

ANÁLISIS MULTIDIMENSIONAL DE UNA BASE DE DATOS DE MUJERES PIMA

MULTIDIMENSIONAL ANALYSIS FROM A DATABASE OF PIMA WOMEN

María Cristina Tarrés^{1,2,3}, Nora Moscoloni³, Hugo Navone³, Alberto Enrique D'Ottavio^{1,2}

¹ Facultad de Ciencias Médicas, Universidad Nacional de Rosario, Rosario, Santa Fe, Argentina.

² Consejo de Investigaciones Científicas y Técnicas, Universidad Nacional de Rosario, Rosario, Santa Fe, Argentina.

³ Programa Interdisciplinario de Análisis de Datos, Universidad Nacional de Rosario, Rosario, Santa Fe, Argentina.

RESUMEN

Se presenta una tipología multidimensional de mujeres estadounidenses Pima con datos de la *Pima Indians Diabetes Database*, analizada mediante Componentes Principales y posterior clasificación. Fueron construidos tres *clusters* individuales con: (1) 64% de positividad para diabetes, glucemia acorde con tolerancia alterada a la glucosa, hiperinsulinemia, obesidad según índice de masa corporal, grosor del pliegue de la piel del tríceps superior al promedio general y presión diastólica cercana a prehipertensión; (2) 50 % de positividad para diabetes, glucemia cercana al límite inferior de tolerancia alterada a la glucosa, índice de masa corporal correspondiente a obesidad y grosor del pliegue de la piel del tríceps, presión diastólica, edad y número de embarazos mayores a la media general; (3) 16% de positividad para diabetes, índice de masa corporal indicando sobrepeso y número de embarazos, función de pedigrí de diabetes, edad, glucemia, insulinemia y grosor del pliegue de la piel del tríceps menores al promedio general. Resulta destacable el mayor grado de antecedentes familiares en el grupo de riesgo y la relevancia del sobrepeso así como la configuración espacial de los individuos revelando posibles fases en el desarrollo diabético de estas mujeres. Finalmente, se subraya la utilidad de la *clusterización* en problemas biológicos.

Palabras clave: Diabetes, obesidad, mujeres EEUU, indios Pima, análisis multidimensional

ABSTRACT

A multidimensional typology of U.S. Pima women is presented with data obtained from *Pima Indians Diabetes Database*, and analyzed employing Principal Components and subsequent classification. Three individual *clusters* were obtained: (1) 64% of positivity for diabetes, glycemia compatible with abnormal glucose tolerance, hyperinsulinemia, body mass index corresponding to obesity, triceps skin fold thickness above the general average and diastolic pressure near to prehypertension; (2) 50% of positivity for diabetes, glycemia near the lower level of abnormal glucose tolerance, body mass index corresponding to obesity and skin fold thickness, diastolic pressure, age and number of pregnancies above the overall average; (3) 16% of positivity for diabetes, body mass index corresponding to overweight and number of pregnancies, diabetes pedigree function, age, glycemia,

insulinemia and skin fold thickness below the overall average. Greater family history in risk groups, relevance of overweight and the spatial configuration of subjects revealing possible developmental phases towards diabetes in these women have to be outlined. The usefulness of *clusterización* in biological disturbances is finally remarked.

Keywords: Diabetes, obesity, U.S. women, Pima Indians, multidimensional analysis

INTRODUCCIÓN

En el continente americano han sucedido serias transformaciones socioculturales cuyos efectos más conspicuos fueron modificaciones en el perfil epidemiológico, en los estilos de vida, en los patrones de consumo alimentario y en las tasas de morbi-mortalidad. Entre ellas, se destaca un incremento de las enfermedades crónico-degenerativas no transmisibles, como la obesidad y la diabetes (Peña y Bacallao, 2000; Ceballos Juárez *et al.*, 2012).

