Universidad Internacional de La Rioja (UNIR)

Escuela Superior de Ingeniería y Tecnología

Máster Universitario en Visual Analytics & Big Data

Desarrollo de un modelo de riesgo de prediabetes

Trabajo Fin de Máster

Presentado por: Eguiguren Eguiguren, Margarita

Director/a: Fuentes Lorenzo, Damaris

Ciudad : Quito, Ecuador
Fecha : febrero de 2019

Índice de contenidos

Índice de ilustraciones

Índice de tablas

Dedicatoria

A mis hijas Emilia y Paula,

mi vital motivación.

A todos aquellos

que creyeron en mí.

Resumen

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. Los esfuerzos para controlarla se centran principalmente en programas de prevención, por lo que, en las últimas décadas, muchos investigadores dedican recursos y tiempo en desarrollar herramientas que permitan realizar un cribado oportuno de la enfermedad. Este proyecto, que forma parte de un estudio macro sobre prediabetes, constituye una gran oportunidad para contribuir con los esfuerzos mundiales de prevención mediante la creación de un score que mida el riesgo de padecer prediabetes, eslabón anterior al desarrollo de la diabetes. Los datos utilizados para el análisis fueron obtenidos de la base datos de acceso libre de la encuesta de salud norteamericana NHANES. Se decidió que el mejor método a adoptar para la realización de esta investigación es el modelo estadístico de análisis factorial que permitió generar el score, mismo que sirvió de base para el desarrollo de una herramienta web amigable que brinde alertas tempranas al usuario final.

Palabras Clave: prediabetes, análisis factorial, NHANES, prevención, score, set de datos.

Abstract

Diabetes is one of the worldwide greatest public health problems, with serious personal, family and social consequences. The prevalence of this chronic degenerative disease has reached unimaginable figures. Efforts to control it are mainly focused on prevention programs, which is why, in recent decades, many researchers devote resources and time to develop tools that allow timely screening of the disease. This project, which is part of a macro study on prediabetes, is a great opportunity to contribute to the global prevention efforts by creating a score that measures the risk of prediabetes, a link prior to the development of diabetes. The data used for the analysis were obtained from the free access database from the North American health survey NHANES. It was decided that the best method to adopt for the realization of this research is the factor analysis statistical model that allowed generating the score, which served as the basis for the development of a friendly web tool that provides early warning to the final user.

Keywords: prediabetes, factor analysis, NHANES, prevention, score, dataset.

1. Introducción

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”.

1.1. Motivación y planteamiento

Según los cálculos de la Federación Internacional de Diabetes (FID), publicados en la octava edición del Atlas de la Diabetes, hay 425 millones de personas que tienen actualmente diabetes (8.8% de los adultos) y más de 352 millones sufren intolerancia a la glucosa con un elevado riesgo de contraer la enfermedad. Para el 2045 se prevé que el número de personas con la enfermedad se incrementará a más de 693 millones. 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 profesor Nam Han Cho 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).

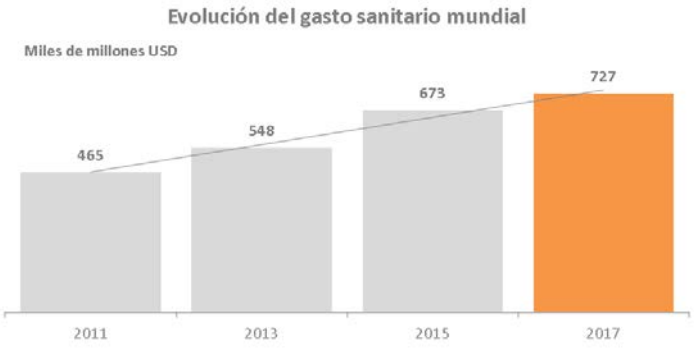


Figura 1. Evolución del gasto sanitario mundial por diabetes. (Statista, 2018)

Como se puede apreciar en la figura 2, las cifras de diabetes en las diferentes zonas geográficas del mundo van en continuo incremento, llegando a pronosticarse para el año 2045, por ejemplo para el África, la extraordinaria cifra de 156% de aumento.

"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).

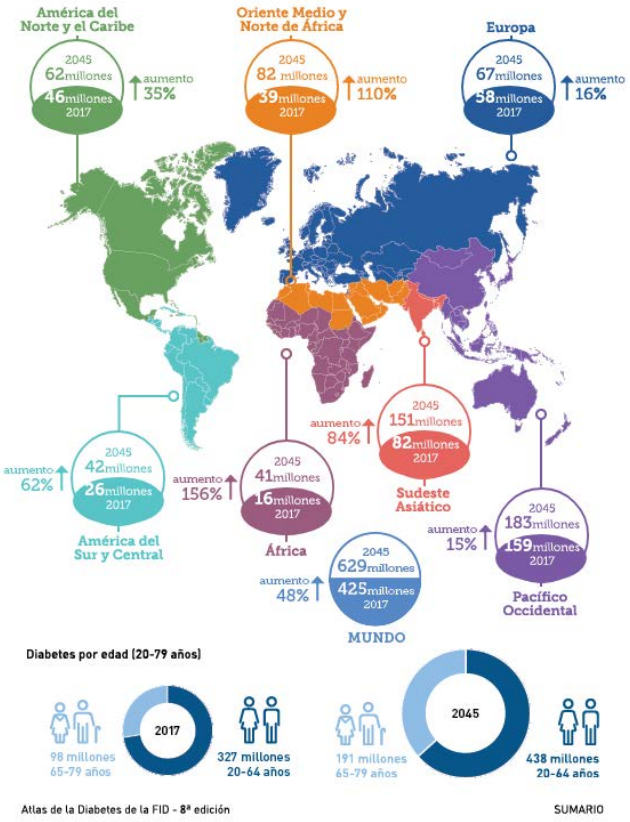


Figura 2. Prevalencia de diabetes a nivel mundial. (FID, 2017)

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.

Desarrollo de un modelo de riesgo de prediabetes

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. (American Diabetes Association, 2018), (Asociación Latinoamericana de Diabetes ALAD, 2017):

- Edad, sexo, etnia
- Antecedentes genéticos
- Niveles de glucosa en plasma
- Índice de masa corporal
- 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 en este trabajo 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, procesado, modelamiento, análisis, visualización, llegando finalmente a extraer conocimiento válido para la toma de decisiones.

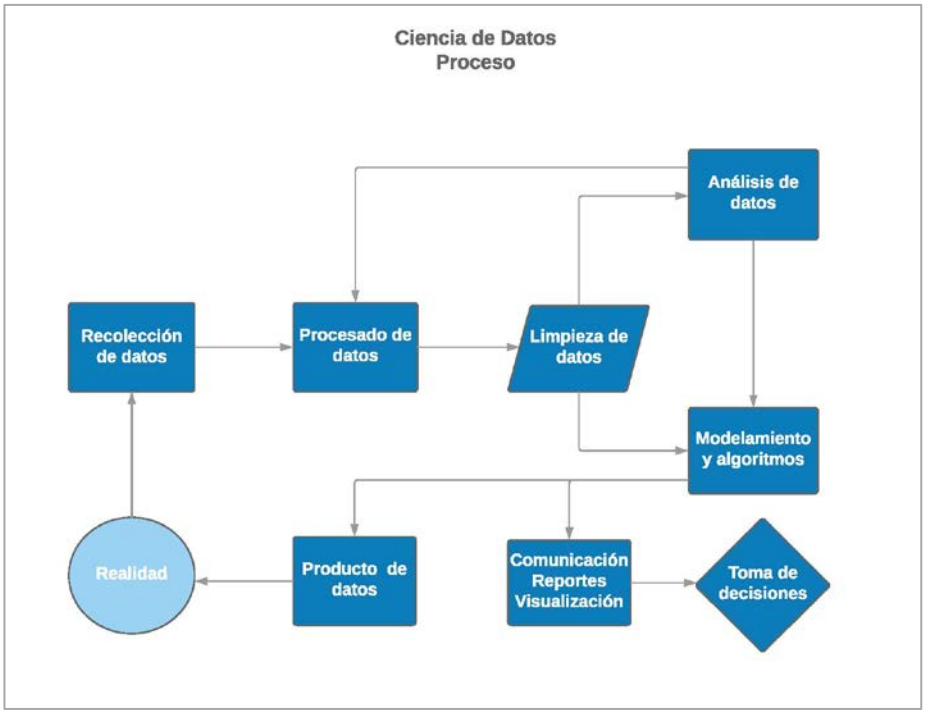


Figura 3. Proceso de ciencia de datos

El presente trabajo, que aporta al proyecto de investigación liderado por Wilmer Danilo Esparza, PhD. en Ciencias y Técnicas de la Actividad Física y del Deporte de la Universidad de Las Américas en Quito-Ecuador, que busca desarrollar una aplicación informática predictiva, tiene como objetivo aplicar un modelo estadístico que genere un score que sirva de apoyo para determinar el riesgo de prediabetes en la población, con base en los resultados de la encuesta de salud norteamericana NHANES (National Health and Nutrition Examination Survey) de los años 2011-2012, 2013-2014 y 2015-2016. Esta propuesta está basada en los conocimientos que engloban los procesos inmersos en la ciencia de datos, desde la recolección de los datos, procesado de los mismos para obtener información relevante y útil, limpieza, exploración, modelaje estadístico y visualización de resultados, con la finalidad de crear conocimiento aplicable al campo de la salud.

Desarrollo de un modelo de riesgo de prediabetes

1.2. Estructura del trabajo

En la figura 4 se esquematiza las secciones de las que está constituida la presente memoria, la misma que consta de 8 capítulos que recogen todo el proceso de investigación realizado.

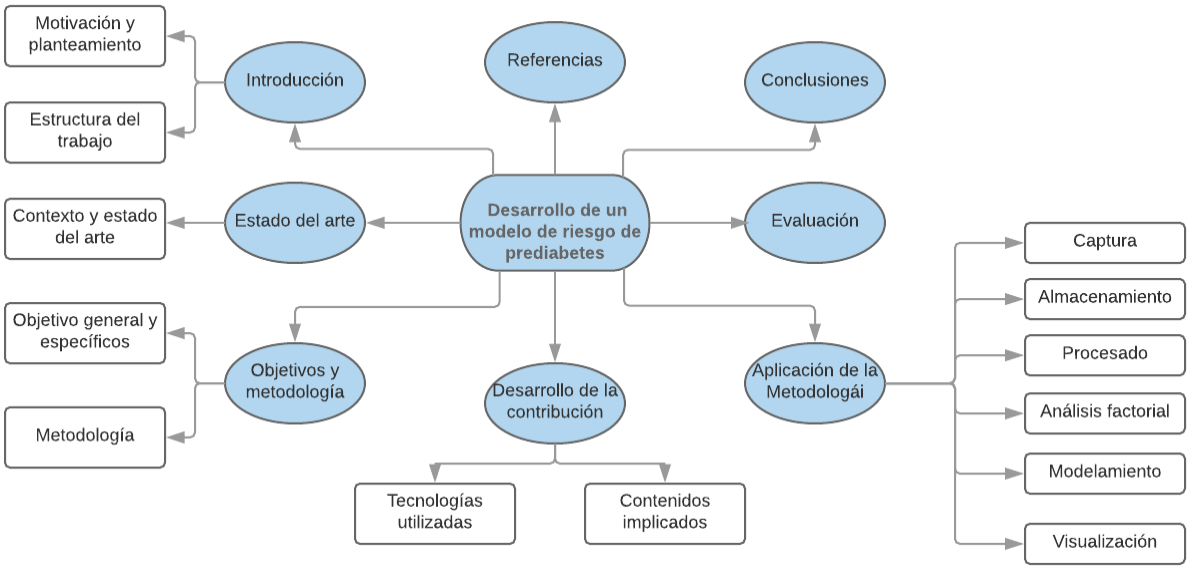


Figura 4. Estructura del trabajo

La Introducción presenta una mirada general de la prevalencia mundial de la diabetes mediante cifras extraídas de distintas organizaciones reconocidas a nivel mundial. Introduce conceptos específicos de diabetes y prediabetes, sus factores de riesgo y padecimientos futuros. También se expone el motivo que soporta la decisión de investigar el tema de la prediabetes.

