

UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
ESCUELA PROFESIONAL DE ESTADÍSTICA



TESIS

**Factores asociados con la diabetes en mujeres en edad reproductiva en la
Macro Región Norte del Perú según ENDES, 2024**

Para optar por el Título Profesional de:

Licenciado en Estadística

Presentado por:

Bach. Lisely Mirella Guadalupe Durand Valiente

Bach. Angie Michell Jibaja Guzman

Asesora:

Dra. Paredes López Lilian Roxana

Lambayeque – Perú

2026



**UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO
FACULTAD DE CIENCIAS FISICAS Y MATEMATICAS
UNIDAD DE INVESTIGACIÓN**



TESIS

**Factores asociados con la diabetes en mujeres en edad
reproductiva en la Macro Región Norte del Perú según ENDES,
2024**

Presentado por:

Bach. Lisely Mirella Guadalupe Durand Valiente

FIRMA

Bach. Angie Michell Jibaja Guzman

FIRMA

Asesor:

Dra. Lilian Roxana Paredes López

FIRMA



**UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO
FACULTAD DE CIENCIAS FISICAS Y MATEMATICAS
UNIDAD DE INVESTIGACIÓN**



TESIS

**Factores asociados con la diabetes en mujeres en edad
reproductiva en la Macro Región Norte del Perú según ENDES,
2024**

Aprobado por:

Dr. Wilver Omero Rodríguez López
Presidente

Dr. Grimaldo Dermalí Benavides Campos
Secretario

M.Sc. Hipólito Macalopú Inga
Vocal

ACTA DE SUSTENTACIÓN



UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS DECANATO

Ciudad Universitaria – Lambayeque

LICENCIADA - RESOLUCIÓN DEL CONSEJO DIRECTIVO N° 015 -2023-SUNEDU / CD



0060



ACTA DE SUSTENTACIÓN N° 33.-2024.-D/FACFyM

Siendo las 11:30 am del día 07 de abril del 2024, se reunieron los miembros del jurado evaluador de la Tesis titulada:

"Factores asociados con la diabetes en mujeres en edad reproductiva en la Macro Región Norte del Perú según ENDES, 2024"

Designados por Resolución N° 597-2025 D/FACFyM de fecha 30 de junio de 2025

Con la finalidad de evaluar y calificar la sustentación de la tesis antes mencionada, conformada por los siguientes docentes:

<u>Dr. Wilber Omer Rodríguez López</u>	Presidente
<u>Dr. Gerardo Bermúdez Benavides Campos</u>	Secretario
<u>H.Sc. Hipólito Macalope Inga</u>	Vocal

La tesis fue asesorada por (el) (la) Dra. Lilian Roxana Paredes López, nombrado por Resolución N° 597-2025 D/FACFyM de fecha 30 de junio 2025

El Acto de Sustentación fue autorizado por Resolución N° 244-2026 D/FACFyM de fecha 23 de marzo de 2026

La Tesis fue presentada y sustentada por (el) (los) Bachiller (es): Durand Valiente Lisely Mirella Guadalupe y Titya Gorman Angie Michelle y tuvo una duración de 50 minutos.

Después de la sustentación, y absueltas las preguntas y observaciones de los miembros del jurado se procedió a la calificación respectiva, otorgándole el Calificativo de dieciocho (18) en la escala vigesimal, mención (Muy Buena).

Por lo que queda(n) apto(s) para obtener el Título Profesional de Licenciado en Estadística, de acuerdo con la Ley Universitaria 30220 y la normatividad vigente de la Facultad de Ciencias Físicas y Matemáticas y la Universidad Nacional Pedro Ruiz Gallo.

Siendo las 12:30 se dio por concluido el presente acto académico, dándose conformidad al presente acto con la firma de los miembros del jurado.

Dr. Wilber Omer Rodríguez López
Presidente

Dr. Gerardo Bermúdez Benavides Campos
Secretario

H.Sc. Hipólito Macalope Inga
Vocal

Dra. Lilian Roxana Paredes López
Asesor

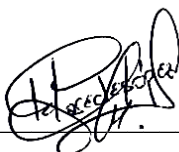
CONSTANCIA DE VERIFICACION DE ORIGINALIDAD DEL ASESOR

Yo, Dra. Paredes López Roxana, Asesor de tesis de las estudiantes, Lisely Mirella Guadalupe Durand Valiente y Angie Michell Jibaja Guzman, del trabajo de investigación titulado “Factores asociados con la diabetes en mujeres en edad reproductiva en la Macro Región Norte del Perú según ENDES, 2024”

Luego de la revisión exhaustiva del documento constato que la misma tiene un índice de similitud de 17% verificable en el reporte de similitud del programa de Turnitin.

El suscrito analizó dicho reporte y concluyó que cada una de las coincidencias detectadas no constituyen plagio, a mi leal saber y entender la tesis cumple con todas las normas para el uso de citas y referencias establecidas por la Universidad Nacional Pedro Ruiz Gallo.

Lambayeque, 9 de Julio del 2026



FIRMA

Lilian Roxana Paredes López

DNI:16655482

REPORTE DE TURNITIN

Factores asociados con la diabetes en mujeres en edad reproductiva en la Macro Región Norte c
según ENDES, 2024

INFORME DE ORIGINALIDAD

17%

INDICE DE SIMILITUD

13%

FUENTES DE INTERNET

8%

PUBLICACIONES

11%

TRABAJOS DEL ESTUD

FUENTES PRIMARIAS

1

hdl.handle.net

Fuente de Internet

2

repositorio.unprg.edu.pe

Fuente de Internet

3

Submitted to Universidad Internacional de la Rioja

Trabajo del estudiante

4

Submitted to Universitat Oberta de Catalunya

Trabajo del estudiante

5

Submitted to Universidad de Manizales

Trabajo del estudiante

6

Submitted to Ilerna Online Blackboard

Trabajo del estudiante

7

Submitted to University College London

Trabajo del estudiante

8

archivos.evidenciasenpediatria.es

Fuente de Internet

9

www.codigosanluis.com

Fuente de Internet

10

Submitted to Universidad Autonoma de Chile

Trabajo del estudiante

11

repositorio.unjfsc.edu.pe

Fuente de Internet

12

dgsa.uaeh.edu.mx:8080

Fuente de Internet

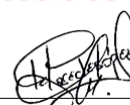
13

Submitted to Universidad Nacional Pedro Ruiz Gallo

Trabajo del estudiante

14

Débora Cristina Martins, Giordana Maronezzi da Silva, Beatriz Maria dos Santos Santiago Ribeiro, Elen Ferraz Teston et al. "Chronic diseases among women of reproductive age primary care: prevalence and associated factors", Revista Gaúcha de Enfermagem, 202



FIRMA

Lilian Roxana Paredes López

DNI: 16655482



Recibo digital

Este recibo confirma que su trabajo ha sido recibido por **Turnitin**. A continuación podrá ver la información del recibo con respecto a su entrega.

La primera página de tus entregas se muestra abajo.

Autor de la entrega:	Lisely Mirella Guadalupe Durand Valiente / Angie Michell Jibaja	...
Título del ejercicio:	Quick Submit	
Título de la entrega:	Factores asociados con la diabetes en mujeres en edad repro	...
Nombre del archivo:	ella_Guadalupe_Durand_Valiente,_Angie_Michell_Jibaja_Guzma	...
Tamaño del archivo:	705.84 K	
Total páginas:	61	
Total de palabras:	13,456	
Total de caracteres:	77,729	
Fecha de entrega:	09-jul-2026 11:42p. m. (UTC -0500)	
Identificador de la entrega:	2995880656	

UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
ESCUELA PROFESIONAL DE ESTADÍSTICA



TESIS

Factores asociados con la diabetes en mujeres en edad reproductiva en la
Macro Región Norte del Perú según ENDES, 2024

Para optar por el Título Profesional de:
Licenciado en Estadística

Presentado por:
Bach. Lisely Mirella Guadalupe Durand Valiente
Bach. Angie Michell Jibaja Guzman

Asesora:
Dra. Paredes López Lilian Roxana

Lambayeque - Perú
2026

FIRMA

Lilian Roxana Paredes López
DNI:16655482

AGRADECIMIENTO

Agradezco a Dios por darme la fortaleza y la claridad necesarias para llegar hasta aquí.

A mi familia, especialmente a mi madre y a mi tía, por su apoyo constante, su cariño y por impulsarme siempre a seguir adelante.

A mi novio, por su paciencia, su comprensión y por acompañarme en los momentos más exigentes de este camino. A mi asesora de tesis, la Dra. Lilian Roxana Paredes López, por su guía continua.

A mis profesores, por su dedicación y por cada enseñanza que contribuyó a mi formación.

Lisely Mirella Guadalupe Durand Valiente

A mis Hermanos, Julio y Alejandra, por su amor incondicional, su compañía y por ser la alegría en mi vida. Ustedes son mi motor para seguir siendo un ejemplo de perseverancia.

A mi Tío William, por su valiosa ayuda, su motivación constante y su apoyo incondicional en los momentos cruciales de este proceso. Finalmente a Dios, por guiar mis pasos, brindarme salud y darme la fuerza y la sabiduría necesarias para culminar este proyecto. Gracias a toda mi familia y a mi pequeño perrito Milancito.

Angie Michell Jibaja Guzman

DEDICATORIA

Dedico esta tesis a mi madre Margarita y a mi tía Cecilia, quienes han sido mi refugio, mi impulso y la luz que me sostuvo incluso en los días en que dudé de mí misma. Gracias por su amor incondicional, por el apoyo de diferentes maneras y por enseñarme a no rendirme nunca.

A mi novio, por su paciencia, por su ternura, motivación diaria y por recordarme, en cada palabra, que soy capaz de alcanzar mis metas, pese a las circunstancias de la vida siempre está pendiente de mí.

Este logro es un pedacito de cada uno de ustedes, y lo llevo conmigo con gratitud profunda.

Gracias por ser parte esencial de este camino que hoy culmina.

Lisely Mirella Guadalupe Durand Valiente

Dedico esta tesis, el fruto de mi esfuerzo, a mis padres quienes me dieron la vida, la educación y cada oportunidad. Gracias por sus innumerables desvelos y por la confianza que me impulsó a llegar hasta aquí. Los amo profundamente.

Con todo mi cariño a mis Abuelos: Angelita, por su ternura, su refugio constante y sus consejos sabios. Y a mi abuelo Julio, que, aunque hoy no esté físicamente, sé que desde donde está me cuida y me protege, siendo mi inspiración de fortaleza. Su legado ha marcado mi camino.

Angie Michell Jibaja Guzman

Índice de Contenido

INTRODUCCIÓN	xi
DISEÑO TEÓRICO.....	5
• Antecedentes	5
• Bases Teóricas.....	7
• Bases Conceptuales.....	15
DISEÑO METODOLÓGICO.....	18
• Diseño de Contrastación de Hipótesis	18
• Población y Muestra	19
• Técnicas, Instrumentos, Equipos y Materiales	19
RESULTADOS	20
DISCUSIÓN	28
CONCLUSIONES	32
RECOMENDACIONES.....	34
REFERENCIAS.....	35
Anexos	40
Anexo 1. Script de ejecución en código R.....	40
Anexo 2. Output de ejecución en código R	43
Anexo 1. Resultados en SPSS.....	48

Índice de tablas

Tabla 1 Operacionalización de variable	16
Tabla 2 Dimensión salud reproductiva y materna en mujeres en edad reproductiva en la Macro Región Norte del Perú según ENDES, 2024	20
Tabla 3 Dimensión salud física en mujeres en edad reproductiva en la Macro Región Norte del Perú según ENDES, 2024	20
Tabla 4 Dimensión salud mental en mujeres en edad reproductiva en la Macro Región Norte del Perú según ENDES, 2024	21
Tabla 5 Dimensión características sociodemográficas en mujeres en edad reproductiva en la Macro Región Norte del Perú según ENDES, 2024	22
Tabla 6 Prueba de chi-cuadrado dimensiones y la variable Diabetes en mujeres en edad reproductiva en la Macro Región Norte del Perú según ENDES, 2024	23
Tabla 7 Modelo logístico binario de los factores asociados con la diabetes en mujeres en edad reproductiva en la Macro Región Norte del Perú según ENDES, 2024	24
Tabla 8 Matriz de confusión del modelo logístico binario de los factores asociados con la diabetes en mujeres en edad reproductiva en la Macro Región Norte del Perú según ENDES, 2024.....	26
Tabla 9 Indicadores de rendimiento del modelo logístico binario de los factores asociados con la diabetes en mujeres en edad reproductiva en la Macro Región Norte del Perú según ENDES, 2024.....	27

Índice de figuras

Figura 1 Curva de ROC del modelo logístico binario de los factores asociados con la diabetes en mujeres en edad reproductiva en la Macro Región Norte del Perú según ENDES, 2024.....	25
--	----

Resumen

El objetivo de la investigación fue determinar las variables de salud reproductiva y materna, salud física, salud mental y características demográficas asociadas con la diabetes en mujeres de 15 a 49 años de la Macro Región Norte del Perú, utilizando la base de datos ENDES 2024. Se desarrolló un estudio cuantitativo, básico, explicativo, de diseño no experimental y transversal, con una muestra de 2 421 mujeres seleccionadas de la ENDES. Se aplicaron tablas de frecuencia, pruebas de chi-cuadrado y un modelo de regresión logística binaria sobre una base balanceada mediante el método híbrido ROSE. Los resultados mostraron que el embarazo y la presión alta incrementan significativamente la probabilidad de presentar diabetes, mientras que ser menor de 35 años, contar con educación superior y reportar depresión se asociaron con menor riesgo; además, la residencia urbana aumentó la probabilidad de diagnóstico. El modelo alcanzó un AUC de 0.713, con sensibilidad de 0.762 y especificidad de 0.564, evidenciando capacidad discriminativa moderada. Se concluye que la diabetes en mujeres en edad reproductiva responde a una interacción multifactorial de factores biológicos, clínicos y sociodemográficos, destacando la utilidad de la regresión logística binaria para su análisis epidemiológico.

Palabras clave: diabetes, mujeres en edad reproductiva, regresión logística, ENDES 2024.

