UNIVERSIDAD BICENTENARIA

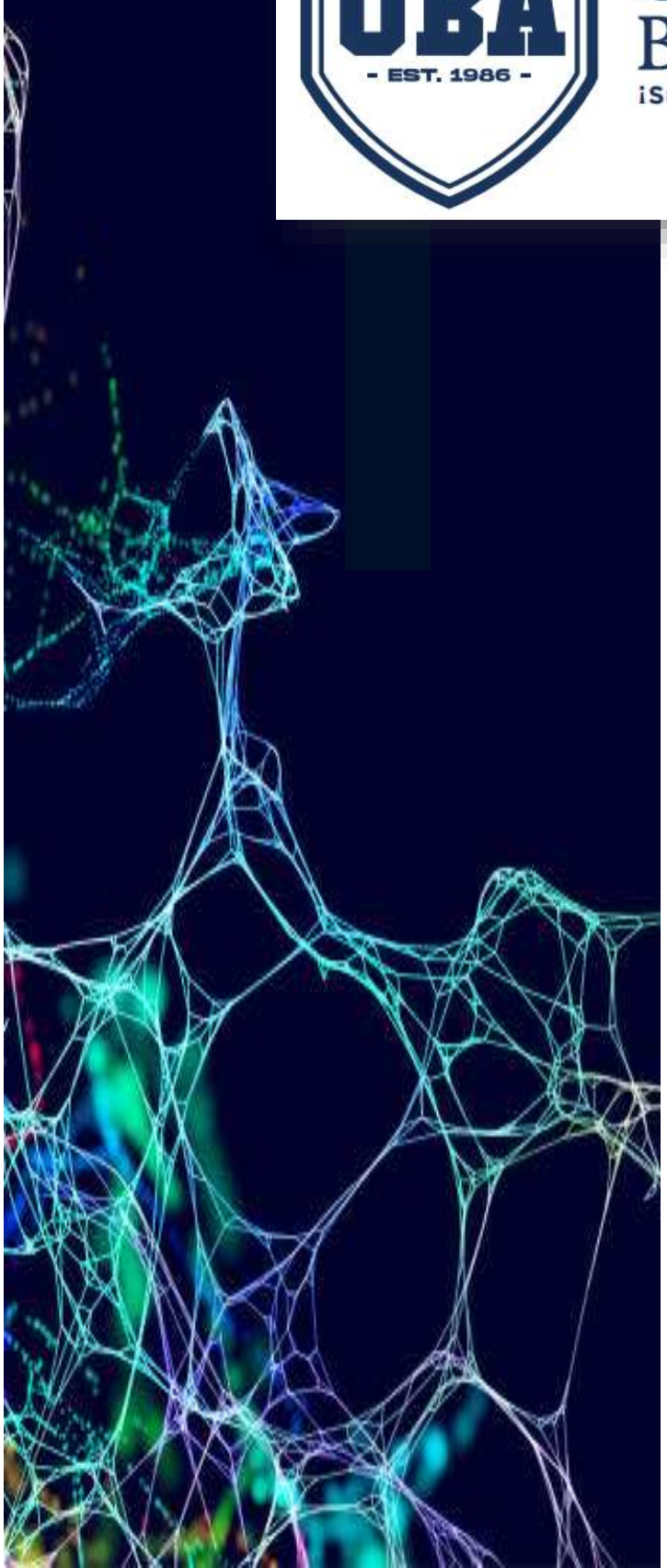
CLAVES PARA ENTENDER LA PREDIABETES

UN MODELO DE CLASIFICACIÓN INNOVADOR

¡SUEÑA, HAZ QUE SUCEDA!

Autores:
Margarita Eguiguren
Gabriel Rivadeneira

Serie Nodo ITC
Vol. 5 Nro. 1, 2025



UNIVERSIDAD BICENTENARIA

¡Sueña, haz que suceda!

AUTORIDADES

Dr. Basilio Sánchez
Presidente

Dr. Gustavo Sánchez
Rector

Dra. Mirian Regalado
Vicerrectora Académica

Dra. Zeyda Padilla
Vicerrectora Administrativa

Dra. Edilia Papa
Secretaria General

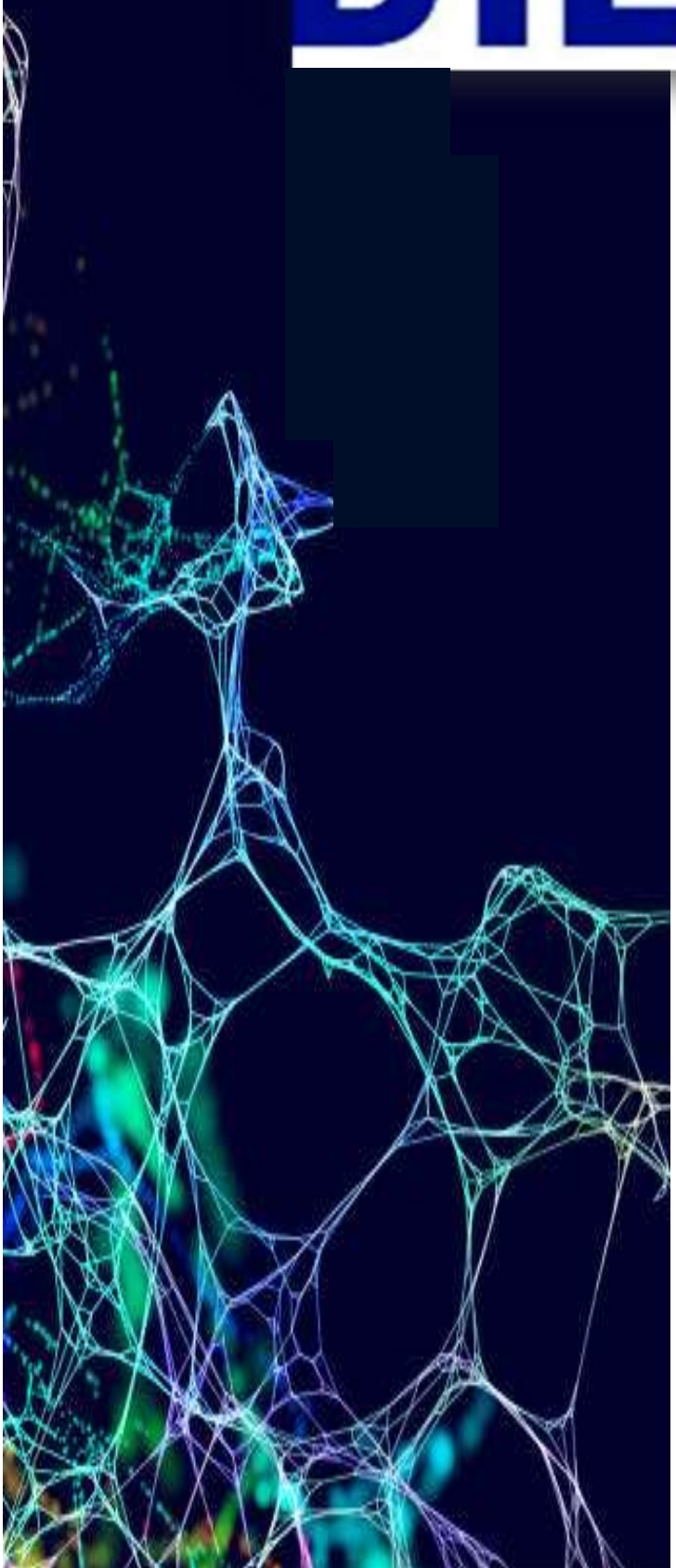
DECANATO DE INVESTIGACIÓN, EXTENSIÓN Y POSTGRADO

Abog. Wilmer Galíndez MSc.
Decano

Abog. María Teresa Ramírez MSc.
Directora de Postgrado

Dra. Maite Marrero
Directora de Investigación

Dra. Yesenia Centeno
Coordinadora del Fondo Editorial



DIEP | Decanato De Investigación, Extensión y Postgrado

®UNIVERSIDAD BICENTENARIA DE ARAGUA

Título:
CLAVES PARA ENTENDER LA PREDIABETES:
Un Modelo de Clasificación Innovador

Autores:
Margarita Eguiguren
Facultad de Formación General,
Escuela de Humanidades,
Universidad de Las Américas (UDLA)
Quito, Ecuador
margarita.eguiguren@udla.edu.ec

Gabriel Rivadeneira
Universidad Internacional del Ecuador (UIDE)
Quito, Ecuador
garivadeneirafu@uide.edu.ec

Fecha de Recepción: junio, 2024
Fecha de Aceptación: septiembre, 2024
Fecha de Publicación: marzo, 2025

Lugar: Turmero, Venezuela,
1ra edición

Depósito Legal: AR2025000051
ISBN: 978-980-6508-92-7

ISBN: 978-980-6508-92-7
9 789806 508927

SERIE NODO ITC
Volumen 5, Número 1 Año 2025

Es una publicación del Fondo Editorial de la Universidad Bicentenaria de Aragua (FE-UBA) en convenio con el Nodo Investigación, Transcomplejidad y Ciencia (NITC) de la Red Internacional InComplex. Tiene como propósito divulgar los avances de estudios, casos o experiencias de interés para el desarrollo de la investigación universitaria y el pensamiento, sistema complejo y ciencias de la complejidad; producto de la actividad de la comunidad académica. Es una publicación arbitrada por el sistema doble ciego, el cual asegura la confidencialidad del proceso, al mantener en reserva la identidad de los árbitros.

ÍNDICE DE CONTENIDOS

INTRODUCCIÓN

La diabetes constituye uno de los grandes problemas de salud pública a nivel mundial, con graves consecuencias personales, familiares y sociales. La prevalencia de esta enfermedad crónica degenerativa ha alcanzado cifras inimaginables. Dado que su aparición suele estar precedida por un estado de prediabetes, los esfuerzos para controlarla se centran principalmente en programas de prevención. En las últimas décadas, muchos investigadores dedican recursos y tiempo a desarrollar herramientas que permitan realizar un cribado oportuno tanto de la diabetes como de su fase previa.

Según la Organización Mundial de la Salud OMS (2016) “La diabetes es un importante problema de salud pública y una de las cuatro enfermedades no transmisibles (ENT), seleccionadas por los dirigentes mundiales para intervenir con carácter prioritario”. Es una enfermedad metabólica de carácter crónico que se origina por un mal funcionamiento del páncreas, especialmente de las células beta que son las encargadas de secretar la insulina, hormona que controla la cantidad de azúcar en la sangre. Esta condición conlleva graves consecuencias en el organismo “ataques cardíacos, accidentes cerebrovasculares, insuficiencia renal, amputación de piernas, pérdida de visión y daños neurológicos”.

Esta investigación forma parte de los tres trabajos investigativos del proyecto en desarrollo sobre prediabetes liderado por Wilmer Danilo Esparza, PhD en Ciencias y Técnicas de la Actividad Física y del Deporte por la Universidad de Las Américas

en Quito. Tiene como objetivo desarrollar una aplicación informática predictiva, esta herramienta se basa en un modelo estadístico que genera un índice de apoyo para determinar el riesgo de prediabetes en la población, utilizando datos obtenidos entre 2011 y 2016 de la Encuesta Nacional de Salud y Nutrición de Estados Unidos (NHANES, por sus siglas en inglés). La propuesta se fundamenta en los principios de la ciencia de datos, abarcando todas sus etapas: recolección y procesamiento de datos para extraer información relevante, limpieza, exploración, modelado estadístico y visualización de resultados, con el propósito de generar conocimiento aplicable al campo de la salud.

El Objetivo General fue generar un índice de riesgo de prediabetes que permita brindar alertas tempranas, a usuarios que deseen conocer de manera rápida cuan urgente es realizar un análisis de su metabolismo para prevenir la prediabetes. Los objetivos específicos fueron:

- Identificar de manera preventiva los factores de riesgo de la prediabetes.
- Ofrecer a los usuarios que cuenten con sus exámenes de laboratorio, valorarse con la herramienta propuesta.
- Incentivar a los usuarios a acudir a servicios de salud preventiva basados en el valor ponderado por el índice de prediabetes alcanzado.
- Generar un manual de usuario que permita aplicar la metodología en poblaciones que cuenten con características similares.

El desarrollo e implementación del índice único de riesgo de prediabetes representan un avance significativo en la prevención de esta condición metabólica. Mediante la identificación temprana de factores de riesgo, se establece una base sólida para alertar a los individuos sobre su posible predisposición. Además, la herramienta propuesta brinda a quienes ya cuentan con análisis de laboratorio la posibilidad de obtener una valoración rápida y personalizada de su estado metabólico. Este acceso directo a una evaluación de riesgo, facilitado por un índice ponderado, cumple el objetivo de ofrecer una alerta temprana y concreta sobre la necesidad de profundizar en el análisis de su salud.

La creación de este índice no solo proporciona una herramienta de autoevaluación, sino que también busca impactar positivamente en la salud pública. Al incentivar a los usuarios —a partir del valor obtenido— a buscar servicios de salud preventiva, se fomenta una cultura de detección temprana y acción proactiva contra la prediabetes. La elaboración de un manual de usuario robusto, diseñado para aplicar la metodología en poblaciones con características similares, asegura la escalabilidad y replicabilidad de esta propuesta, permitiendo que un mayor número de personas y comunidades se beneficien de esta herramienta de prevención y alerta temprana.

Para facilitar la comprensión del trabajo desarrollado, a continuación, se describe la estructura de esta investigación. Cada capítulo aborda aspectos específicos que, en conjunto, permiten alcanzar el objetivo planteado y ofrecer una contribución significativa en el ámbito de la detección temprana de la prediabetes.

En el capítulo I se muestra la situación actual del contexto, avances y perspectivas de la diabetes. Además, se presenta el estado del arte, con un recorrido por la literatura científica que documenta los trabajos de investigación existentes a nivel mundial relacionados con la detección precoz de la diabetes y la prediabetes, mediante el uso de modelos predictivos, modelos clasificatorios y técnicas de *machine learning*. En el capítulo II, se describe la metodología del trabajo, la selección del *dataset* para el estudio, la captura de la información, almacenamiento de la información, procesado de datos, análisis estadístico y el diseño de una visualización en HTML.

El capítulo III, muestra la construcción conceptual, datos, análisis y herramientas. El capítulo muestra la definición de conceptos, los contenidos implicados, descripción de la bases NHANES. Además, el desarrollo de la contribución donde se describen las tecnologías utilizadas, como, por ejemplo, el lenguaje R, el sistema SQL, el paquete estadístico SPSS, el lenguaje HTML, entre otras. También, se explica la técnica estadística multivariante seleccionada conocida como análisis factorial. En el capítulo IV, se explica el proceso de

desarrollo de la metodología aplicada al estudio de la prediabetes, con la generación del modelo, la variables seleccionadas, la parametrización y los criterios para extracción de factores que permitieron establecer la fórmula para el cálculo del índice o *score* propuesto.

Finalmente se presentan las conclusiones con un resumen de la contribución, la relevancia y alcance de la propuesta, recomendaciones y una propuesta de trabajo futuro basado en los resultados de la presente investigación. Y la referencias bibliográficas, necesarias para respaldar el carácter científico de la investigación.

CAPÍTULO I

Panorama de la Diabetes. Progresos y Horizontes

Según los rigurosos cálculos de la Federación Internacional de Diabetes (FID), detallados en la trascendental octava edición de su Atlas de la Diabetes, la prevalencia global de esta condición metabólica alcanza cifras alarmantes. En el momento de la publicación, se estimaba que 425 millones de individuos adultos en todo el mundo vivían con diabetes, lo que representaba un significativo 8.8% de la población adulta mundial. Esta cifra no solo subraya la magnitud del desafío sanitario actual, sino que también pone de manifiesto la necesidad urgente de estrategias efectivas de prevención y manejo.

Además, la FID identificó una población aún mayor, estimada en más de 352 millones de personas, que padece intolerancia a la glucosa. Esta condición, conocida como prediabetes, se caracteriza por niveles de glucosa en sangre más altos de lo normal, pero que no son lo suficientemente elevados como para ser diagnosticada como diabetes tipo 2. Sin embargo, es crucial destacar que estas personas se encuentran en un elevado riesgo de desarrollar la enfermedad en el futuro, convirtiéndose en un grupo de intervención prioritaria para las iniciativas de salud pública.

Las proyecciones para el año 2045 son aún más sombrías, anticipando un incremento drástico en el número de personas con diabetes, superando los 693 millones. Este aumento previsto no solo representa una carga creciente para los sistemas de salud a nivel global, sino que también subraya la imperiosa necesidad de intensificar los esfuerzos en investigación, prevención y educación para mitigar esta creciente epidemia y sus devastadoras consecuencias individuales y socioeconómicas.

1. Situación actual del contexto, avances y perspectivas futuras sobre la diabetes

En la actualidad, la FID considera a la diabetes una epidemia que crece de forma acelerada y será una verdadera amenaza para el desarrollo mundial futuro (FID, 2017). El gasto mundial en atención sanitaria por diabetes y sus complicaciones

ascendió en el año 2017 a 727.000 millones de dólares, como se muestra en la figura 1. Según el presidente del Comité del Atlas de Diabetes de la FID, “todavía se necesitan más investigaciones multidimensionales y multisectoriales para fortalecer la base de datos y reunir más conocimientos que sirvan de base de los métodos y programas para combatir la epidemia de diabetes” (FID, 2017, p.7).

Figura 1 *Evolución del gasto sanitario mundial por diabetes*
(Statista, 2018)

Como se puede apreciar en la figura 2, las cifras de diabetes en las diferentes zonas geográficas del mundo van en continuo incremento. Con pronósticos para el año 2045, en África la extraordinaria cifra de 156% de aumento. Según la OMS y la Organización Panamericana de la Salud (OPS), “el avance de la diabetes puede detenerse a través de una combinación de políticas fiscales, legislación, cambios en el medio ambiente y sensibilización a la población para modificar los factores de riesgo, entre ellos la obesidad y el sedentarismo” (OMS/OPS, 2017). El punto de partida es un diagnóstico precoz: “cuanto más tiempo se tarda en diagnosticar la diabetes, peores pueden ser las consecuencias para la salud” (OMS, 2016).

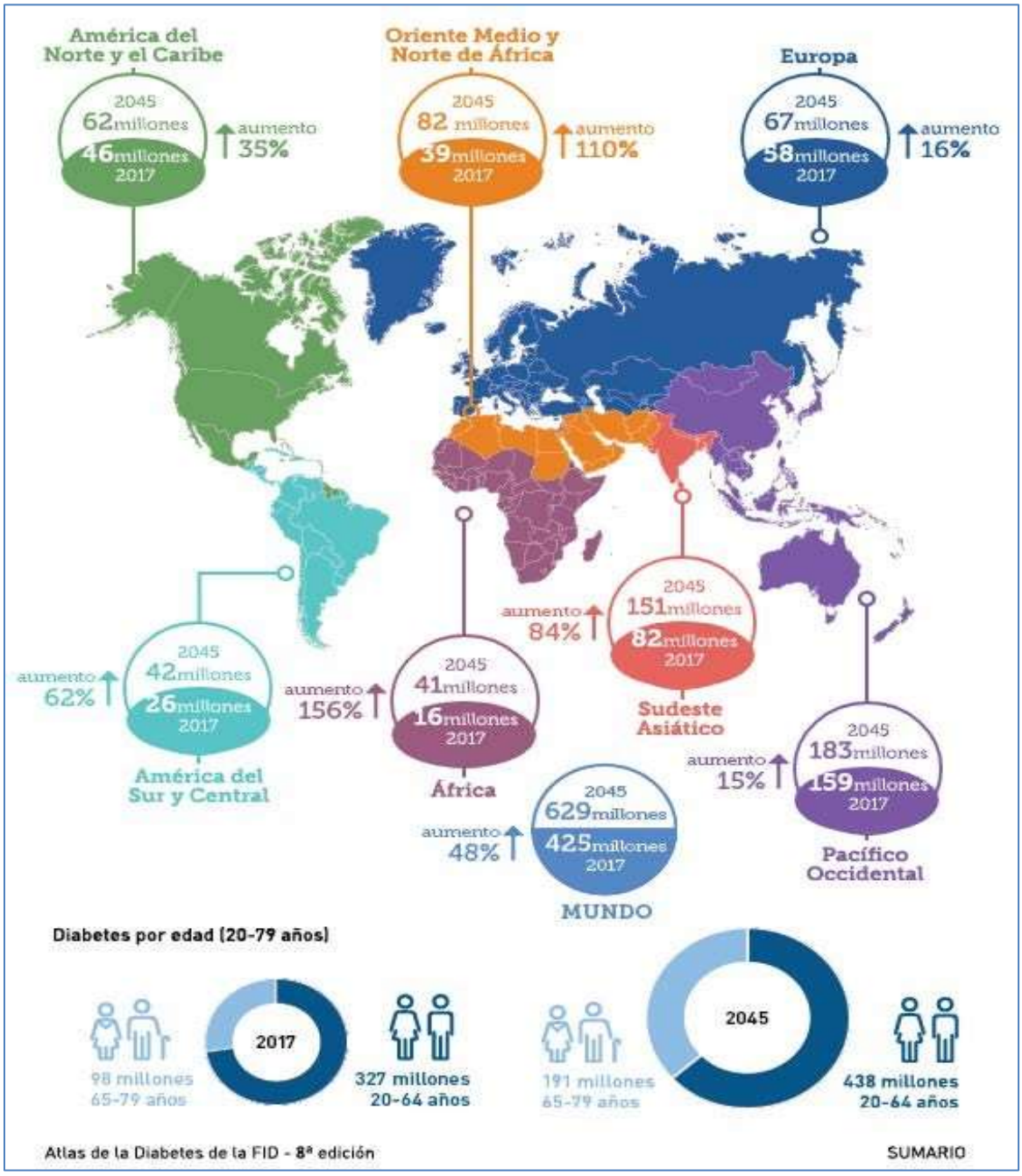


Figura 2 *Prevalencia de diabetes a nivel mundial*
(FID, 2017).

Si se considera que la prevención es el camino más adecuado para bajar las elevadas cifras expuestas anteriormente, es preciso hablar de la prediabetes, también llamada «hiperglucemia intermedia» o «disglicemia», una condición de

salud que está un paso antes que la diabetes Mellitus o diabetes tipo II. La American Diabetes Association (ADA) define a la prediabetes de la siguiente manera:

> Es un estado que precede al diagnóstico de diabetes tipo 2. Esta condición es común, está en aumento epidemiológico y se caracteriza por elevación en la concentración de glucosa en sangre más allá de los niveles normales sin alcanzar los valores diagnósticos de diabetes (ADA, 2018).

Es decir, una “zona gris” entre los niveles normales de glucosa y la diabetes. Algunos estudios han demostrado que se puede prevenir el paso de prediabetes a diabetes tipo II en un 58% de los pacientes, interviniendo de manera oportuna en su estilo de vida (Asociación Latinoamericana de Diabetes ALAD, 2009). La ADA y la ALAD, consideran los siguientes aspectos como factores de riesgo para la detección de prediabetes y diabetes en la población mundial (ADA, 2018; ALAD, 2017):

- Características demográficas como: Edad, sexo, etnia
- Antecedentes genéticos
- Niveles de glucosa en plasma
- Índice de masa corporal (IMC)
- Diámetro sagital abdominal
- Actividad física
- Antecedentes de hipertensión
- Resistencia a la insulina
- Otras pruebas de laboratorio (triglicéridos, colesterol, transaminasas, etc.)
- Otros factores antropométricos (ovarios poliquísticos, diabetes gestacional)

Por la magnitud e importancia del tema, se plantea esta obra con la creación de un modelo de interdependencia de los factores de riesgo que inciden en el diagnóstico de la prediabetes. Generando una propuesta que coadyuve a la detección precoz de la prediabetes, contribuyendo así con los programas de prevención de salud pública. Una correcta identificación de los factores de riesgo

que desembocan en una enfermedad degenerativa crónica como es la diabetes, puede ayudar significativamente a su prevención, mediante la determinación oportuna de niveles de riesgo de contraer la enfermedad.

En la actualidad, se aplican muchas técnicas de análisis de datos y *machine learning*, basados en los procesos de *data science* o ciencia de datos en el campo de la salud. Especialmente diseñadas para el desarrollo de nuevos modelos predictivos, caracterización de fenotipos para identificar grupos de pacientes con mayor riesgo de desarrollar una patología, optimización de técnicas de *screening* y personalización de la atención médica. En la figura 3, se aprecia una panorámica general del procesado de datos, desde su recolección, limpieza, modelamiento, análisis, visualización, llegando finalmente a extraer conocimiento válido para la toma de decisión.