Luego, en el Estado del Arte, se hace un recorrido por la literatura científica documentando los trabajos de investigación existentes a nivel mundial relativos a la detección precoz de la diabetes y prediabetes, utilizando modelos predictivos, modelos clasificatorios y machine learning.

En el tercer capítulo se expone el objetivo general del estudio, es decir, la finalidad que se busca en última instancia con el desarrollo de la investigación; además se presentan varios objetivos específicos que formulan lo que se quiere alcanzar en cada etapa de la

investigación y se expone la metodología que se va a seguir para alcanzar los objetivos propuestos.

En el desarrollo de la contribución se pueden encontrar descritas las tecnologías utilizadas, como por ejemplo el lenguaje R, el sistema SQL, el paquete estadístico SPSS, el lenguaje HTML, entre otras. Por otro lado, se explican en detalle los contenidos implicados describiendo la base de datos utilizada, concepto y factores de riesgo de la prediabetes y el modelo estadístico escogido, "análisis factorial".

En el capítulo subsecuente, se explica paso a paso la elaboración de un manual para la implementación de la metodología, que establece de forma clara y sencilla el proceso en su totalidad, desde la captura de la información hasta la herramienta de visualización propuesta, pasando por el almacenamiento, procesado, análisis y modelamiento.

La evaluación de la implementación de la metodología es un espacio para valorar los resultados obtenidos luego de la ejecución del método propuesto, realizando pruebas del modelo final y contrastándolas contra los resultados obtenidos.

Posteriormente, en el capítulo de conclusiones se encuentra un resumen de la contribución, la relevancia y alcance de la propuesta, conclusiones finales, recomendaciones y una propuesta de trabajo futuro basado en los resultados obtenidos en la presente investigación.

Finalmente se listan las referencias de las fuentes investigadas que sirvieron de fundamento teórico para elaborar esta memoria.

2. Contexto y estado del arte

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.

Como lo indica 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 5, que ayuda a determinar la prediabetes o la diabetes no diagnosticada en poblaciones de riesgo, la utilización de "reglas de predicción clínicas" mediante la revisión de historias clínicas o la aplicación de "escalas de riesgo o cuestionarios" previos a la realización de exámenes de laboratorio confirmatorios.

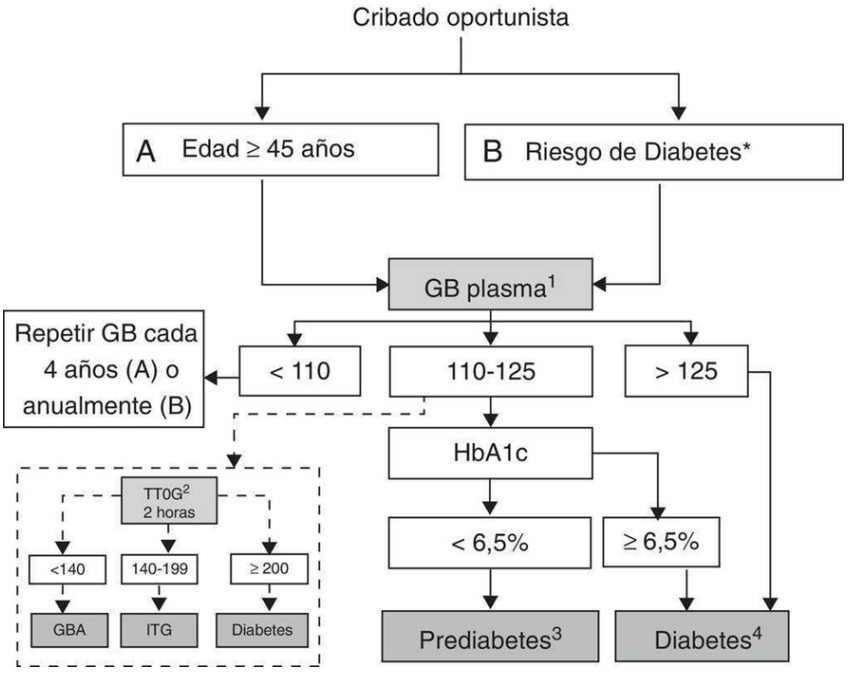
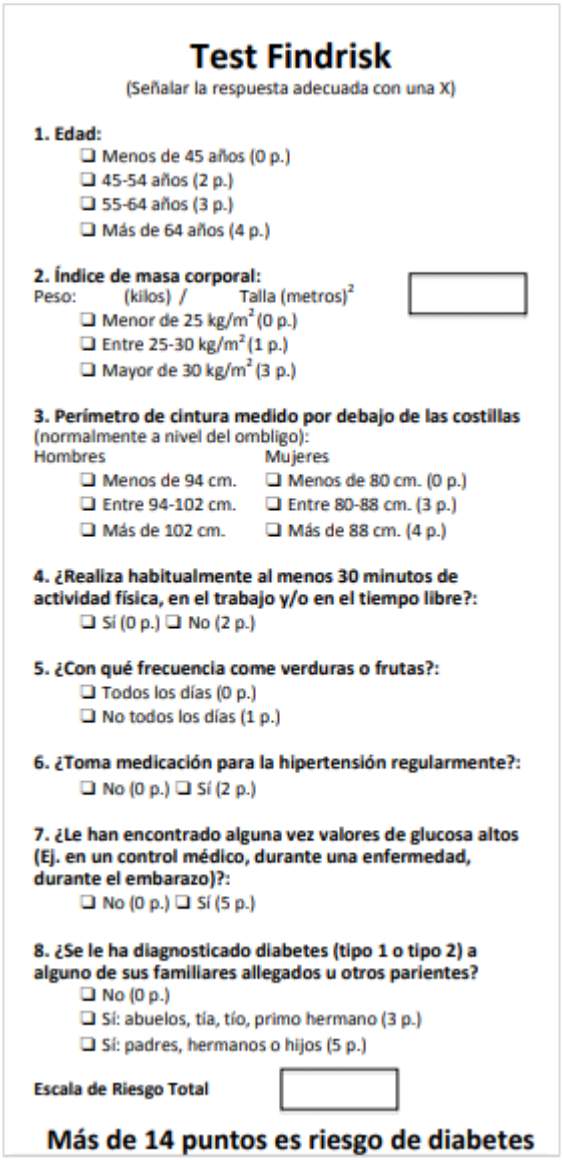


Figura 5. Algoritmo de detección de prediabetes y diabetes. (Mata-Cases et al., 2015)

2.1. Trabajos relacionados

Actualmente existen muchos y variados métodos que permiten determinar de manera oportuna la presencia de factores de riesgo en la población, uno de los más utilizados es el cuestionario **Finnish Diabetes Risk Score (FINDRISC)**, test no invasivo que se basa en un cuestionario de 8 preguntas, el mismo que se desarrolló para el Programa Nacional de Prevención de la Diabetes Tipo 2 de Finlandia, como se muestra en la figura 6; encuesta que se realiza cada cinco años en una muestra aleatoria de la población finlandesa.



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 6. Cuestionario FINDRISK.

Otra herramienta no invasiva para detectar prediabetes y diabetes no diagnosticada es la **“Diabetes risk calculator”** propuesta por Heikes y Asociados (Heikes, Eddy, Arondekar, Desarrollo de un modelo de riesgo de prediabetes

Schlessinger, 2008), mismo que 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 encuesta de salud y nutrición de los Estados Unidos (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 dio la precisión más baja". En esta investigación de dos años de duración, las variables de entrada fueron 12 factores de riesgo 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, Huang, Rao, Zhang, Liu, 2012)

Castrillón, Sarache y Castaño (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 (Rojas, Escamilla, Gómez, Zárate, Aguilar) plantea la necesidad de distinguir hombres de mujeres con el fin de establecer un score que tenga mejor desempeño por género. 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.

En el 2015, los investigadores Zhang Y, Hu G, Zhang L, Mayo R, Chen L., 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 pre-diabetes.

En el 2016, Peng Ouyang, Xitong Guo, Yiting Shen, Naiji Lu, Chenghua Ma, 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 incorpora 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 J., SomannavarM., y AraganjiR., (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 C., Vistisen D., Borch-Johnsen K. y Colagiuri S. en su studio “Risk scores for type 2 diabetes can be applied in some populations but not all” (2006), 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.

Una tabla con un resumen comparativo de otras herramientas de evaluación de riesgo de diabetes se agrega a este trabajo de investigación y se la puede consultar en el anexo I.

Finalmente, se debe mencionar la contribución que a futuro aportarán en este campo los tres trabajos investigativos que son parte integrante 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 más 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 de las mismas al grupo humano estudiado (por sus características específicas), de la simplicidad de su aplicación e interpretabilidad de los resultados.

2.2. Implicaciones sobre el desarrollo del trabajo

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, las cuales utilizan variables antropométricas exclusivamente radica que, en general, éstos 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, Kramer; Miller, Arena, Kriska , 2014).

Según el consenso español de prediabetes “Una gran limitación para el uso del FINDRISC es que el paciente no sabe calcular su propio IMC y que la medición del perímetro de cintura no se realiza habitualmente en nuestro medio” (Mata-Cases et al., 2015).

Las propuestas que utilizan regresiones logísticas para predicción de diabetes, en su mayoría son investigaciones de campo que siguen el proceso de la enfermedad durante ciertos períodos de tiempo (elegidos por el investigador); éstos trabajan con series de tiempo y con diagnósticos ya establecidos que permiten hacer predicciones de desarrollo de la enfermedad a un número de años definido. La regresión logística es un método de regresión no lineal para predecir una variable dependiente categórica conocida. La fórmula del modelo logístico calcula la probabilidad de la enfermedad seleccionada (y = 0 si el sujeto no sufre la enfermedad; de lo contrario, y = 1) en función de los valores de los factores de riesgo predictivo.

Cuando se utilizan modelos de aprendizaje supervisado como árboles de decisión, la finalidad es clasificar el estado de salud del paciente, en categorías conocidas, por ejemplo catalogándolo como sano, prediabético o diabético, mas no reconocer la incidencia de cada uno de los factores (otorgándoles pesos ponderados) sobre el diagnóstico, de manera que puedan ofrecer criterios para anticiparse al desarrollo de la enfermedad y dar recomendaciones oportunas para revertir el proceso a tiempo.

