

# UBICACIÓN DE REVISTAS CIENTÍFICAS EN CUARTILES SEGÚN SJR: PREDICCIÓN A PARTIR DE ESTADÍSTICA MULTIVARIANTE

*Tomás Darío Marín Velásquez\**

Departamento de Ingeniería de Petróleo. Universidad de Oriente.

*Dany Day Josefina Arriojas Tocuyo\*\**

Gerencia de calidad del dato. Petróleos de Venezuela.

**Resumen:** Se analiza el sistema de cuartiles para la clasificación de revistas científicas, mediante clasificación estadística multivariante, utilizando datos de la página oficial de Scopus en 2019. Se tomó una muestra de 5740 registros y se extrajeron cuatro indicadores (CiteScore, Porcentaje de citas, SJR y Percentil) además del cuartil de ubicación (Q). El comportamiento de los indicadores se analizó mediante estadística descriptiva por cuartil, además de la clasificación mediante análisis discriminante y redes neuronales artificiales (RNA). El cuartil con mayor dispersión de indicadores fue Q1 y el cuartil con indicadores más homogéneos fue el Q4. El análisis discriminante arrojó 97,82% de cuartiles correctamente clasificados y 97,23% con RNA. Existen revistas que no se ajustan al cuartil donde se encuentran, de acuerdo a la estadística multivariante. Se observó que el factor más influyente en la clasificación es el percentil y no los indicadores de impacto.

**Palabras clave:** Análisis discriminante; redes neuronales; revistas científicas; impacto; clasificación.

**Title:** LOCATION OF SCIENTIFIC JOURNALS IN QUARTILES ACCORDING TO SJR: PREDICTION FROM MULTIVARIATE STATISTICS.

**Abstract:** The quartile system for the classification of scientific journals is analyzed, through multivariate statistical classification, using data from the official Scopus website in 2019. A sample of 5740 records was taken and four indicators were extracted (CiteScore, Citation Percentage, SJR and Percentile) in addition to the location quartile (Q). The behavior of the indicators was analyzed through descriptive statistics by quartile, in addition to classification through Discriminant Analysis and Artificial Neural Networks. The quartile with the highest indicator dispersion was Q1 and the quartile with the most homogeneous indicators was Q4. Discriminant Analysis showed 97,82% of correctly classified quartiles and 97,23% with RNA. There are magazines that do not adjust to the quartile where it is, according to multivariate statistics. It was observed that the most influential factor in the classification is the Percentile and not the impact indicators.

**Keywords:** Discriminant analysis; neuronal networks; scientific magazines; impact; classification.

**Copyright:** © 2021 Servicio de Publicaciones de la Universidad de Murcia (Spain). Este es un artículo de acceso abierto distribuido bajo los términos de la licencia Creative Commons Reconocimiento 4.0 Internacional (CC BY 4.0).

## 1 INTRODUCCIÓN

La masificación de la difusión del conocimiento científico y el surgimiento de gran cantidad de revistas científicas especializadas, ha llevado a que las mismas sean clasificadas mediante métricas específicas para la medición de su impacto como vehículos de promoción del conocimiento científico. Desde que se introdujo por primera vez el concepto del Factor de Impacto propuesto por Eugene Garfield en 1963 y su uso en la base de datos de Science Citation Index (SCI), el mismo se considera como el principal indicador de la calidad de las revistas científicas (Beltrán, 2006). Este índice se calcula a partir del número de citas recibidas por los documentos de la revista en relación con el número de documentos publicados, ambos medidos en un lapso de tiempo determinado y el mismo mide la relevancia que tiene la revista desde un punto de vista global (Miró y Burbano, 2013).

Las principales indexadoras a nivel mundial han creado su propia versión del Factor de Impacto, así la Web of Science (WoS) calcula para sus revistas el Journal Citation Index (JCI) y Scopus el Scimago Journal & Rank (SJR) a partir de los cuales ubican a las revistas en un sistema de cuartiles, de acuerdo al percentil de ubicación dentro de su área de conocimiento (Vasen y Lujano, 2017). Para la ubicación en cuartiles las revistas de un área de conocimiento son ordenadas de acuerdo al Factor de Impacto de mayor a menor y la lista se divide en cuatro partes, las revistas que ocupan los percentiles inferiores ( $\leq 25\%$ ) se ubican en Q4, las de los percentiles entre 25 y 50% en el cuartil Q3, las de

\* tmarin@protonmail.com

\*\* arriojasd@gmail.com

Recibido: 12-11-2020; 2ª versión: 17-01-2021; aceptado: 20-01-2021.

MARÍN VELÁSQUEZ, T.D. y ARRIOJAS TOCUYO, D.D.J. Ubicación de revistas científicas en cuartiles según SJR: Predicción a partir de estadística multivariante. *Anales de Documentación*, 2021, vol. 24, nº 1. Disponible en: <http://dx.doi.org/10.6018/analesdoc.455951>.

los percentiles entre 50 y 75% en el cuartil Q2 y el percentil superior (> 75%) es ocupado por las revistas Q1 (Orbay *et al.*, 2020).

Aun cuando se ha estandarizado la clasificación por cuartiles como una forma de indicar el impacto y la calidad de las revistas científicas, desde el surgimiento del Factor de Impacto ha sido objeto de críticas por parte de algunos investigadores, que han llegado a dudar del verdadero valor para definir la relevancia de una revista (Miró y Burbano, 2013, Xie *et al.*, 2020). Respecto a lo anterior se han realizado investigaciones donde se han propuesto otras formas de calcular el impacto de las revistas, como el trabajo de Aleixandre-Benavent *et al.* (2007) quienes destacan como en España se han creado factores de impacto propios para áreas específicas. De igual forma Xu *et al.* (2015) propusieron un método de clasificación de revistas basado en el índice h global como respuesta al sistema de clasificación imperante. Meyerholz y Flaherty (2017) investigaron sobre el futuro del factor de impacto. Gómez-Núñez *et al.* (2014) al comparar el SJR y el JCI consideran que existen diferencias que deben ser estudiadas a fondo.

Uno de los aspectos relevantes en la clasificación por cuartiles es que los mismos son dependientes del área de conocimiento, lo que fue resaltado por Ennas y Di Guardo (2015) quienes estudiaron revistas incluidas en la base de datos de Scopus de varios países y notaron que las áreas con mayor cantidad de revistas son las ciencias de la salud, la bioquímica y la agrícola y en áreas como economía, ingeniería química y gerencia existe menor cantidad. De lo anterior es claro que revistas en áreas con mayor cantidad deben tener mayor factor de impacto para ubicarse en los cuartiles superiores, si se comparan con las revistas en áreas minoritarias.

