



Universidad Veracruzana

Facultad de Estadística e Informática

Región Xalapa

Especialización en Métodos Estadísticos

Modelación del tipo de atención médica recibida por causa de dengue en Veracruz, 2020-2023

Reporte de aplicación
para obtener el diploma de Especialista en
Métodos Estadísticos

Presenta:

Juliet Dalila Lozano Rafael

Directora:

Mtra. María Yesenia Zavaleta Sánchez

Co-director:

Dr. Ángel Juan Sánchez García

Agosto de 2025

“Lis de Veracruz: Arte, Ciencia, Luz”



Universidad Veracruzana

Facultad de Estadística e Informática
Región Xalapa

Especialización en Métodos Estadísticos

Modelación del tipo de atención médica recibida por causa de dengue en Veracruz, 2020-2023.

Reporte de aplicación para obtener el diploma de Especialista
en Métodos Estadísticos

Presenta:
Juliet Dalila Lozano Rafael

Directora:
Mtra. María Yesenia Zavaleta Sánchez

Co-director:
Dr. Ángel Juan Sánchez García

UNIVERSIDAD VERACRUZANA

Facultad de Estadística e Informática

El Comité Académico de la Especialización en Métodos Estadísticos y el director de este trabajo recepcional intitulado: **Modelación del tipo de atención médica recibida por causa de dengue en Veracruz, 2020-2023**, autorizan la impresión y la constitución del jurado para la defensa.

COMITÉ ACADÉMICO



Dra. Cecilia Cruz López
Coordinadora
Especialización en Métodos Estadísticos



Dr. Luis Gerardo Montané Jiménez
Director
Facultad de Estadística e Informática



Mtro. Carlos Adrián Alarcón Rojas
Secretario



Dr. Sergio Francisco Juárez Cerrillo
Vocal



Dra. Julia Aurora Montano Rivas

Presidente

GENERACIÓN: 2024-2025

SEDE: Xalapa

TÍTULO: Modelación del tipo de atención médica recibida por causa de dengue en Veracruz, 2020-2023.

AUTOR: Juliet Dalila Lozano Rafael

DIRECTOR y co-director: Mtra. María Yesenia Zavaleta Sánchez, Dr. Ángel Juan Sánchez García.

TIPO DE TRABAJO:

Reporte de aplicación Trabajo Practico-Educativo Desarrollo estadístico Monografía

TIPO DE ESTUDIO:

Exploratorio	<input type="checkbox"/>	Retrospectivo	<input type="checkbox"/>
Descriptivo	<input type="checkbox"/>	Prospectivo	<input type="checkbox"/>
Expostfaco o cuasiexperimental	<input type="checkbox"/>	Transversal	<input checked="" type="checkbox"/>
Experimental	<input type="checkbox"/>	Longitudinal	<input type="checkbox"/>

METODOLOGÍA ESTADÍSTICA:

A) Diseño:

Muestreo	<input type="checkbox"/>
Experimento	<input type="checkbox"/>
Estudio observacional	<input checked="" type="checkbox"/>

B) Análisis

Exploratorio	<input type="checkbox"/>
Descriptivo básico	<input checked="" type="checkbox"/>
Inferencia básica	<input checked="" type="checkbox"/>
Métodos multivariados	<input checked="" type="checkbox"/>
Regresión	<input checked="" type="checkbox"/>
ANOVA y ANCOVA	<input checked="" type="checkbox"/>
Control de calidad	<input type="checkbox"/>
Métodos no paramétricos	<input checked="" type="checkbox"/>
Modelos especiales	<input type="checkbox"/>
Técnicas avanzadas	<input type="checkbox"/>
Series de tiempo	<input type="checkbox"/>

Dedicatoria o agradecimientos

Este trabajo está dedicado a mi familia, amigos, y a mi novio Ángel, por su increíble apoyo en esta temporada. Menciono agradecimientos especiales a las dras. Julia Aurora Montano Rivas y Cecilia Cruz López, por su soporte para culminar mi etapa final, al Dr. Nicandro Cruz Ramírez por brindarme sus consejos durante mi estancia, a los dres. José Ramos Castañeda y Ruth Martínez Vega por su asesoría en el tema del dengue, y a la Lic. Carmín Marín Guerrero por su ayuda incondicional.

Índice

Dedicatoria o agradecimientos	3
Índice	4
1. Introducción	8
1.1. Planteamiento del problema	8
1.2. Objetivos.....	10
1.2.1. Objetivo general	10
1.2.2. Objetivos particulares	10
1.3. Hipótesis de investigación	10
1.4. Justificación	10
2. Marco conceptual y revisión sistemática de la literatura	12
2.1. Marco conceptual	12
2.2. Revisión Sistemática de la Literatura	13
2.2.1. Revisión de proyectos similares	15
2.2.2. Limitaciones de las investigaciones encontradas	18
3. Metodología	20
3.1. Diseño del estudio	20
3.2. Recolección de datos	20
3.3. Preprocesamiento de datos	20
3.4. Análisis estadístico	22
3.5. Descripción de las técnicas estadísticas a utilizar	23
3.5.1. Prueba Ji cuadrada (χ^2) para proporciones.....	24
3.5.2. Prueba Ji cuadrada (χ^2) de independencia	24
3.5.3. Prueba exacta de Fisher	25
3.5.4. Coeficiente de asociación V de Cramer (v)	25
3.5.5. Análisis de Correspondencias Múltiples	26
3.5.6. Validación cruzada por bloques y rendimiento de un clasificador.....	27
3.5.7. Regresión Logística	28
3.5.8. Árboles de decisión	29
3.5.9. Análisis de varianza de una vía (ANOVA) con medidas repetidas.....	31
3.6. Justificación de la elección de estadísticos.....	33
4. Resultados	35
4.1. Características de los casos de dengue	35