Los modelos de aprendizaje no supervisado que utilizan herramientas de clustering solamente agrupan los pacientes por categorías no conocidas con características comunes, sin proveer información suficiente de las variables que inciden en la enfermedad y que podrían ser modificadas con cambios oportunos en el estilo de vida.

Al ser el objetivo de esta propuesta el generar un aplicativo que calcule un score individual que brinde una alerta temprana sobre el riesgo de desarrollar prediabetes con base en los datos proporcionados por la encuesta norteamericana NHANES, se estudian varios posibles modelos estadísticos y de inteligencia artificial en busca del que mejor se adecue a los datos disponibles.

Cabe destacar que la encuesta NHANES no es un estudio longitudinal pues no representa una serie de tiempo sino un estudio transversal en cada periodo indicado; tampoco genera diagnósticos basados en los resultados de las pruebas de laboratorio, sino que pregunta sus percepciones a los encuestados.

Así, al no disponer de datos que generen información en el tiempo y tampoco de un diagnóstico certero de prediabetes, el modelo estadístico que se decide aplicar en este trabajo es el análisis factorial, el mismo que brinda la posibilidad de introducir todas las variables intercorrelacionadas que influyen en el fenómeno, buscando reducir el número de variables a un conjunto menor que, sin perder demasiada información, expliquen de mejor

manera la prediabetes. En el análisis factorial todas las variables cumplen el mismo papel, no existe una variable respuesta ni variables independientes, como en la mayoría de modelos de regresión, sino que todas las variables son analizadas en conjunto, no existe a priori una dependencia conceptual de unas variables sobre otras por lo que tiene gran versatilidad.

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 que le dé a conocer este score con su respectiva interpretación y proveer recomendaciones que le ayuden a tomar acciones preventivas antes que correctivas.

3. 3. Objetivos y Metodología

A continuación se plantean los objetivos generales, objetivos específicos y metodología del presente trabajo de investigación.

3.1. Objetivo General

Generar un índice único 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.

3.2. Objetivos Específicos

- 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.

3.3. Metodología del trabajo

En la figura 7 se detalla el flujo de los procesos que integran la metodología utilizada en la presente investigación.

Tanto la captura de la información como el almacenamiento y limpieza de los datos, son procesos realizados en conjunto con el equipo de trabajo que conforma el macro proyecto sobre prediabetes, mencionado anteriormente en esta memoria.

Las técnicas descritas específicamente en las secciones 5.1, 5.2, 5.3 y subsecciones 5.3.1, 5.3.2, 5.3.3, no son exclusivamente de autoría individual ya que se desarrollaron, como aporte técnico al proyecto general, conjuntamente con el maestrante Gabriel Rivadeneira, miembro del equipo de investigación.

En contraste, la selección de variables especificada en la sección 5.3.4, la codificación de variables descrita en el acápite 5.3.5, el procesamiento estadístico mediante análisis factorial detallado en la sección 5.4 y el diseño de la visualización interactiva narrada en el punto 5.6 son de autoría propia individual.

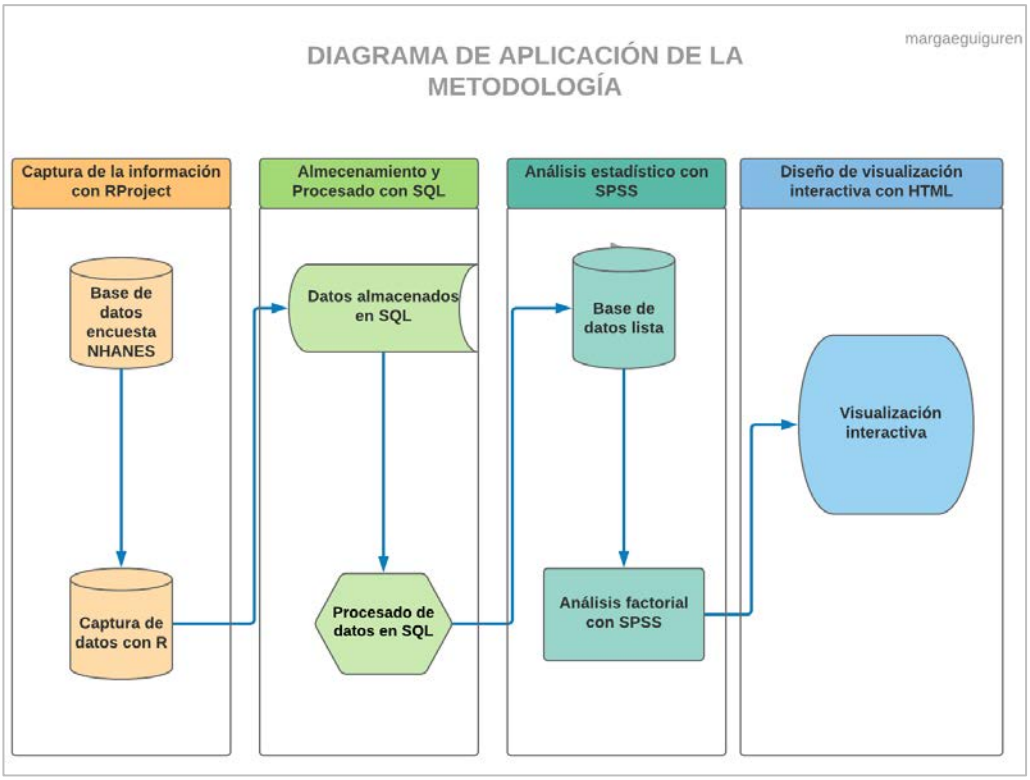


Figura 7. Diagrama de la metodología utilizada.

3.3.1. Selección del dataset para el estudio

Existen varias consideraciones que son de suma importancia y que se deben hacer a la hora de escoger 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 norteamericana NHANES (National Health and Nutrition Examination Survey) de los años 2011 al 2016, la misma que reúne todas las condiciones antes citadas.

3.3.2. 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.

3.3.3. 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.

3.3.4. 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 variables y así se define un grupo de datos con los que se trabaja para investigar el fenómeno de la prediabetes.

3.3.5. 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.

3.3.6. 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.

4. Definición de conceptos

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.

4.1. Contenidos implicados

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

La Encuesta de salud y nutrición NHANES [1] es un estudio realizado por el Centro Nacional 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. Se inició en el año 1990 y se realiza anualmente hasta la actualidad, recopilando información de alrededor de 5000 personas por año, encuestando un total de 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. Al ser una encuesta que contiene variedad, cantidad, calidad y veracidad de los datos, es un dataset confiable para la producción del presente trabajo de investigación.

4.1.2. 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 sub diagnosticada ya que no presenta ninguna sintomatología. “Más de uno de cada tres estadounidenses tiene prediabetes, y el 90 por ciento de ellos ni siquiera saben que la tienen” (Brass, L. 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

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

una manifestación subclínica de un trastorno del metabolismo de los carbohidratos" (Díaz, O., Cabrera, E., Orlandi, N., Araña, M., y Díaz, O., 2011)

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, 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 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, A., et al. 2008).

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.

"A pesar de la base de evidencia bien establecida para el tratamiento de prediabetes, hay un sustancial sub-reconocimiento y tratamiento insuficiente del problema" (Khetan, A., Rajagopalan, S., 2018).

4.2. Análisis factorial

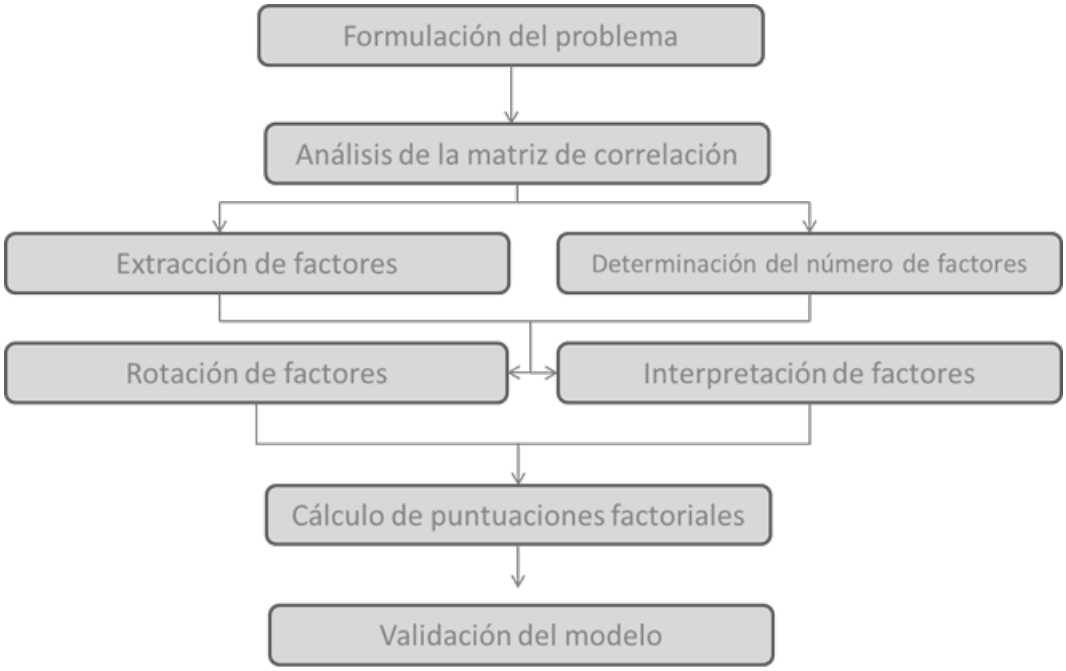


Figura 8. Esquema de análisis factorial. (De La Fuente S., 2011)

El análisis factorial es una técnica estadística que se ocupa de reducir variables agrupándolas en conjuntos homogéneos compuestos por las variables que tienen mayor correlación entre ellas, con ello se pretende explicar el máximo de información que contienen los datos con el menor número de dimensiones (parsimonia o economía de la información); además busca que los factores o dimensiones resultantes tengan un significado reconocible (interpretabilidad). Esto da como resultado una matriz de correlaciones y finalmente un número de factores. Generalmente "un análisis factorial

eficiente es aquel que proporciona una solución factorial sencilla e interpretable" (Ibarra, A. 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, P., Fachelli, S., 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.

Además hay que considerar que existen dos tipos de análisis factorial: el exploratorio en el que el investigador no conoce los factores comunes a priori y el confirmatorio donde el investigador establece los factores que contienen las variables independientes originales.

Como muestra la figura 8, el análisis factorial tiene varias etapas o fases que se describen a continuación:

4.2.1. Formulación del problema

Es el objeto del estudio y debe estar bien expresado y caracterizado y por supuesto tiene que ser viable. "En realidad, plantear el problema no es sino afinar y estructurar más formalmente la idea de investigación" (Hernández Sampieri, R., Fernández Collado, C., y Baptista Lucio, P., 2003, p. 8). En la figura 9 se sintetizan los elementos y criterios necesarios para formular el problema de investigación.

