

Procedimientos de Escalamiento Multidimensional y Mapas Auto Organizados para visualizar información bibliométrica relacionada con Métodos Estadísticos no paramétricos

Magaly Herrera Villafranca¹, Caridad W. Guerra Bustillo², Verena Torres Cárdenas¹, Yeter Caraballo Pérez¹

¹ *Instituto de Ciencia Animal, Km 47½ Carretera Central, San José de las Lajas.*

² *Universidad Agraria de La Habana, Autopista Nacional y Carr. De Tapaste San José de Las Lajas, La Habana*

Introducción

En la actualidad existe un gran cúmulo de información que aparecen en libros, revistas, bases de datos etc., y que serviría como base para dar origen a otras investigaciones, sin embargo al encontrarse tan dispersa, resulta muy engorrosa la búsqueda e identificación de diferentes temas que se desea investigar según el interés del investigador. En este sentido se han desarrolla 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, para construir un mapa de la menor dimensión posible, que reproduzca lo más fielmente posible las disimilitudes originales (Cuesta, 2005)

Las primeras ideas para la aplicación de esta técnica surgieron a mitad del pasado siglo con un trabajo realizado por Torgerson en 1952, quien introdujo las primeras ideas sobre el tema (Linares, 2001).

Los Mapas Auto organizados han sido de utilidad para descubrir y profundizar en el conocimiento de información. Se han empleado en el análisis bibliométrico de grandes volúmenes de información y contenido en las bases de textos, además han sido de gran utilidad en estudios médicos para el mapeo de informaciones cerebrales.

Miret (2005) plantea que dentro de las Redes Neuronales Artificiales, los Mapas Auto-Organizados de Kohonen (SOM del inglés Self Organizing Maps) se caracterizan por su capacidad para clasificar un conjunto complejo de patrones de manera no supervisada.

El objetivo de esta investigación fue aplicar las técnicas de Escalamiento Multidimensional y los Mapas Auto Organizados para analizar el empleo de los métodos estadísticos no

paramétricos en investigaciones científicas publicadas en revistas agropecuarias, en el período del 2000 al 2007.

Materiales y métodos

Se revisaron 1418 artículos, de revistas científicas pertenecientes al fondo bibliotecario del Instituto de Ciencia Animal, las revistas revisadas en el período 2000-2007 fueron:

Nacionales:

- Revista Cubana de Ciencia Agrícola (RCCA).
- Revista Ciencias Biológicas (CENIC).
- Revista de Pastos y Forrajes (EPFIH).

No nacionales

- Agrociencia (México)
- Veterinaria de México (México)
- Archivos de Zootecnia. (España)
- Ciencias Pecuarias (Colombia)

Los resultados fueron tabulados en una base de datos analizándose las variables: nombre de la revista, número de los artículos por revista, nombres de los autores, tipo de variables analizadas (cuantitativas y cualitativas), 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 para el agrupamiento de las revistas por temática y según Métodos no paramétricos. Las medidas de la calidad de ajuste utilizadas fueron el coeficiente de esfuerzo (Stress, 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 cero y los valores del DAF y el coeficiente de congruencia de Tucker son de 0,8 en adelante.

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, lo que es representado en un mapa.

Del análisis se obtiene dos mapas, en estos 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. Para un mejor entendimiento se ofrece en la parte inferior una leyenda que explica las abreviaturas usadas y una escala de matices de colores, del color más intenso al menos intenso, que muestra de derecha a izquierda, en orden decreciente las frecuencias por temática.

En el procesamiento de los datos se trabajó con los valores originales, debido a las características que deben tener los datos de entrada para que sean procesados por el software.

Resultados y discusión

La tabla 1 y 2 y la figura 1 muestran los resultados obtenidos, a partir de la aplicación del método de Escalamiento Multidimensional (PROXSCAL).