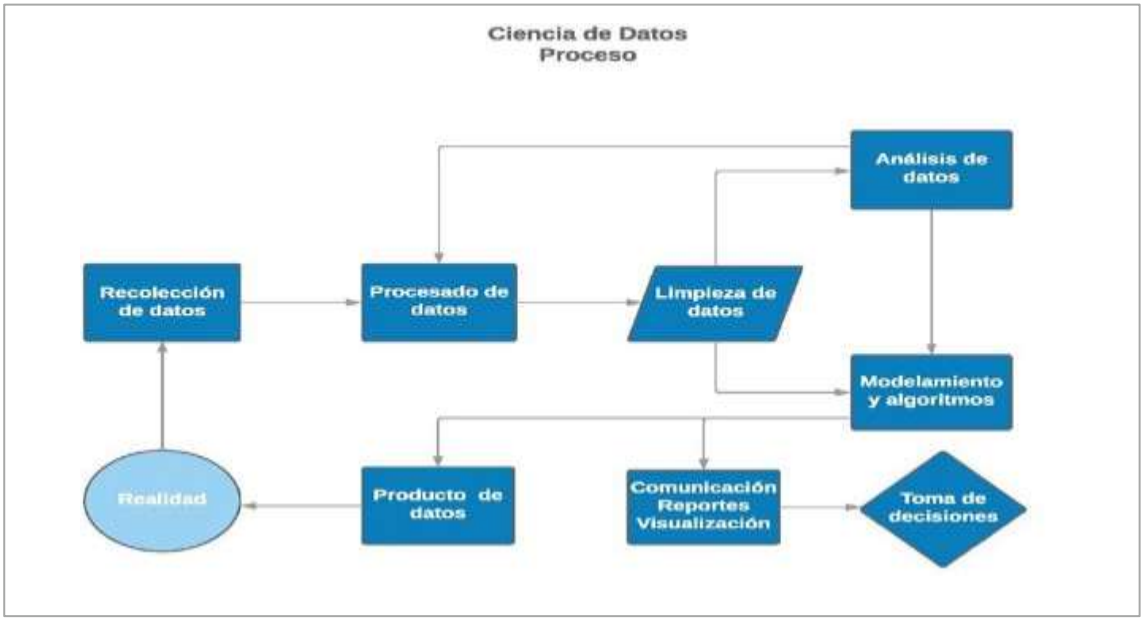


Figura 3 *Proceso de ciencia de datos*

1.1. Estado del arte sobre la prediabetes

La revisión de la literatura relacionada al tema en cuestión ha permitido tener

una mirada amplia del problema y concluir que, la mejor manera de contribuir a los programas de prevención de salud es aportar con herramientas que brinden la posibilidad de realizar intervenciones oportunas en la población. El “inconveniente” observado en las herramientas no invasivas que utilizan variables antropométricas exclusivamente radica que, en general, estos métodos tienen alta sensibilidad (68.9% a 98.5%) para detectar prediabetes, pero baja especificidad (6.7% a 44.5%). Es decir, los métodos no invasivos identifican correctamente a las personas con prediabetes, pero, a su vez, incluyen en este grupo a personas que no padecen la enfermedad (Vanderwood et al., 2014).

El FINDRISC (Finnish Diabetes Risk Score) es una herramienta diseñada para evaluar el riesgo de desarrollar diabetes tipo 2. Este índice se fundamenta en una serie de factores de riesgo, tales como la edad, el índice de masa corporal (IMC), la circunferencia de la cintura, el nivel de actividad física, la historia familiar de diabetes, los niveles de glucosa en sangre y el uso de medicación para la hipertensión. Estos componentes se combinan para generar un puntaje que ayuda a identificar a las personas con mayor riesgo de desarrollar la enfermedad, permitiendo intervenciones preventivas tempranas.

Sin embargo, según el consenso español sobre prediabetes, uno de los principales inconvenientes del uso del FINDRISC en el contexto de atención médica en España es la dificultad para obtener algunos de los datos requeridos, como el IMC y la medición del perímetro de cintura. Muchos pacientes no están familiarizados con su propio IMC, y la medición del perímetro de cintura no es una práctica estándar en todos los centros de salud, lo que limita la aplicabilidad de este índice. Esta situación destaca una barrera importante para la implementación efectiva de herramientas predictivas en la práctica clínica diaria, ya que depende de parámetros que no siempre están disponibles o son fáciles de obtener (Mata-Cases et al., 2015).

Por otro lado, las investigaciones que emplean modelos de regresión logística para predecir la diabetes suelen basarse en estudios de campo que siguen a los pacientes durante períodos de tiempo específicos, en su mayoría con diagnósticos

ya establecidos. Estos estudios, aunque útiles para realizar predicciones sobre el desarrollo de la enfermedad a lo largo de los años, también presentan limitaciones, ya que trabajan con datos retrospectivos y series de tiempo que no siempre reflejan la detección temprana de la condición en pacientes sin diagnóstico previo. Esto pone en evidencia que, aunque los modelos predictivos basados en regresión logística pueden ser útiles para monitorear la progresión de la enfermedad, la integración de herramientas como el FINDRISC para la evaluación temprana del riesgo aún enfrenta desafíos prácticos en su implementación generalizada (Mata-Cases et al., 2015).

Otro método ampliamente utilizado en la predicción de enfermedades crónicas como la prediabetes es la regresión logística. Se trata de una técnica estadística que permite modelar la relación entre una variable dependiente categórica (por ejemplo, presencia o ausencia de enfermedad) y una o varias variables independientes, que suelen ser factores de riesgo clínicos o demográficos. A diferencia de los modelos lineales tradicionales, la regresión logística se basa en la estimación de probabilidades, proporcionando una herramienta sólida para predecir la probabilidad de que un individuo pertenezca a una determinada categoría (por ejemplo, prediabético o no prediabético) en función de sus características (Hosmer et al., 2013).

Uno de los puntos fuertes de esta técnica es su capacidad para interpretar el peso de cada variable predictora, lo cual resulta útil no solo para el diagnóstico, sino también para la identificación de los factores que más influyen en el desarrollo de la enfermedad. Sin embargo, la regresión logística presenta limitaciones cuando se trata de datos complejos o no lineales, ya que su capacidad de adaptación a relaciones más sofisticadas entre variables es reducida (Tanaka et al., 2013).

En este contexto, los modelos de aprendizaje supervisado como los árboles de decisión, los bosques aleatorios (random forest), y las máquinas de soporte vectorial (SVM), han comenzado a ganar terreno como herramientas complementarias o alternativas. Estos modelos no solo son capaces de manejar

grandes volúmenes de datos con múltiples variables, sino que también permiten captar relaciones no lineales y detectar interacciones complejas entre los factores de riesgo. Los árboles de decisión, en particular, han sido ampliamente utilizados por su facilidad interpretativa, ya que simulan un proceso lógico de toma de decisiones que resulta comprensible para los profesionales clínicos (Kavakiotis et al., 2017).

No obstante, una de sus limitaciones más notorias es que, si bien son eficaces para clasificar a los pacientes en categorías como sano, prediabético o diabético, tienden a centrarse más en la precisión de la clasificación que en el análisis profundo del impacto individual de cada variable. Esta característica, aunque útil en términos operativos, restringe su utilidad cuando el objetivo es diseñar intervenciones personalizadas basadas en el peso relativo de cada factor de riesgo (Alghamdi et al., 2017).

Asimismo, los modelos de aprendizaje no supervisado, como los algoritmos de clustering (por ejemplo, K-means o DBSCAN), han sido empleados en el análisis de grandes bases de datos de pacientes para descubrir patrones subyacentes y segmentar a la población en grupos homogéneos. Esta técnica puede resultar valiosa para identificar subconjuntos de individuos con perfiles metabólicos similares, incluso antes de que presenten síntomas evidentes de enfermedad. Sin embargo, su principal desventaja radica en la falta de una variable objetivo-conocida, lo que impide evaluar de manera directa el riesgo individual de padecer prediabetes (Sajida y Shoba, 2020).

Además, estos modelos no proporcionan una interpretación clara de qué factores explican la pertenencia de un paciente a un determinado clúster, ni ofrecen recomendaciones explícitas sobre cómo modificar esas variables para reducir el riesgo. En consecuencia, aunque el clustering es útil para la exploración y segmentación de datos, su valor predictivo y su aplicabilidad clínica directa siguen siendo limitados frente a los modelos supervisados o basados en inferencias estadísticas (Reddy et al., 2020).

En conjunto, estas técnicas reflejan los avances actuales en la aplicación de la inteligencia artificial y el análisis de datos en el ámbito de la salud. No obstante, aún persisten desafíos importantes para traducir estos modelos en herramientas prácticas de uso cotidiano. Entre ellos, destacan la necesidad de bases de datos de alta calidad, la interpretación ética y transparente de los resultados, y la validación externa en poblaciones diversas. Superar estos retos será fundamental para que la predicción y prevención de la prediabetes pueda evolucionar hacia un enfoque más personalizado, efectivo y accesible.

El objetivo principal de esta propuesta es desarrollar una aplicación que calcule un índice individual para alertar tempranamente sobre el riesgo de desarrollar prediabetes, utilizando los datos proporcionados por la NHANES. Para ello, se investigan distintos modelos estadísticos y de inteligencia artificial con el fin de seleccionar el más adecuado según los datos disponibles. Es importante señalar que la encuesta NHANES no es un estudio longitudinal, ya que no sigue una serie de tiempo, sino que se trata de un estudio transversal realizado en cada período específico. Además, NHANES no proporciona diagnósticos definitivos basados en los resultados de las pruebas de laboratorio; en cambio, obtiene la información a partir de las percepciones de los encuestados.

Cuando se utilizan datos transversales, como es el caso de la encuesta NHANES, no se tiene información sobre cómo evolucionan las variables a lo largo del tiempo. En otras palabras, no se pueden seguir a los individuos a lo largo de diferentes puntos en el tiempo para observar cómo cambian sus condiciones de salud. Esto limita el uso de modelos estadísticos que requieren series temporales o un seguimiento de los mismos sujetos a lo largo del tiempo, como las regresiones logísticas o modelos de predicción basados en series temporales, que necesitan datos secuenciales para evaluar cómo un conjunto de factores va influyendo en la evolución de una enfermedad o condición.

Dada la limitación de contar con datos transversales y la ausencia de diagnósticos certeros de prediabetes, se opta por aplicar un análisis factorial. Este modelo estadístico es ideal porque permite integrar todas las variables

intercorrelacionadas que influyen en el desarrollo de la prediabetes. A través de este análisis, se busca reducir el número de variables sin perder información crucial, de manera que se pueda explicar de forma más clara el fenómeno de la prediabetes. Por tanto, la elección del análisis factorial está motivada por la naturaleza de los datos transversales de la encuesta, ya que permite identificar las relaciones entre diferentes factores de riesgo, sin necesidad de seguir a los individuos a lo largo del tiempo. Este enfoque es útil para obtener una visión global y simplificada de las variables involucradas en la prediabetes en un punto específico del tiempo.

A diferencia de otros modelos, como los de regresión, el análisis factorial no asigna roles específicos a las variables, es decir, no existe una variable dependiente ni variables independientes. En cambio, todas las variables se analizan conjuntamente, sin una dependencia conceptual predefinida, lo que otorga una gran versatilidad al modelo. El objetivo final, es conseguir una ecuación que defina los pesos de las variables dentro del fenómeno para calcular un score de riesgo de desarrollar prediabetes. Crear un instrumento amigable para el usuario para conocer este valor con su respectiva interpretación y proveer recomendaciones que le ayuden a tomar acciones preventivas antes que correctiva.

Desde hace un cuarto de siglo, la Organización Mundial de la Salud (1994) ya advirtió que la detección oportunista de la diabetes es el instrumento más utilizado y eficaz para la determinación precoz de la prediabetes y diabetes tipo II, por la sencillez de su aplicación y por los bajos costos que representa para el usuario y los sistemas de salud pública. La detección oportunista se refiere a realizar un cribado mediante un sencillo cuestionario con la finalidad de identificar si están presentes en la población uno o varios factores de riesgo (antecedentes familiares, obesidad, sedentarismo, hábitos alimenticios, etc.) que permita decidir la conveniencia de la aplicación de pruebas de laboratorio.

De acuerdo con la Sociedad Española de Diabetes (2015), en su artículo “Consenso sobre la detección y el manejo de la prediabetes. Grupo de Trabajo de

Consensos y Guías Clínicas de la Sociedad Española de Diabetes" (Mata-Cases et al., 2015), existen varias estrategias para el cribado de diabetes. Como el cribado oportunista, que se muestra en la figura 4, que ayuda a determinar la prediabetes o la diabetes no diagnosticada en poblaciones de riesgo, mediante la utilización de "reglas de predicción clínicas", con la revisión de historias clínicas o la aplicación de "escalas de riesgo o cuestionarios", previos a la realización de exámenes de laboratorio confirmatorios.

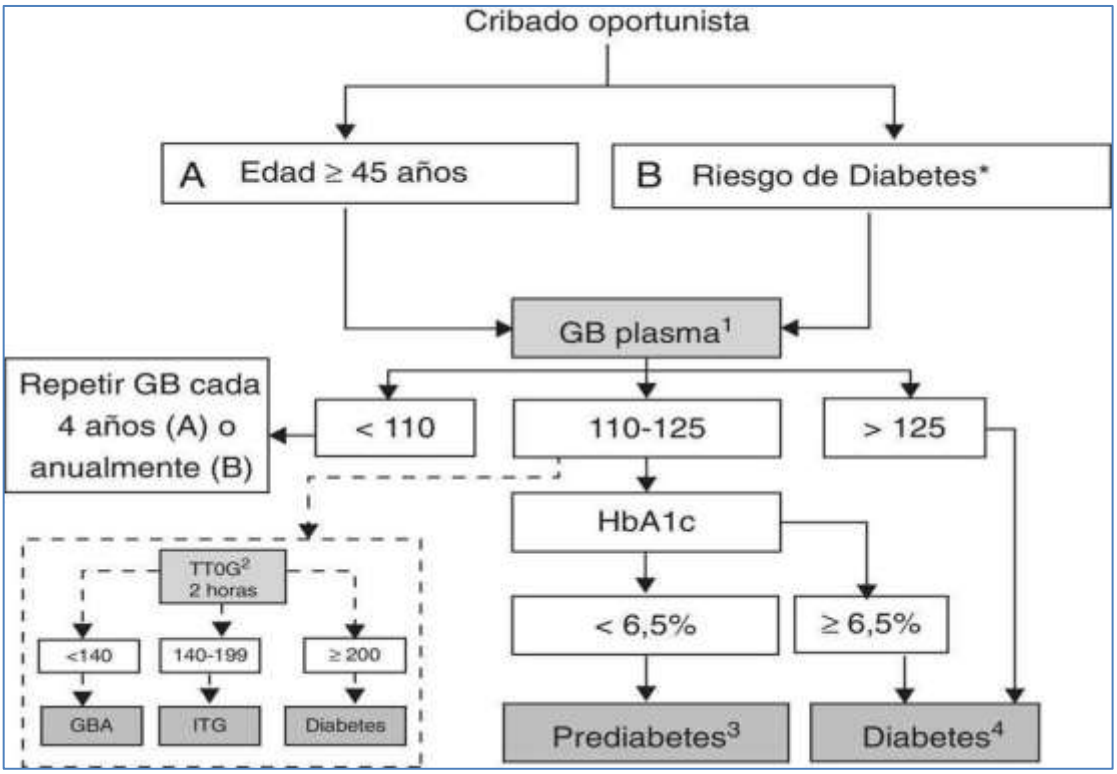


Figura 4 *Algoritmo de detección de prediabetes y diabetes* (Mata-Cases et al., 2015).

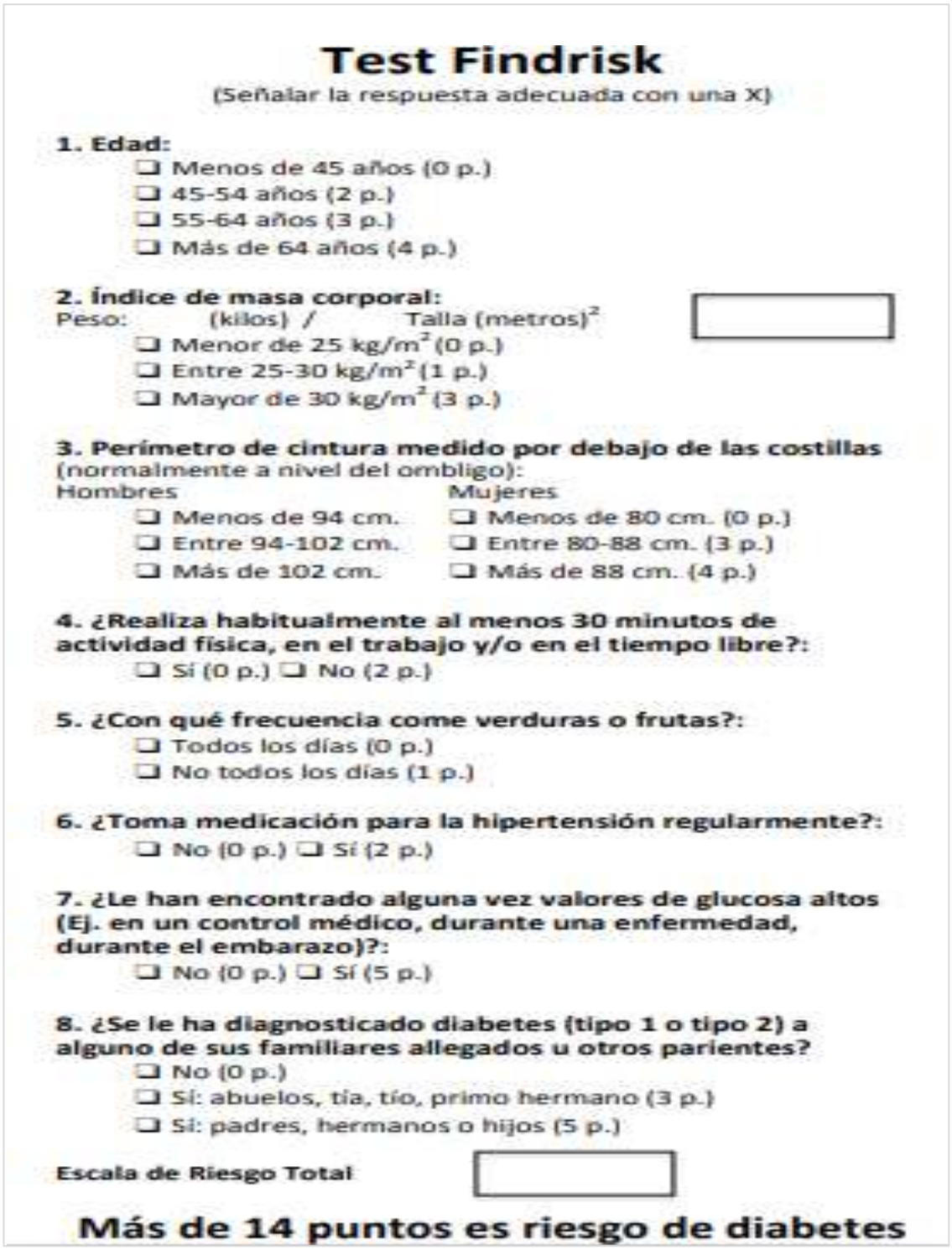
Actualmente, existen numerosos y diversos métodos que permiten identificar de manera temprana la presencia de factores de riesgo asociados a la diabetes tipo 2, siendo uno de los más ampliamente utilizados el cuestionario Finnish Diabetes Risk Score (FINDRISC) (Lindström y Tuomilehto, 2003). Esta prueba, de carácter no invasivo, se compone de ocho preguntas simples que abarcan aspectos clave del estilo de vida y características personales del individuo, como

la edad, el índice de masa corporal (IMC), la circunferencia de la cintura, la actividad física habitual, los hábitos alimenticios, la historia familiar de diabetes y antecedentes de hipertensión o alteraciones en la glucemia. Fue desarrollado en el marco del Programa Nacional de Prevención de la Diabetes Tipo 2 de Finlandia y se aplica cada cinco años a una muestra aleatoria de la población como parte de una estrategia nacional de vigilancia epidemiológica (Salinero-Fort et al., 2016).

La utilidad del FINDRISC radica en su capacidad para predecir con alta sensibilidad el riesgo de desarrollar diabetes tipo 2 en los próximos diez años, permitiendo a los profesionales de salud identificar a personas aparentemente sanas, pero con una probabilidad significativa de desarrollar la enfermedad si no se adoptan cambios en el estilo de vida. Gracias a su formato accesible y la ausencia de procedimientos clínicos o de laboratorio, este cuestionario ha sido adoptado y adaptado por múltiples países, tanto en Europa como en América Latina, como herramienta de cribado poblacional (Costa et al., 2011).

No obstante, su aplicación masiva también ha puesto en evidencia ciertos desafíos, especialmente relacionados con la variabilidad en la recolección de datos antropométricos y la interpretación cultural de algunas preguntas. Por ejemplo, la autoevaluación del IMC o la medición precisa del perímetro de cintura puede representar una dificultad para los pacientes fuera del entorno clínico, lo que afecta la precisión del resultado final (Lindström et al., 2006).

A pesar de estas limitaciones, el FINDRISC sigue siendo una herramienta valiosa para orientar intervenciones preventivas tempranas y promover la concienciación sobre el riesgo metabólico entre la población general. Su simplicidad, bajo costo y validación en diversos contextos lo convierten en un instrumento esencial dentro de las estrategias de salud pública enfocadas en la reducción del impacto de la diabetes tipo 2 y sus complicaciones asociadas (Salinero-Fort et al., 2016). Mediante la figura 5 se muestra el cuestionario.



Test Findrisk

(Señalar la respuesta adecuada con una X)

1. Edad:

- ❑ Menos de 45 años (0 p.)
- ❑ 45-54 años (2 p.)
- ❑ 55-64 años (3 p.)
- ❑ Más de 64 años (4 p.)

2. Índice de masa corporal:

Peso: (kilos) / Talla (metros)^2

- ❑ Menor de 25 kg/m^2 (0 p.)
- ❑ Entre 25-30 kg/m^2 (1 p.)
- ❑ Mayor de 30 kg/m^2 (3 p.)

3. Perímetro de cintura medido por debajo de las costillas (normalmente a nivel del ombligo):

Hombres	Mujeres
❑ Menos de 94 cm.	❑ Menos de 80 cm. (0 p.)
❑ Entre 94-102 cm.	❑ Entre 80-88 cm. (3 p.)
❑ Más de 102 cm.	❑ Más de 88 cm. (4 p.)

4. ¿Realiza habitualmente al menos 30 minutos de actividad física, en el trabajo y/o en el tiempo libre?:

❑ Sí (0 p.) ❑ No (2 p.)

5. ¿Con qué frecuencia come verduras o frutas?:

- ❑ Todos los días (0 p.)
- ❑ No todos los días (1 p.)

6. ¿Toma medicación para la hipertensión regularmente?:

❑ No (0 p.) ❑ Sí (2 p.)

7. ¿Le han encontrado alguna vez valores de glucosa altos (Ej. en un control médico, durante una enfermedad, durante el embarazo)?:

❑ No (0 p.) ❑ Sí (5 p.)

8. ¿Se le ha diagnosticado diabetes (tipo 1 o tipo 2) a alguno de sus familiares allegados u otros parientes?

- ❑ No (0 p.)
- ❑ Sí: abuelos, tía, tío, primo hermano (3 p.)
- ❑ Sí: padres, hermanos o hijos (5 p.)

Escala de Riesgo Total

Más de 14 puntos es riesgo de diabetes

Figura 5 *Cuestionario FINDRISK*

Otra herramienta no invasiva que ha ganado relevancia en los últimos años para la detección temprana de prediabetes y casos de diabetes no diagnosticados es la

“Diabetes Risk Calculator”, propuesta por Heikes y colaboradores en 2008. Este instrumento fue diseñado como una alternativa sencilla y accesible para ser utilizada tanto en entornos clínicos como comunitarios, especialmente en poblaciones con limitado acceso a pruebas de laboratorio. La calculadora se basa en un cuestionario breve que recopila información clave relacionada con factores de riesgo reconocidos, como la edad del paciente, la circunferencia de la cintura, el historial de diabetes gestacional en mujeres, la estatura, la etnia, la presencia de hipertensión arterial, los antecedentes familiares de diabetes y la práctica regular de ejercicio físico (Heikes et al., 2008).

A diferencia de otros modelos, esta herramienta fue validada mediante estudios que analizaron su capacidad predictiva en diferentes poblaciones, mostrando una buena discriminación entre individuos con bajo y alto riesgo de desarrollar diabetes tipo 2. Su estructura permite calcular una probabilidad estimada de padecer la enfermedad, lo que la convierte en una opción útil tanto para tamizajes masivos como para la orientación clínica inicial. Una de sus ventajas es que no requiere conocimientos técnicos para su aplicación, lo cual facilita su implementación por parte de profesionales de la salud, promotores comunitarios o incluso por los propios pacientes a través de plataformas digitales (Heikes et al., 2008).

Sin embargo, como ocurre con la mayoría de las herramientas basadas en autoinforme, la exactitud del resultado puede verse afectada por errores en la percepción o el reporte de ciertas variables, como la medida de la cintura o la frecuencia del ejercicio. Aun así, la Diabetes Risk Calculator representa un aporte significativo al enfoque preventivo, al permitir intervenciones tempranas que pueden frenar o incluso revertir el desarrollo de la enfermedad mediante cambios en el estilo de vida, mucho antes de que se manifiesten alteraciones bioquímicas detectables por métodos invasivos (Heikes et al., 2008). Incluye preguntas sobre edad, circunferencia de la cintura, diabetes gestacional, talla, etnia, hipertensión, antecedentes familiares y ejercicio.

Los investigadores utilizan los datos de la tercera NHANES (1999-2004) para proponer dos modelos de predicción, utilizando análisis de regresión logística y árboles de decisión. La ecuación resultante de esta regresión logística asigna

pesos para cada variable escogida y está expresada de la siguiente manera: valor constante, −21.6343; edad en la entrevista (años), 0.0402; sexo, −0.5042; peso (kilogramos), −0.029; altura (centímetros), 0.0730; relación cintura-cadera, 5.3827; IMC (peso en kilogramos dividido por el cuadrado de altura en metros), 0.2947; dicho tiene presión arterial alta, 0–0.3449; y el padre tiene diabetes, 0.3981.