Abstract

The aim of this research was to determine reproductive and maternal health, physical health, mental health and sociodemographic variables associated with diabetes among women aged 15–49 years in the Northern Macro Region of Peru, using the 2024 Demographic and Family Health Survey (ENDES) database. A quantitative, basic, explanatory, non-experimental and cross-sectional study was conducted on a sample of 2,421 women obtained from ENDES. Frequency tables, chi-square tests and a binary logistic regression model were applied on a balanced dataset generated through the hybrid ROSE method. Results showed that pregnancy and high blood pressure significantly increased the likelihood of diabetes, whereas being younger than 35 years, having higher education and reporting depression were associated with a lower risk; urban residence was also related to a higher probability of diagnosis. The model achieved an AUC of 0.713, with a sensitivity of 0.762 and a specificity of 0.564, indicating moderate discriminative performance. It is concluded that diabetes in women of reproductive age arises from a multifactorial interaction of biological, clinical and sociodemographic factors, and that binary logistic regression is a useful tool for epidemiological analysis in this population.

Keywords: diabetes, women of reproductive age, logistic regression, ENDES 2024.

INTRODUCCIÓN

La diabetes, especialmente en mujeres en edad reproductiva, constituye un problema de salud significativo. Las mujeres diabéticas enfrentan un amplio espectro de complicaciones, como retrasos en la pubertad y la menarquia, trastornos menstruales, subfertilidad y riesgos elevados durante el embarazo, además de la posibilidad de una menopausia precoz. A pesar de los avances en el tratamiento con insulina, los problemas reproductivos persisten, y nuevas afecciones como el síndrome de ovario poliquístico y el hiperandrogenismo continúan emergiendo (Thong et al., 2020).

En México, un estudio sobre el diagnóstico de diabetes mellitus tipo 2 reveló que un 92.7% de las mujeres participantes se encontraban en edad reproductiva. La diabetes en mujeres en edad fértil no solo plantea riesgos metabólicos, sino también complicaciones significativas en su salud reproductiva. Estas mujeres enfrentan una amplia gama de trastornos, como retrasos en la pubertad, alteraciones en el ciclo menstrual, subfertilidad y problemas durante el embarazo, además de un riesgo incrementado de menopausia precoz (Heredia y Gallegos, 2022).

Por otro lado, Ordoñez y Gallardo (2024) en su estudio en Ecuador sobre el conocimiento de las complicaciones del síndrome de ovario poliquístico (SOP) en mujeres con diabetes mostró que un 68.2% de las participantes no tiene conocimiento sobre dichas complicaciones. En cuanto a la relación entre infertilidad y diabetes mellitus, solo un 23.6% reconoció esta conexión, mientras que un 8.2% mencionó insuficiencia y hemorragia como posibles complicaciones. En términos de edad, las mujeres menores de 25 años presentaron el nivel más alto de desconocimiento, con un 39.1%, seguidas por un 26.4% de mujeres de 26 a 37 años. Por otro lado, el grupo de mujeres de 38 a 47 años mostró un 2.7% de desconocimiento sobre el tema.

A nivel nacional, se evidencio que el 73.6% de las mujeres gestantes con diabetes presentó algún tipo de complicación materna, de las cuales el 15.2% correspondió a complicaciones relacionadas directamente con el curso del embarazo. Estos datos reflejan la alta prevalencia de problemas obstétricos en mujeres con diabetes, especialmente en aquellas que se encuentran en edad reproductiva, lo que agrava aún más el impacto de esta enfermedad crónica sobre la salud reproductiva y materna (Victorio et al., 2024).

En la región de Lambayeque, Peña (2022) identifico que el 70% de las mujeres presentan un riesgo moderado o alto de desarrollar diabetes mellitus tipo 2. Esta cifra resulta especialmente preocupante si se considera el impacto que dicha enfermedad puede tener en la salud reproductiva femenina, ya que incrementa la probabilidad de experimentar alteraciones menstruales, subfertilidad, complicaciones durante el embarazo y trastornos hormonales como el síndrome de ovario poliquístico. La elevada proporción de mujeres en situación de riesgo evidencia una problemática de salud pública que afecta de manera particular a la población femenina en edad reproductiva.

En el contexto de la Macro Región Norte del Perú, las complicaciones reproductivas asociadas a la diabetes mellitus tipo 2 en mujeres en edad reproductiva representan una problemática compleja que involucra múltiples factores biológicos, sociales y conductuales. Esta complejidad exige un enfoque analítico que permita comprender cómo distintas variables interactúan y contribuyen a la aparición de dichos desenlaces. En este sentido, se hace evidente la necesidad de utilizar modelos estadísticos como la regresión logística binaria, que permiten analizar la relación entre diversas características individuales y la probabilidad de presentar complicaciones específicas, ofreciendo una base cuantitativa para comprender la magnitud y naturaleza del problema en esta población.

La justificación teórica se presenta en que la diabetes mellitus tipo 2 en mujeres en edad reproductiva constituye un fenómeno multifactorial que requiere ser analizado desde modelos

estadísticos capaces de manejar variables categóricas. La regresión logística binaria se fundamenta teóricamente en el análisis de probabilidades condicionadas y se emplea ampliamente en investigaciones de salud para estudiar relaciones entre factores de riesgo y desenlaces dicotómicos. En el caso de la diabetes, este modelo ofrece una herramienta sólida para explorar cómo variables como la edad, el índice de masa corporal, los antecedentes familiares y el estilo de vida inciden en la aparición de complicaciones reproductivas, contribuyendo a un marco explicativo más preciso y con respaldo científico.

La justificación práctica del uso de regresión logística binaria se presenta en los beneficios concretos para los servicios de salud en la Macro Región Norte del Perú, ya que permite generar perfiles de riesgo reproductivo asociados a la diabetes tipo 2 en mujeres jóvenes. Esta información es clave para el diseño de estrategias preventivas y de intervención temprana, optimizando los recursos disponibles en los establecimientos de salud. Al identificar qué factores tienen mayor peso en la aparición de complicaciones, se pueden establecer prioridades en la atención médica y educativa, permitiendo una respuesta más efectiva ante una problemática que afecta a una proporción considerable de la población femenina en edad reproductiva.

El estudio se justifica metodológicamente, la regresión logística binaria es adecuada para el análisis de datos en estudios donde el resultado es dicotómico, como la presencia o ausencia de complicaciones reproductivas en mujeres con riesgo de diabetes tipo 2. Esta técnica permite controlar múltiples variables independientes de manera simultánea, ofreciendo un análisis multivariado robusto que facilita la identificación de asociaciones significativas. Además, su flexibilidad para trabajar con variables categóricas y continuas permite una modelación más realista del fenómeno estudiado, ajustándose a las características del contexto local y a los objetivos específicos de investigación en la Macro Región Norte del Perú.

Según lo mencionado, nos planteamos como problema de investigación ¿Cuáles son las variables del factor: salud reproductiva y materna, salud física, salud mental y de las características demográficas asociadas con la diabetes en mujeres en edad reproductiva en la Macro Región Norte del Perú según ENDES, 2024? Además, se plantea como hipótesis que los factores, salud reproductiva y materna: Actualmente embarazada, número de hijos, uso de anticonceptivos; salud física: Anemia, diagnóstico de presión alta; Salud mental: Depresión y Características demográficas: Edad, área de residencia, cuenta con seguro de salud, educación superior, están asociadas a la presencia de diabetes en mujeres en edad reproductiva en la Macro Región Norte del Perú según ENDES, 2024.

Como objetivo general se plantea determinar las variables del factor: salud reproductiva y materna, salud física, salud mental y de las características demográficas asociadas con la diabetes en mujeres en edad reproductiva en la Macro Región Norte del Perú según ENDES, 2024. Como objetivos específicos, describir las características de las mujeres en edad reproductiva en la Macro Región Norte del Perú según ENDES, 2024. Evaluar la relación entre las condiciones sociodemográficas y el diagnóstico de diabetes en mujeres en edad reproductiva en la Macro Región Norte del Perú según ENDES, 2024. Elaborar un modelo de regresión logística binaria para identificar los factores asociados con la diabetes en mujeres en edad reproductiva en la Macro Región Norte del Perú según ENDES, 2024.

DISEÑO TEÓRICO

- **Antecedentes**

A nivel internacional, Sun et al. (2022) se estimó que más del 10,5% de la población adulta mundial, es decir, aproximadamente 536,6 millones de personas entre 20 y 79 años, vivían con diabetes, una cifra que podría ascender a 783,2 millones para el año 2045. Esta proyección se obtuvo a partir de 219 fuentes de datos de calidad recopiladas entre 2005 y 2020, que cubrían 215 países y territorios. En los países sin información suficiente, se utilizaron estimaciones basadas en características similares como economía, etnicidad, geografía e idioma. La prevalencia fue comparable entre hombres y mujeres, más elevada en personas de 75 a 79 años, y se observó mayor incidencia en zonas urbanas y en países de altos ingresos. El crecimiento más significativo en prevalencia se proyecta en países de ingresos medios. Además, los costos sanitarios atribuibles a la diabetes ascendieron a 966 mil millones de dólares en 2021, con una proyección de 1,054 mil millones para 2045, evidenciando el creciente impacto económico y sanitario de esta enfermedad a nivel global.

Rojo et al. (2020) tuvo como objetivo determinar la incidencia de diabetes mellitus tipo 2 en una cohorte representativa de la población adulta en España. Se seleccionaron aleatoriamente 5,072 personas mayores de 18 años, a quienes se les recogieron datos sociodemográficos, clínicos, hábitos de vida, medidas antropométricas y presión arterial. También se realizaron pruebas de glucosa en ayunas y de tolerancia oral a la glucosa. En el seguimiento, participaron 2,408 personas, de las cuales 154 desarrollaron diabetes, lo que representa una incidencia acumulada del 6,4% en 7,5 años. La tasa ajustada fue de 11,6 casos por cada 1,000 personas-año, siendo mayor en hombres, personas con prediabetes, obesidad, obesidad central, antecedentes familiares y aumento de peso. Los resultados ofrecen datos valiosos sobre la incidencia de diabetes tipo 2 y sus factores de riesgo en la población española, contribuyendo a la planificación de estrategias de prevención y control de esta enfermedad.

Bustamante et al. (2023) En Chile, en 2003, el 1,2% de las embarazadas tenía diagnóstico de diabetes. Este estudio descriptivo y transversal analizó la tasa de egresos hospitalarios por ambas condiciones entre 2018 y 2021, en una muestra de 42.631 mujeres embarazadas. Se observó una mayor tasa de egreso por diabetes gestacional (117,17) frente a la pregestacional (27,64), especialmente en el grupo de 20-44 años. La estancia hospitalaria fue mayor en la diabetes pregestacional (5,03 días). Tras 2020, ambas tasas disminuyeron, posiblemente debido a la pandemia por COVID-19. Se identifican factores como la edad, el sedentarismo y el mal control metabólico como posibles causas del aumento previo. El estudio destaca la necesidad de mantener controles prenatales y analizar el impacto del COVID-19.

A nivel nacional, Prado et al. (2023) La diabetes mellitus gestacional es la complicación metabólica más común durante el embarazo, con una prevalencia global que oscila entre el 1 % y el 14 %, siendo América Latina una de las regiones más vulnerables. Esta condición representa un problema de salud pública debido a su alto impacto tanto en la madre como en el feto. Entre los factores de riesgo más asociados se encuentran la edad materna avanzada, la obesidad y los antecedentes médicos, los cuales incrementan en 1,2 veces la probabilidad de desarrollar esta patología. Sin embargo, es importante señalar que muchas mujeres que presentan diabetes gestacional no presentan dichos factores. Además, estas pacientes tienen entre un 35 % y 60 % de probabilidad de desarrollar diabetes tipo 2 en los 10 a 20 años posteriores al embarazo. Por ello, resulta fundamental detectar y tratar a tiempo cualquier alteración en los niveles de glucosa durante la gestación, a fin de reducir las complicaciones materno-fetales inmediatas y a largo plazo.

Tamata (2021) en su estudio titulado Factores de Riesgo y Prevalencia de Diabetes Mellitus Tipo II en mujeres adultas atendidas en el consultorio de Endocrinología del Hospital Guillermo Díaz de La Vega de Abancay tuvo como finalidad identificar los factores de riesgo asociados a la diabetes tipo 2 en mujeres adultas. Se analizaron variables como la actividad

física, el índice de masa corporal, la obesidad abdominal, el perfil lipídico, la presión arterial y los niveles de hemoglobina glicosilada. La investigación fue explicativa, de tipo transversal, y se llevó a cabo entre julio y septiembre con una muestra de 49 mujeres. Para evaluar la actividad física se utilizó el Cuestionario Internacional de Actividad Física (IPAQ) en su versión abreviada, complementado con entrevistas, mediciones antropométricas, análisis bioquímicos y toma de presión arterial. Los resultados revelaron que el 55,1 % presentaba baja actividad física, el 79,6 % tenía sobrepeso u obesidad, y el 85,7 % mostraba obesidad abdominal. Además, el 89,8 % tenía colesterol alterado, el 70,6 % triglicéridos elevados, el 69,4 % colesterol HDL alterado, el 64,3 % colesterol LDL alterado, el 36,7 % hipertensión arterial y el 69,4 % una inadecuada gestión clínica de su enfermedad.

Seclen (2024) Las enfermedades crónicas no transmisibles (ECNT) son la principal causa de muerte entre mujeres a nivel mundial, y la gestación representa un momento clave para intervenir en su perfil de riesgo. En Perú, la diabetes mellitus gestacional (DMG) es el trastorno médico más frecuente durante el embarazo, con una prevalencia del 14–16 %. Factores como la edad, antecedentes familiares y un IMC superior a 30 kg/m² incrementan este riesgo. La DMG no solo afecta el embarazo, sino que también eleva el riesgo futuro de diabetes tipo 2, obesidad y enfermedades cardiovasculares tanto en la madre como en el hijo. A pesar de esto, persiste una deficiente implementación de estrategias de prevención y diagnóstico, debido a limitaciones de infraestructura y personal en el primer nivel de atención. La reciente inclusión de la DMG en la Ley General de Protección a personas con DM ofrece una oportunidad para fortalecer el abordaje integral y temprano de esta condición.

- **Bases Teóricas**

La diabetes según Marc Lalonde quien señaló que las enfermedades no siempre están vinculadas directamente a agentes infecciosos, sino que muchas tienen un origen sociodemográfico, es decir, resultan de una combinación de factores personales, sociales,

económicos y ambientales. En el caso de la diabetes mellitus, esta se asocia a diversas condiciones biológicas, sociales y ambientales (Azriel et al., 2008).