Dada la importancia y los debates que se presentan alrededor del factor de impacto y su sistema de clasificación en cuartiles, se propuso el siguiente trabajo donde se utilizan dos métodos estadísticos de clasificación multivariante para predecir la ubicación de las revistas en cuartiles de acuerdo a indicadores bibliométricos obtenidos de la plataforma de Scopus, con la finalidad de analizar que tan bien clasifican las revistas de acuerdo a las metodologías estadísticas, interrelacionado cuatro indicadores de impacto. El uso de la estadística multivariante puede ser una metodología alternativa para una mejor clasificación de las revistas en respuesta a los algoritmos tradicionales ya que ha sido utilizada con resultados relevantes en otras áreas como las ciencias médicas (Lins *et al.*, 2017, Rahayu *et al.*, 2019) ciencias ambientales (Mehdi *et al.*, 2019) y las ciencias agrícolas (Gandhi *et al.*, 2016; Silva *et al.*, 2017).

## 2 METODOLOGÍA

El trabajo se realizó a partir de datos bibliométricos generales obtenidos del portal de Scopus preview (<https://www.scopus.com>) en el que se encuentran registros detallados de las revistas indexadas en la base de datos Scopus. Se obtuvieron los cuartiles de ubicación de las revistas, los cuales se basan en el SJR y son divididos en cuatro grupos iguales (Q1, Q2, Q3 y Q4) así como información de indicadores bibliométricos obtenidos con base en el registro de 2019, la cual se detalla en la Tabla 1.

Indicador bibliométrico	Descripción
CiteScore	Cuenta el número de citas recibidas en 2016-2019 de artículos, documentos de conferencias, capítulos de libros, reseñas y documentos de datos y las divide por el número de publicaciones en ese mismo periodo de tiempo.
Porcentaje de citaciones	Porcentaje de citas totales de la revista.
SJR	La metodología da cuenta del número de citas, así como de la fuente de las mismas, valiendo más las citas de revistas de alto prestigio que las de revistas de menor prestigio. El valor del prestigio depende del campo, la calidad y la reputación de las revistas, fuente en las que se publica el artículo que cita. Solo considera artículos, reseñas y documentos de conferencias revisados por pares.
Percentil	Representa la posición de la revista dentro de su área de conocimiento. Mientras más alto, mejor ubicada se encuentra la revista.

**Tabla 1. Descripción de los indicadores bibliométricos tomados en cuenta para el pronóstico de la ubicación en cuartiles según la base de datos Scopus.**

De un total reportado en la base de datos de 57019 registros, se escogió de forma probabilística una muestra de 5740 registros pertenecientes a todas las áreas temáticas, con los indicadores descritos en la tabla 1, lo que representa el 10,07% de la población total. La muestra se clasificó en cantidades iguales de revistas por cuartil, es decir que cada cuartil representó el 25% de la muestra (1435 registros por cuartil). Lo anterior se hizo para asegurar que se contara

con igual cantidad de revistas ubicadas en los cuartiles, para no crear sesgos hacia ninguno de ellos. Tomando en cuenta que se utilizaron cuatro indicadores bibliométricos por registro, se contó con un total de 22960 datos para el modelado de la ubicación de las revistas mediante la estadística multivariante.

El tratamiento de los datos se realizó mediante dos técnicas estadísticas multivariantes, el análisis discriminante y las redes neuronales artificiales. Ambas técnicas permiten clasificar una serie de datos de acuerdo a una variable clasificatoria de salida, para analizar que tan bien se clasifica la variable con respecto a los datos introducidos.

El análisis discriminante se realiza para establecer la relación que existe entre una variable dependiente de tipo categórica (en este caso Q) con X niveles que están en correspondencia con los grupos que se analizan (Q1, Q2, Q3, Q4) y un conjunto de variables dependientes cuantitativas (indicadores bibliométricos) con base en una serie de ecuaciones discriminantes que son combinaciones lineales de las variables independientes, para buscar las que mejor separan o discriminan a los grupos. Según Mures *et al.* (2005) la expresión de las funciones discriminantes tiene la forma:

$$f_{km} = \mu_0 + \mu_1 X_{1km} + \mu_2 X_{2km} + \dots + \mu_p X_{pkm} \tag{1}$$

Donde:

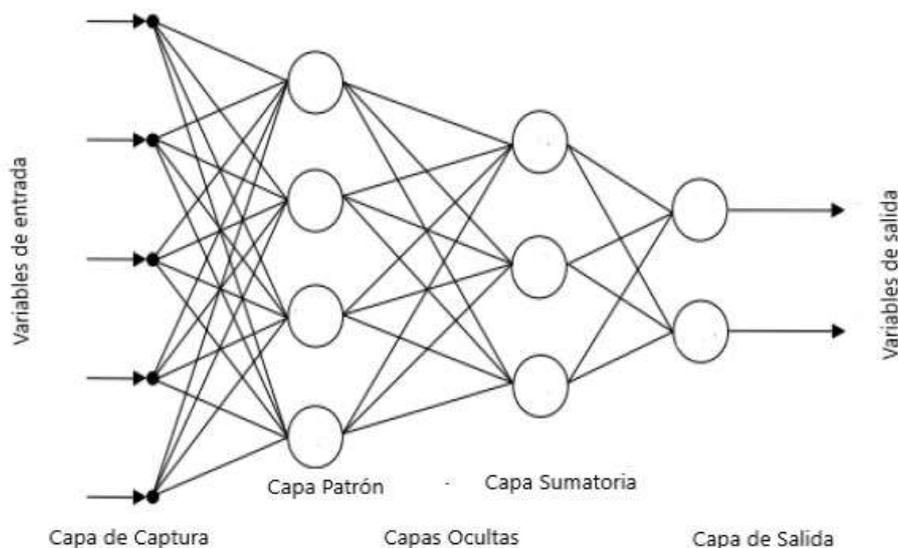
$f_{km}$  = valor o puntuación en la función discriminante para el caso m en el grupo k.

$X_{ikm}$  = valor de la variable discriminante xi para el caso m en el grupo k.

$\mu_i$  = coeficientes o ponderaciones de las variables xi.

