similares, as linhas tracejadas indicam os moderadamente similares. Uma vez que as linhas estão unindo pontos que estão próximos no espaço, a representação espacial é bastante compatível com os dados.

As linhas mostram outras tendências com mais clareza, por exemplo, a similaridade da Espanha com os países da América Latina de um lado e com os da Europa Ocidental por outro lado. Observa-se no entanto que não há linhas conectando os países da América Latina com os da África apesar deles estarem próximos na configuração. Isso mostra uma certa discordância entre as distâncias espaciais e os dados de proximidade (stress = 0,128).

Além do traçado de linhas para auxiliar na interpretação por vizinhanças Kruskal (1978) sugere que se façam novos E.M.D. para subconjuntos de estímulos que estejam próximos no espaço. Quando há um nº suficiente de estímulos envolvidos essas análises podem revelar dimensões que estiveram obscurecidas quando foram analisados todos os estímulos simultaneamente.

7.2.5 - Outras formas de interpretação

Ao se interpretar uma configuração deve-se recorrer também a uma certa imaginação criadora baseada muitas vezes num certo conhecimento prévio dos estímulos que estão sendo estudados.

Um exemplo interessante neste aspecto foi relatado por Kruskal (1978).

EXEMPLO 7.6 - Os dados do experimento foram coletados por Rothkopf (1957) são medidas de confusão entre 36 sinais auditivos do código Morse. Cada sinal do código consiste de uma sequência de pontos e traços tais como ... para a letra k e para o nº 2. Os indivíduos submetidos ao experimento não conheciam o código Morse. Uma máquina produzia pares de sinais com intervalos entre cada sinal de

1,4 segundos, esses pares eram ouvidos pelos sujeitos e era exigido deles que avisassem se os dois sinais ouvidos eram iguais ou diferentes.

Foi construída uma matriz de proximidades Δ onde cada elemento δ_{ij} representava a porcentagem de respostas "igual" quando eram apresentados aos sujeitos o estímulo i seguido do estímulo j , isso para um total de 150 sujeitos. Observe-se que os nomes das letras convencionados para os sinais do código Morse não entraram de maneira alguma no experimento; foram usados na tabela que representa a matriz de dados simplesmente por conveniência (figura 7.10).

A diagonal principal da matriz corresponde aos pares que eram realmente iguais, portanto, espera-se para eles valores de δ_{ij} bem altos e isto realmente ocorre sendo o menor deles 84% correspondente ao par BB. Fora da diagonal os valores correspondem a pares diferentes portanto esperam-se δ_{ij} pequenos e verificam-se vários deles abaixo de 10%.

Essa matriz é grosseiramente simétrica. Isto é, se o valor δ_{ij} é baixo, o correspondente δ_{ji} também o é, o mesmo acontecendo se ele for alto.

Tem-se na figura 7.11 a configuração obtida aplicando-se E.M.D. a esses dados (esta aplicação foi retirada

Figura 7.10 - Matriz de similaridades obtidas por Rothkopf para os símbolos do Código Morse.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	1	2	3	4	5	6	7	8	9	0																																				
02	04	06	13	03	14	10	13	46	05	22	03	25	34	06	06	09	35	23	06	37	13	17	12	07	03	02	07	05	05	04	06	06	02	03	03																																				
06	64	37	31	05	28	17	21	05	19	34	40	06	10	12	22	25	16	18	02	18	34	08	64	30	42	12	17	14	40	32	74	45	17	04	04																																				
04	38	67	17	04	29	13	07	11	19	24	35	14	03	09	51	34	24	19	06	06	kk	k4	32	82	38	13	15	31	14	10	30	28	24	18	12	12																																			
08	62	17	06	07	23	40	36	09	13	81	56	08	07	09	27	09	45	29	06	17	20	27	40	15	33	03	09	06	11	09	08	10	05	06	06																																				
06	13	14	06	97	02	04	04	17	01	05	06	04	04	05	01	05	10	07	67	03	02	05	06	05	04	03	05	02	03	05	02	04	02	03	05																																				
09	41	33	19	02	90	10	29	05	33	16	50	07	06	10	42	12	35	14	02	21	27	25	19	27	13	06	16	47	25	26	24	21	05	03	05																																				
04	18	27	38	01	14	90	06	05	22	33	16	14	13	82	52	23	21	05	03	15	14	32	21	23	39	15	14	05	10	04	10	17	23	20	11																																				
03	45	23	25	09	32	08	87	10	10	09	29	05	08	08	14	08	17	37	04	36	59	09	33	14	11	03	09	15	43	70	35	17	04	03	03																																				
04	07	07	13	10	08	06	12	93	03	05	16	13	30	07	03	05	19	35	16	10	05	08	02	05	07	02	05	08	09	06	08	05	02	04	05																																				
07	09	38	09	02	24	18	05	04	85	22	31	08	03	21	63	47	11	02	07	09	09	09	22	32	28	67	66	33	15	07	11	28	29	26	23																																				
05	24	38	73	01	17	25	11	05	27	91	33	10	12	31	14	31	22	02	02	23	17	33	63	16	18	05	09	17	06	08	18	14	13	05	06																																				
02	69	43	45	10	24	12	26	09	30	27	86	06	12	10	15	20	19	05	16	19	20	31	25	69	12	13	17	15	26	29	36	16	07	03	03																																				
24	12	05	14	07	17	29	08	08	11	23	08	96	62	11	10	15	20	07	09	13	04	21	09	18	08	05	07	06	06	05	07	11	07	10	04																																				
31	04	13	30	08	12	10	16	13	03	16	08	59	93	05	09	05	28	12	10	16	04	12	04	06	11	05	02	03	04	04	06	02	02	10	02	02																																			
07	07	20	06	05	09	76	07	02	39	26	10	04	08	86	37	35	10	03	04	11	14	25	35	27	19	17	07	07	06	18	14	11	20	12	12	12																																			
05	22	33	12	05	36	22	12	03	78	14	46	05	06	21	83	43	23	09	04	12	19	19	41	30	34	44	24	11	15	17	24	23	25	13	13	13																																			
08	20	38	11	04	15	10	05	02	27	23	26	07	06	22	51	91	11	02	03	26	14	12	37	50	63	34	32	17	12	09	27	40	58	37	24	24	24																																		
13	14	16	23	05	34	26	15	07	12	21	37	14	12	12	29	08	87	16	02	06	23	62	14	12	13	07	12	13	04	07	12	07	09	0	02	02	02																																		
17	24	05	30	11	26	05	59	16	03	13	10	05	17	06	06	08	18	96	09	56	24	12	10	06	07	08	02	02	15	28	09	05	05	05	02	02	02																																		
13	10	01	05	46	03	06	06	14	06	14	07	06	05	06	11	04	04	07	96	08	05	04	02	02	06	05	05	03	03	03	08	07	06	14	06	14	06																																		
14	29	12	32	04	32	11	34	21	07	44	32	11	13	06	20	12	40	51	06	93	57	34	17	09	11	06	16	34	10	09	09	07	04	03	03	03	03																																		
09	21	30	22	09	36	25	15	04	25	29	18	15	06	26	20	25	61	12	04	19	20	86	22	25	22	10	22	19	16	05	09	11	06	03	07	04	05																																		
07	64	45	19	03	28	11	06	01	35	50	42	10	08	24	32	61	10	12	03	12	17	21	91	48	26	12	20	24	27	16	57	29	16	17	06	17	06																																		
09	23	62	15	04	26	22	09	01	30	12	14	05	06	14	30	52	05	07	04	06	13	21	44	86	23	26	44	40	15	11	26	22	33	23	16	16	16																																		
03	46	45	18	02	22	17	10	07	23	21	51	11	02	15	59	72	14	04	03	09	11	12	36	42	87	16	21	27	09	10	25	66	47	15	15	15	15																																		
02	05	10	03	03	05	13	04	02	29	05	14	09	07	14	30	28	09	04	02	03	12	14	17	19	22	84	63	13	08	10	08	19	32	57	55	11	11																																		
07	14	22	05	04	20	13	03	25	26	09	14	02	03	17	37	28	06	05	03	06	10	11	17	30	13	62	89	54	20	05	14	20	21	16	11	11	11																																		
03	06	21	05	04	32	06	12	02	23	06	13	05	02	05	37	19	09	07	06	04	16	06	22	25	12	18	64	86	31	23	41	16	17	08	10	10	10																																		
06	19	19	12	06	25	14	16	07	21	13	19	03	03	02	17	29	11	09	03	17	55	08	37	24	03	05	26	44	89	42	44	32	10	03	03	03	03	03																																	
08	45	15	14	02	45	04	67	07	14	04	41	02	00	04	13	07	09	27	02	14	45	07	45	10	10	14	10	30	69	90	42	24	10	06	05	05	05																																		
07	80	30	17	04	23	04	14	02	11	11	27	06	02	07	16	30	11	14	03	12	30	09	58	38	39	15	14	26	24	17	86	69	14	05	14	05	14	14																																	
06	33	22	14	05	25	06	04	06	24	13	32	07	06	07	36	39	12	06	02	03	13	09	30	30	50	22	29	18	15	12	61	85	70	20	13	7	7	7																																	
03	23	40	06	03	15	15	06	02	33	10	14	03	06	14	12	45	02	06	04	06	07	05	24	35	50	42	29	16	16	09	30	60	89	61	26	8	8	8																																	
03	14	23	03	01	06	14	05	02	30	06	07	16	11	10	31	32	05	06	07	06	03	08	11	21	24	57	39	09	12	04	11	42	56	91	78	9	9	9																																	
09	03	11	02	05	07	14	04	05	30	08	03	02	03	25	21	29	02	03	04	05	03	02	12	15	20	50	26	09	11	05	22	17	52	81	94	0	0	0																																	
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	1	2	3	4	5	6	7	8	9	0	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	1	2	3	4	5	6	7	8	9	0

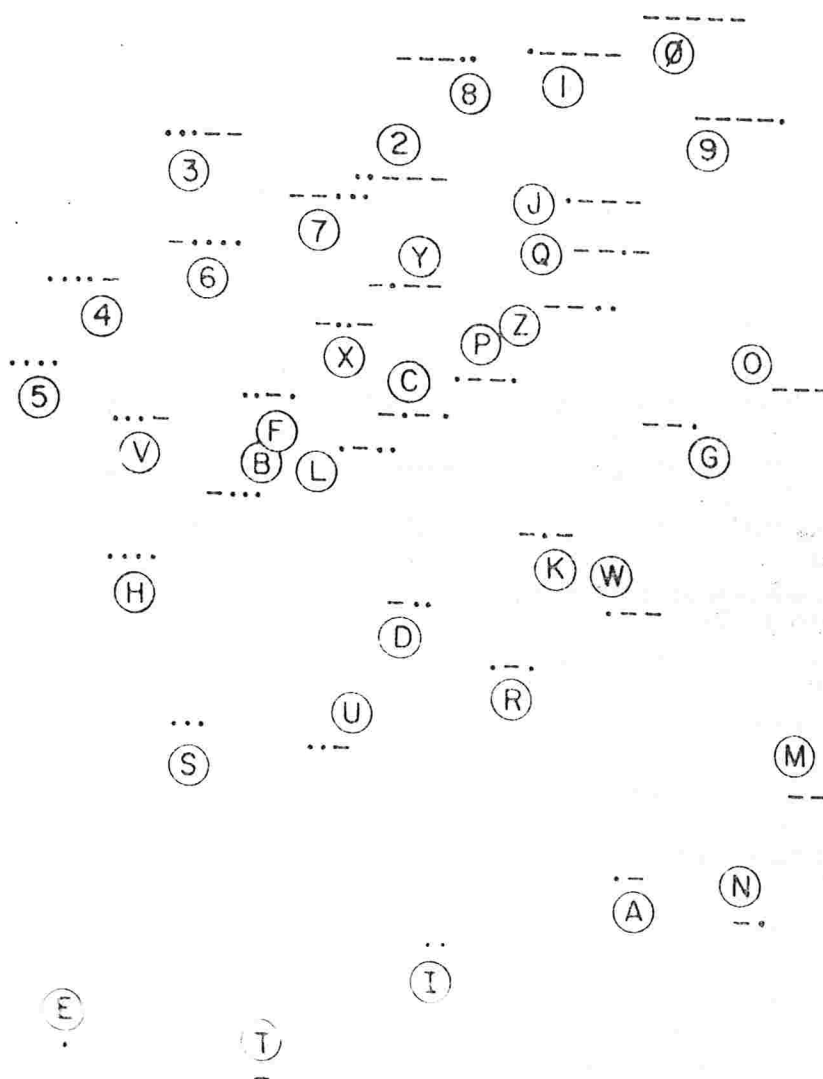


Figura 7.11 - Configuração resultante do E.M.D. das similaridades da Fig. 7.10.

da por Kruskal dos trabalhos de Shepard, 1963, que analisou outros três conjuntos de dados sobre o código Morse). Os 36 pontos obtidos pelo E.M.D. são os centros dos círculos representados na figura. Para melhor entender a configuração ao lado de cada círculo estão representadas as sequências de traços e pontos correspondentes ao estímulos auditivos ouvidos pelos sujeitos.

Pode-se observar uma boa concordância entre a matriz e a configuração. Tomando-se por exemplo as letras B e X tem medidas de similaridade altas 84% e 64% e encontram-se bastante próximas na configuração, por outro lado as letras E e O tem medidas de similaridade baixas 3% e 5% e encontram-se bem afastadas na configuração. Para outros pares de sinais o fenômeno se repete havendo uma correspondência entre medidas de similaridade e distâncias na configuração.

Uma interpretação muito engenhosa para esta configuração foi dada por Shepard e é mostrada na figura 7.12. Ele observou que para esse conjunto de símbolos havia duas variáveis muito correlacionadas que não poderiam ser distinguidas, por esses dados, são "comprimento do sinal" e "número de componentes". No entanto estava claro para ele que alguma variável de comprimento desempenhava um papel importante.

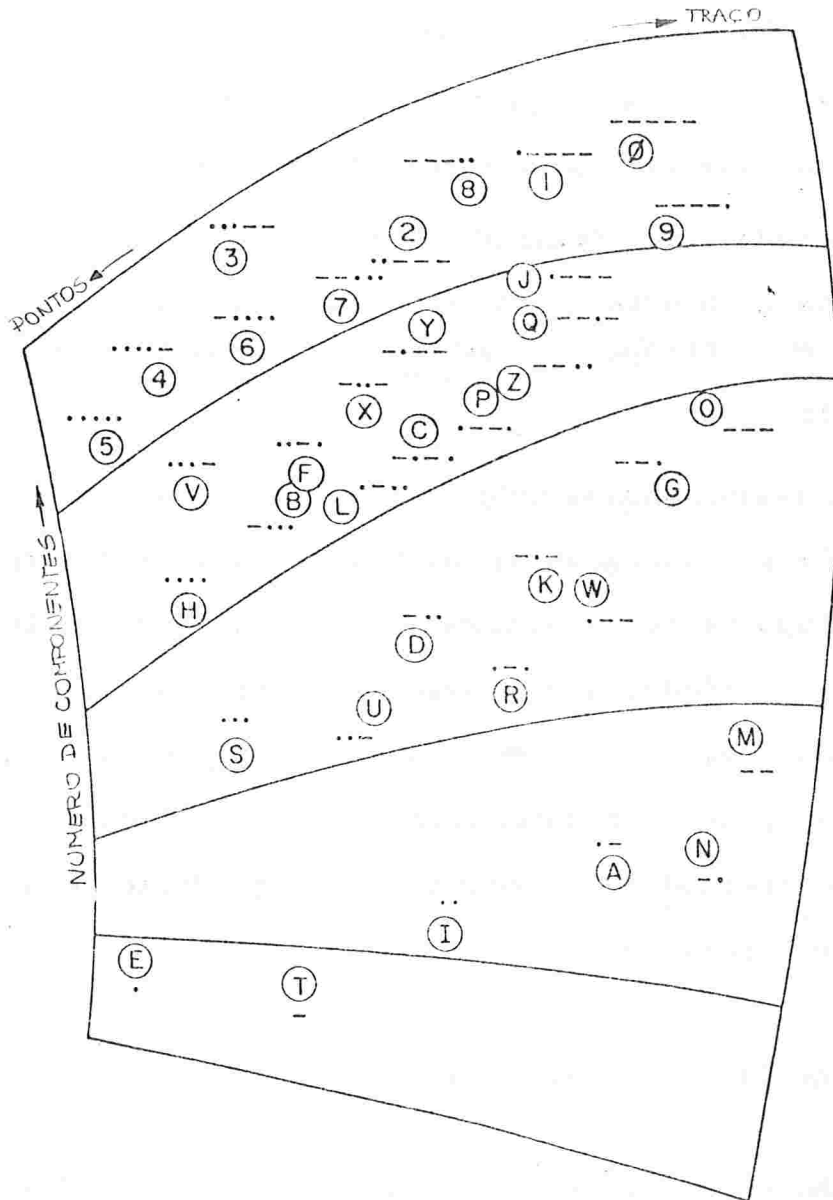


Figura 7.12 - Interpretação de Shepard para a configuração dos símbolos do Código Morse.

Shepard construiu 5 faixas na configuração de modo que os estímulos pertencentes a mesma faixa tivessem o nº de componentes igual e definiu assim uma das dimensões "nº de componentes". Notou também que os estímulos localizados mais a direita na configuração tinham predominância de traços e aqueles mais à esquerda tinham predominância de pontos, estava assim definida a segunda dimensão "tipos de componentes". Observa-se na 5ª faixa bem à direita o \emptyset com 5 traços, em contraposição na 1ª faixa bem à esquerda o E com um único ponto.

Quando Shepard publicou seu trabalho, o fato das duas variáveis mostradas na configuração dele desempenharem um papel importante na percepção dos sinais do código Morse era uma novidade, apesar que o assunto já havia sido muito estudado. Deve-se observar no entanto, que as suas conclusões não podem ser generalizadas, são válidas para sujeitos não treinados na escuta do código Morse, nas condições desse experimento.

7.2.6 - Auxílio na análise da solução

Ao se analisar uma solução do E.M.D. pode-se recorrer a matriz de resíduos $[r_{ij}]_{n \times n}$ onde $r_{ij} = d_{ij} - d_{ij}^*$ para auxiliar na interpretação desta solução. No exemplo 7.2 foi usado esse recurso após ser observado o ajuste da

configuração dos dados. Obteve-se um stress de 0,19 para a configuração no espaço bi-dimensional, apesar deste ser bem melhor do que o stress de 0,39 obtido para a configuração num espaço unidimensional, ele ainda indica que o ajuste está longe de ser perfeito. Dado o pequeno número de sujeitos (18) do experimento não ficou claro se o 0,19 refletia falta de confiabilidade ou a existência de alguma estrutura adicional na matriz de dados. Partiu-se então para a análise da matriz de resíduos.

	BR	CON	CUB	EGT	FR	IND	ISR	JP	CH	RS	USA	YUG
Brasil	-											
Congo	.28	-										
Cuba	.78	-.09	-									
Egito	-.04	-.36	-.40	-								
França	.10	.14	-.07	-.13	-							
India	-.37	-.22	-.27	-.18	-.66	-						
Israel	-.37	.05	.40	.45	-.64	-.05	-					
Japão	.25	.36	-.15	.18	-.57	.35	-.25	-				
China (Mainland)	.00	.20	-.10	-.10	-.11	.30	.11	.52	-			
Russia	-.02	.00	.12	-.01	-.15	.20	.15	-.10	-.10	-		
U.S.A.	.43	.00	-.01	-.34	.17	.20	-.09	.09	.00	.35	-	
Yugoslavia	.03	.38	.12	-.04	-.13	.05	.28	-.18	-.32	.00	-.04	-

Figura 7.13 Resíduos da solução num espaço bidimensional do E.M.D. das similaridades entre 12 nações.

Observando-se a matriz de resíduos (fig. 7.13) percebe-se que o espaço bidimensional superestima as distâncias entre as nações geograficamente próximas e subestima as distâncias entre as nações geograficamente ou culturalmente distantes. Por exemplo, de acordo com as medidas de similaridades Cuba deveria estar próxima do Brasil no espaço ($r_{ij} = 0,78$) e a China deveria estar próxima do Japão ($r_{ij} = 0,52$). Por outro lado, a França deveria estar bem distante da Índia e de Israel ($r_{ij} = - 0,66$ e $r_{ij} = - 0,64$ respectivamente).

O comportamento sistemático da matriz de resíduos indica que as duas dimensões não representam toda a estrutura da matriz de proximidades. Confirmando esta idéia tem-se o stress caindo para 0,11 num espaço tridimensional e para 0,05 num espaço tetradimensional. Se um número maior de nações tivesse sido incluído nesse estudo, seria razoável concluir que uma representação tetradimensional seria a mais adequada. No entanto com apenas 12 estímulos é improvável que seja determinado um espaço de mais de duas ou três dimensões. Foi feito um estudo mais amplo por Wish, Deutsch e Biener (1970, 1972) que obteve além das dimensões da fig. 7.4 outras duas que refletiam as proximidades culturais e geográficas entre as nações.

7.2.7 - Representação de uma configuração num espaço tridimensional.

Para encerrar este capítulo seria interessante observar uma configuração num espaço tridimensional. A configuração apresentada faz parte de um trabalho realizado por Gary A. Mauser (1972) sobre a Previsão de Comportamentos para a Substituição Eleitoral.

A representação é feita por bandeirolas de diversas alturas representando assim a 3ª dimensão. Cada bandeira está representando um político americano de um conjunto de candidatos dos quais eleitores americanos julgaram as similaridades. O programa usado para o E.M.D. foi o MDSCAL, o stress obtido foi 0,183 (fig. 7.14).

