

PROPOSTA PARA GERAÇÃO DE AGRUPAMENTOS NA SOLUÇÃO DO MDS NÃO MÉTRICO DE 3-VIAS

MOACYR MACHADO CARDOSO JUNIOR

Instituto Tecnológico de Aeronáutica – ITA
Praça Mal. Eduardo Gomes, 50 – São José dos Campos
moacyr@ita.br

RODRIGO ARNALDO SCARPEL

Instituto Tecnológico de Aeronáutica – ITA
Praça Mal. Eduardo Gomes, 50 – São José dos Campos
rodrigo@ita.br

RESUMO

A geração do mapa perceptual via escalonamento multidimensional de 3 vias possibilita ao analista a visualização da separação dos objetos no espaço Euclidiano. O método MDSvarext incorpora a esta análise as regiões de confiança dos objetos permitindo assim realizar inferência estatística quanto à real diferença entre objetos, porém as regiões de confiança geradas são muito grandes devido à variabilidade inerente ao grupo de avaliadores. Uma solução para este problema é a geração de agrupamentos antes da realização do MDSvarext, visando a obtenção de subgrupos homogêneos e com maior controle da variância. Neste trabalho pesquisou-se a possibilidade da utilização do algoritmo *K-médias* para geração dos subgrupos antes do MDSvarext, avaliando o processo com dois indicadores de qualidade, um *Ex-ante* e outro *Ex-post*. Os resultados mostraram-se promissores quanto às possíveis interpretações do mapa perceptual e para os indicadores gerados.

PALAVRAS CHAVE. Agrupamentos não hierárquicos, Escalonamento Multidimensional, Mapa Perceptual.

AO – Outras aplicações em PO

ABSTRACT

The generation of the perceptual map via 3-way multidimensional scaling allows the analyst to visualize the separation of objects in a Euclidean space. The MDSvarext method incorporates to this analysis, the confidence regions of the objects allowing statistical inference about the difference between objects, but the confidence regions generated are very large due to the inherent variability of evaluators. A solution to this problem is the cluster generation prior to the MDSvarext in order to obtain homogeneous subgroups and with greater control of the variance. This work investigated the possibility of using K-means algorithm for generation of subgroups before MDSvarext evaluating the process with two quality indicators, an *Ex-ante* and one *Ex-Post*. The results obtained are promising as the possible interpretations of the perceptual map and indicators generated.

KEYWORDS. Multidimensional Scaling, Non-hierarchical clusters, Perceptual Map.

OA - Other applications in OR

1. Introdução

A solução do NMDS (MDS não métrico) de 3 vias, utilizando o algoritmo proposto por Cardoso-Junior e Scarpel (2011), denominado MDSvarext, apresenta como resposta visual um mapa perceptual bidimensional com a representação das regiões de confiança para cada objeto. Desta forma é possível identificar e analisar quais objetos são diferentes entre si, utilizando para tanto a estimação pontual via configuração de consenso obtida pela Análise de Procrustes Generalizada – GPA e a estimativa das regiões de confiança utilizando uma abordagem não paramétrica, o *Jackknife*.

No entanto, esta abordagem resulta em regiões de confiança muito grandes, tornando-se praticamente impossível separar os diferentes objetos estatisticamente.

A proposta para melhoria envolveria a separação dos julgadores em agrupamentos mais homogêneos, de forma que se possa gerar mais de um mapa perceptual, com um controle maior da variabilidade intrínseca ao grupo de julgadores.

O presente estudo tem por objetivos: i) propor um método para obtenção do mapa perceptual de um grupo, com separação em agrupamentos homogêneos, ii) testar a utilização de indicadores de qualidade dos agrupamentos gerados, iii) Verificar o comportamento do método utilizando simulação de Monte Carlo.

A contribuição esperada do trabalho é a de produzir um mapa perceptual consensual de grupo com a separação em subgrupos homogêneos e representação das regiões de confiança.

Este trabalho encontra-se organizado da seguinte forma: na seção de introdução, foi abordada a motivação e os objetivos para desenvolvimento deste trabalho; a Seção 2 uma breve descrição do método MDSvarext. A Seção 3 apresenta a proposta para geração dos agrupamentos no contexto do MDSvarext e para obtenção dos indicadores de qualidade desses agrupamentos; a Seção 4 apresenta os resultados obtidos com um conjunto de dados simulados; e finalmente a Seção 5 com as conclusões finais.

2. Método MDSvarext – MDS Variabilidade Externa

O modelo MDSvarext busca a representação do mapa perceptual de 3 vias, ou seja, incorporar ao mapa perceptual obtido via Escalonamento Multidimensional não métrico a variabilidade inerente a um grupo de avaliadores. Os mapas perceptuais tem inúmeras aplicações nas áreas de marketing, ciências do comportamento, econometria, ciências sociais e políticas e na área de percepção de riscos (SLOVIC, 2001; MOREIRA, 2006; VANLAAR; YANNIS, 2006).

