

CLASIFICACIÓN DE BEBIDAS DEPORTIVAS Y BEBIDAS ENERGÉTICAS MEDIANTE QUIMIOMETRÍA

Classification of sports drinks and energy drinks by chemometrics

 ¹ Robert Cazar

 ¹ Mabel Parada*

 ² Nathaly Maza

¹ Escuela Superior Politécnica de Chimborazo, Facultad de Ciencias, Riobamba, Ecuador.

² Investigador independiente, Riobamba, Ecuador.

* mabel.parada@esPOCH.edu.ec

RESUMEN

Este trabajo tiene como propósito desarrollar un procedimiento confiable y exacto para discriminar entre las bebidas deportivas y energéticas. Ya que estos dos tipos de bebidas difieren en sus propiedades químicas, estas propiedades pueden ser aprovechadas para desarrollar un modelo de clasificación que separe con exactitud ambos grupos. El modelo de clasificación ha sido construido utilizando técnicas de quimiometría tales como el Análisis de Componentes Principales y el Análisis de Agrupamientos, las cuales son sumamente apropiadas para este objetivo. Un conjunto de 11 bebidas (7 bebidas energéticas y 4 bebidas deportivas) fue caracterizado mediante 8 propiedades químicas y el conjunto de datos resultante fue analizado con las técnicas quimiométricas ya mencionadas. Tales análisis proporcionaron un método muy exacto para discriminar entre los dos tipos de bebidas. Se concluye que la quimiometría constituye una herramienta muy eficiente para el tratamiento de datos químicos multivariados, particularmente para propósitos de modelización y clasificación.

Palabras claves: *Bebidas deportivas, bebidas energéticas, quimiometría.*

ABSTRACT

The goal of this work is to create a reliable and accurate method for distinguishing between sports and energy drinks. Since these two types of beverages differ in their chemical properties, these characteristics can be used to develop a classification model that accurately separates those groups. The classification model has been built utilizing chemometric techniques such as Principal Component Analysis and Cluster Analysis that are highly appropriate for this objective. A set of 11 beverages (7 energy drinks and 4 sports drinks) was characterized by 8 chemical properties, and the resulting data set was analyzed with the chemometric techniques mentioned above. Such analyses provided a very accurate method to distinguish between the two types of beverages. It is concluded that chemometrics is a very efficient tool for the treatment of multivariate chemical data, particularly for modeling and classification purposes.

Keywords: *Sports drinks, energy drinks, chemometrics.*

I. INTRODUCCIÓN

Las bebidas deportivas y energéticas se han vuelto omnipresentes en la sociedad moderna (1-4). Ellas se consumen profusamente en todo el mundo (5) y constituyen un mercado de miles de millones de dólares (6-8). Estos dos tipos de bebidas tienen propósitos diferentes y por tanto poseen composiciones químicas diferentes (9-11). En consecuencia, un químico debería ser capaz de discriminarlas en base a sus propiedades químicas. Este estudio tiene como objetivo ilustrar tal tarea. Para este propósito, once marcas de bebidas disponibles comercialmente, siete correspondientes a bebidas energéticas y cuatro a bebidas deportivas, han sido seleccionadas y ocho propiedades químicas han sido medidas sobre cada una de ellas produciendo un conjunto de datos multivariado. En la actualidad, este tipo de datos es común en la química experimental. La quimiometría provee la estrategia más adecuada para la interpretación de conjuntos de datos multivariados ya que posee una colección de métodos matemáticos y estadísticos sumamente potentes diseñados para extraer información relevante de datos complejos (12-14). El análisis de agrupamientos y el análisis de componentes principales (15-18) están entre los métodos más útiles de esta disciplina. El análisis de agrupamientos identifica grupos de muestras con comportamiento similar y el análisis de componentes principales reduce la dimensionalidad de los datos generando nuevas variables altamente informativas, sobre las cuales es posible proyectar tanto las muestras como las variables para inspeccionar la estructura de los datos (19-21). Estas técnicas han sido aplicadas a los datos descritos anteriormente y se ha logrado obtener un modelo que discrimina las bebidas deportivas de las energéticas con gran exactitud (22).

