



Universidad Autónoma del Estado de Hidalgo  
Instituto de Ciencias Básicas e Ingeniería  
Licenciatura en Matemáticas Aplicadas

---

# Análisis multivariado del estado de salud de personas con Diabetes Mellitus y complicaciones asociadas

---

T E S I S

que para obtener el título de

Licenciada en Matemáticas Aplicadas

presenta

Estefany Meneses Meneses

bajo la dirección de

Dr. Roberto Ávila Pozos

Mineral de la Reforma, Hidalgo.

Abril de 2025



Mineral de la Reforma, Hgo., a 6 de junio de 2025

Número de control: ICBI-D/895/2025  
Asunto: Autorización de impresión.

**MTRA. OJUKY DEL ROCÍO ISLAS MALDONADO**  
**DIRECTORA DE ADMINISTRACIÓN ESCOLAR DE LA UAEH**

Con Título Quinto, Capítulo II, Capítulo V, Artículo 51 Fracción IX del Estatuto General de nuestra Institución, por este medio, le comunico que el Jurado asignado a la egresada de la Licenciatura en Matemáticas Aplicadas **Estefany Meneses Meneses**, quien presenta el trabajo de titulación "**Análisis multivariado del estado de salud de personas con Diabetes Mellitus y complicaciones asociadas**", ha decidido, después de revisar fundamento en lo dispuesto en el Título Tercero, Capítulo I, Artículo 18 Fracción IV; dicho trabajo en la reunión de sinodales, **autorizar la impresión del mismo**, una vez realizadas las correcciones acordadas.

A continuación, firman de conformidad los integrantes del Jurado:

**Presidente:** Dra. Alma Sofía Santillán Hernández

**Secretario:** Dra. Lourdes Cristina Carrillo Alarcón

**Vocal:** Dr. Roberto Ávila Pozos

**Suplente:** Mtra. Margarita Tetlalmatzi Montiel

Sin otro particular por el momento, reciba un cordial saludo.

Atentamente  
"Amor, Orden y Progreso"

Mtro. Gabriel Vergara Rodríguez  
Director del ICBI



GVR/YCC

Ciudad del Conocimiento, Carretera Pachuca-Tulancingo Km. 4.5 Colonia Carboneras, Mineral de la Reforma, Hidalgo, México. C.P. 42184  
Teléfono: 771 71 720 00 Ext. 40001  
direccion\_icbi@uaeh.edu.mx, vergarar@uaeh.edu.mx

"Amor, Orden y Progreso"



uaeh.edu.mx

# Índice general

|   |            |
|---|------------|
| <b>Agradecimientos</b>  | <b>III</b> |
| <b>Resumen</b>  | <b>IV</b>  |
| <b>Introducción</b>   | <b>v</b>   |
| <b>1. Antecedentes</b>  | <b>1</b>   |
| 1.1. Distribución Normal Multivariada . . . . .                               | 1          |
| 1.2. Análisis discriminante . . . . .   | 3          |
| 1.2.1. Supuestos del análisis discriminante . . . . .                         | 4          |
| 1.2.2. Formalización del análisis discriminante . . . . .                     | 4          |
| 1.2.3. Estimación de las probabilidades de clasificaciones erróneas . . . . . | 6          |
| 1.3. Regresión logística . . . . .  | 9          |
| <b>2. Material y Métodos</b>  | <b>12</b>  |
| 2.1. Diagramas de correlación . . . . .                                       | 17         |
| <b>3. Resultados</b>  | <b>20</b>  |
| 3.1. Modelo Discriminante Lineal . . . . .                                    | 20         |
| 3.2. Modelo Discriminante Cuadrático . . . . .                                | 28         |
| 3.3. Modelo Discriminante Logit . . . . .                                     | 35         |
| <b>4. Conclusiones</b>  | <b>40</b>  |

## Agradecimientos

El desarrollo de esta tesis no hubiera sido posible sin el apoyo, la paciencia, el compromiso y la guía de muchas personas, a quienes deseo expresar mi más sincero agradecimiento.

En primer lugar, al Dr. Roberto Ávila Pozos, mi director de tesis, quien con gran paciencia me ha guiado, enseñado y apoyado a lo largo de este camino. Su experiencia y acompañamiento han sido fundamentales para alcanzar este logro.

Agradezco también al jurado conformado por la Dra. Alma Sofía Santillán Hernández, la Dra. Lourdes Cristina Carrillo Alarcón y la Mtra. Margarita Tetlalmatzi Montiel, por sus valiosos comentarios, su orientación, y especialmente por su tiempo y disposición durante todo este proceso. Sus conocimientos y consejos fueron clave para la culminación de este trabajo.

A mis amigas, cuyo apoyo y compañía hicieron este camino mucho más llevadero, y a todas aquellas personas que, de manera directa o indirecta, contribuyeron a este trabajo. Llevo en este logro un pedacito de cada uno de ustedes.

Por último, a las personas más importantes de mi vida: mi familia. Mamá, papá y hermana, gracias por ser mi pilar emocional, por creer en mí en todo momento y por brindarme las herramientas necesarias para llegar hasta aquí. Su amor, comprensión y aliento fueron mi mayor fortaleza durante este reto. A mis hermanos, mi sobrina y mis sobrinos, gracias por estar.

Este logro también es de ustedes.

## Resumen

En este trabajo se utilizó información proporcionada por los Servicios de Salud de Hidalgo, específicamente de la Coordinación de Investigación en Salud. Estos registros, reúnen la información de un conjunto de pacientes con diabetes, que formaron parte del programa denominado “Clínica de Diabetes” y que fueron registrados entre los años 2003 a 2010.

Mediante técnicas de estadística multivariada, se analizaron los datos recolectados. Debido a la estructura y naturaleza de la información disponible, se utilizó análisis multivariado. El propósito de este trabajo es utilizar los datos de los registros clínicos de un conjunto de pacientes, para poder clasificar los casos que corresponden a alguno entre tres diferentes tipos de complicaciones clínicas: Retinopatía, Neuropatía y Nefropatía. Esa clasificación se realizó haciendo uso de funciones discriminantes y de funciones logit.

En la introducción se hace una revisión de la literatura, poniendo especial atención al trabajo donde se estudian complicaciones asociadas a la diabetes y las técnicas que emplearon para su análisis.

En el capítulo uno se presenta una revisión de las bases teóricas de estadística multivariada, especialmente de los análisis que se utilizaron en este trabajo. En el capítulo dos presentamos el trabajo que se realizó en el preprocesamiento de los datos, así como un análisis exploratorio de la información mediante un análisis de correlación.

En el capítulo tres presentamos los resultados. Para cada análisis mostramos los parámetros que se estimaron y se discute la relevancia de algunos de estos valores.

Por último, en el capítulo cuatro presentamos algunas conclusiones sobre el trabajo realizado.

## Introducción

La diabetes mellitus constituye uno de los principales desafíos de salud pública a nivel global. Aproximadamente 537 millones de adultos en el mundo padecen esta enfermedad, cifra que continúa en aumento [7]. En México, los datos más recientes indican que alrededor de 12.8 millones de personas viven con diabetes, lo que posiciona al país entre los de mayor prevalencia [20]. A nivel estatal, en Hidalgo, se reportó que cerca de 300,000 habitantes están afectados por esta condición [21]. Estos datos reflejan la urgencia de implementar estrategias efectivas para su prevención y control.

La diabetes mellitus, comúnmente conocida como diabetes, es un trastorno metabólico crónico caracterizado por niveles elevados de azúcar en sangre. Es una condición de salud global prevalente que afecta a millones de personas en todo el mundo. La diabetes ocurre cuando el cuerpo no produce suficiente insulina o no puede utilizar eficazmente la insulina que produce, lo que provoca niveles elevados de glucosa en la sangre [34].

La insulina, una hormona producida por el páncreas, desempeña un papel vital en la regulación de los niveles de azúcar en sangre. Permite que las células del cuerpo absorban y utilicen la glucosa, que es la principal fuente de energía. En las personas con diabetes, este delicado equilibrio de la regulación de la glucosa se altera, lo que resulta en una hiperglucemia persistente (alto nivel de azúcar en sangre).

La diabetes tipo 2, por otro lado, es la forma más común de diabetes y representa aproximadamente entre el 90 % y el 95 % de los casos. Generalmente ocurre en la edad adulta, aunque cada vez se diagnostica más en niños y adolescentes debido al aumento de las tasas de sobrepeso y obesidad. La diabetes tipo 2 se caracteriza por la resistencia a la insulina, lo que significa que las células del cuerpo no responden eficazmente a la insulina, lo que provoca niveles elevados de azúcar en sangre.

La diabetes no controlada puede tener importantes consecuencias para la salud y afectar a varios órganos y sistemas del cuerpo. Con el tiempo, puede provocar complicaciones como enfermedades cardiovasculares, daño renal, daño a los nervios (neuropatía), problemas oculares (retinopatía) y problemas de cicatrización de heridas.

La retinopatía diabética tiene una alta prevalencia en México, afectando a aproximadamente 31.5 % de la población con diabetes. Esta condición es una de las principales causas de pérdida visual y ceguera en los adultos [14]. En México, se estima que la prevalencia de neuropatía diabética periférica es mayor a 80 %, siendo más frecuente en hombres de 51 a 60 años [30]. En un estudio en México, entre 1998 y 2014, el 4.8 % de las muertes estaban relacionadas con

enfermedad renal crónica (ERC), y la mayoría se atribuía a nefropatía diabética (47.9 %) [3].

El control de la diabetes implica una combinación de modificaciones en el estilo de vida, incluida una dieta saludable, actividad física regular y control del peso, junto con medicación y/o terapia con insulina. Es fundamental que las personas con diabetes controlen sus niveles de azúcar en sangre con regularidad, sigan un plan de tratamiento completo y reciban chequeos médicos periódicos para prevenir o controlar las complicaciones [26].

En esta era de avances en el conocimiento y la tecnología médica, la investigación en curso continúa explorando nuevas vías para la prevención de la diabetes, mejores opciones de tratamiento y curas potenciales [27]. Al crear conciencia, promover la detección temprana y fomentar un entorno de apoyo, podemos trabajar para lograr un mejor manejo y una mayor calidad de vida para quienes viven con diabetes [37].

Las complicaciones asociadas a la diabetes han sido estudiadas con diferentes métodos y desde diferentes perspectivas. Jun y su equipo realizaron un estudio respecto a la relación entre la variabilidad de la glucemia y la neuropatía cardiovascular mediante análisis univariante y encontraron que todos los parámetros de variabilidad de la hemoglobina glicosilada (HbA1c) están asociados de forma independiente con la neuropatía cardiovascular. En este estudio, los datos se obtuvieron mediante un monitoreo continuo de la glucosa con 348 pacientes adultos [25].

Carreiro-Lima y colaboradores, realizaron un estudio para determinar los factores de riesgo para desarrollar retinopatía diabética (RD) y encontraron que el sexo, la edad, la altura y el peso no son factores para la aparición de RD. Como factores de riesgo encontraron el control glucémico y la duración de la DM, y relación con la presencia de nefropatía. Encontraron que los pacientes con 10 años o más del diagnóstico de diabetes y con un pobre control de la glicemia, tienen mayor riesgo de padecer retinopatía. Además, encontraron una relación entre nefropatía y el riesgo de retinopatía diabética [5].

Liu y su equipo elaboraron un metanálisis basado en la hipótesis de que la neuropatía periférica asociada a la diabetes tiene como factores la duración de la enfermedad, la edad, el nivel de HbA1c, la retinopatía diabética (RD), el tabaquismo y el índice de masa corporal (IMC). factores de riesgo. Su principal resultado es que la duración de la diabetes, la edad, el nivel de HbA1c y la presencia de retinopatía diabética son factores que aumentan el riesgo de neuropatía periférica, mientras que el IMC, el tabaquismo, los triglicéridos totales y el colesterol total no representan un mayor riesgo de diabetes neuropatía periférica. Este trabajo sugiere que los factores de riesgo para la aparición de neuropatía periférica a consecuencia de la diabetes son, el tiempo que el paciente lleva con la diabetes, la edad, el nivel de hemoglobina glicosilada, la retinopatía diabética, si es fumador y el índice de masa corporal. Sin embargo, en este estudio no se encontró que los triglicéridos totales y el colesterol total resultaran ser factores de riesgo para el desarrollo de neuropatía periférica [29].

A principios del siglo XXI, Gross y su equipo reportaron que para el caso de la Diabetes Mellitus tipo 2, la incidencia de microalbuminuria en el Reino Unido era de 2 % anual y la prevalencia diez años después del diagnóstico estaba entre el 5 y el 20 %. En ese estudio, se encontró que la nefropatía diabética tenía mayor prevalencia entre afroamericanos, asiáticos y americanos nativos, en comparación con los caucásicos [11].

Otros autores, como Lee y colaboradores, han realizado estudios a fin de relacionar diabetes y prediabetes con el riesgo de alguna o algunas complicaciones. En un estudio del 2012, asociaron la pre diabetes con el riesgo de un infarto y cómo esta relación varía con los criterios de diagnóstico de prediabetes [28].

En 2014, Al-Rubeaan y su equipo publicaron un trabajo que estudió los factores de riesgo de nefropatía, en una sociedad con una epidemia de Diabetes Mellitus tipo 2. Usando un registro de 54670 pacientes, encontraron que los factores más relevantes para el desarrollo de nefropatía fueron la hipertensión y el pobre control glicémico. Encontraron un riesgo mayor en pacientes masculinos y en aquellos cuya edad sobrepasaba los 45 años y llevaban más de 15 años con el diagnóstico de Diabetes [1].

En 2018, Aune y su grupo realizaron un meta análisis, y encontraron que la prediabetes y la diabetes incrementan el riesgo de fibrilación atrial, y que hay una relación dosis-respuesta entre el incremento en el nivel de glucosa en sangre y la fibrilación atrial [4].