O MDSvarext se processa em 3 fases: Redução de Dimensão, Alinhamento das configurações e Obtenção das regiões de confiança utilizando métodos de reamostragem.

Na primeira fase, a partir das matrizes de dissimilaridade D individuais é aplicado o algoritmo de solução SMACOF, proposto por De Leeuw (1977) e implementado por De Leeuw e Mair (2009) em um pacote denominado SMACOF do software estatístico R. Uma série de configurações em baixa dimensão são obtidas nesta fase.

Na segunda fase, a partir das configurações finais individuais, obtidas de cada julgador, que são invariantes à rotação, reflexão e translação, aplica-se a Análise Procrustes Generalizada – GPA sobre as mesmas de forma a obter o alinhamento das diferentes configurações e a obtenção de uma configuração de consenso por meio das transformações rígidas admissíveis das configurações. As configurações finais são alinhadas segundo um critério de minimização dos erros das coordenadas.

A GPA segundo Brombin e Salmaso (2009) é uma ferramenta da estatística da forma. O termo forma é definido pelos autores associando as propriedades geométricas de uma configuração de pontos que são invariantes a mudanças de translação, rotação e escala. A análise direta de um conjunto de pontos não é conveniente devido à presença de erros sistemáticos tais como posição, orientação e tamanho, e usualmente para que se possa conduzir uma análise

estatística da forma confiável normalmente a GPA é utilizada para eliminar os fatores não relativos à forma e para alinhar as configurações para um sistema de coordenadas comum.

A GPA é uma técnica estatística multivariada empírica na qual três dimensões estão envolvidas: os objetos de estudo, as pessoas que avaliam os objetos e os atributos nos quais os objetos são avaliados. A GPA é ideal para analisar dados oriundos de diferentes indivíduos (DIJKSTERHUIS; GOWER, 2010).

As transformações permitidas na GPA são translação, rotação/reflexão e escalonamento isotrópico, de forma que as distâncias relativas entre os objetos permaneçam inalterada (RODRIGUE, 1999). A GPA foi realizada utilizando-se o software estatístico R e o pacote SHAPES de autoria de Dryden (2009).

Finalmente na última fase são obtidas as regiões de confiança das coordenadas. Após a GPA obtêm-se as configurações alinhadas, cujas coordenadas são então submetidas a uma estratégia de reamostragem (*randomization tests*), uma vez que não existe garantia de que as coordenadas sigam uma determinada distribuição probabilística (ie.: distribuição Normal), especialmente por se tratar de um caso multivariado. Várias estratégias foram testadas no contexto do MDS, tais como *Jackknife*, *Bootstrap* e Teste de Permuta com resultados satisfatórios (WEINBERG; CARROLL; COHEN, 1984; SABURI; CHINO, 2008; JACOBY, 2009, ABDI; DUNLOP; WILLIAMS, 2009).

3. Método

A partir do método descrito por Cardoso-Junior e Scarpel (2011), a idéia básica desta proposta é a de gerar agrupamentos homogêneos antes do processo de redução de dimensão (via MDS não métrico) e do alinhamento das diferentes configurações via GPA.

Os agrupamentos serão gerados com a utilização do método não hierárquico *K-médias*, que busca maximizar a distância entre agrupamentos diferentes e minimizar as distâncias intra-agrupamento.

A formulação, por programação matemática, do algoritmo *K-médias*, conforme Webb (2002), é

$$\begin{aligned} \text{Min} \quad & \sum_{i=1}^N \sum_{c=1}^K Z_{ic} \left[\sum_{j=1}^P (x_{ij} - m_{cj})^2 \right]^{\frac{1}{2}} \\ \text{S. A.} \quad & \sum_{c=1}^K Z_{ic} = 1, \quad i = 1, \dots, N \end{aligned} \quad (1)$$

$$Z_{ic} = \begin{cases} 1, & \text{se o ponto } i \text{ pertencer ao agrupamento } c \\ 0, & \text{caso contrário} \end{cases}$$

Em que:

$$m_{cj} = \frac{\sum_{i=1}^N Z_{ic} x_{ij}}{\sum_{i=1}^N Z_{ic}} \quad c = 1, \dots, K \text{ e } j = 1, \dots, P$$

sendo m_{cj} o centróide do agrupamento c na dimensão j .

A medida de validação desses agrupamentos gerados foi realizada com uma métrica de validação interna de agrupamentos proposta por Brock *et al.* (2008), denominado de conectividade, gerando assim um indicador *Ex-ante* para a qualidade dos agrupamentos gerados.

A conectividade refere-se à extensão na qual uma instância está alocada a um mesmo agrupamento dos seus vizinhos mais próximos. (EVERITT *et al.*, 2001), sendo definida por:

$$Con(C) = \sum_{i=1}^N \sum_{j=1}^M x_i nn_{i(j)} \quad (2)$$

onde N representa o número total de observações e M o número de vizinhos considerados, e $nn_{i(j)}$ é o *j*-ésimo vizinho mais próximo da instância *i* na dimensão *j*, de forma que $x_i nn_{i(j)} = 0$ se *i* e *j* estiverem no mesmo agrupamento e 1/*j* caso contrário.

