

Escalamiento multidimensional (MDS) no métrico en el análisis del comportamiento del consumidor en la adquisición de productos lácteos



Non-metric multidimensional scaling (MDS) in the analysis of consumer behavior in the purchase of dairy products

Paulina Fernanda Bolaños Logroño.¹, Fernando Ricardo Márquez Sañay.² & Carmen Elena Mantilla Cabrera.³

Recibido: 14-05-2021 / Revisado: 25-05-2021 / Aceptado: 18-06-2021/ Publicado: 05-07-2021

Abstract.

DOI: <https://doi.org/10.33262/concienciadigital.v4i3.1788>

Introduction: The study of consumer behavior has made it possible to address various multivariate techniques applied in the area of consumer psychology, economics, marketing, tourism and education, trying to explain the behavior of consumers when choosing a certain product and its influence on decision-making. **Objective:** In this sense, the present study aims to analyze consumer behavior when buying dairy products in the city of Riobamba, in the middle of a sanitary situation that involves the whole world. **Methodology:** Through the use of non-metric multidimensional scaling (MDS), starting from a hierarchical matrix related to the importance of the most influential attributes when choosing a dairy product, the multidimensional analysis was optimized through a SMACOF algorithm implemented in the free software RStudio. **Results:** Consequently, the results obtained allowed to distinguish 2 dimensions, where the relevance of each

¹ Escuela Superior Politécnica de Chimborazo, Facultad de Ciencias, Proyecto de investigación diseño de un sistema de negocios para impulsar el desarrollo empresarial de la ciudad de Riobamba, Riobamba, Ecuador, paulina.bolanos@esoch.edu.ec, ID ORCID: 0000-0003-3911-0461.

² Escuela Superior Politécnica de Chimborazo, Facultad de Ciencias, Riobamba, Ecuador, fernando.marquez@esoch.edu.ec, ID ORCID: 0000-0001-5549-9572.

³ Escuela Superior Politécnica de Chimborazo, Facultad Recursos Naturales, Grupo de Investigación en Seguridad Telemática (SEGINTE), Riobamba, Ecuador, carmen.mantilla@esoch.edu.ec, ID ORCID: 0000-0001-5422-7073.

variable with the consumer behavior at the time of purchase, the brand is definitely far from the rest of variables, being the aspect of less importance for the consumer, **Conclusions:** it was possible to distinguish a group of factors that associate and confirm that the taste, quality and nutritional value are the most important attributes, the price is a factor that represents a medium importance at the time of purchase of dairy products. It is demonstrated that the Multidimensional Scaling technique can be used in marketing studies on consumer behavior as an alternative to other multivariate techniques.

Keywords: consumer, multidimensional scaling, dairy, algorithm, proximity, metric, non-metric.

Resumen.

Introducción: El estudio sobre el comportamiento del consumidor ha permitido abordar diversas técnicas multivariantes aplicadas en el área de psicología del consumidor, economía, marketing, turismo y educación, intentando explicar el comportamiento de los consumidores al elegir un determinado producto y su influencia para la toma de decisiones. **Objetivo:** En este sentido, el presente estudio pretende analizar el comportamiento del consumidor al comprar productos lácteos en la ciudad de Riobamba, en medio de una situación sanitaria que aborda todo el mundo. **Metodología:** Mediante el uso del escalamiento multidimensional (MDS) no métrico, partiendo de una matriz jerárquica relacionada con la importancia de los atributos más influyentes al momento de elegir un producto lácteo, se optimizó el análisis multidimensional a través de un algoritmo SMACOF implementado en el software libre RStudio. **Resultados:** En consecuencia, los resultados obtenidos permitieron distinguir 2 dimensiones, en donde la pertinencia de cada variable con el comportamiento del consumidor al momento de realizar la compra, se observa que la marca definitivamente se encuentra alejada del resto de variables, siendo el aspecto de menor importancia para el consumidor, **Conclusiones:** en adición se puede distinguir un grupo de factores que asocian y confirman que el sabor, la calidad y el valor nutricional son los atributos de mayor importancia, el precio es un factor que representa una importancia media al momento de realizar la compra de productos lácteos. Se demuestra que la técnica de Escalamiento Multidimensional puede ser utilizada en estudios de Marketing, sobre el comportamiento del consumidor como alternativa a otras técnicas multivariantes.

Palabras clave: consumidor, escalamiento multidimensional, lácteos, algoritmo, proximidades, métrico, no métrico.

Introducción

La industria láctea en la ciudad de Riobamba presenta alta demanda al tratarse de productos que contribuyen en la alimentación de todos quienes la habitan. En tal virtud, es relevante conocer el comportamiento del consumidor y los factores o atributos que influyen en la decisión de compra de productos lácteos.

El objetivo de este trabajo investigativo es analizar el comportamiento del consumidor al comprar productos lácteos, identificando los factores y el grado de importancia conferido por el consumidor en la toma de decisiones, a través de la aplicación de Escalamiento Multidimensional (MDS), generando información confiable y convirtiéndose en un aporte académico en beneficio de la industria.

El escalamiento multidimensional (MDS) es un método de análisis estadístico multivariante que representa mediciones de similaridad (o disimilaridad) entre pares de objetos como distancias entre puntos de un espacio de dimensión reducida. La técnica tiene sus orígenes en los estudios de psicología experimental en la década de 1950, llevados a cabo para descubrir la similaridad entre estímulos aplicados a distintos individuos; y es en el área de las ciencias sociales donde preferentemente se han aplicado muchos de los avances de las investigaciones. No obstante, el MDS ha encontrado aplicación en una amplia gama de disciplinas científicas, entre otras razones porque admite una gran variedad de datos de entrada como tablas de contingencia, matrices de proximidad y correlaciones. (Vera & Mair, 2019)

El objetivo fundamental del MDS consiste en generar un mapa o representación gráfica de los objetos en un espacio de modo que sus posiciones relativas en tal configuración sean el reflejo del grado de proximidad percibida entre los objetos. Otros propósitos de este método que vale la pena destacar (Borg & Groenen, 2005).

Facilitar el análisis exploratorio de los datos generando una representación en un espacio de dimensión reducida haciendo que estos sean accesibles a la inspección visual del investigador, de modo que pueda apreciarse la estructura de los datos y se encuentren reglas que ayuden a describir su distribución. Revelar el número de dimensiones implícitas a las causas de similaridad o disimilaridad, así como también su número e importancia relativa.