El número de funciones discriminantes que se obtienen va a depender, del número de grupos definidos para la variable dependiente (en el caso particular son cuatro grupos) y se obtienen de tal forma que la primera función se genera con las variables explicativas (indicadores bibliométricos) cuyos valores son los que más diferencian los grupos, la siguiente función es la segunda combinación de los indicadores que más discrimina entre los cuartiles, pero cuyos valores no se correlacionan con los obtenidos mediante la primera función y así sucesivamente hasta completar el número total de funciones que clasifican o discriminan el total de los grupos.

Respecto a las RNA, Villalba *et al.* (2012) indican que son una metodología estadística efectiva para la resolución de complicados problemas de clasificación y reconocimiento de patrones, donde las más utilizadas son las llamadas de propagación hacia delante, cuya representación gráfica se observa en la Figura 1.



**Figura 1. Representación gráfica de una red neuronal artificial.**

La capa de entrada tiene un tamaño que es proporcional al número de variables disponibles para la creación de la RNA, en este caso los cuatro indicadores bibliométricos escogidos, los cuales se utilizan para la clasificación de la variable de salida. El número de neuronas de salida coincide con el número de niveles de la variable dependiente a ser

separadas, es decir los cuartiles Scopus (Q1, Q2, Q3 y Q4). Las unidades de cada una de las capas se conectan de forma unidireccional con las pertenecientes a la capa siguiente, creándose las capas ocultas cuyas salidas se diferencian por la multiplicación de un factor que es único y con un valor diferente para todas las conexiones.

La señal que recibe una neurona puede ser calculada mediante la expresión matemática mostrada en la ecuación 2 (Gestal, 2013).

$$NET_i(t) = \sum_{j=1}^{N-1} [W_{ij} * O_j * (t-1)] \quad (2)$$

Donde:

$NET_i(t)$  = conjunto de fuerzas o variables que actúan sobre la neurona t.

N = número de variables.

$W_{ij}$  = peso de la conexión entre una neurona emisora j y una receptora i.

$O_i$  = respuesta de la neurona ante el estado de activación.

Luego de haberse calculado el valor conjunto total de todas las entradas se aplica una función de activación (FA) la cual va a determinar el valor del estado interno de la neurona, así como la respuesta que esta transmita a su salida (Figura 2).

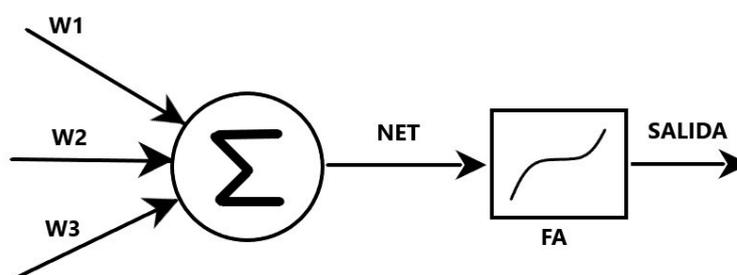


Figura 2. Esquema de una neurona artificial.

Los cálculos fueron realizados mediante el paquete estadístico Statgraphics Centurion XVII, el cual es un software especializado que permite el manejo de gran cantidad de datos y permite de forma rápida y en pocos pasos realizar análisis estadísticos complejos. El paquete estadístico realiza los análisis de clasificación multivariante de acuerdo a lo siguiente: Análisis discriminante se diseña para distinguir grupos de datos con base en un conjunto de p variables cuantitativas observadas e introducidas como variables de entrada. Esto se hace construyendo funciones discriminantes como combinaciones lineales de las variables de entradas. En el caso particular, las variables cuantitativas de entrada fueron los indicadores bibliométricos y los grupos a discriminar los cuartiles.

La clasificación mediante redes neuronales artificiales es realizada por el paquete mediante un método no paramétrico con el cual clasifica observaciones en uno de g grupos (cuartiles) con base en p variables cuantitativas observadas (indicadores bibliométricos) suministradas como entrada. Se realiza una estimación de la función de densidad de los grupos, basada en métodos no paramétricos para cada grupo en una localización específica que se basa en las observaciones, adyacente a ese grupo, usando una ventana de Parzen (García y Sancho, 2010) para ponderar las observaciones de cada grupo con base a su distancia a la localización o nivel especificado para la variable a clasificar, especificadas en los datos de entrada.

Además de los análisis multivariantes, se realizaron análisis descriptivos de los indicadores (promedio, desviación estándar, coeficiente de variación y rango) así como de análisis de correlación entre los indicadores mediante la prueba Rho de Spearman.

### 3 RESULTADOS Y DISCUSIÓN

El análisis descriptivo de los indicadores bibliométricos de la muestra tomada para el estudio, respecto a los cuartiles de ubicación según Scopus se muestra en la Tabla II. Se observa que todos los indicadores muestran una tendencia definida respecto a los cuartiles, siendo los promedios más altos los de las revistas ubicadas en el cuartil Q1.

Indicador bibliométrico	Cuartil	Promedio	Desviación estándar	Coefficiente de variación	Mínimo	Máximo
CiteScore	Q1	6,3	5,9	94,71%	0,3	73,4
	Q2	2,7	1,4	51,68%	0,2	7,5
	Q3	1,4	1,0	67,75%	0,1	5,5
	Q4	0,5	0,4	89,37%	0,1	3,0
SJR	Q1	1,517	1,852	122,08%	0,1	29,997
	Q2	0,598	0,347	58,03%	0,1	2,940
	Q3	0,344	0,224	65,09%	0,1	2,081
	Q4	0,176	0,097	55,14%	0,1	1,194
Porcentaje de citación	Q1	74,0	13,9	18,79%	18,4	100,0
	Q2	61,7	15,9	25,83%	9,3	98,6
	Q3	45,8	17,6	38,40%	3,8	92,7
	Q4	24,0	13,5	56,39%	1,8	73,4
Percentil	Q1	87	7	8,20%	75	99
	Q2	62	7	11,55%	50	74
	Q3	37	7	19,38%	25	49
	Q4	14	6	42,82%	0	24

Tabla II. Estadística descriptiva de los indicadores bibliométricos respecto a los cuartiles Scopus de la muestra de estudio.

Con relación al coeficiente de variación, el cual representa una medida de cuanto varían los datos respecto al valor promedio, en el caso del CiteScore la mayor dispersión de los datos se obtiene en las revistas Q1, lo que indica que existen revistas en este cuartil con CiteScore tan bajos como 0,3 y altos como 73,4. Por el contrario la menor dispersión del indicador se observó en las revistas Q2 con un rango de 7,3. Así mismo, las revistas Q3 y Q4 también presentaron rangos bajos en comparación con las Q1, lo que demuestra que las revistas ubicadas en el cuartil superior presentan una gran variación en cuanto a este indicador.