En la tabla 1 se observan los resultados de la calidad del ajuste, dado por el Stress bruto normalizado y el SStress, que son bajos, por lo que se consideran un ajuste adecuado de los datos.

Tabla 1. Resultado de calidad de ajuste para las temáticas por revistas.

Medida	Resultado
Stress bruto normalizado	0,01539
S Stress	0,04395
Dispersión explicada (D.A.F.)	0,98461
Coeficiente 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, quien ofreció una guía para su interpretación.

Otros de los criterios para validar los resultados obtenidos son el D.A.F. y el coeficiente de congruencia de Tucker los que se encuentran muy cercanos al 100%, dando validez a la configuración obtenida con esta información. Cuesta (2005), indica que se suele multiplicar por 100 el Stress obtenido, lo que facilita la interpretación en términos de porcentaje.

En la tabla 2, se ubican en dos dimensiones los pesos o ponderaciones de las revistas con las temáticas de aplicación, en esta tabla se muestra la distancia que existe entre las revistas analizadas por diferentes temáticas, al indicar el grado de similitud entre éstas, de esta forma, las más similares son aquellas que presentan sus valores más próximos. Los

valores pequeños muestran los individuos que más se parecen, mientras que los valores grandes señalan diferencias entre estos. Entre las revistas que más sobresalen por dimensiones son, por la uno la RCCA y por la dos Ciencias Biológicas y Veterinaria de México con los valores superiores positivos y negativos respectivamente.

Tabla 2. Ubicación en dos dimensiones de las revistas por temáticas

Revistas	Dimensión	
	1	2
Agrociencia	-.015	.025
Archivo de Zootecnia	.001	-.226
Ciencias Biológicas	-.492	.856
Ciencias Pecuarias	-.199	-.034
Pastos y Forraje	.316	.188
RCCA	.939	-.061
Veterinaria	-.550	-.750

21 Según,

Guerrero y Ramírez (2002) el Escalamiento Multidimensional está basado en la comparación de objetos o de estímulos, de forma que si 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.

Los resultados obtenidos a partir de la tabla 2 se pueden graficar, estos se muestran en la figura 1. donde se forma cuatro grupos, dos en ambas dimensiones, en la primera se agrupan las revistas que presentan un comportamiento similar entre las que se encuentran, Agrociencia, Archivo de Zootecnia, Ciencias Pecuarias y Pastos y Forrajes mostrándose sus valores muy cercanos entre si, esta misma dimensión se observar que la revista RCCA se aleja de las demás con un valor diferente a los anteriores, en la segunda dimensión se encuentran las revistas Ciencias Biológicas y Veterinaria de México respectivamente, las cuales forman los otros dos grupos.