El Multidimensional Scaling (MDS) o Análisis de Proximidades técnica multivariante que procura representar medidas de proximidad entre objetos como distancias euclidianas en un espacio baja dimensión. La representación en baja dimensión permite la inspección visual de la estructura de los datos.

El escalamiento multidimensional es una familia de modelos que tienen en común el esquematizar las proximidades mediante distancias entre puntos de un espacio de dimensión k . Las variantes surgen debido a las diferentes suposiciones sobre la escala de medida de las proximidades y al uso de diferentes funciones para calcular la distancia; esto último da lugar a distintas geometrías en los modelos, ya que la distancia euclídea al igual que otras distancias de Minkowski implican una geometría plana, mientras que la distancia geodésica implica una geometría curvada. (Peñaranda Arenas, 2016)

Dentro de los diferentes modelos de escalamiento multidimensional, se encuentra el MDS clásico, métricos mínimos cuadrados, no métrico, unfolding, diferencias individuales entre otros.

Modelo clásico asume que las proximidades se comportan como si fueran distancias medidas en realidad (generalmente euclídeas), suposición que puede aceptarse para aquellos datos que se derivan de matrices de correlación, pero raramente para valoraciones de disimilaridad directas. La ventaja de esta técnica es que proporciona una solución analítica y que tampoco requiere procedimientos iterativos (Wickelmaier, 2003).

Mientras que el modelo no métrico admite que las disimilaridades trabajan escala ordinal, de esta forma, la configuración espacial utiliza exclusivamente información ordinal de las similitudes.

Según (Peñaranda Arenas, 2016), existen dos formas de MDS no-métrico cuya diferencia reside en el modo en el que tratan los empates o valores iguales en los datos. El enfoque primario que es el adoptado por defecto en la mayoría de los programas, consiste en considerar que las proximidades iguales no necesariamente corresponden a distancias iguales; mientras que el enfoque secundario lleva a mantener los empates, es decir, que las proximidades iguales correspondan a iguales distancias en la solución (Borg & Groenen, 2005).

El algoritmo del MDS no-métrico comprende un proceso de optimización dual en el que debe encontrarse primero una transformación monótona óptima de las proximidades y posteriormente debe arreglarse óptimamente la configuración de los puntos, de manera que sus distancias correspondan a las proximidades escaladas o disparidades lo más cerca que sea posible. Básicamente las etapas del algoritmo del MDS no-métrico consiste en hallar una configuración aleatoria de puntos, por ejemplo, mediante una muestra tomada de una distribución normal. Hallar una configuración aleatoria de puntos, por ejemplo, mediante una muestra tomada de una distribución normal. Calcular las distancias d entre los puntos. Luego hallar la transformación monótona óptima de las proximidades, con el fin de obtener datos óptimamente escalados o disparidades $\hat{d}=f(p)$, seguido de minimizar el stress entre las disparidades y las distancias encontrando una nueva configuración de puntos, finalmente se deberá comparar el stress con algún criterio. (Wickelmaier, 2003):

Por otro lado, unfolding es un modelo geométrico que aplica modelos de distancia y técnicas de escalamiento a matrices rectangulares para establecer preferencias y alternativas. Los judge y objetos son representados juntos en el mismo espacio, el orden de rango de las distancias desde i -th judge a los objetos refleja la efectividad i -th del rango judge. Los modelos unfolding se utilizan ampliamente en la escala de la opción preferencial y la escala de la actitud, (Crepaldi, 2020).

Similar al modelo unfolding, el modelo de diferencias individuales tiene m judges y n objetos, pero en este caso m se llevan a cabo diferentes análisis *MDS*, uno por cada judge. Como resultado, una configuración general de puntos representando a los objetos es dada junto con una representación del espacio de los judge's.

El análisis de datos de elección preferencial ha atraído la atención de los metodólogos de las ciencias sociales durante mucho tiempo. El enfoque clásico, a partir del trabajo de Fechner sobre estética experimental, y formulado como teoría de la elección por

Thurstone en su famosa Ley del juicio comparativo (Thurstone, 1927), implica la asunción de una teoría unidimensional.

Así, por ejemplo, dada una matriz de correlaciones entre diversas variables, el MDS permite representar esas variables como puntos de forma que dos puntos se encontrarán tan próximos entre sí como estén de correlacionados los elementos a los que representan. Si esta relación entre correlaciones y distancias es lo suficientemente precisa, conseguiremos una representación que pondrá de manifiesto la estructura intrínseca existente, hecho que de otro modo podría permanecer oculto al investigador puesto que en general resulta mucho más difícil observar una tabla de coeficientes de correlación que una gráfica en un plano. (De Lucas Jaramillo, 2012)

Así mismo nos introducimos al comportamiento del consumidor, conociendo que el ser humano trata de satisfacer necesidades insatisfechas con recursos limitados. Y tal como lo advertía (Socrático, T. 2012) el hombre es un eterno insatisfecho, que se comporta con la naturaleza, su mundo primigenio, como un perfecto inadaptado, tratando de crear nuevos mundos. Mucho de razón tenía (Thomas Hobbes, XVI) al expresar que *homo homini lupus*, el hombre es un lobo para el hombre. Un ser al que solo le interesa su propia supervivencia, sus logros y de naturaleza intrínsecamente egoísta. Ya Torres (2011) recuerda, que en la teoría económica el egoísmo es un instinto fundamental para los seres humanos. Un egoísmo que se satisface en el individualismo como motor de la relación más relevante del mercado: necesidad y satisfacción.

Para el padre de las ciencias económicas modernas, Adam Smith, en 1937, existe la ley fundamental del egoísmo que soporta la estructura de la conducta individual y que busca el lucro versus una competencia de intercambio en los mercados libres. Para (Smith, 1776) el egoísmo como un recurso metódico es la fuerza que impulsa la economía y permite que al ser humano satisfaga su consumo. Una ansiedad que plantea la búsqueda constante hacia el estado de satisfacción.

El consumo es uno de los comportamientos esenciales en la conducta humana. Los diferentes procesos socioeconómicos se pueden comprender de mejor manera gracias al análisis de las preferencias de los individuos a lo largo del tiempo. En algunos casos el consumo llega a convertirse en un elemento de identidad económica entre los diferentes agentes. Al respecto hay que considerar el aporte de (Dubois y Rovira, 1998) al referirse que los consumidores no poseen un conocimiento perfecto de sus necesidades, ni tampoco conocen aquellos productos que servirían para satisfacerlas.

