

Escalamiento Multidimensional y Mapas Auto-organizados para visualizar el uso de los Métodos Estadísticos no paramétricos en la rama de las Ciencias Agraria y Biológica

Magaly Herrera Villafranca
Caridad W. Guerra Bustillo
Verena Torres Cárdenas
Yeter Caraballo Pérez

Con el objetivo de aplicar el Escalamiento Multidimensional y los Mapas Auto-organizados para visualizar las temáticas abordadas y los Métodos no paramétricos en la rama de las Ciencias Agrarias y Biológicas se revisaron 1418 artículos científicos referidos a estos temas y contenidos en las revistas ubicadas en base de datos internacionales. Con la técnica de escalamiento multidimensional PROXSCAL y el software estadístico Viscovery SOMine, 5.0.2., permitieron visualizar 30 artículos de pastos, 48 en nutrición de bovinos, 32 estudios aplicados a pruebas de medicamentos en humanos y 28 enfocados a la salud animal. Los Métodos Estadísticos no paramétricos más empleados fueron, las dójimas X^2 con 69 aplicaciones, Mann Whitney (M-W) con 57, Kruskal-Wallis (K-W) con 54 y Exacta de Fisher (PE Fisher) con 34. Se sugiere continuar investigando en revistas de alto índice de impacto, relacionadas con las diversas ramas de investigación del ICA, el empleo de los Métodos Estadísticos no paramétricos y ampliar el uso de las herramientas utilizadas en este estudio.

Palabras clave: Stress bruto Normalizado, coeficiente de congruencia de Tucker, Estadístico no paramétrico

RESUMEN

ABSTRACT

With the aim of applying the Multidimensional Scaling and Self-organized Maps in order to visualize the topics addressed and the non parametric methods in the fields of agricultural and biological sciences, a total of 1418 scientific articles referred to these topics and contained in journals from international database were checked. Using the PROXSCAL Multidimensional Scaling Technique and the statistical software known as Viscovery SOMine 5.0.2., it was possible to visualize 30 articles on grasslands, 48 on bovine nutrition, 32 studies applied to medicine trials in humans and 28 focused on animal health. The most widely used non parametric statistical methods were the following: the X^2 sampling with 69 applications, Mann Whitney (M-W) with 57, Kruskal-Wallis (K-W) with 54 and Fisher's Exact (PE Fisher) with 34. It is suggested to keep on researching high impact journals related to the diverse research branches of the ICA, the use of the non parametric statistical methods and to extend the use of the tools used in this study.

Keywords: Gross standardized stress, Tucker congruency coefficient, non parametric statistical method.

Introducción

En la actualidad existe gran cúmulo de información que aparecen en libros, revistas, bases de datos y otros soportes, sin embargo al encontrarse tan dispersa, resulta muy engorrosa la búsqueda e identificación de diferentes temas que son importantes para continuar su investigación.

En este sentido se han desarrollado herramientas de análisis que agrupan estas informaciones por temática y otras que permiten visualizar este conocimiento. Estas herramientas se conocen con el nombre de Escalamiento Multidimensional (EMD) y Mapas Auto-organizados.

La técnica de Escalamiento Multidimensional, parte de la matriz de disimilitudes entre objetos o individuos como distancias entre puntos en un espacio de dimensión reducida para construir un mapa de menor dimensión, que reproduzca lo más fielmente posible las disimilitudes originales (Cuesta, 2005).

Los Mapas Auto-organizados han sido de utilidad para descubrir y profundizar en el análisis de datos. Se han empleado en el análisis bibliométrico de grandes volúmenes de información, contenida en bases de datos Sotolongo, Guzmán, Carrillo (2001), además, se han empleado en estudios de la ciencia de la salud humana para el mapeo de informaciones cerebrales.

En las Ciencias Agropecuarias Caraballo, Mesa y Herrera (2009) realizaron un estudio en el Instituto de Ciencia Animal (ICA) relacionado con la aplicación de los Mapas Auto-organizados en centros de investigación agropecuarios con el objetivo de ofrecer una panorámica acerca de la gestión del conocimiento y de sus procesos estratégicos, al llamarlos mapas de conocimiento, entre sus principales resultados se destaca que la gestión del conocimiento es la base para llevar a cabo los demás procesos de gestión del capital intelectual y la adquisición del aprendizaje de la institución mediante un proceso de mejora continua.

