




Transformación y agroindustria

Artículo de investigación científica y tecnológica

Análisis multivariado de índices de consumo en procesos de obtención de leche de soya

Multivariate Analysis of Consumption Indexes During the Production of Soybean Milk

 Edilia Gutiérrez-Cordero ^{1*}  Eduardo García-Noa ²  Yanet Sariego-Toledo ²

¹ Universidad de Camagüey “Ignacio Agramonte Loynaz”, Camagüey, Cuba.

² Universidad Tecnológica de La Habana “José Antonio Echeverría”, La Habana, Cuba.

*Autor de correspondencia: Yanet Sariego Toledo. Universidad Tecnológica de La Habana “José Antonio Echeverría”, calle 114 #11901, ciclovía y rotonda, CP 19390, Mariano, La Habana, Cuba. ysariego@quimica.cujae.edu.cu

Recibido: 01 de marzo de 2022

Aprobado: 27 de julio de 2023

Publicado: 27 de octubre de 2023

Editor temático: Juan Camilo Henao, (Corporación Colombiana de Investigación Agropecuaria [AGROSAVIA]), Antioquia, Colombia.

Para citar este artículo: Gutiérrez-Cordero, E., García-Noa, E., & Sariego-Toledo, Y. (2023). Análisis multivariado de índices de consumo en procesos de obtención de leche de soya. *Ciencia y Tecnología Agropecuaria*, 24(3), e2959. https://doi.org/10.21930/rcta.vol24_num3_art:2959

Resumen: Dada la importancia del ahorro de energía y de materias primas como la vía fundamental para reducir los costos de producción y disminuir el impacto medioambiental, se definió como objetivo establecer la relación funcional entre el consumo de vapor por masa de leche producida y el de masa de leche producida por masa de materia prima utilizada con los parámetros de operación, mediante métodos multivariados en la producción de leche de soya. Se obtuvieron los datos en 25 producciones consecutivas de una planta industrial y el procesamiento incluyó la obtención de los conglomerados y los componentes principales como pasos previos a la modelación. Mediante el programa Statgraphics, se obtuvieron modelos estadísticamente significativos para esos índices, relacionándose con 20 parámetros tecnológicos medidos durante la producción y resultaron más influyentes: la masa de vapor, las temperaturas en los calentadores y la masa de leche en el intercambiador. La función obtenida con el método de mínimos cuadrados parciales con las variables no estandarizadas se empleó para evaluar alternativas de operación y en la simulación del proceso se obtuvo que se puede alcanzar un índice de 9,5 kg de leche por kg de soya empleada, donde este valor es estimado con un error absoluto del modelo ajustado de $4,18 \times 10^{-2}$. Además, como resultado se llegó a la conclusión de que se pueden producir 1050,0 t de leche de soya adicionales al año, por lo tanto, mediante los modelos obtenidos es posible establecer las mejores condiciones de operación en la planta estudiada.

Palabras clave: análisis de componentes principales, análisis de conglomerados, derivados de la soja, mínimos cuadrados parciales, simulación.

Abstract: Energy and raw materials saving is the best way to reduce production's cost and its environmental impact. Therefore, the objective of this paper was to establish the functional relationship among the indexes consumption of vapor for milk mass produced, and also the milk mass produced by raw material, using multivariate methods. The data was obtained in 25 serial productions of an industrial plant, using the Clusters Analysis and The Principal Component method. Models with statistical significance were obtained for those indexes related to 20 technological parameters measured during the productive process using the software Statgraphics. The mass of vapor, the temperatures in the heaters, and the mass of milk in the heat interchanger were the most significant variables. The function obtained with the Partial Least Square Method without standardized variables was used to simulate different operational conditions. We obtained that it could be reached an index of 9.5 kg of milk for each kg of soybean, with an absolute error of 4.18×10^{-2} . An increment of 1050,0 t of soybean milk in a year is feasible to improve the revenues of more soybean milk per kg of raw material processed. Through the obtained models it is possible to establish better operation conditions in the studied plan.

Keywords: Cluster analysis, milky products, partial least square, principal components analysis, simulation.



Introducción

Considerada por sus propiedades como un producto lácteo, la leche de soya es empleada actualmente para elaborar un conjunto amplio de derivados que tienen una gran demanda en la población por su alta digestibilidad y valor nutritivo. Precisamente se reportan varias investigaciones dirigidas al perfeccionamiento de una de las etapas del proceso, entre ellas las realizadas por Acuña Reyes et al. (2011), Chico et al. (2017), Hernández et al. (2015), Rodríguez et al. (2017a) y Hernández Beltrán et al. (2012).

Por su parte, diversos estudios evidencian la intención de generalizar el empleo de la soya como alimento humano, entre los que se encuentra el trabajo sobre los derivados de soya y la reducción de los antinutrientes de este producto presentado por Acuña Reyes et al. (2011), pero también se estudia el procesamiento del haba (Mateos-Aparicio et al., 2008), la cinética de secado de semillas de soya (Vega León et al., 2017), la calidad sensorial del néctar de mango enriquecido con soya (Montalvo-Perdomo et al., 2016) y el empleo de maíz y soya en la elaboración de pan (Hernández et al., 2015). En estos estudios destacan (entre los asociados a la producción de derivados): la elaboración de quesos mezclando leche de soya y de vaca (de Luna Jiménez, 2018), el desarrollo de un queso crema con leche de soya y búfala (Rodríguez et al., 2017b), la elaboración de leche fermentada a partir de arroz, leche de búfala y soya (Rodríguez et al., 2017a), las etapas de la producción de análogos al queso con pastas de soya (Barcenos Ochoa et al., 2017), la elaboración de yogur a base de soya con quinua y camote (Chico et al., 2017) y la estabilidad del aderezo de soya (Vives et al., 2017).

Como se aprecia, los trabajos recientes están dirigidos al mejoramiento del producto o de algunas de las etapas del proceso, pero los problemas actuales en la industria láctea requieren para su estudio de un enfoque multilateral que posibilite la intensificación global de la producción (García-Noa et al., 2015). Por otra parte, es necesario realizar un análisis de la demanda y el consumo de portadores energéticos, en especial para las industrias procesadoras de alimentos, debido a que el uso eficiente de la energía incide tanto en la calidad del producto como en su costo de producción.

La leche de soya es el producto obtenido de la trituración en medio acuoso caliente del cotiledón del frijol de soya, al que se le inactivan las enzimas y se reducen los carbohidratos indeseables mediante un calentamiento severo, además de requerir condiciones especiales de enfriamiento (Resolución N.º 139/08). Debido a sus requerimientos, el proceso es costoso y de ahí la importancia de establecer condiciones de operación que resulten eficientes en el uso de la materia prima y los portadores energéticos, quedando aún mucho potencial por aprovechar en la tecnología en estudio (García-Noa et al., 2015).

Una vía para definir las mejores condiciones de operación mediante un estudio integral es el empleo de las técnicas del análisis de proceso, que es un método general de investigación científica. Este se ha empleado como el fundamento de una estrategia experimental para el aumento de la productividad de los procesos, basado en la modelación matemática y la simulación.