II. MATERIALES Y MÉTODOS

Los datos usados en este análisis provienen de 11 muestras (las bebidas) sobre las que se han medido 8 variables (Carbohidratos, Energía, Azúcar, Sodio, pH, Grados Brix, Densidad y Ácido Cítrico). Las muestras 1 – 7 corresponden a bebidas energéticas y las muestras 8 – 11 son bebidas deportivas. Los valores de carbohidratos, energía, azúcar y sodio fueron recuperados directamente de las tablas de valor nutricional

de los productos. Las variables restantes fueron obtenidas experimentalmente como sigue:

Los valores de pH fueron determinados empleando un pH-metro previamente calibrado con soluciones buffer de pH 4, 7 y 10, respectivamente. Se colocó un volumen adecuado de bebida en un vaso de precipitación, se verificó que su temperatura se encuentre en el rango de 20 a 25 °C, se introdujo el electrodo del instrumento y se registró el pH.

Los valores de Grados Brix fueron obtenidos empleando un refractómetro previamente calibrado con agua destilada. Se colocó una gota de muestra en el prisma del refractómetro cuidando que lo cubra completamente, se cerró la tapa y se procedió a registrar la lectura.

Los valores de densidad fueron recabados como se describe a continuación. Primero, se lavó un picnómetro con agua destilada, se escurrió, y secó en la estufa. Se pesó el picnómetro limpio y seco en una balanza previamente encerada y se registró tal dato. Se colocó agua destilada en el picnómetro hasta alcanzar cierta marca y se pesó nuevamente en la balanza y se registró el resultado. Se vació el picnómetro y se lo secó en la estufa. Finalmente, se colocó la muestra en el picnómetro hasta alcanzar la marca establecida, se pesó en la balanza y se anotó el resultado. La densidad se obtuvo aplicando la siguiente ecuación:

$$\rho = \frac{m(p+s)-m(p)}{m(p+ag)-m(p)} \times \rho(ag) \quad (1)$$

Donde:

ρ = densidad

$\rho(ag)$ = densidad del agua (0.998)

$m(p)$ = masa del picnómetro vacío, en gramos

$m(p + ag)$ = masa del picnómetro con agua destilada

$m(p + s)$ = masa del picnómetro con la muestra

Los valores de ácido cítrico fueron obtenidos mediante cromatografía líquida de alto rendimiento, HPLC, aplicando el procedimiento descrito en el trabajo de Iní y sus colaboradores (23).

La Tabla I recolecta los datos usados en este trabajo.

Muestra	Marca	Carbohidratos	Energía	Azúcar	Sodio	pH	Grados Brix	Densidad	Ácido Cítrico
1	Volt	26	110	17.8	65	3.45	12.00	1.04	6777.22
2	Vive 100	20	90	16.0	70	2.92	9.14	1.03	8843.99
3	V220	13	50	13.0	21	2.75	13.29	1.04	7192.65
4	Amper	19	80	14.0	135	3.16	8.36	1.03	9361.20
5	Monster	30	120	27.0	190	3.44	12.61	1.04	5149.79
6	Speed	28	112	28.0	20	3.04	12.09	1.04	4878.84
7	Red Bull	11	45	11.0	40	3.48	11.65	1.04	9580.04
8	Powerade	19	80	19.0	145	2.95	6.08	1.02	2751.50
9	Gatorade	0	0	0.0	115	2.93	0.48	1.00	3870.19
10	Sporade	15	70	15.0	100	3.36	6.49	1.02	3883.72
11	Profit	10	40	10.0	150	3.60	4.56	1.01	3447.32

Tabla 1. Datos usados en el análisis.

Las técnicas quimiométricas descritas en la sección anterior han sido aplicadas en este conjunto de datos con la finalidad de obtener un procedimiento confiable y exacto para clasificar los dos grupos de bebidas representados en los datos.