El objetivo de esta investigación fue aplicar las técnicas de Escalamiento Multidimensional y los Mapas Auto-organizados para identificar las temáticas más abordadas, así como el uso de los Métodos Estadísticos no paramétricos en investigaciones agrarias y biológicas, en el período 2000-2007.

Materiales y métodos

Se revisaron 1418 artículos para conocer las temáticas más abordadas, así como los Métodos Estadísticos no paramétricos en revistas científicas correspondientes a las Ciencias Agrarias y Biológicas, pertenecientes al fondo bibliotecario del ICA, las revistas revisadas en el período 2000-2007 están contenidas en Bases de datos internacionales y en la Web of Science.

- Cuban Journal of Agricultural Science
- Revista Ciencias Biológicas (CENIC)
- Pastos y Forrajes (EPFIH)
- Agrociencia (México)
- Veterinaria de México (México)
- Archivos de Zootecnia (España)
- Ciencias Pecuarias (Colombia)

Los resultados se tabularon en una base de datos y las variables analizadas fueron: nombre de la revista, número de los artículos por revista, nombres de los autores, tipo de variables

analizadas (continuas, binomiales, ordinales y nominales), métodos estadísticos empleados y temáticas tratadas.

Para procesar la información por revista se empleó el método de Escalamiento Multidimensional PROXSCAL (en inglés Multidimensional Sealing of Proximity) utilizando el software estadístico SPSS versión 16.0 (2007), con esta técnica se obtuvieron las medidas de calidad de ajuste de la configuración de los datos obtenidos de las revistas por temática y según Métodos Estadísticos no paramétricos. Las medidas de la calidad de ajuste utilizadas fueron el coeficiente de esfuerzo (Stress (S), y Stress bruto normalizado), el coeficiente de congruencia de Tucker y la dispersión explicada (DAF, en inglés). Se considera un buen ajuste cuando el resultado obtenido del Stress se encuentra próximo a 0 y los valores del DAF y el coeficiente de congruencia de Tucker son de 0,8 en lo adelante lo cual explica la dispersión obtenida del análisis y que al multiplicar por 100 su resultado se expresan en %. Esta técnica se basa en la comparación de objetos o de estímulos, de forma que sí un individuo juzga a los objetos A y B como los más similares, entonces las técnicas de Escalamiento Multidimensional colocarán a los objetos A y B en el gráfico de forma que la distancia entre ellos sea más pequeña que la distancia entre cualquier otro par de objetos.

Estos indicadores los reportan la bibliografía clásica del tema, los primeros estudios fueron realizados por Torgerson en 1952 quien introdujo el término y esbozó las primeras ideas, posteriormente son aplicados por diferentes autores entre los que se encuentran (Mora y González, 2009).

Para visualizar la información por revista se utilizó el software Viscovery SOMine, (2008), y el método SOM-Wald Clusters, que combina la ordenación de información local del mapa con el algoritmo de Cluster Jerárquico Clásico de Ward (Ward, 1963).

Este método permite agrupar los elementos más representativos de lo general a lo particular y une aquellos con valores similares en cuanto a cercanía hasta formar un grupo llamado cluster, (conglomerados) lo que es representado en un mapa.

Del análisis se obtienen dos mapas, en los que se observan las temáticas y los Métodos Estadísticos no paramétricos usados en las revistas estudiadas. La información se organiza

por cluster donde la cantidad de ellos varía según los datos del estudio. Se ofrece en la parte inferior de cada mapa una escala con matices de colores, del más intenso al menos intenso, que muestra de derecha a izquierda, en orden decreciente las frecuencias (cantidad de investigaciones) por temática y Métodos Estadísticos no paramétricos.

Resultados y discusión

En la tabla 1 se presentan los resultados de bondad de ajuste de las temáticas abordadas por revistas, los valores 0,01539 y 0,04395 corresponden al stress bruto normalizado y al stress (S), según su interpretación se considera como un ajuste bueno cuando los resultados se encuentran cercanos a 0, otros de los criterios para validar el estudio son el D.A.F. y el coeficiente de congruencia de Tucker con valores de 0,98 y 0,99 respectivamente, este último se refiere a la dispersión explicada por el análisis, cuyos valores se encuentran muy cercanos al 100%. Según Cuesta (2005) plantea estos coeficientes se suele multiplicar por 100, lo que facilita la interpretación en términos de porcentaje.

Tabla 1. Resultado de calidad de ajuste de las temáticas por revista.