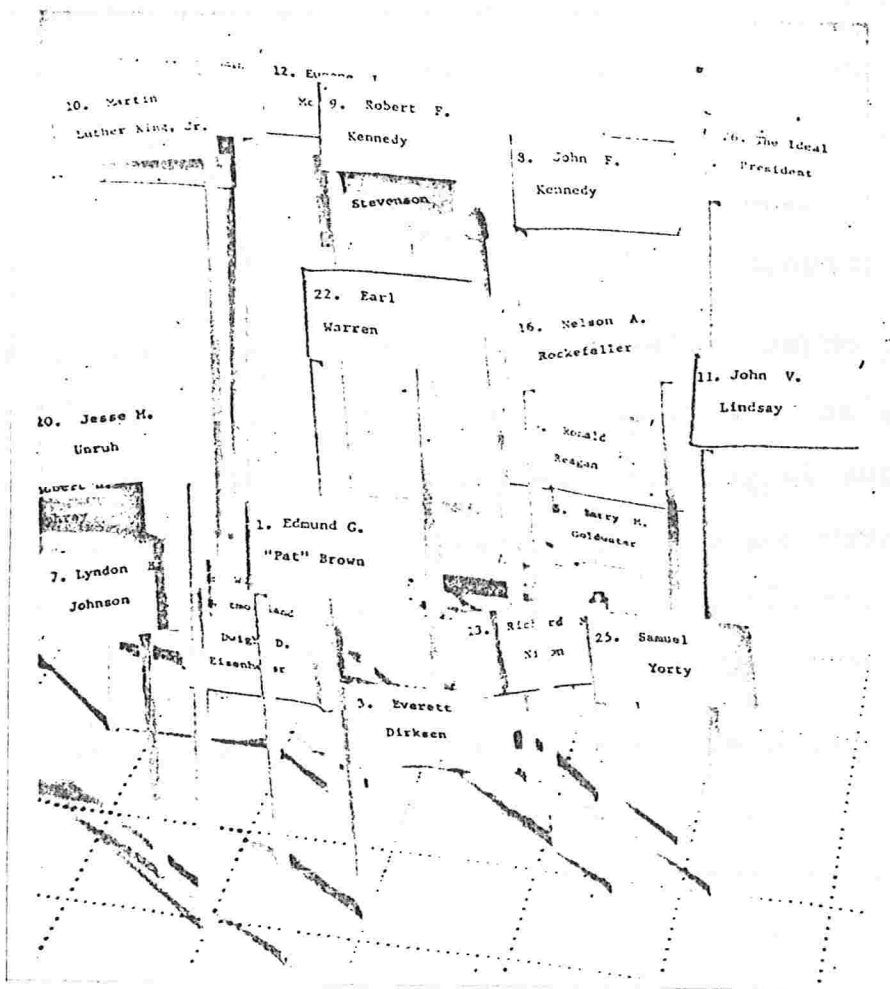


Figura 7.14 - Representação de uma configuração resultante de um EMD em 3 dimensões.

CAPÍTULO 8

APLICAÇÃO

Realizamos a pedido do Prof. Arno Engelmann, um experimento com vários grupos de termos que exprimem relatos verbais de estados subjetivos, estes grupos haviam sido previamente estudados por ele, através de outras técnicas estatísticas. (Engelmann 1978).

O objetivo desse experimento foi a visualização desses grupos de termos, querendo-se verificar se algum termo se distingue do grupo, ou se existe a formação de blocos de termos dentro dos grupos, ou ainda eventualmente interpretar as características que definem as posições dos termos na configuração (interpretação das dimensões do espaço).

Disponhamos inicialmente de 49 grupos de termos, começamos eliminando os grupos que continham 2 ou 3 termos por não haver condições de fazer a análise com tão poucos estí-

mulos. A aplicação do programa ALSCAL acabou mostrando que também deveriam ser eliminados os grupos com 4 termos pois com 4 estímulos fica impossível estimar a configuração (são dadas 6 medidas de proximidade e exigem-se 8 coordenadas num espaço bidimensional). A análise foi feita para 25 grupos com número de termos variável.

Para os grupos com menos que dez estímulos a coleta dos dados foi feita pelo "Método dos Pares" (vide 2.3.2.2) optou-se por esse método pela facilidade de aplicação; o número de pares correspondente a essa quantidade de estímulos não seria muito grande, e a atribuição de uma nota numa escala de similaridade de 1 a 9 não seria difícil nem cansativa para os sujeitos submetidos ao experimento. Além de que a partir dessas notas atribuídas pelos sujeitos a obtenção da matriz de proximidades é extremamente simples.

A pesquisa foi feita sob forma de questionário, à cada par de estímulos o sujeito deveria atribuir uma nota de 1 a 9 numa escala que associava ao 1 muito diferente e ao 9 muito similar. Obtidas essas notas calculou-se a média e a mediana delas para todos os sujeitos, foi construída uma matriz de similaridades com as médias e outra com as medianas; ambas foram processadas pelo ALSCAL. Porém, após o processamento optou-se pelas médias, devido à pouca variabilidade das medianas, usando-se as medianas havia muitas medidas repetidas sendo portanto prejudicial à solução do

E.M.D. .

Quanto aos grupos com mais de dez termos, formar-se-ia um número muito grande de pares, por exemplo, com 11 estímulos obter-se-iam 55 pares, tornando-se portanto extremamente cansativo, para os sujeitos submetidos ao experimento, realizarem todos esses julgamentos. Em vista deste fato a coleta dos dados foi feita pelo "Método das Categorias" (vide 2.3.2.3). Para esses grupos o questionário continha a relação de todos os termos do grupo. A tarefa do sujeito consistia em subdividir esses termos em categorias mutuamente exclusivas, de modo que os estímulos pertencentes a uma mesma categoria fossem semelhantes entre si, sendo que, ele não poderia classificar todos os estímulos numa única categoria. A medida de proximidade foi obtida através da percentagem de vezes que dois estímulos foram classificados, pelos sujeitos, na mesma categoria.

Optou-se pelo E.M.D. não métrico a "2 fatores", uma vez que os dados se encontravam no nível de mensuração ordinal e havia uma única matriz para todos os sujeitos, não sendo portanto levadas em conta as diferenças individuais, para isso foi usado o programa ALSCAL.

Foram submetidos ao experimento alunos de um curso de Psicologia ministrado pelo Prof. Arno Engelmann nos anos de 1979 e 1980 no Instituto de Psicologia Experimental da

Universidade de São Paulo.

Obtiveram-se as configurações desejadas, para alguns grupos em espaços de dimensão um e dois, para outros até em espaços de dimensão três. Além disso foram obtidos os respectivos diagramas de dispersão e os valores do STRESS e do SSTRESS correspondentes a cada solução.

Apresentaremos abaixo apenas dois grupos como exemplo do que foi realizado, um com menos de dez estímulos e outro com mais.

EXEMPLO 8.1: Os estímulos que constituem um dos grupos analisados são: A: Sentir-se bem disposto

B: Sentir confiança	E: Sentir-se deslumbrado
C: Estar contente	F: Estar encantado
D: Sentir uma satisfação	G: Sentir intimidade

A tabela 8.1 apresenta as medidas de proximidade obtidas para esse grupo a partir dos julgamentos de 53 sujeitos, realizados através do Método dos Pares, e usando a média dos julgamentos como medida de similaridade.

Obteve-se a configuração desejada em espaço unidimensional (Figura 8.1) e bidimensional (Figura 8.2).

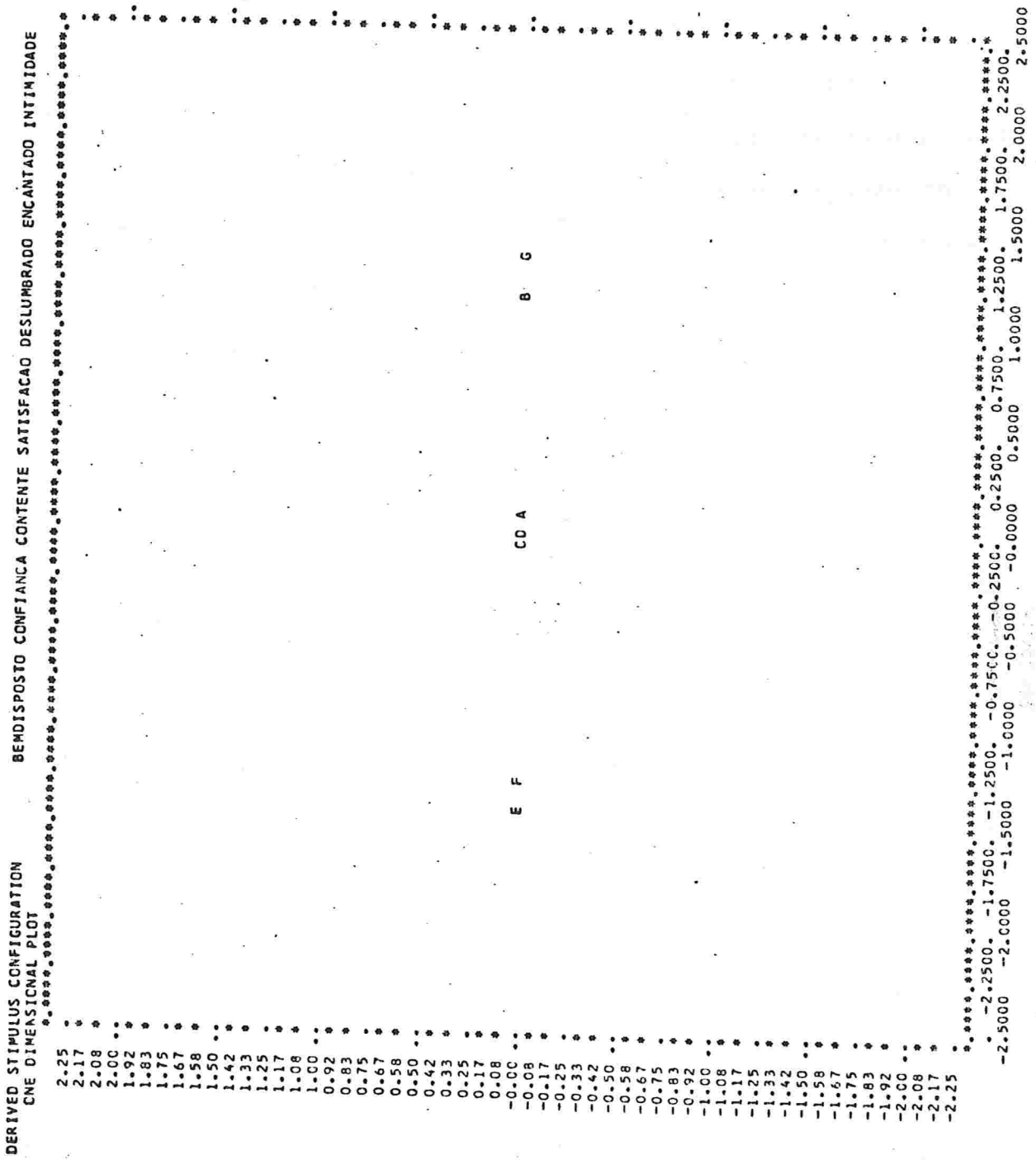


Figura 8.1 - Configuração em espaço unidimensional referente aos dados do exemplo 8.1.

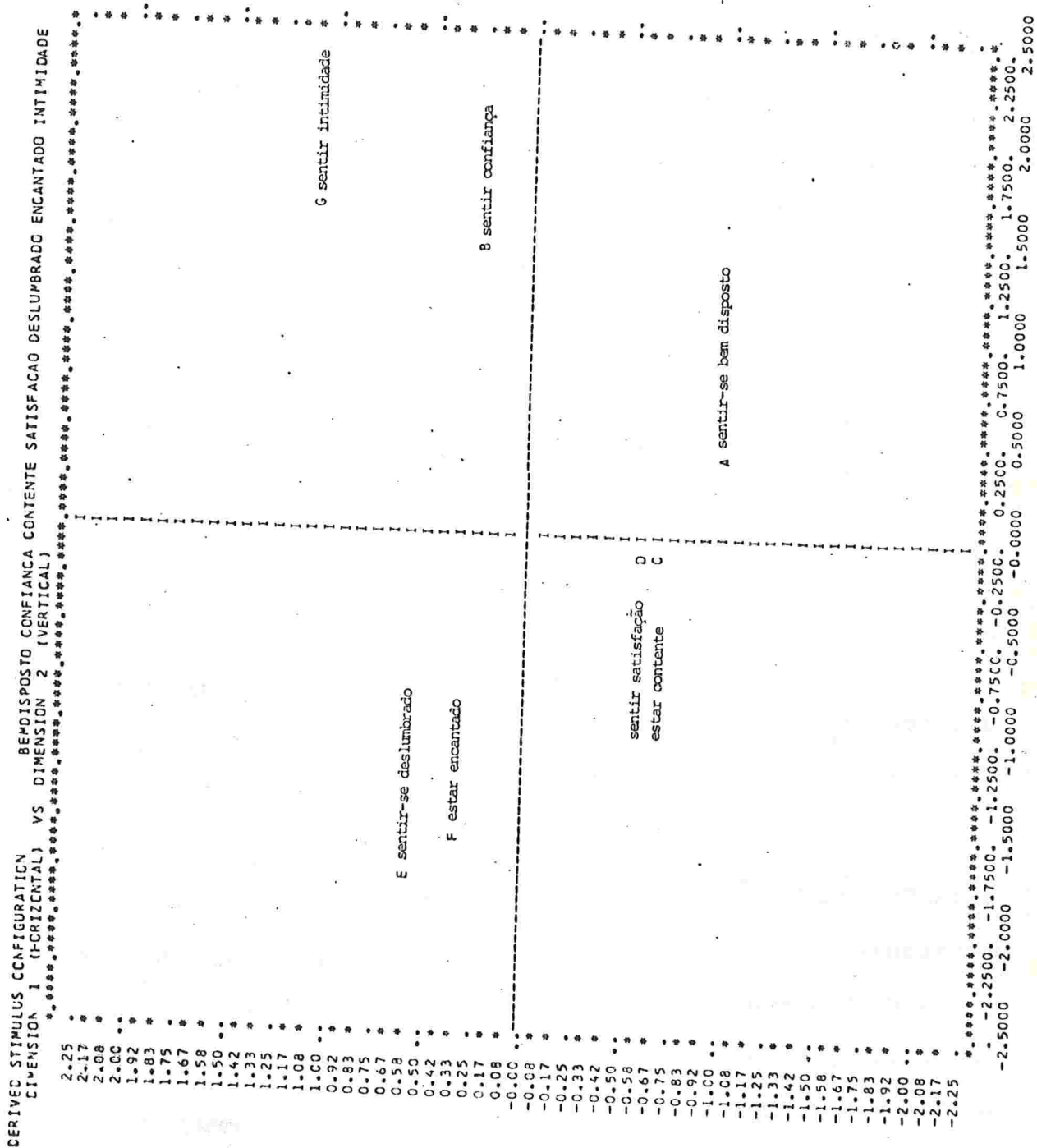


Figura 8.2 - Configuração no espaço bidimensional dos dados referentes ao exemplo 8.1.

Tabela 8.1: Matriz de similaridade referente aos estímulos do exemplo 8.1

	A	B	C	D	E	F	G
A							
B	5,92						
C	7,23	5,36					
D	6,49	5,19	7,45				
E	4,21	3,32	4,87	4,58			
F	4,47	3,37	5,62	5,23	7,81		
G	4,36	6,02	4,28	4,58	2,74	3,13	

Vale a pena observar que a configuração no espaço unidimensional é praticamente a projeção da configuração a duas dimensões sobre o eixo horizontal.

Para a solução no espaço unidimensional obteve-se SSTRESS igual a 0,01503 e STRESS de 0,018. No espaço bidimensional houve uma melhora do SSTRESS, que passou a 0,00540 e o STRESS permanece praticamente igual com valor de 0,020 (o ligeiro acréscimo pode ser explicado pelo fato do programa trabalhar em função do SSTRESS e fornecer apenas o STRESS da solução obtida).

O diagrama de dispersão correspondente à solução no espaço bidimensional (Figura 8.3) apresenta um aspecto que satisfaz bastante a condição de monotonicidade exigida pelo

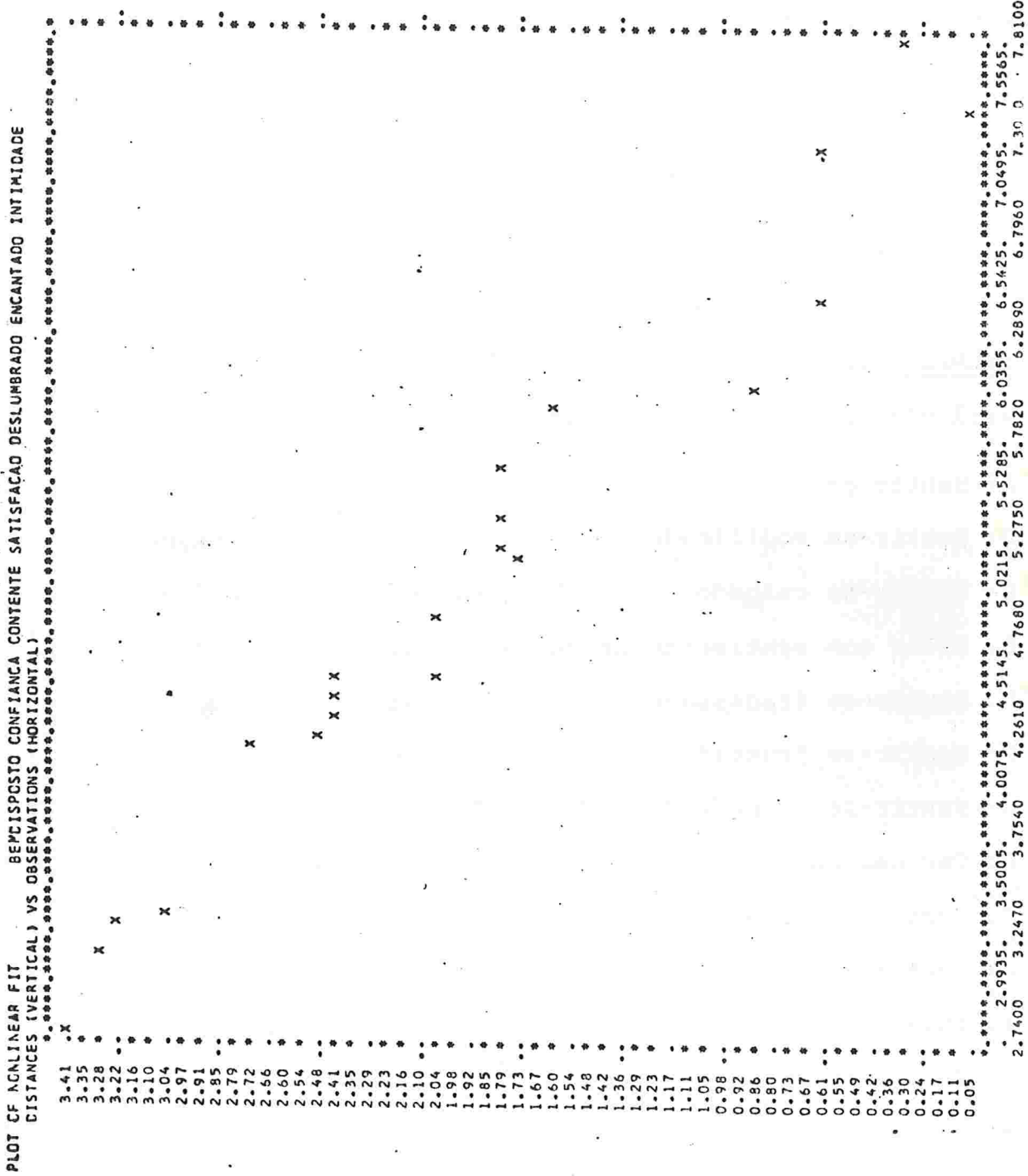


Figura 8.3 - Diagrama de dispersão para a solução no espaço bidimensional referente aos dados do exemplo 8.1.

modelo, e aquele relativo à solução unidimensional (Figura 8.4) apresenta um número maior de patamares.

Para esse grupo não foi possível dar uma interpretação aos eixos, no entanto interessou a visualização da posição relativa dos estímulos.

EXEMPLO 8.2: Neste exemplo vamos observar um grupo com 25 estímulos que são os enumerados abaixo:

A: Sentir angustia	N: Sentir uma decepção
B: Sentir-se modificado	O: Sentir-se desenganado
C: Sentir-se culpado	P: Sentir-se desapontado
D: Estar com sentimento de culpa	Q: Estar aborrecido
E: Sentir-se fracassado	R: Estar fulminado
F: Sentir-se frustrado	S: Sentir-se mal
G: Sentir-se amargurado	T: Sentir vergonha
H: Ter uma magoa	U: Ficar sem graça
I: Sentir-se inconsolável	V: Sentir-se desconcertado
J: Sentir-se lastimando algo	W: Sentir-se tímido
K: Estar triste	X: Sentir-se acanhado
L: Sentir pesar	Y: Sentir-se constrangido
M: Sentir-se desolado	

A tabela 8.2 apresenta as medidas de proximidade obtidas através dos julgamentos que 47 sujeitos efetuaram pelo "Método das Categorias"

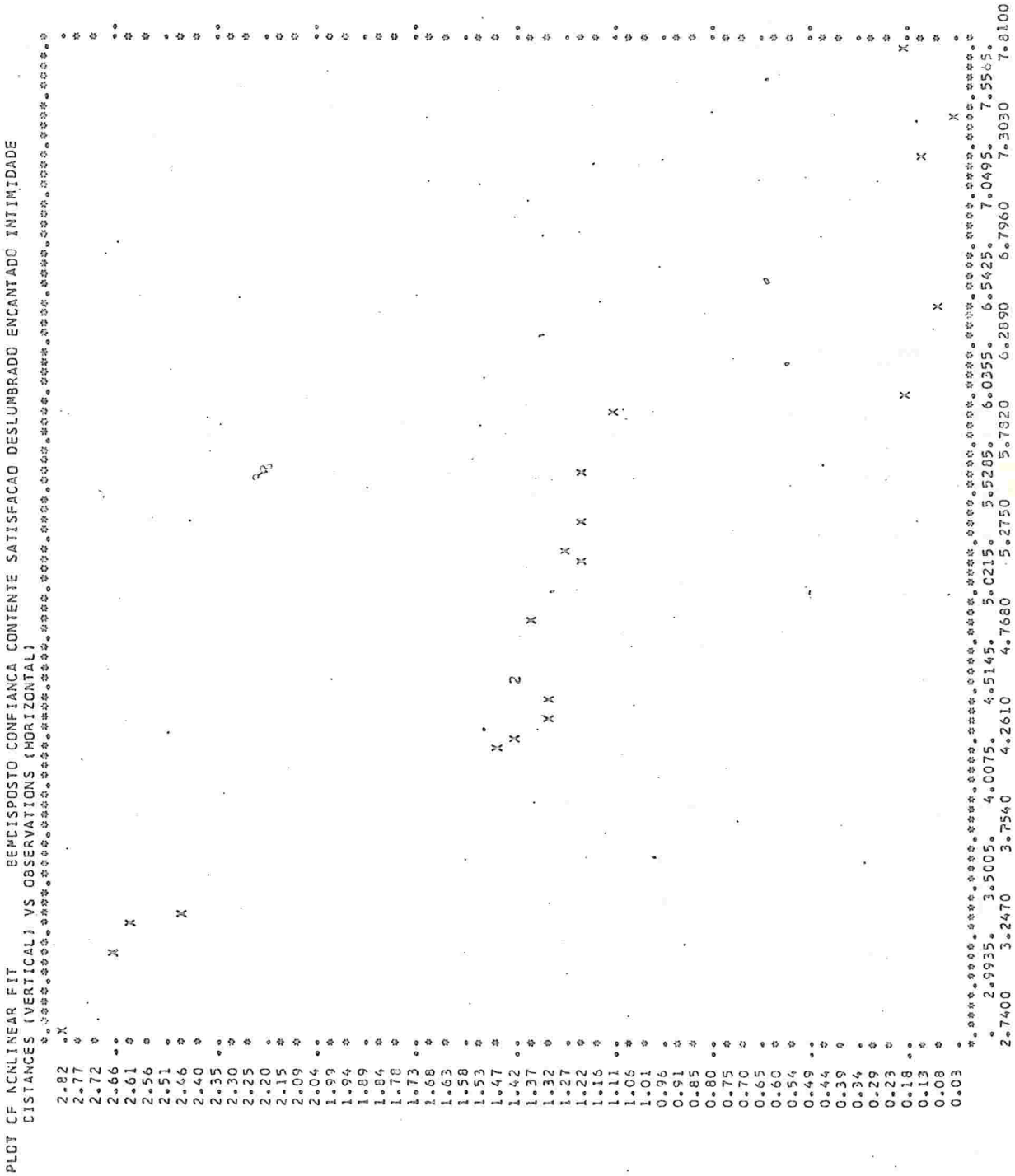


Figura 8.4 - Diagrama de dispersão para a solução no espaço unidimensional referente aos dados do exemplo 8.1.