Hou y su grupo publicaron un trabajo sobre factores de riesgo y prevalencia de retinopatía diabética. Utilizaron bases de datos donde se incluyeran los siguientes factores de riesgo: edad avanzada, genero masculino, duración de la DM, tratamiento de insulina, glucosa en ayunas, hemoglobina glicosilada, índice de masa corporal, colesterol total, triglicéridos y presión sistólica [18].

Por su parte, Ho y Zhang realizaron un estudio enfocado en identificar los factores de riesgo asociados con la nefropatía diabética en personas obesas que padecen diabetes mellitus tipo 2. Sus hallazgos señalaron que la hipertensión, la hiperuricemia, la presión arterial sistólica elevada y los niveles altos de hemoglobina glicosilada constituyen factores de riesgo importantes para el desarrollo de nefropatía diabética en pacientes obesos con diabetes mellitus tipo 2 [19].

Como puede notarse en esta revisión de la literatura, hay trabajo en el análisis de datos relacionados con pacientes con diabetes. Su tratamiento depende de las variables que se incluyan en el registro, además de la historia clínica de cada paciente, así como su seguimiento en el tiempo. Por lo tanto, no existe un procedimiento único para analizar los registros de pacientes con diabetes, que han estado participando en un seguimiento regular de su control de glucosa y su apego a la dieta y a la alimentación. Este seguimiento es de suma importancia para generar evidencia que permita hacer más y mejores inferencias.

El objetivo de este trabajo es emplear el análisis multivariado para clasificar, con base en un conjunto de variables registradas en la Clínica de Diabetes, a los pacientes registrados en algún grupo que desarrolla una complicación asociada a la diabetes (entre retinopatía, neuropatía y nefropatía). Para llevar a cabo esta clasificación, se utilizaron tanto funciones discriminantes como modelos logit. En el primer enfoque, la función discriminante asigna a cada paciente a uno de los grupos definidos (cada una de las complicaciones), en función del valor que toma dicha función para cada observación. En el segundo enfoque, modelo logit, la clasificación se realiza a partir de la estimación de la probabilidad de que un paciente desarrolle alguna de las complicaciones consideradas.



## Antecedentes

### 1.1. Distribución Normal Multivariada

La distribución normal multivariada es un concepto fundamental en estadística. Se utiliza para representar conjuntos de variables aleatorias que están correlacionadas entre sí. Esta distribución es una extensión de la distribución normal para el caso univariado, que describe una sola variable aleatoria.

En el contexto de la distribución normal multivariada, el interés se centra en comprender cómo se distribuyen conjuntamente varias variables aleatorias continuas. Cada conjunto de variables aleatorias multivariadas se puede caracterizar por un vector de medias  $\mu$  y una matriz de covarianza  $\Sigma$ , que describen, respectivamente, la tendencia central y la relación entre las variables.

La distribución normal multivariada es utilizada en muy distintos campos, como la estadística, la econometría, la ingeniería y la ciencia de datos. Se ha utilizado en la modelización de datos en los que las variables están correlacionadas de alguna manera, lo que la hace especialmente valiosa en el análisis de datos multivariados y los pronósticos.

Definición. Decimos que un vector aleatorio de  $p$  dimensiones, es decir

$$\mathbf{X} = (X_1, X_2, \dots, X_p)$$

tiene una distribución normal multivariada si su función generadora de momentos es:

$$M_{\mathbf{X}}(t) = \exp\left(t^T \mu + \frac{1}{2} t^T \Sigma t\right), \quad (1.1)$$

para todo  $t \in \mathbb{R}^p$ , donde  $\Sigma$ , la matriz de varianzas y covarianzas, que es una matriz simétrica y semidefinida positiva y  $\mu \in \mathbb{R}^p$  donde  $\mu$  es el vector de medias de  $\mathbf{X}$ . Abreviaremos esto diciendo que  $\mathbf{X}$  tiene una distribución normal  $p$ -variada, con vector de medias  $\mu$  y matriz de varianzas y covarianzas  $\Sigma$  ( $N_p(\mu, \Sigma)$ ).

$$\mu = (\mu_1, \mu_2, \dots, \mu_p)^T$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{p1} & \sigma_{p2} & & \sigma_{pp} \end{pmatrix}$$

La función de densidad de probabilidad de  $\mathbf{X}$  es:

$$f_X(x) = \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} \exp \left\{ -\frac{1}{2} (x - \mu)^T \boldsymbol{\Sigma}^{-1} (x - \mu) \right\}, \quad \text{para } x \in \mathbb{R}^p. \quad (1.2)$$

**Teorema.** Supongamos que  $\mathbf{X}$  tiene una distribución  $N_p(\mu, \boldsymbol{\Sigma})$ . Sea  $\mathbf{Y} = \mathbf{A}\mathbf{X} + b$ , donde  $A$  es una matriz  $q \times p$  y  $b \in \mathbb{R}^q$ . Entonces,  $\mathbf{Y}$  tiene una distribución  $N_q(\mathbf{A}\mu + b, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T)$ .

**Demostración:** Notemos que para  $t \in \mathbb{R}^q$ , la función generadora de momentos (mgf) de  $\mathbf{Y}$  es

$$\begin{aligned} M_Y(t) &= E[\exp\{t^T \mathbf{Y}\}] \\ &= E[\exp\{t^T (\mathbf{A}\mathbf{X} + b)\}] \\ &= \exp\{t^T b\} E[\exp\{(\mathbf{A}^T t)^T \mathbf{X}\}] \\ &= \exp\{t^T b\} \exp\{(\mathbf{A}^T t)^T \mu + \frac{1}{2} (\mathbf{A}^T t)^T \boldsymbol{\Sigma} (\mathbf{A}^T t)\} \\ &= \exp\{t^T (\mathbf{A}\mu + b) + \frac{1}{2} t^T \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T t\}. \end{aligned}$$

que es la mgf de una distribución  $N_q(\mathbf{A}\mu + b, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T)$ . ■

Si quiere explorar este tema con mayor profundidad, consulte el libro de Hogg [15].

## 1.2. Análisis discriminante

El análisis discriminante lineal es el método más común de Análisis Discriminante. Se trata de una técnica basada en el análisis propio (eigenanalysis) y asume que los datos son distribuidos normalmente. Como su nombre indica, el análisis discriminante basado en la distancia [2] puede aplicarse a cualquier tipo de datos, incluidos los datos a nivel de comunidad y tipos de datos como las variables categóricas que no pueden incorporarse al análisis discriminante lineal. Irigoien y colaboradores [22] describen cómo pueden utilizarse estas técnicas en biología (por ejemplo, el análisis morfométrico para la identificación de especies), en ecología (por ejemplo, la separación de nichos por especies simpátricas) y en medicina (por ejemplo, para el diagnóstico de enfermedades).

El análisis discriminante lineal se ha utilizado para diversas aplicaciones ecológicas, como describen McCune y Grace [?].

La razón más común para utilizar el análisis discriminante lineal es evaluar en qué medida las observaciones de diferentes grupos pueden diferenciarse unas de otras en función de sus respuestas medidas.

El análisis discriminante se utiliza para crear un modelo predictivo que permita elucidar la pertenencia a determinado grupo. El modelo se compone por una función discriminante (en el caso de más de dos grupos, se tiene un conjunto de funciones discriminantes). Esta función discriminante se basa en combinaciones lineales de las variables que se consideran predictoras, y que proporcionan la mejor discriminación posible entre los grupos. Dichas funciones discriminantes se generan a partir de una muestra de casos en los que se conoce con certeza el grupo al que pertenecen. Posteriormente, las funciones discriminantes pueden aplicarse a nuevos conjuntos de datos de mediciones para las variables predictoras en los que se desconoce a qué grupo pertenecen.

Hay características similares en el modelo de regresión lineal y la función (como la forma del modelo como una combinación lineal de las variables). Algunos de los cálculos son parecidos en ambos casos, pero tienen algunas algunas diferencias estructurales.

1. En el análisis de regresión se supone que la variable dependiente tiene distribución normal, y que los regresores son fijos. En cambio, en el análisis discriminante las variables independientes se consideran con distribución normal, y la variable de respuesta se supone fija, y sólo toma los valores cero o uno, dependiendo de la ubicación del objeto en un grupo o en otro.
2. El principal objetivo del análisis de regresión es predecir la respuesta media, a partir de la información de algunos valores fijos de un conjunto de variables, mientras que el análisis discriminante busca encontrar una combinación lineal de variables independientes, tal que se minimice la probabilidad de realizar una mala clasificación.
3. El análisis de regresión propone un modelo, con un conjunto de supuestos, a fin de contar con estimadores de los parámetros con algunas propiedades deseables. El análisis discriminante se centra en la búsqueda de un procedimiento eficiente para la correcta asignación o clasificación de casos a grupos.

### 1.2.1. Supuestos del análisis discriminante

Supongamos que la población puede dividirse en subconjuntos o grupos mutuamente excluyentes. En nuestro caso, la población de pacientes diabéticos se divide de acuerdo a la complicación que presentan: Grupo 1 con retinopatía, Grupo 2 con nefropatía y Grupo 3 con neuropatía.

Como en toda la inferencia estadística, el análisis discriminante debe satisfacer algunos supuestos:

1. Los casos deben ser independientes.
2. El conjunto de variables predictoras debe tener una distribución normal multivariada
3. Las matrices de varianzas y covarianzas intra-grupos, deben ser iguales en todos los grupos.
4. La pertenencia a un grupo es mutuamente excluyente (ningún caso pertenece a más de un grupo).
5. La pertenencia a un grupo es exhaustiva de modo colectivo (todos los casos son miembros del grupo).

### 1.2.2. Formalización del análisis discriminante

El Análisis Discriminante es un conjunto de herramientas utilizados para determinar cómo asignar nuevas observaciones a los grupos de una población.

Una suposición para este análisis es que los grupos tienen distribución normal multivariada, con vector de medias  $\mu_i$  y matriz de varianzas y covarianzas  $\Sigma$ . Sea  $G_j, j = 1, 2, \dots, J$ , los conjuntos de un grupo, y debemos asignar una observación  $x$ , a uno de estos grupos  $G_j$ . En este trabajo se consideran tres grupos; sin embargo, la descripción metodológica se presenta para el caso de dos grupos, dado que puede generalizarse a un número mayor de grupos.

Una regla discriminante es una separación del espacio muestral en conjuntos  $R_j$  tales que si  $x \in R_j$ , se identifica como un miembro de la población  $G_j$ . Algunas reglas discriminantes son:

1. Regla de máxima verosimilitud, dado que los datos son de poblaciones  $G_j$  con densidades  $f_j, j = 1, 2$ , la regla discriminante de máxima verosimilitud asigna una observación  $x$  a esa población  $i$  que tiene la máxima verosimilitud  $L_j(x) = f_j(x) = \max_i f_i(x)$ . Esta regla se fundamenta en combinaciones lineales y forma parte de los métodos de la familia del análisis discriminante lineal.
2. Dadas las probabilidades previas  $\pi_j$  para poblaciones  $G_j$ , la Regla Discriminante de Bayes asigna una observación  $x$  a la población  $G_j$  que maximiza  $\pi_j f_i(x)$  con respecto a  $i$ .

Todas la reglas discriminantes de Bayes son admisibles; además decimos que una regla discriminante es admisible si no existe una regla discriminante mejor.

Suponga que  $G_i \sim N_p(\mu_i, \Sigma)$ ,  $\mu_i$  el vector de medias,  $\Sigma$  la matriz de varianzas y covarianzas y  $\mathbf{X}$  el vector de  $p$  variables, de manera que la verosimilitud está dada por

$$L_i(\mathbf{X}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{X} - \mu)^T \Sigma^{-1} (\mathbf{X} - \mu) \right\}$$

Maximizar la expresión anterior es equivalente a obtener el mínimo de

$$(\mathbf{X} - \mu_i)^T \Sigma^{-1} (\mathbf{X} - \mu_i).$$

Entonces, se asigna  $\mathbf{X}$  al grupo  $G_1$  si

$$(\mathbf{X} - \mu_1)^T \Sigma^{-1} (\mathbf{X} - \mu_1) \leq (\mathbf{X} - \mu_2)^T \Sigma^{-1} (\mathbf{X} - \mu_2),$$

o bien, si se cumple

$$-2\mu_1^T \Sigma^{-1} \mathbf{X} + \mu_1^T \Sigma^{-1} \mu_1 + 2\mu_2^T \Sigma^{-1} \mathbf{X} - \mu_2^T \Sigma^{-1} \mu_2 \leq 0$$

$$2(\mu_2 - \mu_1)^T \Sigma^{-1} \mathbf{X} + (\mu_1 - \mu_2)^T \Sigma^{-1} (\mu_1 + \mu_2) \leq 0$$

$$(\mu_1 - \mu_2)^T \Sigma^{-1} \mathbf{X} - \frac{1}{2} (\mu_1 + \mu_2)^T \Sigma^{-1} (\mu_1 - \mu_2) > 0.$$

A la función

$$\mathbf{Y} = \mathbf{b}^T \mathbf{X}, \quad \text{con} \quad \mathbf{b} = \Sigma^{-1} (\mu_1 - \mu_2) \tag{1.3}$$

se le conoce como *función discriminante lineal*, y es una combinación lineal de las medidas asociadas con las variables para una unidad experimental particular (individuo u objeto).

Asumiendo que las dos poblaciones se conocen mediante sus funciones de distribución de probabilidad, la descripción anterior es de utilidad. Sin embargo, en la práctica los parámetros que determinan esas distribuciones se estiman a partir de muestras aleatorias independientes.

Suponga que tiene una muestra aleatoria de una población con distribución normal  $N_p(\mu_j, \Sigma)$ , si tenemos  $J$  grupos con  $n_j$  observaciones en cada grupo, usamos  $\bar{\mathbf{X}}_j$  para estimar  $\mu_j$  y  $S_j$  para estimar  $\Sigma$ .

