

APLICACIÓN DE TÉCNICAS DE *CLUSTERING* PARA LA ELABORACIÓN DE PERFILES EPIDEMIOLÓGICOS EN ESTUDIOS SANITARIOS

Fernando Massa*
Ramón Álvarez-Vaz**

Resumen. En el contexto de los estudios epidemiológicos, donde se indaga por las enfermedades no transmisibles (ENT), es práctica habitual utilizar variables categóricas para reflejar tanto la presencia de determinadas enfermedades como de comorbilidades o factores de riesgo (FR). En particular, dentro del ámbito de las patologías orales (PO), las ENT también suelen ser consideradas como FR. En este sentido, existe la posibilidad de considerar el estado de salud global de los individuos, atendiendo simultáneamente la presencia de ENT, PO y FR. Entre los años 2015 y 2016 se llevó a cabo un estudio concerniente a las personas que demandaron asistencia en la Facultad de Odontología de la Universidad de la República, Uruguay. Dicho estudio tenía como objetivo investigar el vínculo entre los FR de las ENT más prevalentes y la salud bucal. Se propone la obtención de perfiles epidemiológicos bien diferenciados a partir de la determinación de tipologías de individuos, en base a atributos categóricos, referentes a ENT, las PO más frecuentes y los FR comunes a ambas. Se detectaron dos tipologías definidas, principalmente por las prevalencias de ENT y PO. Una de ellas se caracterizó por la coexistencia de PO, ENT y FR mientras que en la otra se observaron participantes con menores prevalencias de ENT. Adicionalmente se constató que el grupo de mayor carga de enfermedad estuvo compuesto por principalmente por mujeres y personas de edades más avanzadas, lo cual podría explicar la mayor prevalencia de ENT.

Palabras Clave: Clustering; Enfermedades no transmisibles; Factores de riesgo.

* Universidad de la República (UdelaR), Uruguay.

Contacto: fmassa@iesta.edu.uy

** Universidad de la República (UdelaR), Uruguay.

Contacto: ramon@iesta.edu.uy

APPLICATION OF CLUSTERING PROCEDURES FOR THE ELABORATION OF EPIDEMIOLOGICAL PROFILES IN SANITARY STUDIES

Abstract. In the context of epidemiological studies, where non-communicable diseases (NCDs) are investigated, it is common practice the use of categorical variables to reflect the presence of both certain diseases and comorbidities or risk factors (RF). Particularly, within the scope of oral pathologies (PO), NCDs are usually considered as RF. In this sense, it is possible to consider the global health status of individuals while simultaneously taking into consideration the presence of NCDs, PO and RF. During years 2015 and 2016, a study was carried out concerning people who demanded assistance at the Facultad de Odontología of the Universidad de la República, Uruguay, which aimed to investigate the link between the RF of the most prevalent NCDs, and oral health. It is proposed to obtain well differentiated epidemiological profiles from the determination of typologies of individuals, based on categorical attributes, referring to NCDs, the most frequent OPs and the RF, common to both. Two typologies were detected mainly defined by the prevalence of NCD and PO. One of them was characterized by the coexistence of PO, NCD and RF, while in the other participants with lower prevalence of NCD were observed. Additionally, it was found that the group with the highest burden of disease was mainly composed of women and people of more advanced ages, which could explain the higher prevalence of NCDs.

Keywords: Clustering; Non-communicable diseases; Risk factors.

Original recibido el 03/05/2018

Aceptado para su publicación el 30/11/2019

1. Introducción

Las enfermedades no transmisibles (ENT), entre las que pueden agruparse la diabetes, el cáncer, enfermedades cardiovasculares y respiratorias crónicas, son las causas de mortalidad a nivel mundial más importantes, representando el 63% de las muertes globales. Específicamente en países con bajos y medianos ingresos, este tipo de enfermedades son la causa de casi el 86% de las muertes prematuras (WHO, 2013). Este conjunto de enfermedades es responsable de un gran aumento de la discapacidad en varios países del mundo, en particular en países subdesarrollados o en vías de desarrollo, puesto que en ellos el diagnóstico suele producirse en edades más tempranas, acarreando consigo períodos más prolongados de discapacidad.