Un estudio comparativo del desempeño de tres modelos predictivos aplicados a una misma población, regresión logística, redes neuronales y árboles de decisión concluye que “el modelo de árbol de decisión (C5.0) tuvo la mejor precisión de clasificación, seguido del modelo de regresión logística y la ANN evidenció la precisión más baja”. En esta investigación de dos años de duración, las variables de entrada fueron factores de riesgo como género, edad, estado civil, nivel educativo, antecedentes familiares de diabetes, IMC, consumo de café, actividad física, duración del sueño, estrés laboral, consumo de pescado y preferencia por alimentos salados. El estudio incluyó 735 voluntarios que confirmaron que tenían diabetes o prediabetes y 752 voluntarios que por chequeo físico se confirmó que padecían la enfermedad (Meng et al., 2012).

Castrillón et al., (2017) proponen un sistema bayesiano para la predicción de la diabetes. El modelo está basado en los siguientes factores de riesgo: número de embarazos, presión sanguínea diastólica, espesor del pliegue cutáneo del tríceps e índice de masa corporal. Con estas variables el estudio arroja un 87.69% de aciertos; sin embargo, al incorporar la variable insulina en suero el sistema incrementa su desempeño al 98.46%.

Un score para detectar adultos con prediabetes y diabetes no diagnosticada elaborado en el año 2018 por un grupo de investigadores mexicanos planteó la necesidad de distinguir hombres de mujeres con el fin de establecer un score que tenga mejor desempeño por sexo. Este es un modelo no invasivo que utiliza las variables sexo, edad, antecedentes familiares, diámetro de la cintura, índice de masa corporal, hipertensión y sedentarismo (Rojas et al., (2018). En el 2015, varios investigadores utilizando los resultados de la encuesta NHANES del 2005 al 2010, propusieron un modelo de prueba simultánea, basado en la puntuación de riesgo

de diabetes FINDRISC y la medición de hemoglobina glicosilada en sangre HbA1c. Concluyendo que, este método es una herramienta práctica y válida en la detección de diabetes en la población norteamericana ya que mejoró la sensibilidad al 84.2% para la diabetes y 74.2% para la prediabetes (Zhang et al., 2015).

Peng et al., (2016), diseñan un modelo de puntaje simple para evaluar el estado de riesgo de prediabetes basado en factores antropométricos (género, edad, historia de hipertensión, antecedentes familiares, índice de masa corporal, presión diastólica) e incorporan pruebas de laboratorio que miden los niveles de triglicéridos en sangre. La metodología utilizada en este caso fue regresión logística binaria, el modelo propuesto es un piloto que utiliza datos no invasivos de la población estudiada y datos tomados a través de pruebas de laboratorio (método invasivo).

Como indican Appajigol et al., (2015) en su trabajo de investigación "*Performance of diabetes risk scores with or with out point of care blood glucose estimation*": "La sensibilidad y la especificidad de la herramienta de detección de T2DM se puede aumentar al incluir una prueba de glucosa en sangre en el punto de atención", parecería ser que los modelos que utilizan datos antropométricos e incluyen pruebas de laboratorio son modelos más robustos para calcular el riesgo de contraer la enfermedad.

Glumer et al., (2006) en su studio "Risk scores for type 2 diabetes can be applied in some populations but not all" concluyen:

> Este estudio ha demostrado que una herramienta de evaluación de riesgos desarrollada en una población caucásica se desempeña razonablemente bien en otras poblaciones caucásicas con una distribución similar de factores de riesgo, pero no en otras poblaciones de origen étnico diverso. La razón principal de la falta de transferibilidad del puntaje de riesgo es la diferencia en el impacto de especialmente el IMC y la edad en la prevalencia de diabetes no diagnosticada.

En el apartado de anexos de la investigación se agrega una tabla con un resumen comparativo de otras herramientas de evaluación de riesgo de diabetes

se agrega a este trabajo de investigación (Anexo I). Es importante mencionar la contribución que a futuro aportarán en este campo de la salud los tres trabajos investigativos del proyecto en desarrollo sobre prediabetes liderado por el PhD. Danilo Esparza, en la Universidad de Las Américas en Quito-Ecuador.

El primero de ellos, cuya autoría pertenece a Sophía Ortiz, se titula “Análisis evolutivo entre prediabetes y actividad física en el período 2007-2016 utilizando la base de datos NHANES” el cual pretende identificar la evolución de la enfermedad y cómo prevenirla a través de la actividad física. El segundo estudio, “Modelo de clasificación de las condiciones clínicas que componen la prediabetes”, desarrollado por Gabriel Rivadeneira, estudiante de la maestría en Big Data de la UNIR, ofrece un árbol de decisión que permite ubicar al paciente en un nivel específico de diagnóstico de prediabetes.

Y el presente trabajo, desarrolla un score único de riesgo de prediabetes con una herramienta interactiva que ofrece al usuario un pseudo diagnóstico con sus respectivas recomendaciones. En cualquier caso, la aplicación de alguna herramienta de las múltiples existentes de detección precoz o predicción de la prediabetes o diabetes, estará en función de la adaptabilidad al grupo humano estudiado (por sus características específicas), de la simplicidad de su aplicación e interpretabilidad de los resultados.

CAPÍTULO II

Desentrañando la prediabetes: La lente metodológica

El presente capítulo, se adentra en la compleja y multifacética realidad de la diabetes en la actualidad. Con el propósito de ofrecer una visión integral que abarque, tanto los significativos avances científicos y médicos alcanzados en la comprensión, diagnóstico y tratamiento de esta condición metabólica, como las persistentes problemáticas y los prometedores caminos que se vislumbran en el futuro cercano. A través de un análisis exhaustivo del contexto actual, se exploran las tendencias epidemiológicas, los últimos descubrimientos en la fisiopatología de la enfermedad, las innovaciones terapéuticas y las estrategias de prevención que están marcando la pauta en la lucha contra la diabetes a nivel global.

Al desglosar los "progresos", se examinan los hitos recientes en áreas como la insulinoterapia, los nuevos fármacos hipoglucemiantes, los sistemas de monitorización continua de glucosa y las tecnologías emergentes para la gestión de la enfermedad. Sin embargo, no se eluden los "horizontes", es decir, las áreas donde, aún existen desafíos importantes y la investigación futura promete ofrecer soluciones innovadoras. Esto incluye, la búsqueda de terapias más personalizadas, la comprensión profunda de las complicaciones a largo plazo, el desarrollo de estrategias preventivas más efectivas, así como el abordaje de los determinantes sociales y económicos, que influyen en la prevalencia y el manejo de la diabetes. En definitiva, este capítulo sienta las bases para una comprensión profunda del estado actual de la diabetes y las direcciones clave que moldearán su futuro.

La captura de la información, fase inicial crucial para la solidez de este estudio, se llevó a cabo mediante un esfuerzo colaborativo y coordinado con el equipo multidisciplinario, que integra el macroproyecto sobre prediabetes previamente detallado en esta memoria. Esta sinergia permitió la implementación de protocolos estandarizados y la aplicación de diversas técnicas de recolección de datos, asegurando la obtención de información relevante y de calidad. La participación activa de expertos en diferentes áreas, desde la clínica hasta la bioestadística, enriqueció el proceso de identificación de variables clave y la selección de instrumentos de recolección apropiados para los objetivos específicos del

macroproyecto.

De manera similar, las etapas posteriores de almacenamiento y limpieza de los datos fueron abordadas de forma conjunta y sistemática. El equipo de trabajo estableció protocolos rigurosos para la organización y resguardo seguro de la información recopilada, garantizando su integridad y confidencialidad. La fase de limpieza, esencial para asegurar la validez de los análisis posteriores, implicó la identificación y corrección de inconsistencias, valores atípicos y datos faltantes, mediante la aplicación de criterios predefinidos y la discusión consensuada entre los miembros del equipo.

Esta colaboración estrecha no solo optimizó la eficiencia de estos procesos críticos, sino que también permitió aprovechar la experiencia y el conocimiento colectivo para superar los desafíos inherentes al manejo de grandes volúmenes de información, fortaleciendo así la calidad y la fiabilidad de los datos que sustentan las conclusiones de esta investigación y del macroproyecto en su conjunto.

2. Metodología del trabajo

En la figura 7 se detalla el flujo de los procesos que integran la metodología utilizada en la presente investigación. Cabe señalar que, las técnicas descritas específicamente en las secciones 4.1, 4.2, 4.3 y subsecciones 4.3.1, 4.3.2, 4.3.3, no son exclusivamente de autoría individual ya que se desarrollaron, como aporte técnico al proyecto general, conjuntamente con el maestrante Gabriel Rivadeneira, en ese momento miembro del equipo de investigación.

En contraste, la selección de variables especificada en la sección 4.3.4, la codificación de variables descrita en el acápite 4.3.5, el procesamiento estadístico mediante análisis factorial detallado en la sección 4.4 y el diseño de la visualización interactiva narrada en el punto 4.6 son de autoría propia individual.

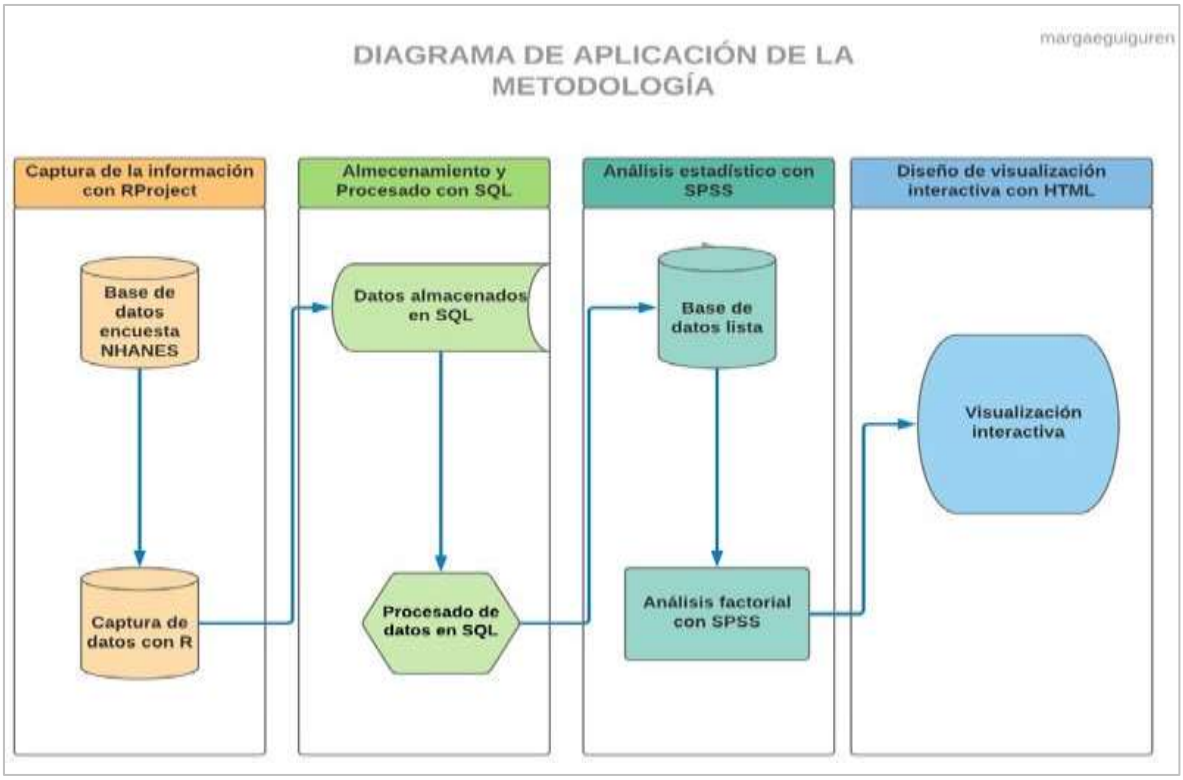


Figura 6 *Diagrama de la metodología utilizada*

2.1. taset para el estudio

Existen varias consideraciones que son de suma importancia y que se deben hacer a la hora de seleccionar una base de datos para desarrollar un estudio. La información contenida en ella debe estar acorde al objetivo del estudio, la fuente de procedencia debe ser totalmente confiable y de libre acceso y por último debe estar totalmente anonimizada. Con estas consideraciones, se elige para la presente investigación, la Encuesta de Salud Nacional de Estados Unidos (NHANES) con un período que comprende los años 2011 al 2016.

2.2. Captura de la información

La captura de los datos se realiza mediante el entorno de programación RStudio para el lenguaje de programación RProject, obteniendo la información

directamente de la página web del NHANES, asegurando de esta manera la calidad, veracidad, completitud y confiabilidad de los datos.

2.3. Almacenamiento de la información

Tras capturar los datos se almacenan en el sistema SQL Server, usando un servidor local, donde se programan diferentes vistas de la información organizándola por año. El siguiente paso es el procesado de los datos realizando un exhaustivo análisis y posterior limpieza.

2.4. Procesado de los datos

El procesado implica verificar la completitud del set procediendo a eliminar los campos que presenten datos nulos o perdidos. Un segundo paso del procesado es verificar los outliers o datos atípicos. Un tercer paso es analizar detenidamente el paquete de datos resultante después del proceso de limpieza, y definir con el aporte de un especialista en la materia las variables que tienen correlación directa con el objetivo del estudio. Verificadas estas variables, se excluyen el resto de las variables y así se define un grupo de datos con los que se trabaja para investigar el fenómeno de la prediabetes.

2.5. Análisis estadístico

Al hacer el análisis de los datos en búsqueda de información relevante para el estudio, se tiende a considerar el mayor número posible de variables. Sin embargo, si se escogen demasiadas variables se puede dificultar la correcta determinación de las relaciones entre las variables obstaculizando la observación efectiva de las correlaciones que puedan existir entre ellas. Por esta razón, se debe seleccionar con cuidado el modelo estadístico que se va a aplicar y posteriormente definir el paquete informático con el que se procesará la data. En este estudio se elige el análisis factorial mediante componentes principales que se ejecuta en el programa estadístico SPSS.

2.6. Diseño de una visualización en HTML

Una vez obtenidos los parámetros del modelo se procede a construir una herramienta de visualización sencilla y amigable, que brinde la posibilidad de que el usuario responda a 10 preguntas que incluyen datos antropométricos como peso, estatura, edad y medidas de exámenes de laboratorio. Una vez que el paciente ingrese sus datos la herramienta arroja un índice conjuntamente con la interpretación del mismo y algunas recomendaciones generales.

CAPÍTULO III

Construcción conceptual: Datos, Análisis y Herramientas

El presente Capítulo se centra en la metodología detallada empleada para la consecución de los objetivos planteados en esta investigación sobre prediabetes. Inicialmente, se describirán los contenidos fundamentales que sustentan el análisis, incluyendo la presentación de la base de datos NHANES, una fuente crucial de información demográfica, nutricional y de salud, así como la definición operativa de prediabetes y la identificación exhaustiva de sus diversos factores de riesgo. Posteriormente, el capítulo se adentrará en la aplicación del análisis factorial, una técnica estadística multivariante esencial para explorar las relaciones subyacentes entre las variables y reducir la dimensionalidad de los datos.

Se detallará cada una de las etapas de este análisis, desde la formulación del problema y el examen de la matriz de correlación, hasta la extracción, determinación, rotación e interpretación de los factores, culminando con la obtención de la matriz de puntuaciones factoriales y la validación del modelo resultante. Finalmente, se presentará el modelo de base de datos diseñado para el almacenamiento y gestión de la información, seguido de una descripción de las tecnologías específicas utilizadas a lo largo del estudio, incluyendo software estadístico, lenguajes de programación y herramientas de desarrollo web.

En este capítulo se describen brevemente las diversas tecnologías utilizadas en el desarrollo de esta propuesta de investigación y se detallan conceptos básicos cuya comprensión es indispensable para desarrollar adecuadamente la presente propuesta.

3. Contenidos implicados

3.1. Base de datos NHANES (National Health and Nutrition Examination Survey)

La Encuesta de salud y nutrición (NHANES)[1] es realizada por el Centro Nacional

[1] https://www.cdc.gov/nchs/nhanes/index.htm

de Estadísticas de Salud (NCHS) de los Estados Unidos. Combina exámenes físicos con una encuesta con preguntas demográficas, socioeconómicas, de hábitos alimenticios y de actividad física. Inició en el año 1990 y se realiza anualmente hasta la actualidad, recopilando información de alrededor de 5000 personas por año y ha encuestado aproximadamente a 190.000 personas. Los datos recolectados por esta encuesta son la base para muchos estudios epidemiológicos y condiciones generales de salud de la población. Con el fin de elaborar políticas públicas sanitarias y determinar la prevalencia de las principales enfermedades y sus factores de riesgo. Gracias a la cantidad, calidad y veracidad de los datos, es un dataset confiable para la investigación.

3.2. Prediabetes y sus factores de riesgo

En la introducción de esta investigación, se hizo un breve preámbulo sobre la prediabetes, en el presente apartado se desarrollará con mayor profundidad este concepto conjuntamente con una explicación amplia de sus factores de riesgo. La prediabetes es una condición de salud que se enmarca en un espacio entre la normalidad de los valores establecidos para las mediciones de glucosa en ayunas, tolerancia a la glucosa y glicohemoglobina y, los valores que determinan la diabetes.

Generalmente, es una patología subdiagnosticada ya que no presenta ninguna sintomatología. Al respecto algunos autores mencionan que, “Más de uno de cada tres estadounidenses tiene prediabetes, y el 90 por ciento de ellos ni siquiera saben que la tienen” (Brass, 2018). “No hay síntomas claros de prediabetes, por lo tanto, puede tenerla y no saberlo” (ADA, 2015). “Dada su alta frecuencia resulta conveniente considerar la prediabetes como un estado de riesgo importante para la predicción de diabetes y de complicaciones vasculares, así como una manifestación subclínica de un trastorno del metabolismo de los carbohidratos” (Díaz et al., 2011).

“Los diagnósticos más tradicionales de la prediabetes se basan en futuras predicciones de riesgo e incluyen mujeres con antecedentes del síndrome de

ovario poliquístico o diabetes gestacional, descendencia de padres con diabetes tipo 2, e individuos con adiposidad abdominal" (Garber et al., 2008). Una revisión minuciosa de la literatura permitió identificar un grupo suficientemente grande de factores de riesgo de la prediabetes, entre ellos se pueden citar: edad, sexo, etnia, antecedentes familiares, niveles de glucosa en plasma, índice de masa corporal, diámetro sagital abdominal, actividad física, antecedentes de hipertensión, resistencia a la insulina.

Además, otras pruebas de laboratorio como glicohemoglobina, tolerancia a la glucosa, triglicéridos, colesterol total, HDL, LDL, transaminasas, así como otros factores antropométricos, ovarios poliquísticos, diabetes gestacional, acantosis nigricans, desórdenes del sueño. Los autores Khetan et al., (2018) mencionan que, "a pesar de la base de evidencia bien establecida para el tratamiento de prediabetes, hay un sustancial escaso reconocimiento y tratamiento insuficiente del problema".

La *American Association of Clinical Edocrinologists* (AACE) propone que las personas que cumplan con cualquiera de los factores de riesgo que se detallan a continuación, acudan a realizarse exámenes para detectar prediabetes o diabetes tipo II:

- Edad ≥ 45 años sin otros factores de riesgo.
- Antecedentes familiares de diabetes
- Sobrepeso u obesidad
- Estilo de vida sedentario
- Miembro de un grupo racial o étnico en riesgo: asiático, afroamericano, hispano, nativo americano, isleño del pacífico
- Colesterol de lipoproteínas de alta densidad (HDL-C) <35 mg / dL (0.90 mmol / L) y / o un nivel de triglicéridos> 250 mg / dL (2.82 mmol / L)
- Deterioro de la tolerancia a la glucosa (IGT), deterioro de la glucosa en ayunas (IFG) y / o síndrome metabólico
- Síndrome de ovario poliquístico (PCOS), acantosis nigricans o

enfermedad del hígado graso

- Hipertensión (presión arterial> 140/90 mm Hg o en terapia antihipertensiva)
- Historial de diabetes gestacional o parto de un bebé que pesa más de 4 kg (9 lb)
- Tratamiento antipsicótico para la esquizofrenia y / o enfermedad bipolar grave
- Exposición crónica a glucocorticoides
- Trastornos del sueño en presencia de intolerancia a la glucosa (A1C> 5.7%, IGT o IFG en pruebas previas), que incluyen apnea obstructiva del sueño (AOS), privación crónica del sueño y ocupación en turnos nocturnos.

3.3. Análisis factorial

El análisis factorial es una técnica estadística que permite simplificar grandes volúmenes de información, agrupando variables en conjuntos más pequeños y homogéneos, basados en la fuerza de su correlación. Su propósito es captar la mayor cantidad de información posible utilizando el menor número de dimensiones, logrando así una representación más clara y manejable del fenómeno que se estudia. Este principio, conocido como parsimonia o economía de la información, ayuda a evitar modelos innecesariamente complicados (De La Fuente, 2011).

Además de buscar esta reducción eficiente, el análisis factorial también se enfoca en que los factores obtenidos tengan sentido práctico: que cada grupo de variables agrupadas refleje una dimensión del problema que pueda ser fácilmente interpretada. A lo largo del proceso, se elabora una matriz de correlaciones entre las variables, la cual sirve de base para identificar los factores comunes subyacentes. El resultado final es un número limitado de factores que no solo resumen la información original, sino que también revelan patrones o estructuras que, de otra forma, quedarían ocultos en los datos dispersos (López-Roldán, 2015).

Esta herramienta es especialmente valiosa en investigaciones exploratorias, donde aún no existe una teoría consolidada que guíe la relación entre las variables,

permitiendo descubrir de manera inductiva cómo se organizan y qué dimensiones principales emergen del fenómeno analizado. Generalmente “un análisis factorial eficiente es aquel que proporciona una solución factorial sencilla e interpretable” (Ibarra, 2001).

“Todo modelo debe procurar ser lo más simple posible en la interpretación o explicación de los datos. La máxima de este tipo de técnicas se expresa en la afirmación “pérdida de información y ganancia en significación” (López-Roldán, 2015). El análisis factorial se puede realizar desde dos enfoques, mediante análisis de componentes principales que analiza toda la varianza de las variables y por análisis de factores de riesgo que se basa solo en la varianza común.

Es importante considerar que existen dos tipos principales de análisis factorial, cada uno con un propósito distinto. Por un lado, el análisis factorial exploratorio (AFE) se utiliza cuando el investigador no conoce previamente los factores comunes; es decir, no tiene una hipótesis clara sobre cómo se agrupan las variables. En este caso, el objetivo es descubrir de manera inductiva las estructuras subyacentes en los datos, permitiendo identificar patrones o agrupaciones de variables que comparten características similares (De La Fuente, 2011).

Por otro lado, el análisis factorial confirmatorio (AFC) parte de un enfoque diferente. Aquí, el investigador ya cuenta con una hipótesis previa sobre cuántos factores existen y qué variables deberían estar asociadas a cada uno. A través del AFC se busca comprobar si los datos empíricos respaldan esa estructura teórica propuesta, validando modelos específicos con base en conocimiento previo o investigaciones anteriores (López-Roldán, 2015).

Mientras que el análisis exploratorio es más flexible y abierto a lo que los datos revelen, el confirmatorio requiere una planificación cuidadosa y suele apoyarse en criterios estadísticos más estrictos para evaluar la calidad del ajuste entre el modelo teórico y los datos reales. Ambos enfoques son herramientas fundamentales en la investigación, dependiendo de si el interés es generar nuevas

teorías o validar hipótesis ya formuladas (Ibarra, 2011; López-Roldán, 2015).

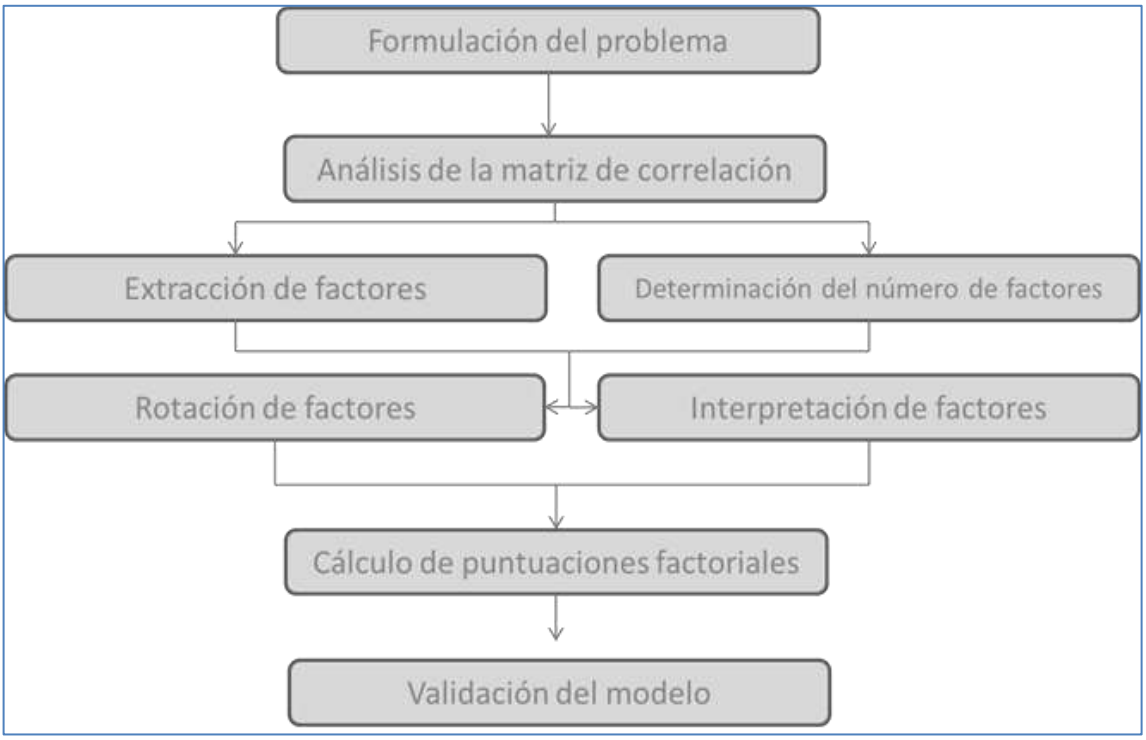


Figura 7 *Esquema de análisis factorial*
(De La Fuente, 2011).

