

ESTADÍSTICA ESPAÑOLA
Vol. 45, Núm. 152, 2003, págs. 5 a 22

Robustez en Multidimensional Scaling

por

PEDRO ANTONIO GARCÍA LÓPEZ
DIEGO TORRECILLA DE AMO
ANDRÉS GONZÁLEZ CARMONA

Departamento de Estadística e I.O. Universidad de Granada

ANTONIO PASCUAL ACOSTA

Departamento de Estadística e I.O. Universidad de Sevilla

RESUMEN

En este trabajo se presenta, mediante un estudio de simulación, una solución del problema de Multidimensional Scaling (MDS), cuando los datos aparecen perturbados por un error aleatorio. En el estudio de simulación se plantea un diseño cuyos factores son los distintos estimadores robustos considerados, el tipo de error y la intensidad del mismo. Asimismo se estudia la adecuación de la solución propuesta y su comparación con algoritmos clásicos de resolución del problema de MDS tratado.

Palabras Clave: Multidimensional Scaling (MDS), Estimadores robustos, Simulación, R, S-Plus.

Clasificación AMS: 65U05.

INTRODUCCIÓN

Los juicios de proximidad que producen los datos básicos (similaridades o disimilaridades) para la mayor parte de las aplicaciones de los procedimientos **MDS**, son sensibles a la existencia de observaciones anómalas, no sólo por la dificultad que presenta el detectar y eliminar errores de transcripción e introducción de datos sino porque, a menudo, el problema del cansancio del sujeto que responde cuando ha de establecer un gran número de juicios, le lleva a respuestas erróneas en muchos casos. Además, algunos juicios de proximidad son intrínsecamente más difíciles de emitir que otros, con lo que es razonable suponer que la distribución de los errores asociados a ellos tenga mayor varianza, lo cual introduce un problema añadido en el proceso de reconstrucción de la configuración.

El primer método de resolución para MDS está basado en los trabajos de Young y Householder (1938) y Torgerson (1958). Es probablemente el método más ampliamente usado porque, aunque los procedimientos no métricos han sustituido en muchos casos a los métodos métricos, casi todos los programas usados en la actualidad emplean alguna variante del procedimiento Young-Householder-Torgerson (Y-H-T) como configuración inicial. Obviamente si el enfoque tradicional es sensible a los datos anómalos, los programas que lo usan para obtener la configuración inicial se ven afectados también por ellos.

En este sentido, Spence y Lewandowsky (1989) demuestran cómo un único dato anómalo puede distorsionar drásticamente una solución MDS cuando se emplea el método métrico tradicional. Debido a esta "fragilidad" resulta necesario, e incluso diríamos imprescindible, encontrar técnicas mucho más resistentes que amortigüen estos problemas. Estas razones justifican sobradamente la definición de procedimientos que protejan a estos métodos de la presencia de estas anomalías. En los trabajos de Spence (1982), Null y Sarle (1982), y Heiser (1988) se argumenta la necesidad, para MDS, de algoritmos que sean resistentes a los efectos de los datos anómalos.

Un programa robusto para MDS debe mostrarse por tanto, poco afectado por la presencia de datos anómalos, incluso cuando existan en número relativamente grande y, además, deberá funcionar de manera adecuada cuando no haya ninguno de ellos presente.

Con esta idea, Spence y Lewandowsky (1989) proponen un programa, basado en la mediana como medida robusta, para resolver la presencia de observaciones anómalas. Esta estrategia supone, obviamente, una drástica eliminación de la información métrica contenida en los datos.

En este trabajo se aborda el problema desde un planteamiento más general, considerando varios estimadores robustos que gozan de unas propiedades teóricas idóneas, en el sentido de intentar utilizar toda (o la mayor parte de) la información métrica posible contenida en los datos. Estos estimadores son las α -trimedias y los estimadores de Huber y de Tukey, estos últimos basados en la definición de una función de peso asociada al vector de datos considerado. Sobre las propiedades teóricas de los mismos nos remitimos a Huber (1981) y a Hampel et al. (1986).

OBJETIVOS

En este trabajo se han perseguido tres objetivos básicos:

1. Desarrollar un método de resolución del problema de MDS que sea resistente a la presencia de datos anómalos

Para ello se ha elaborado un diseño de simulación en el que se contaminan las distancias originales de una configuración dada mediante tres procedimientos:

- a. Una pequeña cantidad de grandes errores,
- b. Una moderada cantidad de pequeños errores (a los que llamaremos errores de fondo)
- c. Ambas contaminaciones simultáneamente.