Los modelos matemáticos para la simulación pueden ser obtenidos al emplear los métodos de la matemática multivariada, la cual agrupa a un conjunto de técnicas que permiten al investigador interpretar y visualizar conjuntos grandes de datos (tanto en individuos como en variables), a partir de su simplificación o reducción. Entre las técnicas multivariadas más utilizadas para hallar la relación entre las variables se encuentran: el análisis de conglomerados, los componentes principales y los mínimos cuadrados parciales (Cerretani et al., 2011; Patras et al., 2011; Burbano & Moreno, 2018). En estos métodos, las variables están correlacionadas entre sí y generalmente se supone que el conjunto de variables que intervienen en el análisis poseen una distribución normal multivariada. Esta suposición permite que el análisis multivariado se desarrolle paralelamente al correspondiente análisis univariado basado en una distribución normal (Lima de Aguiar et al., 2009; Ramírez et al., 2017; Barrera et al., 2020).

Ante la necesidad de definir las condiciones de operación que incrementen la eficiencia del proceso, se realizó el estudio para dar respuesta a la siguiente interrogante: ¿es posible establecer modelos matemáticos que permitan conocer el efecto de los cambios en las variables del proceso sobre los indicadores productivos? A partir de la situación planteada, se definió como objetivo establecer la relación funcional entre el consumo de vapor por masa de leche producida y el de masa de leche producida por masa de materia prima utilizada, con los parámetros de operación mediante métodos multivariados en la producción de leche de soya. Dada la situación, en esta tecnología puede plantearse que el uso poco eficiente de las materias primas y de los portadores energéticos en algunas etapas de la producción de leche de soya provocan pérdidas y encarecen el producto, por lo que a medida en que en cada fase del proceso se trabaje óptimamente, se reducirán los costos y se hará más atractivo desde el punto de vista económico. Estas condiciones de operación pueden determinarse al emplear los modelos ajustados.

Materiales y métodos

El análisis del proceso de obtención de la leche de soya en la línea de producción se desarrolló con una selección previa de 20 variables, por ser las que están relacionadas con el uso de los materiales y de la energía en cada etapa del proceso, por lo que es de esperar que tengan una buena correlación con los índices de consumo que se calculen. Además, se recopilaron valores en 25 producciones en la misma instalación, de ellas 14 con 5600 l de leche producida y el resto fueron producciones de aproximadamente 15.000 l.

Independientemente del volumen de producción, la leche se debe obtener con la misma composición y calidad dada por la norma del proceso, diferenciándose solo en el uso que se le da a la leche. Las producciones de 15.000 l se emplearon posteriormente para producir yogur de soya y con las de 5600 l se produjeron otros derivados, como crema ácida, helados, entre otros, por lo que la relación entre las variables medidas y los índices de consumo fueron las mismas, independientemente del volumen de producción.

La variable medida o calculada con los datos primarios fue la pureza del cotiledón de soya empleado en el proceso (P_s), que es un dato obtenido del laboratorio de control de calidad de la fábrica; para conocer la masa de soya hidratada que entró al molino 1 (m_{sMI}) se midió

directamente la masa de soya que entra al proceso y se le sumó el agua que se alimentó al molino I (FaM_I). Este flujo y el del agua que entró a la canal con tornillo sinfín (Fae) fueron medidos con los flujómetros instalados en línea que poseen una precisión de ± 1 l/min.

Por su parte, las masas de los otros componentes en el proceso se calcularon a partir de la medición del flujo del fluido, medido con un flujómetro en línea (que poseen una precisión de ± 1 l/min), multiplicado por el tiempo de operación. Las variables obtenidas por esta vía fueron: masa del agua que entra al calentador de agua (m_{aCA}), masa del agua que se suministra al molino 1 (m_{aM1}), masa del agua que va a la canal con tornillo sinfín (m_{aTSF}), masa de leche de soya que va al calentador de leche (m_{lCL}) y masa de leche de soya que va al intercambiador de tubo en tubo (m_{lTT}).

Las masas de vapor empleadas en las etapas de calentamiento a lo largo del proceso se calcularon dividiendo la energía transferida por el vapor entre la variación de entalpía del vapor a la presión correspondiente. Estos cálculos se emplearon para conocer las siguientes variables: masa de vapor que entra al calentador de agua (m_{vCA}), masa de vapor que se suministra al calentador de leche (m_{vCL}) y masa de vapor que se suministra a la canal con tornillo sinfín (m_{vTSF}). Así, la masa de vapor total consumida en el proceso (m_{vT}) resulta de la suma de las masas anteriores.

Ahora, la energía transferida utilizada en los cálculos anteriores se obtiene por la ecuación 1:

$$Q_{\text{absorbido}} = M C_p \Delta T \quad (\text{ecuación 1})$$

Donde:

Q absorbido: calor absorbido por el componente (kJ)

M: masa del componente (kg)

C_p: capacidad calórica del componente (kJ/ kg °C)

ΔT: gradiente de temperatura del fluido que se calienta (°C)

Los valores de temperatura de las corrientes se obtuvieron directamente en cada equipo mediante termómetros bimetálicos instalados en línea, los cuales poseen una precisión de ± 1 °C y donde las variables obtenidas por esta vía fueron: temperatura del agua que entra al calentador de agua (T_{aCA}), temperatura del agua que entra al molino 1 (T_{aM1}), temperatura de la leche de soya que entra al calentador de leche (T_{lCL}), temperatura de la leche de soya que entra al tanque *flash* (T_{lTF}), temperatura de la leche de soya que entra al tubo de retención (T_{lTR}), temperatura de la leche de soya que entra al intercambiador de tubo en tubo (T_{lTT}), temperatura de la solución de bicarbonato de sodio (T_{sb}) y temperatura inicial de la soya (T_s).

Se consideró como tiempo de molido (t_m) el periodo transcurrido desde que comienza a caer el grano en el molino hasta que terminó esta operación, donde para la tabulación y la organización de todos estos datos se utilizó el programa Microsoft Excel y en el procesamiento se empleó el programa Statgraphics Centurion.

La modelación multivariada del proceso se realizó siguiendo un procedimiento general a través de cuatro pasos fundamentales, donde se tuvieron en cuenta experiencias anteriores en la aplicación de las técnicas del análisis de procesos (Zuluaga Domínguez, 2011; Hernández Beltrán et al., 2012; Barrera et al., 2020):

1. Análisis de datos y sus correlaciones
2. Análisis de conglomerados
3. Análisis de componentes principales
4. Ajuste de modelos por mínimos cuadrados parciales (MCP)

En el análisis de los datos se describió el comportamiento de las variables, al calcular para cada una de ellas: el promedio, los valores máximos y mínimos y el coeficiente de variación, además se obtuvieron los coeficientes de correlación entre todas las variables analizadas.

Para la obtención de los conglomerados se utilizaron el método del vecino más cercano y la métrica de distancia euclidiana cuadrada y se tuvieron en cuenta experiencias anteriores en producciones lácteas (Avilez et al., 2010; García-Noa et al., 2015; Barrera et al., 2020), además, se estandarizaron los datos para que todas las variables estuvieran dentro de un mismo intervalo de valores. El análisis se realizó para las observaciones (producciones) buscando diferencias y similitudes entre ellas.