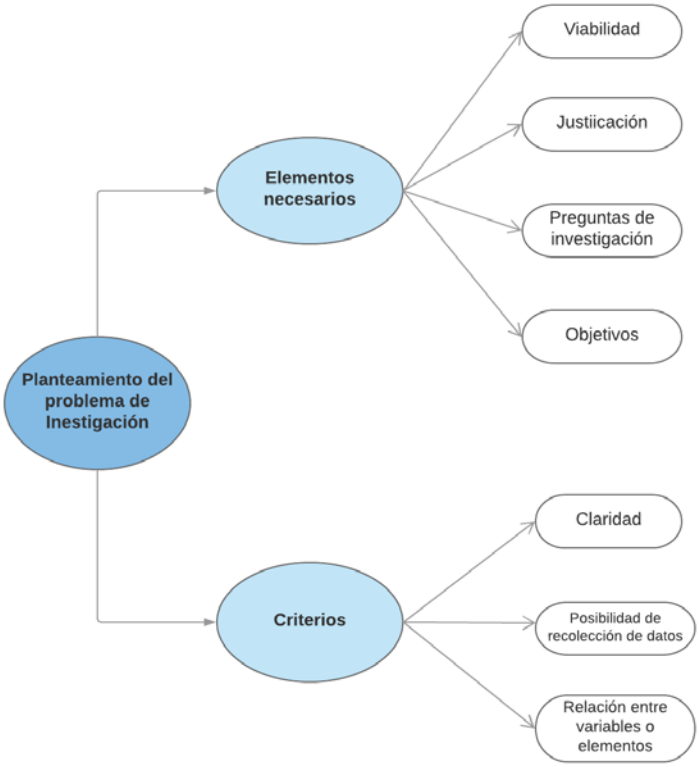


Figura 9. Planteamiento del problema de investigación

Hernández Sampieri, R., Fernández Collado, C., y Baptista Lucio, P. (2003, p. 8)

4.2.2. 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 intercorrelacionadas. 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 KMO (Kaiser-Meyer-Olkin).

El test de esfericidad de Bartlett evalúa la conveniencia de aplicar el análisis factorial, busca contrastar 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), por ende si hay 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 S. (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 10.

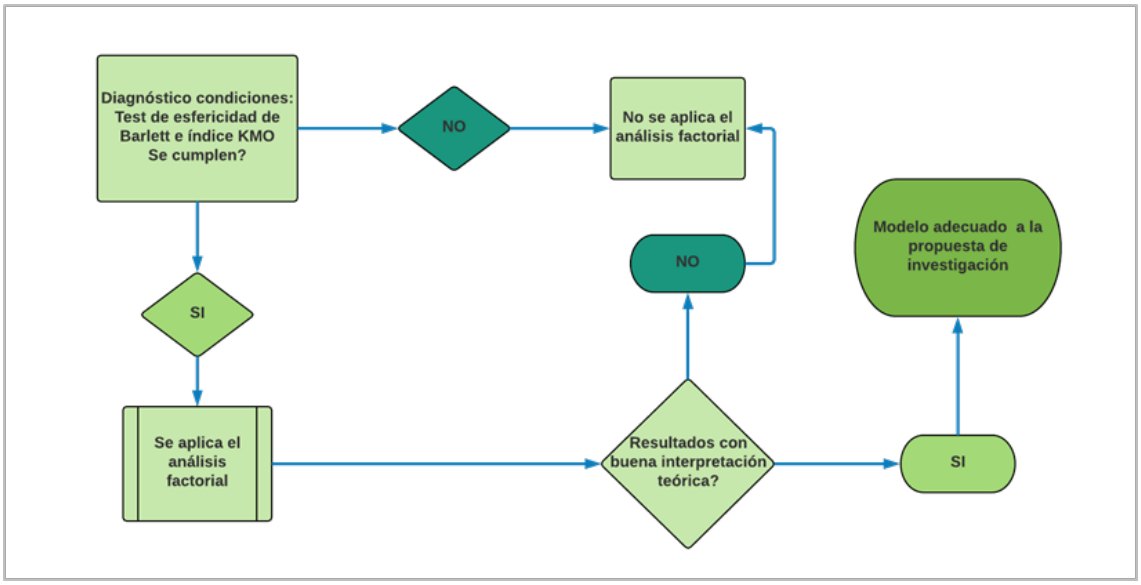


Figura 10. Determinación de la conveniencia de aplicar el análisis factorial

4.2.3. 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 SPSS ofrece las siguientes opciones: Método de las Componentes Principales, Método de los Ejes principales y Método de Máxima Verosimilitud (Pérez, E., Medrano, L. 2010).

- **Método de componentes principales**: Este método 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.

4.2.4. Determinación del número de factores

La selección del número de factores que deseamos obtener es un paso muy importante en el análisis factorial; lo podemos obtener mediante el método que viene por defecto en el paquete SPSS que es la regla Kaiser de extracción de factores con auto valores (eigenvalues) superiores a 1 o mediante la aplicación del screen test que proporciona una gráfica de sedimentación en la que se aprecia claramente el número de factores a extraer, el mismo que está determinado por el primer cambio de la pendiente en la gráfica (criterio de contraste de caída), como muestra la figura 11. (Pérez, E. y Medrano, L. 2010).

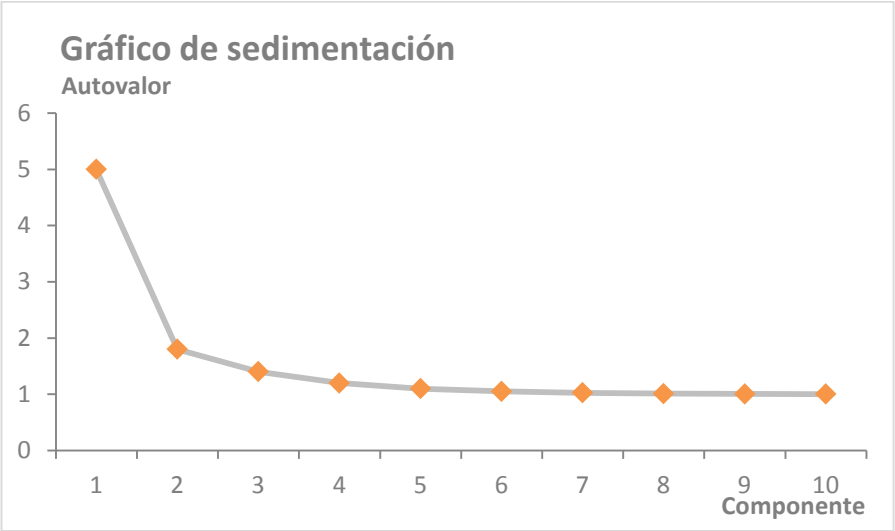


Figura 11. Ejemplo de gráfica de sedimentación (screen test)

4.2.5. 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. 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 factores). La matriz resultante se llama matriz de componentes o factores rotados. Las rotaciones pueden ser oblicuas y ortogonales.

- Rotación ortogonal: “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” (De La Fuente, S. 2011). El método más utilizado es el Varimax, de hecho en el programa SPSS viene establecido por defecto.

Los factores se rotan con el fin de eliminar las correlaciones negativas y disminuir las correlaciones entre las variables de cada factor.

4.2.6. 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.

4.2.7. 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.

4.2.8. Validación del modelo

Los modelos estadísticos, una vez planteados, deben ser validados; con el análisis factorial generalmente se utilizan dos procedimientos para hacerlo, mediante 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 Desarrollo de un modelo de riesgo de prediabetes

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.

4.3. 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.

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 12, 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”.

El uso de uno u otro modelo, 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.

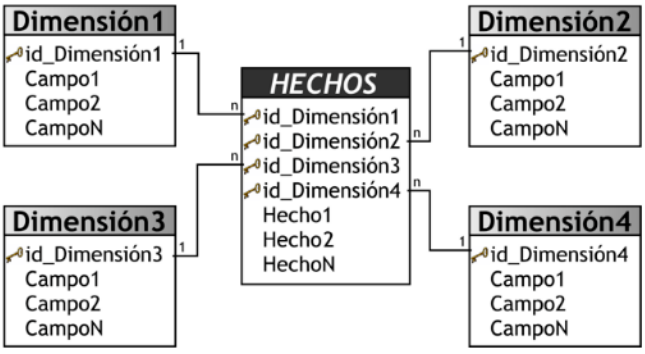


Figura 12. Modelo estrella

4.4. 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.

4.4.1. 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.

4.4.2. R Studio

R Studio[3] es un entorno de desarrollo integrado (IDE) para R. Un IDE es un entorno de programación empaquetado como una aplicación, que incluye un editor de código, una interfaz gráfica (GUI), un compilador y un depurador. R Studio ofrece un entorno de trabajo amigable con la mayoría de los lenguajes de programación y se ejecuta en el escritorio con Windows, Mac o Linux o en un navegador. Contribuye a la comunidad eficazmente en los campos de la investigación, educación e industria, brindando una potente herramienta gratuita para análisis estadísticos, minería de datos, matemáticas financieras, ya que es un software de análisis de datos de código abierto. Para utilizar R Studio se requiere haber instalado previamente R Project.

4.4.3. SQL Server

SQL Server [4] es un sistema de gestión de base de datos relacional (RDBMS) de código abierto, basado en lenguaje de consulta estructurado que se ejecuta en prácticamente todas las plataformas, incluyendo Linux, UNIX y Windows, utiliza múltiples tablas para organizar y estructurar la información. Es un gestor multiusuario, lo que le permite ser utilizado por varias personas al mismo tiempo, e incluso, realizar varias consultas a la vez, además permite acceder rápidamente a las sentencias del gestor de base de datos. Su uso está muy

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

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

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

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.

4.4.4. IBM SPSS

SPSS [5] es un programa estadístico integral muy utilizado, especialmente en el campo de las ciencias sociales, pero su aplicabilidad está extendida actualmente a muchos campos de la investigación. Con esta útil herramienta se pueden hacer múltiples procesos como análisis estadísticos y presentación de informes, modelado para predicción y minería de datos, administración e implementación de decisiones, análisis de big data. Incluye estadísticas descriptivas, estadísticas de dos variables, pruebas T, ANOVA y de correlación. Con SPSS se puede recopilar datos, calcular estadísticas y construir visualizaciones.

4.4.5. HTML

Hyper Text Markup Language (HTML) [6] es un lenguaje de marcado de texto, con un formato y estructura definido, que utiliza etiquetas para ordenar diversos documentos. Tiene una estructura definida por una cabecera (head) y un cuerpo (body), es un lenguaje utilizado para elaboración de páginas web. La cabecera contiene información del documento y no es visible para el usuario, mientras que el cuerpo contiene los contenidos que van a ser vistos por el usuario (textos, imágenes, tablas, gráficos y más).

4.4.6. JavaScript

JavaScript [7] es un lenguaje de programación interpretado, multiplataforma, orientado a objetos, que permite crear contenidos dinámicos para la elaboración de páginas web. Los programas escritos con JavaScript no requieren procesos intermedios para ser probados en cualquier navegador.

4.4.7. Brackets

Brackets [8] es un editor de texto moderno diseñado por Adobe y gestionado a través de GitHub, de código abierto que facilita el diseño de páginas web, construido sobre tecnologías como HTML, CSS y JavaScript. Es una herramienta multiplataforma de fácil uso y de gran ayuda para los desarrolladores web.

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

[6] https://html.com/

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

[8] http://brackets.io/

5. Desarrollo de la aplicación de la Metodología

El presente capítulo tiene como objetivo exponer de forma clara, comprensible y sencilla la aplicación de la metodología utilizada en el presente estudio.