Medida	Resultado
Stress bruto normalizado	0,01539
Stress (S)	0,04395
Dispersión explicada (D.A.F.)	0,98461
Coficiente de congruencia de Tucker	0,99228

Mora y González (2009) indican que en el Escalamiento Multidimensional, el método PROXSCAL minimiza el Stress bruto normalizado, estos valores se obtienen mediante la fórmula introducida por Kruskal en 1964 citado por Mora y González, quien ofreció una guía para su interpretación, valores cercanos a cero se consideran excelentes y por encima de 0,2 muy pobre el ajuste de los datos.

Al tener en cuenta los criterios de bondad de ajuste, se considera que la solución en dos dimensiones es apropiada lo que se muestra en la figura 1. En la primera dimensión se forman dos grupos, en el primero se unen aquellas revistas que tienen un comportamiento similar en cuanto a las temáticas analizadas (Agrociencia, Archivo de Zootecnia, Ciencias Pecuarias y Pastos y Forrajes) estas revistas enfocan sus estudios a diferentes temas como pastos, estadística aplicada, monogásticos

etc., el segundo grupo las que posee un comportamiento diferente, en el caso de la RCCA enmarca sus investigaciones a los pastos y forrajes. En la dimensión dos al igual que la anterior se forman dos grupos, pero cada uno con temáticas diferentes, estas revistas son Ciencias Biológicas que relaciona sus investigaciones a pruebas de medicamentos en humanos y Veterinaria de México que estudia los temas de salud animal.

Según Miret (2005), el Escalamiento Multidimensional (EMD), está basado en procedimientos estadísticos que trabajan con medidas de semejanza o proximidades en el Análisis de Datos, también algunos autores los identifican con el nombre de Análisis de proximidades, como se muestra en la figura 1.

En el figura 2 se visualizan las temáticas más abordadas en el período analizado, a través de la técnica de los Mapas Auto-organizados. Esta permitió identificar aquellas temáticas más predominantes en las revistas. En Agrociencia se destacan los Métodos Estadísticos aplicados a la Ciencia Agraria, las revistas Archivos de Zootecnia y Ciencia Pecuaria enmarcan sus estudios en el tema de reproducción animal y las Ciencias Biológicas a pruebas de medicamentos en humanos. En Pastos y Forrajes y RCCA los temas más estudiados se relacionan con los Pastos y Forrajes y la Veterinaria de México se enfoca a la salud animal.

En la tabla 4 se presentan las medidas de bondad de ajuste, relacionada con la aplicación de los Métodos Estadísticos no paramétricos por las revistas analizadas. Estos representan un ajuste superior en cuanto a la configuración de los datos, en este caso las medidas de calidad del ajuste stress bruto normalizado y stress (S) presentan valores de 0,0009 y 0,00076 respectivamente, casi cero, al tener en cuenta el criterio planteado por Kruskal 1964, se considera un ajuste excelente en los datos, en cuanto a los coeficientes de dispersión y congruencia de Tucker, ambos explican casi el 100 % de la variabilidad de los mismos.

Tabla 4. Resultados de la bondad del ajuste relacionada con la aplicación de los Métodos Estadísticos no paramétricos por revistas.

Medida	Resultado
Stress bruto normalizado	0,00092
Stress (S)	0,00076
Dispersión explicada (D.A.F.)	0,99908
Coefficiente de congruencia de Tucker	0,99954