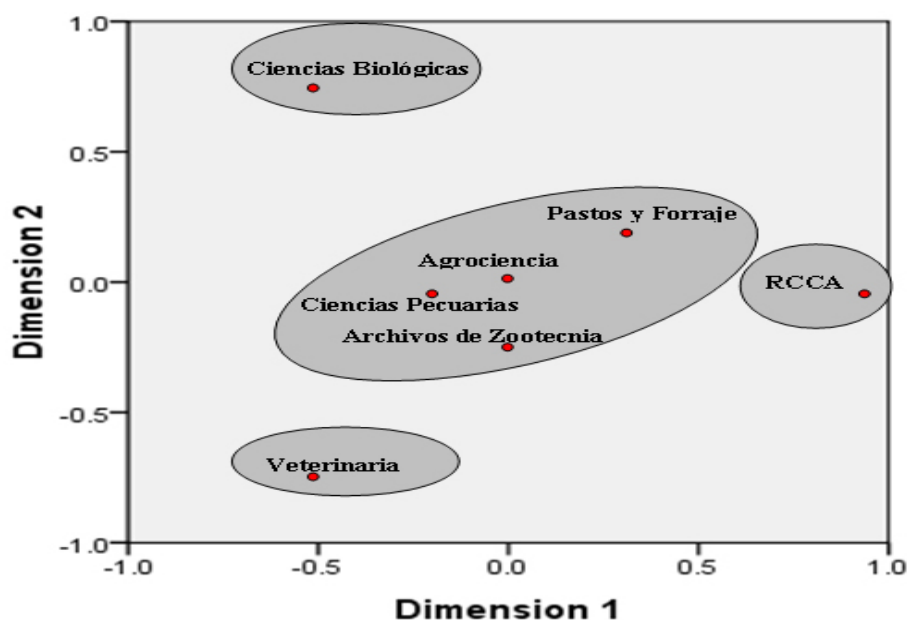


Figura 1. Gráfico en dos dimensiones de las Revistas por temáticas.

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 analizadas. En la Agrociencia se destaca la temática relacionada con los métodos estadísticos aplicados a la Ciencias agropecuarias, en el caso de la revista Archivos de Zootécnica y Ciencia Pecuaria enmarcan sus estudios al tema de Reproducción animal, Ciencias Biológicas a pruebas de medicamentos en humanos. En Pastos y Forrajes y RCCA los temas más estudiados se relacionan con los patos y forraje, haciendo un mayor estudio de la temática la RCCA, y la Veterinaria de México se enfoca a la salud animal.