Los estimadores son

$$\bar{\mathbf{X}}_j = \sum_{i=1}^{n_j} \frac{\mathbf{X}_{ij}}{n_j}$$

$$\mathbf{S} = \frac{1}{n_1 + n_2 - 2} \left[ \sum_{i=1}^{n_1} (\mathbf{X}_{i1} - \bar{\mathbf{X}}_1)(\mathbf{X}_{i1} - \bar{\mathbf{X}}_1)^T + \sum_{i=1}^{n_2} (\mathbf{X}_{i2} - \bar{\mathbf{X}}_2)(\mathbf{X}_{i2} - \bar{\mathbf{X}}_2)^T \right]$$

En estos términos, la función discriminante muestral se puede escribir como

$$\hat{\mathbf{Y}} = \hat{\mathbf{b}}^T \mathbf{X}.$$

Si  $\hat{\mathbf{Y}} = \hat{\mathbf{b}}^T \mathbf{X} \geq \hat{\mathbf{b}}^T \bar{\mathbf{X}}_c$ , se asigna  $\mathbf{X}$  al grupo uno, donde

$$\bar{\mathbf{X}}_c = \frac{1}{2}(\bar{\mathbf{X}}_1 + \bar{\mathbf{X}}_2) \quad \text{y} \quad \hat{\mathbf{b}} = \mathbf{S}^{-1}(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2).$$

Si  $\hat{\mathbf{Y}} = \hat{\mathbf{b}}^T \mathbf{X} < \hat{\mathbf{b}}^T \bar{\mathbf{X}}_c$ , se asigna  $\mathbf{X}$  al grupo dos. Si desea consultar más detalles, los puede encontrar en el libro de Chatfield [6].

### 1.2.3. Estimación de las probabilidades de clasificaciones erróneas

En el análisis discriminante, [13] las probabilidades de clasificación errónea cuantifican la precisión de las reglas de asignación, es decir, miden la probabilidad de que una observación sea asignada a un grupo incorrecto. Estas probabilidades se pueden expresar como:

$$p_{ij} = \int \phi_i(\mathbf{x}) f_j(\mathbf{x}) d\mathbf{x},$$

donde:

- $p_{ij}$  es la probabilidad de asignar incorrectamente una observación de la población  $G_j$  a la clase  $G_i$ .
- $\phi_i(x)$  es la regla discriminante aleatorizada, que indica con qué probabilidad se asigna  $\mathbf{x}$  a la población  $G_i$ .
- $f_j(\mathbf{x})$  es la función de densidad de la población  $G_j$ , es decir, la probabilidad de que una observación  $x$  provenga de  $G_j$ .

Esta ecuación mide la probabilidad de clasificación correcta o errónea comparando una observación con las distribuciones de las poblaciones involucradas.

Dadas dos poblaciones  $G_1$  y  $G_2$  con distribuciones normales multivariadas:

$$G_1 : N_p(\mu_1, \Sigma)$$

$$G_2 : N_p(\mu_2, \Sigma)$$

La regla de máxima verosimilitud asigna una observación  $\mathbf{x}$  a  $G_1$  si:

$$(\mathbf{x} - \mu_1)^T \Sigma^{-1} (\mathbf{x} - \mu_1) \leq (\mathbf{x} - \mu_2)^T \Sigma^{-1} (\mathbf{x} - \mu_2)$$

Esto se simplifica a

$$\mathbf{w}^T \mathbf{x} \geq c$$

donde  $\mathbf{w} = \Sigma^{-1}(\mu_1 - \mu_2)$  y  $c = \frac{1}{2}(\mu_1 + \mu_2)^T \Sigma^{-1}(\mu_1 - \mu_2)$

De modo que la probabilidad de clasificar incorrectamente una observación de  $G_1$  en  $G_2$  es:

$$p_{12} = P(\mathbf{w}^T \mathbf{x} < c \mid \mathbf{x} \in G_1)$$

Dado que  $\mathbf{w}^T \mathbf{x}$  para  $\mathbf{x} \in G_1$  tiene distribución normal:

$$\mathbf{w}^T \mathbf{x} \sim N(\mu_1^T \Sigma^{-1}(\mu_1 - \mu_2), (\mu_1 - \mu_2)^T \Sigma^{-1}(\mu_1 - \mu_2))$$

La varianza es  $\delta^2 = (\mu_1 - \mu_2)^T \Sigma^{-1}(\mu_1 - \mu_2)$ . Estandarizando, obtenemos:

$$p_{12} = P\left(Z < \frac{c - \mu_1^T \Sigma^{-1}(\mu_1 - \mu_2)}{\delta}\right)$$

donde  $Z \sim N(0, 1)$ . Sustituyendo  $c$ :

$$p_{12} = P\left(Z < \frac{\frac{1}{2}(\mu_1 + \mu_2)^T \Sigma^{-1}(\mu_1 - \mu_2) - \mu_1^T \Sigma^{-1}(\mu_1 - \mu_2)}{\delta}\right)$$

Simplificando el numerador:

$$\frac{1}{2}(\mu_1 + \mu_2 - 2\mu_1)^T \Sigma^{-1}(\mu_1 - \mu_2) = \frac{1}{2}(-\mu_1 + \mu_2)^T \Sigma^{-1}(\mu_1 - \mu_2) = -\frac{1}{2}\delta^2$$

Por lo tanto:

$$p_{12} = P\left(Z < -\frac{1}{2}\delta\right) = \Phi\left(-\frac{1}{2}\delta\right)$$

De manera análoga,  $p_{21} = \Phi\left(-\frac{1}{2}\delta\right)$ .

Así que las probabilidades de clasificación errónea se estiman como:

$$\hat{p}_{12} = \hat{p}_{21} = \Phi\left(-\frac{1}{2}\hat{\delta}\right), \quad (1.4)$$

donde:

$$\hat{\delta}^2 = (\bar{x}_1 - \bar{x}_2)^T S^{-1}(\bar{x}_1 - \bar{x}_2),$$

es el estimador de  $\delta^2$ ,  $\Phi$  es la función de distribución acumulativa de la normal estándar,  $\bar{x}_1$  y  $\bar{x}_2$  son los vectores de medias muestrales, y  $S$  es la matriz de covarianza conjunta. Este

estimador, basado en la distancia de Mahalanobis, refleja qué tan separadas están las medias de las poblaciones, lo que influye directamente en la precisión de la clasificación.

Otra manera de estimar las probabilidades de clasificación incorrecta es mediante el método de re-sustitución, en el cual cada observación original se reclasifica usando la regla discriminante. Si  $n_{ij}$  denota el número de observaciones de la población  $G_j$  clasificadas en  $G_i$ , el estimador de las probabilidades de clasificación errónea es:

$$\hat{p}_{ij} = \frac{n_{ij}}{n_j}, \quad (1.5)$$

donde  $n_j$  es el número total de observaciones en  $G_j$ . Aunque este método tiende a subestimar los errores, proporciona una aproximación rápida a la precisión de la regla discriminante. La matriz de confusión, construida a partir de los estimadores  $\hat{p}_{ij}$ , permite visualizar cómo se distribuyen los errores de clasificación entre las distintas categorías, facilitando la evaluación global del rendimiento del clasificador.

## 1.3. Regresión logística

La mayor parte de esta sección fue obtenida del libro de Hosmer [17].

El modelo de regresión logística se utiliza para describir la relación entre una variable dependiente o de respuesta ( $Y$ ) y un conjunto de variables independientes o explicativas ( $x_1 \dots x_k$ ). Este modelo es particularmente útil cuando la variable de resultado es dicotómica, es decir, cuando asume solo dos posibles valores. Para simplificar la notación, denotamos por  $\pi(x) = E(Y|x)$  la media condicional de  $Y$  dado  $x$ , bajo una distribución logística. La forma funcional específica del modelo de regresión logística está dada por la ecuación:

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k}}, \quad (1.6)$$

donde  $\beta_0, \beta_1, \dots, \beta_k$  son los parámetros a estimar.

La transformación logit convierte esta probabilidad en una combinación lineal de las variables independientes:

$$g(x) = \ln \left( \frac{\pi(x)}{1 - \pi(x)} \right) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k. \quad (1.7)$$

Esta transformación es clave porque permite modelar la relación entre las variables predictoras y la probabilidad del evento en una escala lineal. La relevancia de esta transformación radica en que  $g(x)$  posee varias propiedades deseables de los modelos de regresión lineal: el logit es lineal respecto a sus parámetros, puede ser continuo y varía entre  $-\infty$  y  $+\infty$ , dependiendo del rango de  $x$ .

En el modelo de regresión lineal, se asume que una observación de la variable de resultado puede expresarse como  $y = E(Y|x) + \varepsilon$ , donde  $\varepsilon$  representa el error que mide la desviación de una observación respecto a la media condicional. Se supone que  $\varepsilon$  sigue una distribución normal con media cero y varianza constante.

No obstante, cuando la variable de resultado es dicotómica, este enfoque no es adecuado. En este caso, se puede expresar  $y$  como  $y = \pi(x) + \varepsilon$ , donde  $\varepsilon$  sigue una distribución con media cero y varianza  $\pi(x)[1 - \pi(x)]$ .

Supongamos que tenemos una muestra de  $n$  observaciones independientes del par  $(x_i, y_i)$ , para  $i = 1, 2, \dots, n$ , donde  $y_i$  denota una variable de resultado dicotómica y  $x_i$  es el valor de la variable independiente para el  $i$ -ésimo sujeto. El ajuste del modelo de regresión logística a un conjunto de datos requiere la estimación de los parámetros  $\beta_0$  y  $\beta_1$ .

A diferencia de la regresión lineal, en modelos con variables dependientes dicotómicas no es adecuado utilizar mínimos cuadrados, debido a las propiedades no deseables que presenta en este contexto.

Notemos que con el método de mínimos cuadrados ordinarios se pueden obtener predicciones fuera del intervalo  $[0, 1]$ , lo cual resulta problemático al interpretarlas como probabilidades, ya que estas deben estar restringidas a ese rango.

Por otro lado, en la regresión lineal, se asume que los errores tienen distribución normal y varianza constante (homocedasticidad). Pero con variables dicotómicas, los errores no pueden ser normales ya que si la variable es dicotómica y solo toma dos valores (1 o 0), entonces

$$\text{Si } y = 1, \varepsilon_i = 1 - \beta_0 - \beta_1 x_i$$

$$\text{Si } y = 0, \varepsilon_i = -\beta_0 - \beta_1 x_i$$

De modo que  $\varepsilon$  toma valores discretos. Además, hay heterocedasticidad, ya que la varianza del error depende del valor esperado de la variable dependiente:

$$\text{Var}(\Sigma_i) = p_i(1 - p_i), \text{ donde } p_i \text{ es la probabilidad de que la variable sea 1.}$$

Por lo tanto, en su lugar se utiliza el método de máxima verosimilitud, el cual busca los valores de los parámetros que maximizan la probabilidad de observar los datos dados.

Para aplicar este método, construimos la función de verosimilitud. Si  $Y$  está codificada como 0 o 1, entonces la probabilidad condicional de que  $Y = 1$  dado  $x$  es  $\pi(x)$ , y la probabilidad condicional de que  $Y = 0$  es  $1 - \pi(x)$ .

Así, para cada par  $(x_i, y_i)$ , donde  $y_i = 1$ , la contribución a la función de verosimilitud es  $\pi(x_i)$ ; y cuando  $y_i = 0$ , la contribución es  $1 - \pi(x_i)$ . De manera compacta, la función de verosimilitud se expresa como:

$$l(\boldsymbol{\beta}) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \quad (1.8)$$

El principio de máxima verosimilitud establece que nuestra estimación de  $\boldsymbol{\beta}$  (vector de parámetros a estimar) será el valor que maximiza la expresión de la ecuación (1.8). No obstante, es más manejable desde el punto de vista matemático trabajar con el logaritmo de la función de verosimilitud. La expresión resultante, conocida como la log-verosimilitud, se define como:

$$L(\boldsymbol{\beta}) = \ln[l(\boldsymbol{\beta})] = \sum_{i=1}^n \{y_i \ln[\pi(x_i)] + (1 - y_i) \ln[1 - \pi(x_i)]\} \quad (1.9)$$

Para encontrar el valor de  $\boldsymbol{\beta}$  que maximiza  $L(\boldsymbol{\beta})$ . Sea el caso para  $\boldsymbol{\beta} = (\beta_0, \beta_1)$  diferenciamos  $L(\boldsymbol{\beta})$  con respecto a  $\beta_0$  y  $\beta_1$  y luego igualamos las derivadas a cero. Estas expresiones resultantes, conocidas como las ecuaciones de verosimilitud, son:

$$\sum_{i=1}^n [y_i - \pi(x_i)] = 0 \quad (1.10)$$

y

$$\sum_{i=1}^n x_i [y_i - \pi(x_i)] = 0 \quad (1.11)$$

Las ecuaciones (1.10) y (1.11) no son lineales en  $\beta_0$  y  $\beta_1$ , lo que exige métodos iterativos para su solución. El caso general se sigue de este.

El valor de  $\beta$ , obtenido a partir de la solución de las ecuaciones (1.10) y (1.11), se denomina estimación de máxima verosimilitud y se denota por  $\hat{\beta}$ . El símbolo “^” indica la estimación de máxima verosimilitud de la cantidad correspondiente. Por ejemplo,  $\hat{\pi}(x_i)$  es la estimación de máxima verosimilitud de  $\pi(x_i)$ , que proporciona una estimación de la probabilidad condicional de que  $Y = 1$  dado  $x = x_i$ . Este valor representa el ajuste o predicción del modelo de regresión logística.

Una consecuencia interesante de la ecuación (1.10) es que la suma de los valores observados de  $y$  es igual a la suma de los valores predichos o esperados de  $Y$ :

$$\sum_{i=1}^n y_i = \sum_{i=1}^n \hat{\pi}(x_i).$$

Partimos de (1.10):

$$\sum_{i=1}^n [y_i - \pi(x_i)] = 0$$

$$\sum_{i=1}^n y_i - \sum_{i=1}^n \pi(x_i) = 0$$

$$\sum_{i=1}^n y_i = \sum_{i=1}^n \pi(x_i)$$

Dado que  $\hat{\pi}(x_i)$  es la estimación de  $\pi(x_i)$  bajo el modelo ajustado y los estimadores satisfacen las ecuaciones de máxima verosimilitud, se cumple:

$$\sum_{i=1}^n y_i = \sum_{i=1}^n \hat{\pi}(x_i)$$

Esto demuestra que la suma de las respuestas observadas en los datos es igual a la suma de las probabilidades estimadas por el modelo.

