

**OPTIMIZACIÓN DE LOS TEST EDUCATIVOS
MEDIANTE ESCALAMIENTO
MULTIDIMENSIONAL NO-MÉTRICO**

Daniel Patón Domínguez
Facultad de Ciencias. UEx.

Ignacio Chato Gonzalo
IEM Jaranda, Jarandilla de la Vera, Cáceres

María Méndez Suárez
Centro Universitario de Mérida. UEx

Matías Hidalgo Sánchez
Facultad de Ciencias. UEx.

ANTECEDENTES

La realización de pruebas de valoración cuenta con una larga tradición en el ámbito educativo (Abascal y Franco, 2002). Toda encuesta debe basarse en el concepto de idoneidad (Arriaza, 1984). Es decir, debemos plantear qué queremos estudiar, elegir la técnica de muestreo, seleccionar la muestra apropiada, formular las preguntas adecuadas y saber interpretar los resultados. Este proceso está sujeto a posibles errores en todas sus fases, por lo que debe ponerse especial cuidado en todas ellas. Una encuesta solo será idónea cuando mida claramente aquello que queremos valorar. Entonces podremos contrastar hipótesis e incluso formular teorías en el ámbito educativo. Una buena encuesta debe aportar información cuantitativa útil que pueda ser comparada con otros estudios. Además, debe seguir el principio de parsimonia, es decir debe tener el mínimo número de preguntas que permitan valorar lo que se trata de cuantificar. Introducir preguntas superfluas en las encuestas no aporta ningún valor a la mismas e incluso puede ser perjudicial al aumentar la entropía de los datos (Kraemer, 2005). Por otro lado, las encuestas largas aumentan los costos, el cansancio del encuestado y su consiguiente falta de precisión en las respuestas. El problema es mayor en la infancia y la tercera edad.

Para valorar encuestas en el ámbito educativo se han usado muchos métodos (Tavsancil et al., 2010). Los análisis multivariantes (AM) son idóneos por sus propiedades estadísticas (Closas et al., 2013). Por un lado, permiten simplificar la complejidad de las variables (preguntas) a un número reducido y significativo de las mismas. De este modo, evitamos ruido estadístico y simplificamos los test. Por otro lado, nos ayudan a determinar los factores subyacentes (ejes multivariantes) que explican la variabilidad observada en los datos. Estos factores son de gran relevancia en la interpretación de los resultados y en la generación de hipótesis para futuras investigaciones. Además, los AM son una herramienta útil para analizar si la muestra (personas) es homogénea o por el contrario presenta grupos con diferentes respuestas a los test. Finalmente, los AM permiten comparar distintos estudios, ya que dan una información cuantitativa simplificada a unos pocos ejes de variación. Las diferencias en los factores subyacentes pueden ser posteriormente analizadas mediante procesos de rotación matemática (*procrustes*) (Khan et al., 2017).

Existe un gran número de AM disponibles en la mayoría de paquetes estadísticos (Rencher, 2002). La elección de una u otra depende de múltiples factores. Por un lado, el tipo de datos de la muestra condiciona la metodología a emplear (Legendre y Legendre, 2012). Algunas técnicas, como el Análisis de Componentes Principales, solo son apropiadas para variables continuas. Otras, como el Análisis de Correspondencias, son idóneas para frecuencias. Además, del tipo de datos debemos tener en cuenta las propiedades intrínsecas del proceso de extracción de los ejes, que puede ser por descomposición matricial, iteración, etc. De entre la pléyade de técnicas disponibles, el Análisis de Escalamiento Multidimensional No-Métrico (EMNM) se ha revelado como una de las más apropiadas en ciencias sociales (Borg y Groenen, 2005). Este método permite adaptarse a cualquier tipo de datos, ya que se basa en distancias o medidas de similaridad. Estas se construyen sobre las matrices de datos originales. El investigador dispone de un gran arsenal de distancias posibles. El EMNM se diferencia de otras técnicas