Como muestra la figura 7, el análisis factorial tiene varias etapas o fases que se describen a continuación:

3.3.1. Análisis de la matriz de correlación

El objetivo de realizar el análisis de la matriz es comprobar las correlaciones existentes entre las variables y definir si es conveniente realizar el análisis factorial, para ello hay que determinar si las variables están altamente Inter correlacionadas. Los indicadores más importantes para analizar la matriz de correlaciones son el test de esfericidad de Bartlett y la media de adecuación de la muestra de Kaiser-Meyer-Olkin (KMO). El test de esfericidad de Bartlett evalúa la conveniencia de aplicar el análisis factorial, contrastando si la matriz de correlaciones es igual a la

matriz identidad (donde la diagonal es 1 y los valores fuera de la diagonal son 0), que permite identificar un nivel suficiente de multicolinealidad entre las variables. Si existe suficiente multicolinealidad, el p valor resultante debe ser menor a 0,05 y será válido aplicar el modelo de análisis factorial. Según De la Fuente (2011) “el índice KMO se utiliza para comparar las magnitudes de los coeficientes de correlación parcial”, toma valores entre 0 y 1 y su resultado se evalúa con los siguientes criterios:

- KMO ≥ 0,75 ⇒ Muy bien
- KMO ≥ 0, 5 ⇒ Aceptable
- KMO < 0, 5 ⇒ Inaceptable

El índice KMO, conjuntamente con el test de esfericidad de Barlett, son indicadores que ayudan a determinar si se debe o no aplicar el análisis factorial en un determinado estudio, como muestra la figura 8.

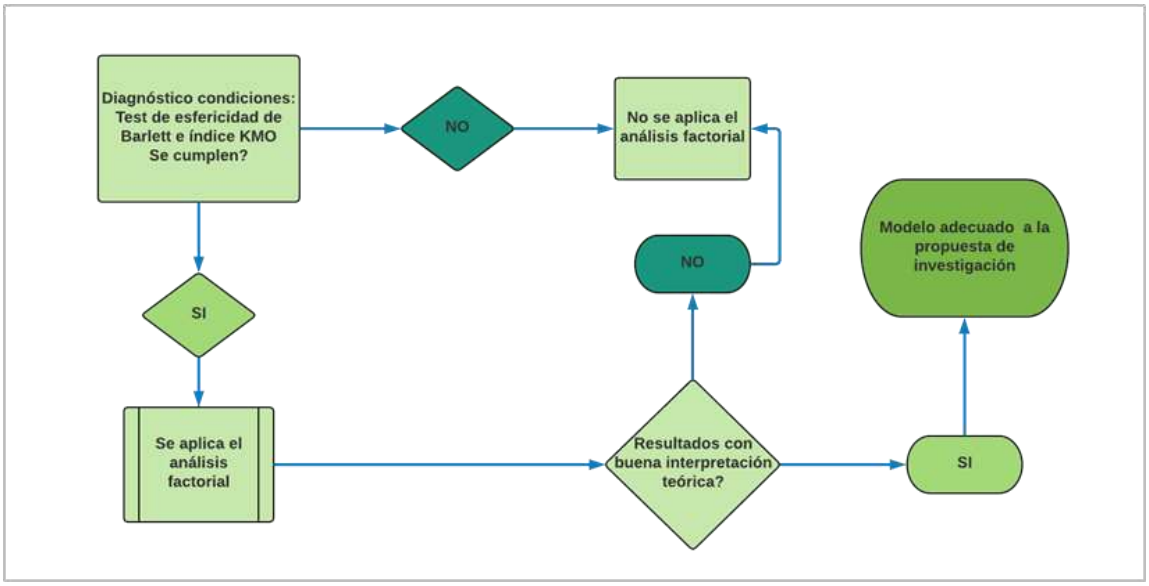


Figura 8 *Determinación de la conveniencia de aplicar el análisis factorial*

3.3.2. Extracción de factores

Extraer factores mediante análisis factorial, significa reducir las variables originales a un número menor de factores que explican de manera similar el fenómeno. Para la extracción se procede a escoger el método que se considere

más adecuado, en este caso el Paquete Estadístico para las Ciencias Sociales (SPSS, por sus siglas en inglés), versión 27 (IBM Corp.), ofrece las siguientes opciones: Método de las Componentes Principales, Método de los Ejes principales y Método de Máxima Verosimilitud. Cada uno de ellos tiene sus particularidades y se elige según los objetivos de la investigación y las características de los datos. El Método de los Ejes Principales, por otro lado, se enfoca exclusivamente en explicar la varianza común, es decir, aquella parte de la varianza que las variables comparten.

El Método de Máxima Verosimilitud no solo estima los factores comunes, sino que también permite realizar contrastes estadísticos para evaluar la adecuación del modelo y compararlo con otros modelos posibles. En la investigación se utiliza el método de componentes principales, que analiza la varianza total que incluye la varianza específica y la varianza de error y, su objetivo es explicar la mayor cantidad de varianza posible en los datos observados (Pérez y Medrano 2010).

3.3.3. Determinación del número de factores

La selección del número de factores a retener es un paso crucial en el análisis factorial, ya que de ello depende que el modelo sea interpretativo y útil. Una de las formas más comunes de hacerlo es mediante la regla de Kaiser, que viene configurada como opción predeterminada en el software SPSS. Esta regla consiste en conservar únicamente aquellos factores cuyos autovalores (o eigenvalues) sean superiores a 1, ya que un valor mayor indica que ese factor explica más varianza que una variable individual (Pérez y Medrano, 2010).

Otra estrategia muy utilizada es el diagrama de sedimentación (*scree test)*, que genera una gráfica donde se representan los autovalores en orden descendente. El número óptimo de factores se identifica visualmente en el punto donde la curva comienza a aplanarse, es decir, donde ocurre el primer gran cambio de pendiente (lo que se conoce como el "codo" de la gráfica). Este criterio, conocido también como criterio de contraste de caída, ayuda a decidir de forma más intuitiva cuántos factores conservar para lograr un modelo más sencillo y representativo.

En general, ambos métodos se suelen aplicar de manera complementaria: la regla de Kaiser aporta un criterio numérico objetivo, mientras que el diagrama de sedimentación ofrece una validación visual que facilita la interpretación.

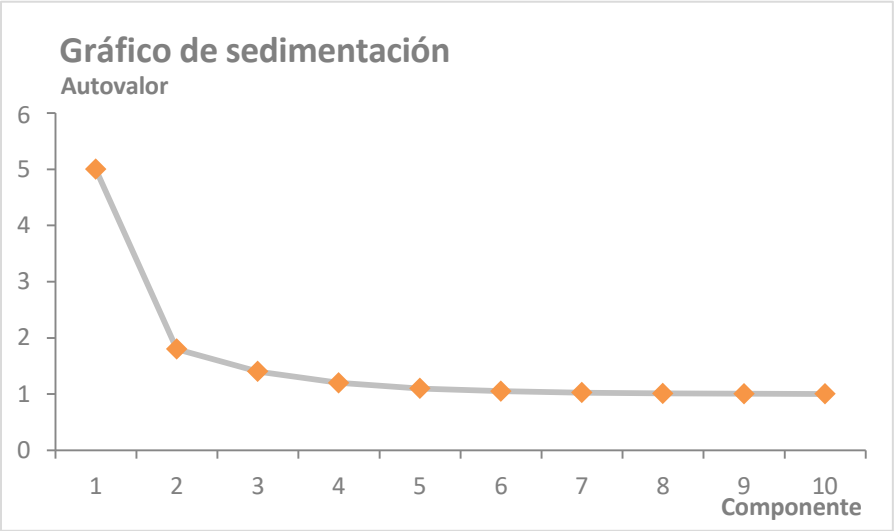


Figura 9 *Ejemplo de gráfica de sedimentación (scree test)*

3.3.4. Rotación de factores

El análisis factorial arroja en su primera etapa una matriz de correlaciones de las variables con los factores (matriz factorial), pero esta matriz es bastante complicada de interpretar por lo que se procede a rotar los factores. Los factores se rotan con el fin de eliminar las correlaciones negativas y disminuir las correlaciones entre las variables de cada factor. Esta rotación tiene como finalidad aproximar la solución factorial a una estructura simple (la correlación de cada variable con uno de los factores se aproxime a 1 y a 0 con el resto de los factores). La matriz resultante se llama matriz de componentes o factores rotados y las rotaciones pueden ser oblicuas y ortogonales.

— La rotación ortogonal donde “los ejes se rotan de forma que quede preservada la incorrelación entre los factores. Es decir, los nuevos ejes (ejes rotados) son perpendiculares de igual forma que lo son los factores sin rotar”,

el método más utilizado es el Varimax, establecido por defecto en el programa SPSS.

- La rotación oblicua, en la cual se permite que los factores se correlacionen entre sí. A diferencia de la ortogonal, aquí los ejes no permanecen perpendiculares después de la rotación, lo que refleja de manera más realista la posible interrelación entre las variables latentes en muchos fenómenos sociales, educativos o de salud (De La Fuente, 2011).

Esta flexibilidad hace que los resultados puedan ser más interpretables cuando, en la práctica, las dimensiones no son completamente independientes. Dentro de la rotación oblicua, los métodos más conocidos son el Oblimin y el Promax. Oblimin es un procedimiento más general que permite diferentes grados de correlación entre factores, mientras que Promax es una opción más rápida y eficiente, especialmente útil cuando se trabaja con grandes conjuntos de datos (De La Fuente, 2011).

Así, mientras la rotación ortogonal (como Varimax) busca mantener los factores independientes para una interpretación más sencilla, la rotación oblicua acepta que los factores puedan estar relacionados, ofreciendo un modelo más flexible y, en muchos casos, más cercano a la realidad de los datos (De La Fuente, 2011).

3.3.5. Interpretación de factores

El análisis factorial arroja uno o más factores o componentes con sus respectivas correlaciones con cada variable. El análisis que se debe hacer, con base en los datos de la matriz rotada, es mirar qué grado de correlación (pesos) tiene cada variable en los diferentes factores apoyándose en el conocimiento teórico de la materia en estudio. Así mismo no se debe perder de vista el porcentaje de la varianza que explica cada factor obtenido. Generalmente, estas saturaciones o pesos resultantes deben ser mayores a 0.30 para ser consideradas, pero esto dependerá de lo que el investigador esté indagando en los datos.

3.3.6. Matriz de puntuaciones factoriales

El siguiente paso luego de la rotación de factores es calcular la matriz de puntaciones factoriales. Existen diversos métodos para calcular las puntuaciones factoriales, entre ellos están los siguientes:

- Método de Regresión: con él se pueden obtener puntuaciones que tengan máxima correlación con las puntuaciones teóricas. Utiliza el método de mínimos cuadrados.
- Método de Barlett: con este método las puntuaciones tienen media 0
- Método de Anderson-Rubin: es una variación del método de Barlett en el que las puntuaciones tienen media 0, desviación estándar de 1 y no tienen correlación entre sí.

3.3.7. Validación del modelo

Los modelos estadísticos, una vez planteados deben ser validados. Con el análisis factorial generalmente, se utilizan dos procedimientos para hacerlo como el análisis de la bondad de ajuste y observando la generalidad de los resultados obtenidos. El análisis de bondad de ajuste se realiza observando los residuos analizando las diferencias entre la matriz de correlación de entrada y las correlaciones reproducidas (matriz factorial). Si estos residuos son pequeños entonces se concluye que el modelo factorial es adecuado. Por el contrario, el modelo no se ajusta a los datos, cuando hay un porcentaje elevado de residuos.

La observación de la generalidad de los resultados consiste en aplicar el análisis factorial nuevamente a una muestra diferente o subgrupo de datos para validar los resultados obtenidos. También, se puede ejecutar nuevamente el modelo sacando variables que presenten bajas relaciones con alguno de los factores o también aquellas que tienen alta correlación para observar los resultados que arroja el procedimiento. Otro aspecto importante es determinar si el número de casos por variable es alto entonces habrá mayor estabilidad en los

resultados.

3.4. Modelo de base de datos

El modelo de base de datos ayuda a visualizar de forma simplificada cómo se organiza la información en una base de datos multidimensional. Tiene tres componentes básicos entidades, atributos y relaciones. Que se describen a continuación:

- Entidades: representan objetos o cosas existentes en el mundo real, pueden ser concretas o abstractas, se distinguen de otras entidades por sus atributos o características individuales.
- Atributos: son las particularidades que diferencian a cada entidad de las otras.
- Relaciones: las relaciones permiten establecer vínculos o asociaciones entre las entidades. Los tipos de relaciones que se dan son uno a uno, uno a varios o varios a varios, esto se denomina correspondencia de “cardinalidades”.

El modelo permite identificar de manera rápida y sencilla el tipo de diseño que tiene el conjunto de datos. Los diseños más utilizados son estrella y copo de nieve, el primero consta de una tabla central de "hechos" y varias "dimensiones", como se aprecia en la figura 10, que se interrelacionan mediante una clave primaria. El segundo, es una derivación del primero, tiene una tabla de “hechos” que está conectada a muchas tablas de “dimensiones” y éstas a su vez se relacionan con otras tablas de “dimensiones”. La selección depende de cómo está organizada la información en la base de datos y hasta qué grado de profundidad se quiera mostrar en el diagrama. En este trabajo se utiliza el modelo estrella.

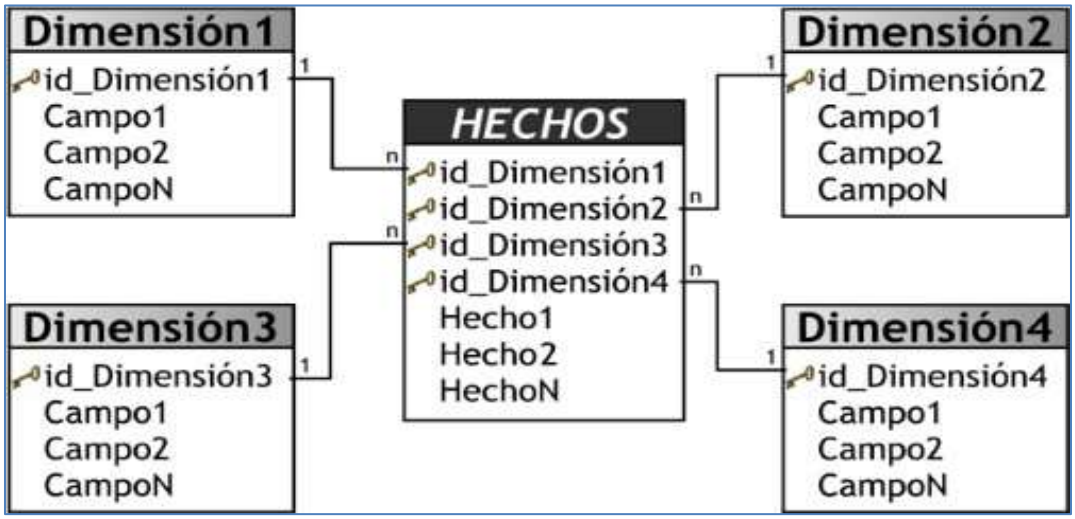


Figura 10 *Modelo estrella de base de datos*

3.5. Tecnologías utilizadas

A continuación, se hace una breve descripción de las herramientas y tecnologías que se utilizan en el presente trabajo de investigación y que permiten capturar, procesar, analizar los datos y presentar los resultados.

3.5.1. R Project

R Project[2] es un lenguaje y entorno de programación libre, bien desarrollado, orientado a objetos basado en comandos y que soporta como tipo de datos básicos valores numéricos, caracteres, cadenas de caracteres y valores booleanos o lógicos. Es utilizado principalmente para análisis estadístico ya que es un lenguaje para análisis de datos muy potente que permite el almacenamiento y manipulación de los datos, posee operadores para cálculo y una gama muy amplia de posibilidades gráficas; funciona con diferentes tipos de hardware y software (Windows, Unix, Linux…), maneja grandes volúmenes de datos y ofrece la posibilidad de cargar bibliotecas y paquetes con diversas funcionalidades.

[2] https://www.r-project.org/

3.5.2. R Studio

R Studio[3] es un entorno de desarrollo integrado (IDE, por sus siglas en inglés) diseñado específicamente para trabajar con el lenguaje de programación R. Un IDE es un entorno de programación que combina, en una sola aplicación, herramientas esenciales como un editor de código, una interfaz gráfica de usuario (GUI), un compilador, un sistema de control de versiones y un depurador, lo que facilita significativamente el desarrollo y ejecución de scripts. En este sentido, RStudio se destaca por ofrecer una plataforma intuitiva, flexible y muy accesible tanto para principiantes como para usuarios avanzados.

Este entorno puede ejecutarse tanto como aplicación de escritorio en sistemas operativos Windows, Mac y Linux, como en versión web a través de un navegador, lo cual amplía su disponibilidad y permite el trabajo colaborativo en la nube. RStudio no solo facilita la escritura y prueba de código, sino que también integra funciones avanzadas para la visualización de datos, elaboración de reportes dinámicos, modelado estadístico, y conexión con bases de datos, lo que lo convierte en una herramienta poderosa para el análisis de datos en múltiples disciplinas.

Uno de los grandes aportes de RStudio es su contribución activa a la comunidad académica, científica e industrial, proporcionando un ecosistema robusto y de código abierto que fomenta la investigación reproducible, el desarrollo de paquetes propios y el intercambio de conocimiento. Su uso es común en campos tan diversos como la biomedicina, las ciencias sociales, la economía, la ingeniería y la educación, siendo ampliamente adoptado tanto en instituciones académicas como en organizaciones privadas.

Es importante destacar que, para poder utilizar RStudio, es necesario instalar previamente R Project —el lenguaje de programación base sobre el cual se construye este entorno. Una vez instalado R, RStudio actúa como una capa

[3] https://www.rstudio.com/

amigable y potente que permite aprovechar al máximo las capacidades del lenguaje, facilitando tanto tareas simples como análisis estadísticos básicos, como procesos más complejos de minería de datos, machine learning o modelado matemático.

3.5.3. SQL Server

SQL Server [4] es un sistema de gestión de bases de datos relacional (RDBMS, por sus siglas en inglés) desarrollado por Microsoft, diseñado para almacenar, organizar y acceder a grandes volúmenes de datos de manera eficiente. Este sistema utiliza el lenguaje de consulta estructurado (SQL) como principal medio de interacción con la base de datos, lo que permite realizar operaciones complejas como la recuperación, inserción, actualización y eliminación de datos de forma rápida y segura.

A diferencia de otros gestores de bases de datos de código abierto como MySQL o PostgreSQL, SQL Server es un software propietario, aunque Microsoft ha liberado versiones como SQL Server Express y ha desarrollado una edición compatible con Linux, lo que ha ampliado su alcance a múltiples plataformas. Se caracteriza por su capacidad para manejar múltiples tablas interrelacionadas, lo que permite estructurar la información de manera lógica y coherente, facilitando su análisis y recuperación. Además, al ser un sistema multiusuario, permite que múltiples personas accedan y trabajen simultáneamente sobre la misma base de datos, sin comprometer el rendimiento ni la integridad de la información.

SQL Server es ampliamente utilizado en entornos empresariales y corporativos por su estabilidad, escalabilidad y características avanzadas como la integración con herramientas de inteligencia empresarial (Business Intelligence), procesamiento de datos en tiempo real, automatización de tareas, y gestión segura de transacciones. Gracias a estas funcionalidades, su uso se ha extendido más allá del ámbito tradicional de los desarrolladores de software, siendo empleado

[4] https://www.microsoft.com/en-us/sql-server/sql-server-downloads

también por analistas de datos, administradores de sistemas y profesionales de distintas disciplinas que requieren manejar grandes volúmenes de información de forma organizada y eficiente. Su uso está generalizado en el campo del desarrollo de aplicaciones web por lo que es popular en grandes sitios web como Google, Facebook, Youtube y muchos otros.

3.5.4. IBM SPSS

SPSS[5] (Statistical Package for the Social Sciences) es un programa estadístico ampliamente reconocido por su robustez, versatilidad y facilidad de uso, especialmente en el ámbito de las ciencias sociales. No obstante, su aplicabilidad ha trascendido los límites de este campo, convirtiéndose en una herramienta indispensable para investigadores de áreas como la salud, la educación, el marketing, la psicología, la economía y otras disciplinas que requieren análisis cuantitativos rigurosos.

Este software permite realizar un amplio abanico de procedimientos estadísticos, desde análisis descriptivos básicos hasta técnicas más complejas como análisis multivariado, modelado predictivo, minería de datos y análisis de big data. Además, ofrece funcionalidades para la creación de modelos de regresión, pruebas T, análisis de varianza (ANOVA), correlaciones y análisis factoriales, todo ello a través de una interfaz intuitiva que no exige conocimientos avanzados de programación, lo cual lo hace accesible para usuarios de diversos niveles de formación técnica.

Uno de los puntos fuertes de SPSS es su capacidad para integrar el procesamiento de datos con la generación de informes y visualizaciones claras, facilitando la interpretación de los resultados y su comunicación efectiva. A través de sus módulos adicionales, como SPSS Amos (para análisis estructural), SPSS Modeler (para minería de datos) o SPSS Text Analytics (para análisis de datos no estructurados), el software se adapta a las crecientes demandas de la

[5] https://www.ibm.com/products/spss-statistics

investigación moderna, incluyendo el análisis de grandes volúmenes de información y la automatización de procesos estadísticos.

Además de permitir la importación y manipulación de bases de datos en múltiples formatos, SPSS facilita la codificación, limpieza y transformación de datos, aspectos fundamentales para garantizar la calidad del análisis. Su capacidad para generar gráficos, tablas y reportes personalizables lo convierte en una herramienta completa tanto para el trabajo exploratorio como para la presentación de resultados finales en contextos académicos o institucionales.

3.5.5. HTML

Hyper Text Markup Language (HTML)[6] es un lenguaje de marcado que se utiliza para estructurar y presentar contenido en la web. A diferencia de los lenguajes de programación tradicionales, HTML no ejecuta funciones lógicas, sino que organiza la información mediante etiquetas que indican cómo debe visualizarse cada elemento dentro de una página. Su estructura básica se compone de dos partes principales: la cabecera (*head*), que contiene metadatos e información técnica del documento, y el cuerpo (*body*), que incluye todos los contenidos visibles para el usuario, como texto, imágenes, enlaces, tablas o videos.

Gracias a su simplicidad y estandarización, HTML se ha convertido en la piedra angular del desarrollo web, siendo compatible con todos los navegadores modernos. Además, puede integrarse con otros lenguajes como CSS para el diseño visual, y JavaScript para funciones interactivas, lo que lo convierte en una herramienta fundamental en la creación de sitios web tanto estáticos como dinámicos.

3.5.6. JavaScript

JavaScript[7] es un lenguaje de programación interpretado, orientado a objetos

[6] https://html.com/
[7] https://www.javascript.com/

y de propósito general, ampliamente utilizado en el desarrollo web para crear contenidos dinámicos e interactivos. Su principal ventaja radica en que los scripts escritos en JavaScript pueden ejecutarse directamente en cualquier navegador moderno, sin necesidad de compilación previa ni de herramientas adicionales, lo que facilita su implementación y prueba en tiempo real.

Gracias a su capacidad para manipular el contenido del documento HTML, responder a eventos del usuario y comunicarse con servidores web, JavaScript permite añadir funcionalidades avanzadas a las páginas, como menús desplegables, validación de formularios, animaciones, actualizaciones automáticas de contenido y mucho más. Además, su compatibilidad multiplataforma lo ha consolidado como una de las tecnologías clave en el desarrollo de aplicaciones web modernas, especialmente cuando se combina con HTML y CSS.

3.5.7. Brackets

Brackets[8] es un editor de texto moderno y de código abierto, desarrollado originalmente por Adobe y mantenido por la comunidad a través de GitHub. Está especialmente orientado al diseño y desarrollo web, ya que fue construido utilizando tecnologías como HTML, CSS y JavaScript, lo que lo convierte en una herramienta ligera, versátil y multiplataforma.