El estilo de vida tiene una influencia significativa en los riesgos o beneficios para la salud. Hábitos como una alimentación inadecuada, la inactividad física, el consumo de sustancias nocivas y una mala higiene son factores que comúnmente se subestiman, pero que afectan negativamente el bienestar. Adoptar un estilo de vida saludable reduce la probabilidad de desarrollar enfermedades como la diabetes mellitus (Pomares et al., 2014).

Dorothea Orem planteó el concepto de autocuidado como un medio para prevenir enfermedades, promoviendo prácticas que favorezcan la salud, ya sea de manera autónoma o con el apoyo del personal sanitario (Bustamante et al., 2023).

El autocuidado es esencial para mantener una vida sana, ya que favorece el desarrollo personal y el bienestar. Orem estructuró su teoría en tres componentes: la teoría del autocuidado, que destaca la capacidad individual e intencionada de mantener la salud; la teoría del déficit de autocuidado, que expone la necesidad de intervención profesional cuando una persona no puede cuidarse por sí misma; y la teoría de los sistemas de enfermería, que busca satisfacer las necesidades del paciente y restablecer su capacidad de autocuidado (Naveiro et al., 2015).

La diabetes es una enfermedad crónica que se caracteriza por niveles elevados de glucosa en la sangre, resultado de una alteración en la producción o en la acción de la insulina, una hormona producida por el páncreas. Esta hormona es esencial para que la glucosa ingrese a las células del cuerpo, donde se utiliza como fuente principal de energía. Cuando este proceso falla, la glucosa se acumula en la sangre, dando lugar a un estado conocido como hiperglucemia (Soriguer et al., 2012).

Existen varios tipos de diabetes, siendo los más comunes el tipo 1, el tipo 2 y la gestacional. La diabetes tipo 1 suele aparecer en la infancia o adolescencia, y se debe a una

destrucción autoinmune de las células beta del páncreas, lo que provoca una deficiencia total de insulina. En cambio, la diabetes tipo 2, que representa la mayoría de los casos, se presenta con mayor frecuencia en adultos y está relacionada con la resistencia a la insulina y un déficit relativo de esta hormona. Por otro lado, la diabetes gestacional ocurre durante el embarazo y, si bien suele desaparecer tras el parto, puede dejar secuelas metabólicas tanto en la madre como en el bebé (Dávila Flores et al., 2023).

El desarrollo de la diabetes está estrechamente vinculado a diversos factores, tanto genéticos como ambientales. La predisposición familiar, el sobrepeso, la obesidad, el sedentarismo y los malos hábitos alimenticios aumentan considerablemente el riesgo de padecer esta enfermedad. Asimismo, el estrés, los trastornos del sueño y otros factores psicosociales también pueden influir en su aparición o en la descompensación de quienes ya la padecen (Rojo et al., 2020).

Las personas con diabetes, si no mantienen un control adecuado de sus niveles de glucosa, están expuestas a múltiples complicaciones a lo largo del tiempo. Estas complicaciones pueden ser agudas, como la hipoglucemia o la cetoacidosis diabética, y crónicas, como enfermedades cardiovasculares, insuficiencia renal, daño ocular (retinopatía), neuropatía diabética e incluso amputaciones en casos severos. Además, la diabetes puede afectar la calidad de vida emocional y social del paciente (Seclen, 2024).

A pesar de su gravedad, la diabetes es una enfermedad controlable si se lleva un tratamiento adecuado que incluya una alimentación saludable, actividad física regular, control médico periódico y, en muchos casos, el uso de medicamentos o insulina. La educación del paciente y su entorno es también fundamental para mejorar la adherencia al tratamiento y reducir los riesgos a largo plazo (Sun et al., 2022).

Existen varios tipos de diabetes, siendo los más comunes la diabetes tipo 1, el tipo 2 y la diabetes gestacional.

La diabetes tipo 1 se presenta generalmente en la infancia o adolescencia, aunque también puede desarrollarse en adultos. En este tipo, el sistema inmunológico ataca por error a las células del páncreas que producen insulina, lo que lleva a una deficiencia total de esta hormona. Por esta razón, las personas con diabetes tipo 1 necesitan inyecciones diarias de insulina para sobrevivir (Díaz et al., 2021).

Por otro lado, la diabetes tipo 2 es la más frecuente y suele aparecer en adultos, aunque actualmente también se diagnostica en jóvenes debido al aumento del sobrepeso y la obesidad. En este tipo de diabetes, el cuerpo no utiliza adecuadamente la insulina, lo que se conoce como resistencia a la insulina, y con el tiempo puede disminuir la producción de esta. La diabetes tipo 2 está estrechamente relacionada con factores como la mala alimentación, la inactividad física y antecedentes familiares (Tamata, 2021).

La diabetes gestacional aparece durante el embarazo y se caracteriza por una intolerancia a los carbohidratos que puede llevar a niveles elevados de glucosa en sangre. Aunque suele desaparecer después del parto, esta condición aumenta el riesgo de desarrollar diabetes tipo 2 en el futuro tanto para la madre como para el hijo (Zavala et al., 2024).

Existen también otros tipos menos comunes, como la diabetes monogénica, que es causada por mutaciones genéticas específicas, y la diabetes secundaria, que se origina como consecuencia de otras enfermedades o del uso de ciertos medicamentos. Todos estos tipos de diabetes requieren atención médica continua, cambios en el estilo de vida y, en muchos casos, tratamiento farmacológico para mantener el control adecuado de la glucosa y prevenir complicaciones a largo plazo (Pérez et al., 2024).

La diabetes está influenciada por una combinación de diversos factores de riesgo. Entre los más relevantes se encuentran los antecedentes familiares, ya que tener parientes cercanos con diabetes aumenta significativamente la probabilidad de desarrollar la enfermedad. La edad

también es un factor determinante, especialmente en la diabetes tipo 2, donde el riesgo incrementa a partir de los 45 años (Prado et al., 2023).

Las mujeres que padecen diabetes enfrentan una serie de complicaciones que pueden afectar tanto su salud general como su calidad de vida. Una de las más comunes es el daño a los vasos sanguíneos, lo que incrementa el riesgo de enfermedades cardiovasculares como infartos y accidentes cerebrovasculares, siendo estas complicaciones más severas en mujeres que en hombres. Además, la diabetes puede alterar el funcionamiento de los riñones, provocando una condición conocida como nefropatía diabética, que en casos graves puede llevar a insuficiencia renal (Ciria, 2021).

También se observa una mayor probabilidad de padecer retinopatía diabética, una afección que daña la retina y puede conducir a la pérdida progresiva de la visión. En el sistema nervioso, la neuropatía diabética causa dolor, entumecimiento o debilidad, principalmente en las extremidades, y puede afectar la función digestiva, urinaria y sexual (Tamata, 2021).

Durante el embarazo, las mujeres con diabetes preexistente o diabetes gestacional están expuestas a riesgos tanto para ellas como para el bebé. Estas incluyen preeclampsia, parto prematuro, infecciones frecuentes y un mayor riesgo de cesárea. El feto, por su parte, puede presentar malformaciones congénitas, macrosomía (bebé de gran tamaño), hipoglucemia al nacer y, a largo plazo, mayor predisposición a desarrollar obesidad y diabetes tipo 2 (Zavala et al., 2024).

Además, la diabetes puede impactar en la salud mental, provocando ansiedad, depresión y fatiga crónica, lo que dificulta el control adecuado de la enfermedad. Por ello, es fundamental que las mujeres con diabetes reciban atención médica integral, con controles periódicos, educación en salud y un enfoque preventivo que permita minimizar estas complicaciones y mejorar su bienestar general (Prado et al., 2023).

El estilo de vida juega un papel crucial. El sedentarismo, la falta de actividad física regular y una alimentación poco saludable, rica en azúcares, grasas saturadas y carbohidratos refinados, contribuyen al desarrollo de sobrepeso y obesidad, condiciones fuertemente asociadas con la diabetes. En especial, la acumulación de grasa abdominal está relacionada con una mayor resistencia a la insulina (Naveiro et al., 2015).

Además, factores como la hipertensión arterial, niveles elevados de colesterol o triglicéridos, y el síndrome de ovario poliquístico en mujeres, también pueden predisponer a esta enfermedad. El tabaquismo y el consumo excesivo de alcohol actúan como agravantes del riesgo metabólico. En el caso de la diabetes gestacional, haber tenido embarazos anteriores con esta condición, bebés de gran tamaño o abortos recurrentes son señales de alerta (Sun et al., 2022).

La regresión logística (RL) constituye una metodología estadística empleada para establecer la relación entre una variable dependiente de naturaleza cualitativa (Y) y un conjunto de variables independientes, las cuales pueden ser tanto cuantitativas como cualitativas (X). La clasificación de esta técnica varía según el tipo de variable dependiente: se distingue así entre regresión logística binaria, cuando la variable es dicotómica; regresión logística multinomial, cuando presenta más de dos categorías; y regresión logística ordinal, cuando las categorías poseen un orden inherente (Ortega et al., 2022). En la actualidad, la regresión logística, al igual que otras formas de regresión, se encuentra comprendida dentro del marco de los modelos lineales generalizados (MLG), los cuales se rigen por la siguiente expresión matemática:

$$f(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n$$

Cuando la variable dependiente es de tipo categórico, se establece que $Y = 0$ representa la ausencia del carácter estudiado, mientras que $Y = 1$ indica su presencia. A diferencia de la regresión lineal, cuyo objetivo es estimar los valores medios de la variable dependiente, la regresión logística se orienta a la predicción de probabilidades (p) de ocurrencia del evento, en

función de los valores que adopten las variables independientes, dentro de un rango comprendido entre 0 y 1. Si bien es factible aplicar un modelo de regresión lineal por mínimos cuadrados en este contexto, los resultados obtenidos carecerían de validez interpretativa, ya que, ante valores extremos de las variables predictoras, las estimaciones de Y podrían situarse fuera del intervalo $[0,1]$, lo cual contradice el concepto estadístico de probabilidad. Para subsanar esta limitación, es necesario emplear una función que permita transformar las salidas del modelo lineal en valores restringidos al intervalo mencionado (Orraca et al., 2021). Existen diversas funciones que cumplen con esta condición, aunque la más comúnmente utilizada es la función logística o sigmoide, cuya formulación se expresa de la siguiente manera:

$$y = \frac{1}{1 + e^{-f(x)}}$$

- Si z tiene un valor muy grande positivo $e^{-f(x)}$ es aproximadamente 0, $y = 1$.
- Si $z = 0$, $e^{-f(x)} = 1$, $y = 0,5$.
- Si z tiene un valor muy grande negativo, $e^{-f(x)} = \infty$, $y = 0$.

Con el propósito de cuantificar el grado de varianza explicada por el modelo ajustado, se han desarrollado indicadores conocidos genéricamente como Pseudo R^2 , los cuales se inspiran en el coeficiente de determinación R^2 empleado en la regresión lineal. Estos indicadores adoptan valores dentro del intervalo de 0 a 1, y su magnitud se interpreta de forma proporcional: a mayor valor, mayor proporción de la varianza explicada por el modelo. No obstante, es importante señalar que estos coeficientes no constituyen pruebas directas de la bondad de ajuste (Ortega et al., 2022).

Entre las variantes existentes, el Pseudo R^2 de McFadden realiza una comparación entre el logaritmo de la verosimilitud del modelo nulo —aquel que incluye únicamente la constante (L_0)— y el correspondiente al modelo que incorpora la variable predictora (LM). Esta medida, sin embargo, es de uso relativamente limitado.

$$R_{MF}^2 = 1 - \frac{l(0)}{l(M)}$$

El PseudoR² de Cox-Snell establece una comparación entre la verosimilitud del modelo nulo (L₀), que solo incluye la constante, y la verosimilitud del modelo que incorpora la variable predictora (LM). Este indicador, al igual que otras medidas de tipo PseudoR², busca ofrecer una estimación del grado de ajuste del modelo, aunque sin constituir una prueba concluyente de su bondad.

$$R_c^2 = 1 - \left(\frac{l(0)}{l(M)} \right)^{\frac{2}{n}}; n = \text{tamaño muestral}$$

El PseudoR² de Cox-Snell permite estimar la proporción de varianza explicada por la variable independiente. Sin embargo, presenta como limitación el hecho de que no alcanza el valor 1, incluso cuando el modelo logra explicar la totalidad de la variabilidad de la variable dependiente.

Por su parte, el PseudoR² de Nagelkerke ajusta esta deficiencia al modificar la fórmula de Cox-Snell, de manera que pueda alcanzar el valor máximo de 1 cuando el modelo explica el 100 % de la varianza. Debido a esta propiedad, se ha convertido en la medida más empleada dentro de este tipo de indicadores.

$$R_N^2 = 1 - \frac{R^2}{R_{max}^2}; R_{max}^2 = 1 - l(0)^{\frac{2}{n}}$$

El análisis basado en las curvas receiver operating characteristic (ROC), comúnmente traducidas como “característica operativa del receptor” —aunque una traducción más precisa sería “curva de eficacia diagnóstica”—, constituye una herramienta estadística empleada para evaluar la precisión diagnóstica de pruebas que generan resultados en escalas continuas. Este tipo de análisis se utiliza con tres finalidades principales: identificar el punto de corte óptimo que maximiza simultáneamente la sensibilidad y la especificidad; valorar la capacidad

discriminativa de una prueba diagnóstica; y comparar el desempeño discriminativo de dos o más pruebas que expresan sus resultados mediante escalas continuas (Martínez y Pérez, 2023).

- **Bases Conceptuales**

Salud reproductiva y materna: La salud reproductiva y materna se refiere al bienestar físico y funcional de las mujeres en relación con su capacidad reproductiva. Abarca aspectos como el embarazo, el uso de métodos anticonceptivos y la maternidad. Esta dimensión considera tanto la posibilidad de concebir como las decisiones relacionadas con la planificación familiar. Evalúa condiciones que pueden influir directa o indirectamente en la salud general de la mujer y su predisposición a enfermedades crónicas, incluyendo la diabetes (Zavala et al., 2024).