Haciendo foco en el grupo étnico akimel o'odham o Pima, asentado actualmente en el sudoeste estadounidense (Estado de Arizona) y en el noroeste mexicano (Estados de Sonora y Chihuahua), corresponde señalar que, excediendo su predisposición genética a la diabetes y a la obesidad (Knowler *et al.*, 1983; Hanson y Ehm, 1998; Narayan, 2016), no evidenciaba estas enfermedades mientras sus integrantes vivieron de la pesca y de la agricultura. Empero, hacia fines del siglo XIX, interrumpieron sus prácticas agrícolas habituales debido a la escasez de agua derivada de la migración de colonos desde el Este. Como consecuencia de ello, el grupo que residía primordialmente en los EEUU occidentalizó sus hábitos y adoptó, después de un breve lapso, costumbres "modernas" ajenas. En ese marco, su alimentación se modificó de manera radical hasta llegar a ser en la actualidad muy similar a la del promedio de la población estadounidense (González Barranco, 2004). Tal relevante cambio dietario, junto con una significativa reducción en el gasto energético, indujo un notorio incremento en la frecuencia de diabetes y obesidad (Knowler *et al.*, 1990; Knowler *et al.*; 1991, Knowler *et al.*, 1993; Krakoff *et al.*, 2001; Schulz *et al.*, 2001; Martorell, 2005).

En ese sentido, fue reportado recientemente que los indios Pima mexicanos, en comparación con los estadounidenses, mostraban menor prevalencia de diabetes tipo 2 y

*Autor para correspondencia: Alberto Enrique D'OttavioCorreo electrónico: aedottavio@hotmail.com

Recibido: 03 de agosto de 2016

Aceptado: 24 de noviembre de 2016

obesidad junto a un disminuido consumo de grasa y mayor nivel de actividad física, corroborando así el papel del estilo de vida en poblaciones genéticamente predispuestas a estas afecciones (Urquidez-Romero *et al.*, 2015).

Todo lo antes referido habla acerca de la persistencia de facetas susceptibles y dignas de ser indagadas en esta población.

Tomando en cuenta que uno de los abordajes de esta problemática se sustenta en el análisis de bases de datos efectuado, desde distintas perspectivas, a fin de diseñar algoritmos clasificadores o diagnósticos, asignando individuos a grupos previamente definidos (Khashei *et al.*, 2012, Gu *et al.*, 2014), y debido a que en la literatura especializada no fuera constatada la construcción de tipologías, el presente trabajo emplea la *Pima Indians Diabetes Database* (PIDD, EEUU) e intenta, mediante análisis exploratorio multidimensional, una categorización multivariada que permita identificar agrupamientos de individuos con mayor riesgo de desarrollar diabetes.

MATERIALES Y MÉTODOS

Se utilizó la *Pima Indians Diabetes Database* (PIDD), cuya propiedad original pertenece al National Institute of Diabetes and Digestive and Kidney Diseases, y los datos fueron obtenidos del UCI Machine Learning Repository - Pima Indians Diabetes Data Set (2016). Las unidades de análisis consistieron en 768 mujeres residentes cerca de Phoenix, Arizona, EEUU, pertenecientes a la etnia Pima y con al menos 21 años de edad. En ellas fueron registradas 9 variables: 8 numéricas y una clasificatoria, detalladas a continuación:

1. Concentración de glucosa plasmática a las 2hs de una prueba de tolerancia oral a la glucosa (G120 mg/dl)
2. Concentración de insulina sérica a las 2hs de una prueba de tolerancia oral a la glucosa (I120 mU/ml)
3. Presión arterial diastólica (PAD mmHg.)
4. Grosor del pliegue de la piel del tríceps (GPPT mm)
5. Índice de masa corporal (IMC= peso /altura al cuadrado= kg/m²)
6. Antecedentes Familiares o función de pedigrí de diabetes (FPD)
7. N° de embarazos (nE)
8. Edad (E años)
9. Variable clasificatoria (D: 0 - 1, donde 1 es interpretado como "test positivo para diabetes"). El diagnóstico estuvo basado en el criterio de la OMS (i.e.: G120 ≥ 200 mg/dl en cualquier examen o evaluación de rutina médica).

ANÁLISIS ESTADÍSTICO

Análisis de Componentes Principales

El enfoque multivariado incluye una serie de técnicas estadísticas que permiten considerar de manera simultánea un conjunto de variables medidas en un conjunto de individuos (Carrasco y Hernán, 1993). Entre ellas, el análisis de componentes principales es una técnica factorial empleada

para resumir un conjunto de variables continuas en un número pequeño de variables sintéticas (Escofier y Pagès, 1995; Aluja y Morineau, 1999).

Este método transforma el conjunto original en uno más pequeño de combinaciones lineales de las variables de origen, las que representan la mayor parte de la variancia del conjunto original. El objetivo es determinar factores o componentes principales que expliquen la mayor variación posible de los datos con la menor cantidad de dichos factores.