4.2.	Tipo de atención médica recibida y características clínicas	38
4.3.	Análisis de Correspondencias Múltiples	45
4.4.	Modelo predictivo.....	47
4.4.1.	Selección del mejor modelo	47
4.4.2.	Representación del modelo predictivo de tipo de atención médica.....	52
5.	Discusión.....	57
6.	Conclusiones	60
7.	Limitaciones y trabajos futuros	61
	Referencias	62

Resumen

La presencia de dengue ha incrementado en los últimos años. Los organismos de salud pública propusieron un protocolo de atención médica para tratar a los pacientes confirmados de dengue, sin embargo, estos son hospitalizados hasta que aparecen signos de alarma en casos con comorbilidad y condiciones especiales, pues se infiere que la enfermedad se ha agravado. En el siguiente trabajo se creó un modelo predictivo para determinar el tipo de atención médica que el paciente recibirá según las características clínicas que presenta, con el fin de evitar agravamientos en la enfermedad por dengue y prevenir posibles fallecimientos. Para esto se tomaron los datos abiertos que presenta la Secretaría de Salud en México, seleccionando solo el estado de Veracruz, y se eligieron variables clínicas, comorbilidades y condiciones especiales en pacientes confirmados de dengue. Se realizó un análisis descriptivo y se compararon cuatro modelos: regresión logística y árboles de decisión con todas las variables y con aquellas que resultaron significativas en el análisis previo. El mejor modelo fue el de regresión logística con las variables significativas, y con esto se encontró una capacidad de predicción del 66.8%. Además, se obtuvo que las comorbilidades diabetes, hipertensión y enfermedad renal, así como la presencia de embarazo y el serotipo, son determinantes significativos para decidir si un paciente será hospitalizado o no. Con los resultados obtenidos, se invita a las instituciones a realizar un seguimiento hospitalario constante de los pacientes confirmados de dengue con las comorbilidades y condiciones especiales mencionadas anteriormente, para evitar la saturación hospitalaria no programada y el agravamiento de cada paciente.

Palabras clave: Dengue, predicción, modelo, regresión logística, árboles de decisión, hospitalizados.

I. Introducción

La razón principal para realizar esta investigación surge del cambio actual de la enfermedad producida por dengue, puesto que en los últimos meses la incidencia de este ha aumentado exponencialmente en el estado de Veracruz. Se desea determinar la significancia del cambio, así como aquellos factores que determinan el tipo de atención y la complicación de la enfermedad a nivel hospitalario, para poder prevenir el crecimiento acelerado de la epidemia y la saturación del entorno hospitalario.

Esta investigación consta de seis capítulos. En el Capítulo 1 se aborda el planteamiento del problema de interés, los objetivos y las hipótesis, así como la justificación del enfoque utilizado en la investigación. En el Capítulo 2, se describe la revisión de la literatura y el marco conceptual. En el Capítulo 3 se muestra la metodología empleada para la recolección, tratamiento y análisis de los datos. En el Capítulo 4 se muestran los resultados descritos detalladamente. En el Capítulo 5 se plantea la discusión de los resultados obtenidos. En el Capítulo 6 se muestran las conclusiones obtenidas. Finalmente, en el Capítulo 7 se describen las limitaciones encontradas en el estudio y los trabajos que se sugieren a futuro.

1.1. Planteamiento del problema

La incidencia de casos de dengue ha aumentado en la actualidad. De acuerdo con la Organización Panamericana de la Salud (OPS, 2024), durante el año 2023 se registró el mayor número de casos en la región de las Américas, pues se registraron 4,594,823 casos. De la misma forma, según registros de la Secretaría de Salud en México (2024), durante la semana epidemiológica 52 de 2023 se registraron 54,406 casos confirmados, es decir, poco más de cuatro veces la cantidad de casos confirmados durante la misma semana del año 2022. Esta infección aún se encuentra desatendida puesto que se dio prioridad a la pandemia por COVID-19, por lo tanto, las medidas de prevención no son las suficientes.

Veracruz es uno de los estados que presenta mayor incidencia de dengue a nivel nacional. En los últimos cinco años, fue uno de los estados donde se concentró el mayor número de casos confirmados (Secretaría de Salud, 2024). Asimismo, durante la semana 52 de 2023 el estado reportó 10,480 casos confirmados de dengue, aproximadamente seis veces

el número de casos confirmados en la misma semana del año 2022 (1,751 casos). Durante este periodo se detectó la circulación de los cuatro serotipos existentes, aumentando la letalidad de 0.87 a 1.23 por cada 100 casos de dengue con signos de alarma y dengue grave.

La OPS en conjunto con la Organización Mundial de la Salud (OMS) definieron un protocolo de atención médica para casos sospechosos de dengue, asignando a cada paciente a un grupo para recibir tratamiento de acuerdo con los síntomas y características de este (OPS, 2020). No obstante, el protocolo de atención adaptado en México refiere que, hasta que aparecen signos de alarma en pacientes con comorbilidad y condiciones especiales, estos son enviados a tratamiento hospitalizado, pues se infiere que la enfermedad se ha agravado (Gobierno de México, 2024).