Se trata, por tanto, de desarrollar un procedimiento que se comporte bien en condiciones normales, y presente resistencia frente a los errores contenidos en los datos hasta, al menos, un porcentaje suficientemente amplio de contaminación.

El método será algorítmico con estimadores robustos y tan general como para que pueda aplicarse a un MDS de cualquier dimensión.

2. Controlar la adecuación de la metodología propuesta

Para ello habrá que medir, por una parte, el grado de reconstrucción de las configuraciones finales conseguidas por medio de los procedimientos propuestos, y por otra comprobar si este grado de reconstrucción depende de la configuración inicial elegida en el proceso de simulación.

3. Contrastar los resultados obtenidos por los procedimientos que se proponen con los que se obtienen, en iguales condiciones, por medio de los procedimientos clásicos (métricos y no métricos) de MDS

Para ello se han utilizado los programas PC-MDS, que incorpora el método clásico KYST, y SPSS (V. 10.0) que incorpora el muy difundido método ALSCAL.

FORMULACIÓN DEL PROBLEMA

En MDS se pretende obtener una configuración geométrica mediante un conjunto de N puntos en un espacio euclídeo de dimensión predeterminada, de tal forma que cada uno de ellos represente a un estímulo y que la diferencia, e_{ij} , entre la distancia de los puntos i y j , d_{ij} , y las disparidades o disimilaridades existentes entre los correspondientes estímulos, δ_{ij} , (i y $j: 1, \dots, N; i \neq j$), sea mínima.

Si designamos por k la dimensión del espacio euclídeo considerado, es posible encontrar la solución del problema de MDS resolviendo el siguiente conjunto de ecuaciones:

$$e_{ij} = \delta_{ij} - d_{ij} = 0 \quad \text{para } i, j: 1, \dots, N \text{ con } i \neq j$$

cuyas incógnitas son las coordenadas de los puntos que se pretenden encontrar. Utilizando la métrica euclídea para los valores de d_{ij} y designando por x_{rs} a la coordenada s -ésima ($s: 1, \dots, k$) del punto genérico, P_r ($r: 1, \dots, N$), de la configuración final, X , se obtienen del sistema anterior $N-1$ ecuaciones en donde aparecerán x_{rs} como incógnitas. Utilizando el método de Newton en cada una de estas ecuaciones, se obtiene la solución iterativa:

$$x_{rs}^{t+1} = x_{rs}^t + \frac{(\delta_{rj} - d_{rj}^t) \cdot d_{rj}^t}{x_{rs}^t - x_{js}^t}$$

Para obtener la coordenada x_{rs} en el paso $t+1$, a partir del correspondiente valor en el paso t , se tienen $N-1$ correcciones, tantas como valores puede tomar j . Si escribimos las $N-1$ ecuaciones como:

$$x_{rs}^{t+1} = x_{rs}^t + (c_{rs}^t)_j \quad j: 1, \dots, N, j \neq r$$

podemos elegir una medida robusta, rob , y calcular su valor a partir de las $N-1$ correcciones $(c_{rs}^t)_j$ para pasar de x_{rs}^t a x_{rs}^{t+1} .

Por tanto, si $(c_{rs}^t) = rob_{j \neq r} (c_{rs}^t)_j$ es una determinación robusta(1) de la corrección, obtenemos:

$$x_{rs}^{t+1} = x_{rs}^t + (c_{rs}^t)$$

(1) En este trabajo hemos usado cinco medidas robustas diferentes: tres L-estimadores (Mediana, Trimedia al 30% y Trimedia al 40%) y dos M-estimadores (el estimador de Huber y el estimador de Tukey)

Pueden realizarse modificaciones del tamaño del paso (c_{rs}^t) en la dirección del vector de correcciones para acelerar la convergencia del procedimiento iterativo. Siguiendo a D.G. Luenberger (1984), el tamaño del paso puede ser modificado multiplicándolo por el factor:

$$\gamma^t = \frac{\alpha^t}{c^t}$$

en donde c^t es la magnitud relativa al vector de correcciones:

$$c^t = \left[\frac{\sum_{i,a} (c_{ia}^t)^2}{\sum_{i,a} (x_{ia}^t)^2} \right]^{1/2}$$

y α^t se obtiene del método δ^2 de Aitken (Ramsay, 1977)

$$\alpha^{t+1} = \alpha^t \left[\frac{\sum_{i,a} (x_{ia}^{t-1} - x_{ia}^{t-2})^2}{\sum_{i,a} (x_{ia}^t - 2x_{ia}^{t-1} + x_{ia}^{t-2})^2} \right]^{1/2}$$

lo cual obliga a generar dos soluciones previas, antes de abordar el procedimiento que aquí se expone. El algoritmo final, queda pues de la forma siguiente:

$$x_{rs}^{t+1} = x_{rs}^t + \gamma^t \cdot (c_{rs}^t)$$

El algoritmo propuesto(2) ha sido programado en R y S-Plus.