Salud física: La salud física engloba el estado general del cuerpo y su funcionamiento, considerando la presencia de enfermedades o condiciones médicas diagnosticadas. Esta dimensión incluye indicadores como la anemia y la presión arterial alta, los cuales pueden comprometer el equilibrio metabólico y aumentar el riesgo de complicaciones. Se relaciona con el adecuado funcionamiento de los sistemas corporales y representa una condición fundamental para la prevención de enfermedades no transmisibles como la diabetes mellitus (Ciria, 2021).

Salud mental: La salud mental hace referencia al estado emocional y psicológico de una persona, y su capacidad para enfrentar el estrés, mantener relaciones saludables y tomar decisiones. La presencia de trastornos como la depresión puede afectar negativamente los hábitos de vida, la adherencia al tratamiento y la percepción del autocuidado. Esta dimensión resulta clave en la comprensión integral del bienestar de un individuo, dado que su alteración puede influir significativamente en el desarrollo y manejo de enfermedades crónicas (Prado et al., 2023).

Datos sociodemográficos: Los datos sociodemográficos corresponden a características estructurales de los individuos que permiten contextualizar su situación social, económica y educativa. Esta dimensión incluye variables como la edad, el lugar de residencia, el nivel educativo y el acceso a servicios de salud. Estos factores inciden en las oportunidades de prevención, diagnóstico y tratamiento de enfermedades. Su análisis es fundamental para identificar grupos vulnerables y establecer relaciones entre el entorno social y la aparición de condiciones como la diabetes (Sun et al., 2022).

Tabla 1
Operacionalización de variable

Variable	Dimensiones	Indicadores	Categoría	Tipo	Escala
Factores asociados con la diabetes	Salud reproductiva y materna	Actualmente embarazada	Si / No	Cualitativo	Nominal
		Número de hijos	Cantidad	Cuantitativo	Discreta
		Uso anticonceptivo	Si / No	Cualitativo	Nominal
	Salud física	Anemia	Si / No	Cualitativo	Nominal
		Diagnóstico de presión alta	Si / No	Cualitativo	Nominal
	Salud mental	Depresión	Si / No	Cualitativo	Nominal
	Características sociodemográficas	Edad	Años	Cuantitativo	Continua
		Área de residencia	Urbana / Rural	Cualitativo	Nominal
		Cuenta con Seguro de salud	Si / No	Cualitativo	Nominal
		Educación superior	Si / No	Cualitativo	Nominal
Presencia de Diabetes	Diabetes	Diagnóstico de diabetes	Si / No	Cualitativo	Nominal

Factores asociados con la diabetes:

Definición conceptual: Los factores asociados con la diabetes comprenden condiciones clínicas, reproductivas, mentales y sociodemográficas que influyen en la probabilidad de desarrollar esta enfermedad. Incluyen aspectos como embarazo, uso de anticonceptivos, anemia, hipertensión, depresión, edad, residencia, nivel educativo y acceso a seguro de salud (Heredia y Gallegos, 2022).

Definición operacional: Se entienden como características observables que permiten identificar condiciones de riesgo presentes en la población. Estas se agrupan en dimensiones específicas y se recogen mediante preguntas directas o antecedentes clínicos, facilitando su análisis en estudios cuantitativos.

Diabetes:

Definición conceptual: La diabetes es una enfermedad metabólica caracterizada por niveles elevados de glucosa en sangre, producto de alteraciones en la secreción o acción de la insulina. Esto constituye un grupo de trastornos que comparten la hiperglucemia como manifestación común (Thong et al., 2020).

Definición operacional: Se mide a partir de antecedentes médicos o resultados clínicos que indican la presencia o ausencia de un diagnóstico previo. Esta variable permite establecer asociaciones con otros factores de interés en investigaciones epidemiológicas.

DISEÑO METODOLÓGICO

- **Diseño de Contrastación de Hipótesis**

En la presente investigación se aplicó un enfoque cuantitativo, ya que se recopiló y analizó datos numéricos con el fin de examinar la relación entre diversos factores y la presencia de diabetes en mujeres en edad reproductiva en la Macro Región Norte del Perú, utilizando la base de datos de la Encuesta Demográfica y de Salud Familiar (ENDES) 2024. Según lo planteado por Arias y Covinos (2021), este enfoque se caracteriza por el uso de mediciones objetivas y técnicas estadísticas que permiten identificar patrones, asociaciones y relaciones entre variables, posibilitando así la obtención de resultados precisos y generalizables.

El estudio fue de tipo básico, pues su finalidad fue ampliar el conocimiento sobre los factores asociados a la diabetes sin que ello implique una intervención directa en la población estudiada. En este sentido, y de acuerdo con Hernández y Mendoza (2018), la investigación básica se orienta hacia la comprensión teórica de fenómenos, sin buscar aplicaciones prácticas inmediatas. Asimismo, el nivel de investigación fue explicativo, ya que se pretende identificar y comprender las relaciones causales entre las variables consideradas, permitiendo establecer posibles factores que inciden significativamente en la aparición de la diabetes en el grupo poblacional analizado.

El diseño metodológico fue no experimental, debido a que no se manipuló las variables de estudio, sino que estas fueron observadas tal como se presentan en su contexto natural. Conforme a lo señalado por Da Cunha y Andrade (2021), en este tipo de diseño no se interviene sobre las condiciones existentes, lo cual es adecuado para analizar fenómenos sociales o de salud en escenarios reales. Finalmente, se empleó un diseño transversal, ya que la información fue recolectada en un solo punto temporal, lo que permitió describir y analizar el comportamiento de las variables durante el año 2024.

- **Población y Muestra**

La población de este estudio estuvo compuesta por mujeres en edad reproductiva, comprendidas entre los 15 y 49 años, según los criterios establecidos por el Instituto Nacional de Estadística e Informática (INEI, 2025) en la Encuesta Demográfica y de Salud Familiar 2024. Esta población estará ubicada en la Macro Región Norte del Perú, la cual incluye los departamentos de Amazonas, Cajamarca, La Libertad, Lambayeque, Piura, San Martín y Tumbes. De acuerdo con Arias y Covinos (2021), la población en una investigación se define como el conjunto total de elementos que comparten características comunes y sobre los cuales se pretende obtener información sistematizada. En este caso, se buscará analizar la asociación entre diversos factores y la presencia de diabetes en dicho grupo poblacional.

Para la recolección de datos, se trabajó con una muestra compuesta por 2,421 mujeres en edad reproductiva que residen en la mencionada macro región. Esta muestra fue tomada directamente de la base de datos de la ENDES 2024, la cual ofrece información representativa a nivel regional y nacional. La selección de la muestra respondió a criterios metodológicos establecidos por el INEI, garantizando así la representatividad de los datos y la pertinencia del análisis estadístico dentro del enfoque cuantitativo planteado.

- **Técnicas, Instrumentos, Equipos y Materiales**

Para el análisis de los datos, se utilizaron tablas de frecuencias con el objetivo de organizar y describir de manera clara la distribución de las variables incluidas en el estudio, facilitando la identificación de patrones generales en la población de mujeres en edad reproductiva. Además, se aplicó métodos de balanceo para equilibrar las categorías de la variable dependiente (presencia o ausencia de diabetes), con el fin de evitar sesgos en los resultados del modelo estadístico. Posteriormente, se realizó una regresión logística binaria utilizando el método de selección hacia atrás de Wald, lo cual permitió identificar los factores asociados de manera significativa a la probabilidad de desarrollar diabetes.

RESULTADOS

Tabla 2

Dimensión salud reproductiva y materna en mujeres en edad reproductiva en la Macro Región Norte del Perú según ENDES, 2024

Variables	Frecuencia	Porcentaje
Embarazada		
No	2363	97.6
Si	58	2.4
Hijos		
Menos de 2 hijos	857	35.4
2 hijos o mas	1564	64.6
Anticonceptivos		
No usa	881	36.4
Si usa	1540	63.6

La Tabla 2 presenta la distribución de las variables relacionadas con la salud reproductiva y materna de las mujeres en edad reproductiva incluidas en el estudio. Los resultados evidencian que solo una proporción mínima de las encuestadas se encontraba embarazada al momento de la encuesta (2.4%), lo cual refleja que la mayoría no estaba atravesando un proceso gestacional. Asimismo, se observa que una mayor proporción de mujeres tenía dos hijos o más (64.6%), lo que sugiere patrones reproductivos asociados a mayor paridad. Respecto al uso de métodos anticonceptivos, el 63.6% reportó utilizarlos, lo que indica una predominancia de prácticas de planificación familiar. En conjunto, estos datos permiten caracterizar el perfil reproductivo de la población, proporcionando un contexto necesario para comprender su posible relación con la presencia de diabetes.

Tabla 3

Dimensión salud física en mujeres en edad reproductiva en la Macro Región Norte del Perú según ENDES, 2024

Variables	Frecuencia	Porcentaje
Anemia		

Sin anemia	1840	76.0
Con anemia	581	24.0
Presión alta		
No	2326	96.1
Si	95	3.9

La Tabla 3 muestra los indicadores relacionados con la salud física de las mujeres analizadas. Se observa que el 24% presenta anemia, mientras que la mayoría (76%) no reporta esta condición, lo que refleja un estado general favorable en cuanto al nivel de hemoglobina. Por otro lado, la prevalencia de diagnóstico de presión alta es baja (3.9%), aunque este factor es clínicamente relevante al ser un conocido predictor de enfermedades metabólicas, incluida la diabetes. El predominio de mujeres sin hipertensión (96.1%) es consistente con la composición general de la muestra, aunque su presencia, pese a ser minoritaria, podría tener un papel importante en el análisis multivariado posterior. En síntesis, esta dimensión contribuye a identificar condiciones físicas que pueden influir en el riesgo de diabetes en esta población.

Tabla 4
Dimensión salud mental en mujeres en edad reproductiva en la Macro Región Norte del Perú según ENDES, 2024

Variables	Frecuencia	Porcentaje
Depresión		
No	570	23.5
Si	1851	76.5

La Tabla 4 presenta la distribución de la variable depresión en la población evaluada. Se aprecia que el 76.5% de las mujeres reportó síntomas de depresión, mientras que el 23.5% indicó no presentar dicha condición. Esta elevada proporción sugiere que la salud mental constituye un aspecto crítico dentro de la población femenina en edad reproductiva de la Macro Región Norte del Perú. Aunque la presencia de depresión no implica necesariamente un impacto directo sobre la diabetes, la literatura señala que los factores emocionales pueden

influir en la adherencia a prácticas saludables y en el riesgo de desarrollar enfermedades crónicas. Por tanto, esta variable se incorpora como un potencial factor asociado que requiere análisis detallado dentro del modelo explicativo multivariado.

Tabla 5

Dimensión características sociodemográficas en mujeres en edad reproductiva en la Macro Región Norte del Perú según ENDES, 2024

Variablen	Frecuencia	Porcentaje
Edad		
35 años o mas	846	34.9
Menos de 35 años	1575	65.1
Residencia		
Rural	850	35.1
Urbano	1571	64.9
Seguro de salud		
No	122	5.0
Si	2299	95.0
Educación superior		
No	1759	72.7
Si	662	27.3

La Tabla 5 detalla las principales características sociodemográficas de la población analizada. Se observa que la mayoría de mujeres tiene menos de 35 años (65.1%), lo que refleja una población predominantemente joven. Asimismo, un porcentaje mayoritario reside en áreas urbanas (64.9%), lo cual podría influir en la exposición a estilos de vida asociados con menor actividad física y mayor riesgo metabólico. Respecto al acceso a servicios de salud, el 95% cuenta con seguro, lo que constituye un indicador favorable para la detección temprana de enfermedades. Finalmente, el 27.3% posee educación superior, lo cual refleja avances en el nivel educativo, aunque aún persiste una proporción considerable sin formación avanzada. Estas características sociodemográficas permiten contextualizar los factores que podrían relacionarse con la presencia de diabetes.

Tabla 6

Prueba de chi-cuadrado dimensiones y la variable Diabetes en mujeres en edad reproductiva en la Macro Región Norte del Perú según ENDES, 2024

Variables	Diabetes				Chi-Cuadrado	
	No		Si			
	f	%	f	%		
Salud reproductiva y materna					Valor	p
Embarazada					2.252	0.133
No	2311	95.5	52	2.1		
Si	55	2.3	3	0.1		
Hijos					2.489	0.115
Menos de 2 hijos	832	34.4	25	1.0		
2 hijos o mas	1534	63.4	30	1.2		
Anticonceptivos					0.000	0.997
No usa	861	35.6	20	0.8		
Si usa	1505	62.2	35	1.4		
Salud fisica						
Anemia					0.147	0.702
Sin anemia	1797	74.2	43	1.8		
Con anemia	569	23.5	12	0.5		
Presión alta					23.100	0.000
No	2280	94.2	46	1.9		
Si	86	3.6	9	0.4		
Salud mental						
Depresión					1.696	0.193
No	553	22.8	17	0.7		
Si	1813	74.9	38	1.6		
Características sociodemográficas						
Edad					17.879	0.000
35 años o mas	812	33.5	34	1.4		
Menos de 35 años	1554	64.2	21	0.9		
Residencia					3.252	0.071
Rural	837	34.6	13	0.5		

Urbano	1529	63.2	42	1.7		
Seguro de salud					0.231	0.630
No	120	5.0	2	0.1		
Si	2246	92.8	53	2.2		
Educación superior					0.389	0.533
No	1717	70.9	42	1.7		
Si	649	26.8	13	0.5		
Total	2366	97.7	55	2.3		

La Tabla 6 presenta los resultados de la prueba de chi-cuadrado aplicada para evaluar la asociación entre cada dimensión analizada y la presencia de diabetes. Los hallazgos indican que, en general, la mayoría de variables no muestra asociaciones significativas con el diagnóstico de diabetes, con excepción de presión alta ($p < 0.001$) y edad ($p < 0.001$), las cuales evidencian diferencias estadísticamente significativas entre mujeres con y sin diabetes. Esto sugiere que la hipertensión y tener 35 años o más podrían desempeñar un rol importante como factores asociados. Las demás variables, incluidas aquellas relacionadas con medidas reproductivas, salud mental y algunas características sociodemográficas, no muestran relaciones significativas en esta prueba. Estos resultados sirven como filtro inicial previo al análisis multivariado mediante regresión logística.