en el proceso de extracción de los ejes multivariantes. Estos son calculados para minimizar el estrés (error global) de los datos y no por un proceso de descomposición matricial en términos de varianza. Se han propuesto diferentes algoritmos para calcular el estrés (De Leeuw y Mair, 2009). El EMNM es idóneo para explicar los factores subyacentes y suele dar soluciones más parsimoniosas que otros AM (Minchin, 1987). A diferencia del ACP y del AC, no estamos sujetos a que los gradientes sean lineales o modales, pudiendo extraerse gradientes curvilíneos, asíntóticos, etc. Finalmente los ejes son re-escalados mediante una regresión monotónica basada en rangos. Esto disminuye aún más las posibles curvilinearidades en los gradientes, lo que simplifica la interpretación de los mismos (Legendre y Legendre, 2012). Por todo ello, resulta un método idóneo en fuentes complejas de datos con variables incluso de diferentes escalas de medida (Kenkel y Orlóci, 1986). Además, es un método altamente recomendable para aplicarlo junto a técnicas de agrupamiento (*clusters*) (Salmasso, 1996). En el presente trabajo exploramos el EMNM con medidas de test educativos realizados por alumnos de la UEx al objeto de determinar las posibilidades prácticas de este AM.

OBJETIVO

Pretendemos determinar un método rápido y potente de optimización de test educativos que pueda ser implementado en conjuntos de instrucciones en el lenguaje R.

PARTICIPANTES

Se encuestó a diversos grupos de alumnos de diversos grados de la UEx dentro de un proyecto de investigación sobre las relaciones inter-generacionales. Diversos profesores de las Facultades de Ciencias de la Educación y Ciencias se coordinaron para la recogida y análisis de datos. También, participaron docentes y alumnos de enseñanzas medias.

MÉTODOS

Los datos de encuestas fueron transferidos a información binaria de presencias (1) y ausencias (0) en una Hoja de Cálculo (LibreOffice Calc) y exportadas a formato CSV para su posterior análisis usando el software R (R Core Team, 2017). Se utilizaron las librerías *vegan* (Oksanen et al., 2017), *smacof* (De Leeuw y Mair, 2009) y *ecodist* (Goslee y Urban, 2007). Si definimos la siguiente tabla de contingencia para dos encuestados:

		Encuestado X	
		Presencia (1)	Ausencia (0)
Encuestado Y	Presencia (1)	A	B
	Ausencia (0)	C	D

Los índices (distancias o similaridades) d_{jk} entre la muestra j y k estudiados serían:

Alt Gower - Canberra. Ambas métricas son distintas en su forma cuantitativa pero coinciden en su forma binaria. Esta medida es independiente de los dobles ceros o dobles ausencias (Cuadras, 1989).

$$d[jk] = \frac{(C+B-2*A)}{(C+B-A)}$$

Binomial. La distancia binomial se deriva de la desviación binomial bajo la hipótesis nula de que las dos muestras comparadas son iguales. Este índice es tolerante a diferencias de tamaños muestrales entre variables:

$$d[jk] = \log(2) * (C+B-2*A)$$

Bray-Curtis - Horn. Este índice expresa una medida de disimilaridad. No es una distancia matemática propiamente dicha por no cumplir la propiedad de desigualdad triangular (Choi et al., 2010). En su forma binaria coincide con la distancia de Horn (Oksanen et al., 2017).

$$d[jk] = \frac{(C+B-2*A)}{(C+B)}$$

Cuando este índice es expresado como medida de similaridad, se conoce también como índice de Steinhaus, Czekanowski o Dice-Sørensen. Entonces su fórmula es:

$$d[jk] = \frac{2*A}{2A+B+C}$$

Chao. Existen diversas variantes de esta métrica según se basen en los índices de Jaccard o Sørensen. La diferencia entre estas dos variantes se debe a si se usan las respuestas presentes o coincidentes entre encuestados. En nuestro caso usaremos la medida análoga al índice de Jaccard, es decir, basada en las presencias (Minchin, 1987).

Euclidea. Se trata de una verdadera distancia matemática y es una de las más empleadas. Es la raíz cuadrada de la suma de los cuadrados de las diferencias entre los valores de los elementos. Ésta es la distancia por defecto que suele usarse para datos de intervalo. Depende de la escala de medida. Sus valores no están acotados.

$$d[jk] = \sqrt{B+C}$$

En su forma cuantitativa, esta distancia es solo apropiada para conjuntos de datos lineales. Si los datos no lo son, pueden linearizarse con alguna transformación. En su forma binaria no tenemos ese problema. Pero esta distancia es muy sensible a los dobles ceros (dobles ausencias), por lo que su uso en encuestas debe hacerse conociendo esta característica.