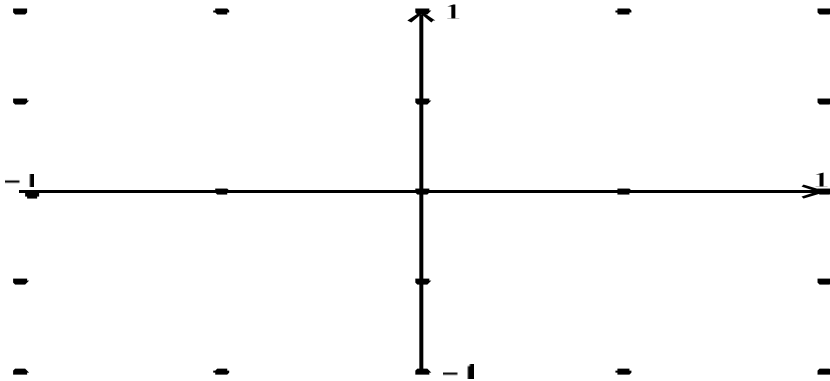
ESTUDIO DE SIMULACIÓN

El estudio consta de tres fases:

- **Primera fase:** Se estudia el comportamiento de los algoritmos presentados aplicándolos a una determinada configuración original en dimensión dos, perturbada convenientemente, y se observa el grado de reconstrucción obtenido.

(2) Puede solicitarse en pagarcia@ugr.es

La configuración inicial seleccionada está formada por 21 puntos del plano, situados 16 de ellos sobre los lados de un cuadrado de lado unidad y equidistantes entre sí, otro en el origen de coordenadas y los cuatro restantes sobre los ejes de coordenadas, a distancia $1/2$ del origen, según se muestra en la figura 1.



Estos 21 puntos determinan una matriz triangular con 210 distancias y las disimilitudes son el resultado de someter a dichas distancias a un efecto distorsionador con un diseño experimental factorial de $5 \times 4 \times 2$ con seis repeticiones por celda, del siguiente modo:

- **Cinco** tipos de medidas robustas: Trimedia al 30%, Trimedia al 40%, Mediana, Huber y Tukey.
- **Cuatro** porcentajes de distancias anómalas: 10%, 20%, 30% y 40%. El método consiste en seleccionar aleatoriamente el porcentaje de distancias correspondiente y contaminar dichas distancias con errores (Ramsay, 1977) lognormales pseudoaleatorios de desviación típica igual a 2. La contaminación se produce multiplicando la distancia por la transformada exponencial del error.
- **Dos** escenarios: Con error de fondo y sin él. El método consiste en contaminar, con el mismo criterio del punto anterior, distancias hasta completar el 50% del total de ellas, con errores pseudoaleatorios procedentes de una distribución lognormal con desviación típica igual a 0.1.

En resumen, se llevan a cabo un total de $5 \times 4 \times 2 \times 6 = 240$ estudios separados.

Puesto que la solución que en este caso proporciona el método Y-H-T es bastante pobre, por cuanto se ve ampliamente afectada por la presencia de valores anómalos, se establece como primera configuración de los procedimientos iterati-

vos una definida por los rangos de las distancias perturbadas, esto es, los datos iniciales (disimilaridades) se reemplazan por sus rangos, lo que atenúa la contaminación anómala de los datos, y a continuación se calcula sobre ellos una solución MDS por el procedimiento Y-H-T.

El problema de la constante aditiva se resuelve utilizando una versión robusta del método de Torgerson. En este caso:

$$c_0 = \text{rob}_i \left\{ \text{rob}_{j \neq i} \left\{ \text{rob}_{h \neq i, j} (\delta_{ij} - \delta_{ih} - \delta_{jh}) \right\} \right\}$$

con lo que la escala original de distancias se reconstruye obligando a que las medianas verifiquen la condición:

$$\text{med}_{i < j} (\delta_{ij} + c_0) = \text{med}_{i < j} (d_{ij}^0)$$

Como medida del índice de reconstrucción se ha utilizado el coeficiente de correlación lineal medio (de las 6 réplicas) entre las distancias de la configuración original y las de la configuración reconstruida por el procedimiento.