III. RESULTADOS

Una rápida inspección de los datos evidencia la presencia de efectos de escala debido a las diversas unidades de medida en la que las variables han sido registradas. Tales efectos de

escala obscurecen la información útil presente en los datos reduciendo el poder discriminante de las variables. En consecuencia, un pretratamiento de los datos es indispensable. Una operación denominada autoscaling fue aplicada a los datos para eliminar este problema (24,25). El autoscaling consiste en restar a cada dato la media de su columna y dividir este valor para la desviación standard de tal columna. A continuación se calculó la matriz de coeficientes de correlación de los datos transformados mediante autoscaling. La Tabla 2 muestra esta matriz, los valores que aparecen en ella corresponden a los coeficientes de correlación de cada pareja de variables.

	Carbohidratos	Energía	Azúcar	Sodio	pH	Grados Brix	Densidad	Ácido Cítrico
Carbohidratos	1.00							
Energía	0.99	1.00						
Azúcar	0.95	0.93	1.00					
Sodio	0.03	0.03	0.001	1.00				
pH	0.15	0.15	0.10	0.34	1.00			
Grados Brix	0.70	0.67	0.68	-0.45	0.06	1.00		
Densidad	0.73	0.70	0.71	-0.44	0.11	1.00	1.00	
Ácido Cítrico	0.11	0.12	-0.04	-0.42	-0.04	0.52	0.52	1.00

Tabla 2. Matriz de coeficientes de correlación de las variables.

Examinando la Tabla 2 se observa que la pareja Densidad/Grados Brix manifiesta un grado de asociación perfecta (coeficiente de correlación igual a 1) esto significa que ambas variables contribuyen con la misma información, y por tanto son redundantes. En tales casos es seguro desechar una de las dos. Se decidió excluir los Grados Brix y conservar la Densidad. Otras asociaciones fuertes corresponden a Energía/Carbohidratos (coeficiente de correlación igual a 0.99) y Azúcar/Carbohidratos (coeficiente de correlación igual a

0.95) de estos valores se deduce que estas tres variables están fuertemente relacionadas y por tanto contribuyen con información muy similar. En este caso, se decidió conservar solamente la variable Azúcar y desechar las otras dos. Entonces, el análisis de la matriz de coeficientes de correlación de las variables permitió reducir su número de 8 a 5 solamente. Las variables que se han conservado son, a saber: Azúcar, Sodio, Densidad, pH y Ácido Cítrico. Tales variables exhiben bajas correlaciones y por consiguiente

garantizan que la información redundante ha sido removida del conjunto de datos.

A continuación, los datos fueron sometidos a un análisis de agrupamientos (clustering) para determinar si las variables seleccionadas son capaces de separar los dos tipos de bebidas presentes en ellos. Una técnica de clustering jerárquico fue empleada, la cual identifica agrupamientos de muestras con comportamiento similar en base a la distancia entre las muestras en el espacio multidimensional generado por las variables. La Figura 1 presenta el dendrograma obtenido al aplicar la técnica de clustering jerárquico de “average linkage” sobre los datos.

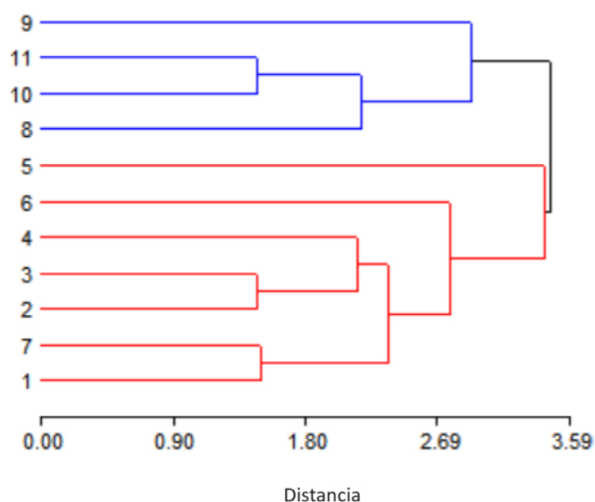


Figura 1. Dendrograma obtenido con la técnica de clustering jerárquico “average linkage”

El dendrograma, como se puede ver, es una representación gráfica tipo árbol de los grupos que gradualmente se van formando en la etapa de Clustering. Éste nos permite identificar la naturaleza de los grupos formados a un cierto nivel de similitud. Por ejemplo, a un nivel de similitud del 25% las muestras se distribuyen en dos grupos bien separados. El grupo en color azul contiene las muestras 8 – 11 que corresponden a las bebidas deportivas y el grupo en rojo contiene las muestras 1 – 7 que corresponden a las bebidas energéticas. De los resultados del Análisis de Agrupamientos se evidencia que es factible discriminar completamente los dos tipos de bebidas utilizados en este estudio.