Para identificar el número óptimo de conglomerados se empleó el método del gráfico de la distancia de aglomeración, que es la herramienta de la que dispone el Statgraphics Centurion para esto (Statgraphics, 2016). Al aplicar este método, el número de conglomerados óptimo es el que produzca la mayor diferencia entre las distancias de los centroides de los grupos sucesivos, lo que se aprecia en el cambio brusco en esta distancia al recorrer los diferentes estados, así, un estado se corresponde con un nuevo grupo formado. Este método es similar al método del codo (*Elbow Method*), el cual se encuentra en otros *softwares* como el R, en el que la selección del número de conglomerados óptimo se basa en la inercia, definida esta como la suma de las distancias al cuadrado de cada objeto del conglomerado a su centroide, donde en este método se produce en el óptimo, un cambio brusco en el gráfico de inercia contra el número de grupos (Tibshirani et al., 2001). En el *software* Statgraphics, el procedimiento de componentes principales está diseñado para extraer los componentes principales de un conjunto de variables cuantitativas y para aplicarlo se estandarizaron los datos experimentales, de forma que todas las variables estuvieran dentro de un mismo intervalo de valores (Statgraphics, 2016).

El objetivo fundamental de aplicar la técnica multivariada del análisis de componentes principales es la reducción de variables mediante la obtención de correlaciones lineales entre las variables, lo que conforma los componentes que explican la mayor variabilidad posible de los datos. Además, se emplea como paso previo al método de mínimos cuadrados parciales para definir el número de componentes a utilizar en los modelos matemáticos a ajustar. Se tiene en cuenta que la selección del número de factores adecuados se realizó mediante el gráfico de sedimentación, el cual muestra los porcentajes de varianza que explican cada uno de los componentes por separado. Así, en este gráfico se busca el número de componentes que explique un porcentaje acumulado superior al 85 % de la varianza total de los puntos experimentales (Piovoso & Hoo, 2002; Statgraphics, 2016).

En el análisis por MCP se ajustó la cantidad de componentes que se obtienen del procedimiento de componentes principales con 20 variables independientes, donde se declara variable dependiente de un modelo la razón que relaciona la masa de vapor total y la masa de leche obtenida (I_1) y del otro modelo la razón que establece la relación entre la masa de leche producida y la masa de soya inicial (I_2).

Para cada variable dependiente (I_1 e I_2) se obtuvieron dos modelos, uno con las variables experimentales estandarizadas centrando por la media, el cual se empleó para identificar las variables significativas estadísticamente y conocer la influencia de cada una de ellas sobre dicho índice. El otro modelo para cada índice se obtuvo con las variables no estandarizadas y se empleó para realizar la simulación con las unidades de medidas originales de cada variable independiente. En total se ajustaron cuatro modelos por MCP y la significación de los modelos se analizó con un 95 % de confiabilidad a través del valor-p de la razón de Fisher. La calidad de predicción se analizó a través del valor Press (sumatoria cuadrada del error de predicción) (Piovoso & Hoo, 2002; Statgraphics, 2016).

Los modelos se validaron por el método “sacar uno a la vez”, en el cual se ajusta el modelo y se deja una producción sin incluir que después se utiliza para calcular el error de predicción. Este procedimiento se repite tantas veces como cantidad de producciones hay y, finalmente, se reporta el promedio de los errores obtenidos en dichos cálculos. El modelo obtenido por el método de MCP con las variables no estandarizadas se empleó para evaluar alternativas de mejora en las condiciones de operación mediante la simulación, donde se busca que se incremente el rendimiento en la etapa y se tiene en cuenta que los valores de las variables de operación que más favorezcan este propósito estén dentro de los valores alcanzados en las producciones estudiadas. Para ello, se consideró el signo de los coeficientes de las variables y los valores experimentales, donde los coeficientes con signo positivo indican que se deben utilizar a los mayores valores de las variables y los de signos negativos a los menores valores de las variables para obtener el mayor rendimiento (Barrera et al., 2020).

Resultados y discusión

El análisis de los datos comenzó al determinar el coeficiente de correlación entre todas las variables medidas durante el proceso productivo, identificándose aquellas con correlaciones estadísticamente significativas, estas se muestran en el correlograma de la figura 1, donde también se aprecia que la mayoría de los coeficientes obtenidos se encontraron entre 0,60 y 0,90, lo que manifiesta un vínculo estrecho entre los valores de estas variables.

El hecho de haber obtenido correlaciones significativas entre un gran número de variables medidas justifica el procesamiento de estas por el método de los mínimos cuadrados parciales. Todas estas variables fueron incluidas en los análisis posteriores al considerar, además, que coinciden con variables empleadas en otros trabajos de evaluación del proceso de este producto (Rodríguez et al., 2017a) y estudios realizados con productos similares (Avilez et al., 2010; García-Noa et al., 2015).

El análisis de conglomerados se realizó para todas las observaciones con el método de vecino más cercano y la métrica de distancia euclidiana cuadrada. En la figura 2 se muestra el gráfico de la distancia de aglomeración que se empleó para identificar el número óptimo de conglomerados a obtener y en ella se aprecia que el cambio brusco en la distancia entre los centroides de los grupos se obtiene al arribar al estado 23, el cual se corresponde con dos conglomerados, por lo que se considera que el número óptimo de conglomerados es dos.