El enfoque de la escuela francesa de análisis de datos posibilita la elección de variables *activas* e *ilustrativas* (*o suplementarias*). Las primeras se incluyen en los cálculos de diagonalización de la matriz mientras que las segundas se proyectan *a posteriori* en el espacio factorial construido con las variables activas.

Como *activas* fueron consideradas las variables continuas: Concentración de glucosa plasmática (G120mg/dl) y de insulina sérica (I120mU/ml), ambas a las 2hs de una prueba de tolerancia oral a la glucosa, Presión arterial diastólica (PAD mmHg), Grosor del pliegue de la piel del tríceps (GPPTmm), Índice de masa corporal (IMC) y Antecedentes Familiares o función de pedigrí de diabetes (FPD).

Como *ilustrativas* lo fueron el N° de embarazos (nE), la edad (E años) y la variable clasificatoria nominal (test positivo o negativo para diabetes) (D).

Construcción de una tipología de individuos

Este método alude a la formación de clases o *clusters* con características semejantes, permite la observación de las variables que se encuentran asociadas (Lebart *et al.*, 1995) y persigue reagrupar los individuos en un número restringido de clases homogéneas y separadas entre sí.

Para la construcción de la tipología fueron tenidas en cuenta las coordenadas de los mismos en los dos primeros ejes factoriales considerando este subespacio como el más significativo, dada la relevancia del efecto tamaño para los objetivos de este trabajo, y se aplicó la estrategia de clasificación mixta de extendida aplicación en muestras grandes (Moscoloni, 2005).

Con posterioridad, fue efectuada una clasificación ascendente jerárquica cuyos elementos terminales del árbol fueron las clases de la partición anterior, construido según el criterio de Ward. La partición final fue definida por corte del árbol correspondiente y la homogeneidad de las clases obtenidas, optimizada por reasignaciones realizadas nuevamente con el procedimiento de agregación alrededor de centros móviles.

Finalmente, se hizo una descripción del contenido de las clases a partir de las variables *activas* e *ilustrativas*. El procedimiento permite la interpretación de las clases formadas en términos de las variables originales, utilizándose un criterio para jerarquizar las modalidades más características basado en el cálculo de valores *tests* en el entorno de una distribución Normal (0.1). Para ello se realizó una comparación de medias, respondiendo a la idea de considerar una variable continua como más característica de un grupo si su

promedio es significativamente diferente del general, que corresponde al del conjunto total de individuos. Para la variable clasificatoria fueron comparados porcentajes siguiendo la idea de considerar una modalidad como más característica de un grupo si su frecuencia relativa es significativamente mayor en una clase y, al mismo tiempo, menor en las otras (Morineau, 1984).

El procesamiento estadístico se llevó a cabo con el software SPAD (Lébart *et al.*, 2000).

RESULTADOS

La descripción simultánea de la estructura de los datos efectuada a través del análisis de componentes principales, puede visualizarse mediante una representación gráfica en ejes factoriales. Sobre un plano dividido en cuadrantes, se ubican las variables activas de acuerdo al valor de las correlaciones obtenidas, observándose en la Figura 1 que, con referencia al primer factor, todas las variables están en el mismo lado respecto del origen de coordenadas. Tal disposición evidencia el "efecto tamaño" característico de los datos biométricos, traduciendo el hecho de que la mayoría están positivamente correlacionadas entre sí por lo que el primer eje factorial separa los individuos según los valores de G120, PAD, GPPT, I120, IMC, FPD, nE y E y el segundo eje factorial opone G120, I120 y FPD a PAD, GPPT e IMC.

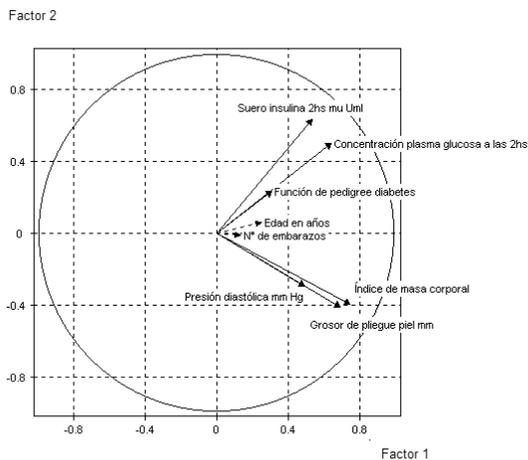


Figura 1. Representación gráfica de las variables activas e ilustrativas en los ejes factoriales