Con el objetivo de contribuir con la investigación sobre prediabetes liderada por el PhD Danilo Esparza, de la cual el presente trabajo es parte integrante, se desarrollaron en conjunto una serie de procesos, específicamente los relacionados con la captura, limpieza y almacenamiento de los datos utilizando R Studio y SQL Server, con la finalidad de crear un Datawarehouse que sea un aporte tecnológico que pueda servir de base para futuras investigaciones.

A continuación se describirá el camino metodológico propuesto.

5.1. 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 14, 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 14. 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

5.2. 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 15.

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

Figura 15. Almacenamiento y respaldo de la información

5.3. 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 macro proyecto sobre prediabetes.

5.3.1. 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 16.

```
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 16. Creación de una clave primaria

5.3.2. 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 a 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 17:

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 17. Limpieza de la información

5.3.3. 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 18, que fue diseñado según los requerimientos de los modelos que se utilizan en esta investigación.

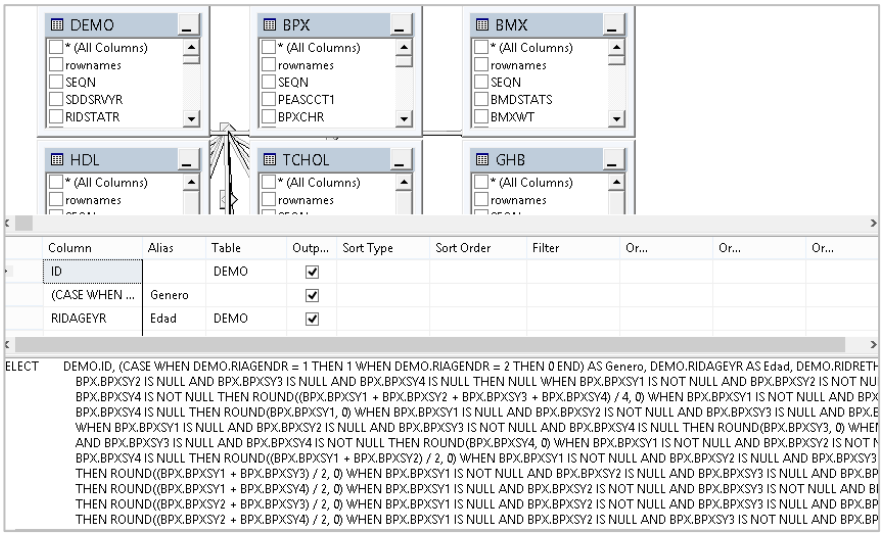


Figura 18. Vista de asociación de tablas

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 19.

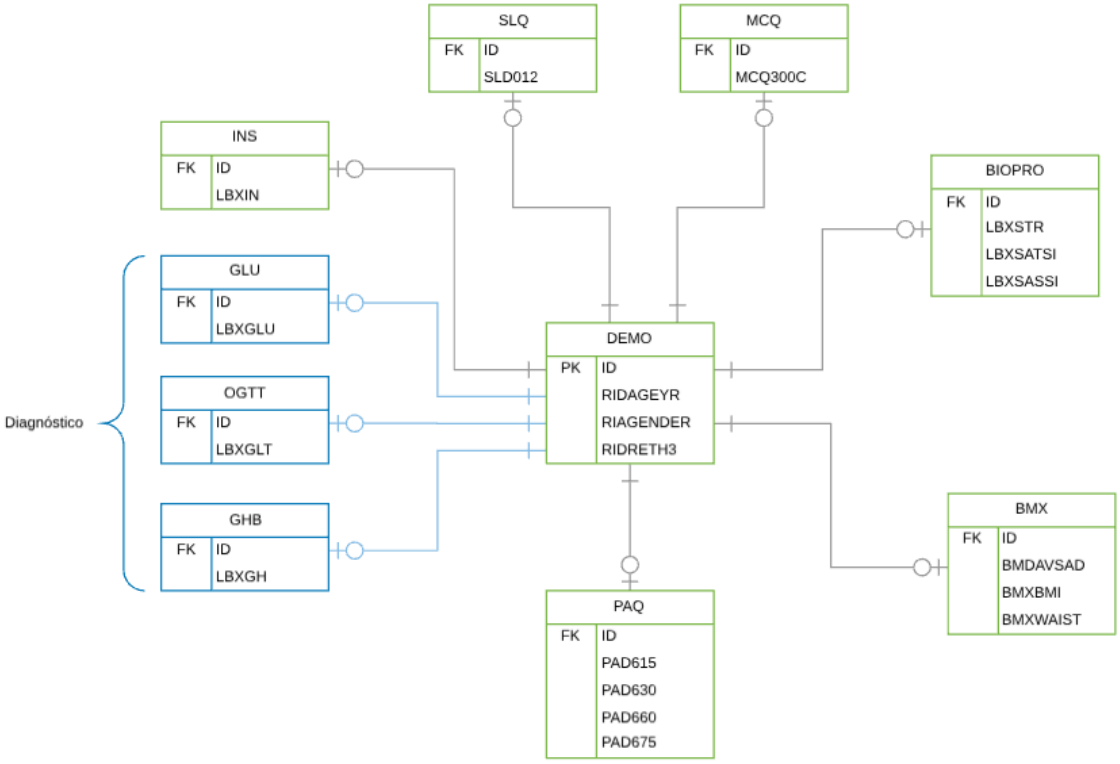


Figura 19. Modelo de base de datos (tablas 2015-2016)

5.3.4. Selección de variables

Habiendo almacenado 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 4.2.2. 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. Éstos 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 puesto que, lo que interesa en este 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.

$$HOMA = \frac{\text{Insulina} * \text{Glucosa}}{405}$$

Ecuación 1

Otro índice que calcula la herramienta es el índice de masa corporal IMC, que al igual que el anterior se lo calcula mediante 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.

$$IMC = \frac{\text{peso}}{estatura^2}$$

Ecuación 2

5.3.5. 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 20:

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 20. Codificación y procesamiento

Este proceso 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 21.

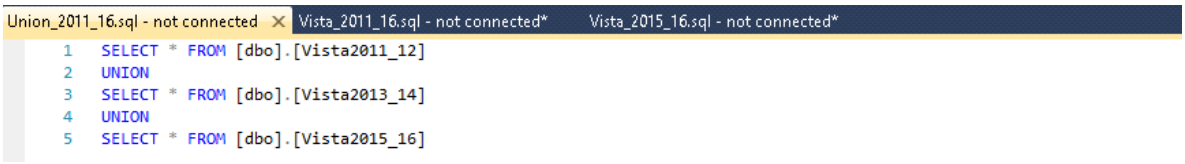


Figura 21. Unión de vistas

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 SPPS el modelo que se mencionó anteriormente. Además aquí se definen ciertos rangos pertinentes como edad, horas de sueño y diagnóstico.

5.4. Análisis Factorial con SPSS

Basándome en los procedimientos realizados en conjunto con el equipo de investigación, mi aporte individual consiste en el procesamiento estadístico mediante análisis factorial con el fin de obtener un índice único que mida el riesgo de padecer prediabetes, mismo que sirva de insumo para posteriormente desarrollar una herramienta interactiva que dé a 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 4 del presente trabajo.

5.4.1. Vista de variables seleccionadas