El consumo y el comportamiento de los individuos no puede ser una mirada aislada y requiere ser analizada desde la perspectiva del eje económico más importante como es la familia y, que es a la vez el núcleo fundamental de la sociedad. En la perspectiva del comportamiento del consumidor el hogar y la familia son consideradas como un solo ente. No participa de la especificación sociológica que podría tener esta referencia. Así lo refiere (Becker, 1987) que incorpora a la influencia de la familia en la toma de decisiones, en sustitución de un solo individuo. De igual manera, la familia se enmarca en la

coordinada del tiempo con el fin de llegar a satisfacer sus necesidades mediante la mayor cantidad de activos.

La teoría moderna ha identificado dos tipos de bienes: duraderos y no duraderos. El análisis de la demanda de estos dos tipos de bienes permite interiorizar las preferencias y los hábitos de los consumidores. De igual manera advierte (Arellano, 2014) existe un perfil que caracteriza al consumidor tradicional y al consumidor no tradicional. El primero lleva una conducta de consumo determinada por las necesidades básicas, el segundo se ve influenciado por la oferta y su información. En este último la promoción juega un papel muy determinante en su juicio de compra.

Metodología.

El modelo de MDS que se adoptó para este estudio es el modelo no métrico y el algoritmo SMACOF, que se describen brevemente:

MDS no métrico

La suposición del MDS métrico de que las proximidades se comportan como distancias puede ser muy restrictiva cuando se aplica el MDS a la exploración del espacio perceptual de los sujetos humanos. Con el fin de resolver este problema, Shepard y Kruskal desarrollaron esta variante del MDS que, en donde se asume que las proximidades están en escala ordinal. (Macías, Rivera, & Vera, 2007)

El problema que aborda el MDS no-métrico es el de hallar una configuración de puntos X tal que las distancias sobre ésta queden ordenadas tan cerca como sea posible a las proximidades y se logren minimizar las diferencias al cuadrado entre las disparidades y las distancias entre los puntos. Este problema es equivalente al de encontrar las coordenadas que minimicen el stress, el cual puede calcularse con la siguiente expresión (Peñaranda Arenas, 2016):

$$STRESS = \sqrt{\frac{\sum(f(p)-d)^2}{\sum d^2}} \quad (1)$$

Tal como se aprecia en la ecuación anterior, la magnitud del stress es proporcional a la diferencia entre las disparidades y las distancias, siendo entonces un indicador de la bondad de ajuste del modelo; en efecto, un valor bajo del stress indica que se obtuvo un buen ajuste con la solución, mientras que un valor alto corresponde a un mal ajuste. Con el fin de interpretar el valor del stress respecto a la bondad de ajuste de la solución suele emplearse la siguiente guía sugerida por Kruskal (Wickelmaier, 2003):

Stress	Bondad de ajuste
> 0.20	pobre
0.1	aceptable
0.05	bueno
0.025	excelente

0.00 perfecto

Tabla 1. Stress y bondad de ajuste

Fuente: (Peñaranda Arenas, 2016)

Dado que la magnitud del stress no proporciona una indicación clara de la bondad del ajuste, existen dos técnicas adicionales que comúnmente se usan para juzgar la idoneidad del modelo: el gráfico de sedimentación y el diagrama de Shepard. El primero de ellos representa la cantidad de stress frente al número de dimensiones de la solución, se busca en éste el menor número de dimensiones asociado a un valor aceptable del stress; un codo en este diagrama indica que la adición de dimensiones a la solución producirá solo una mejora menor en términos del stress, por consiguiente, el mejor ajuste se logra con aquel modelo que utiliza el número de dimensiones que corresponde al codo en esta gráfica. El diagrama de Shepard muestra la relación entre las proximidades y las distancias entre los puntos de la configuración, entre menor sea la dispersión mejor es el ajuste; en el MDS no-métrico la ubicación ideal de los puntos en este diagrama es una línea que aumenta de forma monótona y que describe a las disparidades (Wickelmaier, 2003).

Básicamente las etapas del algoritmo del MDS no-métrico son las siguientes (Wickelmaier, 2003):

- a. Hallar una configuración aleatoria de puntos, por ejemplo, mediante una muestra tomada de una distribución normal.
- b. Calcular las distancias d entre los puntos.
- c. Hallar la transformación monótona óptima de las proximidades, con el fin de obtener datos óptimamente escalados o disparidades $\hat{d}=f(p)$.
- d. Minimizar el stress entre las disparidades y las distancias encontrando una nueva configuración de puntos.
- e. Comparar el stress con algún criterio. En caso de que el stress sea lo suficientemente pequeño se termina el algoritmo y en caso contrario se retorna al paso b.

Algoritmo SMACOF

Este algoritmo minimiza el stress mediante mayorización, en efecto SMACOF significa escalamiento vía mayorización de una función complicada. En sentido estricto, la mayorización no es un algoritmo sino una prescripción para construir algoritmos de optimización. La idea de la mayorización es optimizar una función sustituta más simple que la función original, se garantiza que la función sustituta tiene un valor mayor al de la función original y es igual a esta última en un punto de soporte. En cada iteración, la configuración final es usada como el punto de inicio para la próxima iteración (Leeuw & Mair, 2009).

El algoritmo SMACOF converge a un punto fijo y es equivalente a un algoritmo de gradiente descendente ponderado con tamaño de paso constante. En la versión simple que corresponde al caso de una matriz de disimilaridad Δ simétrica, la función stress (\mathbf{X}) se define así:

$$\sigma(X) = \sum_{i<j} w_{ij}(\delta_{ij} - d_{ij}(X))^2 \quad (2)$$

El algoritmo ubica los $i, j=1, \dots, n$ puntos en un espacio euclídeo de baja dimensión de tal manera que las distancias entre los elementos en la configuración $d_i(X)$ se aproximan a las disimilaridades δ_i .