En la literatura consultada, no se encontraron investigaciones que aborden los casos de dengue en México – y en particular en el estado de Veracruz – desde un punto de vista clínico y predictivo, tratando la gravedad del dengue a través del tipo de atención médica recibida. Por ello, las preguntas de investigación que surgen son:

1. ¿Cuál es el estatus de los casos de dengue en los últimos años en Veracruz?
2. ¿Cuáles son las características clínicas de la población confirmada de dengue en Veracruz de acuerdo con el tipo de atención médica recibida?
3. ¿Qué factores determinan el tipo de atención médica recibida por los pacientes confirmados de dengue en Veracruz?
4. ¿Qué ventajas presenta un modelo estadístico para la predicción del tipo de atención médica recibida contra otros modelos?
5. ¿Cómo ha cambiado la proporción de casos confirmados de dengue en los últimos cuatro años?

1.2. Objetivos

1.2.1. Objetivo general

Construir un modelo para predecir el tipo de atención médica recibida en pacientes confirmados por dengue en Veracruz durante el periodo 2020-2023.

1.2.2. Objetivos particulares

- Describir el estatus de los casos de dengue en el estado de Veracruz en el período 2020 – 2023.
- Presentar las diferencias en la proporción de casos confirmados por año.
- Determinar la asociación que existe entre las variables clínicas y el tipo de atención médica recibida en los casos confirmados de dengue en el estado de Veracruz.
- Modelar el tipo de atención médica recibida en los casos confirmados de dengue en Veracruz usando las variables clínicas.
- Evaluar los modelos construidos con métricas de clasificación (precisión, sensibilidad, especificidad).
- Interpretar el mejor modelo de predicción del tipo de atención médica recibida.

1.3. Hipótesis de investigación

Las enfermedades crónicas influyen en el tipo de atención médica recibida por los pacientes confirmados de dengue en el estado de Veracruz.

1.4. Justificación

Dentro de los lineamientos establecidos en la agenda 2030 para el desarrollo sostenible, se aborda el apartado 3: Salud y Bienestar, que se centra en asegurar el bienestar de las personas y fomentar la salud en todas las edades. Entre los objetivos específicos se encuentra la erradicación de epidemias y control de enfermedades tropicales desatendidas, como el dengue (Organización de las Naciones Unidas [ONU], 2023). En este sentido, resulta

fundamental conocer aquellos factores clínicos que influyen en el agravamiento del dengue a través del tipo de atención médica, con el fin de dar un seguimiento a los pacientes contagiados, prevenir posibles decesos y además reducir la saturación hospitalaria causada por epidemias, sobre todo en las unidades de cuidados intensivos (OPS, 2025).

Por ello, este trabajo presenta una forma de representar el fenómeno en el estado de Veracruz, analizando la incidencia y conociendo aquellas características que producen un cambio en el tipo de atención que el paciente recibe, para mejorar las medidas de cuidado. Al usar este enfoque, se pueden implementar acciones dirigidas a poblaciones específicas basadas en resultados estadísticos.

2. Marco conceptual y revisión sistemática de la literatura

2.1. Marco conceptual

El dengue es una enfermedad viral causada por la picadura de mosquitos pertenecientes al género *Aedes*, mayormente *Aedes aegypti* (Martínez-Torres, 1995). Las infecciones, causadas principalmente por los mosquitos hembra, suelen tener un periodo de incubación de hasta 14 días. El hábitat principal del mosquito es en climas cálidos, sobre territorios a menos de 700 metros sobre el nivel del mar. Sin embargo, dado el cambio climático actual, se han registrado epidemias de mosquitos en zonas no endémicas y territorios elevados (Pizarro, 2009). El dengue es considerado un arbovirus, es decir, un virus transmitido por artrópodos. Los mosquitos sobreviven por la transmisión de la infección de huéspedes infectados a no infectados, por medio de una picadura que suele ocurrir durante el día (Vargas *et al.*, 2021). El mosquito es considerado doméstico por su estrecha relación con los seres humanos, ya que se da principalmente en zonas urbanas y colonizadas por el hombre.

Existen cuatro serotipos de dengue: DENV-1, DENV-2, DENV-3, DENV-4. Si una persona es infectada por uno de ellos, tendrá inmunidad de ese serotipo a largo plazo, mientras que al infectarse por segunda vez o más de dos veces, el riesgo de tener síntomas graves aumenta (OMS, 2023).

Los síntomas principales del dengue definidos por la OMS (2024) son: fiebre, dolor de cabeza, dolor retroocular, dolores musculares y articulares, náuseas y vómitos, glándulas inflamadas, erupciones en la piel, entre otros. Si la infección por dengue se agrava, los síntomas pueden ser dolor abdominal severo, vómitos constantes, sangrado de nariz o encías, fatiga, inquietud, sangre en vómito o heces, deshidratación, debilidad, entre otros. Los síntomas graves aparecen luego de la etapa de fiebre. El periodo general de la infección ocurre de una a dos semanas, pero los síntomas pueden durar de dos a siete días.

El Gobierno de México (2024) aplica el protocolo de atención médica sugerida por la OPS. En este marco, se define un caso probable como aquel que proceda de una región donde existe transmisión de dengue 14 días previos, y que presente los síntomas definidos por la OMS. Una vez estudiado el caso, los pacientes son clasificados en tres grupos principales:

dengue sin signos de alarma (DSSA), dengue con signos de alarma (DCSA) y dengue grave (DG).