Finalmente, un análisis de componentes principales ha sido ejecutado sobre los datos. Este análisis sirve para corroborar los resultados obtenidos del clustering e inspeccionar

visualmente la estructura de los datos. Este método permite proyectar las muestras y las variables sobre unas nuevas variables llamadas componentes principales (PC) las cuales son combinaciones lineales de las variable originales. Así, cada variable original contribuye en cierta proporción a cada componente principal. Tales contribuciones se denominan “loadings” y miden la importancia de cada variable sobre una determinada componente principal. Ya que las componentes principales son los autovectores de la matriz de correlación de los datos, ellas son ortogonales entre sí, lo cual significa que no están correlacionadas unas a otras por lo que cada una contiene una información propia y única respecto a las propiedades de los datos. La Tabla 3 presenta la información retenida por cada componente principal (PC).

PC	Autovalor	Información (%)	Información Acumulada (%)
1	2.10	42	42
2	1.44	29	71
3	0.93	19	89
4	0.47	9	99
5	0.06	1	100

Tabla 3. Información retenida por cada una de las Componentes Principales.

Esta tabla revela que la primera PC explica el 42% de la información contenida en los datos, en tanto que la segunda PC explica un 29% de la información residual. En conjunto, PC1 y PC2 retienen el 71% de la información total, lo que resta se encuentra repartido en las otras componentes.

La Tabla 4 presenta los loadings de las dos primeras componentes principales.

Variable	PC1	PC2
Azúcar	0.39	0.55
Sodio	-0.45	0.47
pH	-0.06	0.58
Densidad	0.46	-0.27
Ácido Cítrico	0.65	0.24

Tabla 4. Loadings de las variables originales sobre las dos primeras componentes principales, las contribuciones más altas aparecen en negrillas

Se observa que las variables Ácido Cítrico, Densidad y Azúcar son las que más contribuyen a PC1; igualmente las variables pH, Azúcar y Sodio son las que más contribuyen a PC2.

La Figura 2 presenta la proyección de la muestras

en el espacio bidimensional generado por PC1 y PC2. Este gráfico contiene el 71% de la información proporcionada por los datos.

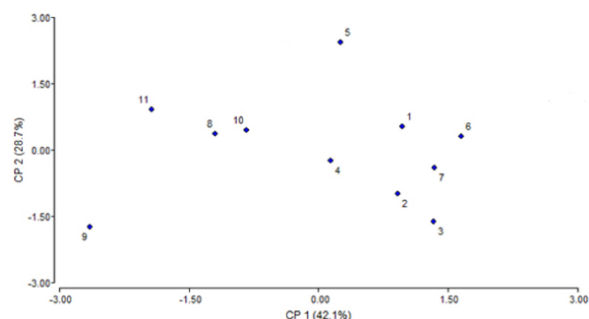


Figura 2. Proyección de las muestras sobre PC1 y PC2. El 71% de la información total es retenida en este gráfico

En la proyección se visualiza que las dos categorías declaradas en el estudio forman grupos separados. En el lado izquierdo del gráfico aparecen las bebidas deportivas (muestras 8 – 11) formando un grupo compacto; en tanto que en el lado derecho del gráfico aparecen las bebidas energéticas (muestras 1 – 7) formando un grupo más grande y disperso. Se observa también que PC1 es la componente discriminante ya que al proyectar las muestras sobre ella las dos categorías resultan bien separadas. Este hecho permite caracterizar a las dos categorías de bebidas como sigue: Las bebidas deportivas se caracterizan por poseer valores bajos a moderados de Ácido Cítrico, Densidad y Azúcar, en tanto que las bebidas energéticas se distinguen por poseer valores altos de Ácido Cítrico, Densidad y Azúcar.

