ARTÍCULO ORIGINAL

Predicción de la diabetes mellitus tipo 2 mediante un modelo de regresión logística

Prediction of Type 2 Diabetes Mellitus using a Logistic Regression Model

Olivia Altamirano Guerrero[1] Ronelsys Martínez Martínez[1] Jhonny Alejandro Rodríguez Gutiérrez[1]

[1] Universidad Regional Autónoma de los Andes, Ambato, Ecuador

Cómo citar este artículo:
Guerrero O, Martínez R, Gutiérrez J. Predicción de la diabetes mellitus tipo 2 mediante un modelo de regresión logística. **Medisur** [revista en Internet]. 2023 [citado 2026 Nov 9]; 21(6):[aprox. 6 p.]. Disponible en: https://medisur.sld.cu/index.php/medisur/article/view/45108

Resumen

Fundamento: la diabetes mellitus tipo 2 constituye una epidemia en aumento y supone una carga económica sustancial para los sistemas de salud. Detectar la enfermedad en una etapa temprana ayuda a reducir los costos médicos y el riesgo de que los pacientes tengan problemas de salud más complicados.
Objetivo: diseñar un modelo matemático para la predicción de la probabilidad de existencia de diabetes mellitus tipo 2 en pacientes atendidos en un hospital de Guayaquil, Ecuador.
Método: se realizó un estudio descriptivo y transversal. La población estuvo conformada por 324 pacientes. El procedimiento estadístico se basó en la aplicación de la regresión logística binaria. Para evaluar la capacidad predictiva del modelo, se empleó la prueba Kappa de Cohen.
Resultados: la hipertensión arterial fue un factor de riesgo positivo para la diabetes mellitus tipo 2, con un coeficiente de probabilidad de 1,415. Los antecedentes familiares positivos incidieron en el aumento de la probabilidad. El consumo de alcohol fue un factor de riesgo positivo y el coeficiente de 0,790 indicó cuánto contribuyó al aumento de la probabilidad. El coeficiente Kappa, tuvo un valor de 0,434; con T aproximada de 7,809 y $p < 0,001$, indicó mayor prevalencia que sesgo y una mayor concordancia entre lo predicho en el modelo y lo observado.
Conclusiones: la presencia de hipertensión arterial, antecedentes familiares positivos y el consumo de alcohol fueron factores significativos que aumentaron la probabilidad de desarrollar diabetes mellitus tipo 2. Es importante la detección temprana y la gestión de estos factores de riesgo en la prevención y el manejo de la enfermedad.

Palabras clave: diabetes mellitus tipo 2, predicción, factores de riesgo, modelos logísticos, diagnóstico, ecuador

Abstract

Foundation: type 2 diabetes mellitus constitutes a growing epidemic and represents a substantial economic burden for health systems. Detecting the disease at an early stage helps reduce medical costs and the risk of patients having more complicated health problems.
Objective: to design a mathematical model to predict the type 2 diabetes mellitus probability of existence in patients treated at a hospital in Guayaquil, Ecuador.
Method: a descriptive and cross-sectional study was carried out. The population was made up of 324 patients. The statistical procedure was based on the binary logistic regression application. To evaluate the predictive capacity of the model, Cohen's Kappa test was used.
Results: high blood pressure was a positive risk factor for type 2 diabetes mellitus, with a probability coefficient of 1.415. Positive family history influenced the increased probability. Alcohol consumption was a positive risk factor and the coefficient of 0.790 indicated how much it contributed to the increased probability. The Kappa coefficient had a value of 0.434; with approximate T of 7.809 and $p < 0.001$, it indicated greater prevalence than bias and greater agreement between what was predicted in the model and what was observed.
Conclusions: the presence of high blood pressure, positive family history and alcohol consumption were significant factors that increased the probability of developing type 2 diabetes mellitus. Early detection and management of these risk factors is important in the prevention and management of the illness.