El indicador SJR también presentó una alta dispersión en las revistas Q1 llegando a ser de más de 100%, con un rango de 29,897 y valores tan distantes como 0,1 y 29,997. Lo anterior demuestra que la ubicación en el cuartil superior está fuertemente condicionada por criterios distintos a este indicador, ya que es ilógico que existan revistas Q1 con SJR = 0,1 y otras con más de 20. El resto de los cuartiles presentó coeficientes de variación más bajos y destaca el Q4 con el menor rango (1,094) siendo el cuartil más homogéneo de acuerdo al SJR.

Respecto al porcentaje de citación los valores favorecen al cuartil Q1, que además presentó la menor dispersión, aunque con un rango amplio de 81,6. Los cuartiles Q2 y Q3 tuvieron rangos de valores similares de 89,3 y 88,9 respectivamente por lo que las revistas en estos cuartiles también muestran porcentajes de citación muy dispersos. Las revistas ubicadas en el cuartil Q4 a pesar de presentar el mayor porcentaje de dispersión, fueron las de menor rango (71,6%) pero existen revistas con porcentajes de citación iguales o mayores a las de los cuartiles superiores ubicadas en este cuartil, lo que indica que la competencia dentro de sus áreas de conocimiento las desfavorece.

La diferenciación se observa más en el indicador percentil, ya que las revistas Q1 son las que se ubican en los valores mayores con baja dispersión respecto a las otras ya que este es el indicador que determina la ubicación de las

revistas dentro de los cuartiles, con las Q1 en los percentiles entre 75 y 99%, las Q2 entre 50 y 74%, las Q3 entre 25 y 49% y las Q4 con menos de 25%. El comportamiento gráfico se puede observar en la Figura 3, donde se toma la relación de los indicadores respecto al porcentaje de citas como principal indicador del que dependen el CiteScore y el SJR y se muestra como Q1 es el cuartil con mayor grado de dispersión. También es notable como se cruzan todos los cuartiles.

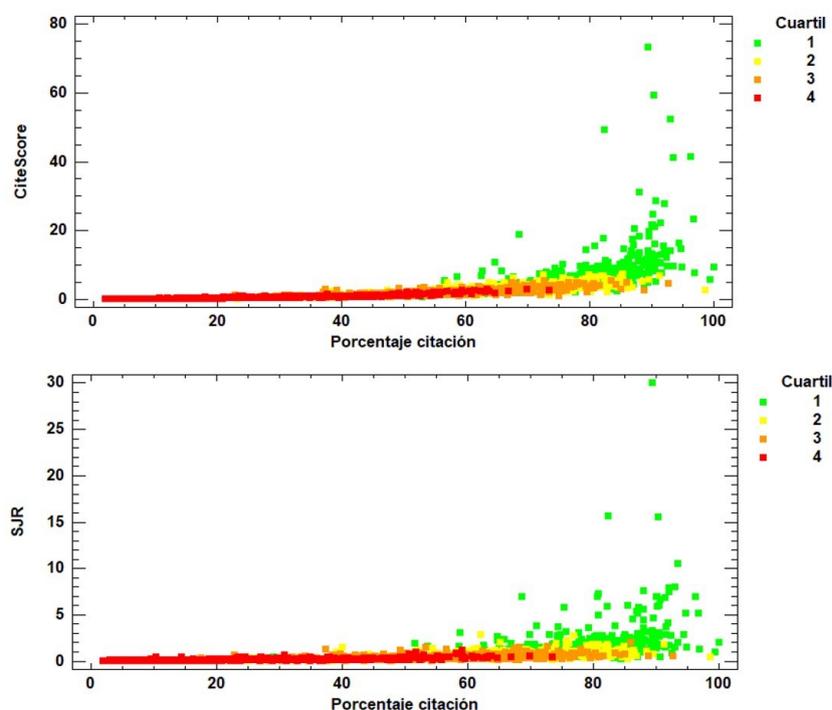


Figura 3. Gráficas de dispersión de CiteScore y SJR respecto al porcentaje de citas.

La mayoría de los estudios realizados, sobre todo a nivel de América Latina y el Caribe, destacan el uso del SJR como indicador de la calidad de las revistas e incluso para la clasificación de los investigadores (Rodríguez *et al.*, 2019). Sin embargo, como se demuestra al analizar los demás indicadores, este factor presenta una alta variabilidad por lo que queda entre dicho el verdadero impacto de las revistas que se ubican dentro de un cuartil que toma el mismo. En un ámbito más global, Grech y Rizk (2018) consideran que la investigación actual se realiza en un entorno complejo de modelos de ciencia interdisciplinaria por lo que las métricas tradicionales comienzan a fallar y se requieren de indicadores innovadores. Las revistas y los académicos deben explorar nuevos enfoques analíticos para evaluar el mérito de la publicación o del investigador, que permitan elaborar una métrica transparente y fiable del contenido científico a nivel de los artículos. De igual forma, el cálculo del factor de impacto para definir la calidad y el cuartil de la revista en la que una persona ha publicado artículos es un tema polémico, que fue advertido desde sus inicios, ya que su uso indebido en la evaluación es evidente debido a la amplia variación de un artículo a otro dentro de una misma revista (Garfield, 2006; Sharma *et al.*, 2014).

Para observar las relaciones entre los indicadores, se realizó un análisis de correlación multivariada no paramétrica ordinal de Spearman, debido a que ninguno de los indicadores sigue una distribución normal según prueba de Kolmogorov-Smirnov la cual arrojó para todos los indicadores un  $p < 0,05$ . Los resultados se muestran en la Figura 4.