A conectividade tem valores entre $0 \leq Con(C) \leq \infty$, e é uma métrica que deve ser minimizada, ou seja, quanto menor o valor, melhor será a estrutura proposta pelo algoritmo, conforme Everitt *et al.* (2001).

A métrica de validação interna, conectividade, utilizada neste trabalho não é absoluta no que se refere à definição do número ideal de agrupamentos a serem gerados conforme afirma Everitt *et al.* (2001). Desta forma, as soluções finais obtidas para cada agrupamento são testadas em relação à efetiva separação de seus centróides, utilizando o teste de hipótese de *Hottelling* – T^2 robusto proposto por Willems *et. al* (2002) e implementado no pacote “RRCOV” do software R (TODOROV; FILZMOSER, 2009). A razão de separação, representada por H_a/H_0 gera um segundo indicador de qualidade da separação dos agrupamentos, desta vez *Ex-post*.

As Hipóteses nula e alternativa estão representadas em (3).

$$\begin{aligned} H_0: \hat{\Theta}_{ic} &= \hat{\Theta}_{id}, \text{ para } i = 1, \dots, 10 \text{ e } c \neq d \\ H_a: \hat{\Theta}_{ic} &\neq \hat{\Theta}_{id} \end{aligned} \quad (3)$$

Onde $\hat{\Theta}_{ic}$, é o centróide do objeto *i* no agrupamento *c*, e $\hat{\Theta}_{id}$ é o centróide do objeto *i* no agrupamento *d*.

Após a separação dos julgadores em agrupamentos, utiliza-se o algoritmo MDSvarext para obtenção das soluções de consenso para cada grupo. Com a utilização de um método não paramétrico de estimação das matrizes de Variância-Covariância, pode-se gerar e representar as regiões de confiança para cada agrupamento gerado. O método utilizado foi o *Bootstrap*, o qual se justifica pelo fato de não ser necessário fazer hipóteses sobre a distribuição de probabilidades das coordenadas obtidas via MDSvarext.

O *Bootstrap* está baseado em computação intensiva em substituição à análise teórica, podendo dar respostas a problemas muito complexos para as abordagens tradicionais e mesmo para problemas mais simples, que devido ao forte declínio do custo computacional passa a ser uma opção adequada (EFRON; TIBSHIRANI, 1986).

A estratégia do *Bootstrap* é implementada pela construção de B amostras aleatórias de tamanho igual ao conjunto original com reposição. O algoritmo de Monte Carlo se processa em três fases:

- a. Utilizando um gerador de números aleatórios, construa um grande número de amostras *Bootstrap*, de forma independente, denominadas $y^*(1), y^*(2), \dots, y^*(B)$;
- b. Para cada amostra *Bootstrap* $y^*(b)$, avalie a estatística de interesse, como $\hat{\Theta}(B) = \hat{\Theta}(y^*(b)), b = 1, 2, \dots, B$;
- c. Calcule o desvio padrão de $\hat{\Theta}^*(b)$

Neste trabalho utilizou-se $B=10.000$, visando garantir a convergência do real valor do desvio padrão.

Finalmente para testar a proposta de incorporação dos agrupamentos ao algoritmo MDSvarext, realizou-se um estudo com dados simulados de um grupo de 10 julgadores, avaliando 10 objetos distintos utilizando 10 características. Dessa forma, obteve-se um vetor com dimensão $10 \times 10 \times 10$.

A partir de uma matriz conhecida de julgamento (10x10), com notas variando de 1 a 7 para cada atributo (coluna), e para cada objeto (linha) gerou-se 10 matrizes aleatórias utilizando distribuição Binomial, em que o $Valor(i, j, k) \sim Bin(MAT_{ij}, p)$, onde $Valor(i, j, k)$ representa cada nota atribuída ao objeto i , na característica j , pelo julgador k , a matriz MAT_{ij} é a matriz geratriz, e p é a probabilidade do valor de MAT_{ij} ser o escolhido. Na Figura 1 está representada a matriz geratriz MAT_{ij} .

$$MAT_{ij} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 6 & 5 & 4 \\ 2 & 3 & 4 & 5 & 6 & 7 & 6 & 7 & 6 & 5 \\ 3 & 4 & 5 & 6 & 7 & 6 & 5 & 6 & 7 & 6 \\ 4 & 5 & 6 & 7 & 6 & 5 & 4 & 5 & 6 & 7 \\ 5 & 6 & 7 & 6 & 5 & 4 & 3 & 4 & 5 & 6 \\ 6 & 7 & 6 & 5 & 4 & 3 & 2 & 3 & 4 & 5 \\ 7 & 6 & 5 & 4 & 3 & 2 & 1 & 2 & 3 & 4 \\ 6 & 5 & 4 & 3 & 2 & 1 & 2 & 1 & 2 & 3 \\ 5 & 4 & 3 & 2 & 1 & 2 & 3 & 2 & 1 & 2 \\ 4 & 3 & 2 & 1 & 2 & 3 & 4 & 3 & 2 & 1 \end{bmatrix}$$

FIGURA 1 - Matriz Geratriz MAT_{ij} .