Otro dato de entrada del algoritmo es la matriz de ponderaciones $W_{n \times n}$, que al igual que la matriz de disimilaridades se asume no-negativa y con elementos iguales a cero en la diagonal. Una de las aplicaciones de esta matriz es facilitar la manipulación de los valores faltantes; por ejemplo, $w_i=0$ si el dato falta y 1 en el caso contrario; otro uso de dicha matriz es definir el énfasis que tendrán en el análisis las disimilaridades. (Arnau, 1996)

El algoritmo inicia asignando en el paso $t=0$ el punto de soporte $Y^{(0)}$ donde $X^{(0)}$ es una configuración inicial. Dentro de cada iteración t se calcula $\bar{X}^{(t)}$, siendo esta última la transformada de Guttman de la configuración, la cual se calcula mediante la siguiente ecuación (Peñaranda Arenas, 2016):

$$\bar{X}^{(t)} = V^+B(Y)Y \quad (3)$$

Donde V^+ representa la inversa de Moore-Penrose: $V^+ = (V + n^{-1}11')^{-1} - n^{-1}11'$. La matriz V se define así:

$$V = \sum_{i<j} w_{ij}A_{ij} \quad (4)$$

Los elementos de la matriz A son iguales a 1 cuando $a_{ii}=a_{jj}$, -1 en el caso que $a_{ij}=a_{ji}$, y 0 en las demás posiciones. Por otro lado, la matriz B evaluada en el punto de soporte Y es igual a:

$$B(Y) = \sum_{i<j} w_{ij}s_{ij}(Y)A_{ij} \quad (5)$$

$$s_{ij}(Y) = \begin{cases} \delta_{ij}/d_{ij}(Y) & \text{si } d_{ij}(Y) > 0 \\ 0 & \text{si } d_{ij}(Y) = 0 \end{cases} \quad (6)$$

En la versión simple de SMACOF la actualización que corresponde a la iteración t , es decir $^{(t)}$, se hace equivalente a $\bar{X}^{(t)}$. El siguiente paso es calcular el stress $\sigma(X^{(t)})$, el proceso de iteración continúa hasta que la diferencia entre el stress calculado en una iteración y la anterior sea menor a una determinada tolerancia, o en otros términos: $\sigma(X^{(t)}) - \sigma(X^{(t-1)}) < \epsilon$, o se alcance un límite en el número de iteraciones. El algoritmo garantiza que en cada iteración el stress disminuye. Otra característica del algoritmo es que a medida que aumenta el número de dimensiones de la solución disminuye la probabilidad de presencia de mínimos locales. (Torgerson, 1952)

Un aspecto a tener en cuenta del algoritmo SMACOF es que cada vez que se calculan las disparidades óptimas \hat{d}_{ij} para las distancias de la actualización de la iteración t , $d_{ij}(X^{(t)})$, éstas se normalizan de la siguiente forma (Borg & Groenen, 2005):

$$\sum_{i < j} w_{ij} (\hat{d}_{ij})^2 = \frac{n(n-1)}{2} \quad (7)$$

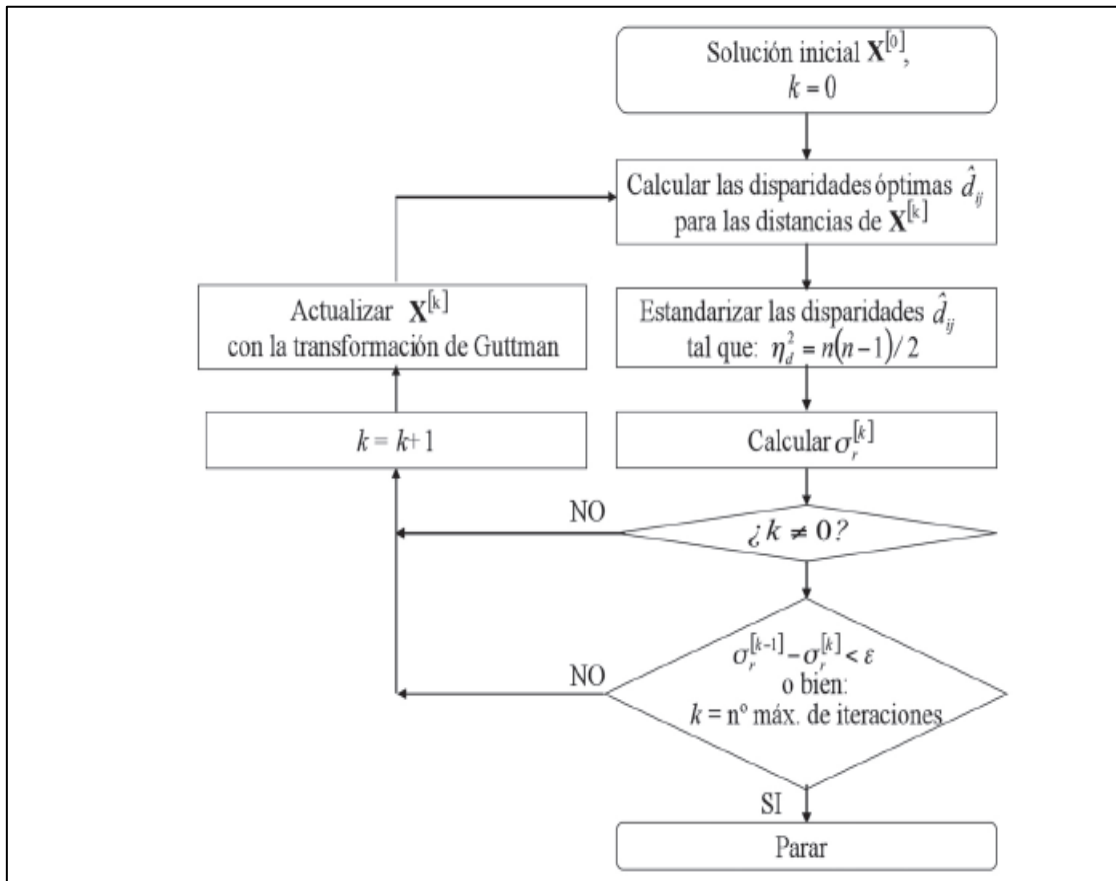


Figura 1. Esquema algoritmo SMACOF

Fuente: (Borg & Groenen, 2005)

La versión implementada de SMACOF se denomina PROXSCAL, permite elegir entre cinco transformaciones implementadas: transformaciones monótonas, lineales, polinómicas, splines o cambios de escala. Se pueden considerar también estas transformaciones para cada fuente por separado (condicionamiento matricial) o a todos los datos globalmente (nivel de medida incondicional). Todas las transformaciones se determinan mediante procedimientos de mínimos cuadrados alternantes. La minimización de esta función se conoce como un problema de proyección métrica, dado que engloba la proyección de las matrices X_r^l , en el espacio de las soluciones restringidas mediante la métrica V_r . La minimización del STRESS se lleva a cabo a través del siguiente algoritmo (De Lucas Jaramillo, 2012):