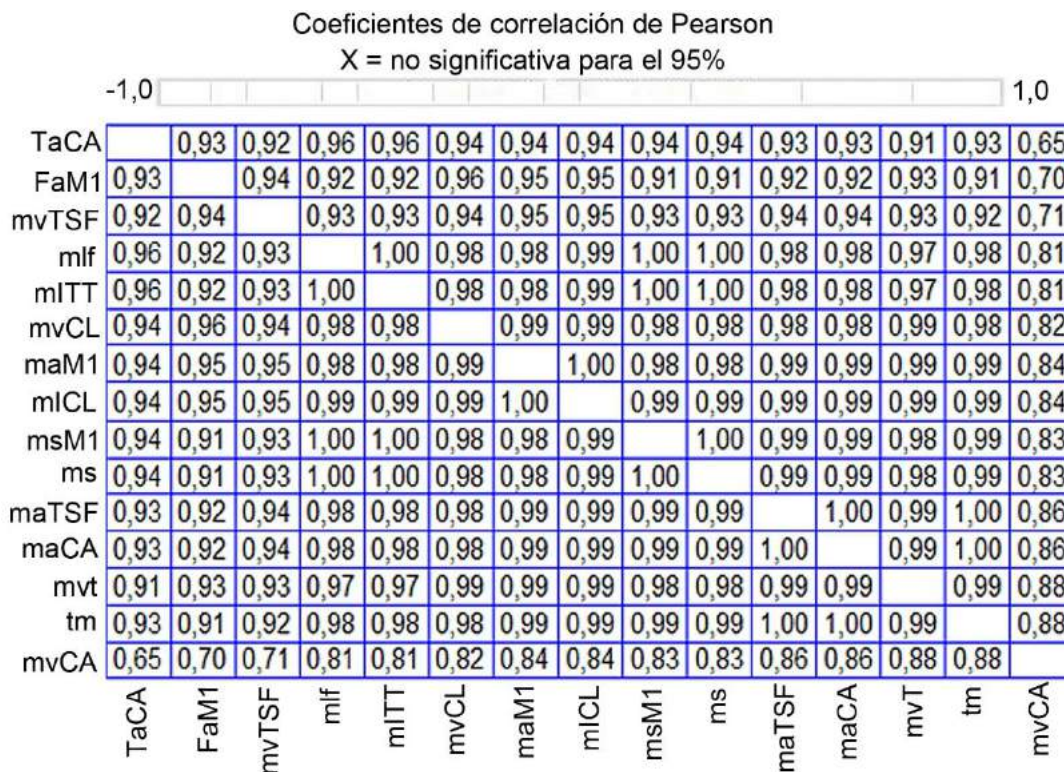


Figura 1. Correlograma de las variables medidas durante el proceso con correlación estadísticamente significativa para un 95 % de confiabilidad.

Fuente: Elaboración propia

Que el número óptimo de conglomerados sea dos es compatible con las características de la base de datos experimental. En el dendrograma obtenido (figura 3) se observan los dos grandes grupos: uno formado por las primeras 14 corridas, que se corresponden con las producciones de 5600 l y otro grupo formado por las últimas 11, es decir, desde la corrida 15 a la 25, que son las producciones de 15.000 l.

Del primer grupo, cinco producciones (8, 9, 10, 11 y 14) se asemejan entre sí, que es un reflejo de que se empleó soya procedente de un suministrador diferente al que la proveyó en el resto de las producciones, lo que está reportado en los registros del laboratorio de control de calidad. En el otro grupo se refleja también este comportamiento, aunque menos marcado. Dentro del primer grupo es de notar que las producciones 12 y 13 son las de mayor distancia, lo que coincide con que son en las que hubo un mayor volumen de producción en las mismas condiciones de

operación antes mencionadas. Del segundo grupo, la producción 24 es la menos semejante al resto de las producciones, lo cual se debe a que el volumen procesado es menor al resto.

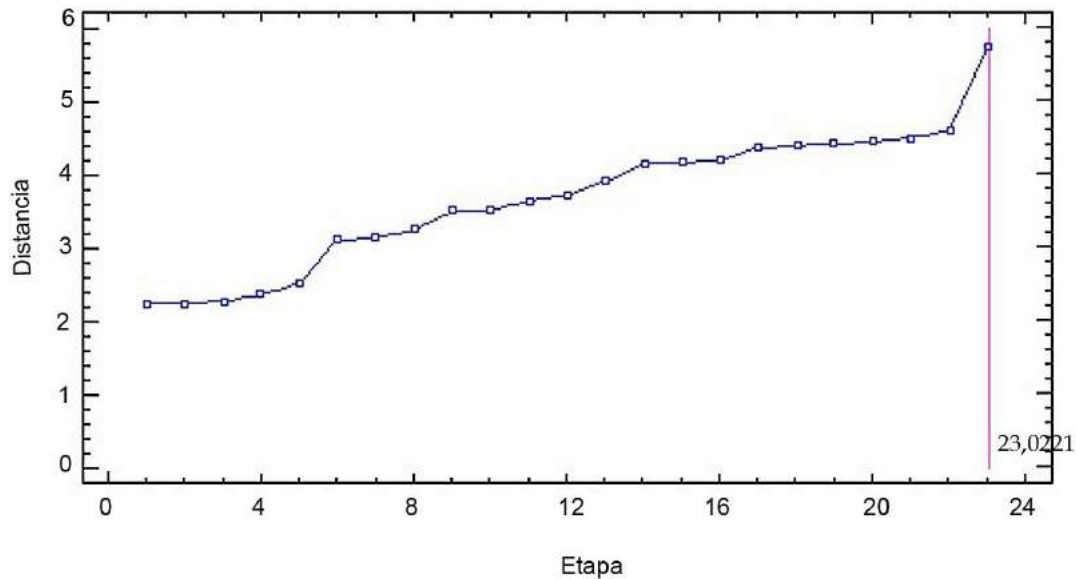


Figura 2. Gráfico de la distancia de aglomeración
Fuente: Elaboración propia

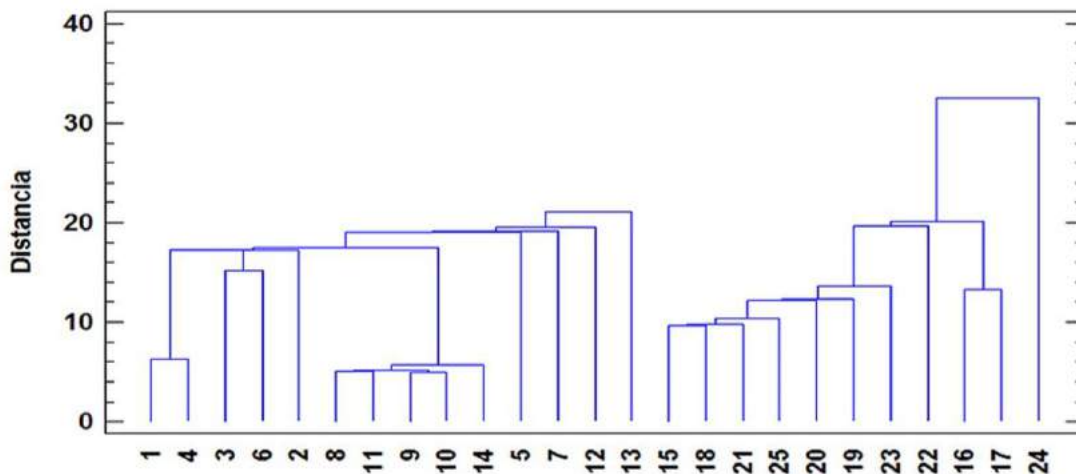


Figura 3. Dendrograma de las observaciones del método del vecino más cercano con métrica distancia euclidiana cuadrada.

Nota aclaratoria: las producciones de la 1 a la 14 corresponden a producciones de 5600 l y de la 15 a la 25 a 15.000 l.

Fuente: Elaboración propia

Es importante destacar la utilidad de este método multivariado para identificar semejanzas y diferencias en una misma línea productiva, lo cual se puede apreciar en el dendrograma obtenido en el análisis del proceso (Ruiz-Tamayo et al., 2021; Zuluaga Dominguez, 2011). Una

herramienta multivariada muy útil para decidir el número de componentes a utilizar en los modelos de componentes principales es el gráfico de sedimentación, el cual se muestra en la figura 4, donde se aprecia que basado en los vectores propios (*eigenvalues*), cuando se emplean cuatro o cinco componentes principales, los vectores propios están muy cerca de la unidad.