A solução da matriz geratriz utilizando o MDS não métrico via SMACOF resulta para duas dimensões no padrão apresentado na Figura 2.

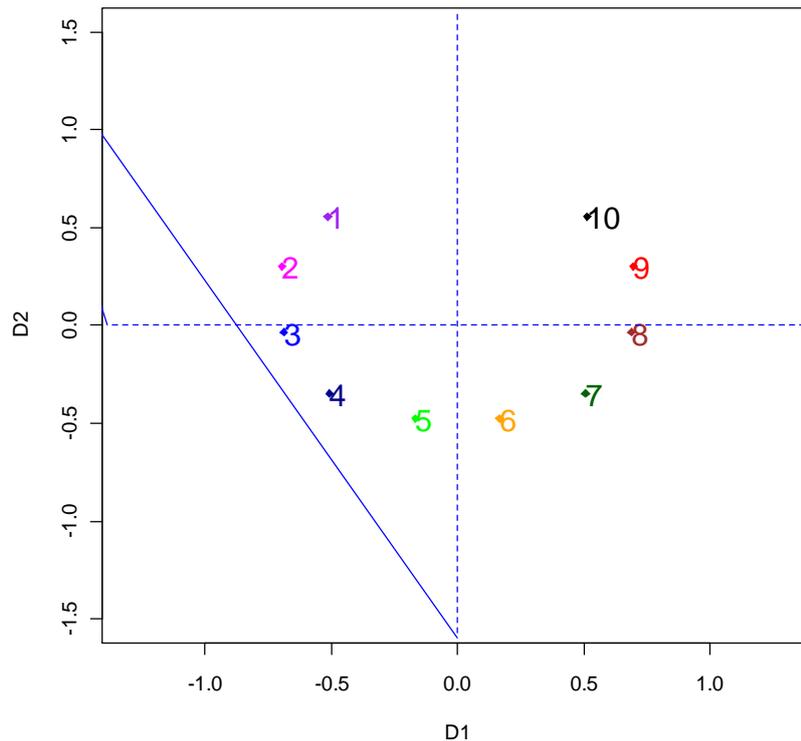


FIGURA 2 - Solução MDS não métrico via SMACOF para a matriz geratriz MAT_{ij} .

O valor de probabilidade p , neste caso, será utilizado com o propósito da geração de um grupo homogêneo com p variando de 0,9 à 0,95 e grupos mais heterogêneos com p variando de 0,50 até 0,75.

4. Resultados

Os resultados serão divididos em duas partes, sendo a primeira uma análise do padrão do mapa perceptual obtido considerando todos os 10 julgadores simultaneamente, e comparando ao resultado da solução com agrupamentos. Na segunda parte serão analisados os indicadores *Ex-ante* e *Ex-post* gerados. Fixou-se para ambos os casos a análise com apenas 2 e 3 agrupamentos.

A primeira simulação foi realizada com o valor de $p=0,95$, para simular uma condição de homogeneidade entre as diferentes matrizes de julgamento. Na Figura 3 apresenta-se a solução considerando os 10 julgadores simultaneamente (3-A) e com a geração de agrupamentos (3-B). Verifica-se na Figura 3-B que a separação em dois agrupamentos destaca a diferença entre os dois grupos, produzindo regiões de confiança menores para os prováveis posicionamentos das configurações de consenso. Na Figura 4 apresenta-se o resultado da separação em 3 agrupamentos, com a mesma tendência do caso anterior. Com o objetivo de verificar qual das configurações é mais adequada gerou-se 100 simulações das matrizes de julgamento tanto para 2 agrupamentos como para 3, e verificou-se o indicador *Ex-ante* para validação do número de agrupamentos com a métrica conectividade. Posteriormente com o indicador *Ex-post*, com a métrica proporção de separação dos centróides dos objetos individuais utilizando o teste de hipótese *Hotteling* T^2 para $\alpha=0,05$.