Tabela 8.2 - Matriz de similaridades entre os estímulos do exemplo 8.2

	A 1	B 2	C 3	D 4	E 5	F 6	G 7	H 8	I 9	J 10
1A	0.0									
2B	21.280	0.0								
3C	10.640	19.150	0.0							
4D	12.760	19.150	93.620	0.0						
5E	12.760	19.150	6.380	6.380	0.0					
6F	14.890	17.020	4.250	4.250	68.080	0.0				
7G	44.680	17.020	12.760	14.850	17.020	8.510	0.0			
8H	8.510	8.510	4.250	4.250	6.380	4.250	36.170	0.0		
9I	12.760	29.790	4.250	4.250	12.760	4.250	27.660	27.660	0.0	
10J	6.380	8.510	2.130	2.130	6.380	10.640	12.760	21.280	38.300	0.0
11K	10.640	10.640	0.010	0.010	6.380	6.380	25.530	51.060	27.660	25.530
12L	12.760	14.890	17.020	19.150	6.380	2.130	23.400	31.910	27.660	23.400
13M	12.760	29.790	6.380	6.380	27.660	10.640	23.400	23.400	59.570	23.400
14N	8.510	2.130	4.250	6.380	21.280	31.510	10.640	14.850	10.640	10.640
15O	6.380	10.640	6.380	6.380	25.530	14.890	14.890	14.890	21.280	8.510
16P	8.510	2.130	6.380	6.380	21.280	25.530	14.890	14.890	10.640	10.640
17Q	10.640	4.250	4.250	4.250	10.640	14.890	19.150	27.660	14.890	19.140
18R	4.250	31.910	4.250	4.250	10.640	4.250	10.640	4.250	17.020	4.250
19S	21.280	8.510	10.640	12.760	8.510	10.640	12.760	4.250	4.250	4.250
20T	0.010	4.250	8.510	8.510	4.250	2.130	2.130	2.130	2.130	0.010
21U	2.130	4.250	6.380	6.380	4.250	2.130	2.130	2.130	2.130	0.010
22V	4.250	6.380	4.250	2.130	8.510	6.380	6.380	6.380	10.640	4.250
23W	2.130	4.250	0.010	0.010	2.130	4.250	0.010	0.010	0.010	0.010
24X	0.010	2.130	2.130	2.130	4.250	2.130	2.130	4.250	2.130	0.010
25Y	2.130	6.380	8.510	8.510	4.250	0.010	4.250	4.250	4.250	4.250
	K 11	L 12	M 13	N 14	O 15	P 16	Q 17	R 18	S 19	T 20
11K	0.0									
12L	53.190	0.0								
13M	42.550	40.420	0.0							
14N	8.510	10.640	12.760	0.0						
15O	19.140	12.760	29.790	36.170	0.0					
16P	17.020	19.150	17.020	80.850	36.170	0.0				
17Q	34.040	23.400	23.400	17.020	10.640	27.660	0.0			
18R	8.510	10.640	17.020	4.250	19.150	4.250	17.020	0.0		
19S	10.640	8.510	4.250	6.380	8.510	4.250	8.510	8.510	0.0	
20T	0.010	6.380	2.130	2.130	2.130	4.250	2.130	2.130	19.150	0.0
21U	0.010	4.250	2.130	2.130	2.130	2.130	6.380	2.130	10.150	76.590
22V	4.250	4.250	12.760	12.760	6.380	10.640	10.640	10.640	14.890	57.450
23W	2.130	0.010	0.010	0.010	0.010	0.010	2.130	0.010	17.020	48.940
	K 11	L 12	M 13	N 14	O 15	P 16	Q 17	R 18	S 19	T 20
24X	2.130	2.130	2.130	2.130	2.130	2.130	4.250	2.130	17.020	59.570
25Y	4.250	10.640	8.510	2.130	2.130	2.130	6.380	2.130	14.890	55.320
	U 21	V 22	W 23	X 24	Y 25					
21U	0.0									
22V	55.320	0.0								
23W	53.190	38.300	0.0							
24X	55.570	46.810	85.110	0.0						
25Y	53.190	48.940	55.320	65.960	0.0					

O grande número de estímulos, a configuração inicialmente obtida num espaço bidimensional, onde se verifica a formação de alguns blocos (Figura 8.5), e a observação do diagrama de dispersão dessa solução (Figura 8.6) que não respeita a monotonicidade exigida, sugeriram a realização do E.M.D. em espaço de dimensão mais alta.

Os valores do SSTRESS e do STRESS obtidos para essas soluções encontram-se na tabela 8.3.

DIM. DO ESPAÇO	1	2	3	4
SSTRESS	0,28198	0,21216	0,14222	0,09997
STRESS	0,298	0,188	0,111	0,080

TABELA 8.3

A partir da tabela 8.3 construiu-se um gráfico, (Figura 8.7), analisando-se este gráfico, observa-se que os valores do SSTRESS e do STRESS decrescem rapidamente até a dimensão três, apresentando uma ligeira tendência para se estabilizar a partir dessa dimensão. Em vista dessa situação, o espaço tridimensional poderia ser escolhido para a análise dos resultados, uma vez que em termos de visualização da configuração ele ainda é viável.

Observando-se os três planos de projeção da confi

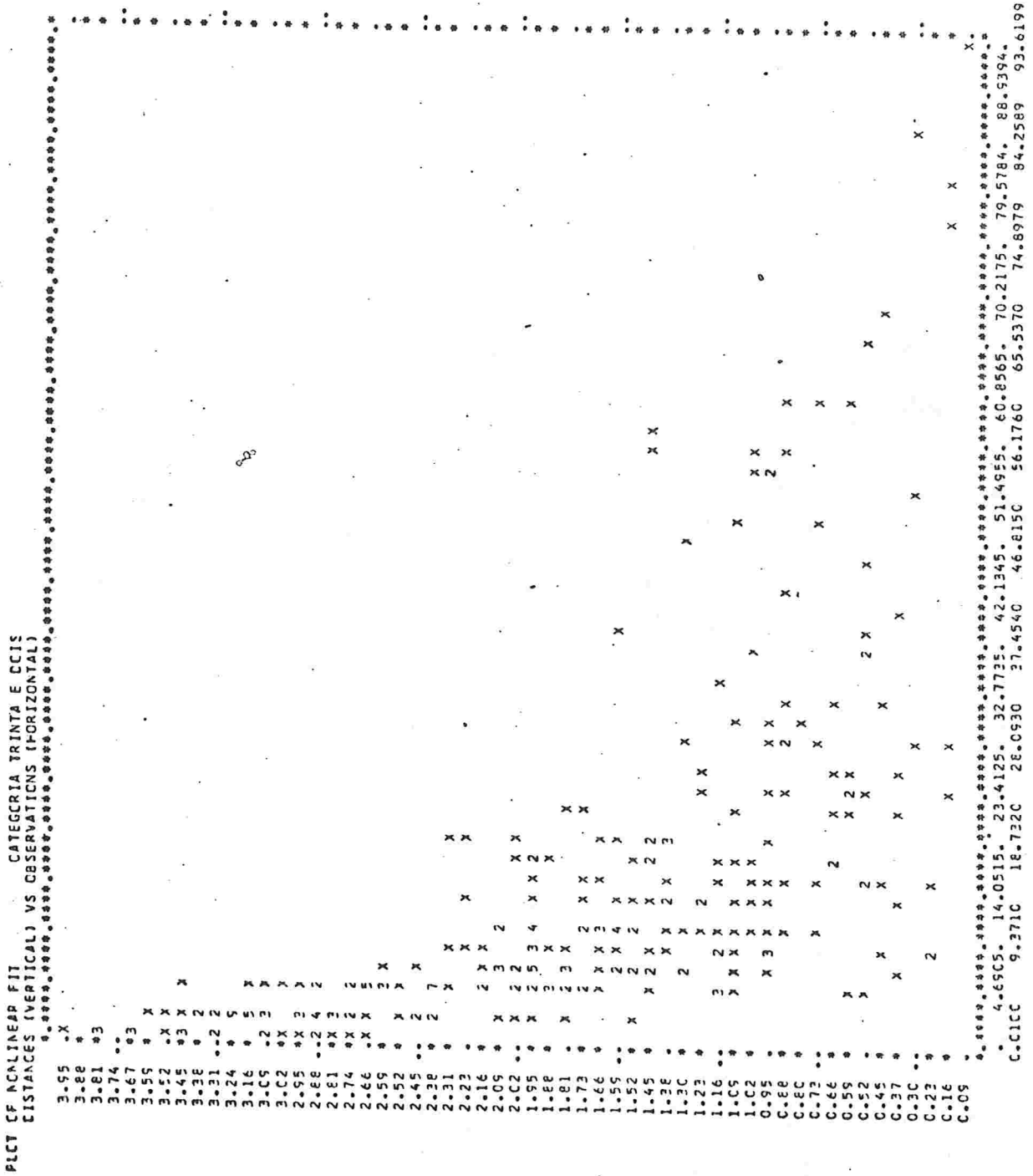


Figura 8.6 - Diagrama de dispersão da solução num espaço bi dimensional para os dados do exemplo 8.2.

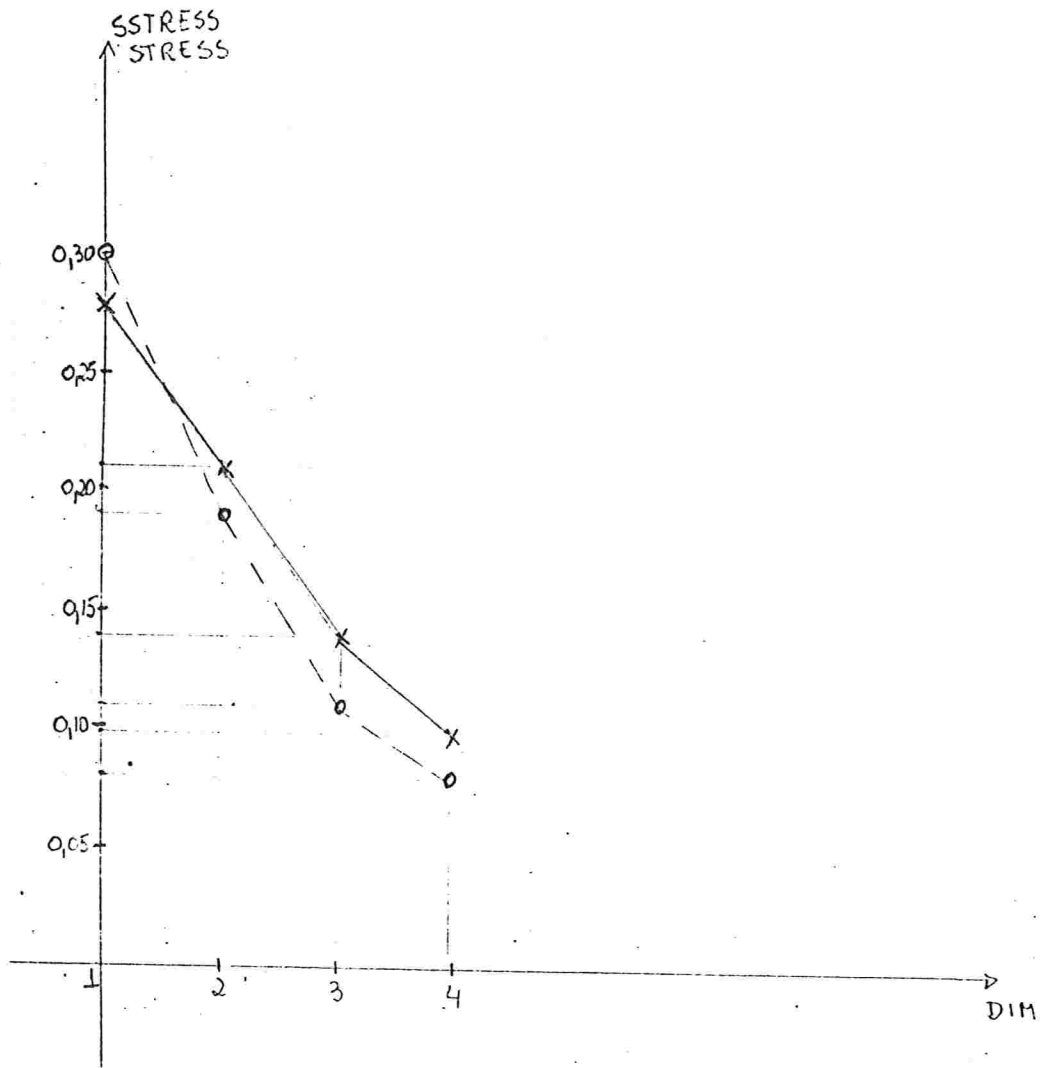


Figura 8.7 - Medidas de ajuste em função da dimensão do espaço para os dados do exemplo 8.2

X : SSTRESS

0 : STRESS

Observando-se os três planos de projeção da configuração no espaço tridimensional figuras 8.8, 8.9, 8.10, verifica-se que dos cinco blocos observados no espaço bidimensional apenas um deles formado pelos estímulos T e U está mantido.

A configuração no espaço bidimensional se assemelha à projeção da configuração tridimensional sobre o plano de dimensões 1 e 3 (Figura 8.9).

A interpretação da solução tridimensional merece um estudo mais detalhado.

O apêndice G apresenta a saída do computador relativa ao exemplo 8.2.

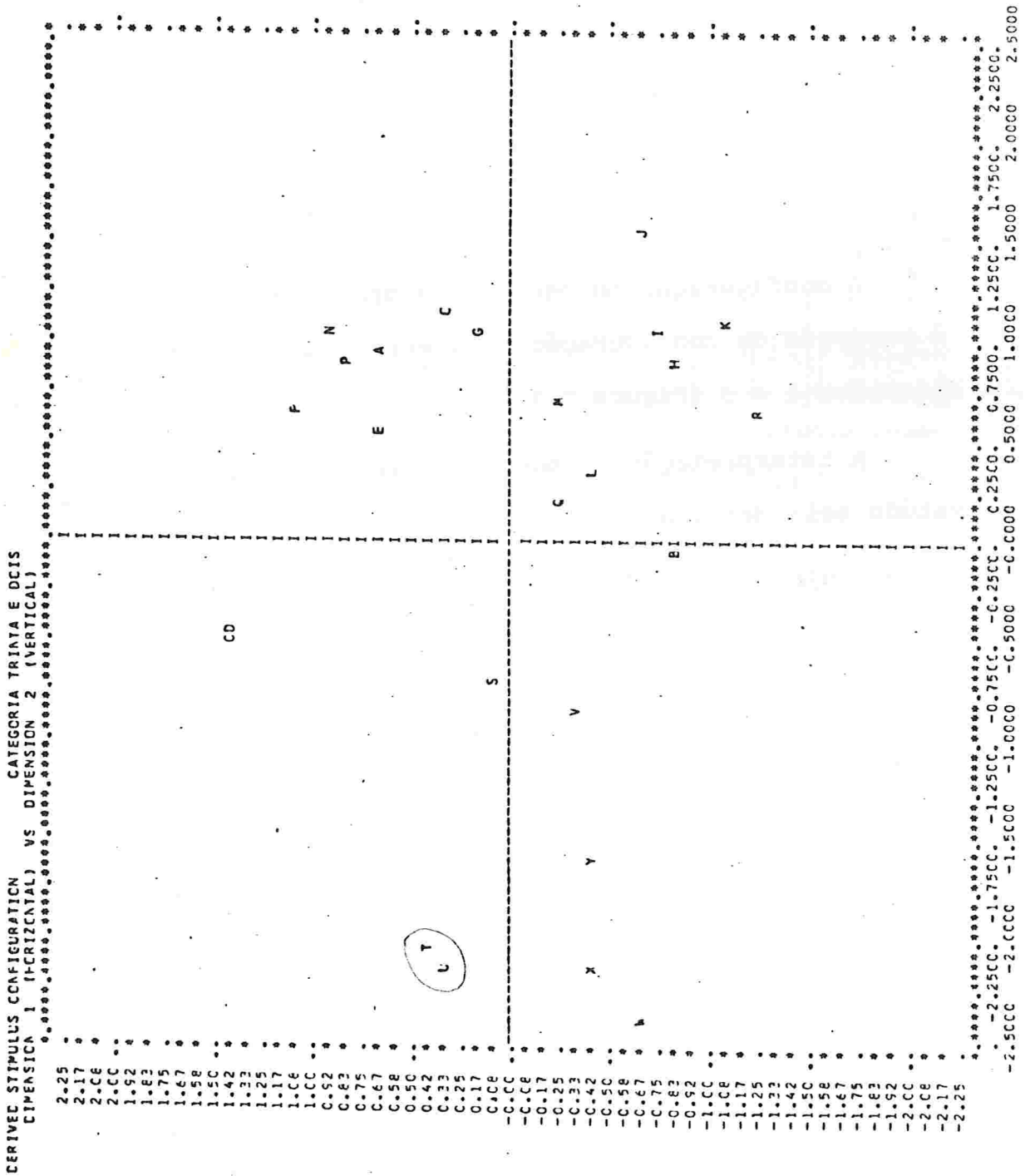


Figura 8.8 - Projeção sobre o plano de dim 1 e dim 2 da configuração tridimensional referente ao exemplo 8.2.

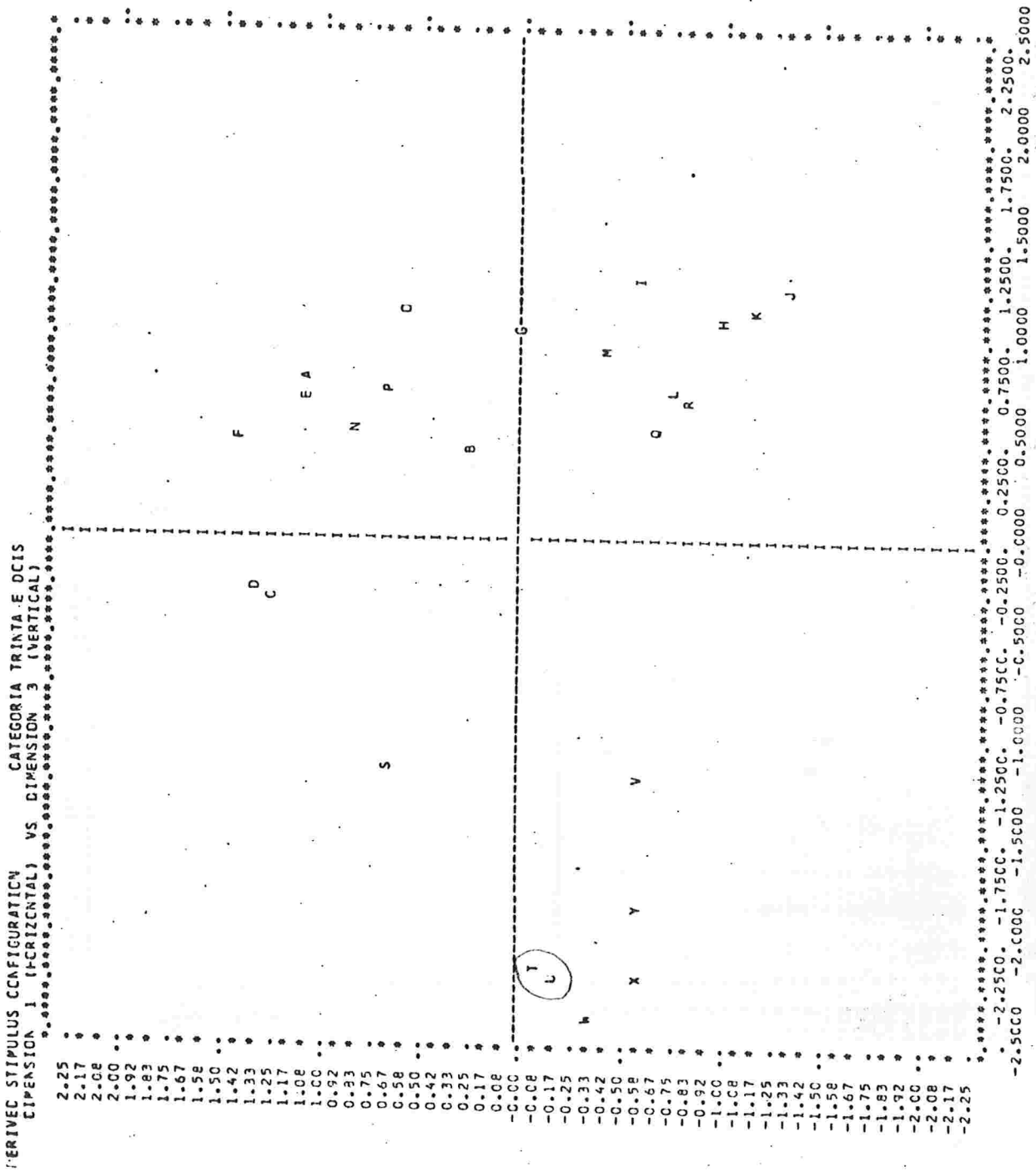


Figura 8.9 = Projeção sobre o plano de dim 1 e dim 3 da configuração tridimensional referente ao exemplo 8.2.