Conductas relacionadas con una alimentación inadecuada, el sedentarismo, el consumo nocivo de alcohol, así como el de tabaco, juegan un papel determinante en el desarrollo del conjunto de ENT, dado que estos propician la aparición de factores de riesgo (FR) tales como la obesidad, los trastornos del metabolismo de los hidratos de carbono, la hipertensión arterial (HTA) o las dislipemias, entre otros (Skapino y Álvarez-Vaz, 2016). La importancia de considerarlos radica en que según se expone en Asamblea General de las Naciones Unidas, (2011) también son FR para enfermedades bucodentales como la caries, paradenciopatías, distintos tipos de cáncer y otras patologías orales (PO). Las enfermedades bucodentales implican un importante gasto económico tanto para los sistemas de salud de varios países como para las personas que los padecen, pero aún más relevante es que estos las afectan durante toda la vida, causando malestar y dolor, desfiguración e incluso pudiendo derivar en la muerte del paciente.

En el contexto de los estudios epidemiológicos, donde se indaga sobre las ENT, es práctica habitual trabajar con variables categóricas, empleándolas para reflejar la presencia tanto de determinadas enfermedades como de comorbilidades o FR. En particular, en el ámbito de las patologías orales, las ENT también suelen ser consideradas como FR (Genco y Genco, 2014) (Bose, 2019) (Petersen, 2003). Considerando este enfoque, existe la posibilidad de considerar la situación de salud global de los individuos, atendiendo simultáneamente la presencia de ENT, PO y FR. En el año 2011 se llevó a cabo en Uruguay el Primer Relevamiento Nacional de Salud Bucal en población joven y adulta uruguaya (Lorenzo, Álvarez-Vaz, Blanco y Peres, 2013). Dicho estudio permitió conocer la prevalencia a nivel nacional de las principales PO así como de algunos FR, incluidas las ENT (Lorenzo *et al.*, 2015; Álvarez *et al.*, 2015; Nicola *et al.*, 2016). La estrategia de análisis empleada consistió en determinar cuáles eran los factores que repercutían sobre cada una de las PO, pero en ningún momento se consideró la salud de los participantes de manera global.

Una posible estrategia de análisis que posibilita el análisis simultáneo de un conjunto de variables es el análisis de *clusters*. En la literatura asociada a las ENT, existen diversas aplicaciones del uso de esta técnica. En Haregu *et al.* (2018) se detectaron cinco tipologías, conformadas principalmente por la obesidad, el consumo de grasas y la hipertensión. Por otro lado, en Nascimento *et al.* (2019) se identificaron tres grupos, basándose en ENT tales como cáncer, enfermedades cardiovasculares y respiratorias. Teniendo en cuenta tanto a las ENT como a sus FR, el estudio de

Zancheta, Machado, Fónias y Bertazzi (2019) llevado a cabo sobre una cohorte de adolescentes, identificó que los FR más relevantes a la hora de determinar los grupos fueron el consumo de alcohol, el uso de tabaco, la actividad física insuficiente y un consumo no saludable de alimentos. En cuanto a la inclusión de las PO, en el trabajo de Austregésilo, Angeiras, de Sena y Costa (2019) se identifican dos tipologías diferenciadas por la presencia de FR y malos hábitos de prevención de aparición de PO. No obstante, no se encontraron estudios donde se atacara la problemática desde las tres perspectivas simultáneamente.

Entre los años 2015 y 2016 se llevó a cabo un estudio concerniente a las personas que demandan asistencia en la Facultad de Odontología, de la Universidad de la República (UdelaR), Uruguay (Lorenzo, Álvarez-Vaz, Fabruccini y Massa, 2020). El mismo tenía como objetivos investigar el vínculo entre los FR de las ENT más prevalentes y la salud bucal, así como generar hipótesis referidas a la relación entre ambas patologías y sus FR común. Teniendo en cuenta esta situación se propone obtener perfiles epidemiológicos bien diferenciados, a partir de la determinación de tipologías de individuos en base a atributos categóricos, (más concretamente, binarios) referentes a ENT, las PO más frecuentes y los FR comunes a ambas, sin discriminar entre variables explicativas o explicadas.