Una de sus características más destacadas es la vista previa en tiempo real, que permite a los desarrolladores ver los cambios en el navegador a medida que editan el código, facilitando así un flujo de trabajo más ágil e intuitivo. Además, Brackets ofrece una interfaz limpia y herramientas útiles como edición en línea, resaltado de sintaxis, autocompletado y soporte para extensiones, lo que lo convierte en una opción muy valorada, especialmente por quienes están comenzando en el mundo del desarrollo web.

[8] http://brackets.io/

CAPÍTULO IV

Proceso de desarrollo de la metodología aplicada al estudio de la prediabetes

El presente capítulo se ha concebido con el propósito fundamental de desglosar y exponer de manera exhaustiva, pero a la vez accesible y desprovista de tecnicismos innecesarios, la aplicación detallada de la metodología que sustenta el presente estudio. Conscientes de la importancia de la transparencia y la replicabilidad en la investigación científica, nos esforzamos por clarificar cada etapa del proceso, desde la concepción de la pregunta de investigación hasta el análisis final de los datos.

Este esfuerzo se inscribe dentro de un marco colaborativo más amplio, cuyo objetivo primordial es contribuir de manera significativa a la investigación pionera sobre prediabetes liderada por el distinguido PhD Danilo Esparza. El presente trabajo, por consiguiente, no debe entenderse como una iniciativa aislada, sino como una parte integrante y esencial de este macroproyecto, aportando elementos metodológicos y resultados específicos que enriquecen la comprensión global de esta condición precursora de la diabetes.

En este contexto de colaboración y con la visión de generar un impacto duradero en la investigación, se desarrollaron en estrecha sinergia una serie de procesos metodológicos críticos, específicamente aquellos concernientes a la captura meticulosa, la depuración exhaustiva y el almacenamiento seguro de los datos. Para llevar a cabo estas tareas fundamentales, se emplearon herramientas tecnológicas de vanguardia como R Studio, un entorno de programación potente y versátil para el análisis estadístico, y SQL Server, un sistema de gestión de bases de datos robusto y confiable.

La integración estratégica de estas plataformas tuvo como finalidad última la creación de un *Datawarehouse* sofisticado. Este repositorio centralizado de datos no solo facilita la gestión y el análisis de la información recopilada en el presente estudio, sino que también se erige como un aporte tecnológico de valor incalculable, destinado a servir como una base sólida y estructurada para futuras investigaciones en el campo de la prediabetes y áreas relacionadas. Se espera que este

Datawarehouse, con su arquitectura bien definida y sus datos depurados, impulse nuevos descubrimientos y acelere el avance del conocimiento en la lucha contra esta creciente amenaza para la salud pública.

4. Desarrollo de la aplicación de la Metodología

A continuación, se describirá el camino metodológico propuesto.

4.1. Captura de la información con RStudio

Para realizar la captura de la información, el primer paso es instalar los paquetes de R necesarios, para lo cual se escribe el código que se aprecia en la figura 14, líneas 1 a la 10. Los paquetes instalados son *foreing* y *RODBC*, el primero descarga los datos en formato de STATA desde la página del NHANES y el segundo permite la opción de trabajar con sentencias SQL dentro del programa RStudio.

El segundo paso es traer a primer plano el paquete RODBC, con el que RStudio se conecta a SQL Server, mediante los comandos que se aprecian en las líneas 11 a 16 de la figura 14. Esta integración facilita la comunicación de doble vía entre los dos programas antes señalados.

De la línea 17 a la 26 de la figura 11, el código trae a primer plano la librería *foreing* con el fin de descargar el archivo de los datos en formato STATA de la página del NHANES. La descarga que se aprecia corresponde a la tabla BIOPRO de los años 2015-2016; este proceso de descarga se repite sucesivamente para todas las tablas de todos los años que componen el estudio.

```r
#PAQUETES Y CONEXIÓN CON SQL SERVER#

# Paquete para descarga de información de STATA

install.packages('foreign')

# Paquete para conexión con SQL Server

install.packages('RODBC')

library(RODBC)

driver <- odbcDriverConnect(connection = "Driver={SQL Server Native Client 11.0}; server=LAPTOP; database=NHANES;
    trusted_connection=yes; ")


#TABLAS 2015-2016

# Descarga BIOPRO_I

library(foreign)

download.file("https://wwwn.cdc.gov/Nchs/Nhanes/2015-2016/BIOPRO_I.XPT", "BIOPRO_I.XPT",mode="wb")

data <- read.xport("BIOPRO_I.XPT")

```

Figura 11 *Código R para captura de información*

Como se puede apreciar en la tabla 1, la denominación de las tablas no es la misma para todos los períodos elegidos, pero la información contenida en ellas es análoga para todos los años.

TABLAS			DESCRIPCIÓN
2011-2012	2013-2014	2015-2016	
DEMO G	DEMO H	DEMO I	DATOS DEMOGRÁFICOS
BMX G	BMX H	BMX I	MEDIDAS CORPORALES
GHB G	GHB H	GHB I	HEMOGLOBINA GLICOSILADA
GLU G	GLU H	GLU I	GLUCOSA
OGTT G	OGTT H	OGTT I	TOLERANCIA A GLUCOSA
BIOPRO G	BIOPRO H	BIOPRO I	PERFIL BIOQUÍMICO
MCQ G	MCQ H	MCQ I	CONDICIONES MÉDICAS
PAQ G	PAQ H	PAQ I	ACTIVIDAD FÍSICA
SLQ G	SLQ H	SLQ I	DESORDEN DE SUEÑO
TRIGLY G	TRIGLY H		TRIGLICÉRIDOS
	INS H	INS I	INSULINA

Tabla 1 *Tablas que componen la encuesta NHANES*

4.2. Almacenamiento de la información en SQL Server

Luego de descargar la información en una variable temporal de RStudio denominada data, se procede a almacenarla de forma permanente en SQL Server. Para prever posibles pérdidas de información, por mal funcionamiento del servidor donde se almacena la información, se crean archivos planos, en formato CSV, de respaldo que se recopilan en un directorio local, como se muestra en la figura 12.

```
27 write.csv(data, file = "C:/NHANES/BIOPRO_I.csv")
28
29 library(RODBC)
30
31 sqlSave(driver,data,tablename = "BIOPRO_I")
32
```

Figura 12 *Almacenamiento y respaldo de la información*

4.3. Procesado de los datos en SQL

Posteriormente se realiza el procesado de los datos, creación de una clave primaria, limpieza de datos y elaboración del modelo de la base de datos, procesos que, como ya se ha mencionado antes, se realizaron en conjunto con el equipo de investigación del macroproyecto sobre prediabetes.

4.3.1. Creación de una clave primaria

A continuación, se crean dos columnas en todas las tablas, una correspondiente a los años a los que pertenecen las tablas y otra con una secuencia que une el ID con el campo de años, creando un nuevo identificador único que cumplirá las funciones de clave primaria, como se puede apreciar en la figura 13.

```
sqlQuery(driver,"
        USE [NHANES]

        ALTER TABLE [dbo].[BIOPRO_I]
        ADD Years Varchar(50)
        ")

sqlQuery(driver,"
        USE [NHANES]

        UPDATE [dbo].[BIOPRO_I]
        SET Years='2015-16'
        ")

sqlQuery(driver,"
        USE [NHANES]

        ALTER TABLE [dbo].[BIOPRO_I]
        ADD ID Varchar(50)
        ")

sqlQuery(driver,"
        USE [NHANES]

        UPDATE [dbo].[BIOPRO_I]
        SET ID=Years+'-'+CONVERT(VARCHAR,SEQN)
        ")
```

Figura 13 *Creación de una clave primaria*

4.3.2. Limpieza de datos

Se procede a crear varias vistas, una por cada período de tiempo, mismas que se utilizan para verificar las condiciones de calidad que debe cumplir la base de datos resultante, de acuerdo con los siguientes parámetros:

- Completitud: Se busca comprobar la cantidad de datos disponibles con respecto a los totales por cada una de las variables, ignorando los registros que presenten campos nulos.
- Precisión: Se analizan los datos en búsqueda de datos erróneos que se salgan de los parámetros reales para cada variable, procediendo a omitir aquellos que presenten valores visiblemente erróneos.
- Consistencia: Se examinan los datos verificando que sean coherentes entre ellos. Al verificar incoherencias se procede a excluir los registros correspondientes.

Esto puede observarse con mayor detalle en la figura 14:

Union_2011_16.sql - not connected | Vista_2011_16.sql - not connected* | Vista_2015_16.sql - not connected*

```sql
SELECT
    [Genero]
    ,[Edad]
    ,[Etnia]
    ,[DiamSagitalAbdominal]
    ,[IMC]
    ,[CircAbdonimal]
    ,[Glicohemoglobina]
    ,[Insulina]
    ,[ToleranciaGlucosa]
    ,[GlucosaPlasma]
    ,[Trigliceridos]
    ,[HistoriaFamiliar]
    ,[ActividadFisica]
    ,[HorasSueno]
    ,[ALT]
    ,[AST]
    ,[HOMA]
    ,(CASE
        WHEN [Diagnostico]='Prediabetes' THEN 1
        ELSE 0
    END) AS DIAGNOSTICO

FROM [dbo].[Union2011_16]
WHERE
Edad>=20 AND
Edad<=65 AND
Genero IS NOT NULL AND
Etnia IS NOT NULL AND
DiamSagitalAbdominal IS NOT NULL AND
IMC IS NOT NULL AND
CircAbdonimal IS NOT NULL AND
Glicohemoglobina IS NOT NULL AND
Insulina IS NOT NULL AND
ToleranciaGlucosa IS NOT NULL AND
GlucosaPlasma IS NOT NULL AND
Trigliceridos IS NOT NULL AND
HistoriaFamiliar IS NOT NULL AND
ActividadFisica IS NOT NULL AND
HorasSueno IS NOT NULL AND
HorasSueno<=12 AND
ALT  IS NOT NULL AND
AST  IS NOT NULL AND
HOMA  IS NOT NULL AND
Diagnostico in ('Normal','Prediabetes')
```

Figura 14 *Limpieza de la información*

4.3.3. Generación del modelo de base de datos

El proceso de asociación de tablas se genera basado en el modelo de la base de datos, que se presenta en la figura 15, que fue diseñado según los requerimientos de los modelos que se utilizan en esta investigación.

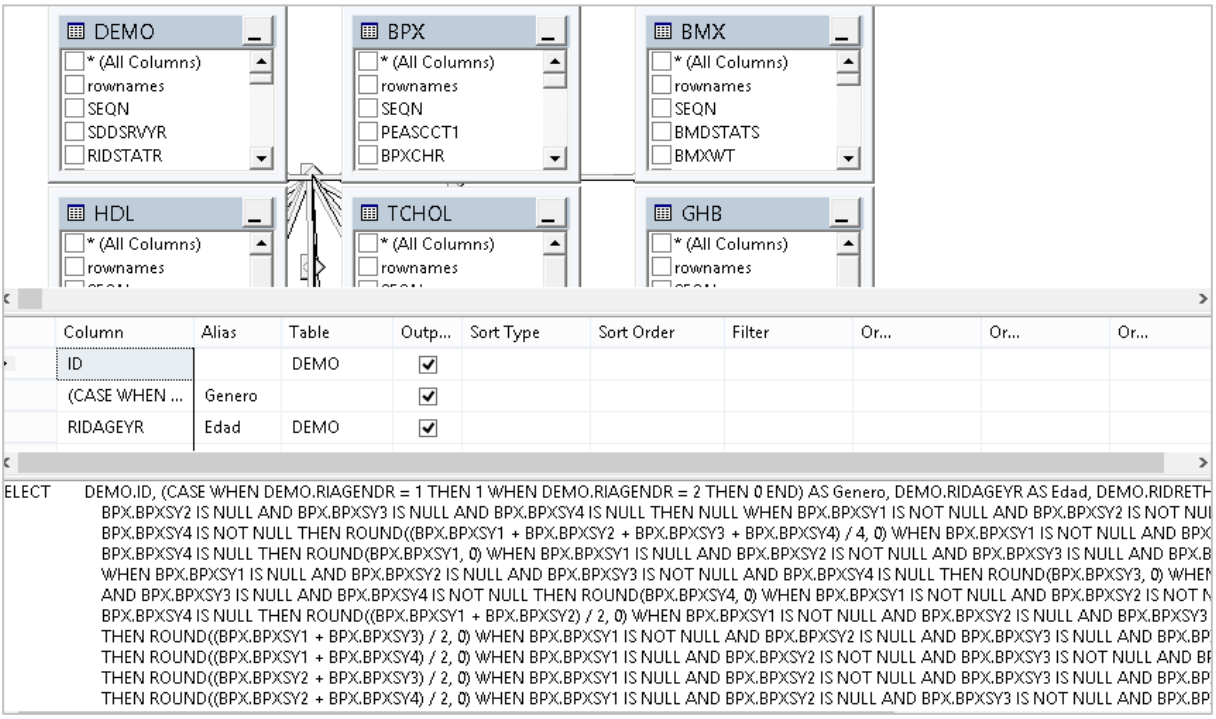


Figura 15 *Vista de asociación de tablas*

Para mejor comprensión, con base en la información desplegada en la vista anterior, se representa un modelo de base de datos, que se muestra en la figura 16.

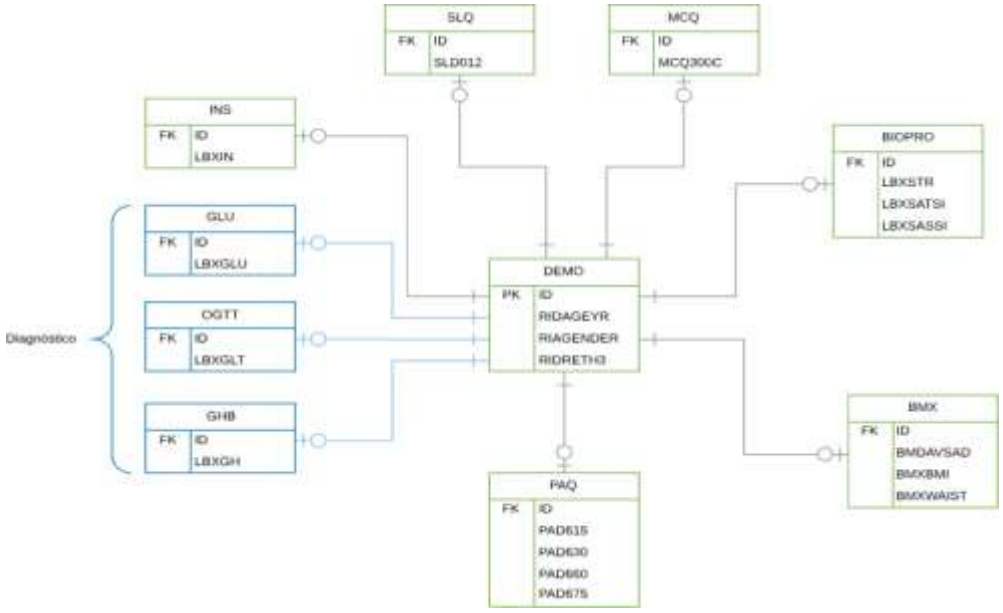


Figura 16 *Modelo de base de datos (tablas 2015-2016)*

4.3.4. Selección de variables

Una vez almacenada toda la información provista por la encuesta NHANES se procede a hacer una selección de las variables a utilizar. Tomando como base los factores de riesgo citados en el acápite 3.2.2. se seleccionan aquellos que se consideran, en base a la literatura médica investigada, mayormente relevantes para el desarrollo de la presente propuesta, como se aprecia en la tabla 2.

Factores de riesgo de prediabetes seleccionados	
Género	Hemoglobina glicosilada
Edad	Insulina
Etnia	Glucosa en plasma
Diámetro sagital abdominal	Tolerancia a la glucosa
Índice de masa corporal	Triglicéridos
Circunferencia abdominal	Transaminasas
Historia familiar	Índice HOMA
Actividad física	Horas de sueño

Tabla 2 *Factores de riesgo de prediabetes*

De éstos se toman los 3 principales factores que permiten, apoyados en los puntos de corte que la literatura médica establece, hacer un cribado de los datos, excluyendo aquellos casos en los que los índices de hemoglobina glicosilada, glucosa en plasma y tolerancia a la glucosa (2 horas), indican presencia de diabetes. Estos factores son los parámetros médicos que se utilizan para el diagnóstico de diabetes y prediabetes, los mismos que se exponen en la tabla 3.

Parámetro	Euglucemia	Prediabetes	Diabetes
Glucosa en plasma (mg/dL)	<100	100-125	>126
2-horas postprandial glucosa (mg/dL)	<140	140-199	>200
Hemoglobina glicosilada (%)	<5.7	5.7-6.4	>6.5

Tabla 3 *Parámetros para diagnóstico de prediabetes y diabetes*
American Diabetes Association (2015)

Este tamizaje se hace debido a que, el interés del estudio es contar con un set de datos que permita analizar los factores de riesgo que anteceden al inicio de la prediabetes. Adelantarse a la aparición de la prediabetes sería una importante arma para revertir a tiempo los factores que la desencadenan interviniendo oportunamente en el cambio de estilo de vida de los pacientes.

El índice HOMA no es una variable que se incluya en la base de datos que soporta este trabajo, pero, según la literatura médica y el criterio expertos en el área, es un factor de suma relevancia para el diagnóstico de la prediabetes; como es un índice resultante de un cálculo mediante la aplicación de una fórmula, se procede a incluir en el código de programación, para que la herramienta propuesta realice su cómputo de forma automática, facilitando al usuario la utilización de la misma.

$$HOMA = Insulina * Glucosa\ 405$$

Ecuación 1

Otro índice que calcula la herramienta es el índice de masa corporal IMC, que al igual que el anterior se calcula mediante una fórmula, dividiendo el peso para la estatura al cuadrado. Como el objetivo es ofrecer una herramienta sencilla, también el cálculo del IMC se incluyó en el código, de modo que el usuario ingrese únicamente información de su estatura y peso, y la herramienta calcule el índice.

$$IMC = \frac{Peso}{Estatura^2}$$

Ecuación 2

4.3.5. Codificación al formato requerido por SPSS

La codificación de la información ajusta la estructura de la data para que el programa SPSS pueda aplicar sobre ella un modelo de análisis factorial sin intervención adicional. Este proceso puede observarse en la figura 17:

Vista_2015_16.sql -...OP\Emilia Peña (52))*

```
SELECT
     DEMO.ID
    ,(CASE
        WHEN DEMO.RIAGENDR=1 THEN 1
        WHEN DEMO.RIAGENDR=2 THEN 0
    END) AS Genero
    ,DEMO.RIDAGEYR AS Edad
    ,DEMO.RIDRETH3 AS Etnia
    ,BMX.BMDAVSAD AS DiamSagitalAbdominal
    ,BMX.BMXBMI AS IMC
    ,BMX.BMXWAIST AS CircAbdonimal
    ,GHB.LBXGH AS Glicohemoglobina
    ,INS.LBXIN AS Insulina
    ,OGTT.LBXGLT AS ToleranciaGlucosa
    ,GLU.LBXGLU AS GlucosaPlasma
    ,BIOPRO.LBXSTR AS Trigliceridos
    ,(CASE
        WHEN MCQ.MCQ300C=1 THEN 1
        WHEN MCQ.MCQ300C=2 THEN 0
        ELSE NULL
    END) HistoriaFamiliar
    ,(CASE
        WHEN PAQ.PAD660>=15 OR PAQ.PAD675>=30 THEN 1
        WHEN PAQ.PAD660<15 OR PAQ.PAD675<30 THEN 1
    END) ActividadFisica
    ,(CASE
        WHEN SLQ.SLD012=77 THEN NULL
        WHEN SLQ.SLD012=99 THEN NULL
        ELSE SLQ.SLD012
    END) AS HorasSueno
    ,BIOPRO.LBXSATSI AS ALT
    ,(CASE
        WHEN BIOPRO.LBXSASSI=832 THEN NULL
        ELSE BIOPRO.LBXSASSI
    END) AS AST
    ,ROUND(INS.LBXIN*GLU.LBXGLU/405,2) AS HOMA
    ,(CASE
        WHEN GLU.LBXGLU>=125 THEN 'Diabetes'
        WHEN GHB.LBXGH>=6.5 THEN 'Diabetes'
        WHEN OGTT.LBXGLT>=200 THEN 'Diabetes'

        WHEN GLU.LBXGLU>=100 AND GLU.LBXGLU<125 THEN 'Prediabetes'
        WHEN GHB.LBXGH>=5.7 AND GHB.LBXGH<6.5 THEN 'Prediabetes'
        WHEN OGTT.LBXGLT>=140 AND OGTT.LBXGLT<200 THEN 'Prediabetes'

        WHEN GLU.LBXGLU<100 THEN 'Normal'
        WHEN GHB.LBXGH<5.7 THEN 'Normal'
        WHEN OGTT.LBXGLT<140 THEN 'Normal'

        ELSE NULL
    END) AS Diagnostico

FROM [dbo].[DEMO_I] AS DEMO
LEFT OUTER JOIN [dbo].[BPX_I] AS BPX
ON DEMO.ID=BPX.ID
LEFT OUTER JOIN [dbo].[BMX_I] AS BMX
ON DEMO.ID=BMX.ID
LEFT OUTER JOIN [dbo].[HDL_I] AS HDL
ON DEMO.ID=HDL.ID
LEFT OUTER JOIN [dbo].[TCHOL_I] AS TCHOL
ON DEMO.ID=TCHOL.ID
LEFT OUTER JOIN [dbo].[GHB_I] AS GHB
ON DEMO.ID=GHB.ID
LEFT OUTER JOIN [dbo].[INS_I] AS INS
ON DEMO.ID=INS.ID
LEFT OUTER JOIN [dbo].[OGTT_I] AS OGTT
```

Figura 17. *Codificación y procesamiento*

Este procedimiento se repite para todos los periodos que conforman la base de datos utilizada para el presente trabajo (2011-2012 y 2013-2014). Posteriormente, se describe la sentencia que procede a unir las vistas de los diferentes periodos, como muestra la figura 18.

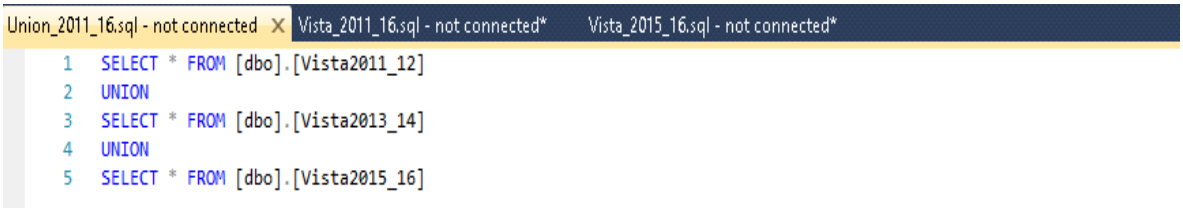


Figura 18. *Unión de vistas*

Debido al proceso anterior, el resultado del script de la Vista_ 2011_16 que corresponde a todos los años analizados tiene la facilidad de correr en SPSS el modelo que se mencionó anteriormente. Además, aquí se definen ciertos rangos pertinentes como edad, horas de sueño y diagnóstico.

4.4. Análisis Factorial con SPSS

Según los procedimientos realizados en conjunto con el equipo de investigación, el aporte de la investigadora consiste en el procesamiento estadístico mediante análisis factorial con el fin de obtener un índice único que mida el riesgo de padecer prediabetes, que servirá de insumo para posteriormente desarrollar una herramienta interactiva que permita conocer al usuario final este score. La información resultante luego de la codificación se carga en el programa SPSS para procesarla con la técnica estadística de análisis factorial, descrita ampliamente en el capítulo 3 del presente trabajo.

4.4.1. Vista de variables seleccionadas

En la figura 19, se presenta la vista de las variables definidas en la tabla 1 con sus características como, tipo de variable, extensión, etiqueta para su identificación posterior y su rol en el modelo. Estas variables son utilizadas para

ejecutar en varias ocasiones el modelo de análisis factorial, hasta definir cuáles de ellas son las que explican de mejor manera el fenómeno de la prediabetes y quedan definitivamente en el modelo.

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	Genero	Numérico	20	2	Genero	{,00, mujer}...	Ninguno	8	Derecha	Nominal	Entrada
2	Edad	Numérico	20	2	Edad	Ninguno	Ninguno	8	Derecha	Escala	Entrada
3	Etnia	Numérico	20	2	Etnia	Ninguno	Ninguno	8	Derecha	Nominal	Entrada
4	DiamSagital...	Numérico	20	2	DiamSagitalAb...	Ninguno	Ninguno	8	Derecha	Escala	Entrada
5	IMC	Numérico	20	2	IMC	Ninguno	Ninguno	8	Derecha	Escala	Entrada
6	CircAbdomi...	Numérico	20	2	CircAbdominal	Ninguno	Ninguno	8	Derecha	Escala	Entrada
7	Glicohemog...	Numérico	20	2	Glicohemoglobina	Ninguno	Ninguno	8	Derecha	Escala	Entrada
8	Insulina	Numérico	20	2	Insulina	Ninguno	Ninguno	8	Derecha	Escala	Entrada
9	ToleranciaG...	Numérico	20	2	ToleranciaGluc...	Ninguno	Ninguno	8	Derecha	Escala	Entrada
10	GlucosaPla...	Numérico	20	2	GlucosaPlasma	Ninguno	Ninguno	8	Derecha	Escala	Entrada
11	Triglicéridos	Numérico	20	2	Triglicéridos	Ninguno	Ninguno	8	Derecha	Escala	Entrada
12	HistoriaFam...	Numérico	20	2	HistoriaFamiliar	{,00, si}...	Ninguno	8	Derecha	Nominal	Entrada
13	ActividadFis...	Numérico	20	2	ActividadFisica	{,00, si}...	Ninguno	8	Derecha	Nominal	Entrada
14	HorasSueno	Numérico	20	2	HorasSueno	Ninguno	Ninguno	8	Derecha	Escala	Entrada
15	ALT	Numérico	20	2	ALT	Ninguno	Ninguno	8	Derecha	Escala	Entrada
16	AST	Numérico	20	2	AST	Ninguno	Ninguno	8	Derecha	Escala	Entrada
17	HOMA	Numérico	20	2	HOMA	Ninguno	Ninguno	8	Derecha	Escala	Entrada
18	DIAGNOSTI...	Numérico	20	2	DIAGNOSTICO	{,00, sano}...	Ninguno	8	Derecha	Nominal	Entrada

Figura 19. *Vistas de variables*

4.4.2. Vista de datos

La figura 20 presenta una vista de los datos que se cargaron en el programa SPSS antes de ejecutar el análisis factorial. Esta base consta de 1334 registros completos que corresponden al mismo número de personas encuestadas con sus respectivas variables. La base completa que se descargó del NHANES y antes de realizarse los procesos de limpieza y filtrado, constaba de 29902 registros.