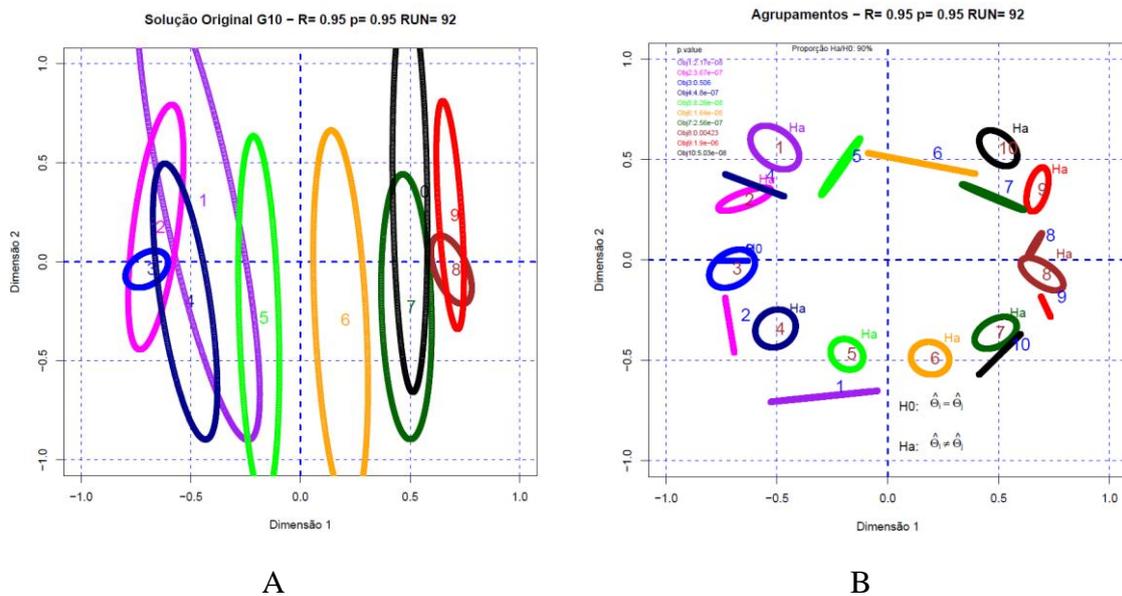


FIGURA 3 - Solução original (A) e com 2 agrupamentos (B) gerados pelo *K-médias*, para $p=0,95$ e região de confiança = 0,95.

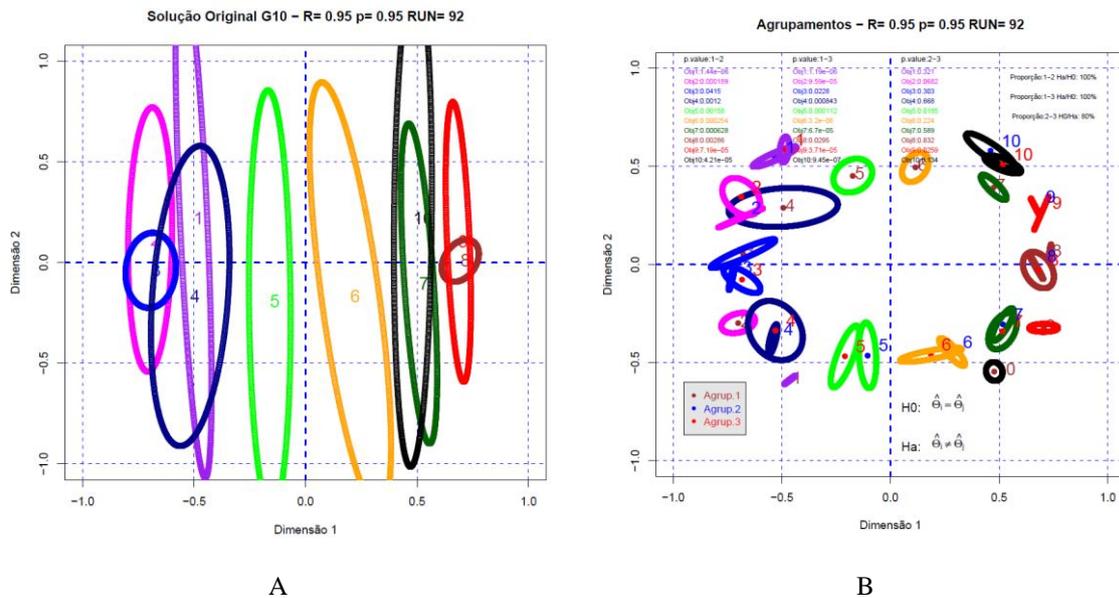


FIGURA 4 - Solução original (A) e com 3 agrupamentos (B) gerados pelo K-médias, para $p=0,95$ e região de confiança = 0,95.

Na Figura 5 pode-se verificar o resultado da simulação.