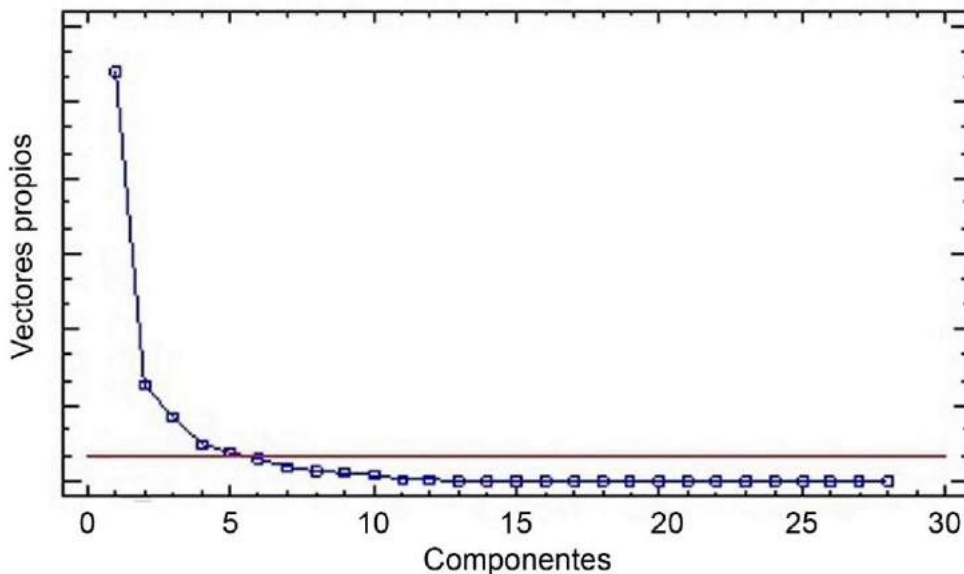


Figura 4. Gráfico de sedimentación basado en los vectores propios

Fuente: Elaboración propia

En el análisis de componentes principales (tabla 1) se aprecia que con cuatro componentes se explica el 85,20 % de la varianza total de los puntos experimentales y con cinco componentes se explica el 89,23 % de la varianza total. Ambos valores son aceptables, ya que se garantizan con valores propios superiores a uno, que es lo requerido por el método. Se puede observar que un quinto componente incrementaría en solo un 4 % el acumulado de la varianza explicada, por lo que se justifica el uso de cuatro componentes en el ajuste por mínimos cuadrados parciales.

Tabla 1. Resultados del análisis de componentes principales

Componente número	Porcentaje de la varianza explicada	Porcentaje acumulado
1	55,1	55,1
2	11,9	67,0
3	10,0	77,0
4	8,1	85,2
5	4,0	89,2
6	3,5	92,7
7	2,4	95,1
8	1,6	96,8
9	1,1	97,9
10	0,9	98,8

Fuente: Elaboración propia

Con vistas a obtener la relación funcional entre las variables del proceso y los índices relacionados con la masa de vapor total y la masa de leche obtenida (I_1) y la relación entre la masa de leche producida y la masa de soya inicial (I_2), y al tener en cuenta los resultados obtenidos en el análisis de componentes principales, se aplicó el método de mínimos cuadrados parciales con cuatro componentes.

Los modelos obtenidos para los índices I_1 e I_2 con los coeficientes estandarizados se muestran en las ecuaciones 2 y 3, respectivamente:

$$I_1 = 0,4392 F_{ac} - 0,1101 F_{aM1} - 0,0140 m_{aCA} - 0,0482 m_{aM1} - 0,0253 m_{aTSF} - 0,2945 m_{ITT} + 0,4459 m_{vCA} - 0,0571 m_{vCL} - 0,0241 m_{vTSF} + 0,0714 P_s + 0,1203 T_{a2} - 0,0677 m_{ICL} - 0,3924 T_{aCA} + 0,2241 T_{aTT} + 0,2220 T_{ICL} - 0,0358 T_{ITF} - 0,1668 T_{ITT} - 0,0358 t_m + 0,0788 T_s - 0,0334 T_{Sb} \quad (\text{ecuación 2})$$

$$I_2 = - 0,1678 F_{ac} + 0,1080 F_{aM1} + 0,0684 m_{aCA} + 0,0879 m_{aM1} + 0,0744 m_{aTSF} + 0,2010 m_{ITT} - 0,1256 m_{vCA} + 0,0920 m_{vCL} + 0,0567 m_{vTSF} - 0,0303 P_s - 0,0204 T_{a2} + 0,0970 m_{ICL} + 0,2303 T_{aCA} - 0,0968 T_{aTT} - 0,0473 T_{ICL} + 0,0473 T_{ITF} + 0,0352 T_{ITT} + 0,838 t_m - 0,0272 T_s + 0,0134 T_{Sb} \quad (\text{ecuación 3})$$

En el análisis de varianza y de calidad de ajuste del modelo para I_1 (tabla 2 y 3) se observa que el modelo de la razón, entre la masa de vapor total y la masa de leche obtenida (I_1), presentó buena calidad de ajuste para un 95 % de confiabilidad, con un valor-p inferior a 0,05. Además, el Press fue un valor del orden de $2,00 \times 10^{-4}$ que indica un error de predicción pequeño. Estos resultados reflejan lo adecuado del modelo obtenido, el cual puede ser útil para la evaluación energética del proceso, sustituyendo los valores de los parámetros de operación que relaciona, como ha sido empleado en otras investigaciones (Valencia García et al., 2007; Sariego Toledo et al., 2017).

Tabla 2. Análisis de varianza y calidad de predicción del ajuste de mínimos cuadrados parciales para la razón masa de vapor total y la masa de leche obtenida (I_1)

Fuente	Suma de cuadrados	Gl	Cuadrado medio	Razón-F	Valor-p
Modelo	0,0130	4	0,003252		
Residuos	0,0019	20	0,000096	33,6228	< 0,001
Total (corr.)	0,0149	24			

Fuente: Elaboración propia

Tabla 3. Calidad de predicción del ajuste de mínimos cuadrados parciales para la razón masa de vapor total y la masa de leche obtenida (I_1)

Componente	% variación en Y	R-Cuadrada	Cuadrado medio (Press)	Predicción R-Cuadrada
1	49,36	49,36	$3,00 \times 10^{-4}$	40,46
2	31,63	81,00	$3,00 \times 10^{-4}$	34,67
3	3,72	84,72	$2,00 \times 10^{-4}$	56,13
4	2,33	87,05	$2,00 \times 10^{-4}$	61,20

Fuente: Elaboración propia

Para el modelo de la relación entre la masa de leche producida y la masa de soya inicial (I_2), también se obtiene una buena calidad de ajuste para un 95 % de confiabilidad, pues el p-valor para la razón de F es inferior a 0,05 (tablas 4 y 5) y posee un Press igual a $4,18 \times 10^{-2}$.

Mediante el análisis de los coeficientes estandarizados fue posible identificar la influencia de cada una de las variables independientes sobre los índices calculados y se observa que las seis variables que inciden en mayor proporción en la razón I_1 son: la masa de vapor que entra en el calentador de agua (m_{vCA}), el flujo de agua de enjuague que se desecha de la canal con tornillo sinfín (F_{ae}), la temperatura del agua que entra al calentador de agua (T_{aCA}), la masa de leche que va al intercambiador de tubo en tubo (m_{ITT}), la temperatura del agua que va al intercambiador de tubo en tubo (T_{aTT}), así como la temperatura de la leche que entra al calentador de leche (T_{ICL}).