Con este diseño se han obtenido los resultados de las tablas 1 y 2, en donde el valor superior de cada fila es el índice medio de reconstrucción (r-medio) y el valor inferior corresponde a la desviación típica para cada r-medio.

Tabla 1

<i>Medida robusta</i>	<i>Porcentaje de datos anómalos (sin error de fondo)</i>			
	<i>10%</i>	<i>20%</i>	<i>30%</i>	<i>40%</i>
	99,65	95,06	86,63	78,23
Trim-30%	0,0023	0,0881	0,0863	0,1201
	99,82	99,89	95,31	77,49
Trim-40%	0,0030	0,0018	0,0391	0,1204
	99,94	99,99	96,59	82,09
Mediana	0,0006	0,0002	0,0307	0,0805
	99,69	95,01	93,11	69,21
Huber	0,0025	0,0636	0,0416	0,0820
	98,23	99,09	87,81	78,57
Tukey	0,0174	0,0124	0,1008	0,1443

Tabla 2

<i>Medida robusta</i>	<i>Porcentaje de datos anómalos (con error de fondo)</i>			
	<i>10%-40%</i>	<i>20%-30%</i>	<i>30%-20%</i>	<i>40%-10%</i>
	92,50	96,44	89,36	51,31
Trim-30%	0,0848	0,0417	0,0993	0,1472
	94,86	99,38	90,06	52,86
Trim-40%	0,0758	0,0058	0,0899	0,2111
	99,79	97,74	96,48	57,55
Mediana	0,0018	0,0370	0,0502	0,02470
	96,18	93,71	90,71	57,43
Huber	0,0691	0,0772	0,0562	0,0646
	99,69	92,94	95,11	66,43
Tukey	0,012	0,1005	0,0591	0,2126

Los gráficos 1 y 2 muestran los valores correspondientes a las tablas 1 y 2

Gráfico 1

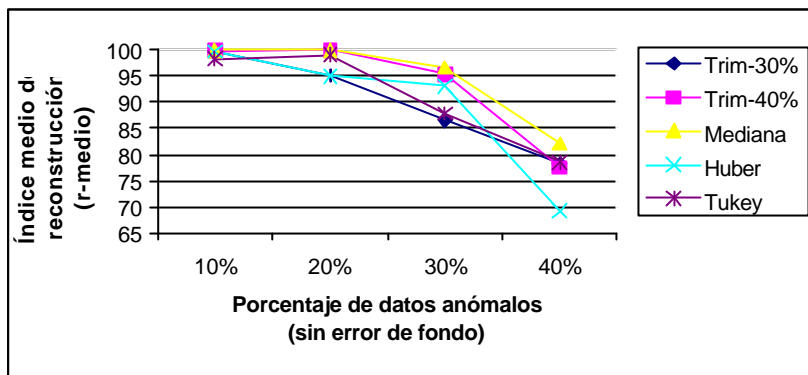
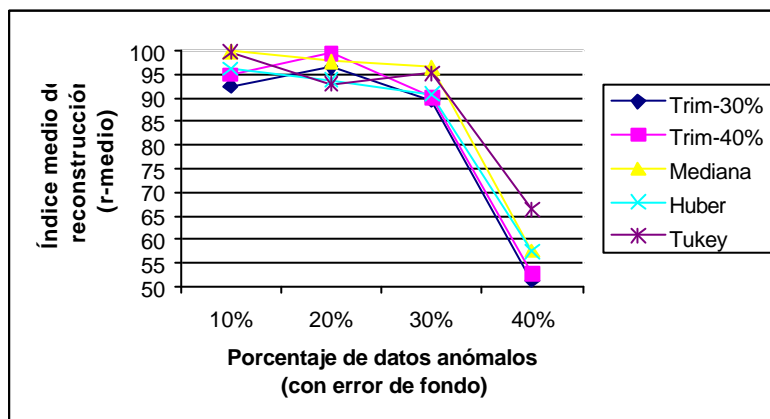


Gráfico 2



A la vista de estos resultados es conveniente hacer notar que es difícil calcular un valor a partir del cual pueda decirse que no se obtiene una reconstrucción idónea. No obstante, podemos afirmar que todos los estimadores robustos obtienen la configuración original de forma **muy buena** cuando la presencia de datos anómalos llega hasta un 30% del total (63 distancias perturbadas de 210). La pérdida mayor se produce en la trimedia al 30%, con un 13.37% cuando están presentes el 30% de anómalos.