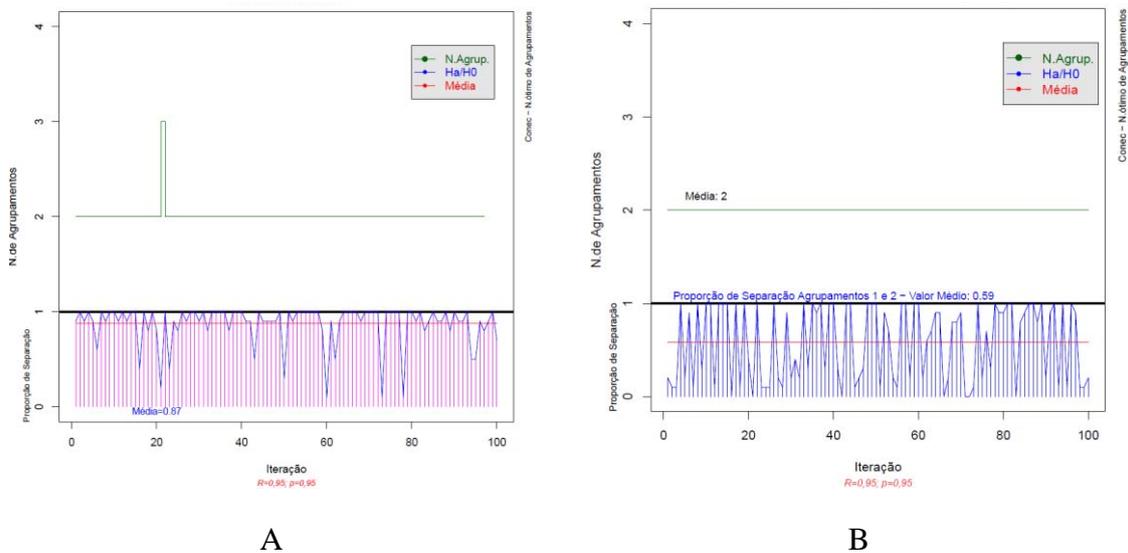


FIGURA 5 - Resultado da simulação para $p=0,95$ do indicador *Ex-ante* conectividade e *Ex-Post* proporção de separação para 2 agrupamentos (A) e para 3 agrupamentos (B).

Verifica-se que na simulação com 2 agrupamentos a conectividade indicou em 99% dos casos 2 agrupamentos como ideal, e a proporção de separação foi de 0,87. No caso da simulação com 3 agrupamentos, novamente o indicador demonstrou que o número ideal de agrupamentos seria 2, e a proporção de separação entre as diferentes combinações dos 3 agrupamentos foi estatisticamente menor teste t ($\alpha=0,05$ de significância, $p\text{-value} = 1,054 \times 10^{-9}$ para Grupo de 3-1.2 e Grupo de 2, $p\text{-value} = 4,325 \times 10^{-9}$ para Grupo de 3-1.3 e Grupo de 2 e finalmente $p\text{-value} = 8,002 \times 10^{-10}$

para Grupo de 3-2.3 e Grupo de 2). Na Figura 6 apresentam-se os histogramas de frequência da proporção de separação entre os grupos, e os respectivos valores médios.

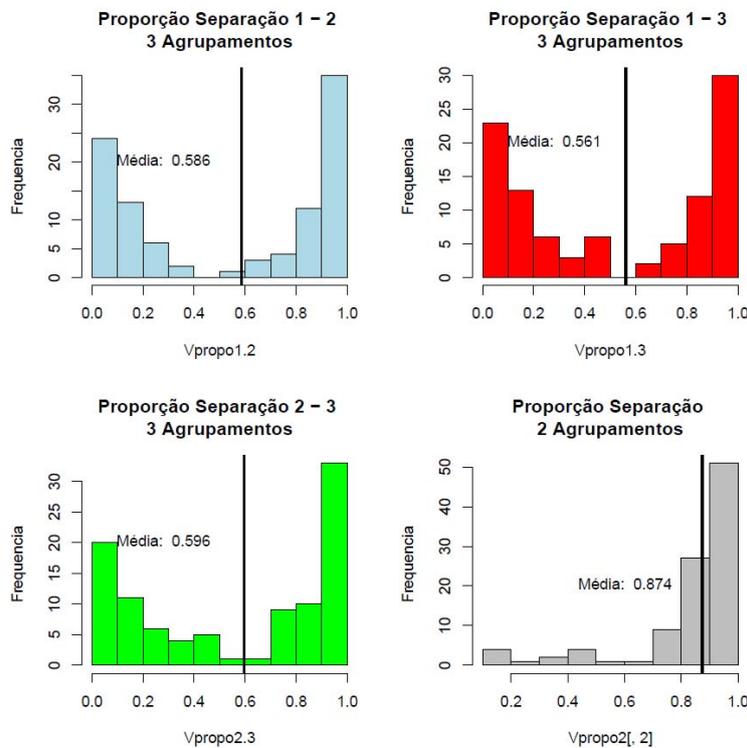


FIGURA 6 - Resultado da simulação para $p=0,95$ do indicador *Ex-post* proporção de separação para 2 agrupamentos e para 3 agrupamentos.

Os resultados mostram que a estrutura dos dados na simulação tende para 2 agrupamentos somente, e que quando a decisão foi pela geração de 2 grupos, a proporção de separação entre os centróides dos objetos individuais foi estatisticamente superior aos casos com 3 grupos.

Para as simulações realizadas com valores de probabilidade “p” de 0,90, 0,75 e 0,50, verificou-se que o indicador *Ex-Ante* do número ideal de agrupamentos apontou na maioria das vezes para 2 agrupamentos. O indicador *Ex-post* apresentou o comportamento de queda conforme a probabilidade “p” foi menor, exceto para o valor $p=0,90$, quando foi maior. Nas Figuras 7 e 8 pode-se observar o comportamento deste indicador.