	Genero	Edad	Etnia	DiamSagitalAbdominal	IMC	CircAbdominal	Glicohemoglobina	Insulina	ToleranciaGlucosa	GlucosaPlasma	Triglicéridos	HistoriaFamiliar	ActividadFisica	HorasSueno	ALT
1	mujer	26,00	3,00	14,50	20,20	73,70	5,20	3,05	99,00	89,00	24,00	si	no	8,00	22,00
2	hombre	52,00	6,00	22,30	23,60	90,30	5,80	6,00	100,00	110,00	93,00	no	no	7,00	20,00
3	mujer	67,00	6,00	29,20	36,30	117,80	5,90	20,93	164,00	107,00	87,00	si	si	7,00	73,00
4	hombre	43,00	3,00	25,30	28,90	102,80	4,90	3,24	95,00	90,00	312,00	si	no	8,00	33,00
5	mujer	54,00	3,00	21,80	32,70	107,80	5,50	7,36	80,00	95,00	77,00	si	no	6,00	30,00
6	mujer	36,00	1,00	20,20	27,30	91,10	5,00	9,86	91,00	88,00	87,00	si	no	7,00	48,00
7	mujer	57,00	3,00	27,10	37,80	113,80	5,50	9,61	120,00	96,00	226,00	si	no	8,00	18,00
8	hombre	25,00	4,00	16,70	21,00	73,20	5,50	4,47	84,00	85,00	39,00	no	no	6,00	16,00
9	hombre	35,00	3,00	29,00	33,00	109,60	4,70	8,02	77,00	108,00	94,00	si	no	7,00	18,00
10	hombre	58,00	3,00	25,60	32,60	113,20	5,50	16,01	198,00	106,00	101,00	no	no	7,00	22,00
11	hombre	26,00	4,00	14,90	19,20	72,50	5,30	2,57	145,00	89,00	29,00	si	no	9,00	14,00
12	mujer	29,00	1,00	24,80	37,50	125,30	5,40	19,20	78,00	88,00	76,00	no	no	8,00	11,00
13	mujer	21,00	1,00	19,00	27,30	89,50	5,00	14,50	185,00	90,00	87,00	no	no	8,00	13,00
14	mujer	39,00	1,00	20,50	23,40	90,00	4,80	1,63	120,00	83,00	38,00	no	no	6,00	29,00
15	hombre	62,00	3,00	16,40	19,10	70,40	5,40	6,69	85,00	90,00	123,00	si	no	8,00	15,00
16	hombre	26,00	3,00	19,70	24,90	86,40	5,20	4,02	99,00	88,00	36,00	si	no	8,00	20,00
17	mujer	46,00	2,00	17,80	23,00	97,90	5,30	7,54	92,00	83,00	60,00	no	no	5,00	17,00
18	hombre	56,00	3,00	23,00	28,60	99,60	5,10	9,82	136,00	107,00	321,00	si	no	7,00	26,00
19	hombre	29,00	3,00	22,60	28,60	104,20	5,30	4,77	85,00	79,00	66,00	si	no	7,00	19,00
20	hombre	43,00	1,00	24,80	30,70	110,30	5,30	9,59	96,00	107,00	179,00	no	no	7,00	69,00
21	hombre	51,00	6,00	18,70	22,20	86,00	5,60	3,01	96,00	99,00	113,00	si	no	7,00	16,00

Figura 20. *Vistas de datos cargados*

4.4.3. Proceso de análisis factorial

Entrando ya en el proceso de análisis factorial, el primer paso es ir a la opción “analizar”, luego se escoge “reducción de dimensiones” y por último la opción “factor”, como se muestra en la figura 24. Posteriormente, en el cuadro de diálogo “Análisis Factorial”, se introducen las siete variables que se escogieron finalmente, tras ejecutar varias veces el modelo, para realizar el análisis factorial.

La figura 21 evidencia las siete variables cuantitativas definidas, entre ellas la variable horas de sueño, que a pesar de que no es significativa en el modelo estadístico, se la incluye por ser un factor que actualmente despierta interés y es objeto de estudio dentro de la comunidad médica. Valenza (21012). A continuación, se listan las variables escogidas:

- Edad
- Diámetro sagital abdominal
- Índice de masa corporal
- Circunferencia abdominal
- Insulina
- Horas de sueño
- HOMA

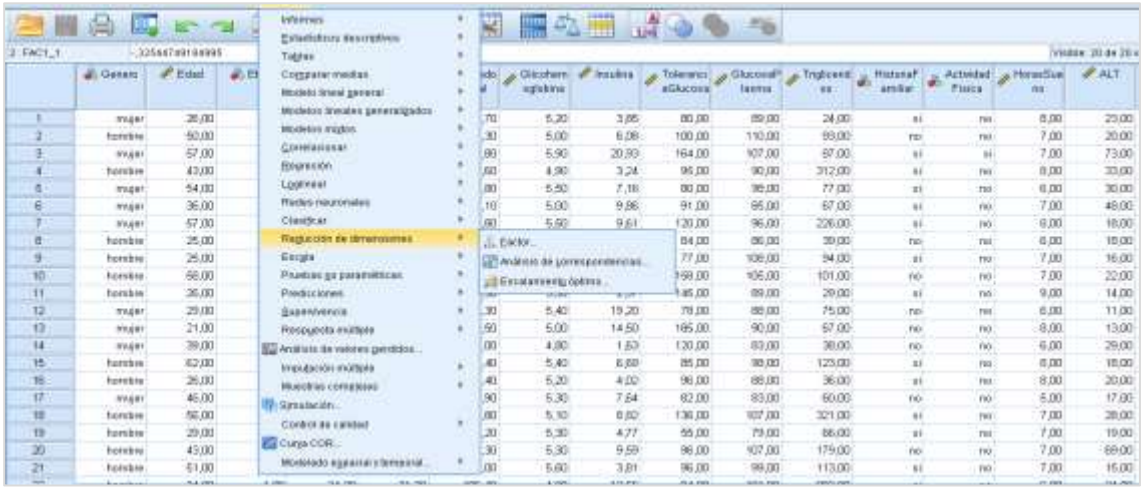


Figura 21. *Reducción de dimensiones*

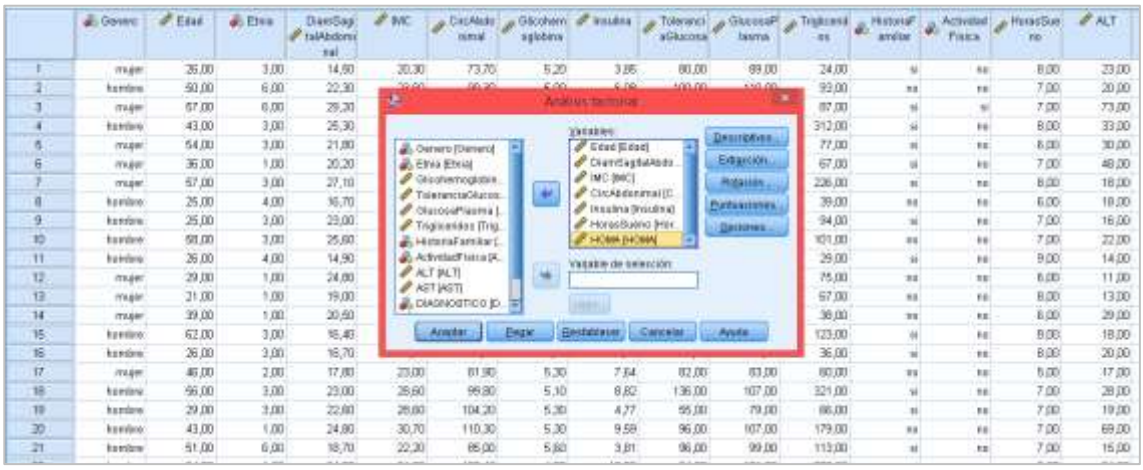


Figura 22. *Variables definidas*

4.4.4. Parametrización del modelo

Existen varias opciones de parametrización antes de ejecutar el análisis factorial. Las que se utilizan en esta propuesta son las que se detallan en las figuras que se explican a continuación:

Descriptivos: en estadísticos se escoge “solución inicial”, en la Matriz de correlaciones se especifican el índice KMO y la prueba de esfericidad de Barllett, como muestra la figura 23.

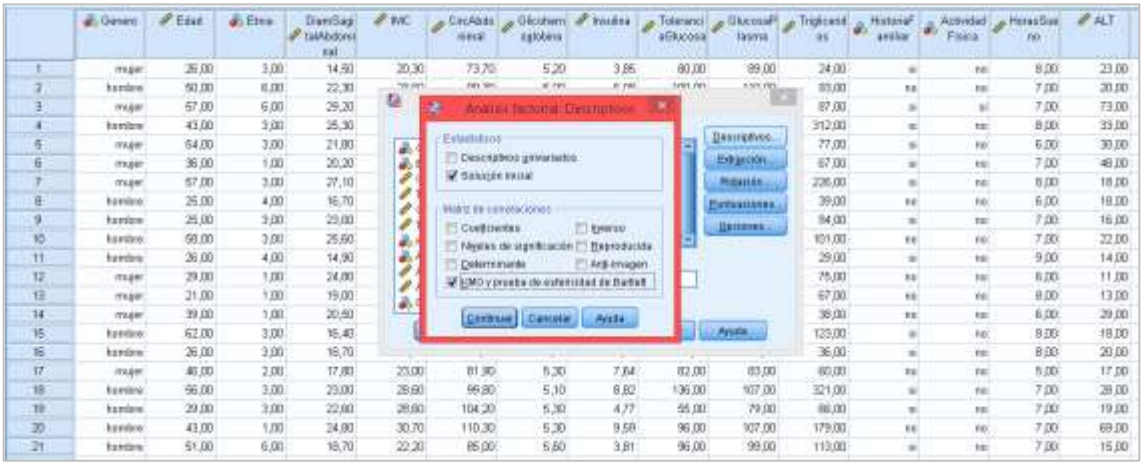


Figura 23. *Descriptivos*

Extracción: Para la extracción se establece el método por “Componentes Principales” y en analizar se escoge “matriz de correlaciones”; en la opción mostrar se elige “solución factorial sin rotar” y además el “gráfico de sedimentación”. Para la extracción de los factores se opta por “autovalores mayores que 1” y se deja la opción por defecto de 25 iteraciones. La figura 24 muestra estas configuraciones del sistema.

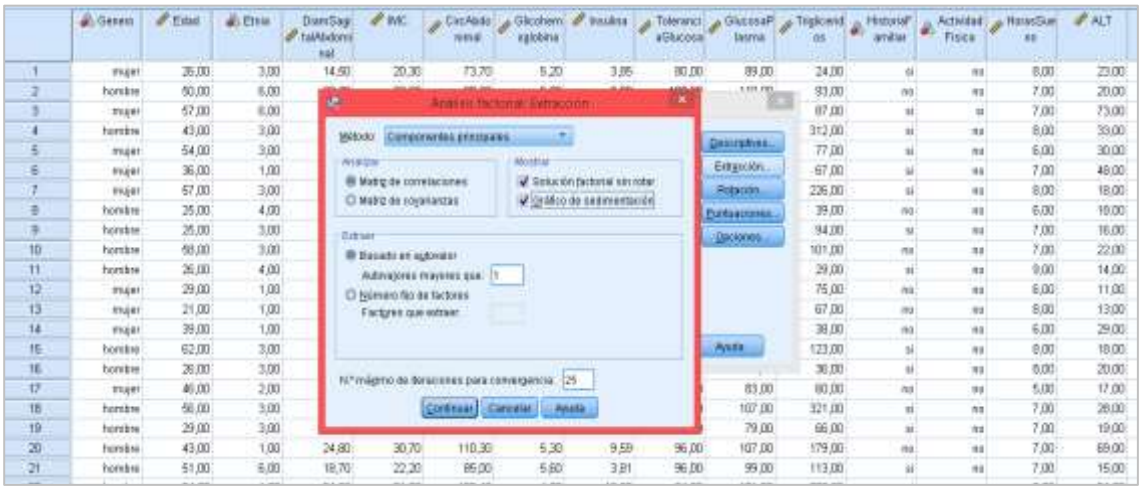


Figura 24. *Extracción*

Rotación: Como se aprecia en la figura 25, el método de rotación de factores escogido es el método ortogonal “Varimax” y en la opción mostrar se elige “solución rotada”.

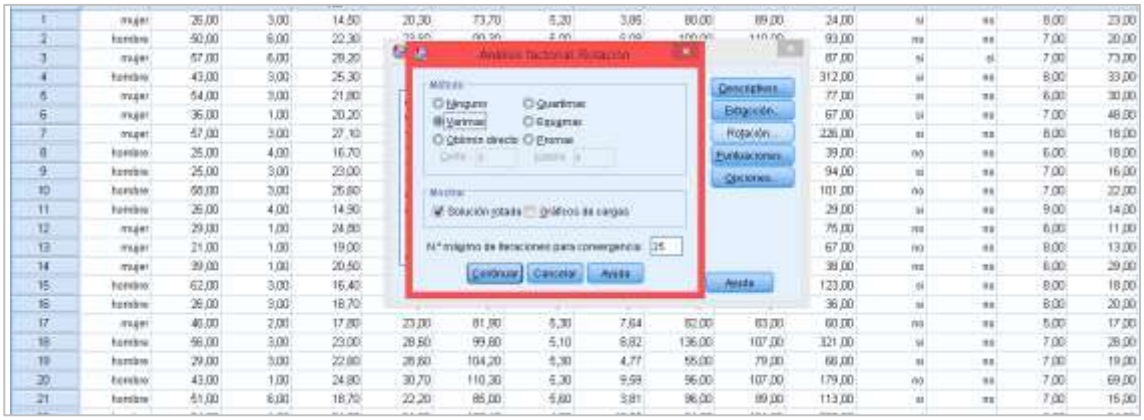


Figura 25. *Rotación de factores*

Puntuaciones factoriales: el método elegido para calcular las puntuaciones factoriales es la “regresión”, las mismas que se muestran en la matriz de coeficientes de puntuaciones factoriales, que se aprecia en la figura 26.

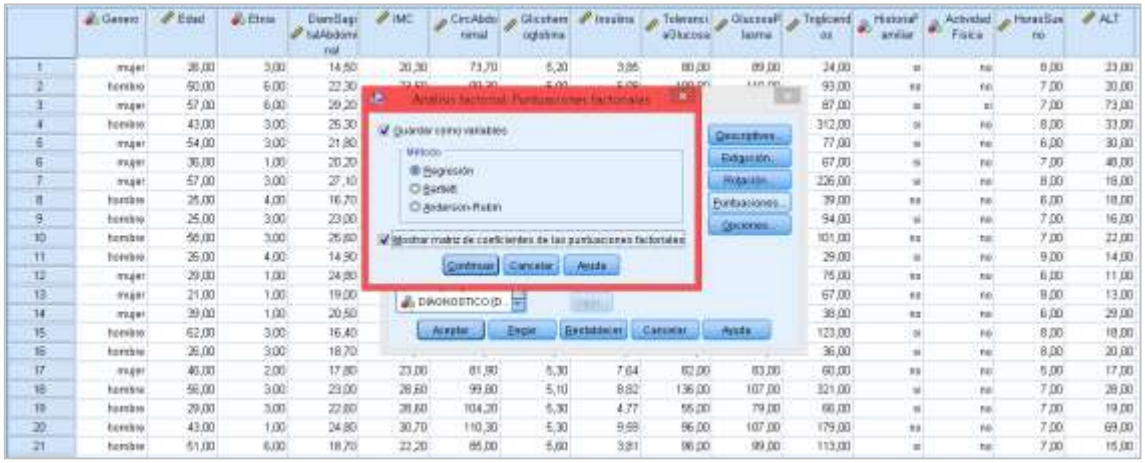


Figura 26. *Puntuaciones factoriales*

Terminada la parametrización, se ejecuta el análisis factorial, obteniendo una salida que se muestra y explica en el apartado 5.5.

4.5. Obtención del modelo de Análisis Factorial con SPSS

4.5.1. Pruebas de KMO y Barlett

El primer análisis que se puede apreciar en la salida del modelo son las pruebas de KMO y Barlett, que como se explica en el capítulo 4, apartado 4.2.3. son índices que ayudan a determinar si se debe o no aplicar el análisis factorial.

En este caso, los resultados obtenidos son un KMO de 0,76 con su aproximación, siendo mayor a 0.75 que, según la literatura revisada, es un resultado muy bueno. Respecto a la prueba de Barlett, la salida arroja un valor de 0.00, siendo menor al p valor propuesto como aceptable que es menor a 0.05.

Con estos resultados, que se pueden apreciar en la figura 27, se concluye que es correcto aplicar el modelo de Análisis Factorial en la presente investigación.

Análisis factorial

Prueba de KMO y Bartlett

Medida Kaiser-Meyer-Olkin de adecuación de muestreo		,757
Prueba de esfericidad de Bartlett	Aprox. Chi-cuadrado	11766,463
	gl	21
	Sig.	,000

Figura 27. *Índice KMO y test de Barlett*

4.5.2. Matriz de varianza total explicada

La figura 28 muestra la matriz de la varianza total explicada. Los dos primeros factores explican en este caso, el 71,8% de la varianza. El factor uno, explica él solo, el 56.1 % de la varianza total. El método utilizado para la extracción de factores es Componentes Principales.

Varianza total explicada

	Autovalores iniciales			Sumas de extracción de cargas al cuadrado			Sumas de rotación de cargas al cuadrado		
Componente	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	3,924	56,064	56,064	3,924	56,064	56,064	3,851	55,016	55,016
2	1,103	15,755	71,819	1,103	15,755	71,819	1,176	16,803	71,819
3	,991	14,163	85,982						
4	,830	11,852	97,834						
5	,090	1,290	99,124						
6	,053	,755	99,879						
7	,008	,121	100,000						

Método de extracción: análisis de componentes principales.

Figura 28. *Varianza total explicada*

4.5.3. Criterio para extracción de factores

Al escoger la opción "autovalores mayores que 1" para la extracción de los factores, el programa arroja dos factores que cumplen con esta condición. En el gráfico de sedimentación, figura 29, se observa claramente que los dos primeros factores cumplen con la condición de que sus autovalores sean mayores a 1.

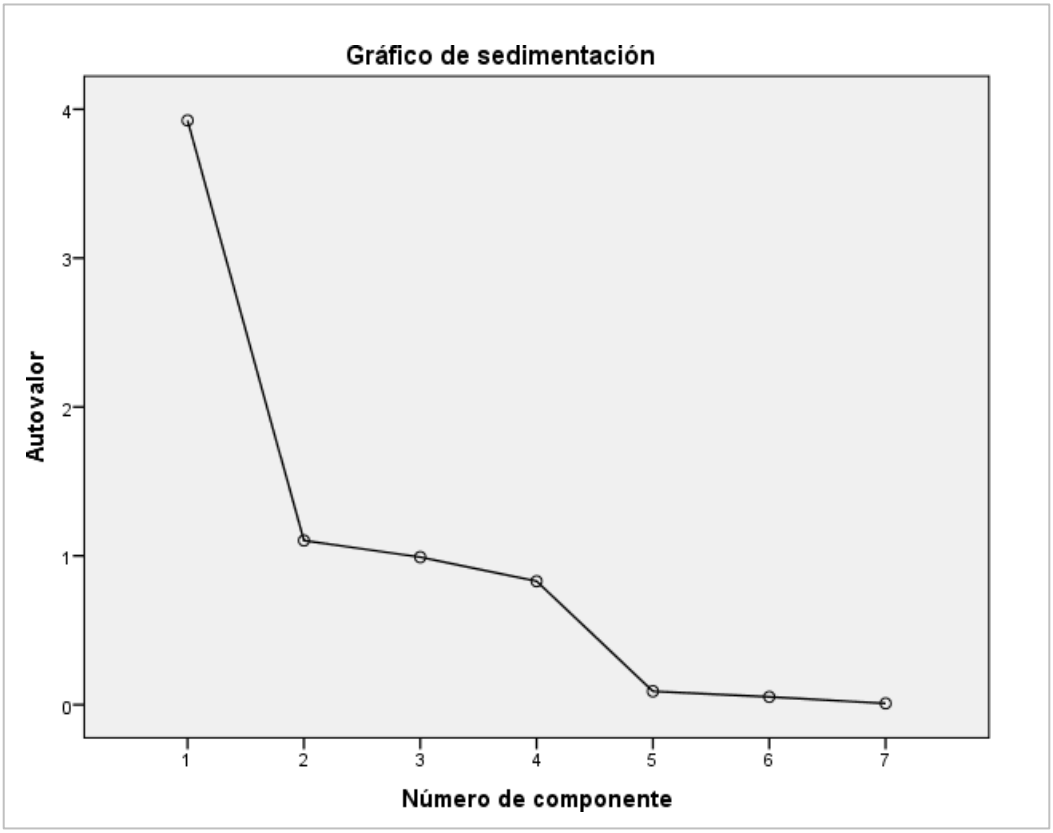


Figura 29. *Gráfico de sedimentación*

4.5.4. Matriz de componentes y matriz de componentes rotados

La figura 30 exhibe la matriz de componentes y la matriz de componentes rotados. Esta última es la que se analiza finalmente ya que es de fácil interpretabilidad

Matriz de componente[a]

	Componente 1	2
DiamSagitalAbdominal	,924	,212
CircAbdonimal	,924	,205
IMC	,905	,165
HOMA	,831	-,377
Insulina	,830	-,393
Edad	,110	,802
HorasSueno	-,074	-,221

Método de extracción: análisis de componentes principales.

a. 2 componentes extraídos.

Matriz de componente rotado[a]

	Componente 1	2
Insulina	,883	-,254
HOMA	,881	-,238
CircAbdonimal	,879	,352
DiamSagitalAbdominal	,878	,358
IMC	,866	,309
Edad	-,021	,809
HorasSueno	-,037	-,230

Método de extracción: análisis de componentes principales.
Método de rotación: Varimax con normalización Kaiser.[a]

Figura 30. *Matriz de componentes rotados*

4.5.5. Matriz de coeficiente de puntuación de componentes

Finalmente, utilizando la matriz de componentes rotado, el programa extrae la matriz de coeficiente de puntuación de los componentes, la misma que arroja las ponderaciones de cada variable dentro de cada uno de los dos factores extraídos. La figura 31 permite visualizar las puntuaciones o pesos de las variables para el factor 1 y para el factor 2 respectivamente.

Componente 1

Matriz de coeficiente de puntuación de componente

	Componente 1	2
Edad	-,090	,722
DiamSagitalAbdominal	,202	,227
IMC	,203	,185
CircAbdonimal	,202	,222
Insulina	,266	-,317
HorasSueno	,014	-,201
HOMA	,264	-,303

Método de extracción: análisis de componentes principales.
Método de rotación: Varimax con normalización Kaiser.
Puntuaciones de componente.

Figura 31. *Matriz de coeficiente de puntuación de componentes*

4.5.6. Determinación del componente que se utilizará para extraer la fórmula para el cálculo del score

Con base en los resultados arrojados por el modelo de análisis factorial ejecutado en el programa SPSS, se define la ecuación que servirá de base para calcular el score de riesgo de prediabetes, objeto primordial del presente trabajo. Analizando los dos componentes resultantes del análisis factorial, se decide escoger el componente 1 que es el que mejor explica el fenómeno de la prediabetes con el menor número de variables.

La ecuación se extrae, con base en los pesos que cada variable tiene dentro del componente elegido, evidenciados en la tabla 4, quedando expresada de la siguiente manera:

$$spd = \beta_1 \times EDD + \beta_2 \times DSA + \beta_3 \times IMC + \beta_4 \times CAB + \beta_5 \times INS + \beta_6 \times HSN + \beta_7 \times HOMA$$

donde:

Variable	Codificación	Peso
Edad	EDD	-0,09
DiamSagitalAbdominal	DSA	0,202
IMC	IMC	0,203
CircAbdonimal	CAB	0,202
Insulina	INS	0,266
HorasSueño	HSN	0,014
HOMA	HOMA	0,264

Tabla 4 *Variables de la ecuación con sus pesos*

4.5.7. Determinación de puntos de corte del score de prediabetes (spd)

Luego se incluye un análisis exploratorio del componente elegido, en este caso el componente 1, y se obtienen los valores de los cuartiles, como muestra la figura 32, los mismos que sirven para definir los puntos de corte o de inflexión del score obtenido.