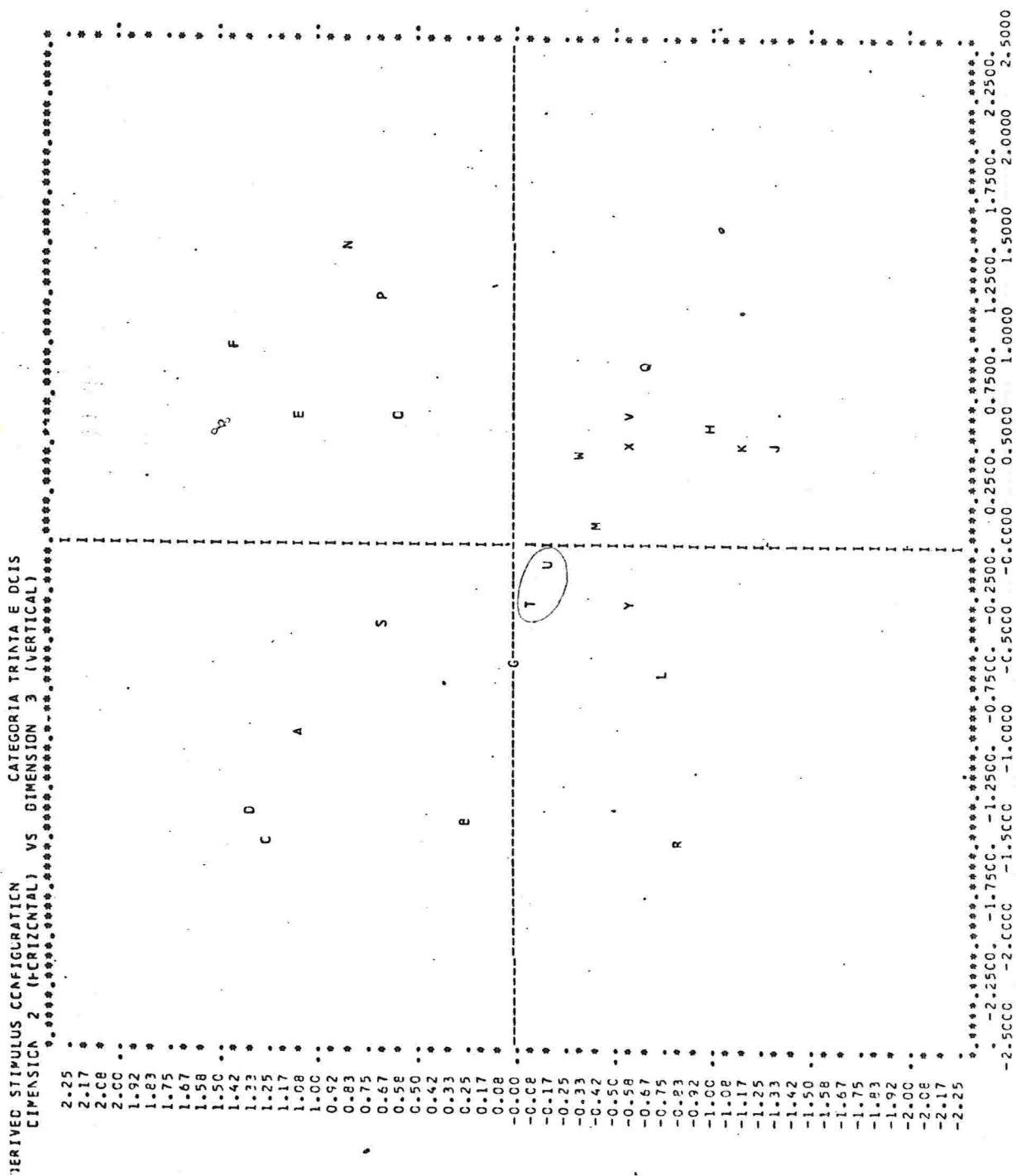


Figura 8.10 - Projeção sobre o plano de dim 2 e dim 3 da configuração tridimensional referente ao exemplo 8.2.

CAPÍTULO 9

COMENTÁRIOS

Neste capítulo pretendemos comentar brevemente alguns aspectos do E.M.D. que não tenham sido abordados nos capítulos anteriores.

9.1 - DISTÂNCIAS NÃO EUCLIDIANAS

Torgerson (1962), Kruskal (1964, a,b), Carroll e Wish (1974) e muitos outros autores comentam a adoção de distâncias não Euclidianas nos modelos de E.M.D., porém a tendência geral, atualmente é a adoção da distância Euclidiana.

Uma classe de distâncias estudadas pelos diversos autores são as r-métricas de Minkowski definidas como

$$d_{ij}^{(r)} = \left(\sum_{t=1}^m |x_{it} - x_{jt}|^r \right)^{1/r}$$

onde para $r = 2$ tem-se a própria distância Euclidiana e para $r = 1$ tem-se a distância "city block" ou "métrica de Manhatttan".

Os programas MDSCAL (Kruskal e Carmone, 1970) e o TORSCA (Young e Torgerson, 1967) estão habilitados a trabalhar com as "r-métricas de Minkowski". Kruskal (1964 a) comenta que essas métricas diferem das distâncias Euclidianas no que diz respeito à rotação. Uma rotação rígida mantém a distância Euclidiana inalterada, enquanto que as únicas rotações rígidas que mantêm as $d_{ij}^{(r)}$ inalteradas são aquelas que transformam eixos de coordenadas em eixos de

Takane e Young (1977) afirmam que para adotar uma r-métrica de Minkowski que não seja Euclidiana devem definir uma nova medida de ajuste o r-STRESS definido por:

$$\sigma^2(X, W, D^*) = \sum_k \sum_i \sum_j^{i-1} (d_{kij}^{*(r)} - d_{kij}^{(r)})^2$$

onde:
$$d_{kij}^{(r)} = \sum_{\alpha=1}^t W_{k\alpha} |x_{i\alpha} - x_{j\alpha}|^r$$

Observam no entanto que não há interesse em adaptar o ALSCAL para essas métricas, principalmente para valores altos de r . Haveria interesse para $r = 1$ que é quando o r-STRESS é idêntico ao STRESS e além do mais a distân

cia "city block" é a distância não Euclidiana mais comumente usada em aplicações do E.M.D. nas ciências sociais.

9.2 - DADOS REPETIDOS

Kruskal (1964 a) apresenta dois procedimentos possíveis quando os dados apresentarem medidas de proximidade repetidas isto é, quando $\delta_{ij} = \delta_{mn}$. Um deles chamado primário não restringe d_{ij}^* e d_{mn}^* , não exige $d_{ij} = d_{mn}$, ou seja, mantém sobre d_{ij}^* apenas a restrição de monotonicidade, ou seja:

$$\delta_{ij} < \delta_{kl} \Rightarrow d_{ij}^* \leq d_{mn}^*$$

O outro, o secundário considera que se $\delta_{ij} = \delta_{mn}$ é evidente que d_{ij} deva ser igual a d_{mn} , para isto impõe $d_{ij}^* = d_{mn}^*$, portanto no procedimento secundário as restrições sobre os d_{ij}^* são:

$$\left\{ \begin{array}{l} \delta_{ij} < \delta_{mn} \Rightarrow d_{ij}^* \leq d_{mn}^* \\ \delta_{ij} = \delta_{mn} \Rightarrow d_{ij}^* = d_{mn}^* \end{array} \right.$$

Na técnica ALSCAL considera-se a transformação t^{co} , contínua ordinal ao se escolher o procedimento primário de Kruskal e a transformação t^{do} , discreta ordinal para o secun

dário.

9.3 - MEDIDAS DE AJUSTE

Kruskal e Carroll (1969) discutem em seu artigo me-
didas de ajuste alternativas para o STRESS. A única delas
que mereceu atenção posterior dos autores é o STRESS fórmu-
la 2. (O STRESS adotado por nós é o também chamado STRESS
fórmula 1) dado por:

$$\phi = \sqrt{\frac{\sum_i \sum_j (d_{ij} - d_{ij}^*)^2}{\sum_i \sum_j (d_{ij} - \bar{d})^2}}$$

onde \bar{d} : média aritmética dos d_{ij} .

Diferindo este do STRESS fórmula 1 apenas pelo fa-
cto de escala do denominador.

Kruskal (1978) afirma que a minimização da última
expressão produz uma solução muito semelhante àquela obti-
da minimizando o STRESS fórmula 1

Young (1970) definiu um índice M que mede o grau
de determinação métrica aferida pelo E.M.D. não métrico se-
ria a correlação quadrática entre as verdadeiras distâncias
e as distâncias obtidas através do E.M.D., estudou o compor-
tamento desse índice com dados simulados, interpretou-o co-

mo uma medida de confiança e não como uma medida apenas descritiva como é o STRESS, porém foi impossível medir M em experimentos reais, tão pouco conseguiu uma função que esti-
masse M.

9.4 - CONSIDERAÇÕES FINAIS

A eficiência e a robustez dos diversos métodos de E.M.D. foi comprovada através de vários estudos usando técnicas de Monte Carlo (Takane e Young (1977), Mac Callum e Cornelius (1977) Spence e Young (1978) ou com os dados sintéticos (Kruskal (1964 a)) ou ainda usando dados dos quais se conhecia de antemão a configuração desejada (Torgerson (1962)), observando-se então que as técnicas de E.M.D. realmente produzem configurações fiéis às medidas de proximidade das quais se parte .

Este trabalho pretende apenas apresentar o E.M.D., na sua versão mais simples, para que sirva de guia introdutório àqueles que pretendem aplicá-lo. Não foi nosso objetivo apresentar todas as abordagens dadas ao E.M.D., no entanto algumas outras devem ser citadas para estudos posteriores como a de Guttman (1968) e aquela apresentada em Fernandez (1980).

A própria técnica ALSCAL pode ser aprofundada visto que existem vários artigos discutindo, aperfeiçoando e a

plicando esta técnica em outras áreas da Estatística (de Leenw, Young e Takane (1976), Takane, Young e de Leeuw (1979)).

Outras alternativas do programa ALSCAL que devem ser estudadas são aquelas que lidam com matrizes retangulares, ou ainda, as que lidam com réplicas no experimento.

A tendência atual é explorar a aplicação do E.M.D. a "3 fatores". Além disto, seria interessante em um trabalho futuro procurar medidas de estabilidade das soluções obtidas através do E.M.D.

APÊNDICE A - DEMONSTRAÇÕES DOS TEOREMAS DO CAPÍTULO 3

APÊNDICE A - DEMONSTRAÇÕES DOS TEOREMAS DO CAPÍTULO 3

A.1 - Demonstração do teorema 3.1

$$(a) \quad d_{rs}^2 = -2 a_{rs} = (z_r - z_s)' (z_r - z_s) \quad (1)$$

Pode-se escrever:

$$B = H A H = A - n^{-1} A J - n^{-1} J A + n^{-2} J A J$$

onde $J = 1 \cdot 1'$

Tem-se:

$$\frac{1}{n} A \cdot J = \begin{bmatrix} \bar{a}_{1.} & \dots & \bar{a}_{1.} \\ \vdots & & \\ \bar{a}_{n.} & \dots & \bar{a}_{n.} \end{bmatrix}$$

$$\frac{1}{n} J A = \begin{bmatrix} \bar{a}_{.1} & \dots & \bar{a}_{.n} \\ \vdots & & \\ \bar{a}_{.1} & \dots & \bar{a}_{.n} \end{bmatrix}$$

$$\frac{1}{n^2} J A J = \begin{bmatrix} \bar{a}_{..} & \dots & \bar{a}_{..} \\ \vdots & & \\ \bar{a}_{..} & \dots & \bar{a}_{..} \end{bmatrix}$$

onde: $\bar{a}_{r.} = \frac{1}{n} \sum_{s=1}^n a_{rs}$ (média da r-ésima linha)

$$\bar{a}_{.s} = \frac{1}{n} \sum_{r=1}^n a_{rs} \quad (\text{média da s-ésima coluna}) \quad (2)$$

$$\bar{a}_{..} = \frac{1}{n^2} \sum_{r=1}^n \sum_{s=1}^n a_{rs} \quad (\text{média global})$$

$$\text{Portanto } b_{rs} = a_{rs} - \bar{a}_{r.} - \bar{a}_{.s} + \bar{a}_{..} \quad (3)$$

Substituindo-se a_{rs} pela relação (1) e usando as relações (2) obtém-se:

$b_{rs} = (z_r - \bar{z})' \cdot (z_s - \bar{z})$ na forma matricial $B = (HZ) \cdot (HZ)'$ portanto $B \geq 0$ estando demonstrada assim a parte (a) do teorema.

(b) Supor $B \geq 0$ e considerar a configuração dada no teorema.

Seja $\Lambda = \text{diag} (\lambda_1, \dots, \lambda_p)$ e seja $\Gamma = X \Lambda^{-1/2}$, assim sendo as colunas de Γ , $\gamma_{(i)} = \lambda_i^{-1/2} x_{(i)}$ são os vetores característicos de B padronizados.

Pela decomposição espectral da matriz B tem-se: $B = \Gamma \Lambda \Gamma' = X \cdot X'$ isto é $b_{rs} = x'_r x_s$, portanto B representa a matriz do produto interno desta configuração.

Deve-se mostrar agora que D representa a matriz das distâncias entre os pontos desta configuração.

Usando a relação (3) para escrever B em função de

A chega-se \tilde{a} :

$$\begin{aligned}
 (x_r - x_s)' (x_r - x_s) &= x_r' x_r - 2 x_r' x_s + x_s' x_s = b_{rr} - 2 b_{rs} \\
 &+ b_{ss} \\
 &= a_{rr} - \bar{a}_{r.} - \bar{a}_{.r} + \bar{a}_{..} - 2 \cdot (a_{rs} - \bar{a}_{r.} - \bar{a}_{.s} + \bar{a}_{..}) + a_{ss} \\
 &- \bar{a}_{s.} - \bar{a}_{.s} + \bar{a}_{..} \\
 &= a_{rr} - 2a_{rs} + a_{ss} \\
 &= - 2a_{rs} \quad \text{pois } a_{rr} = - \frac{1}{2} d_{rr}^2 = 0 \\
 &= d_{rs}^2 \quad \text{e } a_{rs} = - \frac{1}{2} d_{rs}^2
 \end{aligned}$$

Finalizando nota-se que $B:1 = H.A.H.1 = 0$, portanto 1 é um vetor característico de B correspondente à raiz característica 0 .

Portanto 1 é ortogonal às colunas de $X, x'(i) \cdot 1 = 0$, $i = 1, \dots, p$.

Logo $n \cdot \bar{x} = \sum_{r=1}^n x_r = X' \cdot 1 = (x'(i) \cdot 1, \dots, x'(p) \cdot 1) = 0$, portanto o centro de gravidade desta configuração localiza-se na origem.

A.2 - Demonstração do teorema 3.2

(a) Primeiramente nota-se que se $\Delta \geq 0$

$$d_{rs}^2 = \delta_{rr} - 2\delta_{rr} + \delta_{ss} = x' \Delta x \geq 0$$

onde x é um vetor $(x_{ij})_{n \times 1}$ tal que

$$x_{ij} = \begin{cases} 1 & \text{se } i = r \\ -1 & \text{se } i = s \\ 0 & \text{se } i \neq r \text{ e } i \neq s \end{cases} \quad \text{para } r \neq s$$

Portanto a transformação padrão está bem definida e D é uma matriz de distâncias.

(b) Sendo $A = (a_{rs})$, $a_{rs} = -\frac{1}{2} d_{rs}^2$ e $B = H A H$

Uma vez que $H \Delta H$ é também positiva semi definida basta provar que $B = H \Delta H$ para concluir que D é Euclidiana com matriz de produtos internos $H \Delta H$.

Partindo de $B = H A H$, usando a igualdade $b_{rs} = a_{rs} - \bar{a}_{r.} - \bar{a}_{.s} + \bar{a}_{..}$, sendo $a_{rs} = -\frac{1}{2} d_{rs}^2$ w $d_{rs}^2 = \delta_{rr} - 2\delta_{rs} + \delta_{ss}$. Tem-se:

$$-2b_{rs} = d_{rs}^2 - \frac{1}{n} \sum_{i=1}^n d_{ri}^2 - \frac{1}{n} \sum_{j=1}^n d_{js}^2 + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n d_{ij}^2$$

$$-2b_{rs} = \delta_{rr} - 2\delta_{rs} + \delta_{ss} - \frac{1}{n} \sum_{i=1}^n (\delta_{rr} - 2\delta_{ri} + \delta_{ii})$$

$$- \frac{1}{n} \sum_{j=1}^n (\delta_{jj} - 2\delta_{js} + \delta_{ss}) + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (\delta_{ii} - 2\delta_{ij} + \delta_{jj})$$

$$- 2b_{rs} = - 2c_{rs} + 2\bar{c}_{r.} + 2\bar{c}_{.s} - 2\bar{c}_{..}$$

Portanto

$$b_{rs} = c_{rs} - \bar{c}_{r.} - \bar{c}_{.s} + \bar{c}_{..}$$

ou então na forma matricial tem-se $B = H \Delta H$.

A.3 - Demonstração do teorema 3.3

Usando as relações (5) e (6) do capítulo 3 verifica-se que:

$$\theta = \sum_{r,s=1}^n \sum_{ii=K+1}^p (x'_r \ell_{(i)} - x'_s \ell_{(i)})^2$$

$$\theta = \text{tr } L_2' \left\{ \sum_{r,s=1}^n (x_r - x_s) (x_r - x_s)' \right\} L_2$$

$$\theta = 2n^2 \text{tr } L_2' S L_2$$

onde
$$S = \frac{1}{n} \sum_{r=1}^n (x_r - \bar{x}) \cdot (x_r - \bar{x})'$$

Sendo que:

$$\sum_{r,s=1}^n (x_r - x_s) (x_r - x_s)' = 2n \sum_{r=1}^n (x_r - \bar{x}) \cdot (x_r - \bar{x})' -$$

$$- 2 \sum_{r=1}^n (x_r - \bar{x}) \sum_{s=1}^n (x_s - \bar{x}) = 2n \cdot nS - 0 = 2n^2 S$$

Sejam $\lambda_1 \geq \dots \geq \lambda_p$ as raízes características de $n \cdot S$ com os respectivos vetores característicos normalizados $\Gamma = (\gamma_{(1)}, \dots, \gamma_{(p)})$ pode-se escrever:

$$\theta = 2n^2 \operatorname{tr} L_2' S L_2$$

onde $n S = \Gamma \Lambda \Gamma'$ então

$$\theta = 2n \operatorname{tr} L_2' \Gamma \Lambda \Gamma' L_2$$

onde $L_2' \Gamma = F_2'$, matriz coluna ortonormal

Escreve-se:

$$\theta = 2n \operatorname{tr} F_2' \Lambda F_2$$

$\theta = 2n \operatorname{tr} F_2' \Lambda F_2$ é minimizado quando $F_2 = (0, I_{p-k})'$

Para $F_2 = (0, I_{p-k})' = \Gamma' L_2$ tem-se

$$L_2 = (\gamma_{(k+1)}, \dots, \gamma_{(p)})$$

Portanto as colunas de L_1 geram o espaço dos k primeiros vetores característicos de $n \cdot S$ e assim sendo XL_1 representa as principais coordenadas de X em k dimensões.

Nota-se que para essa projeção nas coordenadas principais

$$\theta = 2n (\lambda_{k+1}, \dots, \lambda_p)$$

A.4 - Demonstração do teorema 3.4

Sejam $\lambda_1 \geq \dots \geq \lambda_n$ as raízes características de B , algumas delas podendo ser negativas e sendo Γ a matriz de vetores característicos normalizados. Vamos considerar aqui por simplicidade $\lambda_i > 0$ (a situação $\lambda_i < 0$ pode ser vista em Mardia 1979).

Sejam $\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_n \geq 0$ as raízes características de \hat{B} . Pelo teorema da decomposição espectral das matrizes, pode-se escrever a matriz simétrica $\Gamma' \hat{B} \Gamma$ como:

$\Gamma' \hat{B} \Gamma = G \hat{\Lambda} G'$ onde G é ortogonal. Então:

$$\Psi = \text{tr} (B - \hat{B})^2 = \text{tr} \Gamma' (B - \hat{B}) \Gamma \Gamma' (B - \hat{B}) \Gamma \rightarrow$$

$$\Psi = \text{tr} [\Gamma' B \Gamma - \Gamma' \hat{B} \Gamma] [\Gamma' B \Gamma - \Gamma' \hat{B} \Gamma]$$

$$\Psi = \text{tr} [\Lambda - G \hat{\Lambda} G'] [\Lambda - G \hat{\Lambda} G']$$

Prova-se que para $\hat{\Lambda}$ fixo, Ψ é minimizado quando $G = I$, de modo que:

$$\Psi = \sum_{i=1}^n (\lambda_i - \hat{\lambda}_i)^2$$

Estando \hat{X} em R^k , $B = H \hat{X} \hat{X}'$ H deve ter no máximo k raízes características não nulas que devem ser não negativas. Assim é fácil perceber que Ψ é minimizado quando:

$$\hat{\lambda}_i = \begin{cases} \lambda_i, & i = 1, \dots, k. \\ 0, & i = k + 1, \dots, n \end{cases}$$

Assim $\hat{B} = \Gamma_1 \Lambda_1 \Gamma_1'$ onde $\Gamma_1 = (\gamma_{(1)}, \dots, \gamma_{(k)})$ e $\Lambda_1 = \text{diag} (\lambda_1, \dots, \lambda_k)$ de modo que \hat{X} pode ser tomado igual a $\Gamma_1 \Lambda_1^{1/2}$, que é a solução clássica do problema do E. M.D. em k dimensões.