En la figura 22, se presenta la vista de las variables definidas en la tabla 1 del capítulo 4, acápite 4.2.2. definiendo 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	CircAbdoni...	Numérico	20	2	CircAbdonimal	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	Trigliceridos	Numérico	20	2	Trigliceridos	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 22. Vista de variables

5.4.2. Vista de datos

La figura 23 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 entes de realizarse los procesos de limpieza y filtrado constaba de 29.902 registros.

	Genero	Edad	Etnia	DiamSagitalAbdominal	IMC	CircAbdonimal	Glicohemoglobina	Insulina	ToleranciaGlucosa	GlucosaPlasma	Trigliceridos	HistoriaFamiliar	Actividad Fisica	HorasSueno	ALT
1	mujer	26,00	3,00	14,50	20,30	73,70	5,20	3,85	80,00	89,00	24,00	si	no	8,00	23,00
2	hombre	50,00	6,00	22,30	23,60	99,30	5,00	6,08	100,00	110,00	93,00	no	no	7,00	20,00
3	mujer	57,00	6,00	29,20	38,30	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	102,60	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	7,16	80,00	98,00	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	85,00	67,00	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	86,00	39,00	no	no	6,00	18,00
9	hombre	25,00	3,00	23,00	33,00	109,60	4,70	8,02	77,00	108,00	94,00	si	no	7,00	16,00
10	hombre	58,00	3,00	25,60	33,60	113,20	5,50	16,01	158,00	105,00	101,00	no	no	7,00	22,00
11	hombre	26,00	4,00	14,90	19,20	72,50	5,30	2,97	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	75,00	no	no	6,00	11,00
13	mujer	21,00	1,00	19,00	27,30	89,50	5,00	14,50	185,00	90,00	67,00	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	18,10	78,40	5,40	6,69	85,00	98,00	123,00	si	no	8,00	18,00
16	hombre	26,00	3,00	18,70	24,90	86,40	5,20	4,02	96,00	88,00	36,00	si	no	8,00	20,00
17	mujer	46,00	2,00	17,80	23,00	81,90	5,30	7,64	82,00	83,00	60,00	no	no	5,00	17,00
18	hombre	56,00	3,00	23,00	28,60	99,80	5,10	8,82	136,00	107,00	321,00	si	no	7,00	28,00
19	hombre	29,00	3,00	22,60	28,60	104,20	5,30	4,77	55,00	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	85,00	5,60	3,81	96,00	99,00	113,00	si	no	7,00	15,00

Figura 23. Vista de datos cargados

5.4.3. 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 25 deja ver claramente las siete variables cuantitativas definidas, entre ellas la variable horas de sueño, que a pesar 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

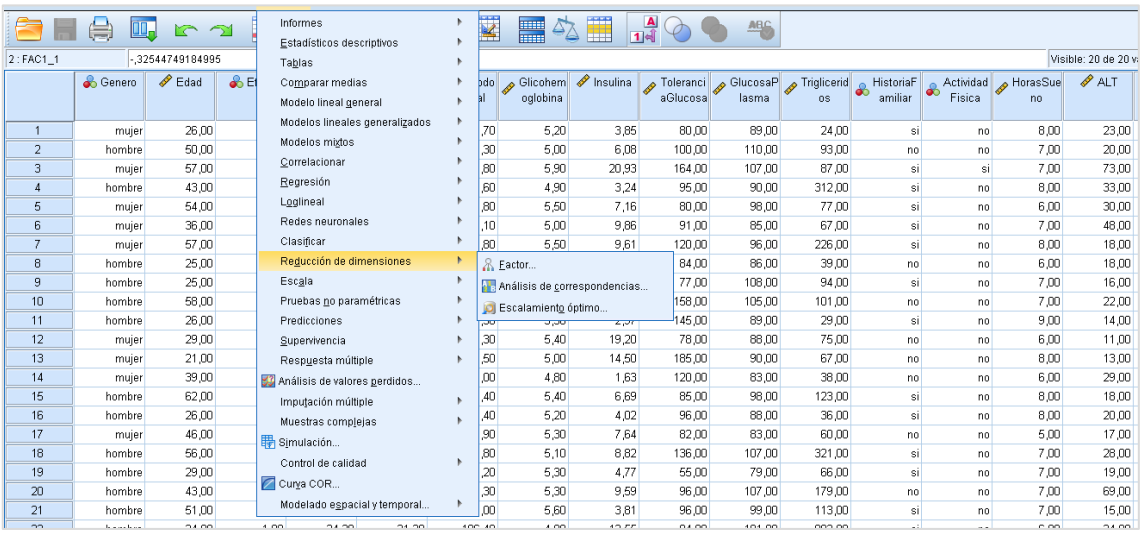


Figura 24. Reducción de dimensiones

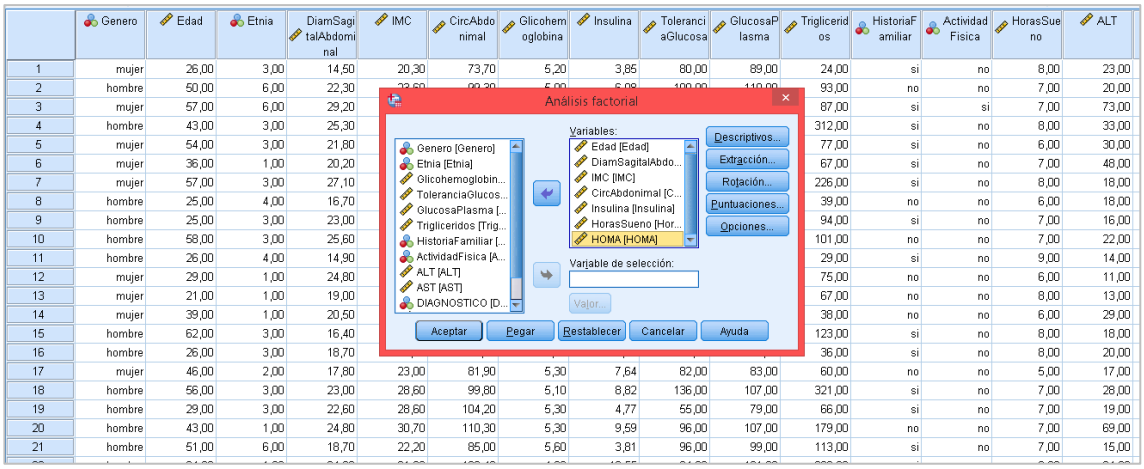


Figura 25. Variables definidas

5.4.4. 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 26.

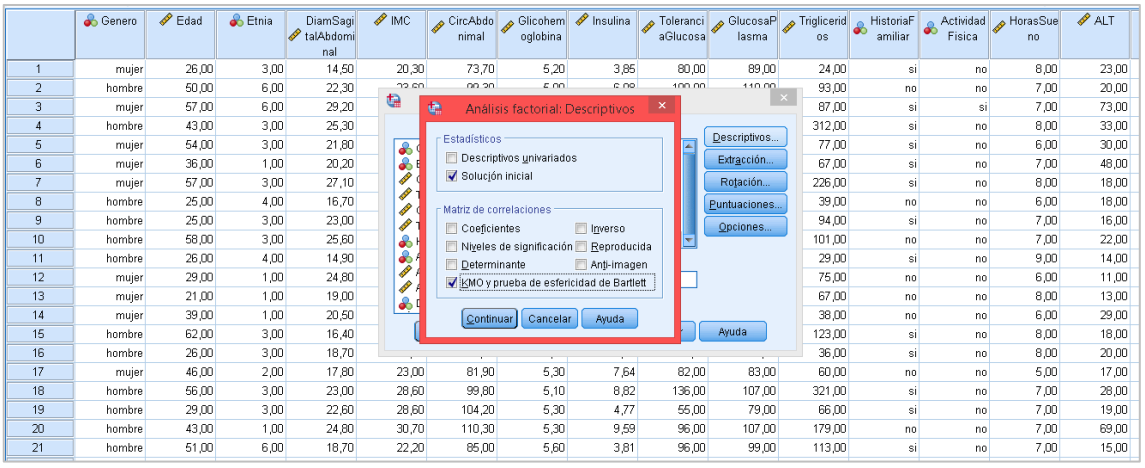


Figura 26. Descriptivos

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 27 muestra estas configuraciones del sistema.

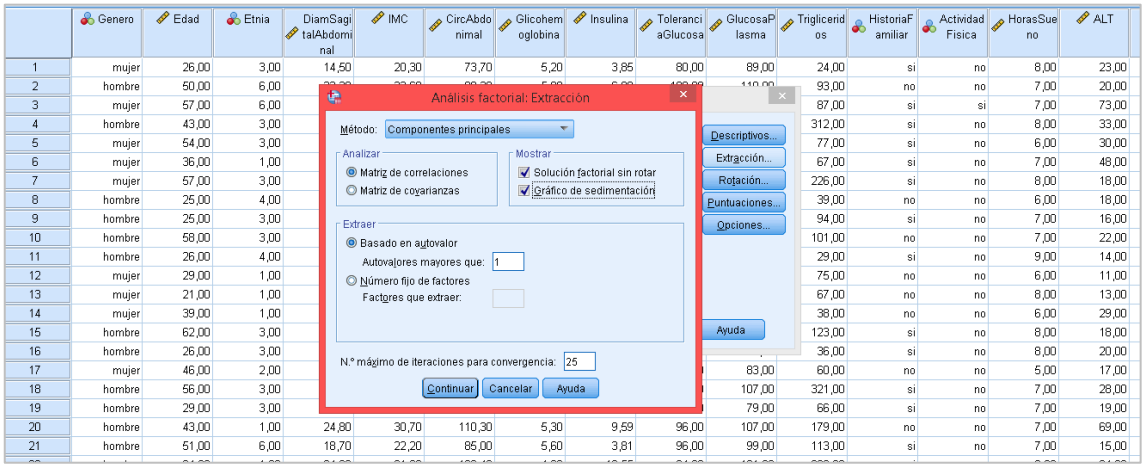


Figura 27. Configuración de extracción

Rotación: Como se aprecia en la figura 28, el método de rotación de factores escogido es el método ortogonal "Varimax" y en la opción mostrar se elige "solución rotada".

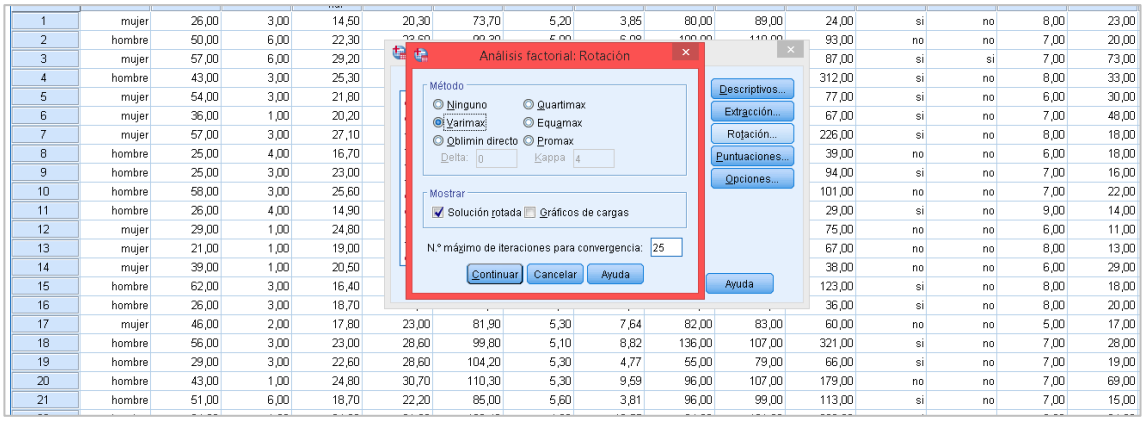
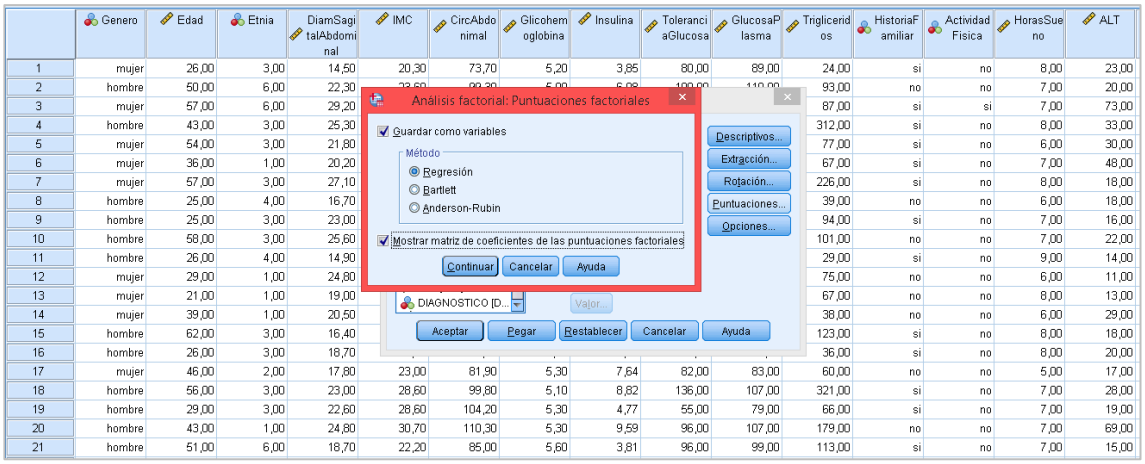


Figura 28. Rotación de factores

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 29.



	Genero	Edad	Etnia	DiamSagitalAbdominal	IMC	CircAbdominal	Glicohemoglobina	Insulina	ToleranciaGlucosa	GlucosaPlasma	Triglicéridos	HistoriaFamiliar	ActividadFisica	HorasSueno	ALT
1	mujer	26,00	3,00	14,50	20,30	73,70	5,20	3,85	80,00	89,00	24,00	si	no	8,00	23,00
2	hombre	50,00	6,00	22,30							93,00	no	no	7,00	20,00
3	mujer	57,00	6,00	29,20							87,00	si	si	7,00	73,00
4	hombre	43,00	3,00	25,30							312,00	si	no	8,00	33,00
5	mujer	54,00	3,00	21,80							77,00	si	no	6,00	30,00