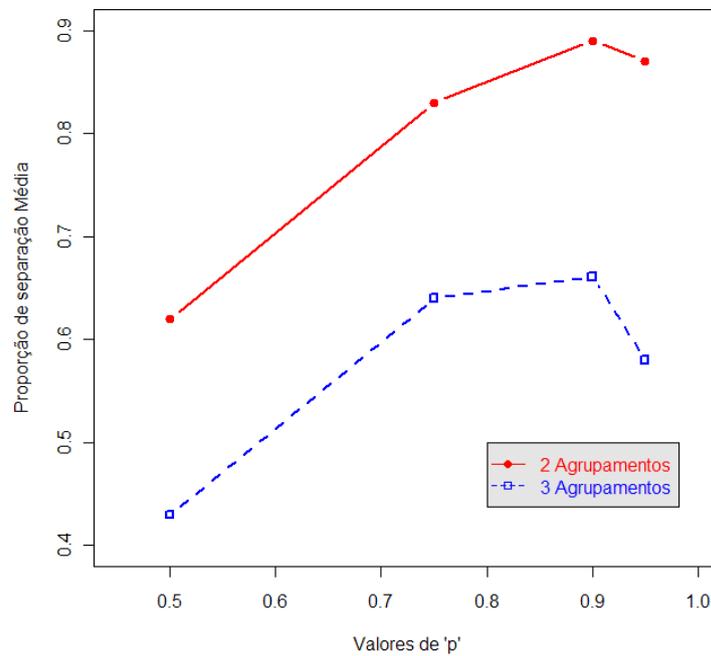


FIGURA 7 - Resultados para o indicador *Ex-post* – proporção de separação média das simulações para $p=0,95$, $p=0,90$, $p=0,75$ e $p=0,50$ para 2 agrupamentos e para 3 agrupamentos.

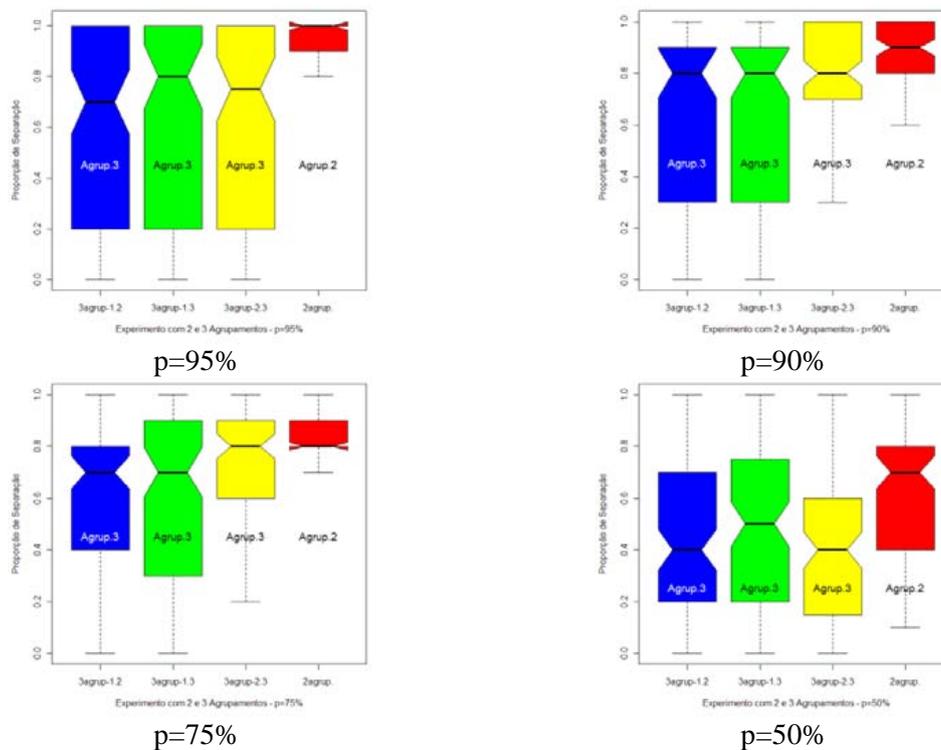


FIGURA 8 – “Boxplot” dos resultados obtidos para o indicador *Ex-post* – proporção de separação média das simulações para $p=0,95$, $p=0,90$, $p=0,75$ e $p=0,50$ para 2 agrupamentos e para 3 agrupamentos.

A análise dos resultados demonstra que a opção para o conjunto de dados simulados com 2 agrupamentos é a indicada em todos os casos, conforme indicado pela métrica de validação de agrupamentos conectividade. O segundo indicador mostra que a opção por 2 agrupamentos também foi superior, pois a proporção de separação dos centróides dos objetos individuais pelo teste de *Hotteling* foi estatisticamente superior em todos os casos ($\alpha=0,05$).

5. Conclusões finais

A proposta de geração de agrupamentos homogêneos antes do processo de redução de dimensão e do alinhamento das diferentes configurações via GPA, utilizando o algoritmo *K-médias* mostrou-se viável, pois possibilita a separação em subgrupos que terão interpretações diferentes no mapa perceptual gerado pelo algoritmo MDSvarext, proporcionando ainda regiões de confiança menores, o que permite diferenciar estatisticamente os diferentes objetos.