Gower. Se trata de otra medida de similaridad.

$$d[jk] = \frac{A}{A+B+C+D}$$

Hamman. Medida de similaridad.

$$d[jk] = \frac{A - (B+C) + D}{A+B+C+D}$$

Índice de Emparejamiento.

$$d[jk] = \frac{(A+D)}{(A+B+C+D)}$$

Jaccard. Medida de similaridad para datos cualitativos. La fórmula se basa en las variables coincidentes por dos encuestados cualesquiera.

$$d[jk] = \frac{A}{(A+B+C)}$$

La versión cuantitativa de esta distancia se conoce como índice de Ružička.

Kulczynski. Medida de similaridad.

$$d[jk] = \frac{A}{B+C}$$

Minkowski. Medida de disimilaridad

$$d[jk] = B+C$$

Manhattan-Hamming. Ambas medidas de disimilaridad son equivalentes para caracteres binarios.

$$d[jk] = B+C$$

Mountford. Medida de similaridad.

$$d[jk] = \frac{A}{\frac{A*B+A*C}{2} + B*C}$$

Ochiai. Medida de similaridad

$$d[jk] = \frac{A}{\sqrt{(A+B)*(A+C)}}$$

Phi de Pearson. Medida de similaridad

$$d[jk] = \frac{A*D - B*C}{\sqrt{(A+B)*(A+C)*(B+D)*(D+C)}}$$

Raup-Crick. Medida de disimilaridad que expresa la probabilidad de respuestas similares de dos encuestados.

$$d[jk] = 1 - \text{prob}(A)$$

Rogers-Tanimoto. Medida de similaridad.

$$d[jk] = \frac{A+D}{A+2(B+C)+D}$$

Sokal-Sneath. Medida de similaridad que se basa en los atributos negativos compartidos. Por ello, es apropiada cuando hay muchos ceros en la matriz (Herrera, 2000). Hay distintas variantes propuestas por estos autores. La primera corresponde al coeficiente S5 de Gower y Legendre (1986).

$$d[jk] = \frac{A}{A+2*(B+C)}$$

La segunda variante corresponde al coeficiente S13 de Gower y Legendre (1986).

$$d[jk] = \frac{A * D}{\sqrt{(A+B)*(A+C)*(D+B)*(D+C)}}$$

Con cada uno de estos índices (distancias o medidas de similaridad/disimilaridad) se realizaron EMNM entre 1 y 20 ejes de variación. En cada caso se determinó la función de estrés de Kruskal, también conocida como *estrés -1* (Hout et al., 2013).

$$\text{Estrés} - 1 = \sigma_1 = \sqrt{\frac{\sum (f(\delta_{i,j}) - d_{i,j})^2}{\sum d_{i,j}^2}}$$

Dicha función de estrés determina las diferencias cuadráticas entre las similaridades/disimilaridades originales de los datos ($\delta_{i,j}$) y las distancias en la ordenación ($d_{i,j}$). En el argot propio de este método se conocen ambos parámetros como proximidades y disparidades (Borg y Groenen, 2005). La selección del NMDS idóneo se realizó observando a qué número de ejes la tendencia de caída del estrés pasa de positiva a negativa. Obviamente se seleccionó la distancia con menor valor de estrés en el menor número de ejes. Una vez seleccionado el mejor NMDS, se testó la calidad de la ordenación mediante el análisis de Shepard (Hout et al., 2013). Este análisis calcula una regresión monotónica entre proximidades ($\delta_{i,j}$) y disparidades ($d_{i,j}$). Se determinó el valor R^2 de dicho modelo y el ajuste a los factores adicionales estudiados (edad, sexo, grado, etc.). Se realizaron medidas adicionales de estrés para determinar las preguntas del test que podrían eliminarse. Se repitió iterativamente el NMDS si se observaba incremento en la eficiencia de la ordenación. Se continuó este proceso iterativo hasta observar que la reducción del ruido estadístico por extracción de variables superfluas se estabilizaba.

RESULTADOS

Del total de distancias observadas, fue la de Horn (=Bray-Curtis en su forma binaria) la que mostró mejores ajustes según la prueba gráfica (Figura 1).