## Material y Métodos

La Coordinación de Investigación en Salud, de los Servicios de Salud de Hidalgo (SSH), nos proporcionó una base de datos que fueron construyendo en el programa que se denominó Clínica de Diabetes. Este programa inició en 2003 y a partir del año 2010 estuvieron disponibles los primeros datos. La base fue creciendo y se fue actualizando para estar acorde con los registros y notación nacional. Los datos corresponden a una observación por paciente que presentan Diabetes Mellitus tipo 2. La base de datos cuenta con un total de 67 variables registradas. Los datos corresponden a una observación por paciente que presentan Diabetes Mellitus. Para el presente trabajo, solo se emplearon 16 variables, que se muestran en la Tabla 2.1. Las variables consideradas son aquellas que contaban con el mayor número de registros completos. Variables como Creatinina, Glucosuria o Cetonuria, tienen apenas unas decenas de datos del total de 53428 registros. Para el caso de la Glucosa, se tomaron en cuenta aquellos registros de glucosa en sangre (en mg/dl), y quedaron fuera aquellas mediciones hechas a través de hemoglobina glicosilada (HbA1c), que mide el nivel promedio de glucosa en la sangre durante los últimos tres meses [35].

En la base de datos original se contaba con dos variables: ANTIDIABETICO1 y ANTIDIABETICO2, las cuales indican los tratamientos farmacológicos antidiabéticos que consume cada paciente. Estas variables pueden contener el nombre de alguno de los siguientes medicamentos: metformina, glibenclamida, tolbutamida, clorpropamida, glimepirida, rosiglitazona, nateglinida, fenformina, acarbosa, glipizida, pioglitazona, Januvia, repaglinida y NPH (insulina). En caso de que el paciente no consuma ningún antidiabético, ambas variables se codifican con cero.

Cada variable puede registrar un medicamento diferente, por lo que un paciente puede estar en una de las siguientes situaciones: no consumir ningún antidiabético, consumir un solo antidiabético (registrado en ANTIDIABETICO1), o consumir dos antidiabéticos distintos (uno registrado en cada variable).

Así que, para este análisis se construyó una nueva variable denominada *Antidiabéticos*, a partir de la combinación de ANTIDIABETICO1 y ANTIDIABETICO2. La variable Antidiabéticos clasifica a los pacientes de acuerdo a si consume o no medicamentos antidiabéticos, toma el valor 1 si consume uno o más antidiabéticos y 0 si no consume antidiabéticos.

Tabla 2.1: Descripción de variables para el estudio

| <b>Variable</b>        | <b>Descripción</b>  |
|------------------------|---|
| Peso                   | Peso del paciente.  |
| Talla                  | Talla del paciente.   |
| IMC                    | Índice de masa corporal.  |
| TAS                    | Presión arterial sistólica.   |
| TAD                    | Presión arterial diastólica.  |
| Glucosa                | Nivel de glucosa del paciente.  |
| TPDIETA                | Variable dicotómica, 1 si el paciente sigue una dieta, 0 si no.   |
| TPMETFO                | Paciente en tratamiento con Metformina, 1 si, 0 si no.  |
| TPSULFO                | Paciente en tratamiento con Sulfonilureas, 1 si, 0 si no.   |
| TPINSUL                | Paciente en tratamiento con Insulina, 1 si, 0 si no.  |
| Úlcera en pies         | Variable dicotómica, 1 si presenta úlceras, 0 si no.  |
| Examen clínico en pies | Variable dicotómica, 1 si ha presentado necesidad de examen clínico, 0 si no.   |
| Antidiabéticos         | Tratamiento farmacológico antidiabético, 1 si consume uno o dos antidiabéticos, 0 si no consume.  |
| NEFROPATÍA             | Variable dicotómica, 1 si el paciente ha sufrido de insuficiencia renal o ha notado algún malestar en algún órgano del sistema renal, 0 si no.                          |
| NEUROPATÍA             | Variable dicotómica, 1 si el paciente presenta leves molestias como hormigueo hasta condiciones severas como pérdida de sensibilidad o amputaciones asociadas, 0 si no. |
| RETINOPATÍA            | Variable dicotómica, 1 si el paciente sufre visión borrosa, daño en la retina o ceguera, 0 si no.   |

---

Fuente: Elaboración propia.

En la tabla 2.2 se presenta la estadística descriptiva de las variables consideradas para el presente estudio.

Tabla 2.2: Descripción de variables para el estudio

|         | Peso   | Talla | IMC   | TAS    | TAD    | Glucosa | TPDIETA |
|---------|--------|-------|-------|--------|--------|---------|---------|
| Mín     | 23.50  | 1.23  | 9.10  | 70.00  | 20.00  | 0.75    | 0.00    |
| 1er Q   | 55.00  | 1.45  | 24.90 | 110.00 | 70.00  | 113.00  | 1.00    |
| Mediana | 62.80  | 1.50  | 27.50 | 120.00 | 80.00  | 142.00  | 1.00    |
| Media   | 63.91  | 1.51  | 27.92 | 123.39 | 76.85  | 159.27  | 0.97    |
| 3er Q   | 71.40  | 1.56  | 30.50 | 130.00 | 80.00  | 190.00  | 1.00    |
| Máx     | 136.00 | 1.95  | 56.90 | 180.00 | 150.00 | 1890.00 | 1.00    |
| DE      | 12.54  | 0.08  | 4.49  | 18.17  | 9.98   | 68.62   | 0.15    |

|         | TPMETFO | TPSULFO | TPINSUL | Ulcera<br>en pies | Examen<br>clínico<br>en pies | Anti<br>diabéticos |
|---------|---------|---------|---------|-------------------|------------------------------|--------------------|
| Mín     | 0.00    | 0.00    | 0.00    | 0.000             | 0.00                         | 0.00               |
| 1er Q   | 0.00    | 0.00    | 0.00    | 0.000             | 1.00                         | 1.00               |
| Mediana | 0.00    | 0.00    | 0.00    | 0.000             | 1.00                         | 1.00               |
| Media   | 0.41    | 0.39    | 0.11    | 0.0045            | 0.90                         | 0.91               |
| 3er Q   | 1.00    | 1.00    | 0.00    | 0.000             | 1.00                         | 1.00               |
| Máx     | 1.00    | 1.00    | 1.00    | 1.000             | 1.00                         | 1.00               |
| DE      | 0.49    | 0.49    | 0.32    | 0.067             | 0.29                         | 0.27               |

Fuente: Elaboración propia.

El análisis discriminante y los modelos de regresión logística constituyen herramientas fundamentales para resolver problemas de clasificación. Entre los métodos que serán utilizados en el presente trabajo son el Análisis Discriminante Lineal, el Análisis Discriminante Cuadrático y el modelo Logit.

Para llevar a cabo los análisis estadísticos en este estudio, se utilizó el software **R**, versión 4.1.0 (R Core Team, 2021). Se emplearon diversas librerías y paquetes especializados para realizar los procedimientos estadísticos y el proceso de datos. En particular, se utilizó la librería **MASS** [39] para ajustar modelos de regresión logística (GLM, por sus siglas en inglés Generalized Linear Models), así como para realizar análisis discriminante lineal (Linear Discriminant Analysis, LDA) y análisis discriminante cuadrático (Quadratic Discriminant Analysis, QDA).

Adicionalmente, se utilizaron otras librerías reconocidas en el ámbito estadístico y de ciencia de datos. La librería **polycor** está especializada en el cálculo de correlaciones, útiles cuando se trabaja con variables mixtas (continuas o categóricas). La función `hetcor`, perteneciente a esta librería, calcula una matriz de correlaciones que tiene en cuenta la naturaleza de las variables presentes en el conjunto de datos. La librería **ggcorrplot** facilita la visualización de matrices de correlación mediante gráficos claros y atractivos, permite mostrar solo el

triángulo inferior de la matriz de correlación. Así tenemos un análisis más preciso y robusto, especialmente en contextos donde las variables no cumplen con los supuestos de normalidad.

En cuanto al análisis de correlación, se calculó el coeficiente de correlación de Pearson para evaluar las relaciones lineales entre las variables de interés. Se estableció un nivel de significancia estadística de  $\alpha = 0.05$  para determinar la relevancia de las correlaciones observadas. Todos los análisis se realizaron siguiendo prácticas estadísticas.

Más adelante se presentan las tablas con los coeficientes de la función discriminante lineal, junto con la tabla que muestra la calidad de la decisión estadística para clasificar en el Grupo 0 (**G0**: Paciente sin complicaciones) o en el Grupo 1 (**G1**: Paciente con alguna de las complicaciones, es decir, retinopatía, nefropatía o neuropatía, según sea el caso). Esta tabla tiene la siguiente forma:

Tabla 2.3: Decisión estadística

|           | <b>D0</b>                  | <b>D1</b>                  | <b>Suma</b> |
|-----------|----------------------------|----------------------------|-------------|
| <b>G0</b> | Decisión correcta $n_{11}$ | error $n_{12}$             | $n_1$       |
| <b>G1</b> | error $n_{21}$             | Decisión correcta $n_{22}$ | $n_2$       |

En este caso, se analiza una función discriminante lineal que clasifica observaciones en dos grupos:  $G_0$  y  $G_1$ .

Donde:

- **G0**: Observaciones que realmente pertenecen al grupo 0.
- **G1**: Observaciones que realmente pertenecen al grupo 1.
- **D0**: Observaciones clasificadas como grupo 0 por el modelo.
- **D1**: Observaciones clasificadas como grupo 1 por el modelo.

Tenemos:

- $n_{11}$ : Número de observaciones correctamente clasificadas en el grupo 0 (decisión correcta para  $G_0$ ).
- $n_{12}$ : Número de observaciones incorrectamente clasificadas en el grupo 1 (error para  $G_0$ ).

- $n_{21}$ : Número de observaciones incorrectamente clasificadas en el grupo 0 (error para  $G_1$ ).
- $n_{22}$ : Número de observaciones correctamente clasificadas en el grupo 1 (decisión correcta para  $G_1$ ).

- $n_1$ : Total de observaciones reales en el grupo 0, calculado como:

$$n_1 = n_{11} + n_{12}$$

- $n_2$ : Total de observaciones reales en el grupo 1, calculado como:

$$n_2 = n_{21} + n_{22}$$

## 2.1. Diagramas de correlación

Se realizó un análisis exploratorio de los datos, a fin de detectar variables con una correlación alta. En las siguientes figuras, se muestran los diagramas de correlación que describe las relaciones lineales entre el conjunto de variables, expresadas mediante coeficientes de correlación de Pearson. Estos coeficientes, que oscilan entre -1 y 1, permiten identificar la dirección y la fuerza de las asociaciones entre pares de variables. Un valor cercano a 1 indica una correlación positiva fuerte, un valor cercano a -1 sugiere una correlación negativa fuerte, y un valor próximo a 0 denota la ausencia de correlación lineal. Es importante recordar que una correlación positiva indica que, al aumentar una variable, la otra también tiende a incrementarse. En cambio, cuando la correlación es negativa, el aumento de una variable se asocia con la disminución de la otra.

En los siguientes diagramas de correlación para cada una de las complicaciones (Retinopatía, Neuropatía y Nefropatía), se observa que hay una correlación positiva fuerte entre Peso e IMC, lo cual es esperable dado que el Índice de Masa Corporal (IMC) se calcula directamente a partir del peso y la talla. Al igual que, la Presión arterial diastólica (TAD) tiene una correlación positiva con la Presión Arterial Sistólica (TAS), lo cual resulta normal dada la naturaleza de las variables.

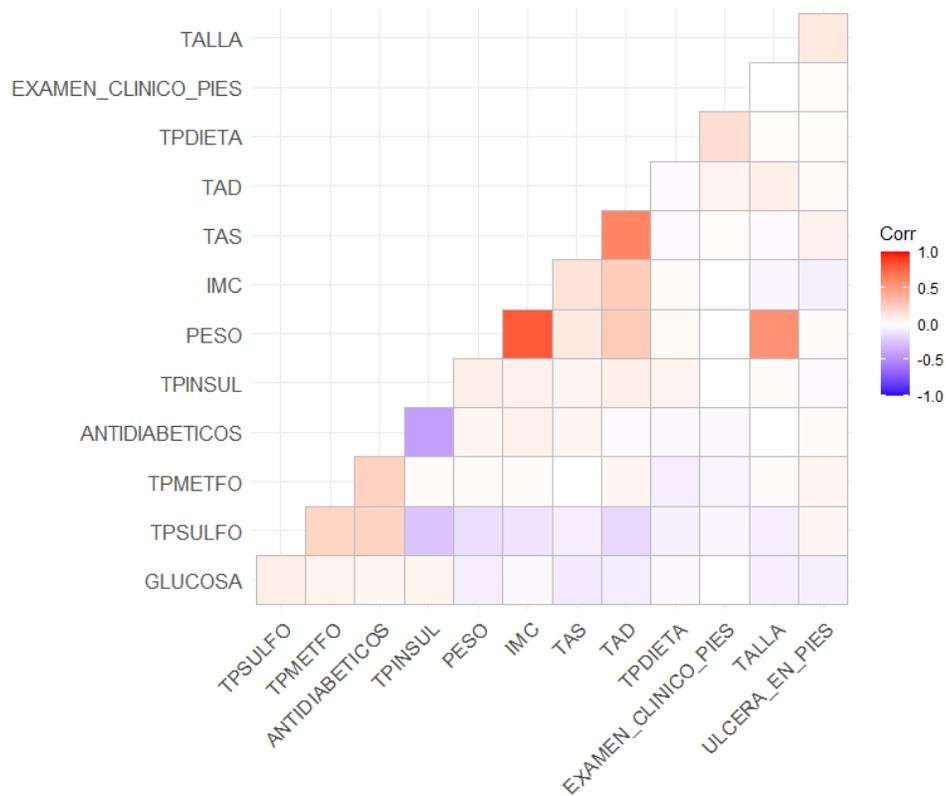


Figura 2.1: Diagrama de correlación entre las variables en pacientes con retinopatía.