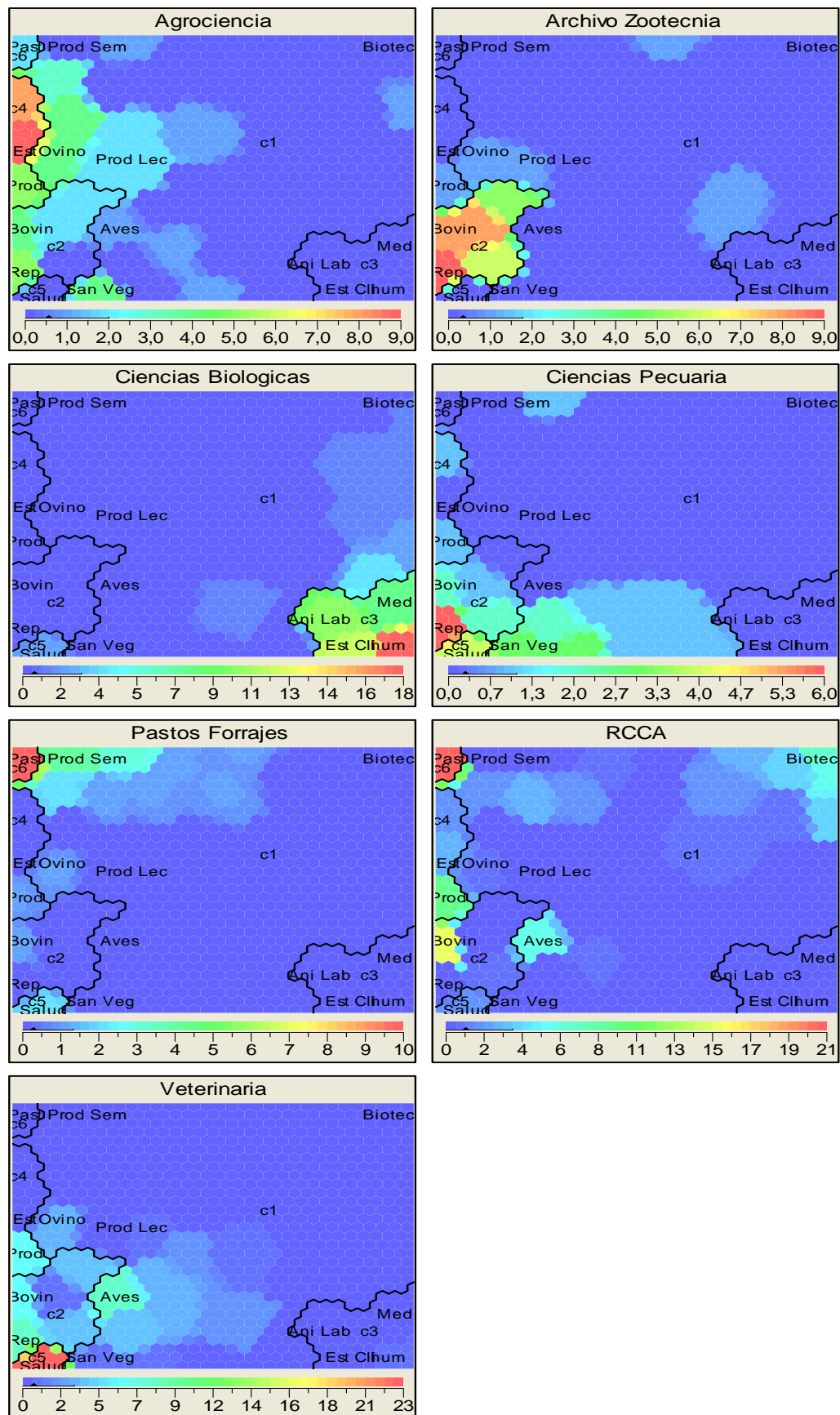


Figura 2. Mapa de temáticas más estudiadas por revistas en el período 2000 - 2007

Leyenda. Temáticas más representativas

Cluster 1 (c_1): investigaciones sobre producción de semillas (Prod Sem), Biotecnología (Bistec), Ovino, Producción de leche (Prod Lec), Aves, sanidad vegetal (San Veg).

Cluster 2 (c_2): estudio en Bovinos (Bovin) y reproducción animal (Rep).

Cluster 3 (c_3): estudios médicos (Med), en animales de laboratorios (Ani Lab) y clínicos en humanos (Est clín hum).

Cluster 4 (C_4): métodos estadísticos aplicados a las Ciencias Agropecuarias (Est) y producción de alimentos (Prod).

Cluster 5 (C_5): estudios en Salud Animal (Salud).

Cluster 6 (C_6): investigaciones relacionadas con los Pastos (Pas).

En la tabla 3 se presentan las medidas de la calidad del ajuste, con la aplicación de los métodos estadísticos no paramétricos en las revistas analizadas. Se consideran los mismos indicadores de Stress, así como los otros dos coeficientes de que ponderan la calidad de ajuste de la configuración dada, la cual resulta superior a la anterior para los indicadores.

Tabla 3. Resultado de la calidad del ajuste de los métodos estadísticos no paramétricos por revistas

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

En la tabla 4, se observa las revistas que tienen una mayor aplicación de los métodos estadísticos no paramétrica, donde se destacan las Revistas Ciencias Biológicas y Veterinaria de México, en la primera y segunda dimensión respectivamente.

Tabla 4 Ubicación en dos dimensiones de las revistas por temáticas que con mayor frecuencia abordan el empleo de los métodos estadísticos no paramétricos

Revistas	Dimensión	
	1	2
Agrociencia	-.302	.109
Archivos de Zootecnia	-.120	.169
Ciencias Biológicas	1.368	-.047
Ciencias Pecuarias	-.232	-.311

Pastos y Forrajes	-.193	-.463
RCCA	-.396	-.088
Veterinaria	-.125	.629

En la figura 3, se aprecia la formación de tres grupos, dos de estos sobresalen de manera independientes con las revistas indicadas en las dimensiones 1 y 2 respectivamente, y el tercero se forma con el resto de las revistas que tienen un comportamiento semejante.