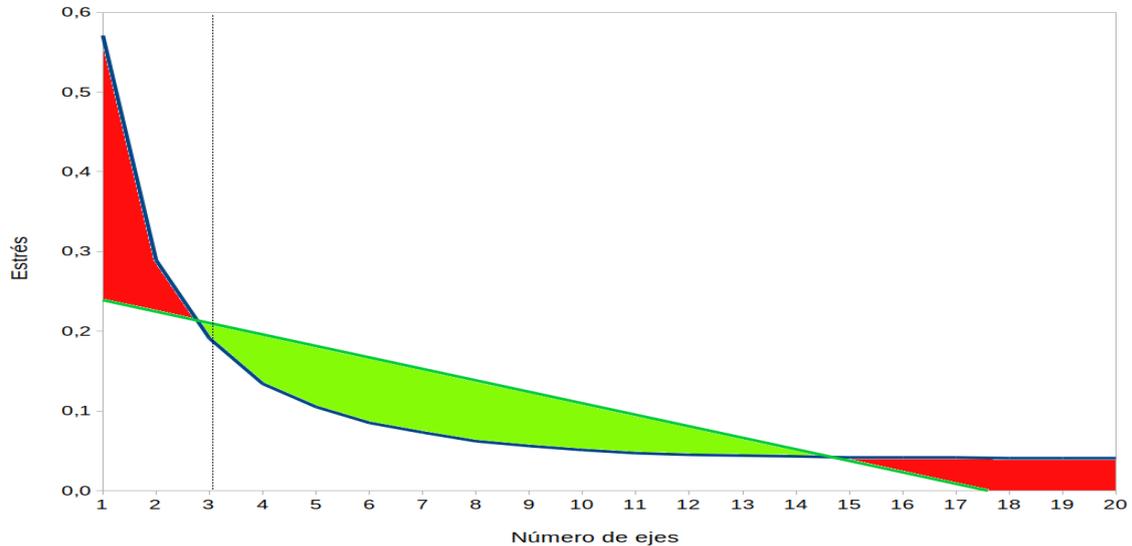


Figura 1. Reducción del estrés estadístico en base al número de ejes para la distancia de Horn sobre datos binarios. Rojo: estrés por encima de la tendencia lineal. Verde: estrés menor que la tendencia. Se seleccionan tres ejes como el número mínimo en el que el estrés baja de la tendencia.

El diagrama de Shepard (Figura 2) mostró un R^2 de 0,974 para la regresión monotónica entre disparidades ($d_{i,j}$) y proximidades ($\delta_{i,j}$), que es altamente satisfactorio. El modelo muestra una escasa curvatura y el re-escalado que se produce al pasar a rangos las distancias de la ordenación (escalones en la Figura 2), siendo no excesivo. Por ello, si asumiéramos un ajuste métrico o lineal, la regresión daría un valor menor de 0,824 y también altamente confiable. Al analizar la contribución en porcentaje de cada pregunta del cuestionario al estrés global, queda claro que preguntas son prescindibles (Figura 3).