Un paciente DSSA puede pasar al grupo DCSA si presenta comorbilidades al momento de su contagio y estas se agravan. Así mismo, si el paciente cuenta con condiciones especiales como embarazo, se maneja como DCSA automáticamente. El tratamiento general para un paciente DCSA es aplicar soluciones cristaloides durante 60 minutos, pero si se han aplicado tres dosis similares y no hay mejora clínica, el paciente es considerado como DG, para lo cual se sigue el mismo tratamiento en condiciones individualizadas y monitoreo estricto. Si el paciente es diagnosticado como DSSA, su tratamiento es ambulatorio, mientras que en la mayoría de los casos DCSA y DG se requiere hospitalización.

2.2. Revisión Sistemática de la Literatura

Para tratar el problema planteado, se realizó una revisión sistemática de la literatura utilizando la metodología de Kitchenham, Budgen y Brereton (2015) y las adecuaciones propuestas por Tebes *et al.* (2020). Los resultados de la búsqueda especializada se muestran en la Tabla 1.

Se seleccionaron las siguientes investigaciones, las cuales provienen de trece países: Brasil, China, Cuba, India, Vietnam, Perú, Camboya, Nicaragua, Venezuela, Tailandia, Puerto Rico, Singapur y México, publicados de 2020 a 2024 en los motores de búsqueda mencionados en la Tabla 1. La mayoría de las investigaciones corresponden a estudios longitudinales, sin embargo, sí se encontraron estudios transversales. Se muestran artículos de carácter descriptivo, explicativo y predictivo. Se encontró que 25% de las investigaciones corresponden a 2020, 6.25% fueron publicadas en 2021, 25% son de 2022, 12.50% son de 2023 y 31.25% fueron publicadas en 2024 (Tabla 2). Las palabras más usadas en las investigaciones dadas las palabras clave fueron: dengue, *machine-learning*, modelo, México, grave, selección, enfermedad, tiempo, series, epidemia, dinámica, serotipo, entre otras (Figura 1). Entre las técnicas más usadas se encuentran las pruebas de asociación Ji cuadrada, la Regresión Logística, las Redes Neuronales, y también modelos más avanzados como el análisis filogenético y los modelos SIR (Suceptibles-Infectados-Recuperados).

Tabla 1. Protocolo de la Revisión Sistemática de la Literatura.

Preguntas de investigación	
<ol style="list-style-type: none"> 1. ¿Cuál es el estatus de los casos de dengue en los últimos cinco años en Veracruz? 2. ¿Cuáles son los factores que influyen en la presencia de dengue en Veracruz? 3. ¿Cuáles son los factores que influyen en el tipo de atención de pacientes con dengue en Veracruz? 4. ¿Qué ventajas presenta un modelo de predicción contra otros modelos? 5. ¿Cuál es la metodología propuesta por otros autores? 6. ¿Cuáles han sido las principales limitaciones de los estudios encontrados? 	
Protocolo de búsqueda	
Cadena	“Dengue” AND (cases OR prediction OR model OR severe OR Mexico OR Veracruz or Time)
Metadatos	Título, resumen, palabras clave
Motores de búsqueda	EBSCO, IEEE Xplore, PLOS ONE, Web of Science, PUBMED
Criterios de selección	<p>Inclusión:</p> <ul style="list-style-type: none"> Que sean artículos de investigación (preferentemente) Que expliquen el estudio de los casos de dengue Que sean de libre acceso (open access) Que sean trabajos publicados en los últimos cinco años <p>Exclusión:</p> <ul style="list-style-type: none"> Estudios duplicados
Formulario de extracción de datos:	
Biblioteca, autor (es), año, país, región, título, palabras clave, objetivos, período de análisis, tipo de investigación, origen de los datos, tamaño de muestra, técnicas aplicadas, principales hallazgos, limitaciones y trabajos futuros.	
Referencia: Elaboración propia, adaptada de Tebes et al. (2020)	

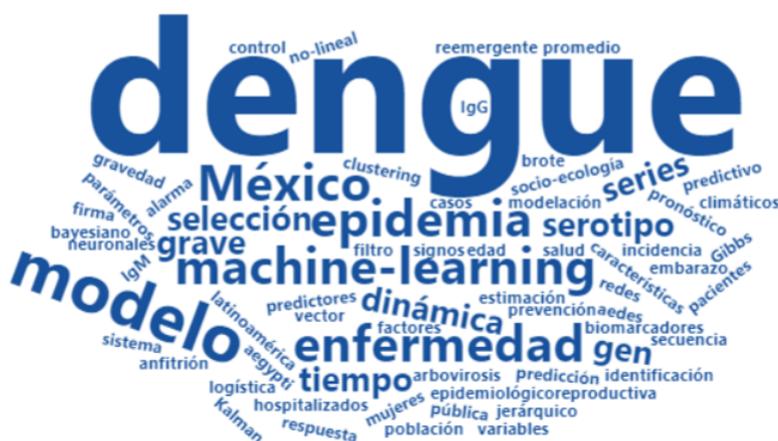


Figura 1. Palabras clave más usadas en las investigaciones encontradas.

Referencia: Elaboración propia.

2.2.1. Revisión de proyectos similares

En Cuba, Sebrango *et al.* (2020) compararon dos brotes de dengue en La Habana durante 2006-2007, y se modeló la incidencia usando los modelos de Richards, *Gibbs Sampling* y Regresión Logística. Se encontró que el modelo de *Gibbs* fue el más eficiente, mientras que el modelo de Richards torna lento el proceso. Ambas técnicas resultaron útiles como herramientas para mejorar las medidas de control en la región de La Habana.