La Figura 3 muestra la proyección de las variables sobre PC1 y PC2.

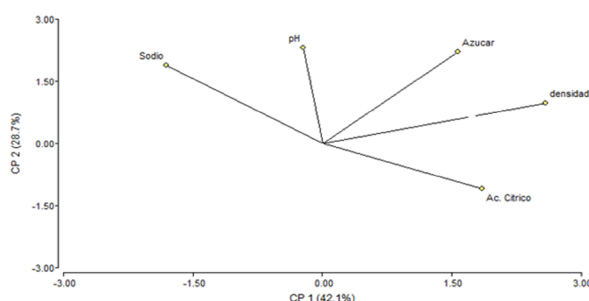


Figura 3. Proyección de las variable sobre PC1 y PC2

Se observa que las 5 variables seleccionadas no manifiestan correlaciones importantes y esto agregado al hecho que ellas contienen la suficiente información para discriminar las 2 categorías bajo

estudio permite establecer que dichas variables son las más apropiadas para este análisis.

IV. DISCUSIÓN

La aplicación de técnicas quimiométricas sobre los datos de este estudio, ha producido resultados que proveen un procedimiento capaz de discriminar de manera eficiente y exacta las bebidas deportivas de las bebidas energéticas a partir de sus propiedades químicas. La operación inicial de autoscaling de los datos permitió retirar los efectos de escala y trabajar con una matriz de datos donde todas las variables originales tienen la misma importancia inicial. El cálculo de la matriz de correlación asistió en la selección de las variables más apropiadas para este estudio. De la inspección de los índices de correlación entre las diferentes parejas de variables se detectó que algunas de ellas contienen información redundante por lo que fueron descartadas y el número de variables se redujo de 8 a 5 conservando las siguientes: Azúcar, Sodio, Densidad, pH y Ácido cítrico. La operación de clustering jerárquico sobre la matriz de datos reducida produjo un dendograma en el cual los dos grupos bajo estudio se separan perfectamente. El análisis de componentes principales ha corroborado tal hallazgo puesto que la proyección de las muestras sobre las dos primeras componentes principales evidencia que las bebidas deportivas y las bebidas energéticas forman dos agrupamientos bien discriminados. Los valores de los loadings para las variables originales sobre las dos primeras Componentes Principales han permitido caracterizar químicamente las dos categorías consideradas en el estudio, a saber, las bebidas deportivas poseen valores bajos a moderados de Ácido Cítrico, Densidad y Azúcar, en tanto que las bebidas energéticas poseen valores altos de Ácido Cítrico, Densidad y Azúcar.

V. CONCLUSIONES

La aplicación de técnicas quimiométricas sobre un conjunto de datos obtenido al medir varias propiedades químicas sobre un número de bebidas deportivas y energéticas ha permitido obtener un procedimiento eficiente y exacto para discriminar entre estas dos categorías de bebidas. El estudio ilustra la utilidad de la quimiometría para extraer información relevante de conjuntos de datos multivariados.