- a. Para $l = 0$ se consideran configuraciones iniciales X_r^0 , ($k=1,2,\dots,R$ y se evalúa la ecuación,

$$f(X_1^l, X_2^l, \dots, X_R^l, \phi) = \frac{1}{R} \sum_{r=1}^R \sum_{i < j} w_{ij} [\phi(d_{ijr}) - d_{ij}(X_r^l)]^2 \quad (8)$$

- b. Para cada r se calcula la transformada de Guttman,

$$X_r = V_r^{-1} B(X_r^l) X_r^l \quad (9)$$

- c. Se resuelve la ecuación del paso 1 para el problema de proyección métrica.
- d. Se evalúa la función de pérdida,

$$f(X_1^{l+1}, X_2^{l+1}, \dots, X_R^{l+1}, \phi) = \frac{1}{R} \sum_{r=1}^R \sum_{i < j}^n w_{ij} [\phi(d_{ijr}) - d_{ij}(X_r^{l+1})]^2 \quad (10)$$

- e. Se considera $l=l+1$. Si la diferencia entre los valores de la función de pérdidas en las dos estimaciones consecutivas es mayor que algún criterio determinado de antemano se vuelve al paso b, en caso contrario, el algoritmo se detiene.

Aunque no se conoce en profundidad como afecta el problema de los mínimos locales a este procedimiento, se sabe que se producen y que en muchas ocasiones estos vienen motivados por la configuración inicial que se considere. Como consecuencia, PROXSCAL dispone de un procedimiento especial para determinar la configuración inicial y comenzar el algoritmo.

Método

Se obtuvo las coordenadas o distancias a través del MDS NO METRICO, ya que los datos de entrada no son proximidades, sino que proceden de una matriz rectangular de dos vías (filas x columnas) y dos modos (sujetos x ítems). Por ello, primero se construye la matriz de proximidades calculando las distancias euclídeas entre los distintos elementos y teniendo en cuenta la métrica ordinal de las respuestas emitidas por los sujetos (una escala Likert).

Población de estudio

Datos

La muestra estuvo constituida por 251 personas, 117 hombres y 134 mujeres, en edades comprendidas entre 24 y 44 años, residentes en la ciudad de Riobamba y sus diferentes parroquias urbanas, la técnica empleada para recopilar los datos fue la encuesta digital.

Variables

La base de datos procede de la valoración que las personas han asignado a cada atributo, de acuerdo con cada uno de los ítems de la Tabla 2 sobre “atributos que influyen en la decisión de compra de productos lácteos”. Las respuestas se establecen con una escala de Likert de 5 puntos (desde 1: “NO ES IMPORTANTE”, hasta 5: “MUY IMPORTANTE”).

Ítem	Atributo
1	Precio
2	Marca
3	Calidad
4	Promociones
5	Sabor
6	Textura

7	Empaquetado
8	Valor nutricional
9	Color

Tabla 2. Atributos que influyen en la decisión de compra

Fuente: Elaboración propia

Selección de Atributos

En el momento de seleccionar los atributos utilizados en el presente trabajo, se consideraron dos elementos: a) por un lado, la bibliografía ya existente sobre esta temática; y b) por otro lado, se realizó un estudio piloto a través de una encuesta digital con una muestra de 120 personas, la cual arrojó como resultado la selección de 5 atributos preferentes al momento de comprar productos lácteos: precio, marca, calidad, sabor y valor nutricional.

Resultados.

En R se implementa el algoritmo SMACOF con el paquete que lleva el mismo nombre. Se crea un fichero de tipo texto con los datos aplicados en la encuesta, donde las filas representan el número de encuestados con el valor de la escala de Likert que cada uno ha seleccionado y las columnas representan las variables de estudio, como se muestra en la tabla 3, una parte de los datos.

	Precio	Marca	Calidad	Sabor	Valor_nutricional
1	4	2	4	4	5
2	4	2	4	4	2
3	4	2	5	4	4
4	2	1	1	1	1
5	3	2	5	5	5
6	3	4	4	4	3
7	3	2	3	3	3
8	3	1	3	3	2
9	2	3	3	3	4
10	1	4	4	4	4
11	1	3	3	3	1
12	4	2	4	5	3
13	2	1	2	4	3
14	2	2	2	3	2
15	2	1	4	5	4
16	2	3	3	3	4
17	3	3	4	5	2
18	3	2	5	4	5
19	4	3	3	5	5
20	4	3	3	4	4
21	3	2	4	5	5
22	1	2	4	4	3
23	3	3	3	3	3
24	2	2	4	5	5
25	4	3	4	3	3

	Precio	Marca	Calidad	Sabor	Valor_nutricional
26	3	3	3	3	3
27	2	4	4	4	4
28	3	1	3	3	3
29	4	3	5	5	4
30	3	2	4	4	5
31	2	4	4	5	3
32	2	3	5	5	5
33	4	5	5	5	5
34	3	4	5	4	5
35	3	3	3	3	3
36	3	3	3	3	3
37	3	3	3	3	4
38	4	5	5	5	4
39	3	3	2	5	5
40	3	2	4	4	2
41	3	4	3	3	5
42	3	3	3	3	3
43	4	1	3	3	4
44	1	1	2	2	3
45	2	2	3	4	5
46	2	2	2	4	3
47	2	4	3	5	5
48	3	3	3	3	5
49	2	2	3	5	5
50	4	4	4	3	3
51	1	5	5	5	5
52	3	3	4	4	4
53	2	2	3	4	5
54	1	1	1	3	4
55	3	3	3	3	3
56	3	3	3	3	5
57	4	1	5	5	3
58	2	2	2	3	3
59	1	3	5	5	5
60	5	4	4	2	4
61	3	3	3	3	5
62	2	2	3	4	3
63	1	2	3	3	1
64	2	1	3	3	5
65	3	2	3	3	3
66	2	1	1	1	2
67	2	2	5	5	4
68	3	1	4	3	4
69	1	2	3	5	5
70	2	2	3	3	3
71	3	1	3	3	5
72	3	3	1	3	3

	Precio	Marca	Calidad	Sabor	Valor_nutricional
73	3	4	4	4	4
74	4	4	4	4	4
75	3	3	3	3	4
76	3	2	3	3	4
77	1	2	3	2	3
78	2	4	4	4	4
79	1	1	2	4	2
80	2	4	4	5	5
81	4	2	4	5	3

Tabla 3. Matriz de datos

Fuente: Elaboración propia

Lo primero que se realiza es la lectura del fichero de datos llamado "DATOS_ENCUE.txt", con la línea de código:

```
atributos2<-read.delim ("~/TFM/DATOS_ENCUE.txt")
```

Luego es instalar el paquete SMACOF de R, mediante las líneas de comando:

```
> install.packages (smacof)
```

Aparecerá un mensaje indicando que Smacof ha sido instalado, para completar el proceso de instalación se debe instalar la librería smacof, tecleando: library (smacof). Con esto el proceso de instalación está completo.