Autores como Liu *et al.* (2022) y Wang *et al.* (2022) buscaron características y factores de riesgo que expliquen la progresión del dengue grave. Se usaron técnicas más avanzadas con aprendizaje automático como el algoritmo XGBoost, así como técnicas propias como el análisis filogenético. Liu *et al.* analizaron este fenómeno utilizando conjuntos de casos de siete países diferentes, mientras que Wang *et al.* lo estudiaron únicamente en China. En ambos casos, los resultados apuntan a una herramienta para la mejora en el diagnóstico previo del paciente y un tratamiento adecuado que disminuya la sobrecarga de los recursos sanitarios.

En México, Annan *et al.* (2023) analizaron cómo el embarazo en pacientes infectadas con dengue interactuaba con el serotipo durante 2012 a 2020 en el país. Se analizaron 2,469 unidades médicas y se aplicó una regresión logística múltiple para explicar la gravedad del dengue. Se concluyó que el embarazo aumenta la probabilidad de agravamiento del dengue, pero este se encuentra moderado de acuerdo con el serotipo. Otro hallazgo significativo fue la identificación de algunas comorbilidades como diabetes e hipertensión como factores influyentes en el aumento de riesgo.

En Delhi, India, Singh y Chaturvedi (2024) buscaron los principales predictores a nivel individual, de hogar y de comunidad, de casos de dengue en los diversos puntos calientes de la ciudad. Obtuvieron sus propios casos sobre pacientes con dengue durante 2015-2018, con una muestra de $n=347$ pacientes. Para analizar los casos se obtuvieron estadísticas descriptivas y pruebas de asociación tales como Ji cuadrada, mientras que para la predicción se realizó una Regresión Logística. Se encontró que un factor importante es el ingreso mensual del hogar, así como la densidad de este (miembros por habitación). Otros

predictores importantes fueron las condiciones de la vivienda, el lugar donde proviene el agua, el drenaje, las prácticas de recolección de basura y el combustible utilizado para cocinar. Una asociación positiva fue el almacenamiento de agua, así como el uso de baños públicos.

Tabla 2. Matriz del estado del arte.				
Autor (es)	Objetivo	Origen de los datos	Técnicas aplicadas	Aportación
Appice et al. (2020)	Predecir la actividad del dengue explorando la incidencia en varios estados de México.	Datos mensuales de la Secretaría de Salud y Comisión Nacional del Agua, 1985-2010.	Comparación de Auto TiC-NN (auto-encoding, count-based window time series clustering, and nearest-neighbor prediction) contra otros métodos de aprendizaje.	La propuesta de los autores superó el poder de predicción contra el resto de las técnicas. Al separar los casos por clústeres se mejora la ganancia en la capacidad predictiva.
Benedum et al. (2020)	Comparar algoritmos de <i>machine learning</i> con modelos de regresión, para predecir casos y brotes con cuatro a doce semanas de anticipación.	Sistemas de vigilancia de Perú, Puerto Rico y Singapur, 2000-2013.	Bosques aleatorios, modelos autorregresivos, regresión Poisson.	Los bosques aleatorios fueron más precisos que la regresión Poisson y ARIMA a corto plazo, y éstos últimos fueron mejores a largo plazo.
Tamayo et al. (2020)	Describir a los pacientes con dengue de acuerdo con la clasificación definida por la OMS y relacionarla con la clasificación tradicional con respecto a la gravedad de los casos.	247 pacientes ingresados en el Instituto de Medicina Tropical Pedro Kourí de La Habana, Cuba en 2012.	Estadísticos descriptivos y pruebas de independencia Ji cuadrada.	Se evidenció que con la identificación tradicional se puede omitir el tratamiento adecuado en pacientes con dengue grave y pacientes con signos de alarma.
Ferdousi, Cohnstaedt y Scoglio (2021)	Mejorar predicciones de casos de dengue agregando casos relevantes de ubicaciones cercanas, en Brasil.	Casos del Proyecto Info Vigilancia del Dengue y del <i>National Oceanic and Atmospheric Administration</i> en Espirito Santo, 2010-2019.	Redes neuronales con series de tiempo.	La metodología propuesta funciona para mejorar la predicción en muchos conjuntos espaciotemporales.

Schaum, Bernal-Jaquez y Sánchez-González (2022)	Predecir la propagación futura del dengue, empleando en mayor profundidad un modelo con respecto a sus propiedades de observabilidad estructural y el diseño de un esquema de monitoreo de predicción basado en un observador de estado no lineal.	Mediciones mensuales de Cuernavaca, México, del Centro de Investigación sobre Enfermedades Infecciosas del Instituto Nacional de Salud Pública, 2012.	Análisis de observabilidad estructural, simulaciones Monte Carlo y modelos SIR.	La propuesta tiene potencial para el monitoreo y predicción de casos de dengue basados en estimaciones iniciales proporcionadas por el filtro de Kalman.
Solís-Navarro et al. (2022)	Predecir la dispersión espacial de los casos de dengue, en México.	Datos del Instituto Nacional de Estadística y Geografía, casos de la Secretaría de Salud del Estado y mediciones climatológicas de la <i>World Meteorological Organization</i> , 2019.	Índice de Morán y regresión Poisson con métodos Monte Carlo.	Los factores climáticos fueron los más relevantes para la presencia y crecimiento de casos positivos de dengue, y esto ayuda a prevenir futuros casos en zonas vulnerables.
Luque et al. (2023)	Conocer datos epidemiológicos y clínicos de pacientes fallecidos por dengue, 2017, en Perú.	24 historias clínicas que proporcionó el Área de Epidemiología de la Dirección Regional de Salud de Piura, Perú.	Estadísticos descriptivos.	La mayoría fueron mujeres, y los tiempos de atención fueron tardíos; las mayores tasas de mortalidad fueron en mayores de 65 años y en personas de 19 a 35 años, así como pacientes con diabetes e hipertensión.
Conde-Gutiérrez et al. (2024)	Proponer un modelo que simule la dinámica del dengue según sus clasificaciones.	Casos registrados por la Secretaría de Salud en México y la Organización Mundial de la Salud durante 2022.	Redes Neuronales Artificiales (ANN).	El modelo ANN funcionó para predecir los casos de dengue con cinco neuronas y una función de activación logística. Los meses con mayores casos