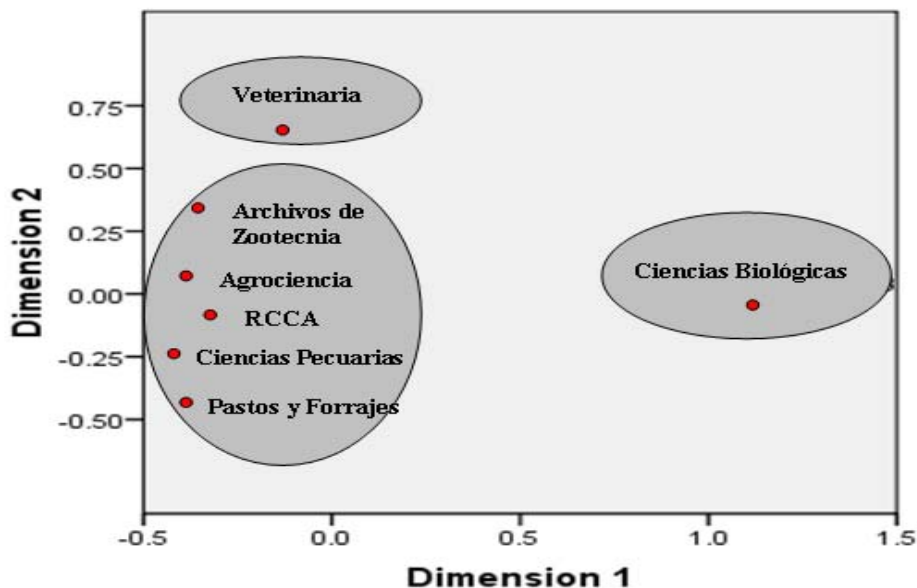


Figura 3. Gráfico en dos dimensiones de las revistas por temáticas de Estadísticas no Paramétricas.

En la figura 4, se reflejan los mapas referentes a la aplicación de los Métodos estadísticos no paramétricos, 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 χ^2 .