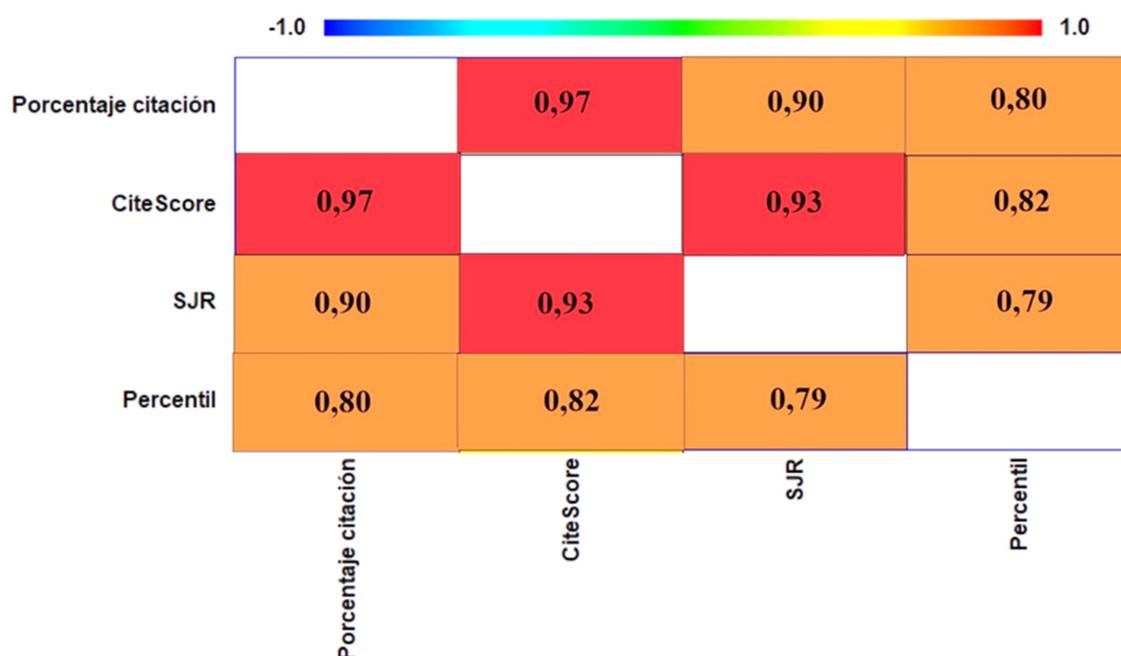


Figura 4. Gráfica de correlación de Spearman de los indicadores bibliométricos de revistas Scopus 2019.

En la figura se observa que todas las relaciones son muy fuertes y positivas, pues son mayores a 0,7 (Hopkins, 2014). Dos de las correlaciones son casi perfectas (CiteScore – Porcentaje de citación y CiteScore – SJR) por ser mayores a 0,9. Las relaciones muy fuertes entre los indicadores eran de esperarse pues tanto el CiteScore como el SJR son dependientes del número de citas que reciben los documentos publicados en las revistas, así mismo, se observa que ambos indicadores también están altamente relacionados con el percentil de ubicación de las revistas. Lo anterior es consistente con lo planteado por Belter (2015) quien menciona que todos los indicadores bibliométricos se basan en medir el impacto de un documento publicado contando la cantidad de otros documentos que lo citaron. Las citas, según la teoría, son consideradas como una forma de importancia o una indicación de la influencia de un artículo respecto a otro. El hecho de que un documento cite a otro es considerado como un indicador de que el documento citado tiene alguna influencia, o impacto.

Al aplicar la metodología estadística de análisis discriminante se obtuvo que existe una función discriminante que predice el 99,18% de la variabilidad de los cuartiles cuando se utilizan los indicadores bibliométricos, con un coeficiente de correlación canónica de 0,97 que indica una relación positiva muy alta y estadísticamente significativa ( $p = 0,000$ ). La función discriminante se muestra en la ecuación 3.

$$Q = -83,2082 - 83,2082 + 0,107681 * \text{Porcentaje citación} - 1,07968 * \text{CiteScore} + 0,640248 * \text{SJR} + 1,86312 * \text{Percentil} \quad [3]$$

El porcentaje de clasificación total fue de 97,82% es decir este es el porcentaje de cuartiles correctamente clasificados según el análisis discriminante. Los resultados se muestran en la Tabla III.

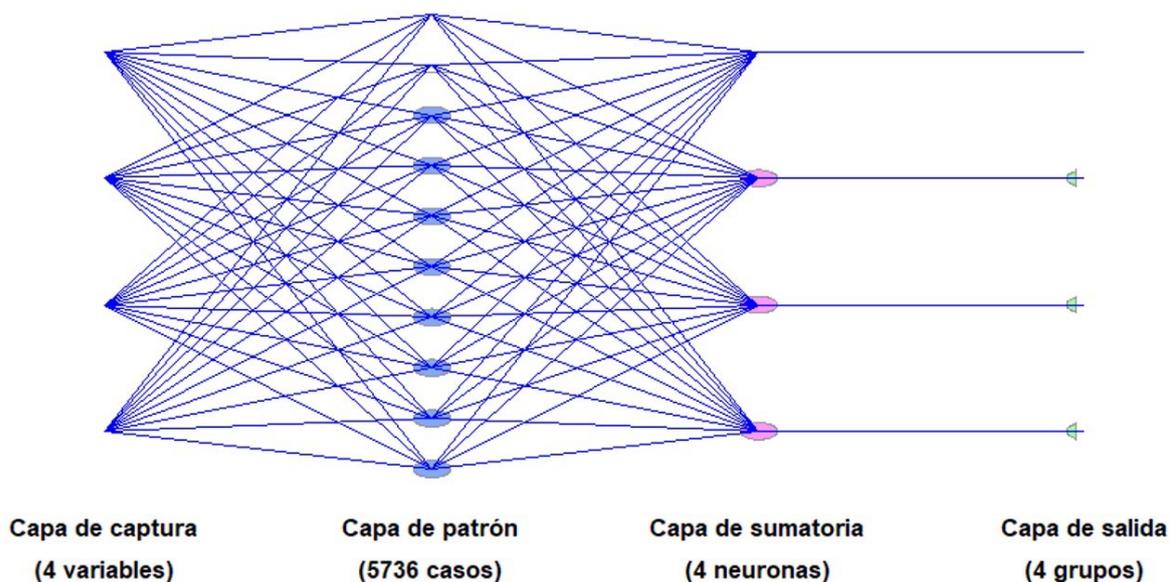
Cuartil Actual	Tamaño de grupo	Cuartil predicho			
		Q1	Q2	Q3	Q4
Q1	1435	1415	20	0	0
		(98,61%)	(1,39%)	(0,00%)	(0,00%)
Q2	1435	18	1402	15	0
		(1,25%)	(97,70%)	(1,05%)	(0,00%)
Q3	1434	0	0	1366	68
		(0,00%)	(0,00%)	(95,26%)	(4,74%)
Q4	1432	0	0	4	1428
		(0,00%)	(0,00%)	(0,28%)	(99,72%)