Matriz de desemejanzas

A través del comando dist(x), que proporciona la matriz de distancias entre las filas de X; para los cual relacionamos con la variable atributos2 y sacamos su transpuesta, para finalmente obtener la matriz de distancias.

Por defecto el comando dist() trabaja con la distancia euclídea, con la que se está trabajando, pero existen la opción de trabajar con otros métodos como por ejemplo maximum, manhattan, canbera o minlowski.

```
#MATRIZ DE DISTANCIAS EUCLIDEAS
```

```
distan_euclidean<- dist(t(atributos2), method = "euclidean" )
```

```
distan_euclidean
```

La variable distan_euclidean, almacena la matriz de distancias y se puede visualizar de la siguiente manera:

```
> distan_euclidean
```

```
Precio  Marca  Calidad  Sabor
```

```
Marca      19.23538
```

```
Calidad    23.10844 24.00000
```

```
Sabor      25.82634 26.09598 14.24781
```

Valor_nutricional 26.21068 25.51470 17.46425 17.43560

MDS no métrico

Por medio de la función `smacofSym()`, se obtiene la salida del escalamiento multidimensional no métrico ordinal en dos dimensiones, ya que en este caso se trabaja con datos ordinales.

```
#algoritmo MDS NO MÉTRICO (ORDINAL)
```

```
mds_lacteos<-smacofSym(distan_euclideanas, type="ordinal")
```

```
mds_lacteos
```

Call:

```
smacofSym(delta = distan_euclideanas, type = "ordinal")
```

Model: Symmetric SMACOF

Number of objects: 5

Stress-1 value: 0.001

Number of iterations: 6

Su salida indica que se han analizado 5 objetos (variables), con 6 iteraciones y se ha obtenido un stress value de 0.001.

Para observar la configuración obtenida se utiliza el comando `summary`, de tal manera que:

```
> summary(mds_lacteos)#detalles en dos dimensiones
```

Configurations:

	D1	D2
Precio	-0.6293	0.4102
Marca	-0.6305	-0.4161
Calidad	0.2237	0.2590
Sabor	0.5823	0.1658
Valor_nutricional	0.4538	-0.4189

Stress per point (in %):

Precio	Marca	Calidad	Sabor	Valor_nutricional
0.00	50.00	24.97	0.00	25.03

Se puede observar el stress por punto, el precio y el sabor del producto tienen 0% de stress value, seguido de la calidad del producto, el valor nutricional y finalmente el más alto porcentaje se presenta en el precio del producto.

También se visualiza la salida MDS no métrico, para tres dimensiones:


```
> mds_lacteos_3<-smacofSym(distan_euclideas, ndim=3,type="ordinal")
```

```
> mds_lacteos_3
```

Call:

```
smacofSym(delta = distan_euclideas, ndim = 3, type = "ordinal")
```

Model: Symmetric SMACOF

Number of objects: 5

Stress-1 value: 0.001

Number of iterations: 11

```
> summary(mds_lacteos_3) #detalles en tres dimensiones
```

Configurations:

	D1	D2	D3
Precio	-0.5819	0.3801	-0.2154
Marca	-0.5981	-0.3931	0.2112
Calidad	0.2293	0.2117	0.1551
Sabor	0.5270	0.1054	0.2562
Valor nutricional	0.4237	-0.3041	-0.4072

Stress per point (in %):

Precio	Marca	Calidad	Sabor	Valor nutricional
42.60	7.40	0.00	40.65	9.35

Donde se puede observar, que el valor del stress no varía de 0.001, se realizan 11 iteraciones, por lo tanto, se seleccionó el MDS no métrico de dos dimensiones.

Graficas en R

```
#mapa perceptual
```

```
>plot(mds_lacteos$conf, pch=7, xlim=range(mds_lacteos$conf, repel=T,
main="Configuración SMACOF"))
```

```
> atributos<-c("Precio", "Marca", "Calidad", "Sabor", "Valor_Nutricional")
```

```
> text(mds_lacteos$conf, pos = 1, labels = atributos)
```

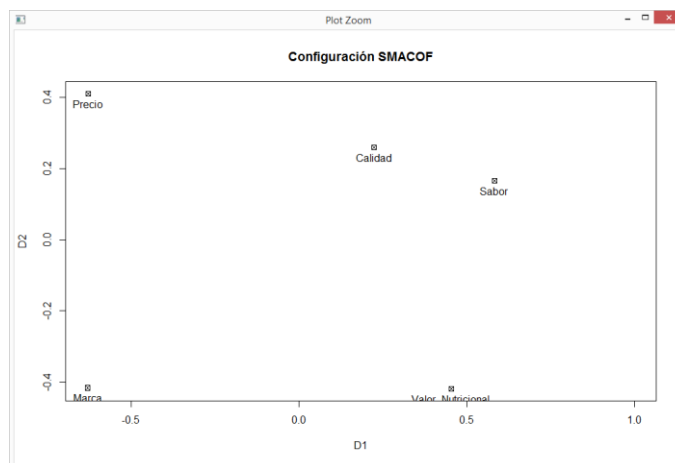


Figura 2. Configuración de variables o atributos por Smacof

Fuente: Elaboración propia

Se puede observar en la figura 2, que la variable (marca) está más aleja del resto de variables siendo el grado de importancia el de menos valor para los consumidores, seguido de la variable (precio), que serían los dos aspectos con menor grado de importancia al momento de realizar la compra de un producto lácteo. Por otro lado, el sabor es la variable que presenta mayor importancia para el consumidor al momento de realizar la adquisición de un producto lácteo, esto al analizarlo de manera individual y contrastarlo con los valores medios de las variables de estudio.

>mean (atributos2)

Precio	Marca	Calidad	Sabor	Valor nutricional
2.85	2.72	3.71	3.77	3.76

En relación con las agrupaciones de variables: Sabor, Valor nutricional y Calidad quedan asociadas con valores medios de 3.77, 3.76 y 3.71 respectivamente.