Figure 1. Graphic representation of active and illustrative variables in factorial axes

Se realizó, luego, un análisis en *clusters* para agrupar los individuos según sus afinidades con las variables estudiadas, aplicando las técnicas de clasificación sobre las coordenadas de los individuos en los ejes factoriales. A partir del árbol jerárquico (o dendrograma) construido mediante el algoritmo ascendente, a cuyo través las clases se van reagrupando progresivamente, fue definida una partición en tres clases.

Los individuos fueron proyectados en el gráfico factorial e identificados por su número de clase, obteniéndose los agrupamientos mostrados en la Figura 2. En ella se observa

también la ubicación de las modalidades de la variable ilustrativa clasificatoria "test positivo o negativo de diabetes" (D).

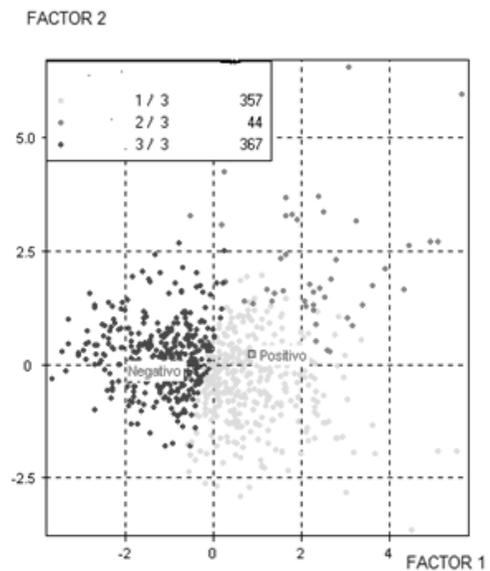


Figura 2. Proyecciones de los individuos según su clase y las modalidades de la variable clasificatoria.

Figure 2. Projections of the subjects in accordance with their class and the modality of the classificatory variable.

La configuración de los *clusters* a través de las variables activas y de la *ilustrativa* clasificatoria así como las comparaciones entre los promedios o porcentajes de las clases respecto del general son observables en la Tabla 1.

En la Tabla se percibe que las variables activas se hallan ubicadas en orden decreciente; esto es, mayores valores en las primeras clases y menores en la última. Las variables ilustrativas número de embarazos, antecedentes familiares y resultados del test de diabetes acompañaron este ordenamiento.

DISCUSIÓN

Las aplicaciones biomédicas del análisis de bases de datos han despertado gran interés en el desarrollo de algoritmos de clasificación basados en lógica difusa y en otras técnicas de inteligencia artificial, como redes neuronales artificiales.

Pima Indians Diabetes Database (PIDDD) es un conjunto de datos que ha sido ampliamente estudiado mediante diferentes procedimientos y resultados, persiguiendo fundamentalmente obtener algoritmos susceptibles de ser empleados como clasificadores o diagnósticos. Entre ellos, el proyecto de Smith *et al* (1988) es considerado uno de los primeros a cuyo través, trabajando redes neuronales para predecir la aparición de la diabetes en la etnia Pima, se obtuvo un porcentaje de predicción de un 76%. Carpenter y Markuzon (1998) lograron un acierto del 81%, y Breault *et al.* (2002), de un 83%. Con el desarrollo de clasificación difusa aplicada al

Tabla 1. Configuración de los *clusters* a través de las variables activas e ilustrativas continuas (promedio \pm desvío estándar) y de los porcentajes de la ilustrativa clasificatoria. Se incluye la probabilidad asociada al valor test que compara el valor de la clase con el general.

Table 1. Configuration of *clusters* through continuous active and classificatory variables (mean \pm standard deviation) and the percentages of the classificatory illustrative one. The probability associated to the test value by comparing class with general values is included.