Fuente: Elaboración propia.

De la Figura 2.1 se observa que en los pacientes con retinopatía, la Presión arterial diastólica (TAD) tiene una correlación positiva moderada con el IMC al igual que con el Peso.

Por otro lado, Paciente en tratamiento con insulina (TPINSUL), tiene una correlación moderada negativa con los pacientes que consumen Antidiabéticos, y más leve con Paciente en tratamiento con Sulfonilureas (TPSULFO). También podemos notar una correlación positiva leve entre Paciente en tratamiento con Metformina (TPMETFO) y Paciente en tratamiento con Sulfonilureas (TPSULFO).

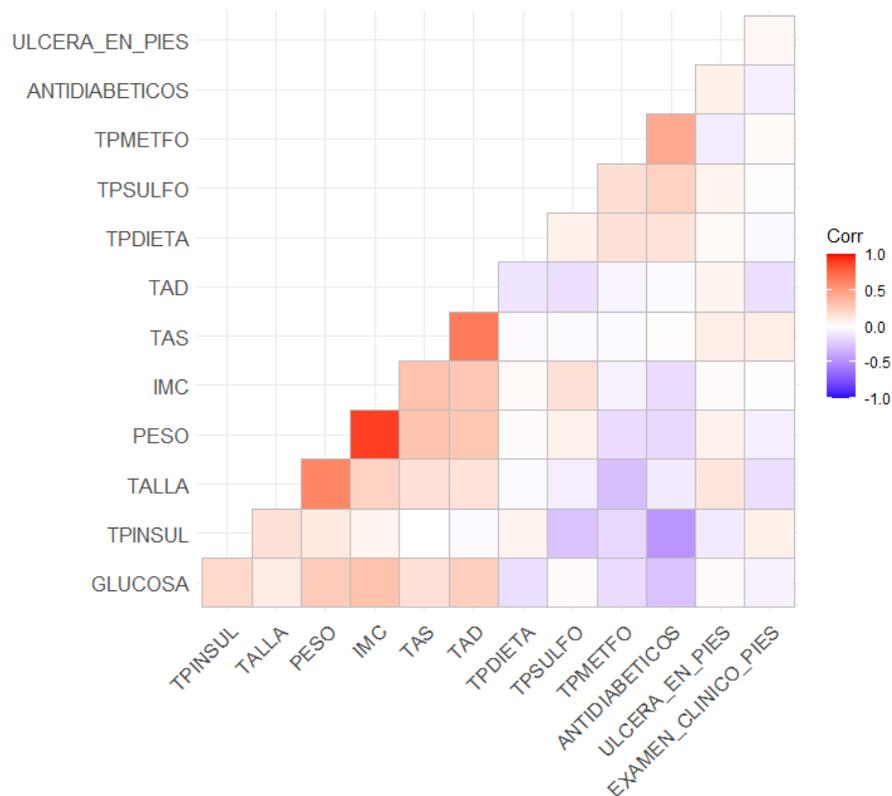


Figura 2.2: Diagrama de correlación entre las variables en pacientes con nefropatía.  
Fuente: Elaboración propia.

De igual manera que en el diagrama anterior, en la Figura 2.2 para los pacientes con nefropatía el Peso y el índice de masa corporal, presentan un relación leve positiva con la Presión Arterial Sistólica (TAS) y también con la Presión arterial diastólica (TAD). Por otra parte, Paciente en tratamiento con insulina (TPINSUL) tiene correlación negativa con con pacientes que consume Antidiabéticos y Paciente en tratamiento con Sulfonilureas (TPSULFO). La Glucosa presenta correlaciones moderadas positivas con Peso e IMC y negativa moderada con Antidiabéticos.

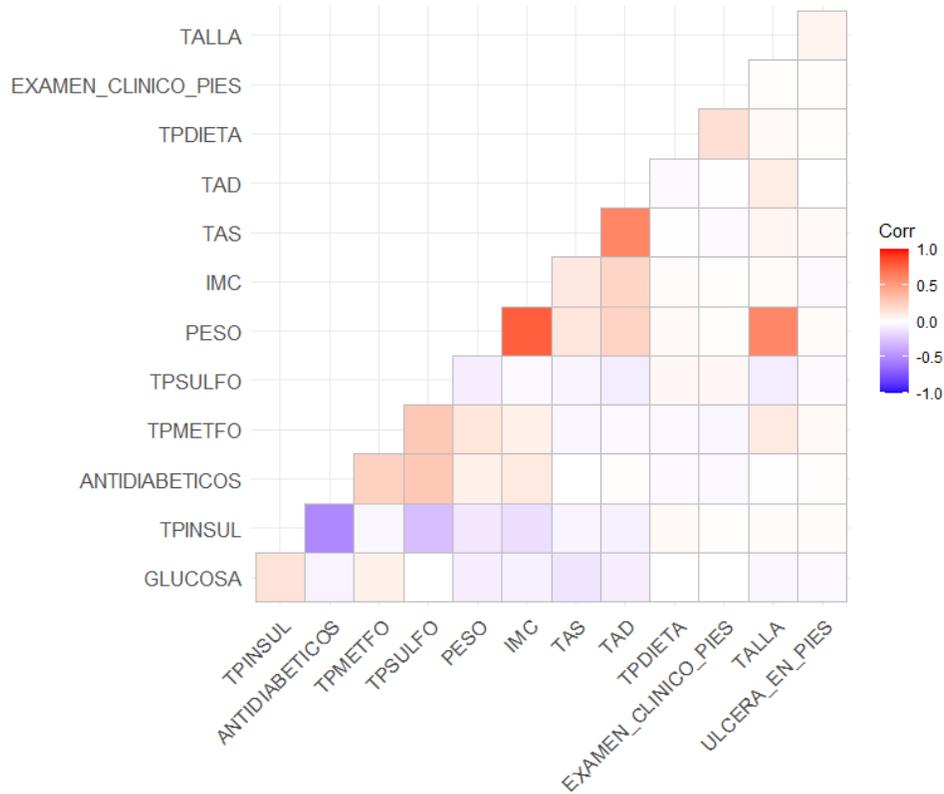


Figura 2.3: Diagrama de correlación entre las variables en pacientes con neuropatía.  
Fuente: Elaboración propia.

En la Figura 2.3, tenemos que para los pacientes que presentan neuropatía, Paciente en tratamiento con insulina (TPINSUL) tiene correlaciones moderadas negativas con pacientes que consumen Antidiabéticos.

Además, Paciente en tratamiento con Sulfonilureas (TPSULFO) presenta una correlación leve positiva con pacientes que consumen Antidiabéticos y Paciente en tratamiento con Metformina (TPMETFO). Otra correlación leve positiva que se observa, es entre la Presión arterial diastólica (TAD) con el Peso e IMC.

## Resultados

A continuación, se presentan los resultados del Análisis Discriminante y el Modelo Logit aplicado a los datos recopilados aproximadamente entre los años 2003 y 2010. Los datos corresponden a pacientes con diabetes que han desarrollado complicaciones adicionales como retinopatía, nefropatía y neuropatía. Para este análisis, se han considerado a los pacientes que presentan sólo una de estas complicaciones. La Tabla 2.1 describe las variables que se utilizan para el análisis que se realizó para presentar en esta tesis.

### 3.1. Modelo Discriminante Lineal

Los resultados del Análisis Discriminante Lineal, se presentan a continuación, con el objetivo de identificar las variables que mejor discriminan entre los grupos definidos en el estudio, así como de evaluar la capacidad de clasificación del modelo.

En la Tabla 3.1, se presentan los coeficientes de los discriminantes lineales utilizados para la clasificación de la retinopatía; se pueden destacar los siguientes resultados.

En primer lugar, los coeficientes de mayor magnitud indican una mayor capacidad de las variables para discriminar entre los grupos analizados. No obstante, también es relevante considerar el signo del coeficiente, ya que un valor positivo implica que ambas variables tienden a aumentar conjuntamente, mientras que un valor negativo sugiere que, a medida que una de las variables incrementa, la otra disminuye. En este sentido, la variable *Úlcera en pies* es la que hace una aportación mayor a la clasificación de la presencia de retinopatía. Le siguen *Talla*, *Paciente en tratamiento con insulina* y *Examen clínico en pies*, cuyos coeficientes también son elevados, lo que sugiere una asociación directa y significativa con la presencia de retinopatía. De manera similar, *Paciente en tratamiento con Sulfonilureas* y *Antidiabéticos* presentan una contribución positiva moderada en la discriminación entre clases. Asimismo, *IMC*, *presión Arterial Sistólica* y *Paciente en tratamiento con Metformina* muestran una influencia positiva, aunque menos pronunciada en comparación con las variables previamente mencionadas.

Por otro lado, las variables con coeficientes cercanos a cero reflejan una baja capacidad para discriminar entre la presencia o ausencia de retinopatía. En particular, *Glucosa* y *Presión Arterial Diastólica* muestran coeficientes próximos a cero, lo que sugiere que su impacto en la clasificación es mínimo.

Tabla 3.1: Parámetros del análisis discriminante lineal para clasificar pacientes con retinopatía

| Variable               | Coefficientes de los discriminantes lineales |
|------------------------|--|
| Peso                   | -0.0572                                      |
| Talla                  | 4.2635                                       |
| IMC                    | 0.0387                                       |
| TAS                    | 0.0206                                       |
| TAD                    | -0.0035                                      |
| Glucosa                | 0.0008                                       |
| TPDIETA                | -0.2923                                      |
| TPMETFO                | 0.0542                                       |
| TPSULFO                | 0.5081                                       |
| TPINSUL                | 1.8949                                       |
| Úlcera en pies         | 5.0985                                       |
| Examen clínico en pies | 1.6695                                       |
| Antidiabéticos         | 0.4085                                       |

Fuente: Elaboración propia.

Finalmente, algunos coeficientes negativos sugieren una asociación inversa con la presencia de retinopatía. En particular, *TPDIETA* y *Peso* presentan coeficientes negativos de pequeña magnitud, lo que indica que valores elevados de estas variables podrían estar asociados con una menor probabilidad de desarrollar retinopatía. Sin embargo, la baja magnitud de estos coeficientes señala que esta relación no es particularmente fuerte.

Tabla 3.2: Calidad de la decisión para el caso de retinopatía

|             | <b>D0</b> | <b>D1</b> | <b>Suma</b> |
|-------------|-----------|-----------|-------------|
| <b>G0</b>   | 25664     | 11        | 25675       |
| <b>G1</b>   | 924       | 6         | 930         |
| <b>Suma</b> | 26588     | 17        | 26605       |

Fuente: Elaboración propia.

El estimador de la probabilidad de clasificación errónea se define según la expresión (1.5). A partir de ello, se obtienen los siguientes valores:

$$\hat{p}_{0,1} = \frac{11}{930} \approx 0,012$$

Este valor indica que aproximadamente el 1.2 % de las observaciones del grupo 1 (con retinopatía) han sido clasificadas incorrectamente en el grupo 0 (sin retinopatía).

$$\hat{p}_{1,0} = \frac{924}{25675} \approx 0,036$$

Este resultado sugiere que alrededor del 3.6 % de las observaciones del grupo 0 han sido clasificadas erróneamente en el grupo 1.

Errores:

- 11 personas de G0 fueron clasificadas como D1 → falso positivo.
- 924 personas de G1 fueron clasificadas como D0 → falso negativo.

Correctas:

- 25664 personas fueron correctamente clasificadas como G0 → D0, es decir 99.96 %

$$\frac{25664}{25675} \approx 0.9996$$

- 6 personas fueron correctamente clasificadas como G1 → D1, es decir 99.35 %

$$\frac{6}{930} \approx 0.9935$$

Tabla 3.3: Parámetros del análisis discriminante lineal para clasificar pacientes con nefropatía

| Variable               | Coefficientes de los discriminantes lineales |
|------------------------|--|
| Peso                   | 0.2244                                       |
| Talla                  | - 18.9022                                    |
| IMC                    | - 0.5847                                     |
| TAS                    | 0.0108                                       |
| TAD                    | 0.0060                                       |
| Glucosa                | - 0.0049                                     |
| TPDIETA                | - 0.0492                                     |
| TPMETFO                | 0.8318                                       |
| TPSULFO                | - 0.4643                                     |
| TPINSUL                | 0.0802                                       |
| Úlcera en pies         | 15.3358                                      |
| Examen clínico en pies | 0.5029                                       |
| Antidiabéticos         | - 0.4411                                     |

Fuente: Elaboración propia.

A partir de la Tabla 3.3, que presenta los coeficientes de los discriminantes lineales para un conjunto de variables clínicas y farmacológicas en un modelo de análisis discriminante lineal orientado a clasificar la presencia de nefropatía, se pueden destacar los siguientes resultados.

En primer lugar, las variables con los coeficientes más altos tienen un mayor peso en la clasificación de los casos. La variable *Úlcera en pies* presenta el coeficiente positivo más elevado, lo que sugiere que es el factor más fuertemente asociado con la presencia de nefropatía en el modelo. Asimismo, destacan el *Examen clínico en pies*, cuyo coeficiente positivo refuerza la asociación entre complicaciones en los pies y la nefropatía, así como *Paciente en tratamiento con Metformina* y el *Peso*.

Por otro lado, las variables con coeficientes cercanos a cero reflejan una baja capacidad discriminativa entre la presencia o ausencia de nefropatía. Por ejemplo, *Paciente en tratamiento*

con *Insulina*, *Presión Arterial Sistólica* y *Presión Arterial Diastólica* presentan coeficientes positivos próximos a cero, lo que indica que su influencia en la clasificación es prácticamente nula.

Finalmente, algunas variables tienen coeficientes negativos, lo que indica una asociación inversa con la nefropatía. La *Talla* muestra un coeficiente negativo con mayor magnitud relativa, lo que sugiere una posible relación inversa relevante. Aunque este hallazgo requiere un análisis más profundo, existen estudios que respaldan esta observación: pacientes en el cuartil más bajo de estatura adulta presentaron mayor riesgo de nefropatía diabética prevalente en el caso de Diabetes Mellitus tipo 1 [40], y otros reportes indican que una baja estatura adulta se asocia con un aumento significativo en la incidencia de enfermedad renal crónica terminal en pacientes con DM2 [16].