VI. REFERENCIAS

1. Higgins JP, Babu K, Deuster PA, Shearer J. Energy drinks: a contemporary issues paper. *Current Sports Med Rep*. 2018;17(2):65–72.
2. Sutehall S, Muniz-Pardos B, Bosch AN, Di Gianfrancesco A, Pitsiladis YP. Sports drinks on the edge of a new era. *Current Sports Med Rep*. 2018;17(4):112–116.
3. Nadeem IM, Shanmugaraj A, Sakha S, Horner NS, Ayeni OR, Khan M. Energy drinks and their adverse health effects: a systematic review and meta-analysis. *Sports Health*. 2021;13(3):265–277.
4. Bleakley A, Ellithorpe ME, Jordan AB, Hennessy M, Stevens R. A content analysis of sports and energy drink advertising. *Appetite*. 2022;174:106010.
5. Simulescu V, Ilia G, Macarie L, Merghes P. Sport and energy drinks consumption before, during and after training. *Science & Sports*. 2019;34(1):3–9.
6. Yang H, Senarath D, Oral C, Ramkumar J. The growing market for energy and sports drinks in the United States: Can chocolate milk remain a contender? *J Food Distrib Res*. 2020;51(2):131–157.
7. Peker Z, Köse DA. Content analysis of locally marketed energy drinks: Turkish market. *Hacettepe J Biol Chem*. 2023;51(4):341–349.
8. Pfender E, Bleakley A, Ellithorpe M, Hennessy M, Maloney E, Jordan A, Stevens R. Perceptions of sports and energy drinks: factors associated with adolescent beliefs. *Am J Health Promot*. 2023;37(1):84–88.
9. Hernández-León A, Artavia G, Cortés-Herrera C, Granados-Chinchilla F. Chromatographic determination of major physiologically active components in energy drinks and sports aids commercialized in Costa Rica. *J Food Nutr Res*. 2022;61(2).
10. García MA, Ruiz Y, Rodríguez JE, Cruz-Socorro A, Casariego A. Isotonic sports drink prepared from pineapple juice: stability during its accelerated storage. *Agroind Sci*. 2023;13(3):135–142.
11. Martins EC, Santana ER, Spinelli A. Nitrogen and sulfur co-doped graphene quantum dot-modified electrode for monitoring of multivitamins in energy drinks. *Talanta*. 2023;252:123836.
12. González-Domínguez R, Sayago A, Fernández-Recamales Á. An overview on the application of chemometrics tools in food authenticity and traceability. *Foods*. 2022;11(23):3940.
13. Wang HP, Chen P, Dai JW, Liu D, Li JY, Xu YP, Chu XL. Recent advances of chemometric calibration methods in modern spectroscopy: algorithms, strategy, and related issues. *TrAC Trends Anal Chem*. 2022;153:116648.
14. Rodionova OY, Oliveri P, Malegori C, Pomerantsev AL. Chemometrics as an efficient tool for food authentication: golden pillars for building reliable models. *Trends Food Sci Technol*. 2024;104429.
15. Cohen-Addad V, Kanade V, Mallmann-Trenn F, Mathieu C. Hierarchical clustering: objective functions and algorithms. *J ACM*. 2019;66(4):1–42.
16. Ezugwu AE, Ikotun AM, Oyelade OO, Abualigah L, Agushaka JO, Eke CI, Akinyelu AA. A comprehensive survey of clustering algorithms: state-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. *Eng Appl Artif Intell*. 2022;110:104743.
17. Schreiber JB. Issues and recommendations for exploratory factor analysis and principal component analysis. *Res Social Adm Pharm*. 2021;17(5):1004–1011.
18. Hasan BMS, Abdulazeez AM. A review of principal component analysis algorithm for dimensionality reduction. *J Soft Comput Data Min*. 2021;2(1):20–30.

19. Gewers FL, Ferreira GR, Arruda HFD, Silva FN, Comin CH, Amancio DR, Costa LDF. Principal component analysis: a natural approach to data exploration. *ACM Comput Surv.* 2021;54(4):1–34.
20. Oyewole GJ, Thopil GA. Data clustering: application and trends. *Artif Intell Rev.* 2023;56(7):6439–6475.
21. Ullmann T, Hennig C, Boulesteix AL. Validation of cluster analysis results on validation data: a systematic framework. *Wiley Interdiscip Rev Data Min Knowl Discov.* 2022;12(3):e1444.
22. Sorochan Armstrong MD, De la Mata AP, Harynuk JJ. Review of variable selection methods for discriminant-type problems in chemometrics. *Front Anal Sci.* 2022;2:867938.
23. Inić S, Ljepović M, Domijan A, Srdarević S, Jablan J, Markov K. HPLC analysis of citric and tartaric acids in fruit nectars and juices. *Croat Chem Acta.* 2020;93(1):57–62.
24. Rustam. Pretreatment methods for enhancing machine learning performance on metabolomics data. *IEEE Access.* 2025;[Epub ahead of print].
25. Sun J, Xia Y. Pretreating and normalizing metabolomics data for statistical analysis. *Genes Dis.* 2024;11(3):100979.