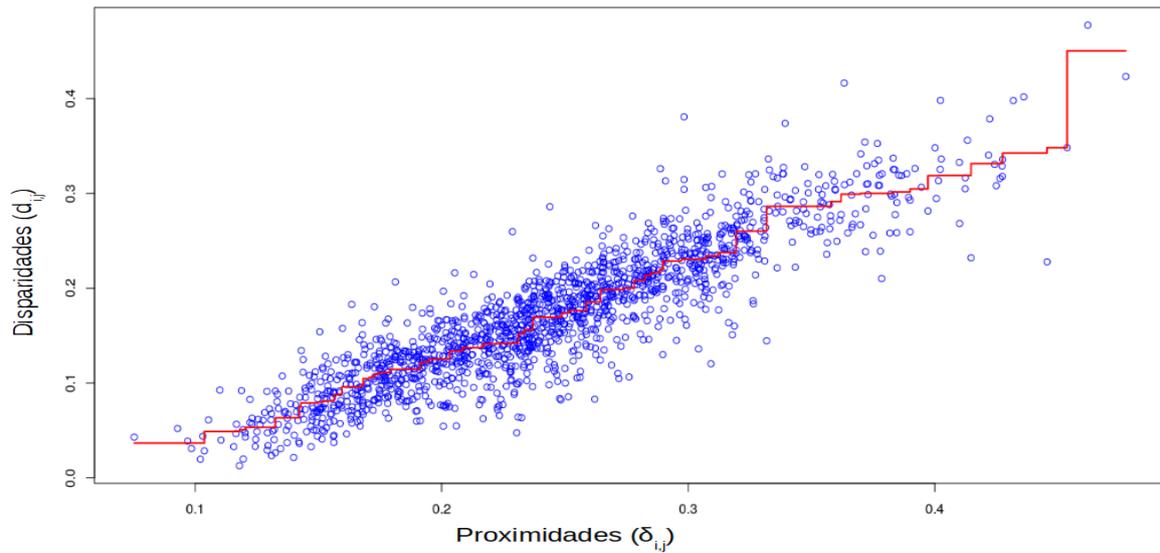


Figura 2. Diagrama de Shepard que determina la relación entre proximidades ($\delta_{i,j}$) y disparidades ($d_{i,j}$) en la ordenación.

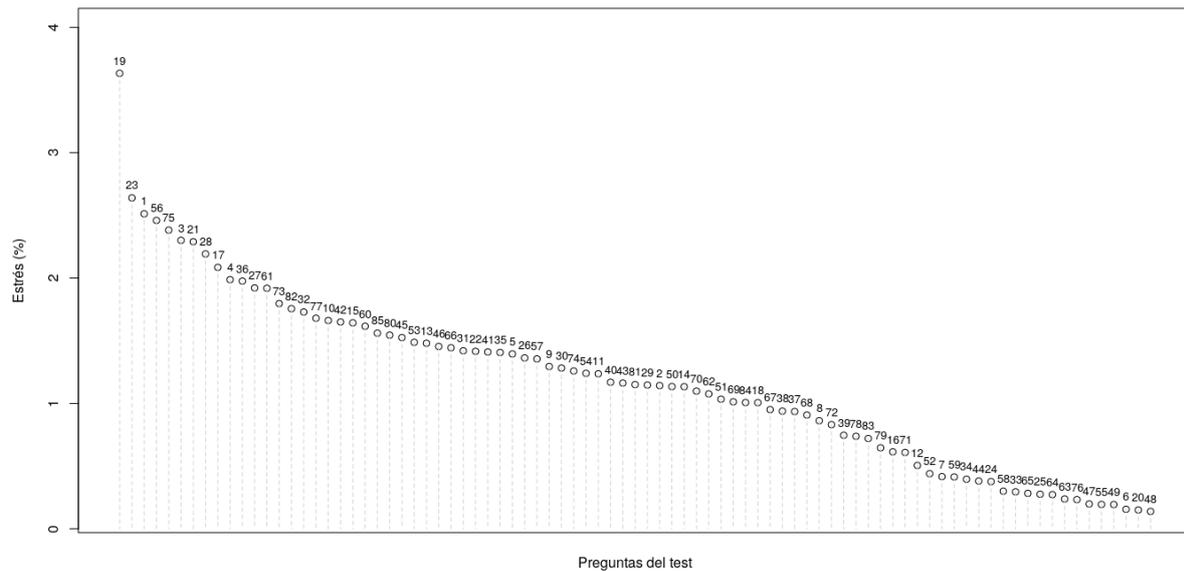


Figura 3. Porcentaje de estrés estadístico de cada pregunta del test en el análisis multivariante.

Al suprimir las preguntas 19 y 23, se obtuvieron similares ajustes por R^2 del modelo NMDS. La supresión de más preguntas no aportó mejoras al modelo. La siguiente pregunta en estrés es la número 1, que no puede ser suprimida ya que forma parte central del estudio.

Finalmente, se realizó la ordenación (Figura 4), indicando con claridad que el eje 1 del NMDS separa el efecto de la encuesta sobre el comportamiento de los encuestados ($r=0,14$; $p<0,05$).

Los ejes 2 y 3 tienen efectos similares al separar el número de abuelos vivos como un factor de importancia ($r=0,14$; $p<0,001$).