El *IMC* presenta un coeficiente negativo moderado, lo que sugiere que un mayor índice de masa corporal podría estar relacionado con una menor probabilidad de nefropatía. Se observa un patrón similar en las variables *TPDIETA* y *Paciente en tratamiento con Sulfonilureas*. Lo mismo sucede con *Glucosa*, lo cual resulta ser contradictorio por la evidencia que se tiene de su importancia en la aparición de complicaciones.

Tabla 3.4: Calidad de la decisión para el caso de nefropatía

|             | <b>D0</b> | <b>D1</b> | <b>Suma</b> |
|-------------|-----------|-----------|-------------|
| <b>G0</b>   | 25624     | 51        | 25675       |
| <b>G1</b>   | 106       | 5         | 111         |
| <b>Suma</b> | 25730     | 56        | 25786       |

Fuente: Elaboración propia.

El estimador de la probabilidad de clasificación errónea se define según la expresión (1.5). A partir de ello, se obtienen los siguientes valores:

$$\hat{p}_{0,1} = \frac{51}{111} \approx 0,459$$

Este valor indica que aproximadamente el 45.9% de las observaciones del grupo 1 (con nefropatía) han sido clasificadas incorrectamente en el grupo 0 (sin nefropatía).

$$\hat{p}_{1,0} = \frac{106}{25675} \approx 0,0041$$

Este resultado sugiere que alrededor del 0.41% de las observaciones del grupo 0 han sido clasificadas erróneamente en el grupo 1.

Errores:

- 51 personas de G0 fueron clasificadas como D1 → falso positivo.
- 106 personas de G1 fueron clasificadas como D0 → falso negativo.

Correctas:

- 25624 personas fueron correctamente clasificadas como G0 → D0, es decir 99.59 %

$$\frac{25624}{25675} \approx 0.9959$$

- 5 personas fueron correctamente clasificadas como G1 → D1, es decir 4.50 %

$$\frac{5}{111} \approx 0.0450$$

La Tabla 3.5 presenta los coeficientes de los discriminantes lineales utilizados para clasificar la presencia de neuropatía. A continuación, se detalla el análisis de los resultados obtenidos.

Las variables con los coeficientes más elevados son aquellas que tienen un mayor impacto en la clasificación de los casos con neuropatía. La variable *Úlcera en pies* presenta el coeficiente más alto, lo que indica que es el factor más fuertemente asociado con la presencia de neuropatía en el modelo.

La variable *Paciente en tratamiento con Sulfonilureas* presenta un coeficiente positivo elevado, lo que sugiere que los pacientes que utilizan este tipo de medicamento tienen una mayor probabilidad de desarrollar neuropatía. De manera similar, el uso de insulina (*TPINSUL*) muestra un coeficiente positivo considerable, lo que también sugiere una mayor probabilidad de neuropatía en pacientes tratados con insulina. Sin embargo, es importante considerar que existe la neuropatía inducida por el tratamiento de la diabetes, la cual puede surgir cuando se logra un control glucémico rápido, especialmente en pacientes con antecedentes de un control glucémico deficiente [8]. La variable *Examen clínico en pies* también presenta un coeficiente positivo notable.

Por otro lado, las variables con coeficientes cercanos a cero tienen un impacto mínimo en la clasificación de la neuropatía. El nivel de *Glucosa* muestra un coeficiente muy bajo, seguido de la *Presión Arterial Sistólica*, que presenta un coeficiente negativo próximo a cero.

Algunas variables presentan coeficientes negativos, lo que implica una relación inversa con la neuropatía. La variable *Talla* indica que una mayor estatura podría estar asociada con una menor probabilidad de neuropatía. El *IMC* muestra un coeficiente negativo moderado, lo que sugiere que un mayor índice de masa corporal podría estar relacionado con una menor probabilidad de neuropatía. De manera similar, las variables *Paciente en tratamiento con Metformina* y *Antidiabéticos* presentan coeficientes negativos.

El estimador de la probabilidad de clasificación errónea se define según la expresión (1.5). A partir de ello, se obtienen los siguientes valores:

Tabla 3.5: Resultados del análisis discriminante para clasificar pacientes con neuropatía

| Variable               | Coefficientes de los discriminantes lineales |
|------------------------|--|
| Peso                   | 0.0520                                       |
| Talla                  | - 7.6994                                     |
| IMC                    | - 0.1697                                     |
| TAS                    | - 0.0009                                     |
| TAD                    | 0.0301                                       |
| Glucosa                | 0.0037                                       |
| TPDIETA                | 0.3019                                       |
| TPMETFO                | -0.4540                                      |
| TPSULFO                | 1.1358                                       |
| TPINSUL                | 0.7055                                       |
| Úlcera en pies         | 8.7128                                       |
| Examen clínico en pies | 1.5450                                       |
| Antidiabéticos         | - 0.2809                                     |

Fuente: Elaboración propia.

$$\hat{p}_{0,1} = \frac{54}{2865} \approx 0,0188$$

Este valor indica que aproximadamente el 18.8% de las observaciones del grupo 1 (con neuropatía) han sido clasificadas incorrectamente en el grupo 0 (sin neuropatía).

$$\hat{p}_{1,0} = \frac{2809}{25675} \approx 0,1094$$

Este resultado sugiere que alrededor del 10.94% de las observaciones del grupo 0 han sido clasificadas erróneamente en el grupo 1.

Tabla 3.6: Calidad de la decisión para el caso de neuropatía

|             | <b>D0</b> | <b>D1</b> | <b>Suma</b> |
|-------------|-----------|-----------|-------------|
| <b>G0</b>   | 25621     | 54        | 25675       |
| <b>G1</b>   | 2809      | 56        | 2865        |
| <b>Suma</b> | 28430     | 110       | 28540       |

Fuente: Elaboración propia.

Errores:

- 54 personas de G0 fueron clasificadas como D1 → falso positivo.
- 2809 personas de G1 fueron clasificadas como D0 → falso negativo.

Correctas:

- 25621 personas fueron correctamente clasificadas como G0 → D0, es decir 99.80 %

$$\frac{25621}{25675} \approx 0.9980$$

- 56 personas fueron correctamente clasificadas como G1 → D1, es decir 1.95 %

$$\frac{56}{2865} \approx 0.0195$$

### 3.2. Modelo Discriminante Cuadrático

Aquí se presentan los resultados del Análisis Discriminante Cuadrático. Los coeficientes de los discriminantes cuadráticos obtenidos en el análisis reflejan la contribución relativa de cada variable en la función discriminante, permitiendo discernir su influencia en la clasificación de los individuos en los grupos de interés.

En este análisis, la variable Glucosa presentó un coeficiente negativo, lo cual resulta contradictorio con la evidencia clínica y epidemiológica. Diversos estudios han establecido una asociación directa entre niveles elevados de glucosa y el desarrollo de complicaciones.

Tabla 3.7: Resultados del análisis discriminante cuadrático, para clasificar pacientes con retinopatía

| Variable               | Coefficientes de los discriminantes cuadráticos |
|------------------------|---|
| Peso                   | -0.082  |
| Talla                  | -13.603   |
| IMC                    | 2.210   |
| TAS                    | 0.050   |
| TAD                    | -0.114  |
| Glucosa                | -0.012  |
| TPDIETA                | -8.397  |
| TPMETFO                | 2.017   |
| TPSULFO                | 2.129   |
| TPINSUL                | 2.616   |
| Úlcera en pies         | 10.458  |
| Examen clínico en pies | - 10.519  |
| Antidiabéticos         | - 4.730   |

Fuente: Elaboración propia.

De la Tabla 3.7 se tiene que, las variables con mayor influencia positiva en la probabilidad de retinopatía incluyen a *Paciente en tratamiento con Sulfonilureas, Insulina y Metformina*. Estos resultados sugieren que los pacientes que requieren estos medicamentos tienen un mayor riesgo de complicaciones como la retinopatía. La presencia de *Úlcera en pies* también muestra una fuerte asociación positiva con la retinopatía. El *IMC* también presenta una influencia positiva considerable y la *Presión Arterial Sistólica* está asociada con una mayor probabilidad de retinopatía, aunque con una influencia menos marcada.

En contraste, varias variables muestran una influencia negativa sobre la probabilidad de retinopatía, la *Talla* tiene el coeficiente negativo más alto, le sigue *Examen clínico en pies* y *TPDIETA*. La variable *Peso* también muestra una asociación negativa, aunque menos pronunciada, al igual que *Presión Arterial Diastólica*.

El coeficiente de *Glucosa* es negativo, lo que indica que niveles más bajos de glucosa están asociados con una mayor probabilidad de retinopatía, esto puede deberse a la asociación que hay entre la variabilidad de glucémica y el hecho de que este es un factor de riesgo para la retinopatía [12].

Tabla 3.8: Calidad de la decisión para el caso de retinopatía

|      | D0    | D1  | Suma  |
|------|-------|-----|-------|
| G0   | 24971 | 704 | 25675 |
| G1   | 835   | 95  | 930   |
| Suma | 25806 | 799 | 26605 |

Fuente: Elaboración propia.

El estimador de la probabilidad de clasificación errónea se define según la expresión (1.5). A partir de ello, se obtienen los siguientes valores:

$$\hat{p}_{0,1} = \frac{704}{930} \approx 0.756$$

Este valor indica que aproximadamente el 75.6% de las observaciones del grupo 1 (con retinopatía) han sido clasificadas incorrectamente en el grupo 0 (sin retinopatía).

$$\hat{p}_{1,0} = \frac{835}{25675} \approx 0.0325$$

Este resultado sugiere que alrededor del 3.25% de las observaciones del grupo 0 han sido clasificadas erróneamente en el grupo 1.

Errores:

- 704 personas de G0 fueron clasificadas como D1 → falso positivo.

- 835 personas de G1 fueron clasificadas como D0 → falso negativo.

Correctas:

- 24971 personas fueron correctamente clasificadas como G0 → D0, es decir 97.26 %

$$\frac{24971}{25675} \approx 0.9726$$

- 95 personas fueron correctamente clasificadas como G1 → D1, es decir 10.22 %

$$\frac{95}{930} \approx 0.1022$$

Tabla 3.9: Resultados del análisis discriminante cuadrático, para clasificar pacientes con nefropatía

| Variable               | Coefficientes de los discriminantes cuadráticos |
|------------------------|---|
| Peso                   | -0.057  |
| Talla                  | -13.930   |
| IMC                    | -1.469  |
| TAS                    | 0.041   |
| TAD                    | 0.116   |
| Glucosa                | -0.020  |
| TPDIETA                | -7.360  |
| TPMETFO                | -2.204  |
| TPSULFO                | -2.494  |
| TPINSUL                | 3.172   |
| Úlcera en pies         | 4.726   |
| Examen clínico en pies | -5.497  |
| Antidiabéticos         | 4.282   |

Fuente: Elaboración propia.

De la Tabla 3.9 se tiene que, el *Peso* presenta un coeficiente negativo, lo que sugiere que un mayor peso está asociado con una menor probabilidad de desarrollar nefropatía. De manera similar, el coeficiente correspondiente a la *Talla*, indica que una mayor estatura está fuertemente asociada con una menor probabilidad de nefropatía. El *índice de masa corporal* (IMC) también muestra un coeficiente negativo, lo que implica que un mayor IMC se relaciona con una menor probabilidad de nefropatía, igual que *Paciente en tratamiento con Sulfonilureas y Metformina*, *Glucosa*, *TPDIETA* y *Exmen clínico en pies*.

En contraste, *Paciente en tratamiento con Insulina*, la *presión arterial diastólica y sistólica*, *Úlcera en pies* presentan coeficientes positivos. Esto indica que un aumento de estas, está

asociado con una mayor probabilidad de nefropatía.

Tabla 3.10: Calidad de la decisión para el caso de nefropatía

|      | D0    | D1  | Suma  |
|------|-------|-----|-------|
| G0   | 25553 | 122 | 25675 |
| G1   | 87    | 24  | 111   |
| Suma | 25640 | 146 | 25786 |

Fuente: Elaboración propia.

El estimador de la probabilidad de clasificación errónea se define según la expresión (1.5). A partir de ello, se obtienen los siguientes valores:

$$\hat{p}_{0,1} = \frac{122}{111} \approx 1.099$$

Este valor indica que aproximadamente el 109.9% de las observaciones del grupo 1 (con nefropatía) han sido clasificadas incorrectamente en el grupo 0 (sin nefropatía).

$$\hat{p}_{1,0} = \frac{87}{25675} \approx 0.0033$$

Este resultado sugiere que alrededor del 0.33% de las observaciones del grupo 0 han sido clasificadas erróneamente en el grupo 1.

Errores:

- 122 personas de G0 fueron clasificadas como D1 → falso positivo.
- 87 personas de G1 fueron clasificadas como D0 → falso negativo.

Correctas:

- 25553 personas fueron correctamente clasificadas como G0 → D0, es decir 99.52%

$$\frac{25553}{25675} \approx 0.9952$$

- 24 personas fueron correctamente clasificadas como G1 → D1, es decir 21.62%

$$\frac{24}{111} \approx 0.2162$$

Tabla 3.11: Resultados del análisis discriminante cuadrático para clasificar pacientes con neuropatía

| Variable               | Coefficientes de los discriminantes cuadráticos |
|------------------------|---|
| Peso                   | - 0.085   |
| Talla                  | 14.566  |
| IMC                    | - 2.190   |
| TAS                    | - 0.053   |
| TAD                    | - 0.123   |
| Glucosa                | - 0.014   |
| TPDIETA                | - 9.180   |
| TPMETFO                | 2.077   |
| TPSULFO                | 2.116   |
| TPINSUL                | 3.215   |
| Úlcera en pies         | 7.332   |
| Examen clínico en pies | - 5.822   |
| Antidiabéticos         | - 4.603   |

Fuente: Elaboración propia.