Observa-se que o valor mínimo de Ψ é dada por:

$$\psi = \lambda_{k+1}^2 + \dots + \lambda_n^2$$

A.5 - Demonstração do teorema 3.5

Derivando-se R^2 com relação a b e igualando - se a zero tem-se:

$$\hat{b} = \bar{x} - A'\bar{y} \quad \text{onde } \bar{y} = \frac{\sum y_r}{n}, \quad \bar{x} = \frac{\sum x_r}{n}$$

Sendo que as duas configurações tem centro na o rigem $\bar{x} = \bar{y} = 0$ e portanto $\boxed{\hat{b} = 0}$

Pode-se então escrever R^2 novamente:

$$R^2 = \min_A \sum_{r=1}^n (x_r - A'y_r)' (x_r - A'y_r)$$

$$R^2 = \min_A \text{tr} (X - YA) \cdot (X - YA)'$$

$$R^2 = \text{tr} XX' + \text{tr} YY' - 2 \max_A \text{tr} X'YA$$

Sobre A há a restrição $AA' = I$ isto é $a'_i a_i = 1$ e $a'_i a_j = 0$, $i \neq j$ onde a'_i é a i-ésima linha de A. Há portanto $\frac{p(p-1)}{2}$ restrições.

Seja $\frac{1}{2} \Lambda$ a matriz simétrica dos multiplicadores de Lagrange para estas restrições. O objetivo é maximizar.

$f(x, y, \Lambda) = \text{tr} \{Z' A - \frac{1}{2} \Lambda (AA' - I)\}$ onde $Z' = X'Y$. Derivando-se diretamente tem-se:

$$\begin{aligned} & \frac{\partial}{\partial A} \text{tr} \{Z'A - \frac{1}{2} \Lambda (AA' - I)\} = \\ = & \frac{\partial}{\partial A} \text{tr} Z'A - \frac{\partial}{\partial A} \frac{1}{2} (\Lambda AA') + \frac{\partial}{\partial A} \frac{1}{2} \Lambda I = \\ & = Z - \frac{1}{2} \cdot 2 \Lambda A = Z - \Lambda A \end{aligned}$$

Igualando-se a zero obtém-se:

$$Z = \Lambda A$$

Sendo $Z = V \Gamma U'$ (pelo teorema da decomposição), Λ simétrica e A ortogonal tem-se:

$$\begin{aligned} Z \cdot A' &= \Lambda AA' \rightarrow ZA' = \Lambda \rightarrow \Lambda' = AZ' \rightarrow \Lambda \Lambda' = ZA'. \quad AZ' \\ \rightarrow \Lambda^2 &= ZA' AZ' \rightarrow \Lambda^2 = ZZ' \rightarrow \Lambda^2 = (V\Gamma U') (U\Gamma V') \rightarrow \Lambda^2 = V\Gamma\Gamma V' \end{aligned}$$

Pode-se portanto tomar $\Lambda = V \Gamma V'$ pois $\Lambda \Lambda = V\Gamma V' V\Gamma V' = V\Gamma\Gamma V'$.

Substituindo-se este valor de Λ em $Z = \Lambda A$ tem-se:

$$Z = V\Gamma V'\hat{A} \rightarrow V\Gamma U' = V\Gamma V'\hat{A} \rightarrow U' = V'\hat{A} \rightarrow VU' = VV'\hat{A}$$

$$\rightarrow \hat{A} = VU'$$

Observa-se que \hat{A} é ortogonal. Usando-se este valor de \hat{A} em:

$$R^2 = \text{tr } XX' + \text{tr } YY' - 2 \max_A \text{tr } X' Y A$$

tem-se:

$$R^2 = \text{tr } XX' + \text{tr } YY' - 2 \text{tr } Z'VU' \rightarrow$$

$$R^2 = \text{tr } XX' + \text{tr } YY' - 2 \text{tr } U' \Gamma V'VU' \rightarrow$$

$$R^2 = \text{tr } XX' + \text{tr } YY' - 2 \text{tr } U' \Gamma U' \rightarrow$$

$$\boxed{R^2 = \text{tr } XX' + \text{tr } YY' - 2 \text{tr } \Gamma}$$

Finalmente para se verificar se \hat{A} maximiza a função $\text{tr} \{Z'A - \frac{1}{2} \Lambda (AA' - I)\}$ ou se é apenas um ponto estacionário deve-se derivar pela segunda vez a função com relação a A. Para esta finalidade é mais conveniente escrever A como um vetor $\underline{a} = (a'_{(1)}, \dots, a'_{(p)})$. Então, a função é uma função quadrática dos elementos de \underline{a} e a sua derivada segunda com relação a A pode ser expressa pela matriz:

$$- I_p \textcircled{X} \Lambda$$

\textcircled{X} : Produto de Kronecker

$$\text{Tem-se: } f(X, Y, \Lambda) = \text{tr} \{Z'A - \frac{1}{2} \Lambda (AA' - I)\}$$

$$\frac{\partial f(X, Y, \Lambda)}{\partial A} = Z - \Lambda A$$

$$\frac{\partial^2 f(X, Y, \Lambda)}{\partial \Lambda} = - \text{Ip} \left(\begin{matrix} X \\ \Lambda \end{matrix} \right)$$

Como $\Lambda = V\Gamma V'$ e os elementos da diagonal de Γ são não negativos, observa-se que a matriz derivada segunda é negativa semi-definida. Portanto A maximiza $f(X, Y, \Lambda)$.

APÊNDICE B - REGRESSÃO MONOTÔNICA

APÊNDICE B: REGRESSÃO MONOTÔNICA

Regressão Monotônica por mínimos Quadrados foi discutida em vários artigos (Barton e Mallons 1961, Bartholomeu 1959 e 1961, Miles 1959 e outros) e é usada como rotina de base em várias técnicas de análise de dados.

Kruskal (1964-b) descreve o seu algoritmo para a Regressão Monotônica utilizada no seu programa de E.M.D. .

Neste apêndice vamos apenas definir matematicamente a Regressão Monotônica, também chamada Regressão Isotônica, mais informações a respeito podem ser encontradas no livro de R.E. Barlow (1972), as propriedades de continuidade e diferenciabilidade dessa Regressão foram discutidas por Kruskal (1971).

Definição B.1 Seja X um conjunto finito $\{x_1, \dots, x_k\}$ com uma ordem simples $x_1 < x_2 < \dots < x_k$. Uma função real de X é isotônica se $x, y \in X$ e $x < y \Rightarrow f(x) \leq f(y)$. Seja g uma dada função em X e ω uma dada função positiva em X . Uma função isotônica g^* de X é uma regressão isotônica de g com pesos ω com relação à ordem simples $x_1 < x_2 < \dots < x_k$ se minimizar na classe das funções isotônicas f de X a soma

$$\sum_{x \in X} [g(x) - f(x)]^2 \omega(x)$$

No E.M.D. a relação isotônica desejada pode ser formalizada como segue: Seja X o conjunto de todos os pares de estímulos (i, j) $1 \leq i < j \leq n$. Define-se uma ordem simples " \leq " em X exigindo-se $(i_1, j_1) \leq (i_2, j_2)$ se $\delta_{i_1, j_1}^* < \delta_{i_2, j_2}^*$ onde $\delta_{i, j}^*$ representa a dissimilaridade do par (i, j) . Espera-se encontrar n pontos num espaço r -dimensional de modo que $(i_1, j_1) \leq (i_2, j_2) \Rightarrow d_{i_1, j_1} \leq d_{i_2, j_2}$ (d_{ij} : distância entre os pontos correspondentes aos estímulos i e j), isto é, de modo que esta distância d , seja uma função isotônica em X .

Na prática nem sempre se encontra uma configuração que satisfaça essas condições; nota-se por \underline{d}^* a regressão isotônica de \underline{d} em X e procura-se minimizar o STRESS

$$\sigma = \sqrt{\frac{\sum \sum (d_{ij} - d_{ij}^*)^2}{\sum \sum d_{ij}^2}}$$

Sendo assim a Regressão Isotônica entra no método somente para encontrar o STRESS de um determinado conjunto de pontos

APÉNDICE C - "STEEPEST DESCENT"

APÊNDICE C - "STEEPEST DESCENT"

C.1 - Métodos de Gradiente

Os métodos de gradiente para minimização de funções usam uma medida da declividade da função para indicar a direção do mínimo. O primeiro ponto é escolhido aleatoriamente e depende da opinião subjetiva do programador a respeito da localização desse mínimo.

Os demais pontos são obtidos através da equação iterativa

$$x^{p+1} = x^p + h^p D^p$$

p : número da iteração

X : vetor n dimensional (x_1, x_2, \dots, x_n)

h : constante positiva

D : vetor direção n-dimensional

O vetor D^p determina a direção a ser tomada a partir do p-ésimo ponto e o módulo de $h^p D^p$ determina quanto grande é o passo a ser dado nessa direção.

Os vários métodos de gradiente diferem na escolha do fator de escala h^p e na direção do vetor D^p .

C.2 - Método do "Steepest Descent"

Os métodos que trocam todas as variáveis independentes a cada iteração usam primariamente o seguinte vetor direção $D^p = - B^{-1} \Delta^p$

onde

$$\Delta^p = \left[\left. \frac{\partial f}{\partial x_1} \right|_p, \left. \frac{\partial f}{\partial x_2} \right|_p, \dots, \left. \frac{\partial f}{\partial x_n} \right|_p \right] \quad e$$

B é uma matriz (nxn) positiva definida de pesos.

A mais antiga e mais simples ponderação do vetor gradiente Δ^p é B = I onde I é a matriz identidade. Este método foi inicialmente descrito por Cauchy em 1847 e é conhecido como método do "Steepest descent" (ver H. A. Spang, III², 1962).

Kruskal (1964-b) usa a seguinte equação iterativa

$$x^{p+1} = x^p + h^p D^p \quad \text{onde}$$

$$D^p = - I \cdot \Delta^p \quad e \quad h^p = \frac{\alpha}{|\Delta^p|}$$

$|\Delta^p|$: módulo de Δ^p

α : é uma constante que varia de iteração para iteração e que Kruskal (1964) determinou após um grande número de experimentos numéricos (ver Kruskal, 1964-b).

APÊNDICE D - CONFIGURAÇÃO INICIAL ALSICAL

APÊNDICE D: FASE INICIAL ALSCAL

O procedimento inicial a ser descrito neste apêndice é muito semelhante ao trabalho apresentado por Schonne mann (1972) em que ele obtém uma solução algébrica para minimizar a função (*) para dados livres de erro no nível de mensuração razão

$$\psi^2 (X, W, P^*) = \sum_{i=1}^N \text{tr} (P_k^* - X W_k X')' (P_k^* - X W_k X') \quad (*)$$

P^* : coleção de matrizes P_k^* ; $k = 1, 2, \dots, N$

W_i : matriz diagonal dos pesos para o sujeito k

P_k^* : matriz de pseudo produtos escalares, obtidos através das dissimilaridades do sujeito k sob hipóteses métricas ou não métricas.

O procedimento apresentado abaixo foi discutido por Young, Takane e de Leeuw em seu artigo de 1977.

Suponhamos que há N matrizes de produtos escalares P_k (uma para cada um dos N sujeitos) cada uma de ordem n (n estímulos) que satisfazem:

$$P_k = X W_k X' \quad (1)$$

O problema que se propõe é obter X e W_k a partir dos P_k sob as hipóteses de X ser de posto coluna completo e

lecionar nossa rotação K de modo que:

$$K' (Y'Y)^{-1} Y' P_k Y (Y'Y)^{-1} K = W_k \quad (3)$$

seja diagonal para todo k . (Observa-se que $K'K = KK' = I$ e que $K^{-1} = K'$).

Seja:

$$C_i = (Y'Y)^{-1} Y' P_k Y (Y'Y)^{-1} \quad (4)$$

$$C_i = K W_k K' \quad (5)$$

A solução algébrica de Schonemann encontra a rotação K que diagonaliza simultaneamente as C_k ($k = 1, \dots, N$) através da obtenção da decomposição espectral das C_k para todo k . No entanto isto só se aplica quando não há erros nos dados. O procedimento SUMSCAL (de Leeuw e Pruzansky, 1975) encontra k que minimiza a soma dos quadrados dos elementos da diagonal dos W_k ($k = 1, \dots, N$) na presença de erros, dando essencialmente a mesma solução de Carroll e Chang (1970) no INSCAL, porém em menos tempo. Trabalhando com SUMSCAL, de Leeuw sugeriu (Young, Takane, Lewyckyj, 1978) que é desejável obter K que diagonalize uma combinação linear das matrizes C_k , onde a combinação linear é tal que a matriz resultante, quando diagonalizada, tem os elementos da diagonal de variância máxima. Ele baseou sua sugestão na observação de que a rotação K não é definida de forma única quando W_k tem diagonal constante. Portanto, pelo menos em algum sentido, K é mais unicamente definida quando as diago

de que os elementos da diagonal da matriz W_k são estritamente positivos pelo menos para um sujeito.

Para alguma matriz R diagonal não singular de ordem r (há r dimensões) temos:

$$P_i = X R (R^{-1} W_i R^{-1}) R X' \quad (2)$$

e conseqüentemente deve-se fazer alguma restrição nas dimensões das W_k , com objetivos de identificação. Portanto definimos:

$$D. = \frac{1}{N} \sum_{K=1}^N W_k$$

e assumimos que $D. = I$ implicando que $P. = XX'$ (onde $P.$ é a média dos P_k). As soluções para esta equação são obtidas a menos de uma rotação.

Nós selecionamos umas dessas soluções arbitariamente, por exemplo usando r passos do processo de Cholesky usando as r raízes características e vetores característicos dominantes de $P.$. Chamamos essa solução arbitária de Y . Segue-se então que:

$$X = Y \cdot K$$

onde K é uma matriz de rotação. Sabemos também que:

$$(X'X)^{-1} X' P_k X (X'X)^{-1} = W_k$$

tem que ser diagonal para todo k . Segue-se que devemos se

nais tem variância máxima.

Para alcançar esse objetivo, nós definimos a combinação linear das matrizes C_k como:

$$\ell(C_k) = \sum_{i=1}^N e_k C_k \quad (6)$$

Seja $\ell(C_k) = K\Delta K'$, onde K é uma matriz de vetores característicos e Δ uma matriz diagonal de raízes características (K diagonaliza $\ell(C_k)$, e Δ diagonalizou os elementos na sua diagonal). A variância dos elementos de entrada da diagonal (raízes características) pode ser escrita:

$$\rho = \text{tr}(\Delta^2) - \frac{\text{tr}^2(\Delta)}{t} = \text{tr}[\ell^2(C_k)] - \frac{\text{tr}^2[\ell(C_k)]}{t} \quad (7)$$

onde t é a ordem de $\ell(C_k)$. Nós maximizamos ρ com relação a $e' = (e_1, \dots, e_N)$ sob a restrição de $e'e = 1$.

Seja:

$$\rho^* = \rho - \lambda (e'e - 1) = \sum_{k,j} e_k e_j \text{tr}(C_k C_j) - \frac{[\sum_i e_i \text{tr}(C_k)]^2}{t} - \lambda (e'e - 1) \quad (8)$$

então:

$$\frac{1}{2} \frac{\partial \rho^*}{\partial e_i} = \sum_k e_k \text{tr}(C_k C_i) - \frac{\text{tr}(C_i) \sum_k e_k \text{tr}(C_k)}{t} - \lambda e_i = 0 \quad (9)$$

Portanto

$$\lambda_{ei} = \sum_k a_{ki} e_k \quad (10)$$

onde

$$a_{ki} = \text{tr} \cdot (C_k C_i) - \frac{\text{tr} (C_k) \text{tr} (C_i)}{t} \quad (11)$$

A equação (10) é uma equação característica para a matriz A cujos elementos são definidos por (11), e \underline{e} (com elementos i_k) é o vetor característico da matriz A. Pode-se prontamente mostrar que \underline{e} é o vetor característico correspondente às raízes características dominantes de A. Uma vez que \underline{e} é determinado, nós definimos a combinação linear $\ell(C_k)$ por (6) para obter a rotação K e a configuração inicial X desejada.

Observa-se que a variância máxima entre as raízes características não implica necessariamente que cada raiz característica é maximamente distinta das outras. A máxima distinção pode ser construída como intervalos iguais entre raízes características adjacentes, quando elas forem ordenadas de acordo com a sua magnitude, dado que a diferença entre a maior e a menor raiz característica é constante (a qual por sua vez, deve ser a maior possível). Portanto em certas situações, a distinção máxima entre as raízes características não pode ser realizada exatamente pelo critério de

variância máxima, mas do ponto de vista prático da obtenção da configuração inicial, o procedimento acima funciona muito bem (Young, Takane e Lewyckyj, 1978).

Em dados com erros deve-se antes de definir e de finir P., isto é facilmente feito, centrando duplamente os elementos de cada matriz Δ_k , com elementos δ_{kij}^2 (e dividindo por -2) para obter a matriz P_k dos produtos escalares para cada sujeito k . (Vide Método Clássico de E.M.D. Cap. 3) Em seguida acha-se a média para todos os sujeitos P. que pode ser decomposta $P. = Y.Y'$ para obter Y a configuração orientada arbitrária que melhor reproduz os produtos escalares médios. Observa-se que:

- (a) se os dados forem assimétricos, nós fazemos a média entre δ_{kij}^2 e δ_{kji}^2 para todos os i e j em cada matriz k antes de centrar;
- (b) se os dados tiverem elementos perdidos, cada elemento é estimado como média dos julgamentos dos sujeitos;
- (c) a condicionalidade dos dados é ignorada.

Apesar do procedimento descrito assumir que os dados são métricos é possível obter-se pesos negativos, especialmente quando as hipóteses de metricidade forem radicalmente violadas (Observa-se que a definição de modelo Euclidiano Ponderado inclui a exigência que todos os pesos sejam

não negativos). Quando forem observados pesos negativos usa-se o seguinte procedimento: adiciona-se o valor absoluto do peso mais negativo a todos os pesos, assegurando assim todos os pesos se tornarão não negativos.

APÊNDICE E - FASE DE ESCALONAMENTO ÓTIMO ALSICAL

APÊNDICE E - FASE DE ESCALONAMENTO ÓTIMO - ALSCAL

A fase de escalonamento ótimo do ALSCAL tem por objetivo escalonar otimamente as observações quadráticas Δ para obter as disparidades D^* de modo que estas disparidades: (1) estejam de acordo com as restrições de mensuração. (2) sejam estimadores de mínimos quadrados das distâncias D obedecidas as restrições de mensuração.

Portanto pretende-se maximizar a correlação entre as observações e o modelo respeitando as características de mensuração das observações. Resolve-se o seguinte problema

$$\text{MIN}_{D^*} [\sigma^2 (D^* / X, W)]$$

Começa-se computando as distâncias D a partir de X e W , parte-se depois para a obtenção de D^* .

E.1 - Obtenção de D^*

Discutiremos aqui brevemente a obtenção de D^* , maiores detalhes podem ser encontrados nos artigos de Takane e Young (1977) e σ^2 , de Leeuw e Takane (1976).

Para a maioria dos dados o procedimento de escalonamento ótimo é a regressão linear ou regressão monotônica (Apêndice B), no entanto há dados que necessitam de outros

procedimentos. No entanto, todos os vários tipos de transformações de escalonamento ótimo podem ser definidas como transformações lineares das distâncias quadráticas. Isto é:

$$d_{kij}^* = \ell(d_{kij}^2)$$

ℓ : transformação linear paralelamente às restrições de mensuração usadas para definir t (Capítulo 6)

Assim sendo ℓ define as d_{kij}^{*2} que minimizam o SSTRESS para valores fixos de W e X segundo um critério de mínimos quadrados.

Uma vez que se vai fazer uma regressão de d_{kij}^2 com relação a δ_{kij}^2 segundo um critério de mínimos quadrados e obedecendo às restrições de mensuração, ℓ pode ser representado por um operador de projeção da forma:

$$\ell : E = Z \cdot (Z'Z)^{-1} Z' \quad (1)$$

Z = matriz de vetores que definem o espaço no qual o vetor das d_{kij}^2 sofrerá regressão.

Para cada tipo de transformação t define-se uma matriz Z .

A - Transformação razão (t^r): $Z = \Delta$ (vetor de observações quadráticas).

B - Transformação intervalar (t^i): Z reduz-se ao caso A

após a estimativa de uma constante aditiva adequada.

Nos casos A e B os estimadores de mínimos quadrados podem ser obtidos por técnicas de regressão bem conhecidas.

C - Transformação ordinal (t^o) e Transformação nominal (t^n):

Z é definida como uma matriz de variáveis dummy que indicam as distâncias que devem ser repetidas para satisfazer às restrições de mensuração.

C₁ - Transformação ordinal contínua (t^{co}): os elementos a serem repetidos envolvem violações de ordem (procedimento primário de Kruskal para dados repetidos - Capítulo 9).

C₂ - Transformação ordinal discreta (t^{do}): os elementos a serem repetidos envolvem também observações que são categoricamente equivalentes (procedimento secundário de Kruskal para dados repetidos - Capítulo 9).

C₃ - Transformação nominal discreta (t^{dn}): Z indica que as distâncias que correspondem a observações categoricamente equivalentes devem ser repetidas. O estimador de mínimos quadrados neste caso envolve simplesmente as médias das categorias.

C₄ - Transformação nominal contínua (t^{cn}): Z indica as distâncias que caem fora do intervalo desejado. Neste ca

so os estimadores de mínimos quadrados são os limites do intervalo para as distâncias que forem violadas e as próprias distâncias para aquelas que não forem violadas (procedimento pseudo ordinal de de Leenw, Young e Takane (1976) para determinar os limites ótimos).

Observe-se que para todas as transformações discretas exceto a discreta ordinal Z é conhecida a priori. Para as demais transformações Z é conhecida após ser feita a análise e nesses casos ela varia de iteração para iteração dependendo da natureza das distâncias.

Uma vez representado o escalonamento ótimo por um operador de projeção da relação (1) para os quatro níveis e dois processos de mensuração, isso permite que uma vez definidos um vetor coluna d contendo os $\frac{N(n-1)}{2}$ elementos d_{kij}^2 e outro vetor coluna d^* contendo os elementos correspondentes d_{kij}^{*2} possa-se observar que

$$d^* = E \cdot d \quad (2)$$

Além disso, essa equação que é implícita em termos de dados não condicionados pode ser facilmente estendida para dados condicionados. Quando os dados forem matriz-condicional define-se uma matriz Z_i para cada indivíduo separadamente e em seguida constroí-se uma super matriz Z com os Z_i 's na diagonal. Para os dados linha-condicionados define-se Z_{ij} para cada linha de cada matriz individual e em se-

guida constroi-se a super-matriz Z com os Z_{ij} 's na diagonal.