2. Metodología

2.1. Datos

El presente estudio considera datos provenientes del estudio transversal llamado “*Salud Bucal y enfermedades no transmisibles. Un abordaje clínico-epidemiológico e interdisciplinario: Primera etapa*”, llevado a cabo entre los años 2015 y 2016 en Uruguay. Dicho estudio contó con la participación de 602 voluntarios/as y fue llevado a cabo para estimar la prevalencia de caries, paradenciopatías y ENT en personas que demandan atención en Facultad de Odontología de UdelaR.

Las variables consideradas para caracterizar la “salud oral” fueron diagnóstico de bolsa periodontal, pérdida de inserción periodontal, presencia de caries y dentición no funcional (menos de 20 piezas dentales). En relación a las ENT, se registraron diagnósticos de diabetes, hipertensión arterial, obesidad y síndrome metabólico (a través de la alteración en la razón cintura/cadera). Adicionalmente se incluyeron FR comportamentales como el consumo de tabaco, consumo nocivo de alcohol y la realización de actividad física de manera insuficiente.

2.2. Análisis estadístico

Se propone el uso de la metodología dirigida a crear grupos a partir de variables categóricas descrita en Tsekouras, *et al.* (2007). Dicha metodología constituye un análogo “difuso” del algoritmo *k-means*, adaptado a variables categóricas, que se basa en tres etapas:

- 1) construcción de una partición inicial rígida, empleando un esquema basado en entropía (Yao, Dash, Tan, y Liu, 2000). Este procedimiento tiene la ventaja de que es de fácil implementación y no asume la existencia de “centros” sino que permite una primera aproximación a la obtención de los mismos.

- 2) a partir de los centros determinados en el paso anterior, emplear el algoritmo *fuzzy c-modes* implementado por Huang y Ng (1999). Someramente, se podría resumir el

algoritmo en los siguientes pasos:

i) a partir de una matriz X , que contiene n individuos y p variables, construir una matriz de distancias D entre los vectores x_1, x_2, \dots, x_n , empleando la distancia *simple matching*,

$$D(x_i, x_j) = \sum_{k=1}^p \delta(x_{ij}, x_{ij'})$$

ii) luego, el algoritmo funciona minimizando iterativamente la función objetivo:

$$J_m(U, V) = \sum_{i=1}^n \sum_{g=1}^c u_{ig}^m D(x_i, v_g) \quad \text{s.a.} \quad \sum_{g=1}^c u_{ig} = 1$$

siendo v_g el g -ésimo centro (definido anteriormente), u_{ig} el grado de pertenencia del i -ésimo individuo al grupo g y m un parámetro empleado para indicar el grado de difusión (*fuzzyness*) de la partición,

iii) el algoritmo funciona de modo tal que en cada iteración se actualizan los centros v_g y los grados de pertenencia a cada grupo u_{ig} , hasta que se cumpla algún criterio de convergencia fijado por el usuario.

3) la determinación del número óptimo de grupos es llevada a cabo mediante la maximización del coeficiente de partición de Dunn (Dunn, 1974), el cual se basa en la difusión total obtenida, en base a la partición difusa generada al final del paso anterior. Este índice, en su versión normalizada, varía entre 0 y 1 correspondiendo estos valores a particiones más difusas o más rígidas. La forma de seleccionar el número óptimo de grupos es maximizando este coeficiente.

La descripción de los perfiles fue llevada a cabo empleando las mismas variables empleadas para la construcción de los mismos así como el sexo y la edad de los y las participantes del estudio. Para constatar cuáles de las variables presentaron diferencias significativas entre los grupos se estimó el *odds-ratio*¹ (OR) correspondiente a cada patología, realizándose en cada caso una prueba de hipótesis a efectos de determinar la diferencia de proporciones entre grupos.

3. Resultados

Del total de 602 participantes, el 58% eran de sexo femenino y la edad media fue de 45 años, con un desvío estándar de 16 años. En cuanto al máximo nivel educativo alcanzado, se observó que el 28% de los participantes alcanzaron la primaria, el 56% la secundaria y sólo el 16% presentó estudios terciarios.