Otro grupo de variables asociadas, aunque con menor grado de importancia seria: el precio y la marca del producto con valores medios de 2.85 y 2.72 respectivamente.

La dimensión 2, muestra la pertinencia de cada variable con el comportamiento del consumidor al momento de realizar la compra de un producto lácteo, se puede observar que la marca definitivamente se encuentra alejada del resto de variables, siendo el aspecto de menor importancia para el consumidor al momento de realizar la compra de un producto lácteo, se puede observar que existe un grupo de factores que asocian y confirman que el sabor, la calidad y el valor nutricional son los valores de mayor importancia para el consumidor y el precio es un factor que representa una importancia media al momento de realizar la compra de productos lácteos.

En la figura 3, se aprecia la contribución de cada atributo en el stress.

> plot(mds_lacteos, plot.type = "stressplot", main="DESCOMPOSICIÓN DEL STRESS ENTRE LOS ATRIBUTOS")



Figura 3. Descomposición del stress entre las variables o atributos

Fuente: Elaboración propia

El atributo que más contribuye (50%) es la marca del producto, pudiendo considerarse como el atributo de más difícil ajuste por el modelo, mientras que el valor nutricional y la calidad del producto lácteo contribuyen con un 25% aproximadamente sobre el stress, y finalmente el precio y sabor son los que presentan una contribución mínima al stress.

En la figura 4, se muestra las distancias frente a los residuos, se puede observar que los errores se generan dentro de la línea de regresión isótoma, lo que sugiere la presencia de heteroscedasticidad. Además de puede observar también que los residuos tienen una dirección ascendente con pendiente de 1, es decir tenemos un buen modelo.

> plot(mds_lacteos, plot.type = "resplot", main="DISTANCIAS FRENTE A LOS RESIDUOS")

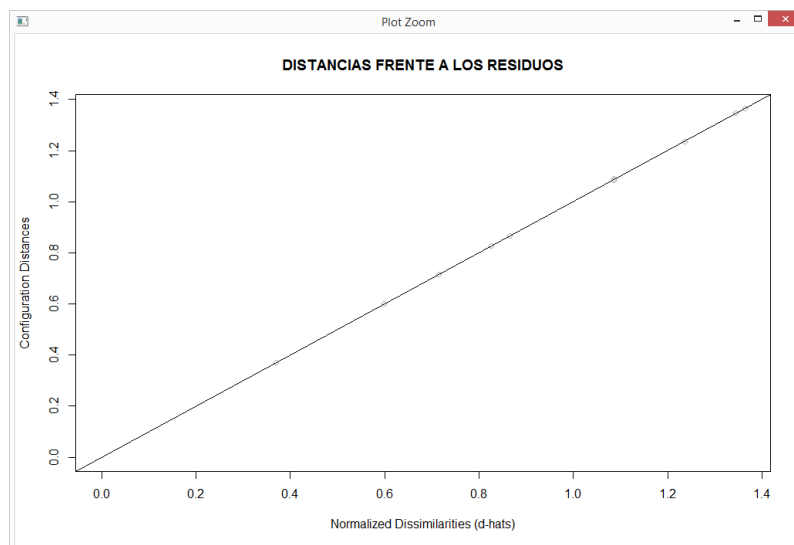


Figura 4. Valores residuales

Fuente: Elaboración propia

En la figura 5, el diagrama de Shepard muestra las disimilaridades frente a las distancias transformadas (disparidades).

> plot (mds_lacteos, plot.type = "Shepard", main="DIAGRAMA DE SHEPARD")

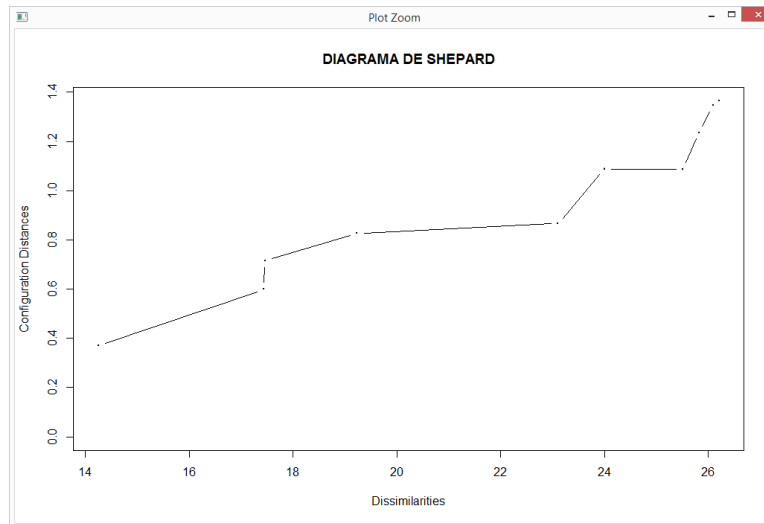


Figura 5. Diagrama de Shepard

Fuente: Elaboración propia

Conclusiones.

- El comportamiento del consumidor se ha visto afectado con los efectos que ha traído la pandemia mundial desde el mes de marzo 2020, tanto así que los consumidores al momento de comprar productos lácteos en la ciudad de Riobamba, han cambiado sus preferencias y comportamiento al decidir qué factor o atributo es el más importante al momento de comprar productos lácteos, siendo su decisión considerar el sabor, la calidad y el valor nutricional como atributos más importantes en la toma de decisión, dejando de lado la influencia de la marca y el precio.
- El comportamiento del consumidor ha descartado atributos o factores como: las promociones, la textura, el empaquetado y el color de los productos lácteos, para dar paso a factores que realmente aportan al cuidado de su salud, por sobre el valor monetario, comportamiento que se ha sido afectado directamente por efectos de la cuarentena extendida a nivel nacional e internacional, a diferencia del comportamiento generado antes de la pandemia donde se consideraba en alto grado la marca y el precio del producto lácteo.
- Con este trabajo se demuestra que la técnica de Escalamiento Multidimensional puede ser utilizada en estudios de Marketing, sobre el comportamiento del consumidor como alternativa a otras técnicas multivariantes o como su complemento.

Referencias bibliográficas

Aldana, S., Vereda, F., Hidalgo-Alvarez, R., & de Vicente, J. (2016). Facile synthesis of magnetic agarose microfibers by directed selfassembly. *Polymer*, 93, 61-64.

Arnau, J. (1996). *Métodos y técnicas avanzadas de análisis de datos en ciencias del comportamiento*. Barcelona: Universitat de Barcelona.