Tabla 4. Análisis de varianza en el ajuste de mínimos cuadrados parciales para la razón la masa de leche producida y la masa de soya inicial (I_2)

Fuente	Suma de cuadrados	Gl	Cuadrado medio	Razón-F	Valor-p
Modelo	15,7201	4	3,9300		
Residuos	0,4062	20	0,0203	193,4640	< 0,001
Total (corr.)	16,1264	24			

Fuente: Elaboración propia

Tabla 5. Calidad de predicción del ajuste de mínimos cuadrados parciales para la razón la masa de leche producida y la masa de soya inicial (I_2)

Componente	% de variación en Y	R-cuadrada	Cuadrado medio (Press)	Predicción R-cuadrada
1	92,61	92,61	$6,13 \times 10^{-2}$	90,48
2	2,17	94,79	$6,89 \times 10^{-2}$	89,31
3	1,42	96,21	$5,99 \times 10^{-2}$	90,70
4	1,26	97,48	$4,18 \times 10^{-2}$	93,51

Fuente: Elaboración propia

Sobre la razón I_2 , las variables que inciden en mayor proporción son: la temperatura del agua que entra al calentador de agua (T_{aCA}), la masa de leche que va al intercambiador de tubo en tubo (m_{ITT}), la masa del vapor que entra al calentador de agua (m_{vCA}), el flujo de agua que entra al molino 1 (F_{aM1}), la temperatura de la leche que va al calentador de leche (T_{ICL}), así como la masa de leche que entra al calentador de leche (m_{ICL}), es decir, son variables que modifican significativamente los índices obtenidos, por lo que se debe tener un control riguroso sobre ellas.

Es de destacar que hay cuatro variables que coinciden en ambos modelos y que por lo tanto son aquellas sobre las cuales hay que tener un mayor control durante el proceso de producción. Este resultado indica la conveniencia de que en cada producción se garantice que la masa del vapor consumida en el calentador de agua, la temperatura de alimentación a este equipo, la temperatura del fluido que alimenta al calentador de leche y la masa de leche que entra al intercambiador cumplan con las especificaciones del control del proceso, pues inciden significativamente en la eficiencia productiva.

El modelo obtenido con las variables no estandarizadas se empleó para evaluar alternativas en las condiciones de operación, al buscar que mejoren los indicadores que caracterizan la eficiencia del proceso y tener en cuenta los valores de las variables de operación que más favorezcan este propósito.

Los modelos obtenidos para los índices I1 e I2 con los coeficientes no estandarizados se muestran en las ecuaciones 4 y 5, respectivamente:

$$I_1 = -0,3129 + 2,00 \times 10^{-3}F_{ac} - 2,0 \times 10^{-4}F_{aMI} - 1,20 \times 10^{-7}m_{aCA} - 3,13 \times 10^{-7}m_{aMI} + 2,90 \times 10^{-7}m_{aTSF} - 1,65 \times 10^{-6}m_{ITT} - 8,40 \times 10^{-5}m_{vCA} - 2,00 \times 10^{-6}m_{vCL} - 1,45 \times 10^{-5}m_{vTSF} - 2,01 \times 10^{-3}P_s + 2,30 \times 10^{-3}T_{a2} - 3,95 \times 10^{-7}m_{ICL} + 9,00 \times 10^{-4}T_{aCA} + 4,80 \times 10^{-3}T_{aTT} + 1,60 \times 10^{-3}T_{ICL} - 7,00 \times 10^{-4}T_{ITF} - 1,80 \times 10^{-3}T_{ITT} - 2,35 \times 10^{-5}t_m + 1,60 \times 10^{-3}T_s - 1,70 \times 10^{-5}T_{sb}$$

(ecuación 4)

$$I_2 = 9,3378 - 2,61 \times 10^{-2}F_{ac} + 9,13 \times 10^{-3}F_{aMI} + 1,9 \times 10^{-5}m_{aCA} + 1,8 \times 10^{-5}m_{aMI} + 2,8 \times 10^{-5}m_{aTSF} + 3,7 \times 10^{-5}m_{ITT} - 7,7 \times 10^{-4}m_{vCA} + 1,4 \times 10^{-4}m_{vCL} + 1,12 \times 10^{-3}m_{vTSF} - 3,01 \times 10^{-2}P_s - 1,28 \times 10^{-2}T_{a2} + 1,8 \times 10^{-5}m_{ICL} + 1,74 \times 10^{-2}T_{aCA} - 6,86 \times 10^{-2}T_{aTT} - 2,51 \times 10^{-2}T_{ICL} + 2,9 \times 10^{-2}T_{ITF} + 1,26 \times 10^{-2}T_{ITT} + 1,7 \times 10^{-3}t_m - 1,86 \times 10^{-2}T_s + 2,2 \times 10^{-3}T_{sb}$$

(ecuación 5)

Se comprobó la posibilidad de aplicar estas condiciones en la práctica al analizar que las variables pueden asumir esos valores sin ninguna contradicción desde el punto de vista técnico.

En la tabla 6 se presentan los valores utilizados para las distintas variables en la simulación, seleccionados entre los valores reales que tomó cada variable. Para coeficientes con signos positivos se escogió el mayor valor de la variable y para coeficiente con signo negativo el menor valor real de la variable.

Como resultado de realizar la simulación del proceso se obtuvo que se puede alcanzar un índice de 9,5 kg de leche producida por cada kg de soya procesada. Este valor, al ser estimado con el modelo ajustado de mínimos cuadrados parciales, posee un error de $4,18 \times 10^{-2}$ y es superior a 8,3 kg de leche producida por cada kg de soya procesada, que es el mayor valor obtenido en las producciones analizadas. Con este resultado se establece que existe un conjunto de condiciones de operación que puede garantizar el incremento de la eficiencia del proceso productivo.

Tabla 6. Condiciones de operación utilizadas para la simulación de I_2

Variables	Unidades de medida	Valores propuestos	Variables	Unidades de medida	Valores propuestos
F_{ac}	L/min	49	T_{a2}	°C	83
F_{aM1}	L/min	95	m_{iCL}	kg	14845
m_{aCA}	kg	11379,6	T_{aCA}	°C	55
m_{aM1}	kg	13087	T_{aTT}	°C	25
m_{aTSF}	kg	8427	T_{iCL}	°C	75
m_{iTT}	kg	14913,8	T_{iTF}	°C	148
m_{vCA}	kg	454	T_{iTT}	°C	92
m_{vCL}	kg	1773	t_m	min	145
m_{vTSF}	kg	243	T_s	°C	25
P_s	%	94,5	T_{sb}	°C	75

Fuente: Elaboración propia

El efecto del cambio en las condiciones de operación calculado por la diferencia entre el mayor índice real (I_2) y el que se obtiene al trabajar bajo las condiciones recomendadas en la tabla 6 resultó ser positivo para un incremento de la producción de leche de soya anual en 1050 t, lo cual debe repercutir favorablemente en el balance económico de la fábrica por concepto de más leche de soya producida por kg de grano de materia prima procesada. Un efecto similar se ha reportado en la producción de queso crema (Valencia García et al., 2007).