6	mujer	36,00	1,00	20,20							67,00	si	no	7,00	48,00
7	mujer	57,00	3,00	27,10							226,00	si	no	8,00	18,00
8	hombre	25,00	4,00	16,70							39,00	no	no	6,00	18,00
9	hombre	25,00	3,00	23,00							94,00	si	no	7,00	16,00
10	hombre	58,00	3,00	25,60							101,00	no	no	7,00	22,00
11	hombre	26,00	4,00	14,90							29,00	si	no	9,00	14,00
12	mujer	29,00	1,00	24,80							75,00	no	no	6,00	11,00
13	mujer	21,00	1,00	19,00							67,00	no	no	8,00	13,00
14	mujer	39,00	1,00	20,50							38,00	no	no	6,00	29,00
15	hombre	62,00	3,00	16,40							123,00	si	no	8,00	18,00
16	hombre	26,00	3,00	18,70							36,00	si	no	8,00	20,00
17	mujer	46,00	2,00	17,80	23,00	81,90	5,30	7,64	82,00	83,00	60,00	no	no	5,00	17,00
18	hombre	56,00	3,00	23,00	28,60	99,80	5,10	8,82	136,00	107,00	321,00	si	no	7,00	28,00
19	hombre	29,00	3,00	22,60	28,60	104,20	5,30	4,77	55,00	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	85,00	5,60	3,81	96,00	99,00	113,00	si	no	7,00	15,00

Figura 29. Puntuaciones factoriales

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

5.5. Obtención del modelo de Análisis Factorial con SPSS

5.5.1. Pruebas de KMO y Barllett

El primer análisis que se puede apreciar en la salida del modelo son las pruebas de KMO y Barllett, 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 al test de Barllett, 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 30, 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 30. Índice KMO y test de Barllett

5.5.2. Matriz de varianza total explicada

La figura 31 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

Componente	Autovalores iniciales			Sumas de extracción de cargas al cuadrado			Sumas de rotación de cargas al cuadrado		
	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 31. Varianza total explicada

5.5.3. 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 32, se observa claramente que los dos primeros factores cumplen con la condición de que sus autovalores sean mayores a 1.

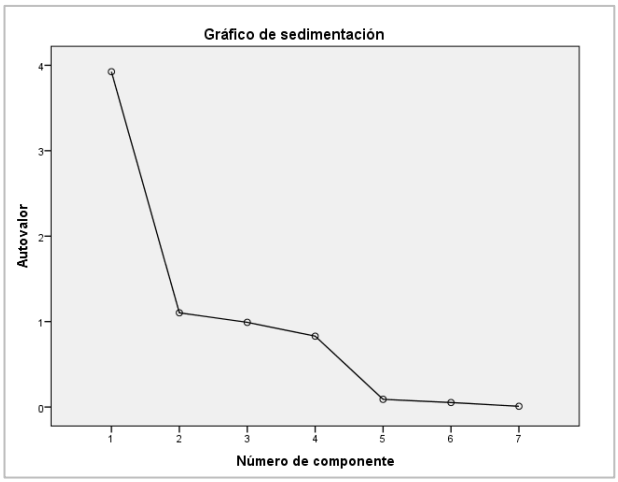
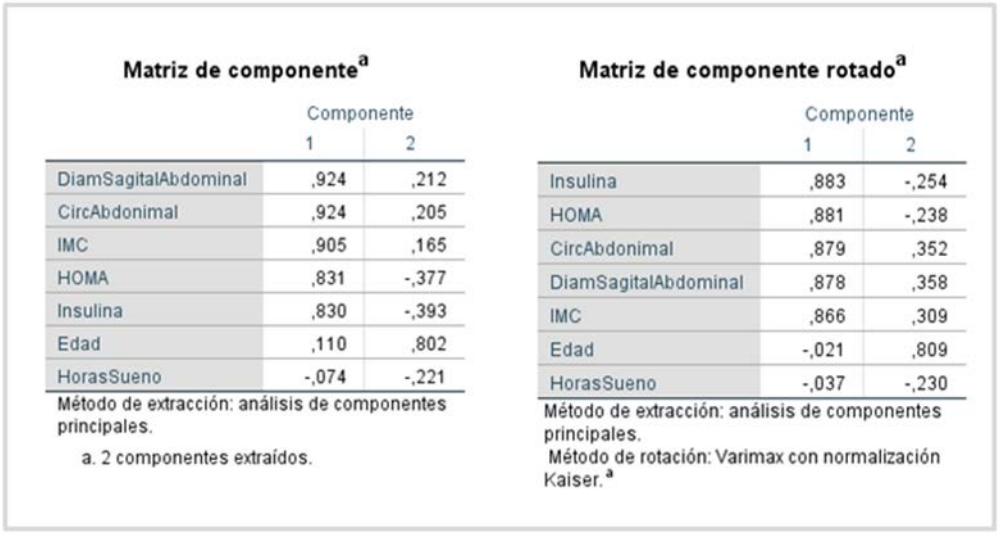


Figura 32. Gráfico de sedimentación

5.5.4. Matriz de componentes y matriz de componentes rotados

La figura 33 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	Componente 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]

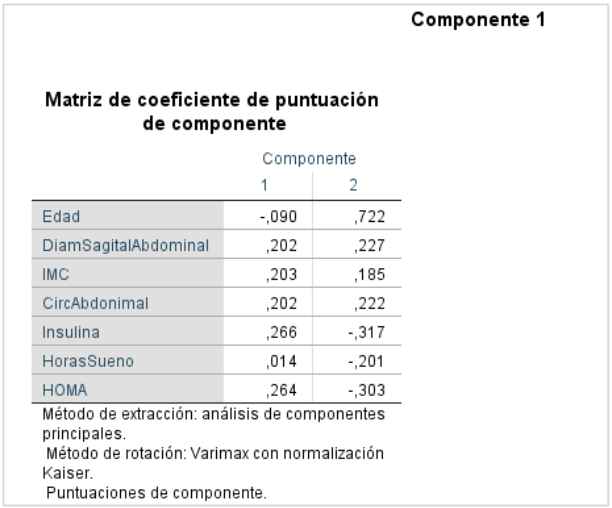
	Componente 1	Componente 2
Insulina	,883	-,254
HOMA	,881	-,238
CircAbdonimal	,879	,352
DiamSagitalAbdominal	,878	,358
IMC	,866	,309
Edad	-,021	,809
HorasSueno	-,037	-,230

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

Figura 33. Matriz de componentes rotados

5.5.5. 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 34 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	Componente 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 34. Matriz de coeficiente de puntuación de componentes

5.5.6. 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

5.5.7. 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 35, 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 35. 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**

5.6. 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 5.5.6., 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 37 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.

Desarrollo de un modelo de riesgo de prediabetes

En la figura 38 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 39), 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 35.

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 36. 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