Key words: diabetes mellitus type 2, prediction, risk factors, logistic models, diagnosis, ecuador

Aprobado: 2023-12-20 14:00:47

Correspondencia: Olivia Altamirano Guerrero. Universidad Regional Autónoma de los Andes. Ambato. Ecuador. ua.oliviaaltamirano@uniandes.edu.ec

INTRODUCCIÓN

La prevalencia de la diabetes ha aumentado significativamente con el envejecimiento de la población mundial y los cambios en los patrones dietéticos. Se espera que la diabetes sea la séptima causa de muerte en todo el mundo para 2030, según la Federación Internacional de Diabetes, mientras que la epidemia está en aumento y supone una carga económica sustancial para los sistemas de salud.[1] La diabetes es una de las principales causas de ceguera, insuficiencia renal, ataques cardíacos, accidentes cerebrovasculares y amputación de miembros inferiores. También se estima que alrededor de 84,1 millones de estadounidenses mayores de 18 años tienen prediabetes.[2]

La diabetes mellitus tipo 2 (DM2) puede aumentar sustancialmente el riesgo de enfermedad macrovascular y microvascular, especialmente en aquellos con un control glucémico inadecuado. La progresión de la DM2 por alteración de la glucosa en ayunas suele ser lenta, y lo que es más importante, sus síntomas pueden pasar inadvertidos durante muchos años. Los retrasos en el diagnóstico son un factor significativo que contribuye al control deficiente y al riesgo de complicaciones.[3]

En el ámbito de las evaluaciones económicas de los tratamientos para la diabetes, es común emplear modelos de rentabilidad que abarcan horizontes temporales prolongados. Por lo tanto, la precisión de las predicciones y la consideración de la incertidumbre estructural se convierten en aspectos cruciales en estos modelos, especialmente al tratar enfermedades crónicas.[4,5]

Investigaciones anteriores han relacionado la diabetes con factores de riesgo debido a los estilos de vida, como el índice de masa corporal (IMC), una dieta poco saludable, el consumo de alcohol, el tabaquismo y la inactividad física.[1] Detectar la enfermedad en una etapa temprana ayuda a reducir los costos médicos y el riesgo de que los pacientes tengan problemas de salud más complicados. Para ello se han propuesto y desarrollado varios métodos estadísticos predictivos, que incluyen modelos de aprendizaje automático: *CATBoost*, *XGBoost*, *Random Forest* (RF) (por sus siglas en inglés), regresión logística y máquinas de vectores de soporte.[1,6]

El objetivo del estudio fue diseñar un modelo matemático que permitiera predecir la probabilidad de DM2 en pacientes atendidos en un hospital de Guayaquil.

MÉTODOS

Se desarrolló un estudio descriptivo y transversal. La población de estudio estuvo conformada por 324 pacientes mayores de 18 años que asistieron a consulta en un hospital ecuatoriano de la ciudad de Guayaquil, con diagnóstico confirmado de DM2. Se excluyeron los pacientes que tuvieron diagnóstico de diabetes mellitus tipo 1 o de otro tipo; enfermedades crónicas graves o trastornos metabólicos no relacionados.

El procedimiento estadístico llevado a cabo en este estudio se centró en la regresión logística binaria, técnica utilizada para predecir valores de una variable categórica dicotómica, en este caso, la presencia o ausencia de DM2. Se construyó un modelo matemático predictivo donde el antecedente personal de hipertensión arterial (HTA), el antecedente familiar de DM2 y el consumo personal de alcohol, se tomaron como variables predictoras. Se utilizó el software SPSS versión 27 para el análisis multivariado para identificar las variables incluidas y calcular los coeficientes β para cada predictor, se incluyó la constante β0.

Para evaluar la capacidad predictiva del modelo, se empleó la prueba Kappa de Cohen como medida de concordancia. Este enfoque metodológico permitió obtener información valiosa sobre la relación entre las variables predictoras y la DM2, así como la capacidad del modelo para predecir la presencia de la enfermedad.