				confirmados fueron octubre y noviembre.
Karasinghe et al. (2024)	Pronosticar la incidencia semanal del dengue en la región de Sri Lanka, India.	Casos de dengue del Ministerio de Salud, 2015-2020.	Medias móviles y modelos autorregresivos.	El modelo pronostica muy bien los casos de dengue en un corto plazo. Con esta información se evita la posible propagación del virus en nuevas regiones.
Rodríguez et al. (2024)	Dar un seguimiento el comportamiento del dengue a través de la descripción de este a lo largo de los años.	Datos de la Plataforma de Información en Salud de las Américas (PLISA), la Organización Mundial de la Salud (OMS), la Dirección General de Epidemiología (DGE) y el Sistema Nacional de Vigilancia Epidemiológica (SINAVE) en México de 2019 a 2023.	Estadísticos descriptivos.	Los hallazgos fueron que la mayoría de los casos se dan entre las semanas 38 y 43, mientras que los estados con mayor incidencia en México fueron Veracruz, Jalisco, Yucatán, Quintana Roo y Morelos.
Domínguez y Valdés (2024)	Describir las características clínico-epidemiológicas de pacientes con dengue y signos de alarma durante la epidemia de 2022.	255 pacientes con dengue y signos de alarma de Santiago de Cuba, 2022.	Estadísticos descriptivos.	Las comorbilidades con prevalencia mayor incluyeron hipertensión arterial, diabetes y asma.
Referencia: Elaboración propia.				

2.2.2. Limitaciones de las investigaciones encontradas

En los trabajos antes mencionados se detectaron las siguientes limitaciones:

- Pocas variables. En algunos artículos se recomienda agregar variables de tipo socioeconómicas o ambientales para obtener una mejor predicción de la incidencia de dengue (Solís-Navarro *et al.*, 2022; Karasinghe *et al.*, 2024; Wang *et al.*, 2022).

- Sobreajuste. Ciertos estudios presentaron este problema, puesto que los modelos creados para predecir la incidencia de dengue solo predijeron con buena precisión los datos de entrenamiento, pero no los de prueba (Benedum *et al.*, 2020).
- Tamaño de la muestra. En un estudio se presentó que el tamaño de muestra fue muy pequeño, por lo que se recomienda abarcar un área mayor de estudio (Wang *et al.*, 2022).

3. Metodología

3.1. Diseño del estudio

El diseño de la investigación es de clase no experimental. Se trabajó con datos pertenecientes a cuatro años diferentes (2020, 2021, 2022 y 2023). El tipo de estudio es descriptivo debido a que se caracteriza a la población confirmada de dengue, es correlacional puesto que se busca la asociación entre las características clínicas y el tipo de atención médica que recibió el paciente, y predictivo porque con las asociaciones encontradas se busca obtener una predicción del tipo de atención médica dadas las características clínicas del paciente.

3.2. Recolección de datos

La población delimitada son los casos de dengue (casos confirmados, descartados y sospechosos). Se tomaron los datos abiertos sobre enfermedades transmitidas por vector, que se encuentran en el portal de documentos de la Secretaría de Salud de México, para los años 2020, 2021, 2022 y 2023 (SALUD, 2025). Cada conjunto de datos contiene un total de 28 variables, que corresponden al seguimiento de los contagios de dengue de acuerdo con las vigilancias epidemiológicas realizadas en todo el país. Para fines del estudio, se tomó el estado de Veracruz, por lo que el conjunto final de datos contiene 57,177 registros, y se seleccionaron 11 variables de interés, correspondientes a datos clínicos del paciente.

3.3. Preprocesamiento de datos

Se unieron los conjuntos de casos, agregando una columna que indica el año de registro. Las variables fueron convertidas a formato factor. La edad de los pacientes fue categorizada en cinco grupos (0-20 años, 21-40 años, 41-60 años, 61-80 años, +80 años). La unidad médica de registro del paciente fue recategorizada con las nuevas etiquetas: IMSS, IMSS-BIENESTAR, ISSSTE, SSA y Otra. Para filtrar los casos confirmados, se seleccionaron aquellos que fueron diagnosticados con uno de los cuatro serotipos existentes; además, se eliminaron dos casos que no registraron el tipo de atención médica recibida (TP), debido a que es la variable dependiente en este estudio (Figura 2). El conjunto final de datos se trabajó con las variables de la Tabla 3.

Tabla 3. Descripción de las variables.