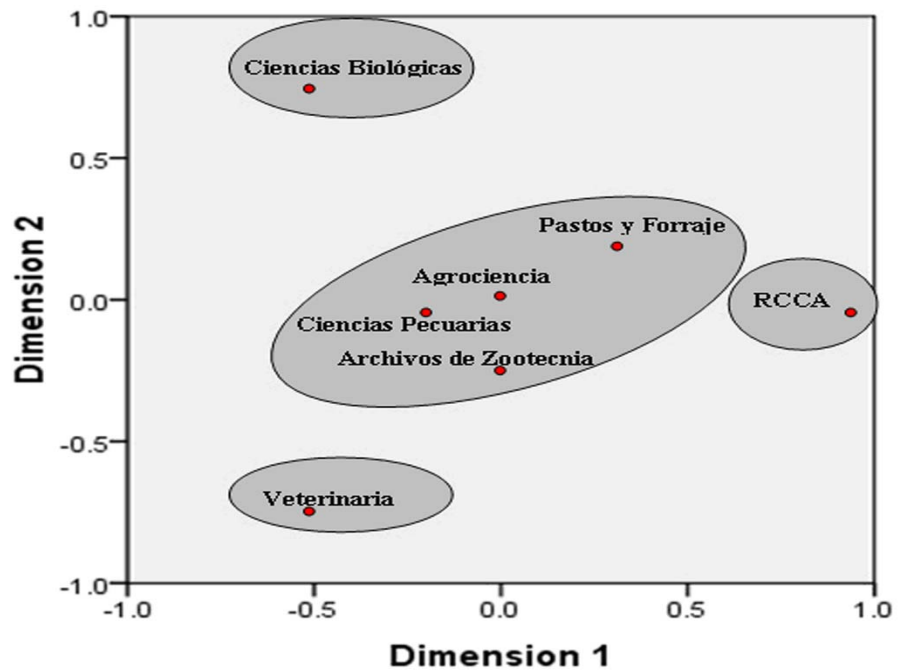
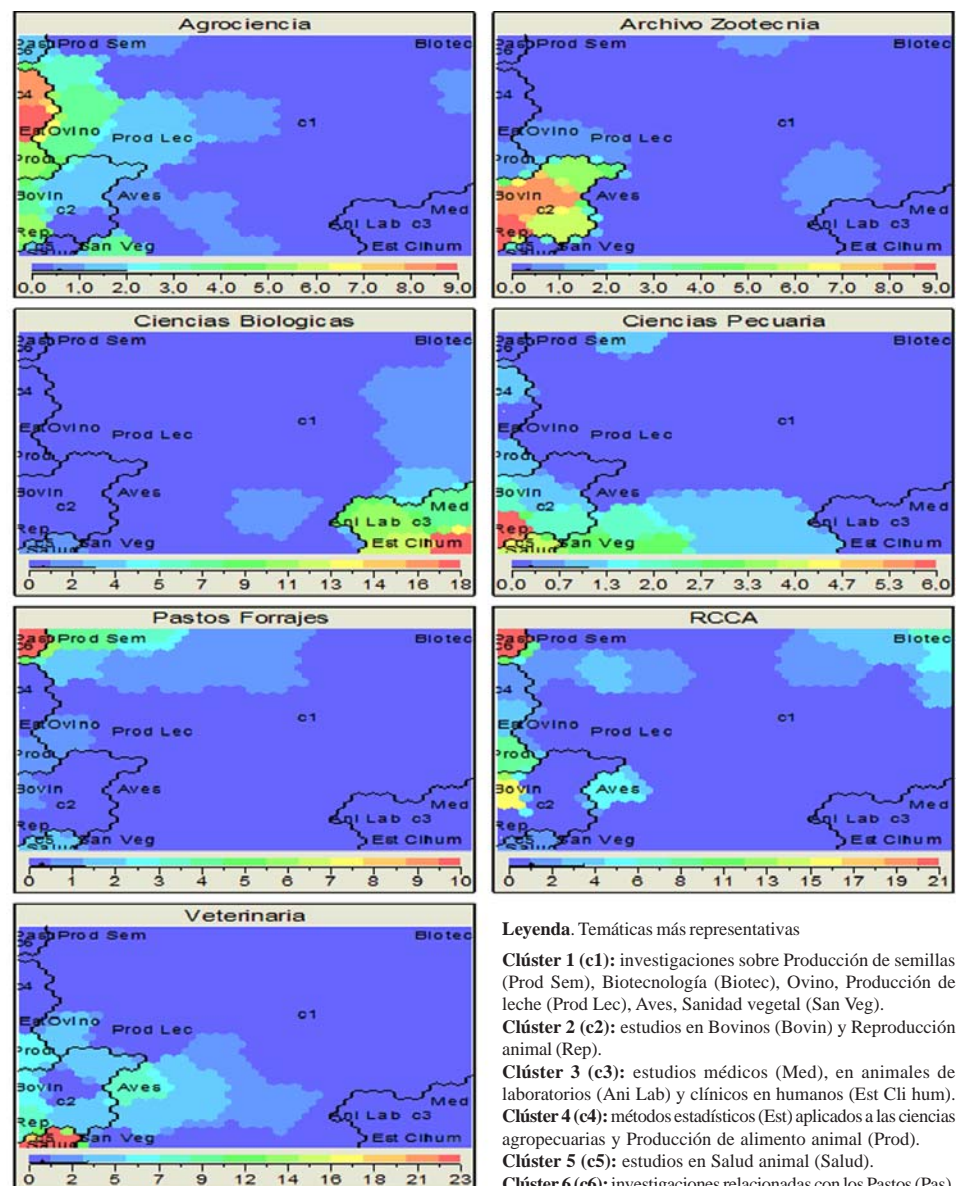


Figura 1. Gráfico en dos dimensiones de las Revistas por temáticas.