En este estudio, se consideraron diversos criterios éticos para garantizar la integridad y el bienestar de los participantes. Estos criterios incluyeron la obtención del consentimiento informado de los pacientes antes de su participación en el estudio, con respeto a su autonomía y derecho a la información, además, se protegió la privacidad y confidencialidad de los datos médicos y personales de los pacientes, cumpliendo con estándares éticos de manejo de información sensible. El estudio se llevó a cabo siguiendo pautas y regulaciones éticas establecidas por la institución hospitalaria. Estas consideraciones éticas fueron fundamentales para asegurar que la investigación se realizara de manera ética y responsable y se respetaran los derechos y la dignidad de los participantes.

RESULTADOS

En el paso inicial para el modelo de regresión logística binaria, se evaluaron las variables predictoras: HTA, antecedente familiar y consumo de alcohol. Sin excepción, todas las variables seleccionadas como predictoras resultaron estadísticamente significativas y se mantuvieron en el modelo de regresión. (Tabla 1).

Al incluir las variables en la ecuación del modelo, todas resultaron estadísticamente significativas, con valores de Wald relativamente altos para la HTA y el antecedente familiar, lo que indica que estas variables tuvieron mayor impacto en la predicción del resultado del modelo. Lo anterior se constata con los valores Exp(B) de estas variables: 4,117 y 3,447 respectivamente; también, fueron los mayores del modelo. (Tabla 2).

Se realizó la prueba Kappa de Cohen como medida de concordancia entre las observaciones reales y las predicciones del modelo de regresión. La prueba arrojó que existió concordancia estadísticamente significativa entre lo predicho por el modelo y lo observado en los datos de la población. El coeficiente Kappa, tuvo un valor de 0,434; con T aproximada de 7,809 y p-valor menor de 0,001. (Tabla 3).

DISCUSIÓN

Se ha demostrado que la regresión logística es uno de los algoritmos eficientes en la construcción de modelos de predicción. La precisión del modelo no solo depende del algoritmo elegido, también de otros factores. El preprocesamiento de datos es uno de esos factores. Eliminar valores redundantes y nulos es esencial para mejorar el rendimiento. La normalización de valores, cuando las características difieren a gran escala, también, juega un papel importante. En este artículo, hemos visto que la selección de funciones juega un rol importante a la hora de aumentar la precisión y reducir el tiempo de ejecución. La combinación de varios algoritmos, como se observó en las técnicas de conjunto, sirve como factor para mejorar el rendimiento del modelo. La validación cruzada, también, desempeña un papel esencial para aumentar la precisión.[2]

La prevalencia de la DM2 está en constante aumento en todo el mundo, resultado que también incluye al Ecuador. Este aumento se atribuye a factores como el envejecimiento de la población, cambios en el estilo de vida y la obesidad. Por lo tanto, la predicción temprana y la identificación de factores de riesgo son esenciales para abordar esta creciente carga de enfermedad. Identificar a las personas en riesgo de desarrollar DM2 permite la implementación de intervenciones preventivas y terapéuticas tempranas, lo que puede ayudar a retrasar o incluso prevenir la enfermedad y sus complicaciones. Esto es particularmente relevante para reducir los costos de atención médica a largo plazo.[1,2,3,4,5]

Con los avances en tecnología y la disponibilidad de datos de salud registrados electrónicamente, la aplicación de modelos de regresión logística binaria para predecir la DM2 se ha vuelto más factible y precisa. Esto refleja la actualidad y pertinencia de utilizar herramientas avanzadas en el campo de la atención médica. La predicción de la DM2 permite una atención más personalizada, donde se pueden adaptar

estrategias de prevención y tratamiento de acuerdo con el riesgo individual de cada paciente, lo que está en consonancia con la tendencia actual en la medicina.[2]

La DM2 tiene un impacto significativo en la calidad de vida de los pacientes. Prevenir o retrasar su desarrollo puede mejorar la calidad de vida de las personas y reducir la carga emocional y económica asociada a la enfermedad. Entre futuros estudios, es esencial investigar la influencia de los productos derivados del tabaco en el desarrollo de la DM2, en particular, los productos sin humo del tabaco.[7]