Tabla 7

Modelo logístico binario de los factores asociados con la diabetes en mujeres en edad reproductiva en la Macro Región Norte del Perú según ENDES, 2024

Variables	B	Std Err	Sig.	Odds	Inf.	Sup.
Embarazada (Si)	1.715	0.247	<0.001	5.560	3.463	9.154
Hijos (2 hijos o más)	0.238	0.106	0.024	1.269	1.032	1.564
Anticonceptivos (Si usa)	0.233	0.097	0.016	1.263	1.044	1.530
Presión alta (Si)	1.272	0.173	<0.001	3.571	2.562	5.062
Depresión (Si)	-0.295	0.098	0.002	0.744	0.613	0.903
Edad (Menor a 35 años)	-1.199	0.100	<0.001	0.301	0.247	0.366

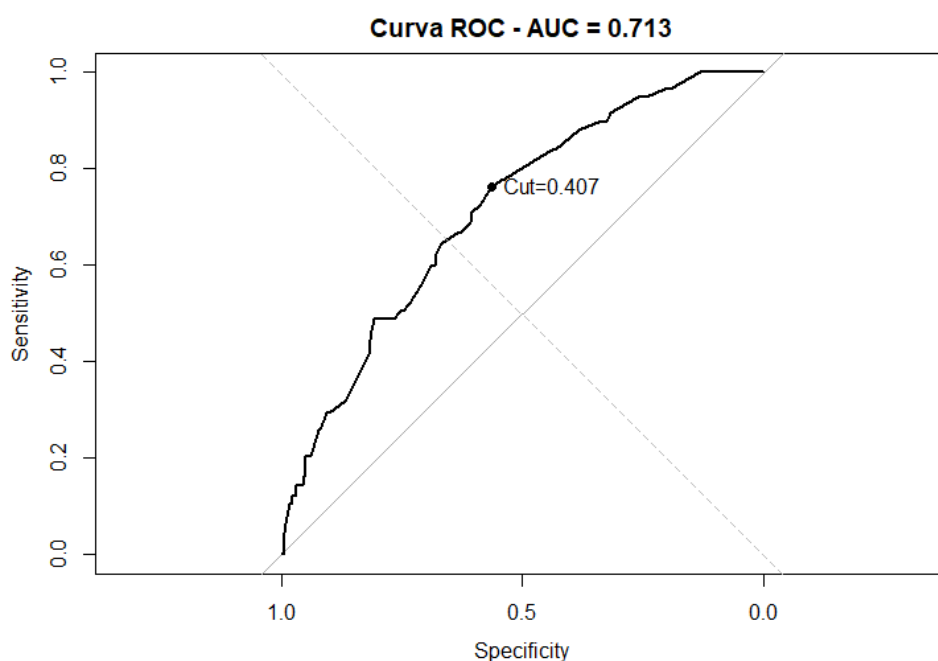
Residencia (Urbano)	0.597	0.100	<0.001	1.818	1.495	2.215
Educación superior (Si)	-0.351	0.105	<0.001	0.703	0.571	0.865
Constante	-0.009	0.146	0.946			

Selección de variables mediante el método de AIC.

La Tabla 7 presenta los resultados del modelo de regresión logística binaria elaborado para identificar los factores asociados con la diabetes en mujeres en edad reproductiva. Es importante señalar que este modelo se construyó empleando una base de datos balanceada mediante un método híbrido de sobremuestreo y submuestreo (ROSE), debido al marcado desbalance inicial de la variable dependiente. Esto permitió corregir la desproporción entre casos con y sin diabetes, garantizando estimaciones más estables y evitando sesgos. Los resultados evidencian que el embarazo (OR = 5.560) y la presión alta (OR = 3.571) se asocian significativamente con mayor probabilidad de diabetes. En contraste, tener menos de 35 años, poseer educación superior y presentar depresión se relacionan con una reducción del riesgo. El modelo final revela un conjunto multifactorial de determinantes.

Figura 1

Curva de ROC del modelo logístico binario de los factores asociados con la diabetes en mujeres en edad reproductiva en la Macro Región Norte del Perú según ENDES, 2024



La Figura 1 muestra la curva ROC correspondiente al modelo logístico estimado sobre la base balanceada mediante el método híbrido ROSE. Esta figura permite evaluar la capacidad discriminativa del modelo para diferenciar entre mujeres con y sin diagnóstico de diabetes. El área bajo la curva ($AUC = 0.713$) indica un desempeño moderado, coherente con modelos epidemiológicos empleados en estudios poblacionales. Asimismo, se identifica un punto de corte óptimo de 0.407 según el índice de Youden, el cual maximiza la sensibilidad y especificidad del modelo. El comportamiento de la curva revela que el balanceo aplicado mejora la capacidad del modelo para detectar casos positivos, reforzando la pertinencia del ajuste realizado para analizar factores asociados en una muestra originalmente desbalanceada.

Tabla 8

Matriz de confusión del modelo logístico binario de los factores asociados con la diabetes en mujeres en edad reproductiva en la Macro Región Norte del Perú según ENDES, 2024

Predicción	Referencia	
	No	Si
No	690	285
Si	533	913

La Tabla 8 presenta la matriz de confusión generada a partir del punto de corte óptimo definido por el índice de Youden (0.407). Esta matriz corresponde al modelo logístico ajustado con la base balanceada híbrida. Los resultados muestran que el modelo clasifica adecuadamente a un número importante de mujeres con diabetes ($n = 913$) y sin la condición ($n = 690$), aunque también se observan falsos positivos ($n = 533$) y falsos negativos ($n = 285$). Esto sugiere que, pese a la mejora obtenida mediante el balanceo, persiste una moderada proporción de errores, especialmente en la identificación de mujeres sin diabetes. Sin embargo, el modelo mantiene un adecuado rendimiento para fines de tamizaje, donde la prioridad suele ser maximizar la detección de casos reales.

Tabla 9

Indicadores de rendimiento del modelo logístico binario de los factores asociados con la diabetes en mujeres en edad reproductiva en la Macro Región Norte del Perú según ENDES, 2024

AUC	Punto de corte	Sensibilidad	Especificidad	Accuracy
0.713	0.407	0.762	0.564	0.662

La Tabla 9 resume los principales indicadores de desempeño del modelo logístico estimado sobre la base balanceada mediante el método híbrido ROSE. El AUC de 0.713 confirma una capacidad discriminativa moderada. Asimismo, la sensibilidad de 0.762 demuestra que el modelo detecta correctamente a más del 76% de mujeres con diabetes, lo cual es adecuado para fines epidemiológicos. La especificidad de 0.564 refleja una capacidad limitada para identificar a quienes no presentan la enfermedad, situación común en modelos ajustados sobre datos balanceados artificialmente. La precisión de 0.631 y la exactitud global de 0.662 indican un rendimiento aceptable. En conjunto, estos indicadores muestran que el balanceo permitió mejorar la identificación de casos positivos sin comprometer excesivamente la estabilidad del modelo.

DISCUSIÓN

De acuerdo al objetivo general se identificó múltiples factores provenientes de distintas dimensiones que se asocian significativamente con la presencia de diabetes en mujeres en edad reproductiva. Los resultados del modelo logístico, ajustado sobre una base balanceada mediante un enfoque híbrido, evidenciaron que variables como el embarazo y la presión alta constituyen los predictores más relevantes, al incrementar considerablemente la probabilidad de presentar diabetes. Asimismo, la residencia urbana también se relacionó con una mayor probabilidad de diagnóstico, lo cual refleja la influencia del entorno en los riesgos metabólicos. En contraste, la edad menor de 35 años, la educación superior y la presencia de depresión se comportaron como factores protectores en el modelo. Estos hallazgos demuestran que la diabetes en este grupo poblacional responde a un conjunto heterogéneo de influencias biológicas, reproductivas, clínicas y sociodemográficas.

Al contrastar estos resultados con la evidencia previa, se observan coincidencias importantes. La asociación entre factores clínicos como la presión arterial y un mayor riesgo de diabetes concuerda con los hallazgos de Rojo et al. (2020), quienes identificaron que las personas con alteraciones metabólicas o hipertensión presentan mayor probabilidad de desarrollar diabetes tipo 2. Asimismo, el riesgo incrementado en entornos urbanos se asemeja a lo reportado por Sun et al. (2022), quienes evidenciaron mayor prevalencia de diabetes en zonas urbanas y países de ingresos medios. Por otra parte, la relevancia de factores reproductivos, como el embarazo, coincide con los planteamientos de Bustamante et al. (2023), Prado et al. (2023) y Seclen (2024), quienes destacan que las condiciones gestacionales pueden aumentar la vulnerabilidad metabólica femenina. No obstante, algunos resultados difieren parcialmente de estudios clínicos como el de Tamata (2021), donde factores emocionales y dislipidemias fueron predominantes, lo que sugiere que las asociaciones pueden variar según el contexto, la metodología y las características sociodemográficas analizadas.

Considerando al primer objetivo específico, los resultados mostraron que la población analizada se caracteriza por ser predominantemente joven, con más de la mitad de mujeres menores de 35 años, una mayoría residente en zonas urbanas y con acceso amplio a seguro de salud. Además, se observó que el uso de métodos anticonceptivos y la presencia de dos hijos o más son comportamientos frecuentes, mientras que la proporción de embarazadas fue reducida. En cuanto a salud física, condiciones como la presión alta se presentaron en un porcentaje bajo, pero clínicamente relevante. Asimismo, la proporción de mujeres con síntomas de depresión fue considerablemente elevada, lo cual destaca la importancia de considerar la salud mental en esta población. Estas características configuran un perfil mixto, donde coexisten factores de riesgo y elementos de protección relacionados con enfermedades crónicas como la diabetes.

La caracterización de la población coincide parcialmente con las tendencias identificadas en estudios internacionales. Sun et al. (2022) mostraron que la diabetes presenta mayor incidencia en zonas urbanas y en países de ingresos medios, lo cual se alinea con el predominio urbano hallado en esta muestra. Sin embargo, la elevada proporción de mujeres jóvenes contrasta con los hallazgos de dicho estudio, donde la prevalencia aumenta en grupos de mayor edad. En el ámbito nacional y regional, investigaciones como las de Prado et al. (2023) y Seclen (2024) destacan la relevancia de factores como edad materna avanzada y sobrepeso en contextos gestacionales, lo que difiere del perfil de esta población, donde predominan mujeres menores de 35 años. No obstante, la presencia elevada de síntomas depresivos guarda relación con factores emocionales asociados al descontrol metabólico descritos en estudios clínicos, lo que sugiere que las características observadas en esta población podrían influir en patrones futuros de riesgo.

Para el segundo objetivo específico, los resultados evidenciaron que variables sociodemográficas como la edad y el área de residencia presentan relaciones relevantes con el diagnóstico de diabetes. La prueba de chi-cuadrado mostró asociaciones significativas con la

edad, indicando mayor presencia de diabetes en mujeres de 35 años o más, lo cual concuerda con el patrón biológico de aumento del riesgo metabólico con la edad. Asimismo, aunque la residencia urbana no mostró una relación significativa en el análisis bivariado, el modelo multivariado reveló que vivir en zonas urbanas incrementa la probabilidad de diabetes, incluso tras el balanceo híbrido aplicado al modelo. Esto sugiere la influencia de estilos de vida urbanos, caracterizados por menor actividad física y mayor exposición a factores de riesgo metabólico. Por tanto, las características sociodemográficas desempeñan un rol relevante en la comprensión del perfil epidemiológico.

Estos hallazgos guardan similitudes con lo reportado por Sun et al. (2022), quienes identificaron mayores prevalencias de diabetes en zonas urbanas y en países de ingresos medios, lo cual refuerza la asociación encontrada entre residencia urbana y mayor probabilidad de diabetes. Asimismo, el incremento del riesgo en mujeres mayores coincide con la evidencia expuesta por Rojo et al. (2020), quienes mostraron mayor incidencia de diabetes en adultos con factores clínicos acumulados, como obesidad y sedentarismo. Sin embargo, los resultados difieren parcialmente de los estudios de Prado et al. (2023) y Seclen (2024), donde la diabetes gestacional —más frecuente en mujeres mayores o con antecedentes familiares— no sigue el mismo patrón observado en la población general de este estudio, compuesta mayoritariamente por mujeres jóvenes. No obstante, todos los antecedentes coinciden en destacar la edad y los entornos urbanos como factores determinantes en la evolución de la diabetes.

Finalmente, para el tercer objetivo específico, el modelo de regresión logística elaborado sobre la base balanceada mediante un método híbrido permitió identificar un conjunto significativo de factores asociados con la diabetes. Entre ellos, el embarazo y la presión alta mostraron las asociaciones positivas más fuertes, sugiriendo que condiciones fisiológicas y clínicas particulares aumentan la probabilidad de diabetes. Por otro lado, ser menor de 35 años, tener educación superior y presentar depresión se relacionaron con menor

riesgo en la muestra balanceada. La residencia urbana también emergió como un factor asociado, reforzando la influencia del contexto geográfico en la salud metabólica. El rendimiento del modelo, representado por un AUC de 0.713, junto con una sensibilidad de 0.762, evidencia una capacidad moderada para discriminar casos, adecuada para fines epidemiológicos en poblaciones amplias.

Los factores identificados muestran consistencias importantes con investigaciones previas. El papel de la presión arterial elevada se alinea con los resultados de Rojo et al. (2020), quienes reportaron mayor incidencia de diabetes en individuos con alteraciones clínicas previas. Asimismo, la relevancia de factores reproductivos concuerda parcialmente con los hallazgos de Bustamante et al. (2023) y Prado et al. (2023), quienes identificaron que la diabetes gestacional y condiciones asociadas al embarazo representan riesgos clave en mujeres en edad fértil. Sin embargo, el menor riesgo asociado a la depresión observado en este modelo difiere de tendencias clínicas reportadas en estudios como los de Tamata (2021), en los cuales problemas metabólicos suelen coexistir con alteraciones emocionales. Finalmente, el mayor riesgo en entornos urbanos es consistente con Sun et al. (2022), lo que reafirma la influencia de los determinantes contextuales en la ocurrencia de diabetes.