VARIABLES	GENERAL	CLASE 1 n=357		CLASE 2 n=44		CLASE 3 n=367	
	$\bar{x} \pm SD$	$\bar{x} \pm DE$	p	$\bar{x} \pm DE$	p	$\bar{x} \pm DE$	p
G ₁₂₀ mg/dL	121.69 ± 30.5	131.43 ± 29.6	0.000	164.55 ± 21.32	0.000	107.07 ± 22.66	0.000
I ₁₂₀ mU/mL	155.55 ± 118.63	---	---	415.21 ± 134.35	0.000	105.42 ± 59.46	0.000
PAD mmHg	72.41 ± 12.37	78.49 \pm 11.77	0,000	---	---	66.37 ± 9.96	0.000
GPPT mm	29.15 ± 10.47	35.91 \pm 10.47	0.000	33.52 ± 8.55	0.002	22.04 ± 6.95	0.001
IMC kg/m ²	32.46 ± 6.92	36.79 \pm 6.04	0.000	36.17 ± 5.70	0.000	27.71 ± 4.28	0.000
FPD	0.47 ± 0.33	---	---	0.61 ± 0.46	0.003	0.43 ± 0.29	0.000
Nº de embarazos (nE)	3.85 ± 3.37	4.29 \pm 3.70	0.000	---	---	3.40 ± 2.96	0.000
EDAD (años)	33.24 ± 11.75	35.78 \pm 11.88	0.000	---	---	30.59 ± 10.93	0.000
TEST DIABETES	+ : 34.90 % - : 65.10 %	+ : 50.42% - : 49.58%	0.000 0.000	+ : 63.64% - : 36.36%	0.000 0.000	+ : 16.35% - : 83.65%	0.000 0.000

campo del diagnóstico de diabetes, Contreras *et al.* (2010), utilizando la PIDD, alcanzaron un 83% de positividad en el diagnóstico, uno de los mejores porcentajes encontrados hasta ahora; Lekkas y Mikhilov (2010) consiguieron 79% de acierto y, recientemente, Gu *et al.* (2014) desarrollaron un nuevo algoritmo que recomiendan para el análisis de bases con problemas de desbalance, ruido y presencia de *outliers*. Cabe destacar que Khashei *et al.* (2012), con el objetivo de lograr una aplicación que colabore a planear procedimientos de prevención, diagnóstico o tratamiento, postulan una combinación de varias técnicas de clasificación con resultados que los llevan a proponer al modelo obtenido como superior a todos los alternativos.

Desde otra perspectiva, en el presente trabajo fue utilizada la PIDD para construir una tipología multidimensional de las mujeres Pima estadounidenses a través de los valores de las variables continuas y la clasificatoria diagnóstica, mediante análisis de Componentes Principales (Aluja y Morineau, 1999) y posterior clusterización. De tal modo, la base fue dividida en grupos subconjuntos diferentes con miembros similares entre sí por medio de la segmentación

de los datos en grupos que no habían sido predefinidos (Moscoloni, 2005).

Así, se obtuvo un cluster, que podría considerarse como el más comprometido, formado por unidades de análisis que presentaron 64% de positividad para diabetes, glucemia promedio compatible con tolerancia alterada a la glucosa ($140 < G_{120} < 199$ mg/dl) (American Diabetes Association, 2014), hiperinsulinemia, IMC correspondiente a sujetos obesos (≥ 30 kg/m²) (Schulz *et al.*, 2006); GPPT superior al promedio general y PAD cercana al límite inferior de la categoría de prehipertensión (80 mmHg) (Moser, 2004). Un segundo grupo, que podría ser catalogado igualmente en situación de riesgo, estuvo constituido por individuos con un 50 % test positivo para diabetes, IMC semejante al de obesos (≥ 30 kg/m²) y GPPT, PAD, edad y número de embarazos superiores a la media general. A su vez, la G₁₂₀, se reveló cercana al límite inferior de tolerancia alterada a la glucosa. El tercer y último cluster se integró con personas en situación menos desventajosa, con 16% de test positivo para diabetes y número de embarazos, FPD, edad, G₁₂₀, I₁₂₀ y GPPT menores al promedio general. Su IMC fue indicador de

sobrepeso ($25 \text{ kg/m}^2 \leq \text{BMI} < 30 \text{ kg/m}^2$). Dada la reconocida influencia de factores genéticos en el desarrollo de diabetes y obesidad dentro de la etnia Pima (Knowler *et al.*, 1983, Hanson y Ehm, 1998), resulta de interés señalar el mayor grado de antecedentes familiares en el grupo de riesgo,

En la obtenida configuración espacial de los individuos, detectora de diferencias multivariadas que bien podrían interpretarse como fases (Hansen y Bodkin, 1986) en el desarrollo de la diabetes de las mujeres Pima estadounidenses, se percibe la relevancia del sobrepeso. Considerando que la magnitud y rápido crecimiento de la prevalencia de obesidad en determinadas poblaciones ocurre en personas cada vez más jóvenes (Violante Ortiz, 2001), lo antedicho habilita, por ende, a la identificación temprana de individuos en riesgo a efectos de aplicar oportunas medidas preventivas (Lakka *et al.*, 2002).