Leyenda. Temáticas más representativas

- Clúster 1 (c1):** investigaciones sobre Producción de semillas (Prod Sem), Biotecnología (Biotec), Ovino, Producción de leche (Prod Lec), Aves, Sanidad vegetal (San Veg).
- Clúster 2 (c2):** estudios en Bovinos (Bovin) y Reproducción animal (Rep).
- Clúster 3 (c3):** estudios médicos (Med), en animales de laboratorios (Ani Lab) y clínicos en humanos (Est Cli hum).
- Clúster 4 (c4):** métodos estadísticos (Est) aplicados a las ciencias agropecuarias y Producción de alimento animal (Prod).
- Clúster 5 (c5):** estudios en Salud animal (Salud).
- Clúster 6 (c6):** investigaciones relacionadas con los Pastos (Pas).

En la figura 3 se observa el análisis de aquellas revistas que hacen un mayor uso de los Métodos no paramétricos, en la primera dimensión se encuentra la revista Ciencias Biológicas y la segunda dimensión se encuentra la revista Veterinaria de México que se separa del resto de las revistas que tuvieron poca aplicación de estos métodos en el período analizado.

En la figura 4, se reflejan los mapas referentes a la aplicación de los Métodos Estadísticos no paramétricos con 214 aplicaciones, en este sentido sobresale la Revista Ciencias Biológicas con las dójimas Prueba Exacta de Fisher, seguida por Mann-Whitney y Kruskal-Wallis, y en la Veterinaria la dójima X^2 .

Un estudio similar realizado en este trabajo fue publicado por De Calzadilla (1999) en Cuba, en el período 1987-1997, al aparecer 65 aplicaciones referidas a estos métodos, en 2291 artículos, en 11 revistas científicas, 6 nacionales y 5 extranjeras. La autora encontró que los métodos estadísticos no paramétricos más empleadas fueron, la prueba de Bondad de ajuste, la correlación por rango Spearman, la prueba Chi-cuadrado de independencia, la prueba Kruskal-Wallis, Mann-Whitney respectivamente, que coinciden en cierta medida con los reportados para las revistas Ciencias Biológicas y Veterinaria de México.

En el análisis realizado se evidenció que la prueba de bondad de ajuste Kolmogorov-Smirnov no ha sido muy empleada, sin embargo Siegel y Castellan (1995) y Salcines (2003) plantean que resulta más potente que la dójima Chi Cuadrado cuando las muestras son más pequeñas.

Uno de los métodos más empleados, es el Análisis de Varianza por rango de Kruskal Wallis, Balzarini, Casanoves, Rienzo, González y Robledo (2001) indican que este procedimiento fue propuesto por Kruskal Wallis en 1952 y permite comparar las esperanzas de dos o más distribuciones sin la necesidad de verificar el supuesto de que los términos de error se distribuyen normalmente.

Sin embargo, en la RCCA, se destaca el trabajo de Mazorra, Borge, Blanco, Borroto y Sorís, (2006), donde se hace referencia al empleo de la dójima de Kruskal-Wallis, porque las variables «cantidad de ramoneo» y «tiempo de ramoneo» incumplen con el supuesto de homogeneidad de varianza. En este artículo el autor emplea la dójima de rangos múltiples