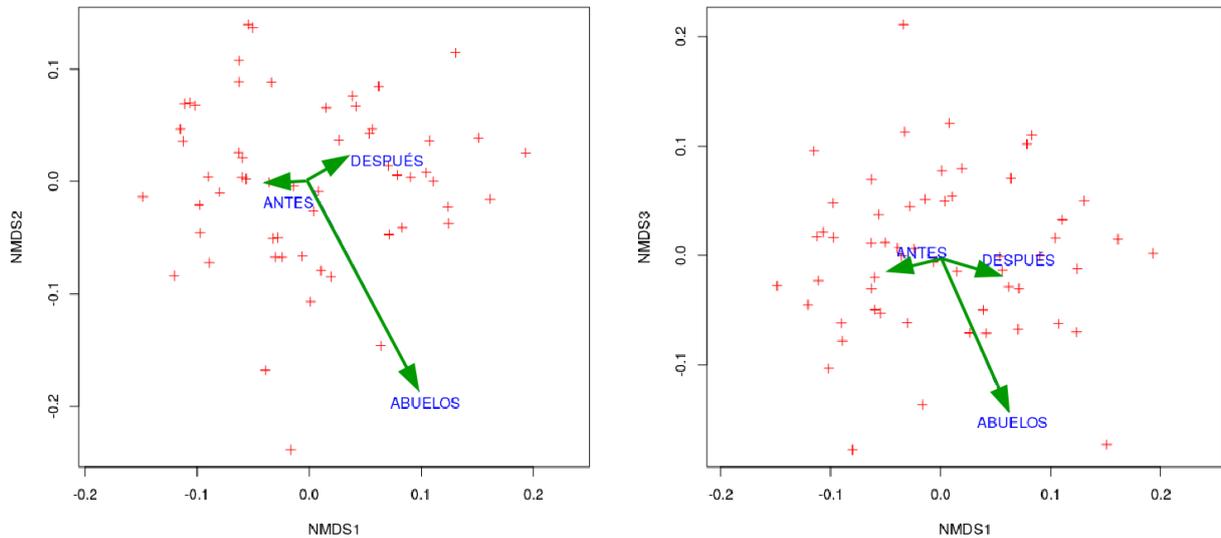


Figura 4. Resultados de la ordenación por escalamiento multidimensional no-métrico. Se muestran los ejes 1, 2 y 3.

DISCUSIÓN

El EMNM es una técnica multivariante que permite adaptarse a cualquier tipo de datos (Borg y Groenen, 2005). Esto incluye datos binarios, frecuencias o variables continuas (Legendre y Legendre, 2012). También, es posible mezclar variables con rangos o tipologías diferentes. La ordenación no-métrica (basada en rangos) nos da posibilidades adicionales de cara a extraer los gradientes principales. Asimismo, el poder usar diferentes distancias matemáticas y un número variable de ejes nos da una gran versatilidad (Cuadras, 1989).

No creemos que pueda recomendarse el uso de una distancia concreta por sus propiedades y el tipo de datos (Gower y Legendre, 1986). Es más, es posiblemente un error el asumir que las distancias matemáticas tienen resultados similares cuando se aplican a los mismos tipos de datos. Creemos que son muchos los factores que pueden hacer que una distancia matemática sea superior a otra para un tipo de datos concreto. La aproximación más apropiada en este sentido debe ser empírica, es decir, usar la distancia que menos estrés produzca, ya que al final se trata de eso, de explicar de manera parsimoniosa una matriz de datos. En este sentido, el EMNM es el método multivariante que mejor explica casi cualquier fuente de datos (Kenkel y Orloci, 1986). Como contrapartida, no podemos explicar los ejes multivariantes en términos de varianza. Sin embargo, esto puede ser obviado si se consigue un valor bajo de estrés estadístico. En este sentido, un 97.4% de explicación total podemos considerarlo altamente satisfactorio. Este valor resulta incluso más sorprendente si consideramos que estamos tratando con datos

procedentes de opiniones de personas y que estas pueden ser altamente variables (Abascal y Franco, 2002).

En la encuesta de ejemplo queda claro que los resultados son esperanzadores, pero se necesita más tamaño muestral y suprimir unas pocas preguntas para aumentar la significación de los ejes. Del mismo modo que se evaluó el estrés de las preguntas, debe evaluarse el estrés aportado por los encuestados cuando se disponga de más datos. Esto puede ayudarnos a determinar singularidades en las respuestas que aporten ideas para explorar nuevas variables.