Cuando los datos originales se contaminan además con un error de fondo, los resultados empeoran ligeramente, pero siguen siendo muy buenos, sobre todo

hasta una contaminación del 30% y 20% de anómalos y error de fondo, respectivamente.

Se confirma que los estimadores robustos utilizados para la reconstrucción de la configuración original, se comportan muy bien, por lo que no necesariamente tendríamos que restringirnos al uso de la mediana.

No obstante lo dicho, conviene señalar que al promediar los resultados con las réplicas podemos ocultar los efectos de determinados patrones de datos anómalos. Consideremos qué pasaría si todos estos factores incidieran en una misma fila o columna de la matriz de datos originales. Es evidente que entonces, ningún procedimiento reconstruiría la configuración con exactitud sobre ese punto.

- Segunda fase: En la segunda fase se trata de comprobar que los resultados del primer estudio no dependen del tipo de configuración inicial usada. Para ello, se ha llevado a cabo un diseño experimental similar al de la primera fase, que consiste en un diseño factorial de $5 \times 3 \times 2 \times 6$, esto es:

- **Cinco** estimadores robustos, coincidentes con los de la primera fase.

- **Tres** porcentajes de datos anómalos: 10%, 20% y 30% por ciento sobre las distancias originales. No se considera en esta fase un 40% de valores anómalos vistos los resultados obtenidos en la primera fase.

- **Dos** escenarios de error: con y sin error de fondo hasta el 50% de igual forma que en la primera fase.

- **Seis** configuraciones aleatorias de 21 puntos cada una, en dos dimensiones, generadas mediante la obtención de una muestra pseudoaleatoria de las coordenadas a partir de una distribución uniforme en el intervalo $(-1,1)$.

Cada procedimiento recibe los mismos datos de entrada, por lo que se llevan a cabo un total de $5 \times 3 \times 2 \times 6 = 180$ estudios. Los resultados obtenidos se muestran en las tablas 3 y 4.

Tabla 3

<i>Medida robusta</i>	<i>Porcentaje de datos anómalos (sin error de fondo)</i>		
	<i>10%</i>	<i>20%</i>	<i>30%</i>
	96,97	90,11	78,11
Trim-30%	0,0367	0,0958	0,1882
	99,96	96,23	83,29
Trim-40%	0,0005	0,0469	0,1269
	99,97	99,46	89,80
Mediana	0,0006	0,0059	0,1163
	97,73	93,75	77,12
Huber	0,0218	0,0888	0,1708
	97,16	96,70	91,20
Tukey	0,0242	0,0274	0,0603

Tabla 4

<i>Medida robusta</i>	<i>Porcentaje de datos anómalos (sin error de fondo)</i>		
	<i>10%-40%</i>	<i>20%-30%</i>	<i>30%-20%</i>
	99,50	94,65	81,05
Trim-30%	0,0022	0,0508	0,1179
	99,49	92,13	93,50
Trim-40%	0,0576	0,00592	0,0415
	99,59	97,98	96,47
Mediana	0,0049	0,0362	0,0365
	99,53	94,39	86,05
Huber	0,0029	0,0612	0,1042
	99,69	93,97	94,59
Tukey	0,0031	0,0604	0,0383