Resulta de interés conocer que un estilo de vida sedentario y una alimentación poco saludable, cuando se suman al sobrepeso, se convierten en factores de riesgo para la DM2, aunque las intervenciones en el estilo de vida que incluyen la pérdida de peso demuestran ser efectivas en la prevención de la DM2, su éxito puede verse obstaculizado por la falta de seguimiento y la presencia de estrés crónico.[8] Igualmente, es interesante el hallazgo de que el consumo de té en cantidades elevadas, mayor o igual a 4 tazas al día, se asocia con el riesgo de desarrollar DM2.[9]

A pesar de la amplia utilización de los medicamentos convencionales en el tratamiento de la DM2, sus resultados terapéuticos son altamente variables. Esta variabilidad en la progresión de la DM2 y en la respuesta al tratamiento se debe en gran medida a la interacción entre el microbioma intestinal y los fármacos hipoglucemiantes. Por un lado, estos medicamentos pueden modificar los componentes del microbioma intestinal. Por otro lado, la composición específica de la microbiota intestinal puede influir en el control de los niveles de glucosa en sangre y en los efectos terapéuticos de estos medicamentos.[10]

Lai y Wilson, en sus respectivos estudios desarrollaron el modelo de puntuación de riesgo de diabetes de Framingham para predecir el riesgo de desarrollar DM en adultos estadounidenses de mediana edad mediante regresión logística. Los factores de riesgo considerados en este modelo simple fueron antecedentes parentales de DM, obesidad, presión arterial alta, niveles bajos de colesterol con lipoproteínas de alta densidad, niveles elevados de triglicéridos y alteración de la glucosa en ayunas. [6, 11]

Aún no se comprenden completamente los procesos metabólicos responsables de la variabilidad en el riesgo de eventos cardiovasculares graves (MACE) (por sus siglas en inglés) en individuos con DM2, sin embargo, se observa que los metabolitos que indican una desregulación en la oxidación de los ácidos grasos mitocondriales se encuentran en concentraciones superiores en personas con DM2 que posteriormente experimentan MACE.[12]

No se ha realizado una investigación exhaustiva sobre el impacto de la exposición al tabaquismo pasivo en la probabilidad de desarrollar DM2, sin embargo, se observa que la exposición al tabaquismo pasivo se asociada con un mayor riesgo de DM2.[13] En el contexto del presente estudio, este hallazgo subraya la importancia de considerar múltiples factores de riesgo, entre los que se incluye el tabaquismo pasivo, al evaluar y prevenir la DM2 en la población estudiada.

El consumo de *cannabis* tiene una serie de efectos en el metabolismo, que incluyen la influencia en la regulación de la glucosa y la secreción de insulina. Algunos estudios sugieren un posible efecto protector del consumo de cannabis en relación con el desarrollo de la DM2, sin embargo, debido a la variabilidad en los resultados de diferentes investigaciones, es necesario llevar a cabo estudios más rigurosos con niveles de evidencia más sólidos. Esto es particularmente relevante en el contexto actual de un aumento en el consumo de *cannabis* y su legalización.[14]

Esencialmente, la DM2 se caracteriza por un complejo deterioro metabólico, la disfunción de las células beta es uno de los mecanismos más difíciles de abordar en su patogénesis. La identificación de factores reversibles que contribuyen al fallo de las células beta podría proporcionar una base para un enfoque de tratamiento personalizado y temprano de la DM2. Este enfoque se relaciona con la importancia de la presente investigación predictiva con pacientes ecuatorianos, ya que resalta la necesidad de identificar y comprender los factores de riesgo reversibles para desarrollar estrategias de prevención y tratamiento más efectivas de la DM2 en una población hospitalaria ecuatoriana.[15]

Predecir la diabetes con la mayor precisión posible es una necesidad imperiosa de las sociedades contemporáneas. Hay muchas características que contribuyen en su predicción, pero es difícil identificarlas y utilizarlas para la

detección en una etapa temprana.[16]