Debido a que se trata de un conjunto de personas que demanda asistencia, se constata una elevada prevalencia de las PO así como de las ENT (Tabla 1).

1 El *odds-ratio* (cociente de chances) indica cuán mayor es la chance de observar un cierto suceso (por ejemplo una enfermedad) en un cierto grupo de interés respecto de otro grupo de referencia.

Tabla 1. Distribución de las variables FR, ENT y PO.

	n	%
Factores comportamentales		
Fuma a diario	199	33,1%
Consumo nocivo de alcohol	59	9,8%
Actividad física insuficiente	269	44,7%
Enfermedades no transmisibles		
Hipertensión arterial	260	43,2%
Diabetes	128	21,3%
Obesidad	345	57,3%
Síndrome metabólico	339	56,3%
Enfermedades orales		
Caries	438	72,8%
Dentición no funcional	359	59,6%
Bolsa periodontal	353	58,6%
Pérdida de inserción periodontal	383	63,6%

Fuente: Elaboración propia.

A partir de la metodología presentada anteriormente se confeccionaron escenarios que consideraran entre dos y cinco grupos. En la Tabla 2 se presenta el tamaño de los grupos obtenidos bajo cada especificación, así como el valor del indicador de validación (IVD) propuesto por Dunn. En base a este último, se decidió que la mejor alternativa consistía en considerar dos grupos ya que de este modo se maximiza el coeficiente de partición. Se evidencia que en todas las especificaciones los grupos presentan niveles de difusión considerables.

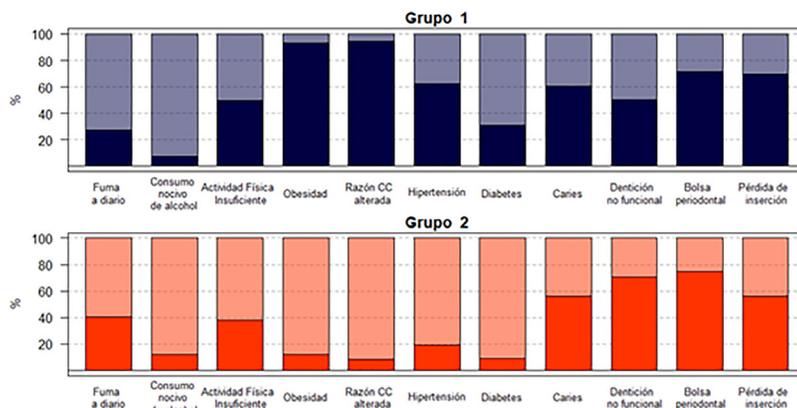
Tabla 2. Elección del número de grupos.

Nº de grupos	n1	n2	n3	n4	n5	IVD
dos	304	298	-	-	-	0,117
tres	200	201	202	-	-	0,041
cuatro	150	144	156	152	-	0,087
cinco	122	124	116	115	125	0,079

Fuente: elaboración propia.

A partir de los valores del IVD se optó por generar dos tipologías y se procedió a examinar la composición de los grupos. En este sentido, la Figura 1 presenta la proporción de participantes con cada ENT, PO y FR comportamentales a la interna de cada grupo.

Figura 1. Proporción de participantes según ENT, PO y FR en cada uno de los grupos



Fuente: Elaboración propia.

En la Figura 1, que presenta la composición de los grupos, puede observarse que los grupos se diferencian principalmente en las prevalencias de las ENT, sobre todo en obesidad, síndrome metabólico e hipertensión. Adicionalmente parecen evidenciarse mayores prevalencias en los tres FR comportamentales, en tanto que las PO parecen presentar valores similares en ambos grupos.

La Tabla 3 presenta los OR cada FR, ENT, PO, sexo, edad y educación entre los grupos considerados, así como el p-valor correspondiente.

A partir de la Tabla 3 se constatan las afirmaciones anteriores. De esta manera, sería posible definir a ambos grupos de la siguiente manera; un grupo donde se observa la coexistencia de ambas patologías, así como sus FR (grupo 1) y otro conformado por pacientes que padecen principalmente PO.