A importância da relação (2) se deve ao fato de agora se poder expressar o SSTRESS sob forma matricial e inteiramente em função das distâncias. Definindo-se $\tilde{E} = I - E$ tem-se:

$$\sigma^2(X, W, D^*) = d' \tilde{E} d \quad (3)$$

Analogamente escreve-se o SSTRESS normalizado como sendo:

$$\sigma'^2(X, W, D^*) = \frac{d' \tilde{E} d}{d'd} = d' \tilde{E} (d'd)^{-1} \tilde{E} d \quad (4)$$

E.1.1 - Constante aditiva

Vamos descrever agora o procedimento apresentado por Takane e Young (1977) para estimar a constante aditiva quando os dados se apresentarem no nível de mensuração intervalar.

Assumindo-se que as observações estão definidas no nível de mensuração intervalar, tem-se:

$$d_{kij} = a(\delta_{kij}) + b \quad (5)$$

para a , b constantes desconhecidas. Se o objetivo fosse otimizar o STRESS o problema de estimação seria um simples pro

blema de regressão envolvendo d_{kij} e δ_{kij} . Porém a situação é mais complicada para se otimizar o SSTRESS pois trabalha-se com uma relação quadrática ao invés de uma simples relação linear

$$d_{kij}^2 = a^2 (\delta_{kij})^2 + 2ab (\delta_{kij}) + b^2 \quad (6)$$

Isto difere de uma simples regressão de d_{kij}^2 sobre δ_{kij}^2 que seria o caso se houvesse uma relação linear entre d_{kij}^2 e δ_{kij}^2 (caso em que $b = 0$; nível razão).

Apesar de ser possível resolver (6) diretamente é mais simples redefinir o problema, tal que:

$$d_{kij}^{*2} = \alpha + \beta (\delta_{kij}) + \gamma (\delta_{kij})^2 \quad (7)$$

deve-se obter as melhores estimativas de α, β e sob a restrição $\beta^2 = 4\alpha\gamma$

(8)

Define-se então: (1) vetor de parâmetros $Y' = [\alpha, \beta, \gamma]$

(2) Matriz de ordem $N \cdot [n(n-1)/2] \times 3$ (S) onde:

coluna 1 : contém unidades

coluna 2 : contém observações

coluna 3 : contém observações quadráticas

(3) Vetor coluna \underline{d} contendo as $N[n(n-1)/2]$ d_{kij}^2 arranjadas da mesma forma que as δ_{kij} em S.

A partir dessas definições pode-se expressar o SSTRESS numa situação de nível intervalar de mensuração como:

$$\sigma^2(Y, \lambda/S, d) = (d - SY)'(d - SY) + (\beta^2 - 4\alpha\gamma) \quad (9)$$

que se minimiza resolvendo por Y e λ (multiplicador de Lagrange)

O estimador de mínimos quadrados para os parâmetros obedecendo à restrição é:

$$Y = (S'S)^{-1} S'd + \lambda \cdot q \quad (10)$$

Para resolver pelos multiplicadores de Lagrange define-se:

$$(S'S)^{-1} \cdot g = \begin{bmatrix} q_1 \\ q_2 \\ q_3 \end{bmatrix} \quad (11)$$

onde $g' = [-2\gamma, \beta, -2\alpha]$ derivadas de (8)

Deve-se em seguida resolver:

$$(\hat{\beta} + \lambda q_2)^2 = 4 (\hat{\alpha} + \lambda q_1) \cdot (\hat{\gamma} + \lambda q_3) \quad (12)$$

Seleciona-se a melhor das duas soluções (isto é, aquela que minimiza o SSTRESS) avaliando o conjunto de \hat{Y} correspondente a cada raiz.

Young, Takane e Lewycky (1978) comentam e modificam o procedimento descrito acima. Afirmam que esse procedimento pode gerar disparidades negativas, sugerem portanto que se estime a constante aditiva sob a condição de não haver disparidades negativas.

Na equação (5) restringem \underline{a} e \underline{b} de modo que $d_{kij}^* \geq 0$. Quando o mínimo d_{kij}^* for negativo através do procedimento sem restrição então definem:

$$d_{kij}^* = a (\delta_{kij} + \tilde{b}) \quad (13)$$

onde $\tilde{b} = \frac{b}{a}$, toma-se $\tilde{b} = -\text{MIN}(\delta_{kij})$

Definem depois:

$$\hat{\delta}_{kij} = \delta_{kij} - \text{MIN}(\delta_{kij}) \quad (14)$$

e estimam as distâncias quadráticas como:

$$d_{kij}^2 \cong d_{kij}^{*2} = a^2 (\delta_{kij}^2)$$

Isto reduz ao caso de nível de mensuração razão.

E.2 - Normalização da solução do E.M.D.

Vamos descrever o processo de normalização utilizado no ALSCAL, Takane e Young (1976) em seu artigo discutem esse processo e o comparam com o de Carroll e Chang (1970).

Há duas considerações separadas no que concerne à normalização da solução do E.M.D., uma com respeito à normalização dos parâmetros (a configuração, os pesos e os valores ótimamente escalonados) e a outra com respeito à função de perda. Enquanto a normalização da função de perda deve ser feita em cada iteração e de maneira específica de modo a impedir certo tipo de soluções degeneradas; a normalização dos parâmetros deve ser feita somente antes de se imprimir a solução final. Desta forma os números dos quais o experimentador dispõe estão em unidades padronizadas para poder compará-los com outras soluções.

A configuração é normalizada de modo que a média das projeções em cada dimensão seja zero e que a variância das projeções em cada dimensão seja unitária.

Normalizam-se os dados otimamente escalonados, dando-se preferência a esses ao invés dos dados simplesmente, uma vez que dados qualitativos não podem ser normalizados e os dados otimamente escalonados são quantitativos mesmo quando os dados são qualitativos. Para qualquer tipo de condicionalidade, normaliza-se para cada partição (linha, matriz ou qualquer) a soma de quadrados dos dados otimamente escalonados quadráticos que é constante.

É interessante normalizar os "pseudo" produtos escalares uma vez que sua soma de quadrados é constante. E em

se fazendo a soma de quadrados dos "pseudos" produtos escalares de um indivíduo igual a 1, a soma dos quadrados dos pesos do indivíduo pode ser interpretada grosseiramente como indicador da proporção de variância nos "pseudo" produtos escalares que é devida ao modelo (elimina-se o grosseiro se $X'X = I$).

Portanto, uma vez completada a análise, faz-se uma normalização final que consiste em computar os "pseudo" produtos escalares a partir dos dados otimamente escalonados e em seguida torna-se cada soma de quadrados dos pseudo produtos escalares de um indivíduo igual a 1.

Deve-se observar, no entanto, que a condicionalidade dos dados e as diferenças resultantes na normalização tem certas implicações na interpretação dos pesos (Takane e Young (1977)).

Vamos considerar agora a normalização da função de perda. Apesar que atualmente a otimização é feita com relação à função de perda não normalizada

$$\phi^2(X, W, D^*) = \sum_k \sum_i \sum_{j=1}^{i=1} (d_{kij}^{*2} - d_{kij}^2) \quad (15)$$

pode-se indiretamente otimizar a função normalizada se para isso forem dados passos convenientes (de Leeuw, Young e Takane (1976)). É conveniente não se trabalhar com a função normalizada uma vez que ela é uma razão entre duas for

mas quadráticas cujas derivadas parciais são consideravelmente mais complicadas que as da função não normalizada.

Ganha-se simplicidade se a função for normalizada com relação aos dados otimamente escalonados. Não há ganho se a normalização for com relação às distâncias. Portanto, Takane e Young (1977) normalizam a função de perda com relação aos dados otimamente escalonados (D^*).

Há mais duas considerações a serem feitas. Uma delas foi enfatizada por (Kruskal e Carroll (1969) e Roskam (1969)) afirma que a normalização deve ser feita dentro das partições, isto é, para dados não condicionados, onde todos os dados formam uma única partição, computa-se um único SSTRESS normalizado

$$\sigma_n^2 = \frac{\sum_k \sum_i \sum_j (d_{kij}^{*2} - d_{kij}^2)^2}{\sum_k \sum_i \sum_j d_{kij}^{*4}} \quad (16)$$

Por outro lado para dados matriz condicionados, onde de cada matriz forma uma partição separada computa-se o valor do SSTRESS normalizado para cada matriz e em seguida obtém-se a sua média.

$$\sigma_m^2 = \frac{1}{N} \sum_{k=1}^N \left[\frac{\sum_i \sum_j (d_{kij}^{*2} - d_{kij}^2)^2}{\sum_i \sum_j d_{kij}^{*4}} \right] \quad (17)$$

Finalmente para dados linha condicionados onde cada linha é uma partição computa-se o SSTRESS normalizado para cada linha e em seguida é feita a média

$$\sigma_r^2 = \frac{1}{Nn} \sum_{k=1}^N \sum_{i=1}^n \left[\frac{\sum_j (d_{kij}^{*2} - d_{kij}^2)^2}{\sum_j d_{kij}^{*4}} \right] \quad (18)$$

Observa-se que se usa a soma de quadrados dos dados otimamente escalonados quadráticos uma vez que a perda é definida em distâncias quadráticas e dados otimamente escalonados quadráticos (sendo este o motivo da potência 4 no denominador).

Pode-se reduzir as fórmulas (16), (17) e (18) a uma única introduzindo-se um novo símbolo:

$$\alpha_{ijkl} = \begin{cases} 1 & \text{se } d_{kij}^* \text{ pertence à partição } l \\ 0 & \text{caso contrário} \end{cases}$$

Escreve-se:

$$\sigma'^2 = \sum_l \frac{1}{\sum_k \sum_i \sum_j \alpha_{kijl}} \left[\frac{\sum_i \sum_j \sum_k (d_{kij}^{*2} - d_{kij}^2)^2 \alpha_{kijl}}{\sum_k \sum_i \sum_j d_{kij}^{*4} \alpha_{kijl}} \right] \quad (20)$$

que é a expressão completa da função SSTRESS normalizado que o programa ALSCAL otimiza.

A outra consideração a ser feita é quanto à normalização pela soma dos quadrados dos dados otimamente escalonados quadráticos como em (20), ou pela variância dos dados otimamente escalonados quadráticos. Essa questão que também foi abordada por Kruskal e Carroll (1969) e Roskam (1969) fica melhor respondida provavelmente afirmando-se que para dados linha condicionados deve-se provavelmente usar a variância, por outro lado para outros tipos de dados deve-se usar a soma de quadrados, Takane e Young (1977) usam apenas a fórmula (20), estão elaborando um plano para incorporar as duas fórmulas ao ALSCAL.

Finalmente, se foi resolvido o problema não normalizado representado pela fórmula (15) e for necessário otimizar o problema normalizado representado pela fórmula (20), pode-se mostrar que basta resolver o problema não normalizado e ajustar o módulo do vetor dos dados otimamente escalonados quadráticos, multiplicando-se todos os seus elementos pelo fator

$$\frac{\sum_{k i j} \sum_{k i j} \sum_{k i j} d_{kij}^4}{\sum_{k i j} \sum_{k i j} \sum_{k i j} d_{kij}^{*2}} \alpha_{kij} \quad \alpha_{kij}$$

APÊNDICE F - FASE DE ESTIMAÇÃO DO MODELO : ALSAL

APÊNDICE F - FASE DE ESTIMAÇÃO DO MODELO : ALSICAL

A fase de estimação do modelo da técnica ALSICAL se propõe a resolver sucessivamente dois problemas de mínimos quadrados condicionados.

$$1^a \text{ sub-fase : } \text{MIN}_W [\varphi^2(W/X, D^*)]$$

$$2^a \text{ sub-fase : } \text{MIN}_X [\varphi^2(X/W, D^*)]$$

F.1: Obtenção dos estimadores dos pesos

Para estimar W obtêm-se as derivadas parciais de $\varphi^2(X, W, D^*) = \sum_k \sum_i \sum_j^{i-1} (d_{kij}^{*2} - d_{kij}^2)^2$ com relação aos elementos de W e igualam-se estas derivadas a zero. Esse sistema de equações é então resolvido com relação a W.

Para simplificar a derivação define-se uma matriz y de ordem $(\frac{n(n-1)}{2} \times r)$ onde as colunas de y contêm as projeções de todas as distâncias em cada dimensão.

$$Y = \begin{matrix} & \begin{bmatrix} a \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} \\ (i,j) & \begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix} \end{matrix}$$

Define-se também uma matriz D* de ordem N x $\frac{n(n-1)}{2}$ cujas linhas contêm as $\frac{n(n-1)}{2}$ observações otimamente esca-

lonadas para cada individuo, seus elementos são arranjados em correspondência com a matriz y (Esta matriz D^* contém as mesmas informações que a matriz D^* descrita no CAP.6 mas é organizada de forma diferente, nesse apêndice nos referiremos a essa nova matriz D^*).

A partir dessas novas definições pode-se escrever o SSTRESS como sendo

$$\sigma^2 (y, w, D^*) = \text{tr} (D^* - wy')' . (D^* - wy') \quad (1)$$

a partir do qual obtêm-se o estimador de mínimos quadrados de W , dado por:

$$W = D^*y (y'y)^{-1} \quad (2)$$

F.2: Restrições de não - negatividade

Uma dificuldade que surge na obtenção da matriz W pelo processo acima descrito é a não observância da restrição de não - negatividade exigidas pelo modelo Euclidiano ponderado utilizado onde:

$$d_{kij}^2 = \sum_{r=1}^R w_{kr} (x_{ir} - x_{jr})^2 \quad w_{kr} \geq 0$$

Vamos descrever abaixo um procedimento que permita incorporar essa restrição dentro do processo de A L S (mínimos quadrados alternados)

Para esse processo é fundamental que se observe que o processo de estimação apresentado pela fórmula (2) é in-

dependente para cada indivíduo. Isto é, os valores estimados para os pesos de um indivíduo não afetam os valores estimados para os pesos de algum outro indivíduo. Isto pode ser observado pelo fato do SSTRESS (1) poder ser decomposto numa soma de componentes separadas, cada uma como função de um único indivíduo.

Uma vez que os pesos para um determinado indivíduo são independentes dos pesos para os outros, pode-se impor a restrição de não - negatividade para alguns indivíduos que tenham pesos negativos sem ter que modificar os pesos para os outros indivíduos. Observa-se, no entanto, que os pesos, para um dado indivíduo, não são independentes entre si, o que significa que não se pode simplesmente igualar os seus pesos negativos a zero e deixar os seus pesos positivos inalterados. Se fizermos isto destruimos as propriedades de mínimos quadrados dos estimadores dos pesos.

A solução para o problema é a seguinte: (Young , De Leeuw e Takane 1977)

1º) Obtêm-se os estimadores de mínimos quadrados de W através da fórmula (2)

2º) Usam-se essas estimativas para os indivíduos cujos pesos são não negativos

3º) Para os demais indivíduos, iguala-se um dos pesos negativos a zero, sendo ele o estimador de mínimos

quadrados sob a restrição de que todos os outros pesos são constantes.

4º) É reestimado o valor de um outro peso sob a hipótese de que todos os outros pesos são constantes. O estimador de mínimos quadrados condicionados para um só peso é dado por:

$$w_{ka} = \frac{(d_k^* - \sum_{b \neq a} w_{kb} y_b)' y_a}{(y'a \quad y_a)}$$

w_{ka} : peso atribuído pelo indivíduo k à dimensão a

y_a : a -ésima coluna da matriz Y

Se este peso estimado for negativo, iguala-se a zero.

5º) Repete-se então esse processo para cada dimensão até que todos os pesos para o indivíduo sejam não negativos.

F.3: Obtenção dos estimadores de mínimos quadrados da configuração

Essa sub-fase é um pouco mais complicada que a estimação dos pesos uma vez que as derivadas parciais do SSTRESS com relação aos elementos da configuração X não são funções lineares das coordenadas (x_{ia}) . Sendo o SSTRESS uma função quarta das coordenadas (x_{ia}) , o sistema obtido

pela derivação é de 3º grau em x_{ia} . Várias são as formas de resolver esse sistema, Young (1977) discute algumas, dividindo-as em dois tipos: as que permitem a estimação simultânea de todas as coordenadas e as que estimam uma coordenada por vez num processo iterativo. Ele opta por um processo intermediário, estimando um bloco de coordenadas x_{ia} simultaneamente; de modo que o número de coordenadas simultaneamente estimadas não seja grande o suficiente, a ponto de tornar o processo tão vagaroso que chegue a cancelar os benefícios derivados da estimação simultânea.

Para esse fim aplica uma modificação do método de Newton - Raphson para obter o novo conjunto de estimadores de mínimos quadrados condicionados de todas as coordenadas de um único ponto simultaneamente, resolvendo a seguir sucessivamente para cada um dos pontos. Isto é, estima $x_{ia} = (x_{i1}, x_{i2}, \dots, x_{ir})$ simultaneamente para um determinado estímulo i e sucessivamente repete o processo para todos os estímulos ($i=1, \dots, n$) (Yates 1972).

O trabalho desenvolvido por Young (1977) consiste em aplicar o método de Newton - Raphson para obter estimadores de mínimos quadrados condicionados que resolvem o problema : $\text{MIN}_{x_i} \left[\phi (x_i/x_j, W, D^*); i \neq j \right]$. Portanto o seu procedimento é colocar o método de Newton - Raphson na forma de trabalho de A.L.S. para resolver um sistema de

equações cúbicas em r incógnitas.

Young (1977) usa a modificação de Gill e Murray (1972) do procedimento de Newton - Raphson que garante que a matriz Hessian vem a ser positiva definida. Isso garante que se está procedendo em direção descendente. Uma vez que a Hessian é sempre pelo menos positiva semi-definida é desejável garantir que isso ocorra durante todo o processo de estimação. Se isso não for "suficiente" então deliberadamente são escolhidos valores e somados na diagonal para forçá-la a ser positiva definida (Gill e Murray 1972). Isso impede a convergência para um máximo ou para algum outro ponto que não seja mínimo.

O processo usado por Young (1977) começa verificando as duas primeiras derivadas do SSTRESS com relação a x_{ia} para um i fixo e a variando de 1 a r .

Para simplificar a derivação observa que:

$$\begin{aligned} d_{kij}^{*2} - d_{kij}^2 &= d_{kij}^{*2} - \sum_{a=1}^r w_{ka} (x_{ia} - x_{ja})^2 \\ &= d_{kij}^{*2} - \sum_{a=1}^r w_{ka} x_{ia}^2 + 2 \sum_{a=1}^r w_{ka} x_{ia} x_{ja} - \sum_{a=1}^r w_{ka} x_{ja}^2 \end{aligned}$$

e introduz uma série de definições.

Primeiro reúne os termos que não envolvem x_{ia} (os termos fixos) e define:

$$h_{kij} = d_{kij}^{*2} - \sum_{a=1}^r w_{ka} x_{ja}^2$$

Organizam-se estes termos num vetor h_i que contém todos os termos h_{kij} para i fixo e $j \neq i$ (este vetor tem $N.(n-1)$ elementos).

Define-se também uma super matriz $G = [G_1, G_2]$ com $N.(n-1)$ linhas e $2r$ colunas. Sendo

$$G_1 = - 2. \begin{bmatrix} w_{11} & x_{11} & \dots & w_{1r} & x_{1r} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ w_{11} & x_{j1} & \dots & w_{1r} & x_{jr} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ w_{11} & x_{n1} & \dots & w_{1r} & x_{nr} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ w_{N1} & x_{n1} & \dots & w_{Nr} & x_{nr} \end{bmatrix} \quad j \neq i$$

$$G_2 = \begin{bmatrix} w_{11} & u & \dots & w_{1r} & u \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ w_{N1} & u & \dots & w_{Nr} & u \end{bmatrix}$$

onde u é um vetor coluna $((n-1) \times 1)$ tal que $u_{\ell m} = 1 \quad \forall \ell, m$

Define também um super vetor de ordem $(2r \times 1)$ α_i que consiste do vetor $\tilde{x}_i = (x_{i1}, \dots, x_{ir})$ e um vetor cujos elementos são os quadrados dos elementos de \tilde{x}_i .

Uma vez que é possível expressar os elementos qua-

Os elementos abaixo da diagonal da matriz Hessian (matriz das derivadas parciais de 2ª ordem) são dados por: (para $a \neq b$)

$$h_{ab} = \begin{bmatrix} e'_a & , & 2 x_{ia} & e'_a \end{bmatrix} G'G \begin{bmatrix} e_b \\ x_{ib} & e_b \end{bmatrix}$$

e_a : vetor com a unidade na a-ésima posição e zero nas outras.

A a-ésima diagonal da Hessian é dada por:

$$h_{aa} = - \begin{bmatrix} 0, & 2 e'_a \end{bmatrix} G' h_i + \begin{bmatrix} e'_a, & 2 x_{ia} & e'_a \end{bmatrix} G'G \begin{bmatrix} e_a \\ 2 x_{ia} & e_a \end{bmatrix} + \begin{bmatrix} 0', & 2 e'_a \end{bmatrix} G'G \begin{bmatrix} x_i \\ x_i & x_i \end{bmatrix}$$

Usando o gradiente e a Hessian com o procedimento de Gill e Murray (1974) para o método de Newton e Raphson, Young (1977) obtêm a ℓ -ésima estimativa de x_i denotada por $x_i^{(\ell)}$ de acordo com:

$$x_i^{(\ell)} = x_i^{(\ell-1)} - \theta \cdot \tilde{H}^{(\ell-1)^{-1}} g^{(\ell-1)}$$

θ : tamanho do passo para garantir $\phi^{2(\ell)} < \phi^{2(\ell-1)}$

$$\tilde{H} = \begin{cases} H & : \text{ se } H \text{ positiva definida} \\ H + F & : \text{ se } H \text{ não é positiva definida} \end{cases}$$

F : matriz diagonal com valores positivos na diagonal.

A matriz F é determinada de acordo com o procedimento de Gill e Murray (1974)

Uma vez tendo minimizado o SSTRESS com relação a um único ponto, repete o procedimento para outro ponto, até que todos os pontos tenham sido submetidos ao processo. Isso define uma iteração. Observa-se que uma vez estimadas as coordenadas de um ponto, as velhas coordenadas são automaticamente descartadas e as novas estimativas são inseridas antes que sejam estimadas as coordenadas do outro ponto. Esta substituição imediata é imprescindível uma vez que cada sub-otimização não é independente das outras.