En última instancia, se puede argumentar que existió una convergencia interesante entre la regresión logística binaria que se utilizó en este estudio y herramientas emergentes como la lógica difusa y la Neutrosofía.[17,18,19] Estas disciplinas están experimentando avances significativos en la gestión de la incertidumbre en diversas aplicaciones, incluidas las predicciones médicas, sin embargo, en el contexto de este estudio, es importante señalar que aún no se ha establecido una vinculación sólida y adecuada entre estas herramientas. Esto representa una oportunidad para futuras investigaciones en el campo médico, ya que la integración de técnicas avanzadas de manejo de la incertidumbre podría mejorar, mucho más, la precisión de las predicciones, especialmente en el contexto de la determinación de la existencia de DM2. El desarrollo de enfoques más sólidos y multidisciplinarios podría conducir a avances significativos en la atención médica y la toma de decisiones clínicas.

En el estudio realizado se constató que la hipertensión arterial, los antecedentes familiares positivos de diabetes mellitus tipo 2 y el consumo de alcohol fueron factores que influyeron significativamente en la probabilidad de desarrollar diabetes mellitus tipo 2 en la población analizada. Estos resultados pueden ser extrapolables y aplicarse en la evaluación integral de los pacientes con diagnóstico probable de diabetes.

Conflicto de intereses

Los autores declaran la no existencia de conflictos de intereses relacionados con el estudio.

Contribuciones de autoría

1. Conceptualización: Olivia Altamirano Guerrero, Ronelsys Martínez, Jhonny Alejandro Rodríguez Gutiérrez.

2. Curación de datos: Olivia Altamirano Guerrero, Ronelsys Martínez, Jhonny Alejandro Rodríguez Gutiérrez.

3. Análisis formal: Olivia Altamirano Guerrero, Ronelsys Martínez Martínez, Jhonny Alejandro Rodríguez Gutiérrez.

4. Adquisición de fondos: Esta investigación no contó con la adquisición de fondos.

5. Investigación: Olivia Altamirano Guerrero, Ronelsys Martínez Martínez, Jhonny Alejandro Rodríguez Gutiérrez.

6. Metodología: Olivia Altamirano Guerrero.

7. Administración del proyecto: Olivia Altamirano Guerrero.

8. Recursos: Olivia Altamirano Guerrero, Ronelsys Martínez Martínez, Jhonny Alejandro Rodríguez Gutiérrez.

9. Software: Olivia Altamirano Guerrero, Ronelsys Martínez Martínez, Jhonny Alejandro Rodríguez Gutiérrez.

10. Supervisión: Olivia Altamirano Guerrero

11. Validación: Ronelsys Martínez Martínez.

12. Visualización: Jhonny Alejandro Rodríguez Gutiérrez.

13. Redacción del borrador original: Olivia Altamirano Guerrero, Ronelsys Martínez Martínez, Jhonny Alejandro Rodríguez Gutiérrez.

14. Redacción, revisión y edición: Olivia Altamirano Guerrero, Ronelsys Martínez Martínez, Jhonny Alejandro Rodríguez Gutiérrez.

REFERENCIAS BIBLIOGRÁFICAS

1. Qin Y, Wu J, Xiao W, Wang K, Huang A, Liu B, et al. Machine Learning Models for Data-Driven Prediction of Diabetes by Lifestyle Type. Int J Environ Res Public Health. 2022 ; 19 (22): 15027.

2. Rajendra P, Latifi S. Prediction of diabetes using logistic regression and ensemble techniques. Computer Methods Programs Biomedicine Update. 2021 ; 1 (10): 32.

3. Kopitar L, Kocbek P, Cilar L. Early detection of type 2 diabetes mellitus using machine learning-based prediction models. Sci Rep. 2020 ; 11 (9): 81.

4. Antoniou M, Mateus C, Hollingsworth B, Titman A. A Systematic Review of Methodologies Used in Models of the Treatment of Diabetes Mellitus. Pharmacoeconomics. 2023 ; 40 (273): 1312-4.