Los gráficos 3 y 4 muestran los datos de las tablas 3 y 4

Gráfico 3

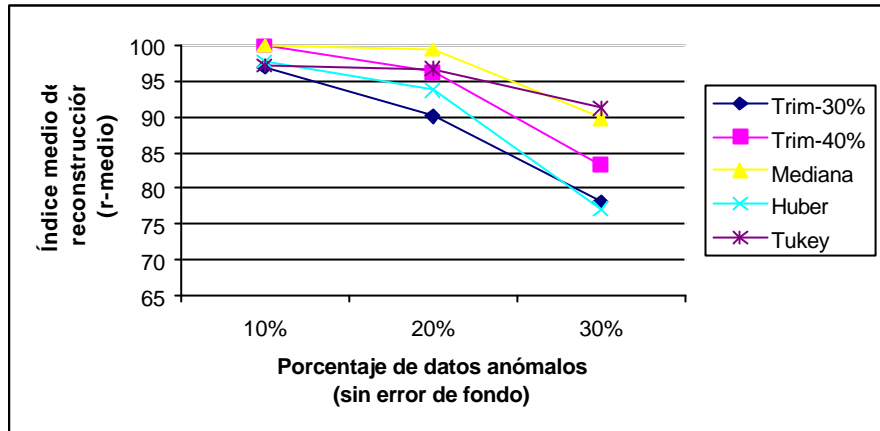
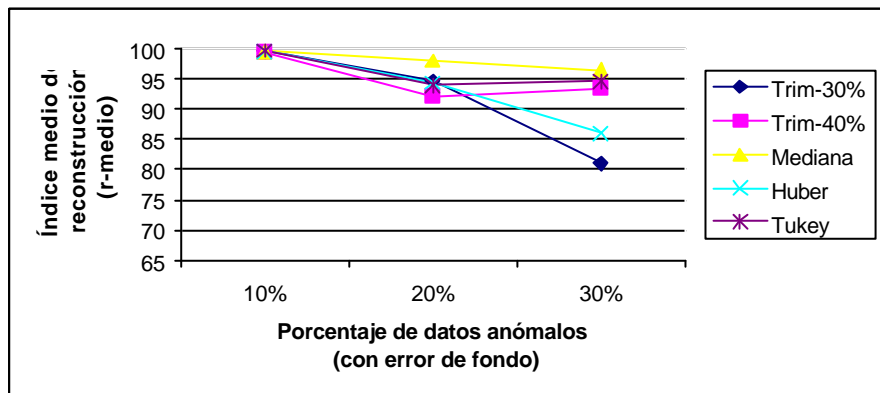


Gráfico 4



Se observa que los niveles de reconstrucción de la configuración (aleatoria en este caso) son bastante altos para todas las medidas robustas planteadas, a excepción hecha de las trimedias, al igual que en la primera fase. Esta conclusión avala la hipótesis de que los procedimientos propuestos no dependen de la configuración inicial considerada.

• Tercera Fase: Se utilizan como procedimientos de solución del problema de MDS, ALSCAL (NM) y KYST (M). El primero, en la versión implementada en el programa SPSS v.10.0 y el segundo en PC-MDS. Cada uno de ellos recibe los mismos datos de entrada que se han utilizado en las dos fases anteriores, con los mismos datos anómalos y errores de fondo.

Los resultados, para la configuración considerada en la primera fase se muestran en la tabla 5.

Tabla 5

<i>Sin error de fondo</i>	<i>Porcentaje de datos anómalos</i>			
	<i>10%</i>	<i>20%</i>	<i>30%</i>	<i>40%</i>
	81,69	65,44	74,75	63,36
ALSCAL	0,03120	0,3896	0,2967	0,2329
	72,21	83,44	59,67	51,09
KYST	0,3558	0,1925	0,3175	0,3057
<i>Con error de fondo</i>	<i>10%-40%</i>	<i>20%-30%</i>	<i>30%-20%</i>	<i>40%-10%</i>
	96,11	60,37	87,13	30,92
ALSCAL	0,0059	0,04265	0,0344	0,3009
	82,38	60,44	64,11	30,37
KYST	0,3417	0,3545	0,2628	0,1455

Los gráficos 5.1 y 5.2 muestran respectivamente cada uno de los datos de la tabla 5 y los correspondientes al peor ajuste robusto de los obtenidos en la primera fase, pudiendo observarse que siempre es mejor este último.

Gráfico 5.1

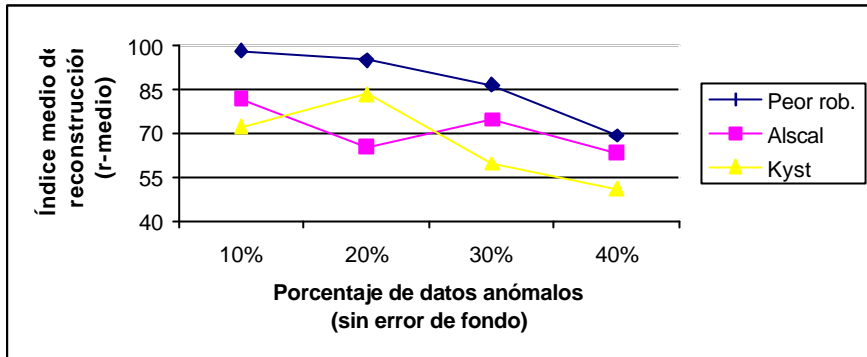
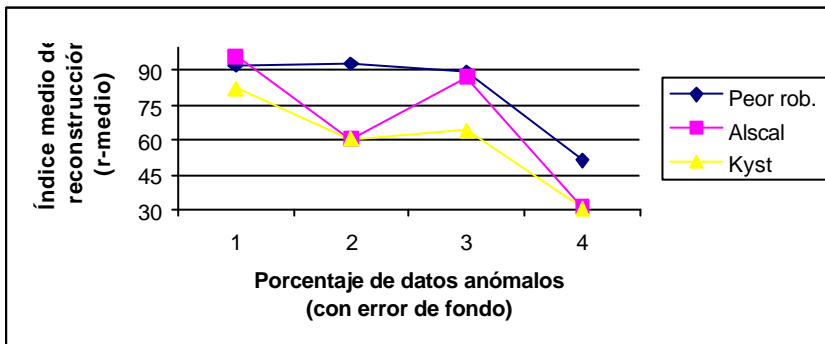