Concluyendo, la aplicación del análisis multidimensional a la PIDD no sólo permitió concretar un perfil de mujeres de la etnia Pima estadounidense a partir de la consideración conjunta de variables sino que, a la par, confirma la utilidad de los algoritmos de *clusterización* para el abordaje de problemas biológicos.

REFERENCIAS

- Aluja, B. y Morineau, A. 1999. Aprender de los Datos: el Análisis de Componentes Principales. Una Aproximación desde el Data Mining. EUB. Barcelona, España. pp. 60-73.
- American Diabetes Association. 2014. Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care* Jan; 37 Suppl. 1: S81-90.
- Breault, J.L., Goodall, C.R., y Fos, P.J. 2002. Data mining a diabetic data warehouse. *Artif Intell Med*. 26(1-2): 37-54.
- Carpenter, G.A. y Markuzon, N. 1998. ARTMAP-IC and medical diagnosis: instance counting and inconsistent cases. *Neural Networks* 11 (2): 323-336.
- Carrasco, J.L. y Hernán, M.A. 1993. Estadística Multivariante en las Ciencias de la Vida. Fundamentos, Métodos y Aplicación. Ciencia 3. Madrid, España. pp. 123-152.
- Ceballos Juárez, C.L., Vizcarra Bordi, I., Diego Acosta, L., Reyes-Ortiz, C.A., y Loza Torres, M. 2012. Sobrepeso y obesidad en preescolares y escolares de una comunidad periurbana de origen otomí del Valle de Toluca, México. *Población y Salud en Mesoamérica* 10, artículo 2 [http://revistas.ucr.ac.cr/index.php/psm] Consulta: 20 julio 2016.
- Contreras, J.A., Martínez, L.B., y Puerta, Y.V. 2010. Clasificador Difuso para Diagnóstico de Enfermedades. *Rev Tecno Lógicas* 25: 201-220.
- Escofier, B. y Pagès, J. 1995. Análisis factoriales simples y múltiples. Objetivos, métodos e interpretación. Servicio editorial de la Universidad de País Vasco (UPV/EHU).
- González Barranco, J. 2004. Influencia de los cambios en el estilo de vida en el desarrollo de la obesidad. El ejemplo de los indios Pimas y otros. *Monografías Humanitas*, 6: 109-123.
- Gu, X., Ni, T., y Wang, H. 2014. New Fuzzy Support Vector Machine for the Class Imbalance Problem in Medical Datasets Classification. *The Scientific World Journal*. Vol. 2014, Article ID 536434, 12 pages [http://www.hindawi.com/journals/tswj/2014/536434/cta/] Consulta 25 julio 2016.
- Hansen, B.C. y Bodkin, N.L. 1986. See comment in PubMed Commons below Heterogeneity of insulin responses: phases leading to type 2 (non-insulin-dependent) diabetes mellitus in the rhesus monkey. *Diabetologia* 29 (10): 713-719.
- Hanson, R.L., y Ehm, M.G. 1998. An autosomal genomic scan for loci linked to type II diabetes mellitus and body-mass index in Pima Indians. *Am J Hum Genet* 63 (4): 1130-1138.
- Khashei, M., Eftekhari, S., y Parvizian, J. 2012. Diagnosing Diabetes Type II Using a Soft Intelligent Binary Classification Model. *Review of Bioinformatics and Biometrics (RBB)* 1: 9-23.
- Knowler, W.C., Pettit, D.J., Bennett, P.H., y Williams, R.C. 1983. Diabetes mellitus in the Pima Indians: genetic and evolutionary considerations. *Am J Physiol Anthropol* 62:107-114.
- Knowler, W.C., Pettit, D.J., Saad, F., y Bennett, P.H. 1990. Diabetes mellitus in the Pima-Indians: incidence, risk factors, and pathogenesis. *Diabetes Metab Rev* 6 (1):1-27.