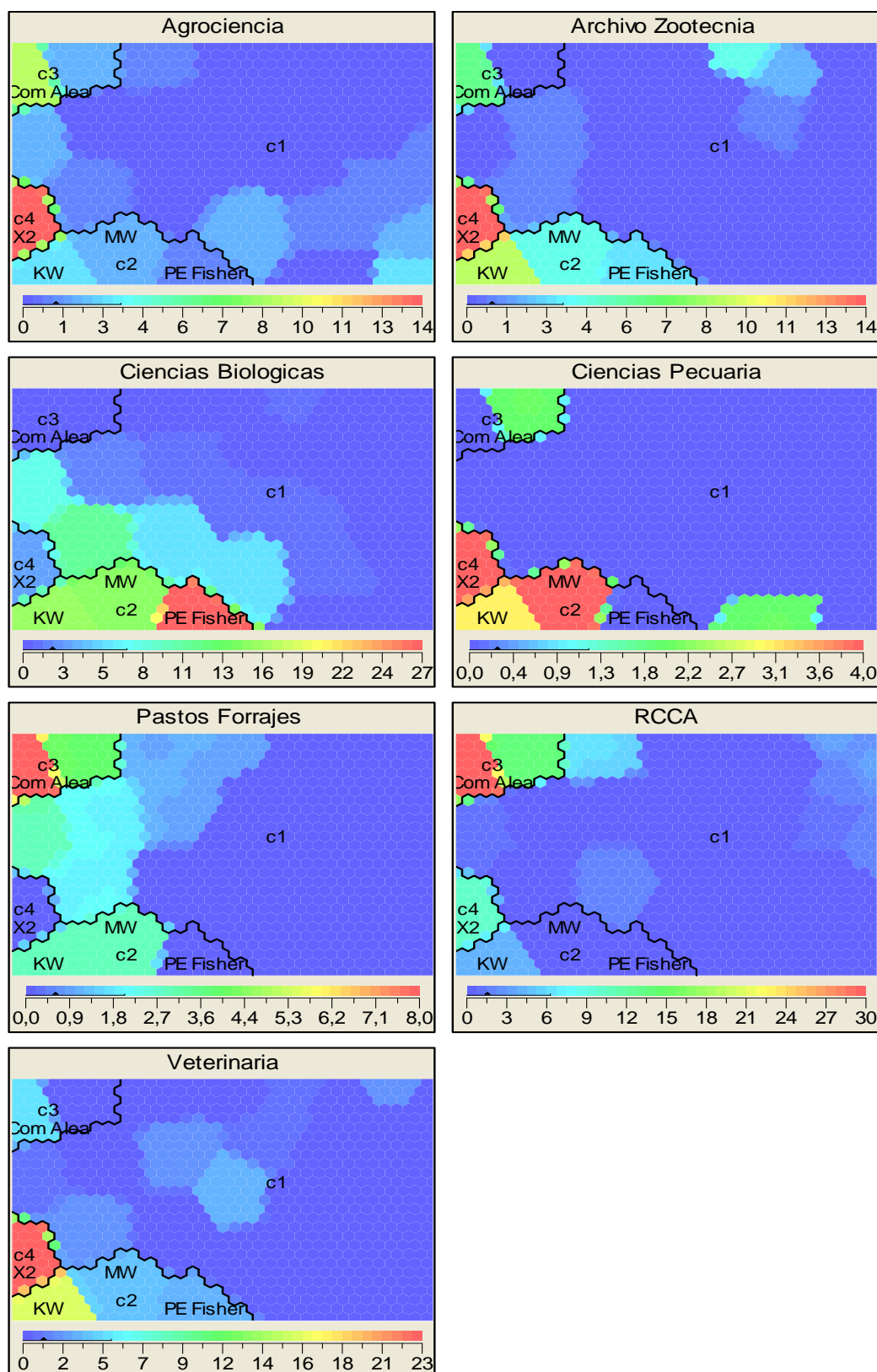


Figura 4. Mapa de Métodos Estadísticos no paramétricos más usado en estudios agropecuarios en el período (2000-2007)

Leyenda. Métodos Estadísticos no paramétricos más representativas
Cluster 1 (c_1): modelos mixtos, dósimas Kolmogorov-Smirnov y Wilconxo
Cluster 2 (c_2): Diseño no paramétrico Completamente Aleatorizado Kruskal Wallis (KW),
Mann Whitney (MW) y Prueba Exacta de Fisher (PE Fisher)
Cluster 3 (c_3): Diseño paramétrico Completamente Aleatorizado (Com Alea) y Diseño
paramétrico Bloque al Azar.
Cluster 4 (C_4): Dócima de X^2 .

Un estudio similar al realizado en este trabajos fue publicado por De Calzadilla (1999), en el período 1987-1997, encontrando 65 aplicaciones referidas a estos métodos, en 2291 artículos, en 11 revistas científicas, 6 nacionales y 5 extranjeras. Esta 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 reportado para las revistas Ciencias Biológicas y Veterinaria de México

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

Uno de los métodos más empleado, es el Análisis de Varianza por rango de Kruskal Wallis, Balzarini et al. (2001), plantean que este procedimiento fue propuesto por Kruskal Wallis en 1952 y permite por una parte, comparar las esperanzas de dos o más distribuciones sin la necesidad de verificar el supuesto de que los términos de error se distribuyen normalmente.

En la RCCA, se destaca el trabajo de Mazorra, (2006), donde se hace referencia al empleo de la dócima de Kruskall-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 dócima de rangos múltiples propuesta por Siegel y Castellan (1995), con el objetivo de detectar diferencias significativas entre los tratamientos.

En la Revista Archivos de Zootecnia, sobresale el trabajo de Delgado et al. (2000), que refiere su estudio a variables nomiles (color de la mucosa, forma de la oreja, etc.) y el de Revalti, et al. (2004) en un estudio similar, aplicando el Análisis de Varianza Kuskal-Wallis.