**Tabla III. Resultado de la clasificación del pronóstico de los cuartiles Scopus mediante análisis discriminante.**

Se observa que el 98,61% de las revistas ubicadas actualmente en el cuartil Q1 fueron correctamente clasificadas por la función discriminante y un 1,39% de ellas se predice que deberían ocupar el cuartil Q2. En el otro extremo las revistas Q4 fueron clasificadas correctamente en un 99,72% y apenas el 0,28% deberían ocupar el cuartil Q3 de acuerdo a la predicción. El cuartil Q2 fue el que mostró mayor dispersión, ya que se clasificaron correctamente 97,70% pero el 1,25% se deberían ubicar en el cuartil Q1 y 1,05% en el cuartil Q3. El cuartil con menor porcentaje de revistas correctamente clasificadas estadísticamente fue el Q3 con 95,26% y el resto (4,74%) se predice que deberían estar ubicadas en el cuartil Q4 de acuerdo a la tendencia de sus indicadores bibliométricos. Es preciso indicar que aun cuando el modelo se alimentó con el total de 5740 datos de entrada, en el análisis se eliminaron cuatro registros, lo cual es común cuando se utiliza un número de datos considerable, debido a posibles redundancias en los mismos detectados por el modelo, sin embargo, el número de registros restantes es representativo.

En términos generales existe una buena coincidencia entre la ubicación actual de las revistas y el pronóstico estadístico, sin embargo, no es menos importante el hecho de que existen revistas que no están bien clasificadas sobre todo en los cuartiles intermedios (Q2 y Q3) lo que debe ser motivo de estudio profundo y reflexión para los sistemas de clasificación y ubicación de revistas, sobre todo a nivel de naciones. Es claro que el sistema de clasificación por cuartiles se ajusta mejor a la predicción del cuartil inferior, lo que se debe principalmente a la menor dispersión por rangos de sus indicadores como se observó en la figura 3. Aun cuando no es posible lograr que todos los grupos sean correctamente clasificados, el análisis discriminante ha demostrado un porcentaje de clasificación alto, lo que también se ha observado en otras investigaciones y en otros campos como lo demostraron Rahayu *et al.* (2019) al utilizar esta técnica para clasificar casos de diabetes, donde concluyeron que se alcanza una precisión razonablemente alta. También Mehdi *et al.* (2019) demostraron que se puede utilizar el análisis discriminante para clasificar zonas por sus índices de evaluación de riesgos ambientales y el índice de carga potencial, además, esta técnica puede utilizarse como herramienta eficaz para clasificar nuevos vertederos de desechos y diseñar nuevos vertederos en función de su nivel de contaminación.

La red neuronal artificial creada a partir de los indicadores bibliométricos para clasificar los cuartiles se muestra gráficamente en la Figura 5.



**Figura 5. Diagrama de red neuronal artificial para la clasificación de los cuartiles.**

Los resultados de la clasificación se muestran en la Tabla IV, donde se observa que el cuartil Q1 se clasificó correctamente en 97,91% y el 2,09% de las revistas en este cuartil fueron incorrectamente clasificadas, por lo que deberían estar ubicadas en el cuartil Q2.

Cuartil actual	Tamaño de grupo	Cuartil predicho			
		Q1	Q2	Q3	Q4
Q1	1435	1405	30	0	0
		(97,91%)	(2,09%)	(0,00%)	(0,00%)
Q2	1435	29	1389	17	0
		(2,02%)	(96,79%)	(1,18%)	(0,00%)
Q3	1434	0	16	1364	54
		(0,00%)	(1,12%)	(95,12%)	(3,77%)
Q4	1432	0	0	13	1419
		(0,00%)	(0,00%)	(0,91%)	(99,09%)

**Tabla IV. Resultado de la clasificación del pronóstico de los cuartiles Scopus mediante RNA.**

En el otro extremo se encuentran las revistas ubicadas en el cuartil Q4, las cuales se clasificaron correctamente en 99,09% y el 0,91% se deberían clasificar como Q3. Los cuartiles intermedios fueron los más dispersos de acuerdo a la técnica estadística de clasificación, donde Q2 tuvo una clasificación correcta del 96,79% con 2,02% que deberían estar clasificadas en Q1 y 1,18% que se ubicaron en Q3. Así mismo, el cuartil Q3 se clasificaron correctamente en 95,12% con 3,77% clasificadas como Q4 y 1,12% en el cuartil Q2. De manera general las redes neuronales artificiales fueron capaces de clasificar correctamente el 97,23% de la ubicación de las revistas en los cuartiles. De forma análoga a lo que se observó en la tabla III, al realizarse el entrenamiento de la RNA se descartaron cuatro registros por presentarse redundancia de datos.

Al igual que se observó en el análisis discriminante, se logra una clasificación relativamente alta, pero es claro que aún existen revistas cuyos indicadores bibliométricos indican que deberían estar clasificadas en otro cuartil, lo que resalta como evidencia de que se deben revisar los algoritmos de cálculo con los que se asignan los cuartiles, pues es claro que hay revistas que podrían estar clasificadas en cuartiles superiores o inferiores.

La técnica de RNA demuestra ser eficiente para la clasificación de las revistas en cuartiles Scopus, lo que coincide con su aplicabilidad en otras áreas del conocimiento como el caso reportado por Gandhi *et al.* (2016) quienes al utilizar RNA para clasificar la productividad de campos de arroz obtuvieron un porcentaje de precisión de 97,5% que es bastante similar al obtenido en la presente investigación, aun cuando fueron datos muy diferentes. De igual forma Silva *et al.* (2017) al utilizar RNA para clasificar cultivares de algodón, concluyeron que la técnica muestra una alta capacidad de clasificación correcta de los 20 genotipos seleccionados en base al índice de calidad de la fibra, lo que ratifica la efectividad de esta técnica estadística multivariante para la clasificación de casos cuando se utilizan los datos correctos.

Aun cuando se obtuvo una mayor cantidad de revistas clasificadas correctamente a través del análisis discriminante, este y las RNA demostraron que existen revistas incorrectamente clasificadas, sobre todo en el cuartil Q2, lo que es debido a la alta dispersión de los principales indicadores entre este cuartil (Tabla II), y que por lo tanto se deben mejorar los algoritmos de clasificación mediante técnicas estadísticas multivariantes. La clasificación de las revistas basada en factores de impacto ha estado presente en el debate científico desde el momento en que el mismo se estandarizó en todo el mundo, por lo que han comenzado a surgir alternativas como contrapartida a la creencia del todopoderoso Factor de Impacto, como lo realizado por varias revistas del grupo *Nature* que han abandonado la publicación exclusiva de la métrica del Factor de Impacto en sus sitios web. En su lugar, esas revistas han comenzado a promover una métrica más diversa para poner de relieve los logros de sus revistas, incluida una nueva métrica denominada mediana de 2 años (*Nature*, 2016). De modo similar, los directores de las revistas de la Sociedad Americana de Microbiología han decidido recientemente eliminar el contenido del factor de impacto de los sitios web de sus revistas para volver a centrar la atención en la publicación de la ciencia de alta calidad (Meyerholz y Flaherty, 2017).