Al considerar los factores sociodemográficos, se puede observar que los grupos presentan una composición de género diferente, donde la proporción de mujeres del grupo 1 es del 66%, mientras que la del grupo 2 es de 48%.

También se pudo detectar que la composición etaria de los grupos es diferente, observándose que por cada año adicional, el *odds*² de pertenecer al grupo 2 disminuye un 5%. En este sentido la edad promedio de los participantes del grupo 1 fue de 50,6 años, mientras que la edad promedio del grupo dos fue de 38,1 años.

2 El *odd* (chance) proporciona una medida de la plausibilidad de la ocurrencia de un cierto suceso respecto la no ocurrencia del mismo.

Tabla 3. OR y p-valor, según prevalencia de FR, ENT, PO y variables sociodemográficas entre los grupos considerados.

	OR	p-valor
Factores comportamentales		
Fuma a diario	0,56	<0,01
Consumo nocivo de alcohol	0,66	0,07
Actividad física insuficiente	1,63	<0,01
Enfermedades no transmisibles		
Hipertensión arterial	7,11	<0,01
Diabetes	4,6	<0,01
Obesidad	100,9	<0,01
Razón cintura-cadera alterada	187,0	<0,01
Enfermedades orales		
Caries	0,84	0,36
Pérdida dentaria	0,42	<0,01
Bolsa periodontal	1,22	0,23
Pérdida de inserción periodontal	1,81	<0,01
Sociodemográficos		
Sexo	2,11	<0,01
Edad	0,95	<0,01
Educación	0,60	0,72

Fuente: Elaboración propia.

4. Conclusiones

A partir de este estudio es posible argumentar que los pacientes que demandaron atención odontológica en la Facultad de Odontología en Uruguay durante los años 2015 y 2016 pueden ser adecuadamente catalogados acorde a dos tipologías.

Por una lado, fue posible detectar un conjunto de pacientes, mayoritariamente mujeres con edades más avanzadas, que padecían tanto ENT como PO y los FR considerados. La otra tipología identificada presentó una composición de género equilibrada, con una edad promedio más baja. Si bien este conjunto de pacientes presentó FR y ENT en menor medida, las prevalencias de las PO fueron similares (en los casos de caries y bolsa periodontal), o presentaron diferencias significativas (en pérdida de inserción y dentición no funcional) pero menos acentuadas que las identificadas en cuanto a ENT.

A partir de estos hallazgos es posible argüir que los pacientes de edades más avanzadas que acuden a Facultad de Odontología de la UdelaR suelen presentar una situación de salud comprometida tanto en el ámbito odontológico como general. Esta situación puede explicarse por los malos hábitos alimenticios y comportamentales resumidos en los FR considerados, que con el devenir del tiempo propician el desarrollo de ENT y PO. Debido a que todos los pacientes que concurren a las clínicas presentan alguna PO, sería oportuno considerar trabajar en prevención con los pacientes más jóvenes (integrantes del grupo 2), no sólo de las PO sino también

de las ENT y los FR analizados, ya que estos últimos están asociados al desarrollo de las patologías consideradas.

A partir de este estudio es posible plantear la necesidad de profundizar el análisis, incluyendo factores relevantes a la dieta, hábitos de sedentarismo así como de higiene oral, sobre todo en la población joven. De esta manera, y teniendo en cuenta la interrelación entre ENT, PO y FR, sería posible definir estrategias de promoción de salud tanto oral como general.