1 Relacionados con la d cima Mann-Whitney se encuentran los trabajos de Carvajal et al.
2 (2001) y Beiro et al. (2001), de la Revista Ciencias Biol gica, los cuales estudian las
3 variables "tiempo de supervivencia" y "actividad antitumoral", despu s de aplicados los
4 tratamientos.

5
6 Resulta importante destacar el empleo de la d cima de la Probabilidad exacta de Fisher,
7 la cual es sumamente  til para analizar datos discretos, nominales u ordinales, cuando las
8 dos muestras independientes son peque as. En este caso se reportan los estudios
9 realizados en la Revista Ciencias Biol gicas, Bacard  et al (2004), que eval a el grado de
10 toxicidad de un estimulador de colonias de granulocitos, y G mez et al. (2005), que
11 estudia las "frecuencias de muertes, aparici n de signos y lesiones cuando es aplicado el
12 tratamiento D-003", en este sentido se observa que las variables analizadas est n en
13 escalas nominal y ordinal.

14
15 Aunque no fue objetivo de este trabajo analizar el empleo de los M todos Estad sticos
16 param tricos, se puede indicar que en la RCCA y Pastos y Forrajes se hace un mayor
17 empleo de estos, en lo fundamental en los An lisis de Varianza asociado a los Dise os
18 Completamente Aleatorizado y Bloque al Azar.

19
20 Se puede concluir que los m todos estad sticos no param tricos m s empleados, fueron
21 las d cimas χ^2 , Mann-Whitney, Kruskal Wallis, y Exacta de Fisher, aunque se observa
22 m s aplicaciones de los m todos estad sticos no param tricos en investigaciones
23 cient ficas publicadas en Revistas Agropecuarias, estos son pocos diversos. El uso del
24 Escalamiento Multidimensional y los Mapas Auto Organizados (SOM), resultaron muy
25  tiles para el procesamiento, visualizaci n y an lisis de la informaci n, al permitir una
26 adecuada identificaci n de las revistas y las tem ticas analizadas y el empleo de los
27 m todos estad sticos no param tricos en las Ciencias Agropecuarias. Se sugiere
28 continuar investigando en Revistas de alto  ndice de impacto, de las diversas ramas
29 investigativas del Instituto de Ciencia Animal, el empleo de los m todos estad sticos no
30 param tricos y ampliar el uso de las herramientas de Escalamiento Multidimensional y los
31 Mapas Auto Organizados, por su utilidad en la visualizaci n y an lisis de informaci n, en
32 particular para variables categor icas.

33 **Bibliograf a**

34
35 Bacard , D., Moreno, N., Cosme, K., Porras; Carrera, I., D az, G.; Su rez, J.; Aldana, L.,
36 Amaya, R., 2004. Evaluaci n de Toxicidad subcr nica Del factor estimulador de
37 colonias de granulocitos. Revista de Ciencias Biol gicas 35:3. Cuba
38