Variable	Descripción	Tipo de variable	Escala	Valores
Año (Año)	Año del caso registrado.	Cualitativa	Ordinal	2020 2021 2022 2023
Sexo (Sex)	Sexo del paciente.	Cualitativa	Nominal	Hombre Mujer
Edad (Ed)	Edad del paciente.	Cualitativa	Ordinal	0-20 años 21-40 años 41-60 años 61-80 años +80 años
Estatus (Est)	Estatus de caso registrado.	Cualitativa	Nominal	Confirmado Descartado Probable
Resultado PCR (PCR)	Resultado de la prueba PCR.	Cualitativa	Nominal	DENV-1 DENV-2 DENV-3 DENV-4
Tipo de paciente (TP)	Tipo de atención médica que ha recibido el paciente.	Cualitativa	Nominal	Ambulatorio Hospitalizado
Paciente hemorrágico (PH)	Si el paciente tuvo dengue hemorrágico.	Cualitativa	Nominal	Sí No
Diabetes (Diab)	Si el paciente sufre de diabetes.	Cualitativa	Nominal	Sí No
Hipertensión (Hip)	Si el paciente sufre de hipertensión.	Cualitativa	Nominal	Sí No
Enfermedad péptica (UCL)	Si el paciente sufre de enfermedad péptica.	Cualitativa	Nominal	Sí No
Enfermedad renal (Ren)	Si el paciente sufre de enfermedad renal.	Cualitativa	Nominal	Sí No
Inmunosupresión (Inmuno)	Si el paciente sufre de inmunosupresión.	Cualitativa	Nominal	Sí No
Cirrosis hepática (CHep)	Si el paciente sufre de cirrosis hepática.	Cualitativa	Nominal	Sí No
Embarazo (Emb)	Si la paciente está embarazada.	Cualitativa	Nominal	Sí No
U.M. Notificación (UM)	Unidad Médica donde se registró el caso.	Cualitativa	Nominal	IMSS IMSS-BIENESTAR ISSSTE SSA Otra

Nota: Se resalta en negritas la variable dependiente.
Referencia: Elaboración propia.

3.4. Análisis estadístico

Se realizó un análisis descriptivo sobre todos los casos de dengue por edad y sexo. Se calculó la proporción de casos confirmados de dengue por año y se aplicaron pruebas Ji cuadrada de homogeneidad para identificar, por parejas de años, diferencias estadísticamente significativas entre las proporciones de casos confirmados de dengue. Asimismo, se describieron los serotipos de dengue encontrados durante el período 2020-2023, y el estatus de los casos reportados en el mismo periodo.

Con los casos confirmados, se describieron las variables clínicas según el tipo de atención recibida, se realizaron pruebas Ji cuadrada de Pearson y pruebas exactas de Fisher (en las variables con frecuencias esperadas menores a 5) para determinar asociación entre cada variable clínica y el tipo de atención, y se calculó el estadístico V de Cramer para definir la magnitud de estas asociaciones.

Se realizó un análisis de correspondencias múltiples (ACM) con todas las variables para encontrar asociaciones o correspondencias a nivel multivariante.

Se ajustaron cuatro modelos para predecir el tipo de atención recibida por el paciente confirmado de dengue (0=ambulatorio, 1=hospitalizado): regresión logística con todas las variables (RL-TV), regresión logística con las variables significativas encontradas en el análisis bivariado (RL-VS), árboles de decisión usando el algoritmo C4.5 con todas las variables (AD-TV) y árboles de decisión usando el algoritmo C4.5 con las variables significativas encontradas en el análisis bivariado (AD-TV). Estos modelos se ajustaron dentro de una validación cruzada estratificada de 10 bloques. Se utilizó la semilla 1234 para determinar los elementos de cada bloque. Para cada bloque se calcularon las métricas de desempeño: exactitud, sensibilidad y especificidad de la clasificación, todas sobre el conjunto de prueba. Posteriormente, se calcularon los promedios y desviaciones estándar de cada métrica de desempeño por modelo, y se aplicó un análisis de varianza (ANOVA) de medidas repetidas para las métricas que cumplieron los supuestos de normalidad y homogeneidad de varianzas, con el fin de encontrar diferencias entre el promedio de las métricas según el modelo ajustado. Para probar los supuestos descritos previamente se usaron las pruebas de Shapiro-Wilk y de Mauchly, respectivamente.

Con el resultado de las pruebas anteriores, se eligió el mejor modelo, se tomó el bloque que obtuvo las mayores métricas de exactitud, sensibilidad y especificidad de la clasificación en el conjunto de prueba, y se describió el modelo completo a través de los coeficientes $\hat{\beta}$ junto con su error estándar, la prueba de su significancia (Z , valor p), así como los *odds ratio* (OR) con respecto a cada nivel de referencia, junto con sus intervalos de confianza al 95%.

Todas las pruebas estadísticas se contrastaron con un nivel de significancia $\alpha=0.05$. Los resultados se presentan en tablas y gráficos. El análisis estadístico se realizó en Microsoft Excel y R versión 4.5.0 (R Core Team, 2025), en un equipo de cómputo Asus TUF Gaming A15 con memoria RAM de 8GB y procesador AMD Ryzen 5 (7,000 series).