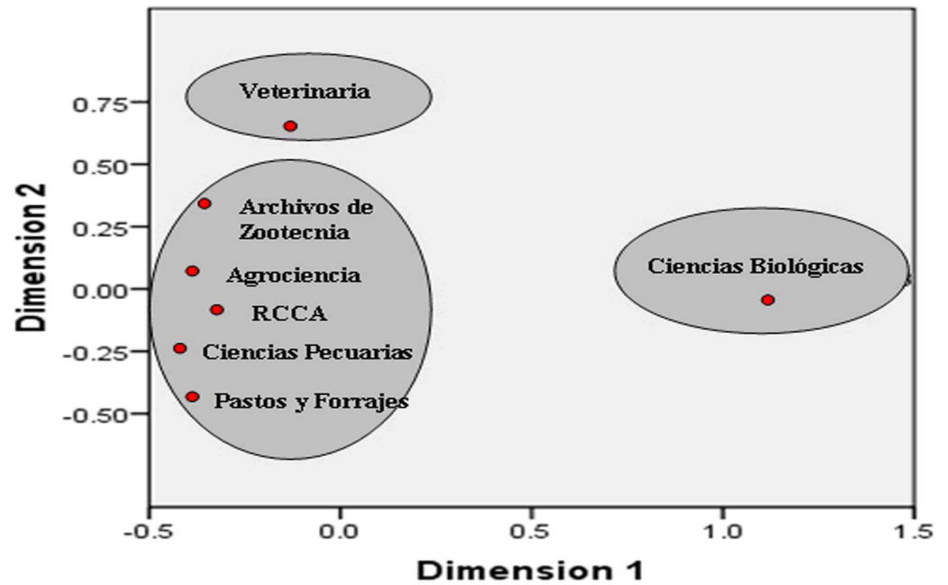
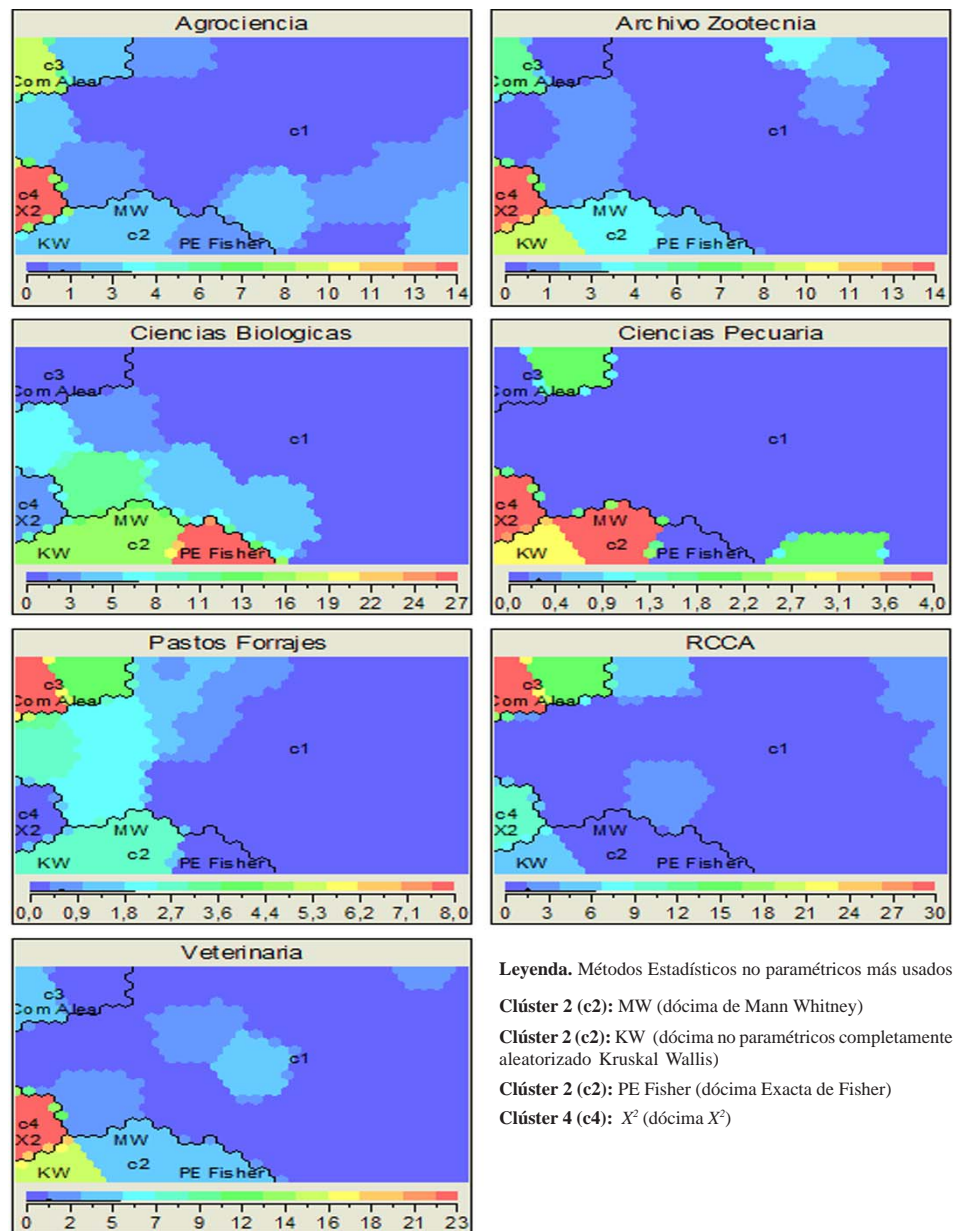


Figura 3. Gráfico en dos dimensiones de las revistas relacionadas con la aplicación de los Métodos Estadísticos no paramétricos.