Referencias Bibliográficas

- Álvarez, L., Liberman, J., Mangarelli, C., Brito, M., Demarco, F., Lorenzo, S., ... Nascimento, G. G. (2015). Dental caries in Uruguayan adults and elders: findings from the first Uruguayan National Oral Health Survey. *Cadernos de Saúde Pública*, 31(8), pp. 1663-1672.
- Asamblea General de las Naciones Unidas. (2011). Declaración Política de la Reunión de Alto Nivel de la Asamblea General sobre la Prevención y el Control de las Enfermedades No Transmisibles. *Seguimiento de los resultados de la Cumbre del Milenio*.
- Austregésilo, S., Angeiras, P., de Sena, M. y Costa, C. (2019). Clustering of oral and general health risk behaviors among adolescents. *Preventive Medicine Reports*, 15.
- Bose, S. (2019). Common Risk Factor Approach for Oral Diseases - A SWOT Analysis. *Acta Scientific Dental Sciences*, 3(9), pp. 7-9.
- Dunn, J. C. (1974). Well-Separated Clusters and Optimal Fuzzy Partitions. *Journal of Cybernetics*, 4(1), pp. 95-104.
- Genco, R. y Genco, F. (2014). Common risk factors in the management of periodontal and associated systemic diseases: the dental setting and interprofessional collaboration. *Journal of Evidence Based Dental Practice*, 14, pp. 4-16.
- Haregu, T. N., Wekesah, F., Mohamed, S., Mutua, M., Asiki, G. y Kyobutungi, C. (2018). Patterns of non-communicable disease and injury risk factors in Kenyan adult population: a cluster analysis. *BMC Public Health*, 18.
- Huang, A. y Ng, M. K. (1999). A fuzzy k-modes algorithm for clustering categorical data. *IEEE Transactions on Fuzzy Systems*, 7(4), pp. 446-452.
- Lorenzo, S., Álvarez-Vaz, R., Andrade, E., Piccardo, V., Francia, A., Massa, F., ... Peres, M. A. (2015). Periodontal conditions and associated factors among adults and the elderly: findings from the first National Oral Health Survey in Uruguay. *Cadernos Saude Pública*, 31(11), pp. 2425-2436.
- Lorenzo, S., Álvarez-Vaz, R., Blanco, S. y Peres, M. (2013). Primer Relevamiento Nacional de Salud Bucal en población joven y adulta uruguaya. *Odontostomatología*, 15.
- Lorenzo, S., Álvarez-Vaz, R., Fabruccini, A. y Massa, F. (2020). Salud Bucal y

enfermedades no transmisibles en pacientes de un centro de enseñanza universitaria, Montevideo, Uruguay. *Odontoestomatología*, 22(36).

- Nascimento, A. P., Prudêncio, C., Vieira, M., Pimenta, R. y Bacelar-Nicolau, H. (2019). Cluster Analysis of Noncommunicable Diseases in Portugal. *14th Iberian Conference on Information Systems and Technologies (CISTI)*. Coimbra, Portugal: IEEE.
- Nicola, P., Silveira, H., Derrégibus, L., Rodríguez, A., Demarco, F. y Lorenzo, S. (2016). Tooth loss and associated factors in elders: results from a national survey in Uruguay. *Journal of Public Health Dentistry*, 76(2), pp. 143-151.
- Petersen, P. (2003). The world oral health report. *WHO Oral Health Programme*. Ginebra, Suiza.
- Skapino, E. y Álvarez-Vaz, R. (2016). Prevalencia de factores de riesgo de enfermedades crónicas no transmisibles en funcionarios de una institución bancaria del Uruguay. *Revista Uruguaya de Cardiología*, 31(2).
- Tsekouras, G., Papageorgiou, D., Kotsiantis, S., Kalloniatis, C. y Pintelas, P. (2007). Fuzzy Clustering of Categorical Attributes and its Use in Analyzing Cultural Data. *International Journal of Computer and Information Engineering*, 1(1), pp. 192-196.
- WHO [World Health Organization]. (2013). *Global action plan for the prevention and control of noncommunicable diseases 2013-2020*. Ginebra, Suiza: WHO Press.
- Yao, J., Dash, M., Tan, S. T. y Liu, H. (2000). Entropy-based fuzzy clustering and fuzzy modeling. *Fuzzy Sets and Systems*, 113(3), pp. 381-388.
- Zancheta, C., Machado, C., Fórnias, L. y Bertazzi, R. (2019). Co-occurrence and clustering of the four major non-communicable disease risk factors in Brazilian adolescents: Analysis of a national school-based survey. *PLoS ONE*, 14(7).