Gráfico 5.2



Los resultados obtenidos considerando la configuración aleatoria de la segunda fase, y los procedimientos KYST y ALSCAL, se muestran en la tabla 6.

Tabla 6

Sin error de fondo	Porcentaje de datos anómalos		
	10%	20%	30%
ALSCAL	73,38	84,05	74,33
KYST	0,3559	0,1045	0,0725
Con error de fondo	84,81	78,30	76,97
	0,0637	0,1567	0,1581
	10%-40%	20%-30%	30%-20%
ALSCAL	93,35	78,53	77,68
KYST	0,0226	0,1303	0,1293
	94,53	88,28	51,70
	0,0317	0,0598	0,1878

Los gráficos 6.1 y 6.2 muestran la comparación entre estos resultados y los obtenidos con el peor ajuste robusto de los obtenidos en la segunda fase. De nuevo se observa que este último es siempre mejor que los procedimientos clásicos considerados.

Gráfico 6.1

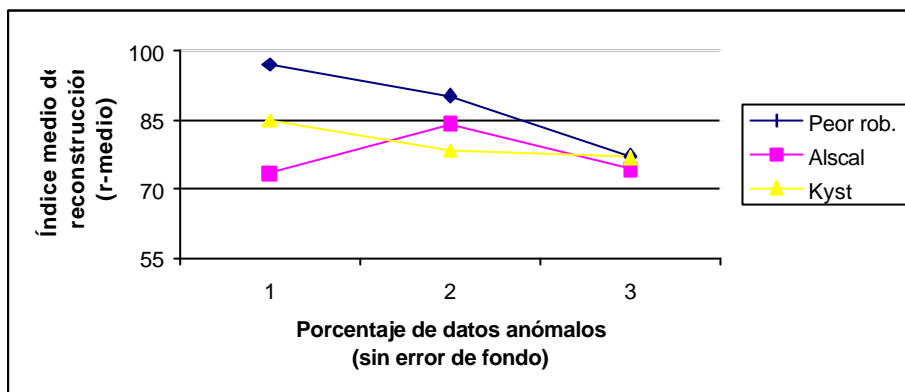
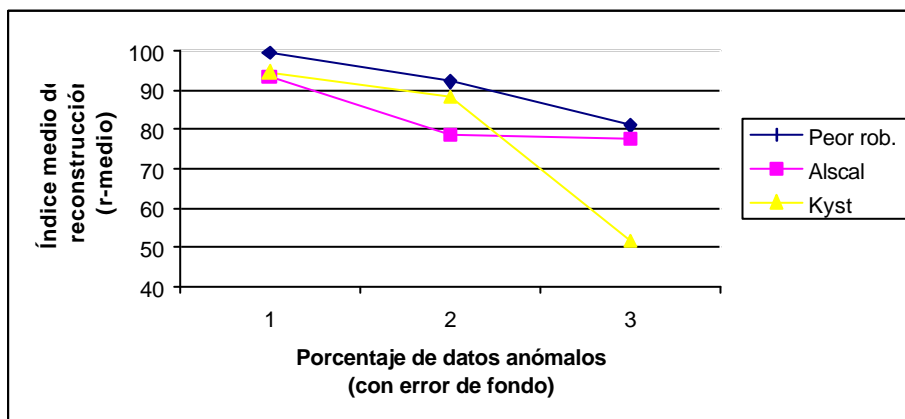


Gráfico 6.2



CONCLUSIONES

1. En el estudio realizado, tanto en la primera fase como en la segunda, los procedimientos robustos ofrecen resultados satisfactorios hasta para un 30% de datos anómalos, con y sin error de fondo.
2. El estudio realizado en la tercera fase, deja claro que cualquiera de los procedimientos expuestos supera, y a veces considerablemente, a los procedimientos clásicos ALSCAL en su versión no métrica y a KYST en su versión métrica.