Las variables con coeficientes positivos, como *Talla*, *Paciente en tratamiento con sulfonilureas*, *metformina e Insulina*, *Úlcera en pies* y *Examen clínico en pies* se identifican como factores de riesgo potenciales para la neuropatía. En particular, la presencia de úlceras en pies destaca por su fuerte asociación con esta condición.

Por otro lado, las variables con coeficientes negativos, como *Peso*, *IMC*, *Presión Arterial Diastólica y sistólica*, *TPDIETA* y *Antidiabéticos* parecen tener un efecto contrario, es decir, en aumento de estas variables, baja la posibilidad de presentar esta complicación.

Tabla 3.12: Calidad de la decisión para el caso de neuropatía

|             | <b>D0</b> | <b>D1</b> | <b>Suma</b> |
|-------------|-----------|-----------|-------------|
| <b>G0</b>   | 25302     | 373       | 25675       |
| <b>G1</b>   | 2661      | 204       | 2865        |
| <b>Suma</b> | 27963     | 577       | 28540       |

Fuente: Elaboración propia.

El estimador de la probabilidad de clasificación errónea se define según la expresión (1.5). A partir de ello, se obtienen los siguientes valores:

$$\hat{p}_{0,1} = \frac{373}{2865} \approx 0.13$$

Este valor indica que aproximadamente el 13 % de las observaciones del grupo 1 (con neuropatía) han sido clasificadas incorrectamente en el grupo 0 (sin neuropatía).

$$\hat{p}_{1,0} = \frac{2661}{25675} \approx 0.103$$

Este resultado sugiere que alrededor del 10.3 % de las observaciones del grupo 0 han sido clasificadas erróneamente en el grupo 1.

Errores:

- 373 personas de G0 fueron clasificadas como D1 → falso positivo.
- 2661 personas de G1 fueron clasificadas como D0 → falso negativo.

Correctas:

- 25302 personas fueron correctamente clasificadas como G0 → D0, es decir 98.55 %

$$\frac{25302}{25675} \approx 0.9855$$

- 204 personas fueron correctamente clasificadas como G1 → D1, es decir 7.12 %

$$\frac{204}{2865} \approx 0.0712$$

### 3.3. Modelo Discriminante Logit

A continuación, se presentan los resultados del modelo, el cual es utilizado para predecir la probabilidad de que se desarrolle alguna de las complicaciones.

Tabla 3.13: Resultados del Modelo Logit para Retinopatía

| Variable               | Estimación | Error estándar | Valor z | Pr(>  z )    |
|------------------------|------------|----------------|---------|--------------|
| Intercepto             | -13.017    | 4.214          | -3.089  | 0.0020 **    |
| Peso                   | -0.060     | 0.033          | -1.796  | 0.0724 .     |
| Talla                  | 4.772      | 2.779          | 1.717   | 0.0859 .     |
| IMC                    | 0.077      | 0.077          | 0.995   | 0.3195       |
| TAS                    | 0.0094     | 0.0021         | 4.338   | 0.000014 *** |
| TAD                    | 0.0015     | 0.0042         | 0.374   | 0.7082       |
| Glucosa                | 0.0002     | 0.0005         | 0.470   | 0.6382       |
| TPMETFO                | 0.044      | 0.075          | 1.588   | 0.5568       |
| TPSULFO                | 0.3016     | 0.075          | 4.022   | 0.000057 *** |
| TPINSUL                | 0.9014     | 0.096          | 9.304   | < 0 ***      |
| Úlcera en pies         | 1.452      | 0.390          | 3.722   | 0.00019 ***  |
| Examen clínico en pies | 2.376      | 0.319          | 7.441   | 0 ***        |
| Antidiabéticos         | -0.009     | 0.075          | -0.113  | 0.0285 *     |

Nivel de significancia: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.'

Fuente: Elaboración propia.

Entre las variables independientes analizadas, se identificaron varios factores que muestran una asociación significativa con la probabilidad de retinopatía. La *Presión Arterial Sistólica* (TAS) presenta un coeficiente positivo y un valor  $p$  altamente significativo, lo que indica que un aumento en la TAS está asociado con un mayor riesgo de retinopatía: En un caso similar está *Paciente en tratamiento con Insulina y sulfonilureas*. De igual forma, la presencia de *Úlcera en pies* y de la mano *Examen clínico en pies* y el uso de *Antidiabéticos* muestra un

coeficiente negativo significativo, lo cual sugiere una ligera relación protectora.

Por otro lado, algunas variables no mostraron una relación significativa con la probabilidad de desarrollar retinopatía. El *Peso*, *Talla*, *IMC*, *Presión Arterial diastólica* y *Glucosa* no presentaron valores  $p < 0.05$ , lo que sugiere que no tienen un efecto estadísticamente significativo en el riesgo de retinopatía dentro de este modelo. Del mismo modo, el tratamiento con *Paciente en tratamiento con Metformina* tampoco mostró una relación significativa con el desarrollo de retinopatía.

Tabla 3.14: Resultados del Modelo Logit para Neuropatía

| Variable               | Estimación | Error estándar | Valor z | Pr(>  z ) |
|------------------------|------------|----------------|---------|-----------|
| Intercepto             | - 0.7136   | 2.4686         | - 0.289 | 0.773     |
| Peso                   | 0.0055     | 0.0196         | 0.280   | 0.779     |
| Talla                  | - 2.380    | 1.6416         | - 1.450 | 0.147     |
| IMC                    | - 0.0424   | 0.0444         | - 0.955 | 0.340     |
| TAS                    | - 0.0003   | 0.0013         | - 0.266 | 0.791     |
| TAD                    | 0.0158     | 0.0025         | 6.281   | 0 ***     |
| Glucosa                | 0.0018     | 0.0002         | 6.922   | 0 ***     |
| TPMETFO                | - 0.2310   | 0.0450         | - 5.131 | 0 ***     |
| TPSULFO                | 0.6454     | 0.0447         | 14.428  | < 0 ***   |
| TPINSUL                | 0.3747     | 0.0684         | 5.475   | 0 ***     |
| Úlcera en pies         | 2.4246     | 0.2018         | 12.024  | < 0 ***   |
| Examen clínico en pies | 1.4063     | 0.1116         | 12.598  | < 0 ***   |
| Antidiabéticos         | - 0.1300   | 0.0853         | - 1.523 | 0.128     |

Nivel de significancia: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.'

Fuente: Elaboración propia.

La *presión arterial diastólica* (*TAD*) y un valor  $p$  altamente significativo, lo que indica que un aumento en la *TAD* está asociado con un incremento en el riesgo de neuropatía. De manera similar, los niveles de *Glucosa* muestran un coeficiente positivo y un valor  $p$  altamente

significativo, sugiriendo que un aumento en la glucosa sanguínea se asocia con una mayor probabilidad de desarrollar neuropatía.

En cuanto a los tratamientos farmacológicos, el uso de metformina (*TPMETFO*) muestra un coeficiente negativo y una significancia estadística alta, lo que sugiere que el tratamiento con metformina está asociado con un menor riesgo de neuropatía. Por el contrario, el tratamiento con sulfonilureas (*TPSULFO*) presenta un coeficiente positivo y un valor  $p$  extremadamente significativo, lo que indica que el uso de sulfonilureas incrementa de manera significativa el riesgo de neuropatía. De manera similar, el tratamiento con insulina (*TPINSUL*) presenta un coeficiente positivo y una alta significancia estadística, sugiriendo que el uso de insulina también está relacionado con un mayor riesgo de neuropatía.

En el caso de *úlceras en pies* tiene un coeficiente positivo y un valor  $p$  extremadamente significativo, lo que indica una fuerte relación entre la presencia de úlceras y un mayor riesgo de neuropatía. De igual manera, el *Examen clínico en pies*.

Por otro lado, algunas variables no mostraron una relación estadísticamente significativa con la probabilidad de neuropatía en este modelo. El *peso*, la *talla*, el índice de masa corporal (*IMC*), la tensión arterial sistólica (*TAS*) y el uso de dos medicamentos *Antidiabéticos* presentaron valores  $p > 0.05$ , lo que sugiere que no tienen un efecto significativo sobre el riesgo de neuropatía.

Tabla 3.15: Resultados del Modelo Logit para Nefropatía

| Variable               | Estimación | Error estándar | Valor z | Pr(>  z )  |
|------------------------|------------|----------------|---------|------------|
| Intercepto             | 31.2107    | 9.6908         | 3.221   | 0.0012 **  |
| Peso                   | 0.2949     | 0.0755         | 3.903   | 0 ***      |
| Talla                  | - 23.5777  | 6.3779         | - 3.697 | 0.0002 *** |
| IMC                    | - 0.7893   | 0.1781         | - 4.430 | 0 ***      |
| TAS                    | 0.0053     | 0.0019         | 2.714   | 0.0066 **  |
| TAD                    | 0.0224     | 0.0099         | 2.262   | 0.0236 *   |
| Glucosa                | - 0.0079   | 0.0020         | - 3.696 | 0 ***      |
| TPMETFO                | 1.1403     | 0.2263         | 5.037   | 0 ***      |
| TPSULFO                | - 0.6162   | 0.2311         | - 2.666 | 0.0076 **  |
| TPINSUL                | -0.0560    | 0.2943         | - 0.191 | 0.8488     |
| Úlcera en pies         | 3.0699     | 0.499          | 6.146   | 0 ***      |
| Examen clínico en pies | 1.2204     | 0.5148         | 2.370   | 0.0177 *   |
| Antidiabéticos         | - 0.8682   | 0.3377         | - 2.571 | 0.0101 *   |

Nivel de significancia: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.'

Fuente: Elaboración propia.

La presión arterial sistólica (*TAS*) y la presión arterial diastólica (*TAD*) muestran una asociación positiva con el riesgo de nefropatía. Además, el modelo indica que un aumento en la *TAS* incrementa el riesgo de nefropatía.

En cuanto a los niveles de *Glucosa*, se observa una relación negativa con el riesgo de nefropatía, con un valor *p* altamente significativo, lo que indica que un aumento en los niveles de glucosa está asociado con una reducción en el riesgo de nefropatía, lo cual como se ha mencionado anteriormente resulta ser una contradicción en el estudio y de igual forma, *Antidiabéticos*.

Respecto a los tratamientos farmacológicos, el uso de metformina (*TPMETFO*) muestra un coeficiente positivo y una significancia estadística elevada, lo que sugiere que el tratamiento con metformina está asociado con un mayor riesgo de nefropatía. Por otro lado, el tratamiento con sulfonilureas (*TPSULFO*) presenta un coeficiente negativo y un valor *p* significativo, lo

que implica que el uso de sulfonilureas está asociado con un menor riesgo de nefropatía.

La presencia de *úlceras en pies* tiene un coeficiente positivo y un valor  $p$  altamente significativo, entonces está estrechamente relacionada con un mayor riesgo de nefropatía, al igual que *Examen clínico en pies*.

Por otro lado, el tratamiento con insulina (*TPINSUL*) no mostró una relación estadísticamente significativa con la probabilidad de desarrollar nefropatía en este modelo.

## Conclusiones

Según el IMSS, en México, la diabetes es la segunda causa de mortalidad y de años perdidos por muerte prematura. Asimismo, es una de las enfermedades más costosas debido a las complicaciones que genera, como retinopatía, nefropatía y neuropatía [32]. En este trabajo se hizo un análisis de estas complicaciones, tomando en cuenta sólo algunas variables de la base de datos. Encontramos algunas limitaciones, puesto que en la base de datos que utilizamos, en los primeros años de registros, la medición del nivel de glucosa se realizaba con pruebas de glucemia plasmática en ayunas. Posteriormente, comenzó a estandarizarse el uso de la prueba de hemoglobina glicosilada (HbA1c) como prueba para el diagnóstico y control de la diabetes [33]. Por lo tanto, en este trabajo esa variable no puede analizarse con la suficiente precisión.

En la base de datos se encontró información acerca del tratamiento farmacológico, por lo cual es importante analizar el efecto del apego en el tratamiento en la evolución del estado de salud del paciente con diabetes. En un trabajo publicado en 2018, se presenta el resultado del efecto del tratamiento farmacológico, la dieta y un cambio en el estilo de vida en lo que los autores presentan como remisión de la diabetes [26, 27]. Desafortunadamente, no contamos con suficientes datos de pacientes con mediciones en distintos momentos, para hacer una evaluación de las mismas variables a lo largo del tiempo. Lo que pudimos hacer fue un análisis con el mayor número de datos de pacientes diferentes, con el propósito de realizar un modelo capaz de clasificar las distintas complicaciones, a partir de un número reducido de variables.

Debido al aumento en la esperanza de vida tanto de la población general como de las personas con diabetes, existe una mayor probabilidad de que desarrollen otras enfermedades asociadas y complicaciones crónicas, lo que afecta negativamente su calidad de vida [23]. Los hallazgos obtenidos en este trabajo aportarían información valiosa para la identificación temprana de pacientes en riesgo y para el diseño de estrategias de prevención y tratamiento de las complicaciones en poblaciones con diabetes.

Sin embargo, el parámetro asociado con la glucosa, contradice toda la evidencia médica disponible, que refiere una relación significativa entre el nivel de glucosa, su control y la aparición de complicaciones crónicas de la diabetes [31, 24, 10]. Este es un punto relevante que debe ser considerado para la validez de un modelo que pueda clasificar adecuadamente y hacer pronósticos a partir de un conjunto de variables fisiológicas y clínicas. Será conveniente buscar otros modelos que sí coincidan con lo estudiado sobre el curso natural de la enfermedad

y sus complicaciones, lo cual es ampliamente conocido.

Para el tratamiento de pacientes con diabetes mellitus tipo 2, existen varios objetivos y los principales serían, el control metabólico y la prevención de complicaciones [36]. En el presente trabajo se excluyeron del análisis los casos en los que los pacientes presentan dos o más complicaciones simultáneas. Esta limitación abre una línea relevante para futuras investigaciones, ya que la coexistencia de múltiples complicaciones puede influir significativamente en la clasificación y el comportamiento de los modelos predictivos. Además, contribuiría a una mayor aplicabilidad clínica, especialmente en contextos donde la multipatología es una condición común entre los pacientes.