Há um pequeno problema teórico a ser considerado com o procedimento descrito acima. A função a ser minimizada (3) é de 4º grau; sendo seu gradiente um sistema de equações do 3º grau (4); esse sistema de equações cúbicas tem no máximo 2^r mínimos, dos quais não mais que $2^r - 1$ são mínimos locais e um é mínimo global. Portanto em uma dimensão deve haver um mínimo local além do mínimo global, em duas dimensões deve haver em torno de 3 mínimos locais, em 3 dimensões sobe a 7 o número de mínimos locais, e assim por diante. O procedimento proposto converge para um dos mínimos sem assegurar que seja o mínimo global. Segundo Young (1977) re

sultados de análises numéricas indicam que com mais frequência converge-se ao mínimo ótimo (especialmente se tiver uma boa estimativa inicial e se o mínimo ótimo tiver um valor bem melhor que o mínimo local, situação esta que pode ou não existir) e pelo menos ocasionalmente converge-se a um mínimo não ótimo.

Young (1977) é de opinião que esse problema teórico é de pequeno efeito prático, ele discute o caso e apresenta soluções alternativas em seu artigo

APÊNDICE G - SAÍDA DO PROGRAMA ALSCAL

ALSCAL-40 DYNAMIC CORE VERSION.
V E R S I O N 4.01C
AUGUST 1974
FINAL CHANGE 11/17/76

ALTERNATING LEAST SQUARES SCALING

YOSHIO TAKANE, FORREST
ROSTYSLAW J. LEHYCKYJ,
PSYCHOMETRIC LABORATORY
THE UNIVERSITY OF NORTH
CHAPEL HILL, N.C. 2751

COPYRIGHT 1977 FORREST W. YOUNG, YOSHIO TAKANE & ROSTYSLAW J. LEHYCKYJ

JOB TITLEO CATEGORIA TRANTA E DOIS

DATA SPECIFICATIONS-

NRCM - NUMBER OF ROW STIMULI
NCCL - NUMBER OF COLUMN STIMULI
NS - NUMBER OF MATRICES
NDTYP - MEASUREMENT LEVEL
NSIP - DATA TYPE
NPS - MEASUREMENT PROCESS
NWC - MEASUREMENT CONDITIONALITY
NDVX - NUMBER OF CELLS FOR TIED OBSERVATIONS

25 ROW STIMULI
25 COLUMN STIMULI
1 MATRIX
3 = GRIDIN L
1 = SYMMETRIC-SIMILARITY
2 = CONTINUOUS
1 = MATRIX CONDITIONAL
300 CELLS

ANALYSIS SPECIFICATIONS-

MDE - MODEL TYPE
NDIM - NUMBER OF DIMENSIONS (MAXIMUM)
NCPN - NUMBER OF DIMENSIONS (MINIMUM)
NCC - NEGATIVE WEIGHTS PERMITTED
MAXIT - MAXIMUM NUMBER OF ITERATIONS
EPSI - CONVERGENCE CRITERION

C = SIMPLE EUCLIDIAN MODEL (DEFAULT)
4 DIMENSIONS (MAXIMUM)
1 DIMENSIONS (MINIMUM)
C = NEGATIVE WEIGHTS NOT PERMITTED
30 ITERATIONS (MAXIMUM)
0.0010000 = MINIMUM STRESS IMPROVEMENT

I/O OPTIONS-

NCT - PRINT DATA, DISTANCES AND DISPARITIES
NPT - PLOT RESULTS
NPH - PUNCH RESULTS
INDATA - DATA INPUT UNIT NUMBER
INITX - INITIAL STIMULUS COORDINATES
INITRC - INITIAL COLUMN STIMULUS COORDINATES
INITW - INITIAL SUBJECT WEIGHTS
INITMS - INITIAL STIMULUS WEIGHTS

1 = DO PRINT
1 = DO PLOT
2 = PUNCH INITIAL AND DERIVED CONFIGURATIONS
1 = READ DATA FROM CARDS
1 = COMPUTE AND PRINT
C = COMPUTE
C = COMPUTE
C = COMPUTE

INPUT DATA FORMAT-

(13FG.2)

INPUT DATA

MATRIX 1

	1	2	3	4	5	6	7	8	9	10
1	C.O									
2	21.280	0.0								
3	10.640	19.150	0.0							
4	12.760	19.150	93.620	0.0						
5	12.760	19.150	6.380	6.380	C.O					
6	14.890	17.020	4.250	4.250	68.080					
7	44.680	17.020	12.760	14.890	17.020	0.0				
8	8.510	8.510	4.250	4.250	6.380	0.0				
9	12.760	29.790	4.250	4.250	12.760	36.170	0.0			
10	6.380	8.510	2.130	2.130	4.250	27.660	27.660	0.0		
11	10.640	10.640	0.010	0.010	6.380	10.640	21.280	38.300	0.0	
12	12.760	14.890	17.020	19.150	6.380	6.380	51.060	27.660	25.530	0.0
13	12.760	29.790	6.380	6.380	27.660	23.400	31.910	27.660	23.400	23.400
14	8.510	2.130	4.250	6.380	21.280	10.640	14.890	59.570	10.640	10.640
15	6.380	10.640	6.380	6.380	25.530	14.890	14.890	21.280	8.510	
16	8.510	2.130	6.380	6.380	10.640	14.890	14.890	10.640	19.140	
17	10.640	4.250	4.250	4.250	21.280	25.530	27.660	14.890	10.640	
18	4.250	31.910	4.250	4.250	10.640	4.250	4.250	17.020	4.250	
19	21.280	8.510	10.640	12.760	8.510	10.640	12.760	4.250	4.250	
20	C.O	4.250	8.510	8.510	4.250	2.130	2.130	2.130	0.010	
21	2.130	4.250	6.380	6.380	4.250	2.130	2.130	2.130	0.010	
22	4.250	6.380	4.250	2.130	8.510	6.380	6.380	10.640	4.250	
23	2.130	4.250	0.010	0.010	2.130	0.010	0.010	0.010	0.010	
24	C.O	2.130	2.130	2.130	4.250	2.130	2.130	4.250	2.130	
25	2.130	6.380	8.510	8.510	4.250	0.010	4.250	4.250	4.250	
11	C.O									
12	53.130	0.0								
13	42.550	40.420	0.0							
14	5.510	10.640	12.760	0.0						
15	17.140	12.760	29.790	36.170	C.O					
16	17.020	19.150	17.020	80.850	C.O					
17	34.040	23.400	23.400	17.020	36.170					
18	8.510	10.640	17.020	17.020	10.640	0.0				
19	10.640	8.510	4.250	6.380	27.660	27.660	0.0			
20	C.O	6.380	2.130	2.130	19.150	4.250	0.0			
21	C.O	4.250	2.130	2.130	8.510	4.250	8.510	0.0		
22	4.250	4.250	12.760	12.760	2.130	4.250	2.130	19.150	0.0	
23	2.130	0.010	0.010	0.010	6.380	10.640	6.380	19.150	76.590	
24	2.130	2.130	2.130	2.130	C.O	0.010	0.010	14.890	57.450	
25	4.250	10.640	8.510	2.130	2.130	2.130	4.250	17.020	48.940	
21	21	22	23	24	25					
22	C.O									
23	53.320	0.0								
24	53.130	38.300	0.0							
25	59.570	46.810	85.110	0.0						
25	53.190	48.940	55.320	65.560	0.0					

ALSCAL MESSAGED MATRIX 1 HAS 0 MISSING OBSERVATIONS.

ITERATION HISTORY FOR THE 3 DIMENSIONAL SOLUTION
SSTRESS (IN SQUARE DISTANCES) FORMULA 1 IS USED.

ITERATION	SSTRESS	IMPROVEMENT
1	C-28347	
2	C-17150	0-11197
3	C-15458	0-01692
4	C-14788	0-00670
5	C-14483	0-00305
6	C-14320	0-00163
7	C-14222	0-00098

ALSICAL MESSAGE ITERATIONS STOPPED BECAUSE SSTRESS IMPROVEMENT LESS THAN MINIMUM PERMITTED.

STRESS AND SQAURED CORRELATION (RSC) IN DISTANCES
KRUSKALS STRESS FORMULA 1 IS USED.

STRESS = 0-111 RSC = 0-913

CONFIGURATION DERIVED IN 3 DIMENSIONS

STIMULUS COORDINATES

STIMULUS NUMBER	PLCY SYMBOL	1	2	3
1	A	0-8173	-0-9292	1-0771
2	B	0-4396	-1-4216	0-2454
3	C	-0-2766	-1-4794	1-2373
4	D	-0-2267	-1-3563	1-3547
5	E	0-7110	0-6516	1-1046
6	F	0-4846	0-9854	1-4271
7	G	1-0365	-0-5980	-0-0168
8	H	1-0593	0-6234	-1-0037
9	I	1-2840	-0-2759	-0-6237
10	J	1-2414	0-4801	-1-3139
11	K	1-1250	0-5249	-1-1564
12	L	0-7262	-0-6526	-0-7262
13	M	0-9587	0-0959	-0-3927
14	N	0-5643	1-4752	0-8387
15	O	1-1628	0-6456	0-5875
16	P	0-7405	1-2708	0-6474
17	Q	0-5522	0-0922	-0-7057
18	R	0-7170	-1-4790	-0-8224
19	S	-1-1328	-0-3903	0-6305
20	T	-2-1706	-0-2897	-0-1152
21	U	-2-1759	-0-0866	-0-1464
22	V	-1-2068	0-6599	-0-5451
23	W	-2-4122	0-4495	-0-3623
24	X	-2-1978	0-5220	-0-6235
25	Y	-1-8611	-0-3180	-0-5960

INITIAL CONFIGURATION

INITIAL STIMULUS SPACE

STIMULUS	DIMENSION		
	1	2	3
1	0.3740	C.2677	-0.4437
2	C.3528	0.1628	-0.6438
3	0.1008	0.8749	-1.4283
4	C.1202	C.8826	-1.4367
5	0.4007	0.8669	0.5547
6	0.3704	1.0284	0.7000
7	0.5832	-C.2102	-0.3573
8	C.5969	-0.8268	0.0547
9	0.6567	-0.7700	-0.0981
10	0.5143	-C.6085	0.0805
11	0.7070	-1.0488	0.1013
12	C.5896	-0.7133	-0.3550
13	0.7160	-0.6455	0.0426
14	0.4810	0.8002	1.0465
15	C.5409	C.3253	0.5659
16	0.5425	C.6318	0.9985
17	0.4580	-0.3467	0.3170
18	0.2775	C.0109	-0.1868
19	-0.2880	C.2567	-0.2714
20	-1.4152	-C.0802	0.0349
21	-1.4185	-0.1098	0.0702
22	-1.0693	-C.1309	0.2953
23	-1.4136	-C.1727	0.1760
24	-1.4882	-C.2173	0.1947
25	-1.2904	-C.2476	-0.0114

DISTANCES

MATRIX 1

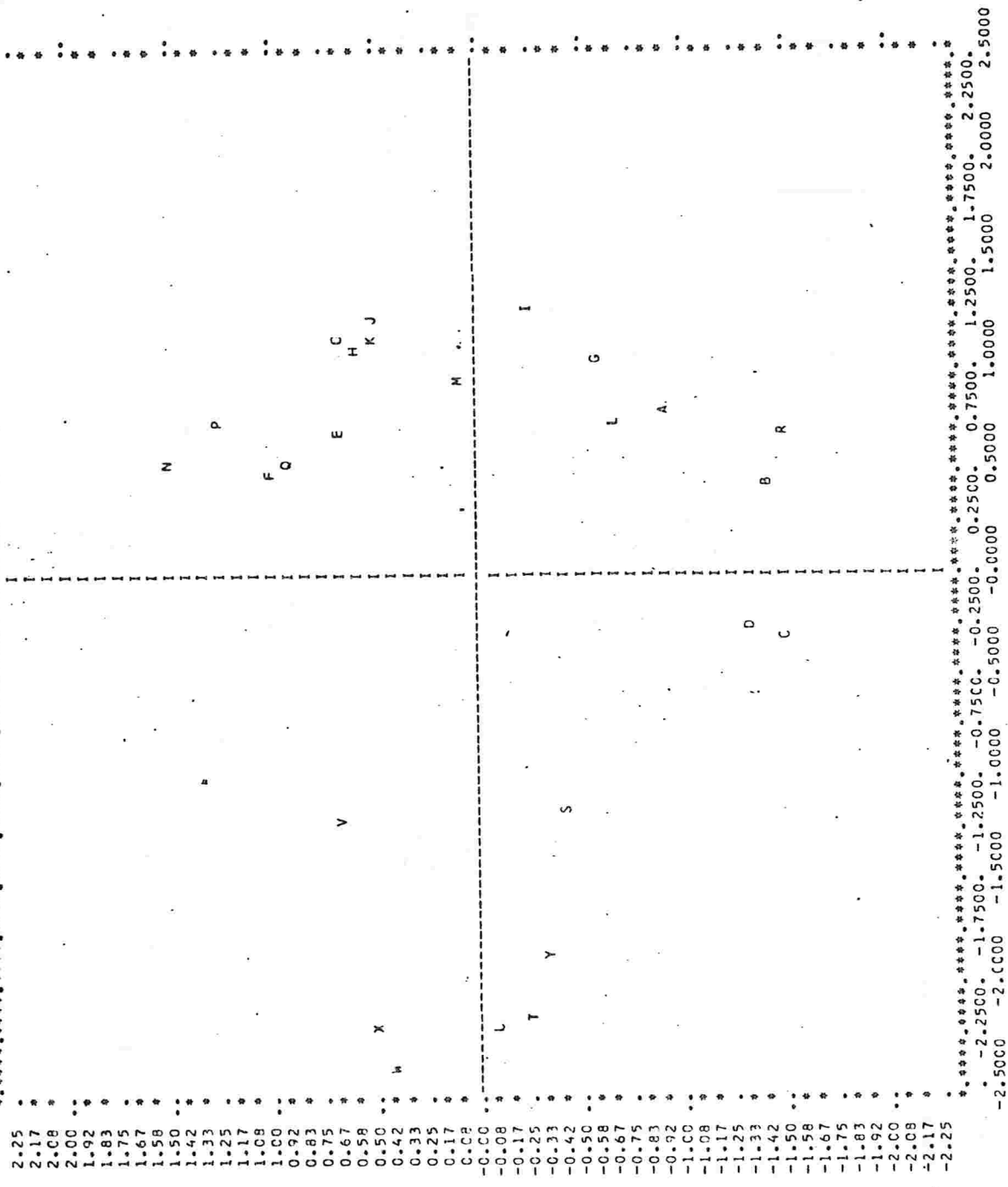
	1	2	3	4	5	6	7	8	9	10
1	C.O									
2	1.038	0.0								
3	1.235	1.225	0.0							
4	1.162	1.256	0.177	0.0						
5	1.585	2.260	2.352	2.230	C.O					
6	1.975	2.682	2.587	2.448	C.516					
7	1.164	1.050	2.018	2.013	2.144	0.0				
8	2.611	2.485	3.367	3.352	2.711	2.213	0.0			
9	1.891	1.668	2.711	2.714	2.043	2.537	1.572	0.0		
10	2.808	2.587	3.557	3.556	2.482	2.888	0.730	0.994	0.0	
11	2.683	2.495	3.422	3.416	2.302	2.701	1.602	0.370	1.024	0.0
12	1.827	1.272	2.355	2.354	2.248	2.716	0.776	0.183	0.975	0.201
13	1.798	1.726	2.582	2.563	1.616	2.081	0.793	1.358	0.681	1.376
14	2.427	2.960	3.078	2.985	0.878	0.770	2.292	0.819	0.545	1.037
15	1.685	2.217	2.555	2.555	0.687	1.131	1.388	2.099	2.332	2.466
16	2.243	2.739	2.991	2.887	0.770	0.869	2.005	1.593	1.527	1.910
17	2.562	2.504	3.176	3.148	0.770	0.869	2.005	1.809	2.074	2.173
18	1.980	1.105	2.287	2.376	1.833	2.136	1.712	0.678	1.381	1.007
19	1.919	1.919	1.512	1.510	2.170	2.268	2.273	2.145	1.365	2.087
20	3.280	2.868	2.614	2.660	3.268	3.325	3.223	2.946	1.725	3.190
21	3.342	2.962	2.732	2.769	3.232	3.272	3.255	3.509	3.492	3.697
22	3.042	2.769	2.936	2.939	2.530	2.618	2.626	3.459	3.498	3.655
23	3.795	3.465	3.292	3.314	3.457	3.447	3.621	3.352	2.662	2.572
24	3.793	3.385	3.341	3.365	3.386	3.408	3.476	3.574	3.776	3.776
25	3.217	2.687	2.687	2.749	3.232	3.361	2.968	3.320	3.572	3.508
								3.133	3.146	3.283
11	C.O									
12	1.316	0.0								
13	0.891	0.852	0.0							
14	2.240	2.646	1.891	0.0						
15	1.749	1.898	1.142	0.0						
16	1.899	2.364	1.584	1.053	C.O					
17	0.816	1.555	0.947	1.651	C.757	0.0				
18	2.072	0.832	1.650	3.393	2.589	1.418	0.0			
19	3.021	2.316	2.379	2.531	2.519	3.118	2.380	0.0		
20	3.951	2.983	3.165	3.352	3.533	2.504	2.504	0.0		
21	3.506	3.013	3.150	3.304	3.496	3.390	3.026	2.592	0.0	0.0
22	2.414	2.344	2.243	2.391	2.626	3.313	2.952	3.281	1.282	0.206
23	3.626	3.346	3.390	3.370	3.704	2.364	1.782	2.890	1.336	1.420
24	3.365	3.153	3.193	3.267	3.574	3.411	3.017	3.704	1.578	1.816
25	3.153	2.612	2.857	3.340	3.387	3.288	2.776	3.541	1.881	0.958
						3.292	2.702	2.837	1.428	0.572
21	0.0									
22	1.287	0.0								
23	0.624	1.237	0.0							
24	0.774	1.004	0.346	0.0						
25	0.596	1.178	0.973	0.905	C.O					

CISPARITIES (OPTIMALLY SCALED DATA)

MAPIX 1

	1	2	3	4	5	6	7	8	9	10
1	0.0									
2	1.103	0.0								
3	1.826	1.485	0.0							
4	1.826	1.485	0.177	0.0						
5	1.826	1.826	2.599	2.480	0.0					
6	1.026	1.826	1.826	1.826	0.516	0.0				
7	2.480	2.480	3.189	3.189	2.480	2.480	0.0			
8	1.826	1.103	2.714	2.714	1.826	1.103	1.103	0.0		
9	2.599	2.480	3.473	3.473	2.480	2.599	1.826	1.103	0.0	
10	2.480	2.480	3.473	3.473	2.480	2.599	1.826	1.103	1.026	0.0
11	1.826	1.826	3.473	3.473	2.480	2.599	1.826	1.103	1.103	1.103
12	1.826	1.103	2.582	2.582	2.480	2.599	1.826	1.103	1.103	1.103
13	2.480	3.189	3.098	2.599	1.103	2.081	1.103	1.103	0.747	1.103
14	2.480	2.217	2.599	2.599	1.103	1.103	2.292	1.826	2.392	2.466
15	2.480	3.185	2.599	2.599	1.103	1.826	1.826	1.826	1.485	2.480
16	2.480	2.599	3.176	3.176	1.103	1.103	1.826	2.074	2.074	2.173
17	2.599	1.103	2.599	2.599	1.833	1.826	1.712	1.103	1.826	1.826
18	1.485	2.480	1.826	1.826	2.480	3.189	1.826	2.599	1.826	2.599
19	3.473	2.868	2.480	2.480	2.480	2.268	1.826	2.946	2.735	3.189
20	3.473	2.962	2.599	2.599	3.185	3.325	3.223	3.473	3.473	3.697
21	3.042	2.599	2.936	2.599	3.185	3.272	3.255	3.459	3.473	3.655
22	3.473	2.599	3.189	3.189	2.480	2.599	2.599	2.480	2.480	2.599
23	3.753	3.365	3.473	3.473	3.457	3.189	3.621	3.574	3.776	3.776
24	3.217	2.599	2.480	2.480	3.189	3.408	3.473	3.189	3.473	3.508
25										
11	0.0									
12	0.951	0.0								
13	1.026	1.026	0.0							
14	2.480	2.480	1.826	0.0						
15	1.826	1.826	1.103	1.826	0.0					
16	1.826	1.826	1.826	1.826	1.026	0.0				
17	1.103	1.103	1.103	1.826	1.826	1.103	0.0			
18	2.480	1.826	1.826	3.189	1.826	3.118	1.826	0.0		
19	2.480	2.480	2.599	2.599	2.599	2.599	2.480	2.480	0.0	
20	3.551	2.599	3.189	3.392	3.473	3.189	3.189	3.202	1.485	0.0
21	3.506	3.013	3.189	3.364	3.473	3.189	3.189	3.281	1.485	0.301
22	2.599	2.599	1.826	1.626	2.599	2.364	1.826	2.480	1.826	0.951
23	3.473	3.473	3.473	3.473	3.704	3.473	3.189	3.704	1.826	0.951
24	3.365	3.193	3.267	3.267	3.473	3.288	2.776	3.473	1.826	0.951
25	3.153	2.480	2.480	3.340	3.387	3.292	2.599	3.189	1.826	0.951
21	0.0									
22	0.951	0.0								
23	0.951	1.026	0.0							
24	0.774	1.026	0.301	0.0						
25	0.951	1.026	0.951	0.747	0.0					