Percentiles

		Percentiles						
		5	10	25	50	75	90	95
Promedio ponderado (Definición 1)	REGR factor score 1 for analysis 1	-1,1752036	-1,0263757	-,6880949	-,1789525	,4512464	1,2733399	1,7307290
Bisagras de Tukey	REGR factor score 1 for analysis 1			-,6874965	-,1789525	,4508577		

Figura 32. *Medidas de posición*

Con estos valores de los puntos de corte se establece la escala de riesgo de desarrollar prediabetes, para la herramienta que se propone como aplicación práctica del análisis factorial; estos puntos de corte definen cuatro posibles escenarios al momento que el usuario ingresa los valores solicitados por la aplicación:

- Riesgo bajo de padecer prediabetes: **score ≤ -0,68800949**
- Riesgo medio de padecer prediabetes:**-0.6880949< score ≤ -0,1789525**
- Riesgo medio-alto de padecer prediabetes: **-0,1789525< score≤ 0,4512464**
- Riesgo alto de padecer prediabetes: **score > 0,4512464**

4.6. Diseño de la visualización interactiva con Brackets en HTML

Todo trabajo de investigación, en última instancia, pretende contribuir de manera práctica a resolver el problema planteado; por este motivo, al finalizar el análisis factorial, se propone una herramienta basada en la ecuación obtenida y

explicada en el acápite 4.5.6., que está direccionada a cualquier usuario que desee conocer el posible riesgo de desarrollar prediabetes.

La herramienta de visualización se diseña con el programa Brackets utilizando los lenguajes HTML y JavaScript. El código utilizado, que se muestra parcialmente como ejemplo en la figura 34 y de forma completa en el anexo II, permite crear varios cuadros de texto donde el usuario debe ingresar los datos que se le solicita (basados en las variables ingresadas al modelo). Las variables categóricas actividad física y antecedentes familiares de diabetes no ingresaron al modelo de análisis factorial; sin embargo, las respuestas de los usuarios respecto a éstas, se las toma en consideración en la herramienta para hacer recomendaciones de prevención válidas.

En la figura 35 se exhibe la herramienta de visualización propuesta, donde se puede ver la estructura de la misma. En un primer plano se explica qué es la prediabetes, luego se cuestiona al usuario si desea conocer su riesgo de padecer la enfermedad, finalmente se le solicita ingresar 10 datos relacionados con datos antropométricos, valores de pruebas de laboratorio, actividad física y antecedentes familiares de diabetes. Para mejor comprensión de la información requerida, en un recuadro se presenta de forma sencilla la manera de calcular ciertos parámetros que podían no ser de conocimiento general.

Con los datos proporcionados por el usuario (ver ejemplo en la figura 35), la herramienta calcula el score de riesgo, normalizando primero los valores y aplicando los coeficientes obtenidos en el análisis factorial; la ecuación de normalización se muestra en la figura 33.

Al pulsar el botón “calcular índice” el instrumento devuelve un mensaje indicando el valor obtenido y su interpretación, suministrando además una sugerencia respecto a actividad física y necesidad de acudir a chequeo médico. Las posibles respuestas que la herramienta ofrece al usuario se resumen en la tabla 5.

$$z = \frac{X_I - \bar{x}}{s}$$

Figura 33. *Ecuación de normalización*

Condición	Respuesta
índice<=-0.6880949	"Felicidades, tienes un riesgo bajo de padecer prediabetes."
índice>-0.6880949 & índice<=- 0.1789525 & actividad física="SI" & historia familiar="NO"	"Tienes riesgo medio de padecer prediabetes. Por favor controla tus índices anualmente."
índice>-0.6880949 & índice<=- 0.1789525& actividad física ="NO" & historia familiar ="NO"	"Tienes riesgo medio de padecer prediabetes. Por favor controla tus índices anualmente y realiza 30 minutos diarios de actividad física moderada."
índice>-0.6880949 & índice<=- 0.1789525 & actividad física ="SI" & historia familiar ="SI"	"Tienes riesgo medio de padecer prediabetes. Por favor controla tus índices semestralmente."
índice>-0.6880949 & índice<=- 0.1789525 & actividad física ="NO" & historia familiar ="SI"	"Tienes riesgo medio de padecer prediabetes. Por favor controla tus índices semestralmente y realiza 30 minutos diarios de actividad física moderada."
índice>-0.1789525 & índice<=0.4512464 & actividad física ="SI" & historia familiar ="NO"	"Tienes riesgo medio alto de padecer prediabetes. Por favor considera realizarte un chequeo médico."
índice>-0.1789525 & índice<=0.4512464 & actividad física ="NO" & historia familiar ="NO"	Tienes riesgo medio alto de padecer prediabetes. Por favor considera realizarte un chequeo médico y realiza 30 minutos diarios de actividad física moderada."
índice>1.15 & índice<=2.89 & actividad física ="SI" & historia familiar ="SI"	"Tienes riesgo medio alto de padecer prediabetes. Por favor considera realizarte un chequeo médico y comentar tu historia familiar de diabetes."
índice>-0.1789525 & índice<=0.4512464 & actividad física ="NO" & historia familiar ="SI"	"Tienes riesgo medio alto de padecer prediabetes. Por favor considera realizarte un chequeo médico comentando tu historia familiar de diabetes y realiza 30 minutos diarios de actividad física moderada."
índice>-0.1789525 & índice<=0.4512464 & actividad física ="NO" & historia familiar ="SI"	"Tienes riesgo medio alto de padecer prediabetes. Por favor considera realizarte un chequeo médico comentando tu historia familiar de diabetes y realiza 30 minutos diarios de actividad física moderada."
índice>0.4512464 & historia familiar ="SI"	"Tienes riesgo alto de padecer prediabetes. Por favor acude a un especialista y comenta tu historia familiar de diabetes."
índice>0.4512464	"Tienes riesgo alto de padecer prediabetes. Por favor acude a un especialista."

Tabla 5 *Interpretaciones del índice*

```
107             if(indice<=-0.6888949)
108              {
109                 console.log("3");
110
111                 document.getElementById("resultado").innerHTML="Tu indice es de "+indice;
112
113                 document.getElementById("diagnostico").innerHTML="Felicidades, tienes un riesgo bajo de padecer prediabetes.";
114              }
115
116             else if(indice>-0.6888949 & indice<=-0.1789525 & field8=="SI" & field9=="NO")
117              {
118                 console.log("4");
119
120                 document.getElementById("resultado").innerHTML="Tu indice es de "+indice;
121
122                 document.getElementById("diagnostico").innerHTML="Tienes riesgo medio de padecer prediabetes. Por favor
                    controla tus indices anualmente.";
123              }
124
125             else if(indice>-0.6888949 & indice<=-0.1789525 & field8=="NO" & field9=="NO")
126              {
127                 console.log("5");
128
129                 document.getElementById("resultado").innerHTML="Tu indice es de "+indice;
130
131                 document.getElementById("diagnostico").innerHTML="Tienes riesgo medio de padecer prediabetes. Por favor
                    controla tus indices anualmente y realiza 30 minutos diarios de actividad fisica moderada.";
132              }
133
134             else if(indice>-0.6888949 & indice<=-0.1789525 & field8=="SI" & field9=="SI")
135              {
136                 console.log("6");
137
138                 document.getElementById("resultado").innerHTML="Tu indice es de "+indice;
139
140                 document.getElementById("diagnostico").innerHTML="Tienes riesgo medio de padecer prediabetes. Por favor
                    controla tus indices semestralmente.";
```

Figura 34. *Código HTML y JavaScript*

Figura 35. *Estructura de la herramienta de visualización*

4.7. Evaluación de la metodología propuesta

En términos generales, el experimento realizado se puede resumir en los pasos que se detallan a continuación. Se capturó la información mediante un proceso de ETL, desde la base de datos de la encuesta NHANES, ingresando a la página web de la misma, mediante el software RProject; posteriormente utilizando el paquete *Foreing*, se descargaron los datos en formato de STATA y con el paquete RODBC

se guardó la información en SQL Server, generando un *Datawarehouse* o almacén de datos.

Una vez almacenada la información en el Data *Warehouse*, se procedió a procesarla, realizando un proceso de limpieza profunda de las tablas que permita garantizar la completitud y comprensión de los datos. Posteriormente, se crearon una serie de vistas que se ajustan al modelo de base de datos planteado para este proyecto de investigación y que juntan las tablas de dimensiones con la tabla de hechos.

Terminado este proceso, se creó una clave primaria que permita la integración de la data correspondiente al resto de años a analizarse en una sola tabla que servirá para alimentar de manera directa el software estadístico. Los procesos antes descritos se desarrollaron en conjunto con los miembros del equipo investigador perteneciente al grupo liderado por Danilo Esparza, PhD de la Universidad de Las Américas en Quito-Ecuador.

Posteriormente, se codificaron los datos al formato requerido por el software SPSS permitiendo ingresar la información de manera apropiada al programa estadístico, iniciando así una serie de procesos que, en su conjunto, conforman mi aporte individual al presente trabajo de investigación. La información resultante se cargó en el programa SPSS para procesarla con la técnica estadística de análisis factorial. Las variables definidas con anticipación fueron utilizadas para ejecutar en varias ocasiones el modelo de análisis factorial, conservando aquellas que explicaban satisfactoriamente la prediabetes.

El siguiente paso fue realizar la parametrización del modelo, escogiendo los descriptivos: índice KMO y prueba de esfericidad de Barlett, cuyos valores permitieron decidir si era aplicable o no el análisis factorial. Para la extracción de factores se utilizó el método por “Componentes Principales” y para la selección del número de factores se eligió el método de autovalores mayores a 1.

Se escogieron los dos primeros factores que explicaban el 71,8% de la varianza y que cumplían con la condición del autovalor mayor a 1. Finalmente, el programa extrajo la matriz de coeficiente de puntuación de los componentes, con las ponderaciones de cada variable, es decir con los pesos de cada variable en cada factor.

Se optó por el factor 1 porque era el que explicaba el 56.1 % de la varianza total. Con las puntuaciones de las variables de este factor se elaboró la ecuación que sirvió para calcular el score de riesgo de prediabetes y con base en éste se construyó la herramienta de visualización que permitirá al usuario, respondiendo 10 preguntas, conocer su índice de riesgo y las respectivas recomendaciones. Los puntos de corte del índice fueron los valores de los cuartiles del componente escogido, siendo el valor del primer cuartil el riesgo más bajo y los valores sobre el valor del tercer cuartil el riesgo más alto de desarrollar prediabetes.

Se realizó luego un análisis exploratorio del componente 1 arrojado por el análisis factorial, en SPSS, factor con el que se trabaja para calcular el score de riesgo de desarrollar prediabetes. En este análisis se puede observar que el factor posee una media igual a 0,00 y una desviación estándar de 1. Además, se pueden ver los valores mínimos y máximos, que en este caso son – 1,81761 y 8,13055 respectivamente, como se observa en la Figura 36.

Descriptivos

			Estadístico	Error estándar
REGR factor score 1 for analysis 1	Media		,0000000	,02737928
	95% de intervalo de confianza para la media	Límite inferior	-,0537112	
		Límite superior	,0537112	
	Media recortada al 5%		-,0831332	
	Mediana		-,1789525	
	Varianza		1,000	
	Desviación estándar		1,00000000	
	Mínimo		-1,81761	
	Máximo		8,13055	
	Rango		9,94815	
	Rango intercuartil		1,13934	
	Asimetría		1,921	,067
	Curtosis		8,003	,134

Figura 36. *Descriptivos del componente 1*

Si se analiza el valor de curtosis obtenido y la forma del histograma expuesto en la figura 37, se concluye que es una distribución con característica leptocúrtica, por lo tanto, permite predecir con facilidad, puesto que tiene un coeficiente de variabilidad bajo.

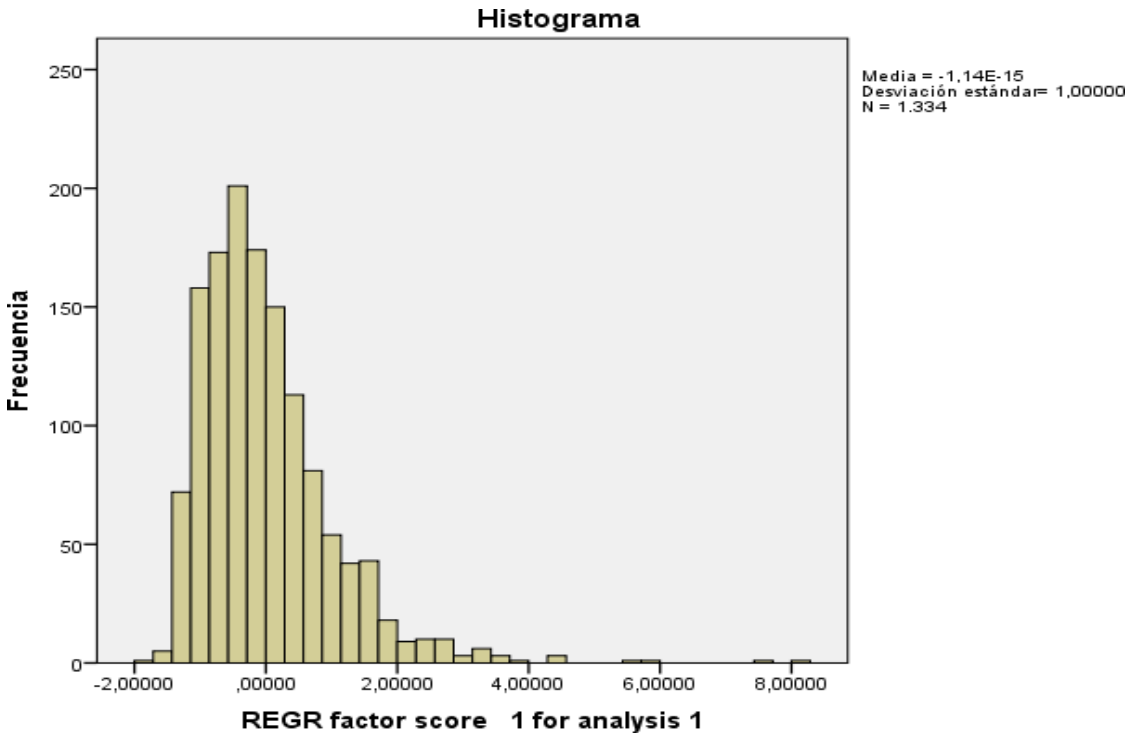


Figura 37. *Histograma*

Si bien los valores fluctúan entre – 1,81761 y 8,13055, se puede ver que la distribución se asemeja a una distribución normal con un rango aproximado entre -2 y + 2, ignorando los valores *outliers*, los mismos que se pueden identificar claramente en la figura 38; estos valores atípicos se los conservó para salvaguardar la confiabilidad de los datos originales.

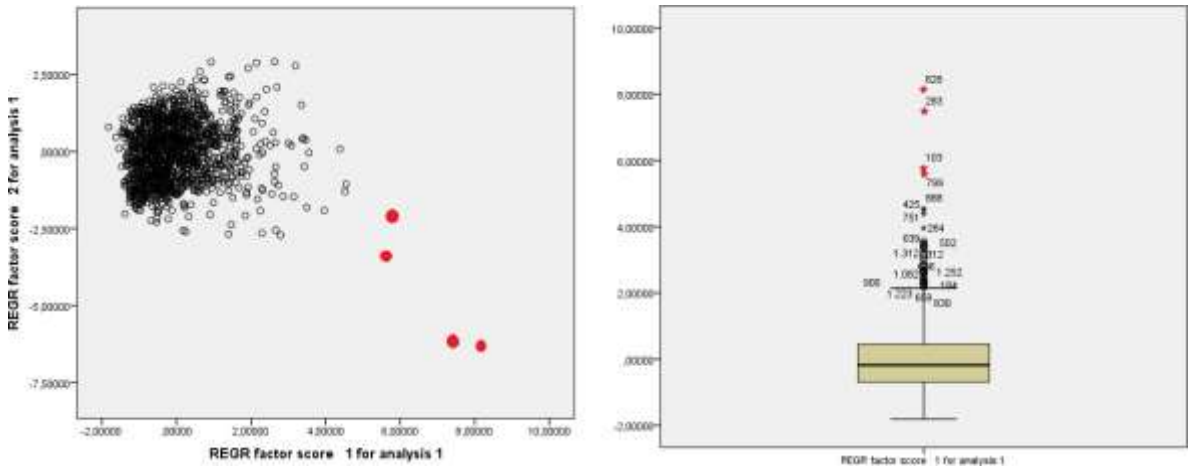


Figura 38. *Reconocimiento de outliers*

Para hacer una evaluación de la herramienta propuesta se utilizó una muestra de 25 casos reales con sus respectivas mediciones. La tabla 6 muestra los valores de las variables cuantitativas de 25 individuos que se ingresaron en el modelo estadístico propuesto, obteniendo índices que explican diferentes niveles de riesgo de padecer prediabetes.

El valor mínimo obtenido para los casos validados fue de – 1,18 y el máximo 2,56, valores que están dentro del rango arrojado por el análisis exploratorio del factor 1; así también la media de la muestra de validación es de 0,02 muy cercana a la media del componente 1 con valor 0 y error estándar de 0,2737928

Edad	DSA	IMC	Circ. Abdominal	Insulina	HorasSueno	HOMA	ScorePrediabetes	Riesgo
40	19,6	23,5	83,9	4,74	6	1,17	-0,71	BAJO
55	21,9	26,6	93,3	14,38	8	3,44	0,15	MEDIO ALTO
22	25,2	34,8	111,8	12,52	7	3,22	0,93	ALTO
20	17,5	21,9	82,7	6,15	5	1,43	-0,67	MEDIO
51	19,6	27,4	90,8	9,03	8	2,65	-0,22	MEDIO
36	18,8	26,7	91,5	13,94	8	2,99	0,04	MEDIO ALTO
20	30,4	40,2	119	29,26	7	7,51	2,56	ALTO
62	23,1	30,1	106,3	10,55	8	2,79	0,23	MEDIO ALTO
21	16,7	21,2	76,6	9,3	6	2,43	-0,58	MEDIO
33	21,5	25,8	95,7	14,45	7	3,75	0,32	MEDIO ALTO
52	18,5	24,3	87,7	9,55	7	2,26	-0,46	MEDIO
20	21,9	30,8	100,6	12,97	8	3,23	0,54	ALTO
41	17,1	21,7	72,3	6,75	9	1,62	-0,89	BAJO
44	30,1	40,6	132,9	11,78	8	3,2	1,46	ALTO
31	25	29,9	109	13,26	7,5	3,5	0,74	ALTO
59	18,7	22,2	82,5	2,41	6,5	0,57	-1,09	BAJO
40	19,7	25,6	94,5	7,69	9	2,13	-0,25	MEDIO
57	26,1	36,4	114,1	5,97	6,5	1,67	0,40	MEDIO ALTO
20	17	19,6	70,6	5,35	8	1,24	-0,94	BAJO
22	16	20,3	70,7	3,52	8,5	0,79	-1,10	BAJO
54	22,3	31,3	97,2	11,07	6,5	2,47	0,13	MEDIO ALTO
47	22,7	30	103	6,75	6	1,63	-0,03	MEDIO ALTO
50	24,9	33,5	103,4	7,21	6	1,5	0,18	MEDIO ALTO
50	15,9	20,5	73,6	4,65	8	1,08	-1,18	BAJO
22	23,8	32,7	109,7	16,16	9	3,55	0,95	ALTO

Tabla 6 *Muestra para validación*

La tabla 6 muestra el nivel de riesgo por individuo junto con el score y una alerta cromática dependiente de su ubicación en el espectro del índice.

CONCLUSIONES

En este capítulo se exponen las conclusiones obtenidas y futuras líneas de investigación. El presente estudio pretende contribuir eficazmente a la investigación de enfermedades crónicas degenerativas como la diabetes, adelantándose un paso a la aparición de la prediabetes, condición previa a la diabetes. Para ello plantea la creación de un score de riesgo de prediabetes y con base en éste, la construcción de una herramienta práctica para que el usuario conozca su estado de salud respecto al posible progreso de esta grave y complicada enfermedad, apoyando a los estudios sobre prediabetes desarrollados por los otros dos miembros que conforman el macroproyecto sobre prediabetes.

El fin último de estas investigaciones, es asegurar una contribución significativa a los programas de prevención de salud pública, aportando con alertas tempranas sobre el riesgo del desarrollo de la enfermedad, brindando sugerencias válidas referentes a la práctica de actividad física y visitas médicas preventivas.

Las organizaciones mundiales de salud exhortan a los gobiernos y entidades a difundir programas de salud preventiva, la herramienta propuesta en esta investigación tiene una aplicabilidad real y eficaz dentro de cualquier iniciativa de salud que se plantee en el marco de la prevención de enfermedades crónicas degenerativas como la prediabetes.

Tras la realización de este trabajo se han obtenido las siguientes conclusiones:

- Es viable generar un índice único de riesgo de prediabetes con el que se brinde alertas tempranas a la población.
- Es factible identificar de manera preventiva los factores de riesgo de la prediabetes e intervenir oportunamente en el estilo de vida de la población.
- La herramienta propuesta ofrece a los usuarios un método rápido y sencillo de valoración del riesgo de desarrollar prediabetes.
- Se puede ofrecer a los usuarios un índice de riesgo que los estimule a acudir a servicios de salud preventivos y cambios en sus rutinas diarias.
- Se puede aplicar la metodología propuesta en este trabajo para otros estudios relacionados con programas de salud preventiva.
- Los modelos basados en técnicas de análisis de datos son aplicables a la resolución de importantes problemas en el campo de la salud pública.

Lecciones aprendidas

El aprendizaje es una parte fundamental de cualquier trabajo de investigación; a continuación, se detallan algunas lecciones aprendidas durante este proceso:

- Para lograr obtener resultados confiables y aplicables, todo trabajo de investigación debe iniciar con una base de datos confiable y correctamente anonimizada, para ello se debe hacer un primer análisis de calidad de los datos.
- Cuando se trabaje con problemas relacionados con ciertas áreas específicas y el investigador no tenga el conocimiento suficiente es indispensable contar con el apoyo de expertos en la materia.
- Antes de iniciar el trabajo investigativo, al momento de plantear el problema, debe verificarse que exista suficiente respaldo teórico en el campo investigado.
- Si se trabaja con modelos estadísticos es extremadamente importante la elección del modelo a aplicar, especialmente hay que considerar las características de los datos con los que se trabajará y el objetivo del estudio.
- Las herramientas tecnológicas, en lo posible deberán ser de uso libre, y de

fácil integración entre ellas.

- Los resultados de la investigación deberán ser aplicables y que contribuyan al bienestar público.
- Si se genera una herramienta para uso del usuario final, ésta tiene que ser lo más amigable y sencilla.

Trabajo futuro

Con base en los resultados y conocimientos obtenidos de la presente investigación, se planteará a futuro la aplicación del modelo y de la herramienta propuestos a la población ecuatoriana, observando qué tan adaptables son a un entorno diverso. Los resultados que se obtengan de este primer proceso servirán para realizar los ajustes necesarios al modelo, de modo que la metodología se adapte a la realidad del medio estudiado.

Se considera que, de ser posible la aplicación del modelo al entorno ecuatoriano, se aportaría de forma eficaz a los programas de salud pública mediante campañas de prevención de la prediabetes y diabetes Mellitus.

El objetivo a futuro es realizar un trabajo conjunto aplicando las herramientas informáticas propuestas dentro del macroproyecto de prediabetes, con el fin de disminuir los niveles alarmantes de incremento de la enfermedad en el Ecuador, desplegando programas preventivos que apliquen los resultados obtenidos en los tres trabajos desarrollados.

Asimismo, se podría adecuar el modelo a programas de prevención de otras patologías con alto grado de morbimortalidad en la población ecuatoriana, como problemas coronarios y desórdenes del metabolismo, entre otros

REFERENCIAS

Alghamdi, M., Al-Mallah, M. H., Keteyian, S. J., Brawner, C. A., Ehrman, J. K., & Sakr, S. (2017). *Predicting diabetes mellitus using SMOTE and ensemble machine learning approach: The Henry Ford Exercise Testing (FIT) project.* PLoS ONE, 12(7), e0179805. https://doi.org/10.1371/journal.pone.0179805.

American Association of Clinical Edocrinologists. (s.f.). Screening and Monitoring of Prediabetes. Recuperado el 28 de noviembre de 2018 de http://outpatient.aace.com/prediabetes/screening-and-monitoring-prediabetes.

American Diabetes Association. (2018). Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes. *Diabetes Care*, 41(1), 513-524. doi: 10.2337/dc18-S002.

American Diabetes Association. (2015). Diagnosing Diabetes and Learning About Prediabetes Recuperado el 29 de noviembre de 2018 de: http://www.diabetes.org/are-you-at-risk/prediabetes/.

Appajigol J., Somannavar M., y Araganji R.(2015). Performance of diabetes risk scores with or without point of care blood glucose. *Journal of the Scientific Society*, 42.1 (January-April 2015): p24. From Academic OneFile.Medknow Publications and Media Pvt. Ltd. Recuperado el 30 de noviembre de 2018 de http://www.jscisociety.com/aboutus.asp.