Leyenda. Métodos Estadísticos no paramétricos más usados
Clúster 2 (c2): MW (dójima de Mann Whitney)
Clúster 2 (c2): KW (dójima no paramétricos completamente aleatorizado Kruskal Wallis)
Clúster 2 (c2): PE Fisher (dójima Exacta de Fisher)
Clúster 4 (c4): X^2 (dójima X^2)

propuesta por Siegel y Castellan (1995), con el objetivo de detectar diferencias significativas entre los tratamientos.

En la Revista Archivos de Zootecnia, sobresalen los trabajos de Delgado, Dieguez y Cañuelo (2000) y Revalti, Capellari y Delgado (2005), referentes a variables nominales (color de la mucosa, forma de la oreja, etc.), dichas variables fueron transformadas a ordinales, en ambos estudios aplicaron el Análisis de Varianza no paramétrico Kuskal-Wallis para diseño completamente aleatorizado.

Relacionados con la d6cima Mann-Whitney se encuentran los trabajos de Carvajal, Arruzazabala, Molina y Vald6s (2001) y Beiro, Hern6ndez, Meri6no, 6lvarez y Quintero (2001), de la Revista Ciencias Biol6gicas, los cuales estudian las variables «tiempo de supervivencia» y «actividad antitumoral», despu6s de aplicados los tratamientos.

Resulta importante destacar el empleo de la d6cima de la Probabilidad exacta de Fisher, que es sumamente 6til para analizar datos discretos, nominales u ordinales, cuando las dos muestras independientes son peque6as. Los estudios realizados por Bacard6 y cols. (2004) en la Revista Ciencias Biol6gicas, eval6a el grado de toxicidad de un estimulador de colonias de granulocitos, y G6mez y cols. (2005), que estudia las «frecuencias de muertes, aparici6n de signos y lesiones cuando es aplicado el tratamiento D-003», en este sentido se observa que las variables analizadas est6n en escalas nominal y ordinal.

Aunque no fue objetivo de este trabajo analizar el empleo de los M6todos Estadísticos param6tricos, se puede indicar que en la RCCA y Pastos y Forrajes se hace un mayor empleo de estos, en lo fundamental en los An6lisis de Varianza asociados a los Dise6os Completamente Aleatorizado y Bloque al Azar.

Conclusiones

1.- El uso del Escalamiento Multidimensional y los Mapas Auto-organizados, resultaron muy 6tiles para el procesamiento, visualizaci6n y an6lisis de la informaci6n, al permitir una adecuada identificaci6n de las tem6ticas m6s abordadas, as6 como el empleo de los M6todos Estadísticos no param6tricos por las revistas analizadas.

2.- Las tem6ticas m6s estudiadas fueron los pastos, nutrici6n en bovinos, estudios aplicados a pruebas para medicamentos en

humanos y salud animal y los M6todos Estadísticos no param6tricos m6s empleados, fueron las d6cimas X^2 o Chi Cuadrado, Mann-Whitney, Kruskal Wallis y Exacta de Fisher.

3.- Se observan m6s aplicaciones de los m6todos estadísticos no param6tricos, pero existe poca diversidad de ellos en investigaciones cient6ficas publicadas en revistas agrarias y biol6gicas.

Referencias

Bacard6, D., Moreno, N., Cosme, K., Porras, D., Carrera, I., Diaz, G., Su6rez, J., et al. (2004). Evaluaci6n de Toxicidad subcr6nica Del factor estimulador de colonias de granulocitos. Revista de Ciencias Biol6gicas, 35(3), 203-209.

Balzarini, M. G., Casanoves, F., Di Rienzo, J. A., Gonz6lez, L. A., Robledo, C. W. (2001). Manual del Usuario Insostat. Argentina.

Beiro, O., Hern6ndez, D., Ar6z, L., Meri6no, N., 6lvarez X. Quintero, Y. (2001). Subnitrito de bismuto en la actividad antitumoral y la nefrotoxicidad por Cisplatinio en ratones trasplantados con leucemia p.388. Revista de Ciencias Biol6gicas, 32(2), 103-106.