A utilização das métricas propostas para a simulação das matrizes de julgamento permitiu definir o melhor número de agrupamentos para o conjunto de dados. A métrica proporção de separação dos centróides dos objetos apresentou comportamento esperado, com a queda na proporção de separação com valores menores de probabilidade “p”, exceto para o ponto $p=0,90$.

Como trabalhos futuros sugerem-se a utilização do modelo proposto em diferentes conjuntos de dados obtidos a partir da coleta de dados com julgadores reais, para verificar as diferentes interpretações que se podem extrair do mapa perceptual, bem como testar as métricas de qualidade *Ex-ante* e *Ex-post*.

Referências Bibliográficas

- Abdi, H.; Dunlop, J. P.; Williams, L. J.** (2009), How to compute reliability estimates and display confidence and tolerance intervals for pattern classifiers using the bootstrap and 3-way multidimensional scaling (distatis). *NeuroImage*, v.45, n. 1, p. 89 – 95.
- Brock G.; Pihur, V.; Datta, S.; Datta, S.** (2008), cIValid: Validation of Clustering Results. Acesso em: 13 Mar. 2010. Disponível em: <<http://www.louisville.edu/~g0broc01/research>>
- Brombin, C.; Salmaso, L.** (2009), Multi-aspect permutation tests in shape analysis with small sample size. *Comput. Stat. Data Anal.*, Elsevier Science Publishers B. V., Amsterdam, The Netherlands, The Netherlands, v.53, p. p.3921–3931.
- Cardoso-Junior, M. M.; Scarpel, R.A.** (2011), Proposal for Development of the Perceptual Map of Socio-environmental Risks using Multivariate Tools and Resampling Methods. In: 19th Triennial Conference of the International Federation of Operational Research Societies - IFORS, 2011, Melbourne. World OR: Global Economy and Sustainable Environment. Melbourne: www.ifors2011.org. p. 1-156.
- De Leeuw, J.** Applications of Convex Analysis to Multidimensional Scaling."In Recent Developments in Statistics, pp. 133-145. [S.l.]: North Holland Publishing Company, Amsterdam., 1977.
- De Leeuw, J.; Mair, P.** (2009), Multidimensional scaling using majorization: Smacof in R. *Journal of Statistical Software*, v.31, p. p.1–30. Disponível em: <<http://www.jstatsoft.org/v31/i03/>>.
- Dijksterhuis, G.; Gower, J.** *The interpretations of generalized Procrustes analysis and allied methods*. [S.l.]: Urtecht: Oliemans Punter and Partners, 2010.
- Dryden, I.** Shapes: Statistical shape analysis. [S.l.], 2009. R package version 1.1-3. Disponível em: <<http://CRAN.R-project.org/package=shapes>>.
- Efron, B.; Tibshirani, R.** (1986), Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, v.1, n.1, p. p.54–77.
- Everitt B. S, Landau S. e Morven L.** Cluster Analysis. London : Arnold. - Vol. 4ª Ed., 2001.
- Jacoby, W. G.** (2009), Public opinion during a presidential campaign: Distinguishing the effects of environmental evolution and attitude change. *Electoral Studies*, v.28, n. 3, p. 422 – 436.

- Moreira, J. A. S.** Mapas Perceptuais e Variações na Participação de Mercado. Dissertação (Mestrado) — Fundação Getúlio Vargas(FGV), 2006.
- R Development Core Team** (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, Disponível: <<http://www.R-project.org>>.
- Rodrigue, N.** A comparison of the performance of generalized procrustes analysis and the intraclass coefficient of correlation to estimate interrater reliability. Dissertação (Mestrado) — McGill University, 1999.
- Saburi, S.; Chino, N.** (2008), A maximum likelihood method for an asymmetric mds model. *Comput. Stat. Data Anal.*, v.52, p. p.4673–4684.
- Slovic, P.** *Risk Perception*. London: Earthscan, 2001.
- Todorov, V., Filzmoser, P.** (2009), An Object-Oriented Framework for Robust Multivariate Analysis. *Journal of Statistical Software*, 32(3), 1-47. Disponível em: <http://www.jstatsoft.org/v32/i03/> Acesso em: 26 Mar. 2012.
- Vanlaar, W.; Yannis, G.** (2006), Perception of road accident causes. *Accident Analysis and Prevention*, v. 38, p. 155–161.
- Webb, A.** *Statistical Pattern Recognition*, West Sussex: John Wiley & Sons, Inc., 2a Ed, 2002.
- Weinberg, S.; Carroll, J.; Cohen, H.** (1984), Confidence regions for indscal using the jackknife and bootstrap techniques. *Psychometrika*, V. 46 n. 4, p. p.475–491.
- Willems G., Pison G., Rousseeuw P. and Van Aelst S.** (2002), A robust hotelling test, *Metrika*, 55, 125–138.