En esta tesis, se han utilizado tres modelos de clasificación, discriminante lineal, discriminante cuadrático y regresión logística, para clasificar complicaciones específicas asociadas a la diabetes mellitus tipo 2, las cuales son, retinopatía, nefropatía y neuropatía, considerando únicamente pacientes que presentan una sola de estas complicaciones.

Se identificaron variables clínicas y farmacológicas que resultaron ser significativas para la clasificación de las complicaciones. En particular, la presencia de *úlceras en pies* y los hallazgos del *examen clínico en pies* emergieron como factores determinantes en los tres tipos de complicaciones, lo que resalta su valor como marcador clínico de riesgo en la evaluación y seguimiento de pacientes diabéticos. Asimismo, el tratamiento con *insulina* y *sulfonilureas* mostró una asociación positiva con la presencia de complicaciones, lo que podría reflejar tanto la gravedad subyacente de la enfermedad como posibles efectos adversos del tratamiento [9]. Este resultado es contradictorio con la evidencia médica reportada, pues está documentado que el empleo de terapia combinada reduce la presencia de complicaciones crónicas.

Por otra parte, el uso de metformina evidenció un comportamiento más complejo: protector en el caso de la neuropatía, pero con posibles efectos desfavorables en nefropatía, lo cual amerita una investigación más profunda. Se ha documentado que en algunos casos disminuye el riesgo de retinopatía, mientras que en otros casos los aumenta [38].

En términos de rendimiento de los modelos, el modelo discriminante lineal se destacó por su estabilidad y precisión general, especialmente en la clasificación de retinopatía y neuropatía. Por el contrario, el modelo discriminante cuadrático presentó serias limitaciones, particularmente en la clasificación de nefropatía, donde alcanzó tasas de error mayores. Finalmente, la regresión logística permitió no solo una buena capacidad de clasificación, sino también una interpretación clara de la asociación de cada variable, fortaleciendo su utilidad como modelo explicativo y como herramienta clínica de apoyo en la toma de decisiones.

Desde una perspectiva clínica y de salud pública, los hallazgos de este estudio subrayan la relevancia de incorporar variables fácilmente obtenibles en la práctica médica como antecedentes de úlceras o el uso de ciertos medicamentos dentro de esquemas de evaluación sistemática del riesgo. La identificación temprana de pacientes propensos a desarrollar complicaciones permitiría implementar intervenciones preventivas más eficaces y personalizadas, mejorando los desenlaces a largo plazo en la población diabética.

## Bibliografía

- [1] K. AL-RUBEAAAN, A. YOUSSEF, S. SUBHANI, N. AHMAD, A. AL-SHARQAWI, AND ET AL., *Diabetic nephropathy and its risk factors in a society with a type 2 diabetes epidemic: A saudi national diabetes registry-based study*, PLOS One, 9 (2014), p. e88956.
- [2] M. J. ANDERSON AND J. ROBINSON, *Generalized discriminant analysis based on distances*, Australian & New Zealand Journal of Statistics, 45 (2003), pp. 301–318.
- [3] E. R. ARGAIZ, L. MORALES-JUÁREZ, C. RAZO, L. ONG, Q. RAFFERTY, R. RINCÓN-PEDRERO, AND G. GAMBA, *La carga de enfermedad renal crónica en méxico. análisis de datos basado en el estudio global burden of disease 2021*, Gaceta médica de México, 159 (2023), pp. 501–508.
- [4] D. AUNE, T. FENG, S. SCHLESINGER, I. JANSZKY, T. NORAT, AND E. RIBOLI, *Diabetes mellitus, blood glucose and the risk of atrial fibrillation: A systematic review and meta-analysis of cohort studies*, Journal of Diabetes and its Complications, 32 (2018), pp. 501–511.
- [5] V. CARRIERO-LIMA, G. COUTINHO-CAVALIERI, M. CARRIERO-LIMA, N. NAZARIO, AND G. CARRIERO-LIMA, *Risk factors for diabetic retinopathy: a case-control study*, International Journal of Retina and Vitreous, 2 (2016), pp. 1–7.
- [6] C. CHATFIELD, *Introduction to multivariate analysis*, Routledge, 2018.
- [7] I. D. FEDERATION, *Idf diabetes atlas, 10th edition*, 2021. Accessed: 2025-05-25.
- [8] M. FERREIRA, G. CAMOES, J. F. DE FERREIRA GOMES, AND D. M. FERREIRA, *Treatment-induced diabetes neuropathy: reminder of an important clinical lesson*, BMJ Case Reports CP, 14 (2021), p. e241849.
- [9] A. GODAY ARNO, J. FRANCH NADAL, M. MATA CASES, F. ÁLVAREZ GUIASOLA, J. DÍEZ ESPINO, I. FERNÁNDEZ FERNÁNDEZ, D. TÓRTOLA-GRANER, D. ACOSTA-DELGADO, M. AGUILAR-DIOSDADO, J. HERRERA-POMBO, ET AL., *La terapia combinada en la diabetes mellitus tipo 2. criterios y pautas*, Elsevier [Internet], 38 (2001), pp. 270–89.
- [10] E. GOUVERI AND N. PAPANAS, *The emerging role of continuous glucose monitoring in the management of diabetic peripheral neuropathy: a narrative review*, Diabetes Therapy, 13 (2022), pp. 931–952.

- [11] J. GROSS, M. DE AZEVEDI, S. SILVEIRO, L. CANANI, M. CARAMORI, AND T. ZELMANOVITZ, *Diabetic nephropathy: Diagnosis, prevention, and treatment*, Diabetes Care, 28 (2005), pp. 176–188.
- [12] C. GUO, M. DESHPANDE, Y. NIU, I. KACHWALA, M. FLORES-BELLVER, H. MEGARITY, T. NUSE, S. BABAPOOR-FARROKHRAN, M. RAMADA, J. SANCHEZ, ET AL., *Hif-1 $\alpha$  accumulation in response to transient hypoglycemia may worsen diabetic eye disease*, Cell reports, 42 (2023).
- [13] W. K. HÄRDLE AND L. SIMAR, *Applied Multivariate Statistical Analysis*, Springer, 2015.
- [14] N. HERNÁNDEZ NAVA, M. F. MENDOZA CORONEL, M. D. R. ROCHA RODRÍGUEZ, M. B. SILVA CÁZARES, L. F. FABELA SÁNCHEZ, AND R. E. FOSADO QUIROZ, *Estudio de la discapacidad visual por diabetes mellitus tipo 2 en el adulto mayor*, Acta universitaria, 30 (2020).
- [15] R. V. HOGG, J. W. MCKEAN, A. T. CRAIG, ET AL., *Introduction to mathematical statistics*, Pearson Education Inc., 2005.
- [16] Y. A. HONG, K.-D. HAN, J.-S. YUN, E. S. SIL, S.-H. KO, AND S. CHUNG, *Short adult height increases the risk of end-stage renal disease in type 2 diabetes*, Endocrine Connections, 9 (2020), pp. 912–921.
- [17] D. W. HOSMER JR, S. LEMESHOW, AND R. X. STURDIVANT, *Applied logistic regression*, John Wiley & Sons, 2013.
- [18] Y. HOU, Y. CAI, Z. JIA, AND S. SHI, *Risk factors and prevalence of diabetic retinopathy: A protocol for meta-analysis*, Medicine, 99 (2020), p. e22695.
- [19] F. HU AND T. ZHANG, *Study on risk factors of diabetic nephropathy in obese patients with type 2 diabetes mellitus*, International Journal of General Medicine, 13 (2020), pp. 351–360.
- [20] INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA (INEGI), *Estadísticas a propósito del día mundial de la diabetes (2022)*, 2022. Accessed: 2025-05-25.
- [21] INSTITUTO NACIONAL DE SALUD PÚBLICA (INSP), *Encuesta nacional de salud y nutrición 2021: Resultados estatales, hidalgo*, 2021. Accessed: 2025-05-25.
- [22] I. IRIGOIEN, F. MESTRES I NAVAL, AND C. ARENAS SOLÀ, *Weighted distance based discriminant analysis: The r package wedibadis*, The R Journal, 2016, vol. 8, num. 2, p. 434-450, (2016).
- [23] A. JIMÉNEZ-CORONA, C. A. AGUILAR-SALINAS, R. ROJAS-MARTÍNEZ, AND M. HERNÁNDEZ-ÁVILA, *Diabetes mellitus tipo 2 y frecuencia de acciones para su prevención y control*, Salud pública de México, 55 (2013), pp. S137–S143.

- [24] Y.-P. JIN, X.-F. SU, G.-P. YIN, X.-H. XU, J.-Z. LOU, J.-J. CHEN, Y. ZHOU, J. LAN, B. JIANG, Z. LI, ET AL., *Blood glucose fluctuations in hemodialysis patients with end stage diabetic nephropathy*, *Journal of Diabetes and its Complications*, 29 (2015), pp. 395–399.
- [25] J. JUN, S.-M. JIN, J. BAEK, S. OH, K. HUR, M.-S. LEE, M.-K. LEE, AND J. KIM, *The association between glycemic variability and diabetic cardiovascular autonomic neuropathy in patients with type 2 diabetes*, *Cardiovascular Diabetology*, 14 (2015), pp. 1–9.
- [26] M. E. LEAN, W. S. LESLIE, A. C. BARNES, N. BROSNAHAN, G. THOM, L. MCCOMBIE, C. PETERS, S. ZHYZHNEUSKAYA, A. AL-MRABEH, K. G. HOLLINGSWORTH, ET AL., *Primary care-led weight management for remission of type 2 diabetes (direct): an open-label, cluster-randomised trial*, *The Lancet*, 391 (2018), pp. 541–551.
- [27] ———, *Durability of a primary care-led weight-management intervention for remission of type 2 diabetes: 2-year results of the direct open-label, cluster-randomised trial*, *The lancet Diabetes & endocrinology*, 7 (2019), pp. 344–355.
- [28] M. LEE, J. L. SAVER, K.-S. HONG, S. SONG, K.-H. CHANG, AND B. OVBIAGELE, *Effect of pre-diabetes on future risk of stroke: meta-analysis*, *BMJ*, 344 (2012), p. e3564.
- [29] X. LIU, Y. XU, M. AN, AND Q. ZENG, *The risk factors for diabetic peripheral neuropathy: A meta-analysis*, *PLoS ONE*, 14 (2019), p. e0212574.
- [30] S. A. A. LONGINOS, E. D. G. TAMAY, AND M. B. H. MIRANDA, *Prevalencia de neuropatía diabética en pacientes con diabetes mellitus tipo 2 en una clínica regional del estado de México*, *Atención Familiar*, 25 (2017).
- [31] M. R. MARTÍNEZ, J. A. S. OTERO, O. B. LABRADA, H. N. S. SÁNCHEZ, AND A. L. B. ZAMORA, *Prognostic value of glycated hemoglobin hba1c in diabetic retinopathy in patients with type ii diabetes mellitus*, *Multimed*, 24 (2020), pp. 399–415.
- [32] J. H. MEDINA-CHÁVEZ, M. VÁZQUEZ-PARRODI, D. L. SANTOYO-GÓMEZ, J. AZUELA-ANTUNA, J. C. GARNICA-CUELLAR, A. HERRERA-LANDERO, AND D. A. BALANDRÁN-DUARTE, *Protocolo de atención integral: complicaciones crónicas de diabetes mellitus 2*, *Revista Médica del Instituto Mexicano del Seguro Social*, 60 (2022), p. S19.
- [33] M. I. MÚNERA-JARAMILLO, M. A. RESTREPO-LOZADA, L. M. GÓMEZ-BAHAMÓN, D. D. R. MESA-SUAREZ, AND B. S. RAMIREZ-PUERTA, *Hemoglobina glicosilada a1c vs. glucemia plasmática en ayunas de pacientes ambulatorios de un laboratorio médico*, *Revista de Salud Pública*, 13 (2011), pp. 980–989.
- [34] U. D. OF HEALTH, H. SERVICES, ET AL., *National diabetes statistics report, 2020*, *National Diabetes Statistics Report*, 2 (2020).
- [35] K. A. F. POVEDA, K. J. Q. GARCÍA, D. L. F. SUBÍA, AND C. A. C. CHOEZ, *Utilidad de hemoglobina glicosilada en diabetes tipo 2*, *RECIAMUC*, 4 (2020), pp. 118–126.

- [36] F. A. REYES SANAMÉ, M. L. PÉREZ ÁLVAREZ, E. ALFONSO FIGUEREDO, M. RAMÍREZ ESTUPIÑAN, AND Y. JIMÉNEZ RIZO, *Tratamiento actual de la diabetes mellitus tipo 2*, Correo científico médico, 20 (2016), pp. 98–121.
- [37] M. R. ROONEY, M. FANG, K. OGURTSOVA, B. OZKAN, J. B. ECHOUFFO-TCHEUGUI, E. J. BOYKO, D. J. MAGLIANO, AND E. SELVIN, *Global prevalence of prediabetes*, Diabetes Care, 46 (2023), pp. 1388–1394.
- [38] Y. TOMITA, D. LEE, K. TSUBOTA, K. NEGISHI, AND T. KURIHARA, *Updates on the current treatments for diabetic retinopathy and possibility of future oral therapy*, Journal of clinical medicine, 10 (2021), p. 4666.
- [39] W. N. VENABLES AND B. D. RIPLEY, *Modern Applied Statistics with S*, Springer, New York, fourth ed., 2002. ISBN 0-387-95457-0.
- [40] J. WADÉN, C. FORSBLOM, L. M. THORN, M. SARAHEIMO, M. ROSENGÅRD-BÄRLUND, O. HEIKKILÄ, K. HIETALA, K. ONG, N. WAREHAM, AND P.-H. GROOP, *Adult stature and diabetes complications in patients with type 1 diabetes: the finndiane study and the diabetes control and complications trial*, Diabetes, 58 (2009), pp. 1914–1920.