Dadas las semejanzas del proceso de producción de la leche de soya con otros productos lácteos y en general con otros muchos productos alimentarios y, al tener en cuenta la necesidad de realizar los estudios con un enfoque multilateral (García et al., 2015, Juárez-Barrientos et al., 2016; Pérez-Sánchez et al., 2020), puede considerarse que el procedimiento a través de los cuatro pasos fundamentales utilizados en esta investigación tiene un grado de generalidad, lo que lo hace atractivo para servir de base metodológica en el empleo de las técnicas del análisis de proceso para el aumento de la productividad de las tecnologías. Así, será necesario, ante todo, un estudio detallado de la tecnología a analizar y definir las etapas y las variables de operación que intervienen en el proceso para luego realizar las mediciones y procesarlas para la obtención de los modelos que permitan evaluar condiciones de operación que mejoren la eficiencia.

Conclusiones

El análisis de conglomerados es una técnica multivariada que permitió clasificar las producciones más semejantes de la línea de producción de leche de soya y discriminar aquellas con diferentes volúmenes de producción, siendo dos el número óptimo de conglomerados para esta base de

datos. El modelo multivariado obtenido con cuatro componentes principales mediante el método de mínimo cuadrado parciales para estimar la relación entre la masa de vapor total y la masa de leche de soya obtenida (I_1), en función de 20 variables de operación, presentó una buena calidad de ajuste para un 95 % de confiabilidad, con un error de predicción pequeño del orden de $2,00 \times 10^{-4}$, el que es útil para la evaluación energética del proceso.

El modelo de regresión por mínimos cuadrados parciales, que establece la relación funcional entre la razón de la masa de leche producida, la masa de soya inicial (I_2) y las 20 variables de operación, presenta una buena calidad de ajuste y posee un error cuadrático de predicción igual a $4,18 \times 10^{-2}$, lo que indica la posibilidad de su empleo para simular este proceso. Bajo las condiciones de operación propuestas y evaluadas con el modelo multivariado ajustado, se estima que se pueden obtener 9,5 kg leche por cada kg de granos de soya empleada en la producción, lo que mejora el balance económico de este proceso.

Agradecimientos

Los autores agradecen al personal de la unidad empresarial de Base Ciego de Ávila de Iproyaz por su valiosa y desinteresada colaboración.

Descargos de responsabilidad

Todos los autores hicieron contribuciones significativas en el trabajo, no existen conflictos de intereses y están de acuerdo en la publicación de estos resultados.

Contribución de los autores

Edilia Gutiérrez Cordero: registro de información en campo, construcción de las bases de datos, análisis de la información y elaboración del manuscrito; Yanet Sariego Toledo: análisis de la información y elaboración del manuscrito; Eduardo García Noa: diseño de las metodologías, registro de la información en campo, supervisión de las actividades, análisis de la información y elaboración del manuscrito.

Implicaciones éticas

El presente artículo cuenta con el aval que consta en el acuerdo 3 del acta con fecha 10/12/2019 del Comité del programa de Análisis y Control de Procesos de la Facultad de Ingeniería Química de la Universidad Tecnológica de La Habana. Se obtuvo el consentimiento de la planta objeto de estudio para usar la información suministrada en la documentación del proceso presentado en el artículo.

Conflicto de interés

Los autores declaran que no existe conflicto de intereses. Todos los autores del artículo declaramos que estamos de total acuerdo con lo escrito en este informe y aprobamos la versión final.

Financiación

La investigación se desarrolló sin contar con financiamiento de ninguna organización.

Referencias

- Acuña Reyes, M. J., Noguera, F., & Aude, I. (2011). Compuestos derivados de soja presentes en alimentos industrializados, estudio del mercado montevideano. *Alimentación Latinoamericana*, 295, 11. <https://www.colibri.udelar.edu.uy/jspui/handle/20.500.12008/9472>
- Avilez, J. P., Escobar, P., von Fabeck, G., Villagran, K., García, F., Matamoros, R., & García Martínez, A. (2010). Caracterización productiva de explotaciones lecheras empleando metodología de análisis multivariado. *Revista Científica*, XX(1), 74-80. <http://www.redalyc.org/articulo.oa?id=95915974011>
- Barcenas Ochoa, M. G., García Santaolalla, M. D. L., Montes de Oca Serrano, M. J., Paredes Ruiz, F., & Úrquiza Sánchez, R. (2017). *Análogos de Queso* [Presentación digital]. México: Universidad Nacional Autónoma de México. <https://docplayer.es/34213798-Productos-lacteos-analogosde-queso.html>
- Barrera, Y., García, E., & Solis, K. (2020). Análisis de procesos en la cuba quesera empleando el método de los mínimos cuadrados parciales. *Revista Centro Azúcar*, 47(2), 1-10. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2223-48612020000200001
- Burbano-Vallejo, E. L., & Moreno, E. (2018). Análisis de conglomerados del norte del valle del cauca. Caso de estudio: Cartago, zarzal y la unión. *Ingeniería Industrial*, XXXIV(1), 78-91. <http://scielo.sld.cu/pdf/rii/v39n1/rii09118.pdf>
- Cerretani, L., Maggio, R. M., Barnaba, C., Toschi, T. G., & Chiavaro, E. (2011). Application of partial least square regression to differential scanning calorimetry data for fatty acid quantitation in olive oil. *Food Chemistry*, 127(4), 1899-1904. <https://doi.org/10.1016/j.foodchem.2011.02.041>
- Chico, M., Romero, B., Hernández, E., & García, A. (2017). Elaboración de yogurt a base de soya enriquecido con quinoa y camote. *Investigación y Desarrollo en Ciencia y Tecnología de Alimentos*, 2(7), 410-416. <http://www.fcb.uanl.mx/IDCyTA/files/volume2/3/8/68.pdf>
- de Luna Jiménez, A. (2018). Elaboración de quesos nutritivos y económicos mediante el mezclado de leche de soya y de vaca. *Centro Agropecuario*, 1. <https://investigacion.uaa.mx/RevistaIyC/archivo/revista7/Articulo%201.pdf>
- García-Noa, E., Cuellar-Pérez, O., Sevilla-Videaux, F., Comas-Castillo, L., & Sariego-Toledo, Y. (2015). Evaluación de los riesgos y la criticidad de los fallos en la tecnología del yogurt de