CONCLUSIONES

El estudio permitió determinar que diversos factores pertenecientes a las dimensiones de salud reproductiva y materna, salud física, salud mental y características demográficas presentan asociación significativa con la diabetes en mujeres en edad reproductiva de la Macro Región Norte del Perú. El modelo de regresión logística ajustado sobre una base balanceada demostró que el embarazo, la presión alta y la residencia urbana incrementan la probabilidad de diabetes, mientras que ser menor de 35 años, poseer educación superior y presentar depresión reducen este riesgo. En conjunto, los resultados evidencian que la diabetes en esta población responde a una interacción multifactorial, destacándose la importancia de analizar simultáneamente condiciones biológicas, clínicas y sociodemográficas para comprender mejor esta problemática de salud pública.

La caracterización de la población evidenció que la mayoría de mujeres es joven, reside en zonas urbanas, tiene acceso a seguro de salud y presenta un nivel educativo predominantemente básico. Asimismo, se identificó una alta proporción de uso de métodos anticonceptivos y de mujeres con dos o más hijos, mientras que el porcentaje de embarazadas fue reducido. En cuanto a salud física, la anemia fue frecuente, pero la presencia de presión alta fue baja; a su vez, la depresión destacó como un problema frecuente. Este perfil demográfico, reproductivo y clínico constituye la base para comprender los factores asociados a la diabetes, mostrando una población heterogénea con condiciones que pueden incrementar o reducir su riesgo metabólico.

El análisis bivariado y multivariado permitió identificar que la edad y la residencia urbana son las condiciones sociodemográficas con mayor relación con el diagnóstico de diabetes. Se evidenció que las mujeres de 35 años o más presentan mayor probabilidad de la enfermedad, reafirmando el rol que cumple el envejecimiento en el deterioro metabólico. Asimismo, aunque la residencia urbana no fue significativa en pruebas bivariadas, sí mostró

asociación en el modelo final, sugiriendo la influencia de estilos de vida urbanos sobre el riesgo de diabetes. Estas relaciones demuestran que los factores sociodemográficos influyen de forma importante en la presencia de diabetes y deben ser considerados conjuntamente con características clínicas y reproductivas.

El modelo logístico elaborado sobre una base balanceada permitió identificar los factores con mayor poder explicativo sobre la presencia de diabetes. El embarazo y la presión alta fueron los predictores de mayor riesgo, mientras que la educación superior, la edad menor a 35 años y la depresión se comportaron como factores protectores. Además, la residencia urbana incrementó la probabilidad de diabetes. El modelo alcanzó un AUC de 0.713, indicando una capacidad moderada de discriminación. Estos resultados confirman que el balanceo híbrido y la selección mediante AIC permitieron obtener un modelo estable y coherente con la naturaleza multicausal de la diabetes en mujeres en edad reproductiva.

RECOMENDACIONES

Se recomienda evaluar otros métodos de balanceo para datos desbalanceados, como SMOTE, ADASYN o undersampling sistemático, y comparar su impacto sobre la estabilidad del modelo. Esto permitiría determinar cuál técnica ofrece mejores niveles de discriminación y precisión en estudios epidemiológicos con baja prevalencia.

Se sugiere complementar el análisis con modelos como regresión LASSO, Random Forest o Gradient Boosting para comparar los factores seleccionados y evaluar si la importancia de las variables se mantiene bajo distintos enfoques estadísticos, mejorando la validez del estudio.

Para futuras investigaciones, se propone analizar posibles interacciones entre variables, especialmente entre factores clínicos y sociodemográficos, y explorar ajustes no lineales mediante splines o modelos aditivos. Esto permitiría identificar relaciones más complejas que podrían estar presentes en la aparición de diabetes.

Se recomienda aplicar validación cruzada k-fold al modelo logístico y reportar métricas como F1-score, curva PR y Brier Score. Estas medidas permitirán evaluar con mayor profundidad la estabilidad del modelo y su capacidad para generalizar a otras poblaciones semejantes.

REFERENCIAS

- Arias Gonzales, J. L., & Covinos Gallardo, M. (2021). *Diseño y metodología de la investigacion* (1.^a ed.). Enfoques Consulting E.I.R.L. https://gc.scalahed.com/recursos/files/r161r/w26022w/Arias_S2.pdf
- Azriel, S., Camaño, I., Montañez, D., Navas, A., & García-Burguillo, A. (2008). Un caso especial de diabetes mellitus gestacional. *Endocrinología y Nutrición*, 55(9), 433-435. [https://doi.org/10.1016/S1575-0922\(08\)75080-0](https://doi.org/10.1016/S1575-0922(08)75080-0)
- Bustamante Saavedra, C., Valencia Ibieta, C., Salazar Vergara, E., Canelo Nigoevic, C., & Camiruaga Jurgensen, I. (2023). Tasa de egresos hospitalarios de mujeres en edad fértil por Diabetes Mellitus Gestacional y Diabetes Mellitus Pregestacional entre 2018-2021, Chile. *Revista Confluencia*, 6(1), 49-53. <https://doi.org/10.52611/confluencia.num1.2023.895>
- Ciria Mallada, M. C. (2021). Diabetes desde la perspectiva de género. *DIABETES PRÁCTICA*, 40-48. <https://doi.org/10.26322/2013.7923.1505400610.03>
- Da Cunha E Silva, L. C., & Andrade Romero, J. F. (2021). Metodología híbrida para la estimación del nivel de llenado en un molino de bolas. *Revista Iberoamericana de Automática e Informática industrial*, 19(2), 210-220. <https://doi.org/10.4995/riai.2021.13064>
- Dávila Flores, J. X., Montenegro Morán, E. E., Macías Gaytán, Á. M., & Tayupanda Martinez, J. L. (2023). La diabetes mellitus y diabetes gestacional, en adolescente, en el mundo y en el Ecuador, manejo, prevención, tratamiento y mortalidad. *RECIMUNDO*, 7(2), 33-48. [https://doi.org/10.26820/recimundo/7.\(2\).jun.2023.33-48](https://doi.org/10.26820/recimundo/7.(2).jun.2023.33-48)
- Díaz Vera, J. P., Ruiz Ramírez, A. K., & Egüez Cevallos, C. (2021). Impacto de las TIC: Desafíos y oportunidades de la Educación Superior frente al COVID-19. *Revista Científica UISRAEL*, 8(2), 113-134. <https://doi.org/10.35290/rcui.v8n2.2021.448>

- Heredia, M., & Gallegos Cabriales, E. C. (2022). Riesgo de diabetes mellitus tipo 2 y sus determinantes. *Enfermería Global*, 21(1), 179-202. <https://doi.org/10.6018/eglobal.482971>
- Hernández Sampieri, R., & Mendoza Torres, C. P. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta* (First edition). McGraw-Hill Education.
- Instituto Nacional de Estadística e Informatica. (2025). *Encuesta Demográfica y de Salud Familiar 2024* (p. 369) [Informe]. INEI. https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib2016/libro.pdf
- Martínez Pérez, J. A., & Pérez Martín, P. S. (2023). La curva ROC. *Medicina de Familia. SEMERGEN*, 49(1), 101821. <https://doi.org/10.1016/j.semerg.2022.101821>
- Naveiro Fuentes, M., Jiménez-Moleón, J. J., Olmedo Requena, R., Amezcua Prieto, C., Bueno Cabanillas, A., & Mozas Moreno, J. (2015). Resultados perinatales en función de 3 criterios diagnósticos de diabetes gestacional. *Clínica e Investigación en Ginecología y Obstetricia*, 42(2), 66-71. <https://doi.org/10.1016/j.gine.2013.12.008>
- Ordoñez Sedamanos, R. A., & Gallardo Heras, L. M. (2024). Nivel Cognitivo sobre Ovario Poliquístico de Mujeres en Edad Reproductiva. *Ciencia Latina Revista Científica Multidisciplinar*, 8(3), 7818-7830. https://doi.org/10.37811/cl_rcm.v8i3.11971
- Orraca, O., León, M., Orraca, M., Lardoezt, R., & Miló, C. (2021). Modelo de regresión logística binaria para el cálculo de riesgo de asma en Pinar del Río. *Revista de Ciencias Médicas de Pinar del Río*, 25(5). <https://www.medigraphic.com/cgi-bin/new/resumenI.cgi?IDARTICULO=112624>
- Ortega, E., Ochoa, C., & Molina, M. (2022). *Regresión logística binaria simple*. 18(11), 9.

- Peña, E. (2022). *Valoración del riesgo de padecer diabetes mellitus tipo 2 en población joven del distrito de Lambayeque, Enero-Marzo 2020* [Tesis de pregrado, Universidad de San Martín de Porres]. <https://hdl.handle.net/20.500.12727/9719>
- Pérez Unanua, M. P., López Simarro, F., Novillo López, C. I., Olivares Loro, A. G., & Yáñez Freire, S. (2024). Diabetes y mujer, ¿por qué somos diferentes? *Medicina de Familia. SEMERGEN*, 50(2), 102138. <https://doi.org/10.1016/j.semerg.2023.102138>
- Pomares, L., Gómez Martín, C., Maccio, N., Geller, M., Bertona, C., Migueles, J., Arinovich, B., Schindler, A., Garrido, N., Larreburo, S., Blanco, M., De Loredó, S., & González, L. (2014). TRASTORNOS DE LA FERTILIDAD EN LA MUJER CON DIABETES TIPO 1 Y 2. *Revista de la Sociedad Argentina de Diabetes*, 48(2), 51. <https://doi.org/10.47196/diab.v48i2.185>
- Prado Herrera, M. F., Guerrero-Aguilar, A. S., Alatrística-Gutiérrez-Vda. Bambarén, M. D. S., Vela-Ruiz, J. M., & Lama-Morales, R. A. (2023). Diabetes gestacional: Impacto de los factores de riesgo en Latinoamérica. *Revista Peruana de Investigación Materno Perinatal*, 12(1), 33-43. <https://doi.org/10.33421/inmp.2023317>
- Rojo-Martínez, G., Valdés, S., Soriguer, F., Vendrell, J., Urrutia, I., Pérez, V., Ortega, E., Ocón, P., Montanya, E., Menéndez, E., Lago-Sampedro, A., González- Frutos, T., Gomis, R., Goday, A., García-Serrano, S., García-Escobar, E., Galán-García, J. L., Castell, C., Badía-Guillén, R., ... Calle-Pascual, A. (2020). Incidence of diabetes mellitus in Spain as results of the nation-wide cohort di@bet.es study. *Scientific Reports*, 10(1), 2765. <https://doi.org/10.1038/s41598-020-59643-7>
- Seclen, S. (2024). Importancia de la salud de la mujer en Latinoamérica (Perú). *Revista de la sociedad Argentina de Diabetes*, 58(35). <https://revistasad.com/index.php/diabetes/article/view/1013>

- Soriguer, F., Goday, A., Bosch-Comas, A., Bordiú, E., Calle-Pascual, A., Carmena, R., Casamitjana, R., Castaño, L., Castell, C., Catalá, M., Delgado, E., Franch, J., Gaztambide, S., Girbés, J., Gomis, R., Gutiérrez, G., López-Alba, A., Martínez-Larrad, M. T., Menéndez, E., ... Vendrell, J. (2012). Prevalence of diabetes mellitus and impaired glucose regulation in Spain: The Di@bet.es Study. *Diabetologia*, 55(1), 88-93. <https://doi.org/10.1007/s00125-011-2336-9>
- Sun, H., Saeedi, P., Karuranga, S., Pinkepank, M., Ogurtsova, K., Duncan, B. B., Stein, C., Basit, A., Chan, J. C. N., Mbanya, J. C., Pavkov, M. E., Ramachandaran, A., Wild, S. H., James, S., Herman, W. H., Zhang, P., Bommer, C., Kuo, S., Boyko, E. J., & Magliano, D. J. (2022). IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. *Diabetes Research and Clinical Practice*, 183, 109119. <https://doi.org/10.1016/j.diabres.2021.109119>
- Tamata Córdova, N. I. (2021). FACTORES DE RIESGO, PREVALENCIA DE DIABETES MELLITUS TIPO II EN MUJERES ADULTAS ATENDIDAS EN EL CONSULTORIO DE ENDOCRINOLOGIA DEL HOSPITAL GUILLERMO DIAZ DE LA VEGA. *Revista de Investigaciones*, 10(2), 99-114. <https://doi.org/10.26788/riepg.v10i2.2575>
- Thong, E. P., Codner, E., Laven, J. S. E., & Teede, H. (2020). Diabetes: A metabolic and reproductive disorder in women. *The Lancet Diabetes & Endocrinology*, 8(2), 134-149. [https://doi.org/10.1016/S2213-8587\(19\)30345-6](https://doi.org/10.1016/S2213-8587(19)30345-6)
- Victorio-Huamán, L., Rivera-Felix, L., López-Sánchez, M., Alfonzo-Flores, I., & Cuya-Candela, E. (2024). Complicaciones maternas asociadas a sobrepeso y obesidad en pacientes atendidas en un hospital público peruano, 2021. *Revista Internacional de Salud Materno Fetal*, 01-06. <https://doi.org/10.47784/rismf.2024.9.4.354>

Zavala-Hoppe, A. N., Suárez-Saá, N. N., Ureta-Guerrero, J. M., & Villacreses-Cordova, L. M. (2024). Epidemiología y medidas de prevención de la diabetes gestacional en Latinoamérica y Europa. *MQRInvestigar*, 8(1), 1390-1408. <https://doi.org/10.56048/MQR20225.8.1.2024.1390-1408>