Por su parte, Orbay *et al.* (2020) presentaron un estudio sobre revistas ubicadas por cuartiles de acuerdo a las métricas aportados por WoS en el que reportaron que a pesar de la inclusión de las revistas en los núcleos principales de la indexadora (SCIE, SSCI y A&HCI) se evidencia un porcentaje de artículos que no son citados llegando a ser de hasta el 27,48% en revistas Q1 y de 48,10% en las revistas Q4. Lo anterior también ratifica que pareciera que se deben evaluar mejores metodologías de impacto donde la calidad de las publicaciones sea revisada a profundidad para

asegurar una ubicación real. Así mismo, Campanario (2018) al estudiar las razones por las cuales algunas revistas han logrado mejorar su posición en los cuartiles en un lapso corto de tiempo (6 años) concluye que se debe principalmente a que solo se toma en cuenta el Factor de Impacto, el cual logran aumentar disminuyendo el número de documentos publicados, lo que hace que aunque bajen las citas, el indicador se mantiene o aumenta, una estrategia que pareciera encaminada a que las revistas publiquen menos artículos pero de mayor calidad, sin embargo no está del todo comprobado. Por su parte, Asan y Aslan (2020) consideran que, aunque la ubicación de las revistas científicas en cuartiles proporciona información valiosa en términos de citación y clasificación en sus categorías científicas, a los efectos de los nombramientos de personal académico puede ser malinterpretado, por lo que su uso debe hacerse con la cautela debida. En tal sentido, un trabajo como el presente aporta información valiosa sobre el uso de técnicas estadísticas para tener una mejor ubicación de las revistas, más allá del cálculo simple del Factor de Impacto.

Es evidente que la clasificación de las revistas por cuartiles basada en el Factor de Impacto y el percentil dentro del área de conocimiento influye de manera desfavorable para algunos investigadores de áreas de conocimiento más competitivas, en comparación con áreas donde las revistas tienen menor competencia. Esto se observa claramente cuando se analiza el rango de variación de los SJR de las revistas Q1 que va desde un valor tan bajo como 0,100 hasta uno alto de 29,997. Lo anterior se puede observar cuando se analiza, por ejemplo, que una revista del área de literatura cuyos indicadores bibliométricos son SJR = 0,100; CiteScore = 0,3 y Porcentaje de citas = 21,6% ocupa el Percentil 79 y está ubicada en Q1, en contraste una revista del área de ingeniería biomédica con indicadores bibliométricos superiores SJR = 0,657, CiteScore = 1,1 y Porcentaje de citas = 43,3% ocupa el Percentil 21 y se ubica en Q4. Lo anterior hace que, en términos de clasificación de investigadores, un investigador del área de literatura publica en una revista de menor impacto, pero es mejor clasificado por ser esta Q1 en comparación con un investigador que sea del área biomédica. De allí la importancia de ajustar las metodologías de clasificación de revistas e investigadores, para lo cual los métodos estadísticos multivariantes pueden ser una alternativa.

#### 4 CONCLUSIONES

El análisis descriptivo de los indicadores bibliométricos con relación a los cuartiles de ubicación de las revistas, de acuerdo a la metodología aplicada por Scopus mostró que existe una alta dispersión de los indicadores de impacto (SJR y CiteScore) en el cuartil superior (Q1) con el rango más amplio dentro de los cuartiles. Lo anterior lleva a que existan revistas con indicadores de impacto bajos que se ubican en Q1, por lo que se concluye que no es el Factor de Impacto lo que determina la ubicación, sino el percentil, el cual a su vez está en función de la cantidad de revistas que se encuentren en el área específica.

La predicción de los cuartiles mediante las metodologías estadísticas de clasificación multivariante demuestra que las mismas son aplicables, por presentar porcentajes altos de revistas correctamente clasificadas, los cuales fueron de más de 97%. Sin embargo, el porcentaje de revistas incorrectamente clasificadas es un indicativo de que existen fallas en la metodología de ubicación de las revistas que puede favorecer o desfavorecer a las mismas, dependiendo de si su ubicación incorrecta es en un cuartil superior o inferior. Los cuartiles con la mayor cantidad de revistas incorrectamente clasificadas son los Q2 y Q3 lo que amerita de un estudio que se centre en estos cuartiles para establecer el porqué de esta anomalía. Se comprobó que el cuartil más homogéneo en cuanto a los rangos de indicadores y al porcentaje de revistas correctamente clasificadas fue el Q4, aunque no deja de haber revistas en este que deberían estar en el cuartil Q3.

Además de las revistas incorrectamente clasificadas por los métodos estadísticos, se observa que la gran dispersión de los indicadores de impacto en el cuartil superior (Q1) genera inconsistencias a la hora de la clasificación de revistas e investigadores a nivel de las naciones, ya que se favorecen o ubican en posiciones mayores a investigadores que publican en revistas de menor impacto, pero ubicadas en cuartiles superiores por la menor competencia en su área. Lo anterior amerita de una revisión del verdadero impacto y visibilidad de las revistas y su ubicación menos sesgada a los percentiles por área que permita mejorar los sistemas de clasificación, por lo que se recomienda aplicar los métodos estadísticos multivariantes a las revistas incluyendo todos los indicadores de impacto y disminuir la importancia que se asocia al percentil.

#### 5 BIBLIOGRAFÍA

ALEIXANDRE-BENAVENT, R.; VALDERRAMA-ZURIÁN, J.C. y GONZÁLEZ-ALCAIDE, G. El factor de impacto de las revistas científicas: limitaciones e indicadores alternativos. *El profesional de la información*, 2007, vol. 16, n° 1, p. 4-11. Disponible en: <http://doi.org/10.3145/epi.2007.jan.01>.