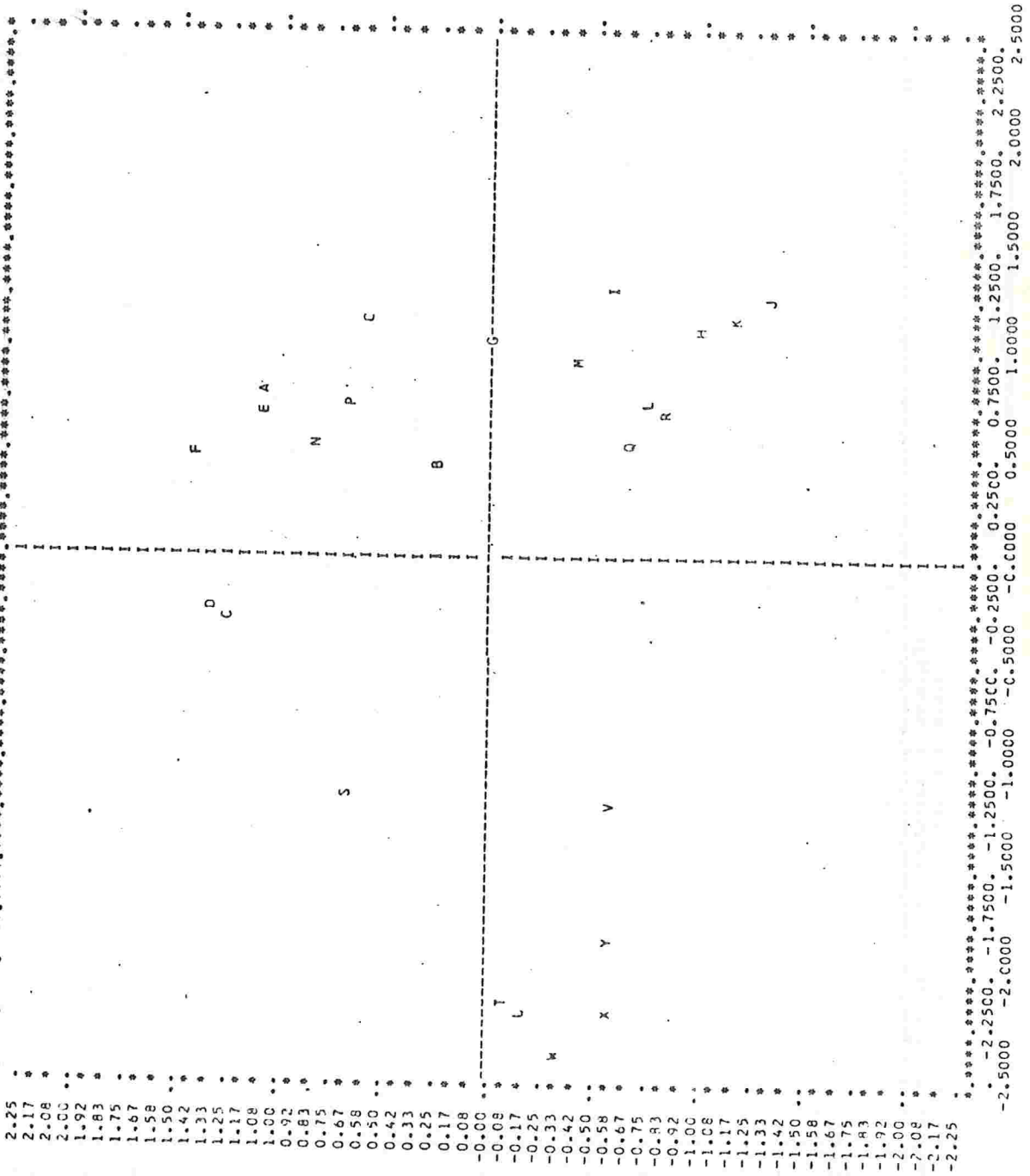
DERIVED STIMULUS CONFIGURATION CATEGORIA TRINTA E DCIS
DIMENSION 1 (HORIZONTAL) VS DIMENSION 2 (VERTICAL)



2-25
2-17 *
2-08 *
2-00 *
1-92 *
1-83 *
1-75 *
1-67 *
1-58 *
1-50 *
1-42 *
1-33 *
1-25 *
1-17 *
1-08 *
1-00 *
0-92 *
0-83 *
0-75 *
0-67 *
0-58 *
0-50 *
0-42 *
0-33 *
0-25 *
0-17 *
0-08 *
0-00 *
-0-08 *
-0-17 *
-0-25 *
-0-33 *
-0-42 *
-0-50 *
-0-58 *
-0-67 *
-0-75 *
-0-83 *
-0-92 *
-1-00 *
-1-08 *
-1-17 *
-1-25 *
-1-33 *
-1-42 *
-1-50 *
-1-58 *
-1-67 *
-1-75 *
-1-83 *
-1-92 *
-2-00 *
-2-08 *
-2-17 *
-2-25

.....
-2.5000 -2.0000 -1.5000 -1.0000 -0.5000 0.0000 0.5000 1.0000 1.5000 2.0000 2.5000
.....

DERIVED STIMULUS CONFIGURATION VS DIMENSION 3 (VERTICAL)
CATEGORIA TRINTA E DCIS
DIMENSION 1 (HORIZONTAL)

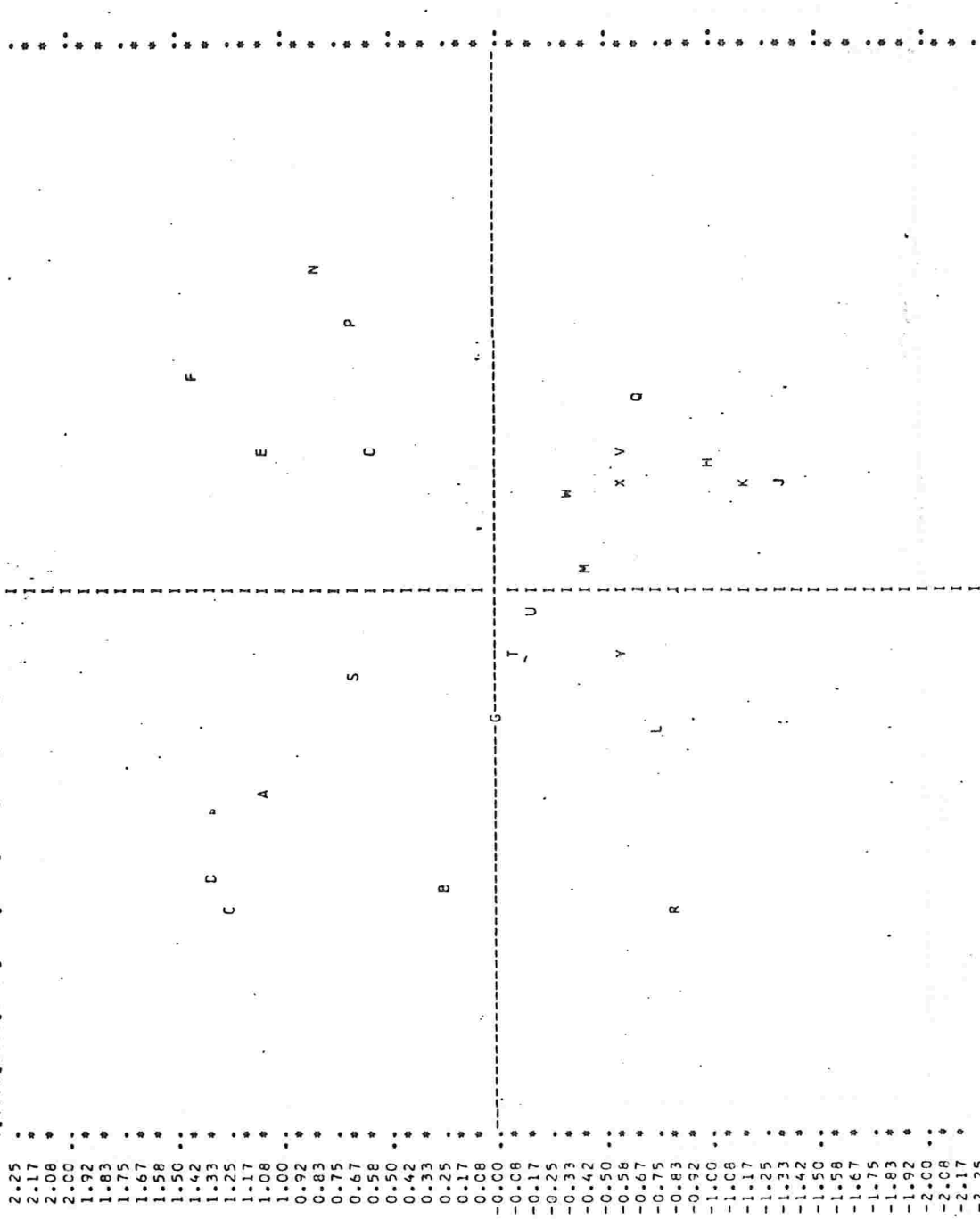


2.25 *
2.17 *
2.08 *
2.00 *
1.92 *
1.83 *
1.75 *
1.67 *
1.58 *
1.50 *
1.42 *
1.33 *
1.25 *
1.17 *
1.08 *
1.00 *
0.92 *
0.83 *
0.75 *
0.67 *
0.58 *
0.50 *
0.42 *
0.33 *
0.25 *
0.17 *
0.08 *
-0.00 *
-0.08 *
-0.17 *
-0.25 *
-0.33 *
-0.42 *
-0.50 *
-0.58 *
-0.67 *
-0.75 *
-0.83 *
-0.92 *
-1.00 *
-1.08 *
-1.17 *
-1.25 *
-1.33 *
-1.42 *
-1.50 *
-1.58 *
-1.67 *
-1.75 *
-1.83 *
-1.92 *
-2.00 *
-2.08 *
-2.17 *
-2.25 *

-2.5000 -2.0000 -1.5000 -1.0000 -0.5000 -C.0000 0.5000 1.0000 1.5000 2.0000 2.5000

-2.5000 -1.7500 -1.2500 -0.7500 -0.2500 0.2500 0.7500 1.2500 1.7500 2.2500

DERIVED STIMULUS CONFIGURATION VS DIMENSION 3 (VERTICAL)
CATEGORIA TRINTA E DCIS



.....
-2.2500. -1.7500. -1.2500. -0.7500. -0.2500. 0.2500. 0.7500. 1.2500. 1.7500. 2.2500.
-2.5000 -2.0000 -1.5000 -1.0000 -0.5000 0.5000 1.0000 1.5000 2.0000 2.5000

BIBLIOGRAFIA

1. Banks, A.S. and Gregg, P.M. (1965) - Grouping political systems: Q-factor analysis of a cross-polity survey- Amer. Behavioral Scientist, 9, pg 3-6.
2. Barlow, R.E. (1972) - Statistical Inference under order restrictions - Wiley.
3. Bloxom, B. (1968) - Individual differences in multidimensional scaling - Research Bulletin pg 68-45 - Princeton, N.J., Educational Testing Service.
4. Bloxm, B. (1974) - An alternative method of fitting a model of individual differences in multidimensional scaling - Psychometrika, 39, pg 365-367.
5. Burton, M. (1972) - Semantic Dimensions of Occupation Names - pg 55-71 in R.N.Shepard, A.K. Romney and S. Nerlove (Eds.), Multidimensional Scaling (vol 2).New York, Seminar Press.
6. Carroll, D.J. and Chang, J.J. (1970)- Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition - Psychometrika, 35, pg 238-319.

7. Carroll, J.D. and Chang, J.J. (1974) - Some methodological advances in INDSCAL - Artigo apresentado no encontro da primavera da Psychometric Society, Stanford, California.
8. Coombs, C.H. (1958) - An application of a nonmetric model for multidimensional analysis of similarities - Psychological Reports, 4, pg 511-518.
9. Coombs, C.H. (1964) - A theory of data - New York, Wiley.
10. De Leeuw, J. (1975) - Canonical discriminant analysis of relational data - Research Bulletin RB004-75 - Leiden, The Netherlands: Data theory, University of Leiden.
11. De Leeuw, J., Young, F.W. and Takane, Y. (1976) - Additive structure in qualitative data: An alternating least squares method with optimal scaling features - Psychometrika, 41, pg 471-503.
12. Engelmann, A. (1978) - Os Estados Subjetivos: Uma Tentativa de Classificação de seus Relatos Verbais - São Paulo, Ática .
13. Fernandez, P. y Yohai, V. (1980) - Analisis de Datos Multivariados - IMPA - Rio de Janeiro.

14. Fisher, R.A. (1946) - Statistical methods for research workers - Edinburgh, Oliver and Boyd.
15. Funk, S.G., Horowitz, A.D., Lipshitz, R. and Young, F. W. (1976) - The Perceived Structure of American Ethnic Groups: The Use of Multidimensional Scaling in Stereotype Research - Sociometry, 39, 2, pg 116-130.
16. Gnanadesikan R. (1977) - Methods for statistical data analysis of multivariate observations - New York , Wiley.
17. Graef, J. and Spence, I. (1976) - Using prior distance information in multidimensional scaling - artigo apresentado na reunião conjunta da Psychometric Society e do Mathematical Psychology Group - Bell Laboratories, Murray Hill, N.J.
18. Green, P.E., Carmone, F.J. and Robinson, P.J. (1968) - Tricon: An IBM 360/65 Fortran IV Program for the Triangularization of Conjoint Similarities Data-Journal of Marketing Research, 5, pg 219-220.
19. Green, P.E., Carmone, F.J. and Fox, B. (1969) - Television programme similarities: and application of

- subjective clustering - Journal of the Market Research, 2, 1.
20. Green, P.E. and Maheshwari A. (1969) - Common stock perception and preference: An application of Multi-dimensional Scaling - Journal of Business, 42, pg 439-457.
 21. Green, P.E. and Carmone, F.J. (1972) - Marketing Research Applications of Nonmetric Scaling Methods - pg 185-210 in R.N.Shepard, A.K. Romney and S.Nerlove (Eds.), Multidimensional Scaling (vol 2), New York, Seminar Press.
 22. Guttman, L. (1966) - Order analysis of correlation matrices - pg 438-458 in R.B.Cattell (Ed.), Handbook of Multivariate Experimental Psychology, Chicago, Rand McNally.
 23. Guttman, L. (1968) - A general nonmetric technique for finding the smallest coordinate space for a configuration of points - Psychometrika, 33, pg 469-506.
 24. Guttman, L. (1969) - Smallest space analysis by the absolute value principle - artigo apresentado no simpósio sobre "Teoria e Prática da Mensuração" no XIX

Congresso Internacional de Psicologia, Londres.

25. Guttman, L. (1971) - Measurement as a structural theory
Psychometrika, 36, pg 329-347.
26. Hodson, F.R., Sneath, P.H.A. and Doran, J.E. (1966) Some
experiments in the numerical analysis of archaeological
data - Biometrika, 53, pg 311-324.
27. Horan, C.B. (1969) - Multidimensional Scaling: Combining
observations when individuals have different percep-
tual structures - Psychometrika, 34, pg 139-165.
28. Johnson, R.M. (1973) - Pairwise nonmetric multidimen-
sional scaling - Psychometrika, 38, pg 11-18.
29. Johnson, S.C. (1967) - Hierarchical clustering schemes -
Psychometrika, 32, pg 241-254.
30. Klingberg, F.L. (1941) - Studies in measurement of the
relations among sovereign states - Psychometrika, 6,
pg 335-352.
31. Kruskal, J.B. (1964a) - Multidimensional Scaling by
optimizing goodness of fit to a nonmetric hypothesis
- Psychometrika, 29, pg 1-27.

32. Kruskal, J.B. (1964b) - Nonmetric multidimensional scaling: a numerical method - *Psychometrika*, 29, pg 115-129.
33. Kruskal, J.B. (1971) - Monotone Regression: Continuity and Differentiability Properties - *Psychometrika*, 36, 1, pg 57-62.
34. Kruskal, J.B. and Carroll, J.D. (1969) - Geometric models and badness-of-fit functions - pg 639-670 in P. R. Krishnaiah (Ed) - *Multivariate Analysis*, (vol 2), New York, Academic Press.
35. Kruskal, J.B. and Wish, M. (1978) - *Multidimensional Scaling* - California, Sage.
36. Mac Callum, R.C., Cornelius III, E.T., (1977) - A Monte Carlo Investigation of Recovery of Structure by ALSCAL - *Psychometrika*, 42, pg 401-428.
37. Mardia, K.V. (1978) - Some properties of classical Mul dimensional Scaling - *Comm. Statist. - Theor. Meth A*, 7, pg 1233-1241.
38. Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979) - Mul tivariate Analysis, London, Academic Press.

39. Mauser, G.A. (1972) - A Structural Approach to Predicting Patterns of Electoral Substitution - pg 249-289 in R.N.Shepard, A.K. Romney and S.Nerlove (Eds), Multi-dimensional Scaling (vol 2), New York, Seminar Press.
40. Mc Gee, V.C. (1968) - Multidimensional scaling of n sets of similarity measures: A nonmetric individual differences approach - Multivariate Behavioral Research, 3, pg 233-248.
41. Messick, S.J. and Abelson, R.P. (1956) - The additive constant problem in Multidimensional scaling - Psychometrika, 21, pg 1-15.
42. Mosteller, F. and Tukey, J.W. (1977) - Data analysis and regression: a second course in statistics-Reading, Mass : Addison Wesley.
43. Neidell, L.A. (1969) - The use of Nonmetric Multidimensional Scaling in Marketing Analysis - Journal of Marketing, 33, pg 37-43.
44. Press, S.J. (1972) - Applied multivariate analysis - New York, Holt.
45. Richardson, M.W. (1938) - Multidimensional psychophysics - Psychological Bulletin, 35, pg 659-660.

46. Roscoe, J.T. (1965) - Fundamental research statistics for the behavioral sciences - Holt, Rinehart and Winston, Inc.
47. Rosemberg, S., Nelson, C. and Vivekanauthan, P.S. (1968) - A multidimensional approach to the structure of personality impressions - J. of Personality and Social Psychology, 9, pg 283-294.
48. Rothkopf, E.Z. (1957) - A measure of stimulus similarity and erros in some paired - associate learning tasks - J. of Experimental Psychology, 53, pg 94-101.
49. Rummell, R.J. (1969) - Some empirical findings on nations and their behavior - World Politics, 21, pg 226-241.
50. Sawyer, J. (1967) - Dimensions of nations: size, wealth, and politcs - Amer. J. of Sociology, 73, pg 145-172.
51. Schönemann, P.H. (1972) - An algebraic solution for a class of subjective metric models - Psychometrika , 37, pg 441-451.
52. Shepard, R.N. (1962) - The analysis of proximities: mul tidimensional scaling with an unknow distance func tion - Psychometrika, 27, pg 125-140, 219-246.

53. Shepard, R.N. (1963) - Analysis of proximities as a technique for the study of information processing in man - Human Factors, 5, pg 33-48.
54. Shepard, R.N. (1974) - Psychological representation of speech sounds in E.E. David and P.B. Denes (Eds) Human Communication: A Unified View - New York, Mc Graw-Hill.
55. Shepard, R.N., Romney, A.K. and Nerlove, S.B. (Eds) (1972) - Multidimensional Scaling: Theory and Applications in the Behavioral Sciences - New York, Seminar Press.
56. Sibson, R. (1978) - Studies in the robustness of multi dimensional scaling: Procrustes statistics - J. R. Statist. Soc. B, 40, pg 234-238.
57. Spang, H.A. (1962) - A review of minimization techniques for nonlinear functions - SIAM Rev., 4, pg 343-365.
58. Spence, I. and Young, F.W. (1978) - Monte Carlo Studies in Nonmetric Scaling - Psychometrika, 43, 1, pg 115-117.
59. Takane, Y., Young F.W. and De Leeuw (1976) - Nonmetric Individual Differences Multidimensional Scaling: An Alternating Least Squares Method with Optimal Scaling

Features - Psychometrika, 42, pg 6-67.

60. Takane, Y., Young, F.W. and De Leeuw J. (1979) - Nonmetric Common Factor Analysis: An Alternating Least Squares Method with Optimal Scaling Features - Behaviormetrika, 6, pg 45-56.
61. Torgerson, W.S. (1962) - Theory and Methods of Scaling - New York, Wiley.
62. Tucker, L.R. (1964) - The extension of factor analysis to three dimensional matrices - pg 109-127 in N. Fredriksen, and H. Gulliksen (Eds) Contribution to Mathematical Psychology - New York, Holt, Rinehart and Winston.
63. Tucker, L.R. (1972) - Relations between multidimensional scaling and three-mode factor analysis - Psychometrika, 37, pg 3-27.
64. Weisberg, H.F. and Rusk, J.G. (1970) - Dimensions of candidate evaluation - Amer. Political Science Review, 64, pg 1167-1185.
65. Wish, M. (1967) - A model for the perception of Morse code like signals - Human Factors, 9, pg 529-539.

66. Wish, M. (1971) - Individual differences in perceptions and preferences among nations - pg 312-328 in C. W. King and D. Tigert (Eds). Attitude Research Reaches New Heights, Chicago: American Marketing Association.
67. Wish, M. and Carroll, J.D. (1973) - Concepts and Applications of Multidimensional Scaling - Sensory Evaluations of Appearance of Materials, ASTM, STP 545, American Society for Testing and Materials, pg 91-108.
68. Wish, M. and Carroll, J.D. (1974) - Applications of individual differences scaling to studies of human perception and judgment - pg 449-491 in E.C. Carterette and M.P. Friedman (Eds) Handbook of Perception (vol 2), New York, Academic Press.
69. Wish, M., Deutsch, M. and Biener, L. (1970) - Differences in conceptual structures of nations: an exploratory study - J of Personality and Social Psychology, 16 , pg 361-373.
70. Yates, A. (1972) - Nonmetric individual - differences multidimensional scaling with balanced least squares monotone regression - artigo apresentado no encontro da Primavera da Psychometric Society - Princeton, N. J.

71. Young, F.W. (1970) - Nonmetric Multidimensional Scaling, Recovery of Metric Information - Psychometrika, 35, pg 455-474.
72. Young, F.W. (1970) - Nonmetric scaling of line lengths using latencies, similarity and same-different judgments - Perception and Psychophysics, vol 8 (5B).
73. Young, F.W. (1975) - Methods for describing ordinal data with cardinal models - Journal of Mathematical Psychology, 12, pg 416-436 (a).
74. Young, F.W. (1975) - Scaling replicated conditional rank order data - Sociological Methodology, 12, pg 129-170 (b).
75. Young, F.W., de Leeuw, J. and Takane, Y. (1976) - Regression with qualitative and quantitative variables An alternating least squares method with optimal scaling features - Psychometrika, 41, pg 505-529.
76. Young, F.W., Takane, Y. and Lewyckyj, R. (1978) - Three Notes on ALSCAL - Psychometrika, 43, pg 433-435.
77. Young, F.W. and Torgerson, W.S. (1967) - TORSCA, a Fortran IV program for Shepard - Kruskal multidimensional scaling analysis - Behavioral Sciences, 12, pg

498-503.

78. Young, G. and Householder, A.S. (1938) - Discussion of a set of points in terms of their mutual distances - *Psychometrika*, 3, pg 19-22.