Anexos

Anexo 1. Script de ejecución en código R

```
# Script R: Análisis completo para "DiabetesUnion.xlsx"
# Requisitos: instalar paquetes si no los tienes
required <-
c("readxl", "dplyr", "ROSE", "MASS", "pROC", "caret", "broom", "knitr", "stats")
install.packages(setdiff(required, installed.packages()[,1]), dependencies
= TRUE)

library(readxl)
library(dplyr)
library(ROSE)      # para balanceo (ovun.sample con method="both" = híbrido)
library(MASS)      # stepAIC
library(pROC)      # ROC
library(caret)     # confusionMatrix
library(broom)     # tidy
library(knitr)

# 0) Cargar datos
# Ajusta la ruta si es necesario
df <- read_excel("DiabetesUnion.xlsx")

# 1) Asegurarse que variables tienen codificación correcta
# Convertir variables a factores con niveles claros (0 = No, 1 = Si) según
tu codificación
df <- df %>%
  mutate(
    V213 = factor(V213, levels = c(0,1), labels = c("No", "Si")),
    V201 = factor(V201, levels = c(0,1), labels = c("Menos de 2 hijos", "2
Hijos o mas")),
    V312 = factor(V312, levels = c(0,1), labels = c("No", "Si")),
    HA57A = factor(HA57A, levels = c(0,1), labels = c("No", "Si")),
    QS102 = factor(QS102, levels = c(0,1), labels = c("No", "Si")),
    Dep = factor(Dep, levels = c(0,1), labels = c("No", "Si")),
    Edad = factor(Edad, levels = c(0,1), labels = c("Mayor de
35", "<=35")),
    HV025 = factor(HV025, levels = c(0,1), labels = c("Rural", "Urbano")),
    QS25N = factor(QS25N, levels = c(0,1), labels = c("No", "Si")),
    QS26 = factor(QS26, levels = c(0,1), labels = c("No", "Si")),
    QS109 = factor(QS109, levels = c(0,1), labels = c("No", "Si"))
  ) %>%
  dplyr::select(V213, V201, V312, HA57A, QS102, Dep, Edad, HV025, QS25N,
QS26, QS109)

# Comprueba proporciones originales
prop_table <- prop.table(table(df$QS109))
cat("Distribución original de QS109 (No/Si):\n")
print(round(prop_table, 3))

# 2) Balanceo híbrido de QS109 (ROSE::ovun.sample method = "both")
# ovun.sample con method="both" realiza sobremuestreo + submuestreo
(balanceo híbrido)
set.seed(123)
balanced <- ovun.sample(QS109 ~ ., data = df, method = "both", N =
nrow(df), p = 0.5)$data
cat("\nDistribución después de balanceo híbrido:\n")
print(round(prop.table(table(balanced$QS109)), 3))
```

```

# 3) Preparar formula para regresión logística (usar las descripciones en
tablas de salida)
# Excluir QS109 de predictores
predictors <-
c("V213", "V201", "V312", "HA57A", "QS102", "Dep", "Edad", "HV025", "QS25N", "QS26")
formula_full <- as.formula(paste("QS109 ~", paste(predictors, collapse = "
+ ")))

# 4) Ajustar modelo logístico inicial con todos los predictores
model_full <- glm(formula_full, data = balanced, family = binomial(link =
"logit"))

# 5) Selección de variables por AIC (stepwise both directions)
model_step <- stepAIC(model_full, direction = "both", trace = FALSE)

cat("\nModelo final seleccionado por AIC:\n")
print(summary(model_step))

# 6) Tabla de ORs (odds ratios) con intervalos de confianza 95%
est <- broom::tidy(model_step, exponentiate = TRUE, conf.int = TRUE)
est <- est %>%
  rename("Termino" = term,
         "OR" = estimate,
         "IC_2.5" = conf.low,
         "IC_97.5" = conf.high,
         "p_valor" = p.value) %>%
  mutate_at(vars(OR, IC_2.5, IC_97.5, p_valor), round, 3)

cat("\nTabla de odds ratios (OR) del modelo final:\n")
kable(est, digits = 3)

# 7) Resumen interpretativo (automatizado) - texto para incluir en tesis
# Generar hilo de texto breve con los load-bearing resultados
coefs <- summary(model_step)$coefficients
# construir texto resumido (ejemplo)
cat("\nResumen interpretativo sugerido (resumido):\n")
# Build sentences
for(i in 2:nrow(coefs)){
  name <- rownames(coefs)[i]
  beta <- round(coefs[i,1],3)
  se <- round(coefs[i,2],3)
  z <- round(coefs[i,3],3)
  p <- signif(coefs[i,4],3)
  cat(sprintf("- El término '%s' tiene beta = %s (SE=%s), z=%s, p=%s. Una
OR estimada = %.3f (IC95%%: %.3f - %.3f).\n",
              name, beta, se, z, p,
              exp(beta), exp(beta - 1.96*se), exp(beta + 1.96*se)))
}
cat("\nNota: sustituye las menciones de término por la 'descripción' de la
variable al incorporar al texto final.\n")

# 8) Curva ROC y punto de corte por Youden
# Probabilidades predichas
probs <- predict(model_step, type = "response")
roc_obj <- roc(balanced$QS109, probs, levels = c("No", "Si"), direction =
"<")
auc_val <- auc(roc_obj)
cat(sprintf("\nAUC (Área bajo la curva ROC): %.3f\n", auc_val))

# Punto de corte óptimo (Youden)

```

```

opt <- coords(roc_obj, "best", best.method = "youden", transpose = FALSE)
cat("\nPunto de corte óptimo según Youden:\n")
print(opt)

# Graficar ROC (si ejecutas en RStudio se mostrará)
plot(roc_obj, main = paste0("Curva ROC - AUC = ", round(auc_val,3)))
abline(a=0,b=1,lty=2,col="gray")
# marcar punto óptimo
points(opt["specificity"], opt["sensitivity"], pch = 19)
text(opt["specificity"], opt["sensitivity"], labels = paste0("Cut=",
round(opt["threshold"],3)), pos=4)

# 9) Matriz de confusión al punto de corte óptimo
cutoff <- as.numeric(opt["threshold"])
pred_class <- factor(ifelse(probs >= cutoff, "Si", "No"), levels =
c("No", "Si"))

####
cm <- confusionMatrix(pred_class, balanced$QS109, positive = "Si")
cat("\nMatriz de confusión (con el punto de corte óptimo):\n")
print(cm$table)
cat("\nMedidas (Precision = Exactitud, Sensibilidad = Recall,
Especificidad):\n")
print(cm$byClass[c("Sensitivity", "Specificity", "Precision")])
cat("\nExactitud global:\n")
print(cm$overall["Accuracy"])

# FIN del script

```

Anexo 2. Output de ejecución en código R

```
R version 4.5.2 (2025-10-31 ucrt) -- "[Not] Part in a Rumble"
Copyright (C) 2025 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64
```

R es un software libre y viene sin GARANTIA ALGUNA.
Usted puede redistribuirlo bajo ciertas circunstancias.
Escriba 'license()' o 'licence()' para detalles de distribución.

R es un proyecto colaborativo con muchos contribuyentes.
Escriba 'contributors()' para obtener más información y
'citation()' para saber cómo citar R o paquetes de R en publicaciones.

Escriba 'demo()' para demostraciones, 'help()' para el sistema on-line de ayuda,
o 'help.start()' para abrir el sistema de ayuda HTML con su navegador.
Escriba 'q()' para salir de R.

```
[Workspace loaded from ~/.RData]
```

```
> library(readxl)
> library(dplyr)
```

Adjuntando el paquete: 'dplyr'

The following objects are masked from 'package:stats':

```
filter, lag
```

The following objects are masked from 'package:base':

```
intersect, setdiff, setequal, union
> library(ROSE) # para balanceo (ovun.sample con method="both" =
híbrido)
Loaded ROSE 0.0-4
> library(MASS) # stepAIC
```

Adjuntando el paquete: 'MASS'

The following object is masked from 'package:dplyr':

```
select
> library(pROC) # ROC
Type 'citation("pROC")' for a citation.
```

Adjuntando el paquete: 'pROC'

The following objects are masked from 'package:stats':

```
cov, smooth, var
> library(caret) # confusionMatrix
Cargando paquete requerido: ggplot2
Cargando paquete requerido: lattice
> library(broom) # tidy
> library(knitr)
>
> # 0) Cargar datos
> # Ajusta la ruta si es necesario
> df <- read_excel("DiabetesUnion.xlsx")
>
```

```

> # 1) Asegurarse que variables tienen codificación correcta
> # Convertir variables a factores con niveles claros (0 = No, 1 = Si)
según tu codificación
> df <- df %>%
+   mutate(
+     V213 = factor(V213, levels = c(0,1), labels = c("No", "Si")),
+     V201 = factor(V201, levels = c(0,1), labels = c("Menos de 2 hijos", "2
Hijos o mas")),
+     V312 = factor(V312, levels = c(0,1), labels = c("No", "Si")),
+     HA57A = factor(HA57A, levels = c(0,1), labels = c("No", "Si")),
+     QS102 = factor(QS102, levels = c(0,1), labels = c("No", "Si")),
+     Dep = factor(Dep, levels = c(0,1), labels = c("No", "Si")),
+     Edad = factor(Edad, levels = c(0,1), labels = c("Mayor de
35", "<=35")),
+     HV025 = factor(HV025, levels = c(0,1), labels = c("Rural", "Urbano")),
+     QS25N = factor(QS25N, levels = c(0,1), labels = c("No", "Si")),
+     QS26 = factor(QS26, levels = c(0,1), labels = c("No", "Si")),
+     QS109 = factor(QS109, levels = c(0,1), labels = c("No", "Si"))
+   ) %>%
+   dplyr::select(V213, V201, V312, HA57A, QS102, Dep, Edad, HV025, QS25N,
QS26, QS109)
>
>
> # Comprueba proporciones originales
> prop_table <- prop.table(table(df$QS109))
> cat("Distribución original de QS109 (No/Si):\n")
Distribución original de QS109 (No/Si):
> print(round(prop_table,3))

      No      Si
0.977 0.023
>
> # 2) Balanceo híbrido de QS109 (ROSE::ovun.sample method = "both")
> # ovun.sample con method="both" realiza sobremuestreo + submuestreo
(balanceo híbrido)
> set.seed(123)
> balanced <- ovun.sample(QS109 ~ ., data = df, method = "both", N =
nrow(df), p = 0.5)$data
> cat("\nDistribución después de balanceo híbrido:\n")

Distribución después de balanceo híbrido:
> print(round(prop.table(table(balanced$QS109)),3))

      No      Si
0.505 0.495
>
> # 3) Preparar formula para regresión logística (usar las descripciones en
tablas de salida)
> # Excluir QS109 de predictores
> predictors <-
c("V213", "V201", "V312", "HA57A", "QS102", "Dep", "Edad", "HV025", "QS25N", "QS26")
> formula_full <- as.formula(paste("QS109 ~", paste(predictors, collapse =
" + ")))
>
> # 4) Ajustar modelo logístico inicial con todos los predictores
> model_full <- glm(formula_full, data = balanced, family = binomial(link =
"logit"))
>
> # 5) Selección de variables por AIC (stepwise both directions)
> model_step <- stepAIC(model_full, direction = "both", trace = FALSE)
>

```

```

> cat("\nModelo final seleccionado por AIC:\n")

Modelo final seleccionado por AIC:
> print(summary(model_step))

Call:
glm(formula = QS109 ~ V213 + V201 + V312 + QS102 + Dep + Edad +
     HV025 + QS25N, family = binomial(link = "logit"), data = balanced)

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)    -0.009763   0.145926  -0.067  0.94666
V213Si          1.715652   0.247281   6.938 3.98e-12 ***
V2012 Hijos o mas 0.238353   0.106058   2.247  0.02462 *
V312Si          0.233830   0.097453   2.399  0.01642 *
QS102Si         1.272859   0.173373   7.342 2.11e-13 ***
DepSi          -0.295451   0.098715  -2.993  0.00276 **
Edad<=35       -1.199990   0.100220 -11.974 < 2e-16 ***
HV025Urbano     0.597870   0.100269   5.963 2.48e-09 ***
QS25NSi        -0.351819   0.105770  -3.326  0.00088 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 3356.0 on 2420 degrees of freedom
Residual deviance: 3006.1 on 2412 degrees of freedom
AIC: 3024.1

Number of Fisher Scoring iterations: 4

>
> # 6) Tabla de ORs (odds ratios) con intervalos de confianza 95%
> est <- broom::tidy(model_step, exponentiate = TRUE, conf.int = TRUE)
> est <- est %>%
+   rename("Termino" = term,
+           "OR" = estimate,
+           "IC_2.5" = conf.low,
+           "IC_97.5" = conf.high,
+           "p_valor" = p.value) %>%
+   mutate_at(vars(OR, IC_2.5, IC_97.5, p_valor), round, 3)
>
> cat("\nTabla de odds ratios (OR) del modelo final:\n")

Tabla de odds ratios (OR) del modelo final:
> kable(est, digits = 3)

```

Termino	OR	std.error	statistic	p_valor	IC_2.5	IC_97.5
(Intercept)	0.990	0.146	-0.067	0.947	0.744	1.318
V213Si	5.560	0.247	6.938	0.000	3.463	9.154
V2012 Hijos o mas	1.269	0.106	2.247	0.025	1.032	1.564
V312Si	1.263	0.097	2.399	0.016	1.044	1.530
QS102Si	3.571	0.173	7.342	0.000	2.562	5.062
DepSi	0.744	0.099	-2.993	0.003	0.613	0.903
Edad<=35	0.301	0.100	-11.974	0.000	0.247	0.366
HV025Urbano	1.818	0.100	5.963	0.000	1.495	2.215
QS25NSi	0.703	0.106	-3.326	0.001	0.571	0.865

```

>
> # 7) Resumen interpretativo (automatizado) - texto para incluir en tesis

```

```

> # Generar hilo de texto breve con los load-bearing resultados
> coefs <- summary(model_step)$coefficients
> # construir texto resumido (ejemplo)
> cat("\nResumen interpretativo sugerido (resumido):\n")

Resumen interpretativo sugerido (resumido):
> # Build sentences
> for(i in 2:nrow(coefs)){
+   name <- rownames(coefs)[i]
+   beta <- round(coefs[i,1],3)
+   se <- round(coefs[i,2],3)
+   z <- round(coefs[i,3],3)
+   p <- signif(coefs[i,4],3)
+   cat(sprintf("- El término '%s' tiene beta = %s (SE=%s), z=%s, p=%s. Una
OR estimada = %.3f (IC95%%: %.3f - %.3f).\n",
+               name, beta, se, z, p,
+               exp(beta), exp(beta - 1.96*se), exp(beta + 1.96*se)))
+ }
- El término 'V213Si' tiene beta = 1.716 (SE=0.247), z=6.938, p=3.98e-12.
Una OR estimada = 5.562 (IC95%: 3.428 - 9.026).
- El término 'V2012 Hijos o mas' tiene beta = 0.238 (SE=0.106), z=2.247,
p=0.0246. Una OR estimada = 1.269 (IC95%: 1.031 - 1.562).
- El término 'V312Si' tiene beta = 0.234 (SE=0.097), z=2.399, p=0.0164. Una
OR estimada = 1.264 (IC95%: 1.045 - 1.528).
- El término 'QS102Si' tiene beta = 1.273 (SE=0.173), z=7.342, p=2.11e-13.
Una OR estimada = 3.572 (IC95%: 2.544 - 5.013).
- El término 'DepSi' tiene beta = -0.295 (SE=0.099), z=-2.993, p=0.00276.
Una OR estimada = 0.745 (IC95%: 0.613 - 0.904).
- El término 'Edad<=35' tiene beta = -1.2 (SE=0.1), z=-11.974, p=4.89e-33.
Una OR estimada = 0.301 (IC95%: 0.248 - 0.366).
- El término 'HV025Urbano' tiene beta = 0.598 (SE=0.1), z=5.963, p=2.48e-
09. Una OR estimada = 1.818 (IC95%: 1.495 - 2.212).
- El término 'QS25NSi' tiene beta = -0.352 (SE=0.106), z=-3.326, p=0.00088.
Una OR estimada = 0.703 (IC95%: 0.571 - 0.866).
> cat("\nNota: sustituye las menciones de término por la 'descripción' de
la variable al incorporar al texto final.\n")