REFERENCIAS

- ANDREWS, D. F.; BINCKEL, P. J.; HAMPEL, F. R.; HUBER, P. J.; ROGERS, W. H. & TUKEY, J. W. (1972).- «Robust Estimates of Location». Princeton, New Jersey: Princeton University Press.
- CLARKSON, D. B. (1988).- «Robustness of Multidimensional Scaling Estimates» *Proceedings of the Statistical Computing Section, Annual Meeting of the A.S.A.*, 175-180.
- CLARKSON, D. B. & KIM, J.- «Robustness of Nonmetric Multidimensional Scaling Estimates». *IMSL Inc., Houston, TX*, 175-180.
- COX, T.F. & COX, M.A.A. (1994).- «Multidimensional Scaling». *Chapman & Hall*
- GARCÍA-LÓPEZ, P.A., GONZÁLEZ, A. & VERA, J.F. (1993).- «Contributions to the Non-metric Maximum Likelihood MDS from Directional Rankings of Similarities Method. Proceedings of the 1993» *European Meeting of the Psychometric Society*.
- GOODALL, C. (1982).- «M-Estimators of Location: an outline of the theory. Understanding Robust and Exploratory Data Analysis». *New York: Wiley*, 339-400.
- HAMPEL, F.R.; RONCHETTI, E.M.; ROUSSEEUW, P.J. & STAHEL, W.A. (1986).- «Robust Statistics. The Approach Based on Influence Functions» *Wiley Series in Probability and Mathematical Statistics*.
- HEISER, W. J (1988).- «Multidimensional scaling with least absolute residuals». *Classification and Related Methods of Data Analysis. H.H. Bock (Editor)*, 455-462

- HUBER, P. J. (1981).- «Robust Statistics». Wiley Series in Probability and Mathematical Statistics.
- KRUSKAL, J. B. (1964a).- «Multidimensional scaling by optimizing goodness-of-fit to a nonmetric hypothesis» *Psychometrika*, 29, 1-27.
- KRUSKAL, J. B. (1964b).- «Nonmetric multidimensional scaling: a numerical method» *Psychometrika*, 29, 115-129.
- LUENBERGER, D. G. (1984).- «Linear and nonlinear programming». Reading, MA: Addison-Wesley.
- NULL, C. H. & SARLE, W. (June 1982).- «Robust multidimensional scaling». Paper presented at the Joint Meeting of the Psychometric and Classification Societies. Montreal, Canada.
- R DEVELOPMENT CORE TEAM (2002).- «R: A Programming Environment for Data Analysis and Graphics».
- RAMSAY, J. O. (1977).- «Maximum Likelihood Estimation in Multidimensional Scaling». *Psychometrika*, 42, 241-266.
- ROSEMBER, J. L. & GASKO, M. (1982).- «Comparing location estimators: Trimmed Means, medians and trimean. Understanding Robust and Exploratory Data Analysis». New York: Wiley, 297-336.
- SPENCE, I. (1982).- «Robust multidimensional scaling». Paper presented at the Joint Meeting of the Psychometric Society and the Classification Society. Montréal, Canada.
- SPENCE, I. (1983).- «Monte Carlo simulation studies». *Applied Psychological Measurement*, 7, 405-425.
- SPENCE, I. & LEWANDOWSKY, S. (1989).- «Robust Multidimensional Scaling». *Psychometrika*, 54, 501-513.
- THISTED, R. A. (1988).- «Elements of Statistical Computing. Numerical Computation». Chapman & Hall.
- TORGERSON, W. S. (1958).- «Theory and methods of scaling» New York: Wiley.
- TORRECILLA DE AMO, D. (1999).- «Estudios de Robustez en Multidimensional Scaling». Tesis Doctoral. Universidad de Granada.
- YOUNG, G. & HOUSEHOLDER, A. S., (1938).- «Discussion of a set of points in terms of their mutual distances». *Psychometrika*, 3, 19-22.

ROBUSTNESS IN MULTIDIMENSIONAL SCALING

SUMMARY

In this paper we show, by means a simulation study, a method to solve the perturbational error Multidimensional Scaling (MDS) problem. We work with a factorial design in which it has been considered several robust estimators, and two types of random error with several intensities. Furthermore we study the fit of the obtained configuration and its comparison with other classical MDS methods.

Key-words: Multidimensional Scaling (MDS), Robust estimators, Simulation, R, S-Plus 2000.

AMS Clasification: 65U05.