El enfoque metodológico que proponemos está altamente optimizado y ha sido usado en contextos muy diferentes. En este sentido queremos llamar la atención sobre la similitud de técnicas de investigación entre las Ciencias Sociales y ciertas ramas de las ciencias ambientales, como es la Ecología Numérica. Creemos que son múltiples las posibles sinergias que pueden explorarse en ambos sentidos. No podemos olvidar que el EMNM es una técnica procedente de la psicometría y que tradicionalmente ha sido obviada en otros muchos campos de investigación (Borg y Groenen, 2005).

REFERENCIAS BIBLIOGRÁFICAS

- Abascal, E., Franco, M.A.** (2002). Análisis textual de encuestas: Aplicación al estudio de las motivaciones de los estudiantes en la elección de su titulación. *Metodología de Encuestas* 4(2), 195-209.
- Arriza, M.** (1984). *Guía práctica de análisis de datos*. Consejería de Educación. Junta de Andalucía, España, 200 pp.
- Borg, I., Groenen, P.J.F.** (2005). *Modern multidimensional scaling. Theory and applications*. Springer, 612 pp.
- De Leeuw, J., Mair, P.** (2009). Multidimensional Scaling Using Majorization: SMACOF in R. *Journal of Statistical Software*, 31(3), 1-30.
- Choi S.S., Cha S.H., Tappert C.C.** (2010) A Survey of Binary Similarity and Distance Measures. *Systemics, Cybernetics and Informatics* 8 (1): 43-48
- Closas, A.H., Arriola, E.A., Kuc, C.I., Amarilla, M.R., Jovanovich, E.C.** (2013). Análisis multivariante, conceptos y aplicaciones en Psicología Educativa y Psicometría. *Enfoques* 25(1), 65-92.
- Cuadras M.C.** (1989). *Distancias estadísticas*. Estadística española 30 (119), 295-378
- Goslee, S.C., Urban, D.L.** (2007). The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software* 22(7), 1-19.
- Gower, J.C., Legendre, P.** (1986). Metric and euclidean properties of dissimilarity coefficients. *Journal of Classification* 3, 5-48.

- Herrera, A. (2000).** *La clasificación numérica y su aplicación en la ecología*. Instituto Tecnológico de Santo Domingo, 89 pp.
- Hout C.M., Papesh H.M., Goldinger D.S. (2013).** *Multidimensional scaling*. Wiley Interdiscip Rev Cogn Sci 4(1), 93-103.
- De Leeuw, J., Mair, P. (2009).** Multidimensional scaling using majorization: SMACOF in R. *Journal of Statistical Software*, 31(3), 1-30.
- Kenkel, N.C., Orloci, L. (1986).** Applying metric and non-metric multidimensional scaling to ecological studies: some new results. *Ecology* 67(4), 919-928.
- Khan, A., Jana, M., Bera, S, Das, A. (2017).** Subject choice in educational data sets by using principal component and procrustes analysis. *Modeling Earth Systems and Environment* 2(4), 1-5.
- Kraemer, D.J. (2005).** Why studying standardized tests with our students is important. *The English Journal*, 94(4), 88-92.
- Legendre, P., Legendre, L. (2012).** *Numerical ecology*. Elsevier, 1006 pp.
- Minchin, P.R. (1987).** An evaluation of the relative robustness of techniques for ecological ordination. *Vegetatio* 69: 89-107.
- Oksanen, J., Blanchet, G., Friendly, M., Kindt, R., Legendre, P., McGlenn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H. (2017).** *Vegan: Community Ecology Package*. R package version 2.4.2 <https://CRAN.R-project.org/package=vegan>.
- R Core Team (2017).** *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Rencher, A.C. (2002).** *Methods of multivariate analysis*. Ed. Wiley InterScience, Canadá, 727 pp.
- Salmaso, N. (1996).** Seasonal variation in the composition and rate of change of the phytoplankton community in a deep subalpine lake (Lake Garda, Northern Italy) . An application of nonmetric multidimensional scaling and cluster analysis. *Hydrobiologia* 337, 49-68.
- Tavsancil, E., Erdem, D., Yalcin, N., Yildirim, O, Bilican, S. (2010).** Examination of data analyses used for master's theses in educational sciences. *Procedia Social and Behavioral Sciences* 9, 1467-1474