```
187                if(indice<=-0.6880949)
188                  {
189                     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>-0.6880949 & 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>-0.6880949 & 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 física moderada.";
132                  }
133
134                 else if(indice>-0.6880949 & 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 37. Código HTML y JavaScript

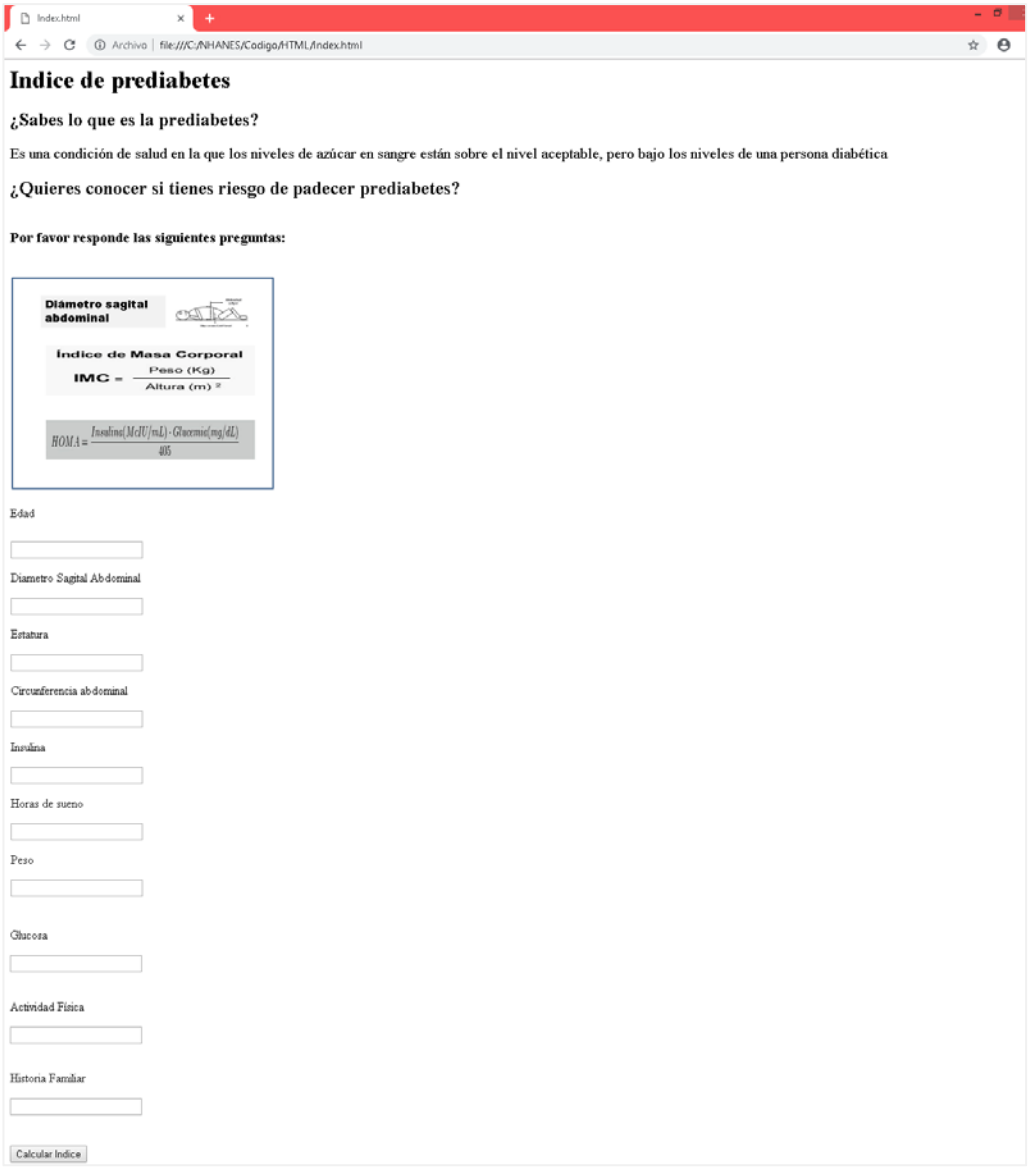


Figura 38. Estructura de la herramienta de visualización

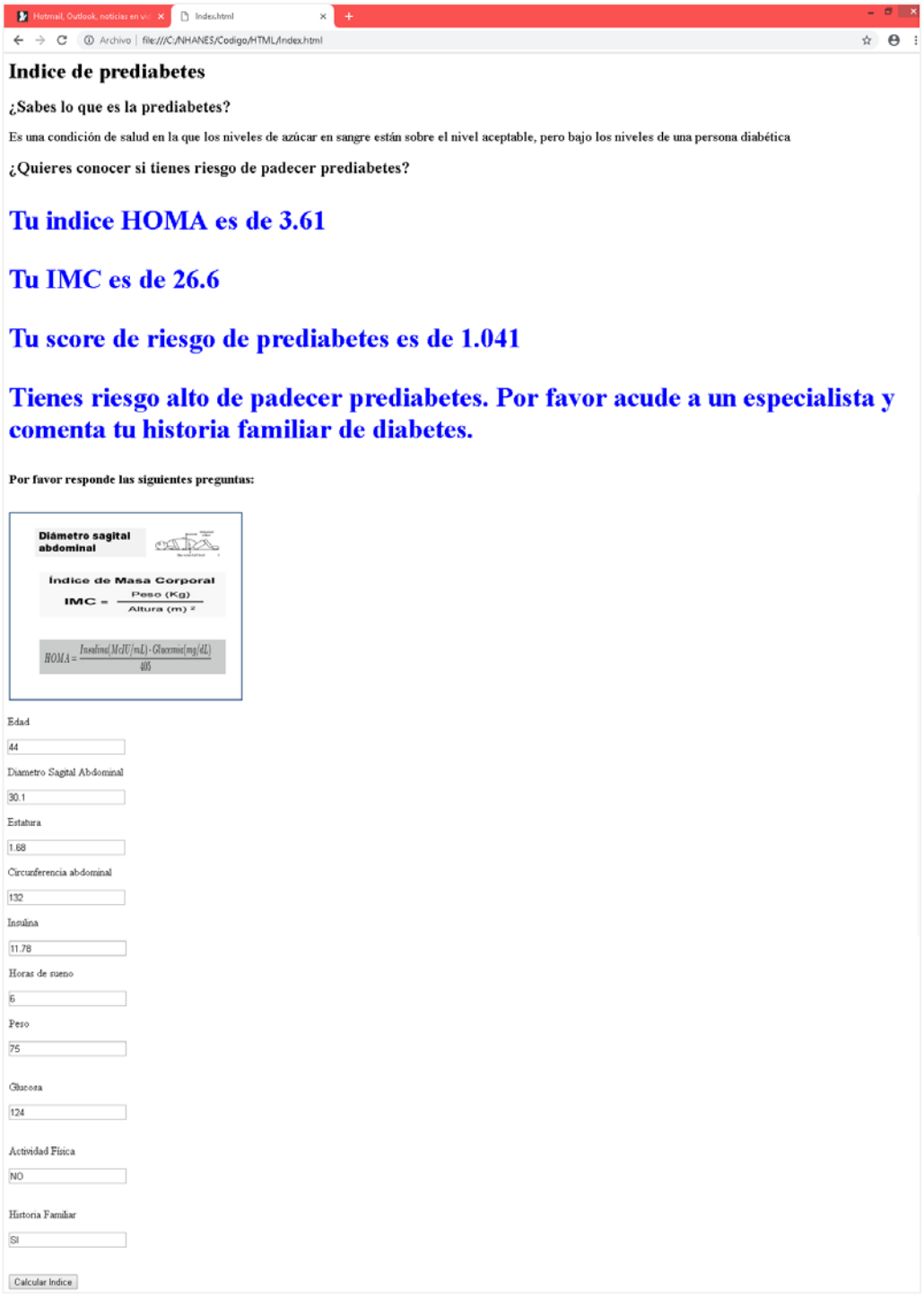


Figura 39. Ejemplo de uso de la herramienta de visualización

6. 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 Barlet, 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 40.

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 40. Descriptivos del componente 1

Si se analiza el valor de curtosis obtenido y la forma del histograma expuesto en la figura 41, se concluye que es una distribución con característica leptocúrtica, por lo tanto permite predecir con facilidad, puesto que tiene un coeficiente de variabilidad bajo.

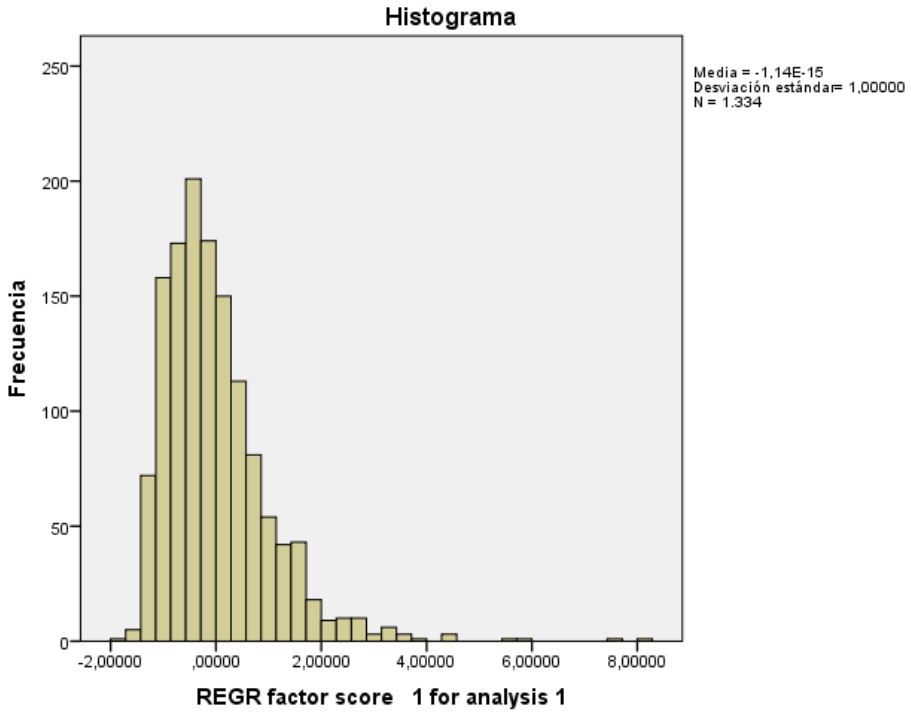


Figura 41. Histograma

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 42; estos valores atípicos se los conservó para salvaguardar la confiabilidad de los datos originales.

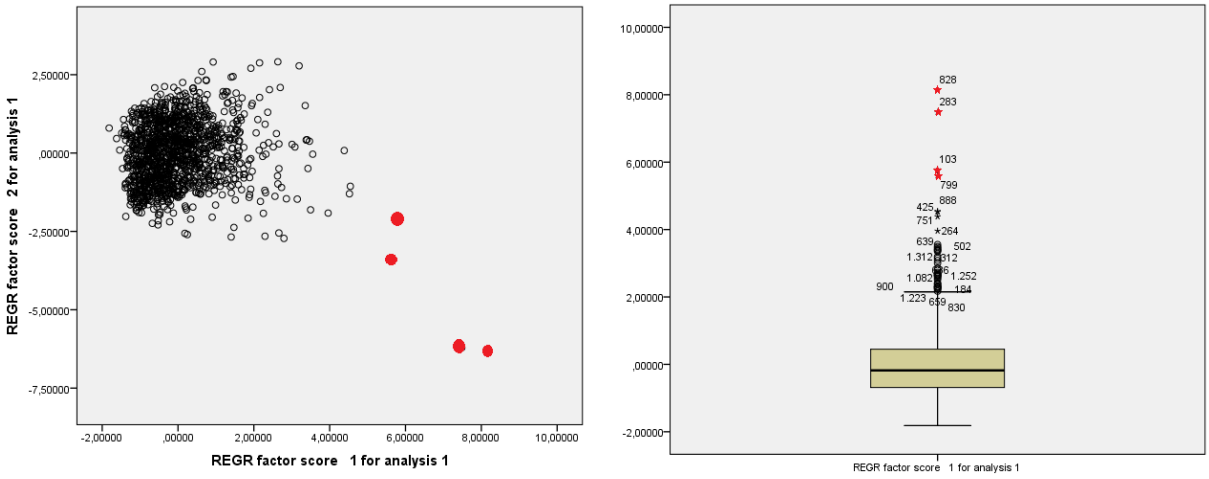


Figura 42. Reconocimiento de outliers

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.

7. Conclusiones y trabajo futuro

En este capítulo se exponen las conclusiones obtenidas y futuras líneas de investigación.

7.1. Relevancia y alcance de la contribución

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 macro proyecto 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.

7.2. Conclusiones

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.

Desarrollo de un modelo de riesgo de prediabetes

7.3. 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.

7.4. 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 Miellitus.

El objetivo a futuro es realizar un trabajo conjunto aplicando las herramientas informáticas propuestas dentro del macro proyecto 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.

Desarrollo de un modelo de riesgo de prediabetes

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

8. Referencias

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.

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), 30+.

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/FACTORIAL/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.

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., Bloomgarden, Z., Fonseca, V. Garvey, T., Gavin III, J., Grunberger, G., Horton, E., Jellinger, P., Jones, K., Lebovitz, H., Levy, P., McGuire, D., Moghissi, E., 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).

Glumer, C., Vistisen, D., Borch-Johnsen, K., &Colagiuri, S. (2006, February). Risk scores for type 2 diabetes can be applied in some populations but not all. *Diabetes Care*, 29(2), 410+.

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

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.

Khetan, A., Rajagopalan, S. (2018). Prediabetes.*Canadian Journal of Cardiology,* 34 (2018) 615e623.

Kolberg, J. A., Gerwien, R. W., Watkins, S. M., Wuestehube, L. J., & Urdea, M. (2011). Biomarkers in Type 2 diabetes: improving risk stratification with the PreDx.sup.[R] Diabetes Risk Score. *Expert Review of Molecular Diagnostics*, 11(8), 775.

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.

Mata-Cases, M., Artola, S., Encalada, J., Ezkurra-Loyola, J., Ferrer-García, J., Girbés, J., Rica, I. (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. *Semergen*, 45(1), 279-281.

Meng, X., Huang, Y., Rao, D., Zhang, Q., y Liu, Q. (2012). Comparison of three data mining models for predicting diabetes or prediabetes by risk factors. *Kaohsiung Journal of Medical Sciences, 29* (2), 93-99.

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 de : https://www.paho.org/hq/index.php?option=com_content&view=article&id=13918:obesity-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.pdf;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.

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.

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/

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.

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>

    <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="HHOMA"></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="HHOMA"></h3>

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

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

        <img src="grafico.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("insulina").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((-0.090*(field1-40.09)/13.11+0.202*(field2-21.40)/4.16+0.203*(indimc-27.85)/6.43+0.202*(field4-
                94.72)/15.47+0.266*(field5-9.92)/8.08+0.014*(field6-7.18)/1.33+0.264*(indhoma-2.43)/2.09)*1000)/1000;

               if(indice<=-0.6880949)
                 {
                    console.log("3");

                    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="Felicidades, tienes un riesgo bajo de padecer prediabetes.";
                 }

                else if(indice>-0.6880949 & indice<=-0.1789525 & field8=="SI" & field9=="NO")
                 {
                    console.log("4");

                    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 de padecer prediabetes. Por favor
```

```
                    controla tus índices anualmente.";
                }

            else if(indice>-0.6880949 & indice<=-0.1789525 & field8=="NO" & field9=="NO")
                {
                    console.log("5");

                    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 de padecer prediabetes. Por favor
                    controla tus índices anualmente y realiza 30 minutos diarios de actividad física moderada.";
                }

            else if(indice>-0.6880949 & indice<=-0.1789525 & field8=="SI" & field9=="SI")
                {
                    console.log("6");

                    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 de padecer prediabetes. Por favor
                    controla tus índices semestralmente.";
                }

            else if(indice>-0.6880949 & indice<=-0.1789525 & field8=="NO" & field9=="SI")
                {
                    console.log("7");

                    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 de padecer prediabetes. Por favor
                    controla tus índices semestralmente y realiza 30 minutos diarios de actividad física moderada.";
                }

            else if(indice>-0.1789525 & indice<=0.4512464 & 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>-0.1789525 & indice<=0.4512464 & field8=="NO" & field9=="NO")
                {
                    console.log("9");

                    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 y realiza 30 minutos diarios de actividad física moderada.";
                }

            else if(indice>1.15 & indice<=2.89 & field8=="SI" & field9=="SI")
                {
                    console.log("10");

                    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 y comentar tu historia familiar de diabetes.";
                 }

               else if(indice>-0.1789525 & 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>-0.1789525 & 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>
```