- BELTRÁN, O.A. Factor de Impacto. *Revista Colombiana de Gastroenterología*, 2006, vol. 21, n° 1, p. 57-61. Disponible en: <<http://www.scielo.org.co/pdf/rcg/v21n1/v21n1a09.pdf>> [Consulta: 11 de septiembre de 2020]
- BELTER, C. Bibliometric indicators: opportunities and limits. *Journal of the Medical Library Association*, 2015, vol. 103, n° 4, p. 219-221. Disponible en: <https://dx.doi.org/10.3163/1536-5050.103.4.014>.
- CAMPANARIO, J.M. Journals that Rise from the Fourth Quartile to the First Quartile in Six Years or Less: Mechanisms of Change and the Role of Journal Self-Citations. *Publications*, 2018, vol. 6, n° 47, p. 1-15. Disponible en: <https://dx.doi.org/10.3390/publications6040047>.
- ENNAS, G. y DI GUARDO, M.C. Features of top-rated gold open access journals: An analysis of the scopus database. *Journal of Informetrics*, 2015, vol. 9, n° 1, p. 79-89. Disponible en: <https://doi.org/10.1016/j.joi.2014.11.007>.
- GANDHI, N.; PETKAR, O. y ARMSTRONG, L.J. Rice crop yield prediction using artificial neural networks. En *2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR)*, 2016, Chennai, India (p. 105-110). Disponible en: <https://doi.org/10.1109/TIAR.2016.7801222>.
- GARCÍA, P. y SANCHO, J. Estimación de densidad de probabilidad mediante ventanas de Parzen. En *Jornadas de introducción a la investigación de la UPCT*. 2010. Disponible en: <<https://dialnet.unirioja.es/servlet/articulo?codigo=3709476>> [Consulta: 14 de septiembre de 2020]
- GARFIELD, E. The History and Meaning of the Journal Impact Factor. *JAMA*, 2006, vol. 295, n° 1, p. 90-93. Disponible en: <http://doi.org/10.1001/jama.295.1.90>.
- GESTAL, M. Introducción a las redes neuronales [en línea]. 2013. Disponible en: <<https://www.researchgate.net/publication/242099672>> [Consulta: 24 de septiembre de 2020]
- GÓMEZ-NÚÑEZ, A.J. *et al.* Optimizing SCImago Journal & Country Rank classification by community detection. *Journal of Informetrics*, 2014, vol. 8, n° 2, p. 369-383. Disponible en: <https://doi.org/10.1016/j.joi.2014.01.011>.
- GRECH, V. y RIZK, D. Increasing importance of research metrics: Journal Impact Factor and h-index. *International Urogynecology Journal*, 2018, vol. 29, p. 619-620. Disponible en: <https://doi.org/10.1007/s00192-018-3604-8>.
- HOPKINS, W. A New View of Statistics [en línea]. 2014. Disponible en: <<https://complementarytraining.net/wp-content/uploads/2013/10/Will-Hopkins-A-New-View-of-Statistics.pdf>> [Consulta: 25 de septiembre de 2020]
- LINS, A.J.C.C. *et al.* Using artificial neural networks to select the parameters for the prognostic of mild cognitive impairment and dementia in elderly individuals. *Computer Methods and Programs in Biomedicine*, 2017, vol. 152, p. 93-104. Disponible en: <https://doi.org/10.1016/j.cmpb.2017.09.013>.
- MEHDI, Z.; ALI, M.; ABOLGHASEM, K.R. y FARAMARZ, D.A. Classification of environmental geochemical data using discriminant analysis and neural network in carbonate-sulfide waste dumps of lead and zinc mines. *Iranian Journal of Mining Engineering (IRJME)*, 2019, vol. 14, n° 44, p. 12-25. Disponible en: <<https://www.sid.ir/en/journal/ViewPaper.aspx?ID=727041>> [Consulta: 18 de septiembre de 2020]
- MEYERHOLZ, D.K. y FLAHERTY, H.A. The Evolving Significance and Future Relevance of the Impact Factor. *Veterinary Pathology*, 2017, vol. 54, n° 4, p. 721-722. Disponible en: <https://doi.org/10.1177/0300985817690209>.
- MIRÓ, Ò. y BURBANO, P. El factor de impacto, el índice h *et al.* indicadores bibliométricos. *Anales del Sistema Sanitario de Navarra*, 2013, vol. 36, n° 3, p. 371-377. Disponible en: <http://dx.doi.org/10.4321/S1137-66272013000300001>.
- MURES, M.J.; GARCÍA, A. y VALLEJO, M.E. Aplicación del análisis discriminante y Regresión Logística en el estudio de la morosidad en las entidades financieras. Comparación de resultados. *Pecunia*, 2005, vol. 1, p. 75-199.
- NATURE EDITORIAL. Time to remodel the journal impact factor. *Nature*, 2016, vol. 535. Disponible en: <https://doi.org/10.1038/535466a>.
- ORBAY, K.; MIRANDA, R. y ORBAY, M. Building Journal Impact Factor Quartile into the Assessment of Academic Performance: A Case Study. *Participatory Educational Research (PER)*, 2020, vol. 7, n° 2, p. 1-13. Disponible en: <http://dx.doi.org/10.17275/per.20.26.7.2>.
- RAHAYU, W.; SANTI, V.M. y PUTRI, B.S. Classification of diabetes events using discriminant analysis. En *Journal of Physics: Conference Series 1402, 077102*. 2019. Disponible en: <https://dx.doi.org/10.1088/1742-6596/1402/7/077102>.
- RODRÍGUEZ, R.; SOCORRO, A. y ESPINOZA, C. Análisis de Scimago Journal & Country Rank, utilidad para el desarrollo bibliométrico en la Universidad Metropolitana del Ecuador. *Revista Publicando*, 2019, vol. 6, n° 21, p. 58-68.
- SHARMA, M. *et al.* Journal Impact Factor: Its Use, Significance and Limitations. *World Journal of Nuclear Medicine*, 2014, vol. 13, n° 2, p. 146. Disponible en: <http://doi.org/10.4103/1450-1147.139151>.
- SILVA, E.G. *et al.* Cotton genotypes selection through artificial neural networks. *Genetics and Molecular Research*, 2017, vol. 16, n° 3, gmr16039798. Disponible en: <http://dx.doi.org/10.4238/gmr16039798>.
- VASEN, F. y LUJANO, I. Sistemas nacionales de clasificación de revistas científicas en América Latina: tendencias recientes e implicaciones para la evaluación académica en ciencias sociales. *Revista mexicana de ciencias políticas y sociales*, 2017, vol. 62, n° 231, p. 199-228. Disponible en: <<http://www.scielo.org.mx/pdf/rmcp/v62n231/0185-1918-rmcp-62-231-00199.pdf>> [Consulta: 30 de septiembre de 2020]