Asociación Latinoamericana de Diabetes ALAD, (2009). Consenso de Prediabetes. *Documento de posición de la Asociación Latinoamericana de Diabetes* (ALAD) Recuperado el 29 de noviembre de 2018, de http://www.revistaalad.com/pdfs/0904_ConsPred.pdf

Brass, L. (2018, March-April). THE Invisible EPIDEMIC: Why Prediabetes is Sweeping the Country and How You Can Avoid It. *Vibrant Life*, 34(2), https://reunir.unir.net/bitstream/handle/123456789/16676/Eguiguren%20Eguiguren%2C%20Margarita%20Maria.pdf?sequence=1&isAllowed=y.

Costa, B., Cabré, J. J., Solà-Morales, O., Barrio, F., Piñol, J. L., Cos, X., ... & Sagarra, R. (2011). *Rendimiento del test FINDRISC en el cribado de la diabetes tipo 2 en atención primaria. Estudio transversal poblacional.* BMC Public Health, 11(1), 775. https://doi.org/10.1186/1471-2458-11-775

De La Fuente, S. (2011). Análisis Factorial. Facultad de Ciencias Económicas y Empresariales UAM. Recuperado el 30 de noviembre de 2018 de http://www.fuenterrebollo.com/Economicas/ECONOMETRIA/MULTIVARIANTE/FACT ORIAL/analisis-factorial.pdf

Díaz, O., Cabrera, E., Orlandi, N., Araña, M., y Díaz, O. (2011). Aspectos epidemiológicos de la prediabetes, diagnóstico y clasificación. *Revista Cubana de Endocrinología*, 22(1), 3-10. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1561-29532011000100003&lng=es&tlng=es.

Federación Internacional de la Diabetes FID, (2017). Atlas de la Diabetes de la FID-8va edición. Recuperado el 29 de noviembre de 2018, de: http://www.diabetesatlas.org/IDF_Diabetes_Atlas_8e_interactive_ES/

Garber, A., Handelsman, Y., Einhorn, D, Bergman, D., … , and Nesto, R. (2008). Diagnosis and Management of Prediabetes in the Continuum of Hyperglycemia—When Do the Risks of Diabetes Begin? A Consensus Statement From the American College of Endocrinology and the American Association of Clinical Endocrinologists. *Endocrine Practice*,14(7). doi: 10.4158/EP.14.7.933.

Glümer C, Vistisen D, Borch-Johnsen K, Colagiuri S; DETECT-2 Collaboration. Risk scores for type 2 diabetes can be applied in some populations but not all. Diabetes Care. 2006 Feb;29(2):410-4. doi: 10.2337/diacare.29.02.06.dc05-0945.

Hernández Sampieri, R., Fernández Collado, C., y Baptista Lucio, P. (2003). *Metodología de la Investigación*. Tercera Edición. Recuperado de https://investigar1.files.wordpress.com/2010/05/sampieri-hernandez-r-cap3- planteamiento-del-problema.pdf.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression* (3rd ed.). Wiley. https://doi.org/10.1002/9781118548387

Ibarra, A. (2001). *Análisis de las dificultades financieras de las empresas en una economía emergente: las bases de datos y las variables independientes en el sector hotelero de la bolsa mexicana de Valores*. (Tesis doctoral). Universitat Autónoma de Barcelona. https://ddd.uab.cat/record/38524

Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). *Machine learning and data mining methods in diabetes research.* Computational and Structural Biotechnology Journal, 15, 104–116. https://doi.org/10.1016/j.csbj.2016.12.005

Khetan AK, Rajagopalan S. Prediabetes. Can J Cardiol. 2018 May;34(5):615-623. doi: 10.1016/j.cjca.2017.12.030

Kolberg JA, Gerwien RW, Watkins SM, Wuestehube LJ, Urdea M. Biomarkers in Type 2 diabetes: improving risk stratification with the PreDx ® Diabetes Risk Score. Expert Rev Mol Diagn. 2011 Nov;11(8):775-92. doi: 10.1586/erm.11.63.

Lindström, J., & Tuomilehto, J. (2003). *The diabetes risk score: A practical tool to predict type 2 diabetes risk.* Diabetes Care, 26(3), 725–731. https://doi.org/10.2337/diacare.26.3.725.

Lindström, J., Peltonen, M., Eriksson, J. G., Ilanne-Parikka, P., Aunola, S., Keinänen-Kiukaanniemi, S., ... & Tuomilehto, J. (2006). *High-fibre, low-fat diet predicts long-term weight loss and decreased type 2 diabetes risk:*

The Finnish Diabetes Prevention Study. Diabetologia, 49(5), 912–920. https://doi.org/10.1007/s00125-006-0198-z.

López-Roldán, P.; Fachelli, S. (2016). Análisis factorial. En P. López-Roldán y S. Fachelli, *Metodología de la Investigación Social Cuantitativa.* Bellaterra (Cerdanyola del Vallès): Dipòsit Digital de Documents, Universitat Autònoma de Barcelona. 1ª edición, versión 3. https://ddd.uab.cat/pub/caplli/2016/163564/metinvsoccua_a2016_cap1-2.pdf

Mata-Cases, M., Artola, S., Escalada, J., Ezkurra-Loyola, P., Ferrer-García, J. C., Fornos, J. A., Girbés, J., Rica, I., & Grupo de Trabajo de Consensos y Guías Clínicas de la Sociedad Española de Diabetes (2015). Consenso sobre la detección y el manejo de la prediabetes. Grupo de Trabajo de Consensos y Guías Clínicas de la Sociedad Española de Diabetes [Consensus on the detection and management of prediabetes. Consensus and Clinical Guidelines Working Group of the Spanish Diabetes Society]. *Atencion primaria*, *47*(7), 456–468. https://doi.org/10.1016/j.aprim.2014.12.002.

Meng, X. H., Huang, Y. X., Rao, D. P., Zhang, Q., & Liu, Q. (2013). Comparison of three data mining models for predicting diabetes or prediabetes by risk factors. *The Kaohsiung journal of medical sciences*, *29*(2), 93–99. https://doi.org/10.1016/j.kjms.2012.08.016.

Organización Mundial de la Salud OMS / Organización Panamericana de la Salud, (2017). *La obesidad, uno de los principales impulsores de la diabetes.* Recuperado el 28 de diciembre de 2018 https://www.paho.org/hq/index.php?option=com_content&view=article&id=13918:obe sity-a-key-driver-of-diabetes&Itemid=1926&lang=es

Organización Mundial de la Salud OMS,(2016). *Informe Mundial sobre Diabetes.* Resumen de Orientación. Recuperado el 23 de diciembre de 2018 de: http://apps.who.int/iris/bitstream/handle/10665/204877/WHO_NMH_NVI_16.3_spa.pd

f;jsessionid=64F71579FB8A0488C112CD9C73F22157?sequence=1

Organización Mundial de la Salud, 1994. Prevención de la Diabetes Mellitus. *Informe de un Grupo de Estudio de Ia OMS.* Recuperado el 29 de noviembre de 2018 de: http://apps.who.int/iris/bitstream/handle/10665/41935/9243208446_es.pdf ;jsessionid= 6C3B8926B6879CEFC6DBF3AC977E46ED?sequence=1

Pérez, E., Medrano, L. (2010). Análisis Factorial Exploratorio: Bases Conceptuales y Metodológicas. *Revista Argentina de Ciencias del Comportamiento, 2* (1), 58-66. https://doi.org/10.32348/1852.4206.v2.n1.15924

Rosas-Saucedo, J., Caballero, E., Brito-Córdova, G., García, H., Costa, J., Lyra, R., y Rosas-Guzman, J. (2017). Consenso de Prediabetes. Documento de posición de la Asociación Latinoamericana de Diabetes (ALAD). *Revista de la ALAD, 7* (4), 186- 187. DOI: 10.24875/ALAD.17000307

Salinero-Fort, M. Á., Burgos-Lunar, C., Lahoz, C., Mostaza, J. M., Abánades-Herranz, J. C., Laguna-Cuesta, F., & Rodríguez-Artalejo, F. (2016). *Performance of the Finnish Diabetes Risk Score and a simplified Finnish Diabetes Risk Score in a community-based, cross-sectional programme for screening of undiagnosed type 2 diabetes mellitus and dysglycaemia in Madrid, Spain: The SPREDIA-2 study.* PLoS ONE, 11(7), e0158489. https://doi.org/10.1371/journal.pone.0158489.

Statista (2018). Gasto sanitario en personas con diabetes a nivel mundial 2010-2017. Recuperado el 30 de noviembre de 2018 de : https://es.statista.com/estadisticas/702527/gasto-sanitario-en-personas-con-diabetes- a-nivel-mundia/.

Tanaka, T., Okamura, T., Miura, K., Kadowaki, T., Ueshima, H., & NIPPON DATA80 Research Group. (2013). *A logistic regression model for predicting incidence of type 2 diabetes mellitus in a Japanese population: The NIPPON DATA80 study.* Journal of Atherosclerosis and Thrombosis, 20(4), 358–370. https://doi.org/10.5551/jat.14783

Valenza, M., Martín, L., González, E., Aguilar, C., Botella, M. Muñoz, T. &

Valenza, T. (2012). Factores de riesgo para el síndrome metabólico en una población con apnea del sueño; evaluación en un grupo de pacientes de Granada y provincia; estudio GRANADA. Nutrición Hospitalaria, (27)4. Recuperado el 10 de noviembre de 2018 de http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S0212-16112012000400042

Zhang, Y., Hu, G., Zhang, L., Mayo, R., & Chen, L. (2015). A novel testing model for opportunistic screening of pre-diabetes and diabetes among U.S. adults. *PloS one*, *10*(3), e0120382. https://doi.org/10.1371/journal.pone.0120382.

Anexo I. Comparación del puntaje de riesgo de diabetes PreDx® con otras herramientas de evaluación de riesgo de diabetes

Diabetes risk assessment tool[dagger]	Study (year)	Score range	AUROC (p-value)[double dagger]	Comment	Ref.
Diabetes risk assessment tools requiring testing of blood samples only					
PreDx DRS	Urdea *et al.* (2009) Lyssenko *et al.*(2011) Kolberg *et al.*(2010) Watkins *et al.*(2010)	PreDx DRS ranging from 1 (least risk) to 10 (highest risk) indicates an individual patient'âs likelihood of developing diabetes within the next 5 years	0.84	Available through the Tethys Bioscience CLIA-certified laboratory. Provides a superior assessment of diabetes risk than other measures, including FPG, HbA1c, measures of insulin resistance and clinical risk factors	[71] [36] [111] [114]
FPG	ADA (2010)	FPG 100 mg/dl (5.6 mmol/l)-125 mg/dl (6.9 mmol/l) indicates increased risk of diabetes	0.77 ($p < 0.0003$)	Low specificity - approximately 26% of adults have IFG [10] , while the annual incidence of diabetes among IFG patients is only 1.95% [11]	[63]
HbA1c	ADA (2010) International ExpertCommittee (2009)	Elevated HbA1c indicates increased risk of diabetes - HbA1c range of 5.7-6.4% per ADA criteria [63] ; HbA1c range of 6.0-6.4% per International Expert Committee recommendation[61]	0.68 ($p < 0.0001$)	Low sensitivity - at the cutoff values recommended by the ADA and International Expert Committee, elevated HbA1c levels are insensitive and racially discrepant, failing to identify most adults with undiagnosed diabetes and prediabetes [15,16]	[63] [61]

OGTT	ADA (2010)	2-h glucose values of 140 mg/dl (7.8 mmol/l) to 199 mg/dl (11.0 mmol/l) indicates increased risk of diabetes.	0.83 (p = 0.982)	Requires measurement of plasma glucose levels in response to 75-g glucose load over a 2-h time period. The test is time consuming and unpleasant for many patients. High within-person variability in 2-h glucose measurements has been well documented [13]	[63]
HOMA-IR	Wallace *et al.*(2004)	HOMA-IR provides an estimate of insulin resistance based on fasting plasma insulin and FPG	0.72 (p < 0.0001)	Only indicates insulin resistance, whereas defects in both insulin secretion ([beta]Î²-cell dysfunction) and insulin resistance play a pivotal role in the pathogenesis of diabetes	[120]
Diabetes risk assessment tools based on clinical factors					
Model based on the Framingham Offspring Study	Wilson *et al.* (2007)	Model based on clinical measures plus FPG and lipids is used to estimate the risk of developing Type 2 diabetes within 7 years	0.80 (p = 0.0005)	Individual scores must be calculated by the physician The simple clinical model requires several measures, including age, sex, parental history of diabetes, BMI, blood pressure, HDL cholesterol, triglycerides, waist circumference and FPG The complex clinical models require in addition 2-h glucose values from an OGTT, fasting insulin level, CRP level, log Gutt insulin sensitivity index, log HOMA- IR and/or log HOMA-%B	[28]
Model based on the San Antonio Heart Study	Stern *et al.* (2002)	Model based on clinical measures plus FPG and lipids is used to estimate the risk of developing Type 2 diabetes within 7-8 years	0.80 (p = 0.0012)	Individual scores must be calculated by the physician Requires a complete set of measures, including age, sex, ethnicity (Hispanic or non-Hispanic white), fasting	[14]

Diabetes risk assessment tool[dagger]	Study (year)	Score range	AUROC (p-value)[double dagger]	Comment	Ref.
				glucose, systolic blood pressure, HDL cholesterol, and history of a parent or sibling with diabetes	
FINDRISC	Lindstrom and Tuomilehto (2003)	Five risk categories in FINDRISC estimate the risk of diabetes occurring over the next 10 years: 20 (very high)	§	Individual scores from the questionnaire are easily tallied, but may be subject to bias introduced by recall of self-reported data Requires information on age, BMI, waist circumference, history of antihypertensive drug treatment, previously measured high blood glucose, physical activity, consumption of fruits, berries or vegetables, and family history of diabetes	[24]
Other risk assessment tools based on laboratory-developed tests					
LP-IR score	No peer-reviewed publications that evaluate or validate the LP-IR score as a predictor of future diabetes	LP-IR score ranging from 0 (most insulin sensitive) to 100 (most insulin resistant) indicates a patient'âs insulin resistance level based on NMR results of six lipoprotein particle numbers and sizes. Doesnotdistinguishbetweeninsulinresistance and diabetes risk	§	The LP-IR score has not been validated against gold-standard measures of insulin resistance. Moreover, the pathogenesis of diabetes involves defects in both insulin secretion ([beta]Î²-cell dysfunction) and insulin resistance. The report does not provide an assessment of absolute risk of developing diabetes	
GenovaDiagnostics	No peer-reviewed publications available	The PreDGuide[trademark] includes an average inflammation score and measures of metabolic markers. The MetSyn Guide[trademark] also includes lipid particle concentration and size, and standard cholesterol measures	§	The validation of the composites of markers and other measures in the PreD Guide and MetSyn Guide has not been published in a peer-reviewed manuscript. The reports do not provide an assessment of absolute risk of developing diabetes	

Anexo II. Código HTML herramienta de visualización

```html
<html>

    <head>


    </head>


    <body>

        <h1>Indice de prediabetes</h1>

        <h2>¿Sabes lo que es la prediabetes?</h2>

        <p style="font-size:20px">Es una condición de salud en la que los niveles de azúcar en sangre están sobre el nivel aceptable, pero bajo los niveles de una persona diabética</p>

        <h2>¿Quieres conocer si tienes riesgo de padecer prediabetes?</h2>

        <h3 style="color:blue; font-size:40px" id="resCMA"></h3>

        <h3 style="color:blue; font-size:40px" id="IIMC"></h3>

        <h3 style="color:blue; font-size:40px" id="resultado"></h3>

        <h3 style="color:blue; font-size:40px" id="diagnostico"></h3>

        <h3 style="color:blue; font-size:40px" id="resCMA"></h3>

        <h3 style="color:blue; font-size:40px" id="IIMC"></h3>

        <h3>Por favor responde las siguientes preguntas:</h3>
        <br>

        <img src="grafica.png" alt="formulas" height="280">
        <br>

        <datalist id="opcion">
            <option value="SI">
            <option value="NO">
        </datalist>

        <p>Edad</p>
        <input type="text" id="edad">
        <br>

        <p>Diametro Sagital Abdominal</p>
        <input type="text" id="diametro">
        <br>

        <p>Estatura</p>
        <input type="text" id="estatura">
        <br>

        <p>Circunferencia abdominal</p>
        <input type="text" id="circunf">
        <br>

        <p>Insulina</p>
        <input type="text" id="insulina">
        <br>

        <p>Horas de sueno</p>
        <input type="text" id="sueno">
        <br>

        <p>Peso</p>
        <input type="text" id="peso">
        <br>
        <br>

        <p>Glucosa</p>
        <input type="text" id="glucosa">
        <br>
```

```
        <br>

        <p>Actividad Física</p>
        <input list="opcion" type="text" id="actividad">
        <br>
        <br>

        <p>Historia Familiar</p>
        <input list="opcion" type="text" id="familia">
        <br>
        <br>
        <br>

        <button onclick="calculate()">Calcular Indice</button>

        <script>

            indice=0;

            console.log("1");

            function calculate()
            {
                console.log("2");

                var field1=document.getElementById("edad").value;

                var field2=document.getElementById("diametro").value;

                var field3=document.getElementById("estatura").value;

                var field4=document.getElementById("circunf").value;

                var field5=document.getElementById("resultmx").value;

                var field6=document.getElementById("sueno").value;

                var field7=document.getElementById("peso").value;


                var field8=document.getElementById("actividad").value;

                var field9=document.getElementById("familia").value;

                var field10=document.getElementById("glucosa").value;

                indimc=Math.round(field7/(field3*field3)*10)/10;

                indhoma=Math.round((field5*field10)/405*100)/100;

                indice=Math.round((-8.998*(field1-48.80)/13.11+8.282*(field2-21.48)/4.16+8.293*(indimc-27.85)/6.43+8.202*(field4-
                94.72)/15.47+8.266*(field5-8.02)/8.88+8.814*(field6-7.10)/1.33+8.264*(indhoma-2.43)/2.09)*1000)/1000;

               if(indice<=-8.6888549)
                {
                    console.log("3");

                    document.getElementById("HOMA").innerHTML="Tu Indice HOMA es de "+indhoma;

                    document.getElementById("IIMC").innerHTML="Tu IMC es de "+indimc;

                    document.getElementById("resultado").innerHTML="Tu score de riesgo de prediabetes es de "+indice;

                    document.getElementById("diagnostico").innerHTML="Felicidades, tienes un riesgo bajo de padecer prediabetes.";
                }

                else if(indice>-8.6888549 & indice<=-8.1780525 & field8=="SI" & field9=="NO")
                {
                    console.log("4");

                    document.getElementById("HOMA").innerHTML="Tu Indice HOMA es de "+indhoma;

                    document.getElementById("IIMC").innerHTML="Tu IMC es de "+indimc;

                    document.getElementById("resultado").innerHTML="Tu score de riesgo de prediabetes es de "+indice;

                    document.getElementById("diagnostico").innerHTML="Tienes riesgo medio de padecer prediabetes. Por favor
```

```
                    controla tus índices anualmente.";
                }

            else if(indice>-8.6888948 & indice<=-8.1789525 & field8=="NO" & field9=="NO")
                {
                    console.log("5");

                    document.getElementById("%HOMA").innerHTML="Tu índice HOMA es de "+indhoma;

                    document.getElementById("IIMC").innerHTML="Tu IMC es de "+indimc;

                    document.getElementById("resultado").innerHTML="Tu score de riesgo de prediabetes es de "+indice;

                    document.getElementById("diagnostico").innerHTML="Tienes riesgo medio de padecer prediabetes. Por favor
                    controla tus índices anualmente y realiza 30 minutos diarios de actividad física moderada.";
                }

            else if(indice>-8.6888948 & indice<=-8.1789525 & field8=="SI" & field9=="SI")
                {
                    console.log("6");

                    document.getElementById("%HOMA").innerHTML="Tu índice HOMA es de "+indhoma;

                    document.getElementById("IIMC").innerHTML="Tu IMC es de "+indimc;

                    document.getElementById("resultado").innerHTML="Tu score de riesgo de prediabetes es de "+indice;

                    document.getElementById("diagnostico").innerHTML="Tienes riesgo medio de padecer prediabetes. Por favor
                    controla tus índices semestralmente.";
                }

            else if(indice>-8.6888948 & indice<=-8.1789525 & field8=="NO" & field9=="SI")
                {
                    console.log("7");

                    document.getElementById("%HOMA").innerHTML="Tu índice HOMA es de "+indhoma;

                    document.getElementById("IIMC").innerHTML="Tu IMC es de "+indimc;

                    document.getElementById("resultado").innerHTML="Tu score de riesgo de prediabetes es de "+indice;

                    document.getElementById("diagnostico").innerHTML="Tienes riesgo medio de padecer prediabetes. Por favor
                    controla tus índices semestralmente y realiza 30 minutos diarios de actividad física moderada.";
                }

            else if(indice>-8.1789525 & indice<=8.4612464 & field8=="SI" & field9=="NO")
                {
                    console.log("8");

                    document.getElementById("resultado").innerHTML="Tu score de riesgo de prediabetes es de "+indice;

                    document.getElementById("diagnostico").innerHTML="Tienes riesgo medio alto de padecer prediabetes. Por favor
                    considera realizarte un chequeo médico.";
                }

            else if(indice>-8.1789525 & indice<=8.4612464 & field8=="NO" & field9=="NO")
                {
                    console.log("9");

                    document.getElementById("%HOMA").innerHTML="Tu índice HOMA es de "+indhoma;

                    document.getElementById("IIMC").innerHTML="Tu IMC es de "+indimc;

                    document.getElementById("resultado").innerHTML="Tu score de riesgo de prediabetes es de "+indice;

                    document.getElementById("diagnostico").innerHTML="Tienes riesgo medio alto de padecer prediabetes. Por favor
                    considera realizarte un chequeo médico y realiza 30 minutos diarios de actividad física moderada.";
                }

            else if(indice>1.18 & indice<=2.89 & field8=="SI" & field9=="SI")
                {
                    console.log("10");

                    document.getElementById("%HOMA").innerHTML="Tu índice HOMA es de "+indhoma;

                    document.getElementById("IIMC").innerHTML="Tu IMC es de "+indimc;
```

```
                    document.getElementById("resultado").innerHTML="Tu score de riesgo de prediabetes es de "+indice;

                    document.getElementById("diagnostico").innerHTML="Tienes riesgo medio alto de padecer prediabetes. Por favor
                    considera realizarte un chequeo médico y comentar tu historia familiar de diabetes.";
                 }

                else if(indice>-0.1709525 & indice<=0.4512464 & field8=="NO" & field9=="SI")
                 {
                    console.log("11");

                    document.getElementById("HHOMA").innerHTML="Tu indice HOMA es de "+indhoma;

                    document.getElementById("IIMC").innerHTML="Tu IMC es de "+indimc;

                    document.getElementById("resultado").innerHTML="Tu score de riesgo de prediabetes es de "+indice;

                    document.getElementById("diagnostico").innerHTML="Tienes riesgo medio alto de padecer prediabetes. Por favor
                    considera realizarte un chequeo médico comentando tu historia familiar de diabetes y realiza 30 minutos
                    diarios de actividad física moderada.";
                 }

                else if(indice>-0.1709525 & indice<=0.4512464 & field8=="NO" & field9=="SI")
                 {
                    console.log("12");

                    document.getElementById("HHOMA").innerHTML="Tu indice HOMA es de "+indhoma;

                    document.getElementById("IIMC").innerHTML="Tu IMC es de "+indimc;

                    document.getElementById("resultado").innerHTML="Tu score de riesgo de prediabetes es de "+indice;

                    document.getElementById("diagnostico").innerHTML="Tienes riesgo medio alto de padecer prediabetes. Por favor
                    considera realizarte un chequeo médico comentando tu historia familiar de diabetes y realiza 30 minutos
                    diarios de actividad física moderada.";
                 }

                else if(indice>0.4512464 & field9=="SI")
                 {
                    console.log("13");

                    document.getElementById("HHOMA").innerHTML="Tu indice HOMA es de "+indhoma;

                    document.getElementById("IIMC").innerHTML="Tu IMC es de "+indimc;

                    document.getElementById("resultado").innerHTML="Tu score de riesgo de prediabetes es de "+indice;

                    document.getElementById("diagnostico").innerHTML="Tienes riesgo alto de padecer prediabetes. Por favor acude a
                    un especialista y comenta tu historia familiar de diabetes.";
                 }

                else if(indice>0.4512464)
                 {
                    console.log("14");

                    document.getElementById("HHOMA").innerHTML="Tu indice HOMA es de "+indhoma;

                    document.getElementById("IIMC").innerHTML="Tu IMC es de "+indimc;

                    document.getElementById("resultado").innerHTML="Tu score de riesgo de prediabetes es de "+indice;

                    document.getElementById("diagnostico").innerHTML="Tienes riesgo alto de padecer prediabetes. Por favor acude a
                    un especialista.";
                 }

                else
                 {
                    console.log("15");

                    document.getElementById("diagnostico").innerHTML="Por favor ingresa tu información de manera correcta. En caso
                    de ser de utilidad, te comentamos que los decimales se ingresan acompañados de un punto.";

                 }

            }
```

```
        </script>

    </body>


</html>
```

Índice de ilustraciones

Índice de tablas

UNIVERSIDAD
BICENTENARIA

¡Sueña, haz que suceda!