```

Nota: sustituye las menciones de término por la 'descripción' de la variable al incorporar al texto final.

```

>
> # 8) Curva ROC y punto de corte por Youden
> # Probabilidades predichas
> probs <- predict(model_step, type = "response")
> roc_obj <- roc(balanced$QS109, probs, levels = c("No", "Si"), direction =
"<")
> auc_val <- auc(roc_obj)
> cat(sprintf("\nAUC (Área bajo la curva ROC): %.3f\n", auc_val))

```

AUC (Área bajo la curva ROC): 0.713

```

>
> # Punto de corte óptimo (Youden)
> opt <- coords(roc_obj, "best", best.method = "youden", transpose = FALSE)
> cat("\nPunto de corte óptimo según Youden:\n")

```

Punto de corte óptimo según Youden:

```

> print(opt)
  threshold specificity sensitivity
1 0.4071425 0.5641864 0.7621035
>
> # Graficar ROC (si ejecutas en RStudio se mostrará)
> plot(roc_obj, main = paste0("Curva ROC - AUC = ", round(auc_val,3)))

```

```

> abline(a=0,b=1,lty=2,col="gray")
> # marcar punto óptimo
> points(opt["specificity"], opt["sensitivity"], pch = 19)
> text(opt["specificity"], opt["sensitivity"], labels = paste0("Cut=",
round(opt["threshold"],3)), pos=4)
>
> # 9) Matriz de confusión al punto de corte óptimo
> cutoff <- as.numeric(opt["threshold"])
> pred_class <- factor(ifelse(probs >= cutoff, "Si", "No"), levels =
c("No", "Si"))
>
> #####
> cm <- confusionMatrix(pred_class, balanced$QS109, positive = "Si")
> cat("\nMatriz de confusión (con el punto de corte óptimo):\n")

Matriz de confusión (con el punto de corte óptimo):
> print(cm$table)
      Reference
Prediction No  Si
      No 690 285
      Si 533 913
> cat("\nMedidas (Precision = Exactitud, Sensibilidad = Recall,
Especificidad):\n")

Medidas (Precision = Exactitud, Sensibilidad = Recall, Especificidad):
> print(cm$byClass[c("Sensitivity", "Specificity", "Precision")])
Sensitivity Specificity Precision
 0.7621035   0.5641864   0.6313970
> cat("\nExactitud global:\n")

Exactitud global:
> print(cm$overall["Accuracy"])
Accuracy
0.6621231
>
> # FIN del script

```

Anexo 1. Instrumentos

REPÚBLICA DEL PERÚ INSTITUTO NACIONAL DE ESTADÍSTICA E INFORMÁTICA ENCUESTA DEMOGRÁFICA Y DE SALUD FAMILIAR ENDES - 2024 CUESTIONARIO DEL HOGAR AMPARADO POR EL D. L. N° 604-SECRETO ESTADÍSTICO																	
CUESTIONARIO DE				<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 33%;">CONGLOMERADO</td> <td style="width: 33%;">VIVIENDA</td> <td style="width: 33%;">HOGAR</td> </tr> <tr> <td> </td> <td> </td> <td> </td> </tr> </table>								CONGLOMERADO	VIVIENDA	HOGAR			
CONGLOMERADO	VIVIENDA	HOGAR															

IDENTIFICACION											
A. HOGARES EN LA VIVIENDA											
Sr. (a): Si hogar es la persona o grupo de personas que se alimentan de una misma olla y atienden en común otras necesidades básicas.				¿Cuántos hogares ocupan esta vivienda?.....		TOTAL DE HOGARES 		HOGAR N°..... /			

NOMBRE DEL JEFE DEL HOGAR _____

VISITAS DE LA ENTREVISTADORA											
	1a.	2a.	3a.	4a.	5a.	VISITA FINAL					
FECHA						FECHA					
						DÍA.....					
NOMBRE DE LA ENTREVISTADORA						MES.....					
						AÑO.....					
RESULTADO***						EQUIPO NÚMERO.....					
PRÓXIMA VISITA:	FECHA					ENTREVISTADORA.....					
	HORA					RESULTADO FINAL.....					
						NÚMERO TOTAL DE VISITAS.....					

***CÓDIGOS DE RESULTADO: 1 COMPLETA 2 HOGAR PRESENTE PERO ENTREVISTADO COMPETENTE AUSENTE 3 HOGAR AUSENTE 4 APLAZADA 5 RECHAZADA ☐ NO DESEAN LA ENTREVISTA ☐ YA FUE ENTREVISTADA POR LA ENDES ☐ YA FUE ENTREVISTADA POR OTRA ENCUESTA DEL INEI				DATOS DE CONTROL TOTAL DE PERSONAS EN EL HOGAR NÚMERO DE MUJERES DE 15 A 49 AÑOS NÚMERO DE MUJERES DE 12 A 14 AÑOS NÚMERO DE NIÑAS Y NIÑOS MENORES DE 12 AÑOS NÚMERO DE NIÑAS Y NIÑOS MENORES DE 6 AÑOS NÚMERO DE NIÑAS Y NIÑOS MENORES DE 5 AÑOS NÚMERO DE NIÑAS Y NIÑOS MENORES DE 3 AÑOS N° DE ORDEN DEL INFORMANTE DEL CUESTIONARIO DEL HOGAR N° DE ORDEN DEL INFORMANTE DEL CUESTIONARIO DE SALUD							
--	--	--	--	--	--	--	--	--	--	--	--

NOMBRE Y FECHA DE SUPERVISION _____ NOMBRE Y FECHA DE REVISIÓN _____	SUPERVISORA LOCAL _____ _____	SUPERVISORA NACIONAL _____ _____	DIGITADOR (A) _____ _____
---	---	--	---

SECCIÓN 5. PRUEBA DE HEMOGLOBINA

Como parte de este estudio, estamos investigando la presencia de Anemia en mujeres y niños como un problema de orden nutricional en nuestro país, cuyas principales consecuencias son: fatiga, mareo, dolor de cabeza, palidez y palpitaciones, que también se expresa en un bajo rendimiento intelectual, alteraciones de la conducta y baja productividad. La anemia se presenta cuando la cantidad de hierro disponible en la sangre es insuficiente para satisfacer las necesidades individuales; es decir, sea por la carencia de suficientes glóbulos rojos o de una baja concentración de hemoglobina en la sangre.

Esta investigación ayudará al gobierno a desarrollar programas de prevención y tratamiento de la Anemia. Con tal motivo, solicitamos que usted (y los niños menores de 6 años) participen en la prueba de Anemia que consiste en dar una gotita de sangre de su dedo; prueba que realizamos con equipos probados y garantizados; e, insumos seguros y descartables para cada persona. Inmediatamente después de la prueba le estaremos dando los resultados, los cuales serán confidenciales.

Sin embargo, si usted decide no hacerse la prueba está en su derecho y nosotros respetaremos su decisión. Ahora, por favor dígame ¿Acepta hacerse la prueba de hemoglobina?

MUJERES DE 12 A 49 AÑOS DE EDAD

Nº DE ORDEN CIRCULADO EN COL. 9 Y 10	MUJER POR MUJER VEA PGTA 202. LUEGO, CIRCULE EN PGTA 208 SEGÚN CORRESPONDA	PGTE POR RESPONSABLE DE ESTA PERSONA Y ANOTE EL Nº DE ORDEN QUE EL RESPONSABLE TIENE EN EL L. H. SI NO ESTÁ EN ÉSTE, ANOTE "00".	REGISTRE FECHA Y HORA DE INICIO DE LA PRUEBA CONSIDERAR EL INICIO DESDE LA LECTURA DEL CONSENTIMIENTO Y/O EXPLICACIÓN A CADA PERSONA	LEA LA DECLARACIÓN DE CONSENTIMIENTO A CADA MUJER O PERSONA RESPONSABLE, DE ACUERDO AL CASO CIRCULE CÓDIGO	NIVEL DE HEMOGLOBINA (G/DL)	ACTUALMENTE EMBARAZADA (VEA PGTA 206 DEL C.I. LUEGO CIRCULE S. C.)	RESULTADO 1 MEDIDA 2 NO PRESENTE 3 RECHAZÓ 6 OTRO (ESPECIFIQUE)	REGISTRE FECHA Y HORA DE TÉRMINO DE LA PRUEBA CONSIDERAR EL TÉRMINO HASTA EL REGISTRO DEL RESULTADO DE LA PREGUNTA 213
(208)	(209)	(209A)	(210)	(211)	(212)		(213)	(213A)
12 - 17 AÑOS ... 1 18 - 49 AÑOS ... 2 PASE A 209A		DIA ... HORA ... MES ... MINUTOS ...	ACEPTÓ 1 RECHAZÓ / OTRO 2 PASE A 213		SI 1 NO/INS... 2			DIA ... HORA ... MES ... MINUTOS ...
12 - 17 AÑOS ... 1 18 - 49 AÑOS ... 2 PASE A 209A		DIA ... HORA ... MES ... MINUTOS ...	ACEPTÓ 1 RECHAZÓ / OTRO 2 PASE A 213		SI 1 NO/INS... 2			DIA ... HORA ... MES ... MINUTOS ...
12 - 17 AÑOS ... 1 18 - 49 AÑOS ... 2 PASE A 209A		DIA ... HORA ... MES ... MINUTOS ...	ACEPTÓ 1 RECHAZÓ / OTRO 2 PASE A 213		SI 1 NO/INS... 2			DIA ... HORA ... MES ... MINUTOS ...
12 - 17 AÑOS ... 1 18 - 49 AÑOS ... 2 PASE A 209A		DIA ... HORA ... MES ... MINUTOS ...	ACEPTÓ 1 RECHAZÓ / OTRO 2 PASE A 213		SI 1 NO/INS... 2			DIA ... HORA ... MES ... MINUTOS ...

Niños MENORES DE 6 AÑOS DE EDAD

Nº DE ORDEN CIRCULADO EN COL. 10	NIÑA /O A NIÑA /O VEA PGTA 203. LUEGO, CIRCULE EN PGTA 208 SEGÚN CORRESPONDA	PGTE POR RESPONSABLE DE ESTA PERSONA Y ANOTE EL Nº DE ORDEN QUE EL RESPONSABLE TIENE EN EL L. H. SI NO ESTÁ EN ÉSTE, ANOTE "00".	REGISTRE FECHA Y HORA DE INICIO DE LA PRUEBA CONSIDERAR EL INICIO DESDE LA LECTURA DEL CONSENTIMIENTO Y/O EXPLICACIÓN A CADA PERSONA	LEA LA DECLARACIÓN DE CONSENTIMIENTO A CADA MUJER O PERSONA RESPONSABLE, DE ACUERDO AL CASO CIRCULE CÓDIGO	NIVEL DE HEMOGLOBINA (G/DL)	HORA DE LA TOMA DE HEMOGLOBINA	FECHA DE LA MEDICIÓN	RESULTADO 1 MEDID(A) 2 NO PRESENTE 3 RECHAZÓ 6 OTRO (ESPECIFIQUE)	REGISTRE FECHA Y HORA DE TÉRMINO DE LA PRUEBA CONSIDERAR EL TÉRMINO HASTA EL REGISTRO DEL RESULTADO DE LA PREGUNTA 213
(208)	(209)	(209A)	(210)	(211)	(212 A)	(212B)	(213)	(213A)	
4 MESES Y MAS 1 OTRO 2 PASE SGTE. NIÑA /O		DIA ... HORA ... MES ... MINUTOS ...	ACEPTÓ 1 RECHAZÓ / OTRO 2 PASE A 213		HORA ... MINUTOS ...	DIA ... MES ...		DIA ... HORA ... MES ... MINUTOS ...	
4 MESES Y MAS 1 OTRO 2 PASE SGTE. NIÑA /O		DIA ... HORA ... MES ... MINUTOS ...	ACEPTÓ 1 RECHAZÓ / OTRO 2 PASE A 213		HORA ... MINUTOS ...	DIA ... MES ...		DIA ... HORA ... MES ... MINUTOS ...	
4 MESES Y MAS 1 OTRO 2 PASE SGTE. NIÑA /O		DIA ... HORA ... MES ... MINUTOS ...	ACEPTÓ 1 RECHAZÓ / OTRO 2 PASE A 213		HORA ... MINUTOS ...	DIA ... MES ...		DIA ... HORA ... MES ... MINUTOS ...	
4 MESES Y MAS 1 OTRO 2 PASE SGTE. NIÑA /O		DIA ... HORA ... MES ... MINUTOS ...	ACEPTÓ 1 RECHAZÓ / OTRO 2 PASE A 213		HORA ... MINUTOS ...	DIA ... MES ...		DIA ... HORA ... MES ... MINUTOS ...	
4 MESES Y MAS 1 OTRO 2 PASE SGTE. NIÑA /O		DIA ... HORA ... MES ... MINUTOS ...	ACEPTÓ 1 RECHAZÓ / OTRO 2 PASE A 213		HORA ... MINUTOS ...	DIA ... MES ...		DIA ... HORA ... MES ... MINUTOS ...	
4 MESES Y MAS 1 OTRO 2 PASE SGTE. NIÑA /O		DIA ... HORA ... MES ... MINUTOS ...	ACEPTÓ 1 RECHAZÓ / OTRO 2 PASE A 213		HORA ... MINUTOS ...	DIA ... MES ...		DIA ... HORA ... MES ... MINUTOS ...	