- Bhat, S., Tripathi, A., & Kumar, A. (2010). Supermacro porous chitosan-agarose-gelatin cryogels. *in vitro* characterization and *in vivo* assessment for cartilage tissue engineering. *Journal of the Royal Society Interface*, 1-15.
- Borg, I., & Groenen, P. (2005). *Modern Multidimensional Scaling. Theory and Applications*. Springer.
- Bossis, G., Marins, J., Kuzhir, P., Volkova, O., & Zubarev, A. (2015). Functionalized microfibers for field-responsive materials and biological applications. *Journal of Intelligent Material Systems and Structures*, 1-9.
- Cortés, J., Puig, J., Morales, J., & Mendizábal, E. (2011). Hidrogeles nanoestructurados termosensibles sintetizados mediante polimerización en microemulsión inversa. *Revista Mexicana de Ingeniería Química*, 10(3), 513-520.
- Crepaldi, M. (2020). Location regions for Interval Multidimensional Scaling. Granada.
- De Lucas Jaramillo, A. B. (2012). UGR. Obtenido de <https://masteres.ugr.es/>
- Dias, A., Hussain, A., Marcos, A., & Roque, A. (2011). A biotechnological perspective on the application of iron oxide magnetic colloids modified with polysaccharides. *Biotechnology Advances* 29, 29, 142–155.
- Estrada Guerrero, R., Lemus Torres, D., Mendoza Anaya, D., & Rodriguez Lugo, V. (2010). Hidrogeles poliméricos potencialmente aplicables en Agricultura. *Revista Iberoamericana de Polímeros*, 12(2), 76-87.
- García-Cerda, L., Rodríguez-Fernández, O., Betancourt-Galindo, R., Saldívar-Guerrero, R., & Torres-Torres, M. (2003). Síntesis y propiedades de ferrofluidos de magnetita. *Superficies y Vacío*, 16(1), 28-31.
- Ilg, P. (2013). Stimuli-responsive hydrogels cross-linked by magnetic nanoparticles. *Soft Matter*, 9, 3465-3468.
- Leeuw, J., & Mair, P. (2009). Multidimensional scaling using majorization: SMACOF in R. *Journal of Statistical Software*.
- Lewitus, D., Branch, J., Smith, K., Callegari, G., Kohn, J., & Neimark, A. (2011). Biohybrid carbon nanotube/agarose fibers for neural tissue engineering. *Advanced Functional Materials*, 21, 2624-2632.
- Lin, Y.-S., Huang, K.-S., Yang, C.-H., Wang, C.-Y., Yang, Y.-S., Hsu, H.-C., . . . Tsai, C.-W. (2012). Microfluidic synthesis of microfibers for magnetic-responsive controlled drug release and cell culture. *PLoS ONE*, 7(3), 1-8.
- Løland, A., & Høst, G. (2003). Spatial covariance modelling in a complex coastal domain by multidimensional scaling. *Environmetrics*, pp. 307-321.
- López González, E., & Hidalgo Sánchez, R. (2010). Escalamiento Multidimensional No Métrico. Un ejemplo con R empleando el algoritmo SMACOF. *Universidad de Navarra*.

- Macías, R., Rivera, C., & Vera, J. (2007). *Dialnet*. Obtenido de <https://dialnet.unirioja.es/servlet/articulo?codigo=3152246>
- Mardia , K., Kent , J., & Bibby, J. (1979). *Multivariate Analysis*. Londres. Gran Bretaña: Academic Press.
- Motterlini, M. (2006). *Economía Emocional. En qué nos gastamos el dinero y por qué*. Milán: Paidós Contextos.
- Peñaranda Arenas, I. D. (2016). *UGR*. Obtenido de https://masteres.ugr.es/moea/pages/curso201516/tfm1516/peaarandaarenas_tfm
- Ruiz Estrada, G. (2004). *Desarrollo de un Sistema de liberación de fármacos basado en nanopartículas magnéticas recubiertas con Polietilén glicol para el tratamiento de diferentes enfermedades*. Madrid: Universidad Autónoma de Madrid. Departamento de Física Aplicada.
- Song , J., King, S., Yoon , S., Cho, D., & Jeong, Y. (2014). Enhanced spinnability of carbon nanotube fibers by surfactant addition. *Fibres and Polymers*, 15(4), 762-766.
- Tartaj, P., Morales, M., González-Carreño, T., Veintemillas-Verdaguer, S., & Serna, C. (2005). Advances in magnetic nanoparticles for biotechnology applications. *Journal of Magnetism and Magnetic Materials*, 290, 28-34.
- Torgerson , W. (1952). *Multidimensional scaling: I. Theory and method*. Torgerson, W. (1952)., 17, 401-419.: Psychometrika.
- Turstone, L. (1927). *A law of comparative judgment*. Psychological Review.
- Vera, J., & Mair, P. (2019). *Researchgate*. Obtenido de https://www.researchgate.net/publication/331213974_SEMDS_An_R_Package_for_Structural_Equation_Multidimensional_Scaling/download
- Wickelmaier, F. (2003). Obtenido de <https://homepage.uni-tuebingen.de/florian.wickelmaier/pubs/Wickelmaier2003SQRU.pdf>
- Wulff-Pérez , M., Martín-Rodríguez, A., Gálvez-Ruiz, M., & de Vicente, J. (2013). The effect of polymer surfactant on the rheological properties of nanoemulsions. *Colloid and Polymer Science*, 291, 709–716.
- Zamora Mora, V., Soares, P., Echeverría, C., Hernández , R., & Mijangos, C. (2015). Composite chitosan/Agarose ferrogels for potential applications in magnetic hyperthermia. *Gels*, 1, 69-80.

PARA CITAR EL ARTÍCULO INDEXADO.

Bolaños Logroño, P. F., Márquez Sañay, F. R., & Mantilla Cabrera, C. E. (2021). Escalamiento multidimensional (MDS) no métrico en el análisis del comportamiento del consumidor en la adquisición de productos lácteos . ConcienciaDigital, 4(3), 156-176. <https://doi.org/10.33262/concienciadigital.v4i3.1788>



El artículo que se publica es de exclusiva responsabilidad de los autores y no necesariamente reflejan el pensamiento de la **Revista Conciencia Digital**.

El artículo queda en propiedad de la revista y, por tanto, su publicación parcial y/o total en otro medio tiene que ser autorizado por el director de la **Revista Conciencia Digital**.