- Knowler, W.C., Pettit, D.J., Saad, F., Charles, M.A., Nelson, R.G., Howard, B.V., Bogardus, C., y Bennet, P.H. 1991. Obesity in Pima-Indians: its magnitude and relationship with diabetes. *Am J Clin Nutr* 53 (6 Suppl.): 1453S-1551S.
- Knowler, W.C., Saad, F., Pettit, D.J., Nelson, R.G., y Bennett, P.H. 1993. Determinants of diabetes mellitus in the Pima Indians. *Diabetes Care* 16(1):216-227.
- Krakoff, J., Hanson, R.L., Kobes, S., Knowler, y W.C. 2001. Comparison of the effect of plasma glucose concentrations on microvascular disease between Pima Indian youths and adults. *Diabetes Care*. 24 (6):1023-1028.
- Lakka, H.M.; Laaksonen, D.E.; Lakka, T.A.; Niskanen, L.K.; Kumpusalo, E.; Tuomilehto, J. *et al.* 2002. The metabolic syndrome and total and cardiovascular disease mortality in middle-aged men. *JAMA* 288 (21), 2709-2716
- Lébart, L., Morineau, A., Lambert, T., y Pleuvret, P. 2000. *Système Pour l'Analyse des Données SPAD. Versión 4.51*. CD. CISIA-CERESTA. París, Francia.
- Lébart, L., Morineau, A., y Piron, M. 1995. *Statistique Exploratoire Multidimensionnelle*. Dunod. París, Francia. pp. 1-10, 67-142, 148-184.
- Lekkas, S. y Mikhailov, L. 2010. Evolving fuzzy medical diagnosis of Pima Indians diabetes and of dermatological diseases. *Artif Intell Med* 50(2):117-126.
- Martorell, R. 2005. La diabetes y los mexicanos: ¿Por qué están vinculados? *Prev Chronic Dis* [serial online] Jan [date cited]. [https://www.cdc.gov/pccd/issues/2005/jan/04_0100_es.htm] Consulta: 15 julio 2016.
- Morineau, A. 1984. Note sur la caractérisation statistique d'une classe et les valeurs-test. *Bull. Tech. Centre Stat. Inf. Appl.* 2: 20-27.
- Moscoloni, N. 2005. *Las Nubes de Datos. Métodos para Analizar la Complejidad*. UNR. Rosario, Argentina. pp. 41-52.
- Moser, M. 2004. Update on the management of hypertension: recent clinical trials and the JNC 7. *J Clin Hypertens (Greenwich)* 6 (10 Suppl. 2):4-13.
- Narayan, K.M. 2016. Type 2 Diabetes: Why We Are Winning the Battle but Losing the War? 2015 Kelly West Award Lecture. *Diabetes Care* 39(5):653- 663.
- Peña, M., y Bacallao, J. 2001. La obesidad y sus tendencias en la Región. *Rev Panam Salud Publica* 10 (2): 45-78.
- Schulz, L.O., Bennett, P.H., Ravussin, E., Kidd, J.R., Kidd, K.K., Esparza, J., y Valencia, M.E. 2006. Effects of traditional and western environments on prevalence of type 2 diabetes in Pima Indians in Mexico and the U.S. *Diabetes Care* 29(8): 1866-1871.

Smith, J.W., Everhart, J.E., Dickson, W.C., Knowler, W.C., y Johannes, E.S. 1988. Using the ADAP Learning Algorithm to Forecast the Onset of Diabetes Mellitus. *Proc Annu Symp Comput Appl Med Care* 261–265 [<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC 2245318/>] Consulta 27 julio 2016.

UCI Machine Learning Repository - Pima Indians Diabetes Data Set. (2016). National Institute of Diabetes and Digestive and Kidney Diseases. Donor of database: Vincent Sigillito. [<https://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes>] Consulta 30 junio 2016

Urquidez-Romero, R., Esparza-Romero, J., y Valencia, M.E. 2015. Interacción entre genética y estilo de vida en el desarrollo de la diabetes mellitus tipo 2: el estudio en los indios Pima *BIOtecnia* XVII (1):40-46.

Violante Ortiz, R.M. 2001. Obesidad y diabetes tipo 2 en el niño. Una nueva epidemia. *Rev Endocrinol. Nutr.* 9 (2): 103-106.