- 1 Balzarini, M. G., Casanoves, F.; Di Rienzo, J. A.; González, L. A., Robledo, C. W. 2001:
2 Manual del Usuario Insostat. Argentina.
- 3
- 4 Beiro, O., Hernández, D., Arúz, L.; Meriño, N., Álvarez X. y Quintero, Y. 2001. Subnitrato
5 de bismuto en la actividad antitumoral y la nefrotoxicidad por Cisplatin en ratones
6 trasplantados con leucemia P-388. Revista de Ciencias Biológicas 32:2 pp. 103.
- 7
- 8 Carvajal, D., Arruzazabala, L., Molina, V., y Valdés S. 2001. Efecto de D-002 en modelos
9 experimentales de anafilaxia. Revista Ciencias Biológicas 32:2 pp.93. Cuba.
- 10
- 11 Cuesta, A., 2005. Análisis Multivariante. Dpto. Matemática, Estadística y Computación.
12 Universidad de Cantabria España 55 p.
- 13
- 14 De Calzadilla, J.1999. Procedimientos de la Estadística no paramétrica. Aplicaciones
15 en las Ciencias Agropecuarias. Tesis presentada en opción al título de Master en
16 Matemática Aplicada a las Ciencias Agropecuaria.
- 17
- 18 Delgado, J., Dieguez, E.; Cañuelo, P. 2000. Caracterización exteriorista de las variedades
19 de cerdo ibérico basadas en características cualitativas 49:185-186. Revista Archivos
20 de Zootecnia pp. 201-207. España
- 21
- 22 Gámez, R.; Más, R., Noa, M.; y García, H.; Felip, E; Goicochea, E, Gómez, A.;
23 Rodríguez; M. y González, J. 2005. Carcinogénesis del D-003 en ratas Sprague
24 Dawley. Reporte de 1 año. Revista Ciencias Biológicas: 36:3 pp. 201. Cuba
- 25
- 26 Guerra, C. W., De Calzadilla, J. y Torres, V. 2000. Índice de eficiencia en relación con
27 procedimientos de la estadística no paramétrica. Revista Cubana de Ciencia Agrícola
28 34:1 pp.1-4. Cuba
- 29
- 30 Guerrero, F. M., Ramírez, J. M. 2002. El análisis de escalamiento multidimensional: una
31 alternativa y un complemento a otras técnicas multivariantes. Asociación Española de
32 Profesores Universitarios de Matemática para la Economía y la Empresas. X Jornada
33 2002
- 34
- 35 Kohonen, T. Self-organizing maps. Berlin: Springer, 3. ed; 2001.
- 36
- 37 Linares, G. 2001. Escalamiento multidimensional. Conceptos y enfoques. Revista
38 Investigación Operacional. Vol. 22. No. 2. P. 173_183.
- 39
- 40 Mazorra, C.; Borge, G.; Blanco, M., Borroto, A.; Sorís, A., L 2006. Influencia de la dosis de
41 cloruro de litio en la conducta de ovinos condicionados que pastorean en plantaciones
42 de cítricos. Revista Cubana de Ciencia Agrícola 40:4 pp. 425-431. Cuba.
- 43
- 44 Mederos, M.V.; Linares, G.; Miret, E. 2004. Una comparación de dos métodos de
45 gradiente en el Escalamiento Multidimensional. Revista Ciencias Matemáticas. Vol. 22.
46 No. 1. p. 44-56.
- 47
- 48 Miret, E. 2005. Un enfoque unificado para técnicas de representación euclidiana. Tesis
49 para optar por el grado de Doctor en Ciencias Matemática. UH. 148 p.
- 50
- 51 Mora, M.C.; González, M.I. 2009. Caracterización del fracaso empresarial en Colombia:
52 Un enfoque con Escalamiento Multidimensional. Universidad de Salamanca. España.
53 Documento de Trabajo 11/09. 37p.
- 54

- 1 Revialti, M. A.; Capellari, A.; Delgado, J. V. 2005. Recurso genético porcino autóctono en
2 el nordeste de la republica de Argentina. Revista Archivos de Zootecnia 54:205.
3 España. http://www3.uva.es/empresa/uploads/dt_11_09.pdf
4
- 5 Salcines, J. 2003. La normalidad en el gasto de los consumidores: Un análisis de los
6 hogares de España. Momento económico 126: p .2-14.
7
- 8 Siegel, S., Castellan, N.J. 1995. Estadística no paramétrica aplicada a las Ciencias de la
9 Conducta. Cuarta edición. Editorial Trillas, México. 437p.
10
- 11 Visauta, B. 2000. SPSS para Window versión 16.0.
12
- 13 Viscovery Software GmbH. 2008. Viscovery SOMine 5.0.2. [Software de cómputo].
14 Viscovery Software GmbH.
15
- 16 Ward, J. 1963. Hierarchical grouping to optimize an objective function. Journal of the
17 American Statistical Association. Vol. 58. No. 4 p. 236_244.