- soya. *Revista Ciencia y Tecnología de Alimentos*, 25(1), 62-68. <https://www.revcitecal.iiia.edu.cu/revista/index.php/RCTA/article/view/311>
- Hernández, G., Álvarez, M., Nuñez de Villavicencio, M., & Rosas, B. (2015). Empleo de la premezcla fortificada de maíz y soya en la elaboración de pan. *Ciencia y Tecnología de Alimentos*, 25(5), 41-45. <https://www.revcitecal.iiia.edu.cu/revista/index.php/RCTA/article/view/305>
- Hernández Beltrán, Y., Valdés Paneca, G. R., Ledesma Plasencia, M. L., & García Noa, E. (2012). Procedimientos para el mejoramiento continuo del queso crema Nela. *Revista Tecnología Química*, XXXII (3), 313-332. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2224-61852012000300010
- Juárez-Barrientos, J. M., Díaz-Rivera, P., Rodríguez-Miranda, J., Martínez-Sánchez, C. E., Hernández-Santos, B., Ramírez-Rivera, E., Torruco-Uco, J. G., & Herman-Lara, E. (2016). Caracterización de la leche y clasificación de calidad mediante análisis Clúster en sistemas de doble propósito. *Revista Mexicana de Ciencias Pecuarias*, 7(4), 525-537. <https://doi.org/10.22319/rmcp.v7i4.4280>
- Lima de Aguiar, C., Teixeira dos Santos, S., Moraes Freixo, J., & Hiromu Okuda, M. (2009). Multivariate statistical method in optimization of protein extraction from soybean flour with dairy whey. *Bioscience Journal*, 25(2), 171-180. https://repositorio.usp.br/bitstream/handle/BDPI/18868/art_AGUIAR_MULTIVARIATE_STATISTICAL_METHOD_IN_OPTIMIZATION_OF_PROTEIN_2009.pdf?sequence=1
- Mateos-Aparicio, I., Redondo Cuenca, A., Villanueva-Suárez, M. J., & Zapata-Revilla, M. A. (2008). Soybean, a promising health source. *Nutrición Hospitalaria*, 23(4), 305-312. http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S0212-16112008000500001
- Montalvo-Perdomo, Y., Panadés-Ambrosio, G., Sardiñas-Reynaldo, L., Guevara-Albo, Y., & Nuñez de Villavicencio, M. (2016). Calidad Sensorial de néctar de mango enriquecido con soya. *Ciencia y Tecnología de Alimentos*, 26(3), 5. <https://www.revcitecal.iiia.edu.cu/revista/index.php/RCTA/article/view/178>
- Patras, A., Brunton, N. P., Downey, G., Rawson, A., Warriner, K., & Gernigon, G. (2011). Application of principal component and hierarchical cluster analysis to classify fruits and vegetables commonly consumed in Ireland based on in vitro antioxidant activity. *Journal of Food Composition and Analysis*, 24(2), 250-256. <https://doi.org/10.1016/j.jfca.2010.09.012>
- Pérez-Sánchez, A., Crespo-Zafra, L., Ramos-López, L., García-Noa, E., & Matos-Mosqueda, L. (2020). Estudio técnico-económico de la producción de leche de cabra en polvo. *Revista EIA*, 17(33), 1-13. <https://doi.org/10.24050/reia.v17i33.1330>
- Piovoso, M. J., & Hoo, K. A. (2002). Multivariate Statistics for Process Control. *IEEE Control Systems*, 22(5), 8. <https://doi.org/10.1109/MCS.2002.1035213>
- Ramírez-Anormaliza, R., Guevara-viejo, F., D'Armas Regnault, M., Pena-Holguín, R., Farias-Lema, R., Bravo-Duarte, F., Diaz-Montenegro, J., Calderon-Cisneros J., Franco-Arias, O., Ramírez-Granda, F., Carrasquero Rodriguez, E., Vargas-Decimavilla, D., Basurto-Quilligana, R., Vargas-Ortiz, O., Bermeo-Paucar, J., & Castelo-Gonzalez, J. (2017). *Análisis Multivariante: Teoría y práctica de las principales técnicas*. Ecuador: Ediciones Holguín S.A. y Universidad Estatal de Milagro. <http://repositorio.unemi.edu.ec/handle/123456789/3855>

- Resolución N.º 139/08. *Norma técnica NRLAL 174 sobre yogur de soya*. Cuba; 2008. <http://faolex.fao.org/docs/pdf/cub82669.pdf>
- Rodríguez, O., Cardoso-Castañeda, F., González-Ríos, J., Íñiguez, C., & Nuñez de Villavicencio, M. (2017a). Elaboración de leche fermentada con probióticos a partir de arroz y leches de búfala y soya. *Ciencia y Tecnología De Alimentos*, 27(1), 44-49. <https://www.revcitecal.iiaa.edu.cu/revista/index.php/RCTA/article/view/160>
- Rodríguez, O., Cardoso-Castañeda, F., González-Ríos, J., Íñiguez, C., & Nuñez de Villavicencio, M. (2017b). Desarrollo de un queso crema probiótico con leches de soya y búfala. *Ciencia y Tecnología de Alimentos*, 27(3), 1-6. <https://www.revcitecal.iiaa.edu.cu/revista/index.php/RCTA/article/view/109>
- Ruiz-Tamayo, J., Vázquez-Lopez, J. A., Ruelas-Santoyo, E. A., Hernández-Lopez, A., López-Juarez, I., & Ríos-Lira, A. J. (2021). Multivariate Pattern Recognition in MSPC Using Bayesian Inference. *Mathematics*, 9(4), 301-306. <https://doi.org/10.3390/math9040306>
- Sariego Toledo, Y., Perez Ybanez, S., Garcia Noa, E., & Rodríguez, L. (2017). Evaluación energética de una planta de helados. *Ingeniería Energética*, 38(1), 42-53. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S181559012017000100006
- Statgraphics. (2016). *Centurion XVII Manual del Usuario*. Estados Unidos: StatPoint Technologys Inc. <https://www.statgraphics.com>
- Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of cluster in a data set via the gap statistic. *Journal of the Royal Statistic Society Series B: Statistical Methodology*, 63(2), 411-413. <https://doi.org/10.1111/1467-9868.00293>
- Valencia García, F. E., Millán Cardona, L. d. J., Restrepo Morales, C. A., & Jaramillo Garcés, Y. (2007). Efecto de sustitutos de grasa en propiedades sensoriales y texturales del queso crema. *Revista Lasallista de Investigación*, 4(1), 20-26. <https://www.redalyc.org/articulo.oa?id=69540104>
- Vega León, M., Castillo Toro, D., Cañet Prades, F. M., Molina, R., & Obregón, V. (2017). *Cinética del secado de semillas de dos variedades de soya (Glycine max (L) Merrill)*. Cuba: Instituto de Investigaciones Fundamentales en Agricultura Tropical "Alejandro de Humboldt". <https://docplayer.es/70344601-Cinetica-de-secado-de-semillas-de-dos-variedades-de-soya-glycine-max-l-merrill-humboldt-inifat.html>
- Vives, Y., Rodríguez, J. L., & Lorenzo, Y. (2017). Efecto antioxidante de extractos de plantas aromáticas en la estabilidad del aderezo de soya. *Ciencia y Tecnología de Alimentos*, 27(1), 1-6. <https://www.revcitecal.iiaa.edu.cu/revista/index.php/RCTA/article/view/152>
- Zuluaga Dominguez, C. M. (2011). Análisis Estadístico Multivariado: una herramienta estratégica para el control de procesos y calidad en la Industria Agroalimentaria. *Publicaciones e Investigaciones*, 5(1), 143-157. <https://doi.org/10.22490/25394088.587>