Caraballo, Y., Mesa, D. y Herrera, J. A. (2009). Herramientas de gesti6n del conocimiento: convergencias hacia un aprendizaje organizacional. Revista Cubana de Ciencia Agr6cola, 43(1), 1-11.

Carvajal, D., Arruzazabala, L., Molina, V., Vald6s S. (2001). Efecto de D-002 en modelos experimentales de anafilaxia. Revista Ciencias Biol6gicas, 32(2), 93-96.

Cuesta, A., (2005). An6lisis Multivariante. Dpto. Matem6tica, Estadística y Computaci6n. Universidad de Cantabria, Espa6a, 55.

De Calzadilla, J. (1999). Procedimientos de la Estadística no param6trica. Aplicaciones en las Ciencias Agropecuarias. Tesis presentada en opci6n al t6tulo de Master en Matem6tica Aplicada a las Ciencias Agropecuaria. Cuba

Delgado, J., Dieguez, E., Cañuelo, P. (2000). Caracterizaci6n exteriorista de las variedades de cerdo ib6rico basadas en característica

cuantitativas. Revista Archivos de Zootecnia, 49(185-186), 201-207.

G6mez, R., M6s, R., Noa, M. Garc6a, H., Felip, E., Goicochea, E, et al. (2005). Carcinog6nesis del D-003 en ratas Sprague Dawley. Reporte de 1 a6o Revista Ciencias Biol6gicas, 36(3), 205-211.

Linares, G. (2001). Escalamiento multidimensional. Conceptos y enfoques. Revista Investigaci6n Operacional, 22(2), 173-183.

Mazorra, C., Borge, G., Blanco, M., Borroto, A., Soris, A., L. (2006). Influencia de la dosis de cloruro de litio en la conducta de ovinos condicionados que pastorean en plantaciones de c6tricos. Revista Cubana de Ciencia Agr6cola, 40(4), 425-431.

Miret, E. (2005). Un enfoque unificado para t6cnicas de representaci6n euclidiana. Tesis para optar por el grado de Doctor en Ciencias Matem6ticas. UH. La Habana. Cuba. 148p.

Mora, M. y Gonz6lez, M.I. (2009). Caracterizaci6n del fracaso empresarial en Colombia: Un enfoque con Escalado Multidimensional. Universidad de Salamanca. Espa6a. Documento de Trabajo 11/09. 37.

Revalti, M. A., Capellari, A., Delgado, J. V. (2005). Recurso gen6tico porcino aut6ctono en el nordeste de la republica de Argentina. Revista Archivos de Zootecnia, 54 (205), 97-100

Salcines, J. (2003). La normalidad en el gasto de los consumidores: Un an6lisis de los hogares de Espa6a [Versi6n electr6nica], Revista Momento Econ6mico S/V (126), 2-14.

Siegel, S y Castellan, N.J. (1995). Estadística no param6trica aplicada a las Ciencias de la Conducta. Cuarta edici6n. Editorial Trillas, M6xico.

Sotolongo, G., Guzm6n, M. V., Carrillo, H. (2002). Vibliosom: visualizaci6n de informaci6n bibliom6trica mediante el mapeo autoorganizado. Revista Espa6ola de Documentaci6n Cient6fica, 25(4), 477-487.

Visauta, B. (2007). SPSS para Window versión 16.0.

Viscovery Software GmbH. (2008).
Viscovery SOMine 5.0.2.
[Software de cómputo].
Viscovery Software GmbH.

Ward, J. (1963). Hierrarchical grouping to optimize an objective function. *Jornal of the Amarican Statistical Association.*, 58(4), 236-244.

Recibido: 5 de julio de 2011.
Aprobado en su forma definitiva:
11 de enero de 2012

Lic. Magaly Herrera Villafranca
Instituto de Ciencia Animal
País: Cuba
Correo electrónico: <mvillafranca@ica.co.cu>

Dra.C. Caridad W. Guerra Bustillo
Instituto de Ciencia Animal
País: Cuba
Correo electrónico: <mvillafranca@ica.co.cu>

Dra.C. Verena Torres Cárdenas
Instituto de Ciencia Animal
País: Cuba
Correo electrónico: <vtorcar@yahoo.es>

Ms.C. Yeter Caraballo Pérez
Instituto de Ciencia Animal
País: Cuba
Correo electrónico: <mvillafranca@ica.co.cu>