5. Amine M, Ali ZC, El Alami EK, Ghazal HN, Bakkali BA, Hamdi S. FTO gene variants (rs9939609, rs8050136 and rs17817449) and type 2 diabetes mellitus risk: A Meta-Analysis. Gene. 2023 ; 8 (8): 147791.

6. Lai H, Huang H, Keshavjee K. Predictive models for diabetes mellitus using machine learning techniques. BMC Endocr Disord. 2019 ; 19 (101): 2-9.

7. Sawane HB, Shetiya SH. Smokeless Tobacco as a Risk Factor for Type 2 Diabetes Mellitus in South East Asia Region: Systematic Review and Meta-Analysis. Indian J Community Med. 2023 ; 48 (4): 579-87.

8. Huttunen M, Raben A, Adam T, Macdonald I, Taylor MA, Stratton G, et al. Socio-economic factors, mood, primary care utilization, and quality of life as predictors of intervention cessation and chronic stress in a type 2 diabetes prevention intervention (PREVIEW Study). BMC Public Health. 2023 ; 23 (1): 1666.

9. Li X, Zeng J, Chen B, Yan Q, Cui Y, Xu W, et al. Daily higher tea consumption is associated with a reduced risk of type 2 diabetes: A cohort study and updated systematic review and meta-analysis. Nutr Res. 2023 ; 11 (8): 116-27.

10. Li R, Shokri F, Rincon AL, Rivadeneira F, Medina C, Ahmadizar F. Bi-Directional Interactions between Glucose-Lowering Medications and Gut Microbiome in Patients with Type 2 Diabetes Mellitus: A Systematic Review. Genes. 2023 ; 14 (8): 1572.

11. Wilson PWF, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino RB. Prediction of Incident Diabetes Mellitus in Middle-aged Adults: The Framingham Offspring Study. Arch Intern Med. 2007 ; 167 (10): 1068-74.

12. Regan JA, Mentz RJ, Nguyen M, Green JB, Truby LK, Ilkayeva O, et al. Mitochondrial metabolites predict adverse cardiovascular events in individuals with diabetes. JCI Insight. 2023 ; 8 (17): e168563.

13. Qin GQ, Chen L, Zheng J, Wu XM, Li Y, Yang K, et al. Effect of passive smoking exposure on risk of type 2 diabetes: a systematic review and meta-analysis of prospective cohort studies. Front Endocrinol. 2023 ; 14 (9): 53-4.

14. Mousavi SE, Tondro F, Sanaie S. Association between cannabis use and risk of diabetes mellitus type 2: A systematic review and meta-analysis. Phytother Res. 2023 ; 37 (11): 5092-8.

15. Stojanovic J, Andjelic M, Vuksanovic M, Marjanovic M, Jojic B, Stojanovic M, et al. The effects of early short-term insulin treatment vs. glimepiride on beta cell function in newly diagnosed type 2 diabetes with HbA1c above 9. Turk J Med Sci. 2023 ; 53 (2): 552-62.

16. Saxena R, Sharma SK, Gupta M, Sampada GC. A Comprehensive Review of Various Diabetic Prediction Models: A Literature Survey. J Healthc Eng. 2022 ; 81 (6): 97.

17. Centeno PA, Puertas Y, Escobar GS, Inca JD. Neutrosophic statistics methods applied to demonstrate the extracontractual liability of the state from the Administrative Organic Code. Neutrosophic Sets and Systems. 2019 ; 26 (1): 5.

18. Smarandache F, Estupiñán J, González E, Leyva MY, Batista N. Delphi method for evaluating scientific research proposals in a neutrosophic environment. Neutrosophic Sets and Systems. 2020 ; 34 (1): 26.

19. Guallpa LG, Paillacho HJ, Yaguar J, Aguilar M. Neutrosophic Iadov for measuring user satisfaction in a virtual learning environment at UNIANDES Puyo. Neutrosophic Sets and Systems. 2020 ; 34 (1): 16.