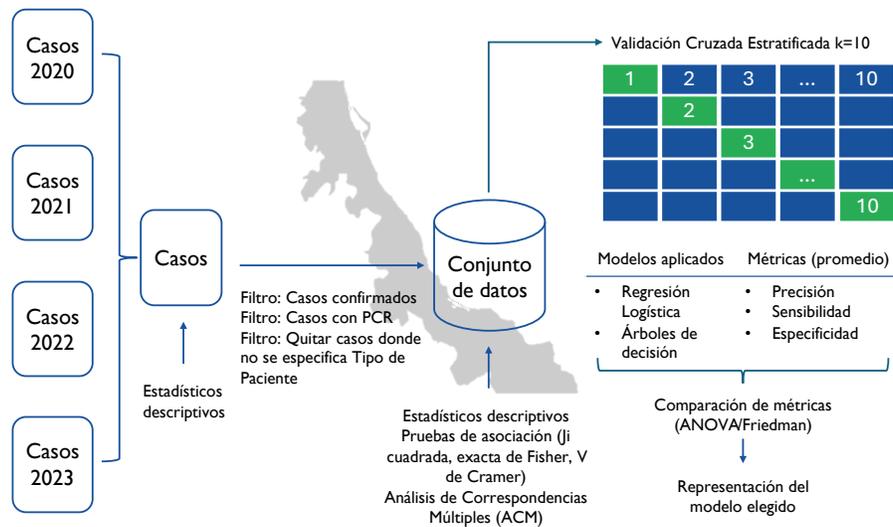


Figura 2. Preprocesamiento y análisis estadístico de los datos.
Referencia: Elaboración propia.

3.5. Descripción de las técnicas estadísticas a utilizar

A continuación, se describen las pruebas y técnicas estadísticas aplicadas en el estudio por orden de uso: prueba J_i cuadrada de homogeneidad, prueba J_i cuadrada para independencia de variables, prueba exacta de Fisher, coeficiente de asociación V de Cramer (v), Análisis de Correspondencias Múltiples (ACM), validación cruzada por bloques y métricas de rendimiento, Regresión Logística, Árboles de decisión (algoritmo C4.5) y ANOVA de medidas repetidas.

3.5.1. Prueba Ji cuadrada (χ^2) para proporciones

La prueba χ^2 de homogeneidad se aplica para determinar si existen diferencias entre dos o más distribuciones, y se utiliza frecuentemente para comparar proporciones. Se basa en las hipótesis:

$$H_0: p_1 = p_2 = \dots = p_k$$

H_1 : Al menos una p_k es diferente

El estadístico de prueba es

$$\chi^2 = \sum_{i=1}^k \frac{(x_i - n_i \hat{p})^2}{n_i \hat{p} (1 - \hat{p})}$$

Donde $\hat{p} = \frac{x_1 + \dots + x_k}{n_1 + \dots + n_k}$, x_i es el número de éxitos en cada grupo k y n_i el tamaño de muestra en cada grupo k . Este estadístico sigue una distribución de probabilidad Ji cuadrada con $k - 1$ grados de libertad. El valor p obtenido de la prueba se compara con un nivel de significancia α definido por el investigador. Si el valor $p < \alpha$ se rechaza la hipótesis nula, indicando que al menos una de las proporciones estimadas es diferente (Freund, Miller y Miller, 2000).

3.5.2. Prueba Ji cuadrada (χ^2) de independencia

La prueba Ji cuadrada para independencia consiste en evaluar la asociación entre dos variables cualitativas que pueden representarse mediante una tabla de contingencia, donde las filas (m) y columnas (n) indican las categorías de las variables y las celdas las frecuencias observadas. Esta prueba se aplica en tablas de contingencia $m \times n$, contrastando las frecuencias observadas con las frecuencias esperadas (Quevedo, 2011). La prueba parte de un juego de hipótesis:

H_0 : Las variables son independientes.

H_1 : Existe una dependencia entre las variables.

Posteriormente se calcula el estadístico de prueba:

$$\chi^2 = \sum \frac{(O_i - e_i)^2}{e_i}$$

donde O_i son las frecuencias observadas y e_i son las frecuencias esperadas.

Una vez obtenido el estadístico de prueba y su valor p correspondiente, el valor p se compara contra un nivel de significancia α definido por el investigador. Si $p < \alpha$ se rechaza la hipótesis nula, mostrando evidencia de que las variables son dependientes o tienen asociación.

3.5.3. Prueba exacta de Fisher

Esta prueba es la alternativa a la prueba Ji cuadrada cuando las frecuencias esperadas de la tabla de contingencias entre las dos variables de interés son menores a 5, incluso el juego de hipótesis a probar es el mismo. La prueba de Fisher calcula la probabilidad de ocurrencia de todas las posibles tablas de contingencia manteniendo los totales marginales fijos, a través de la distribución hipergeométrica multivariada:

$$P(X) = \frac{\prod_{j=1}^C \frac{C_j!}{x_{1j}! x_{2j}! \dots x_{rj}!}}{\frac{T!}{R_1! R_2! \dots R_r!}}$$

donde $T = \sum_{i=1}^r R_i$ es el total global de observaciones en la tabla, C es el total marginal de las columnas y R es el total marginal de las filas. El cálculo del valor p se realiza sumando todas las probabilidades que sean menores o iguales a la tabla observada, y éste se compara con un nivel de significancia definido por el investigador, conocido como α . Para rechazar la hipótesis nula, debe cumplirse que $p < \alpha$, concluyendo que las variables muestran una dependencia estadísticamente significativa (Mehta y Patel, 1983).

3.5.4. Coeficiente de asociación V de Cramer (v)

Este coeficiente se usa para medir la magnitud de la asociación entre dos variables cualitativas. Se recomienda aplicarlo después de las pruebas Ji cuadrada y exacta de Fisher. El cálculo del coeficiente es:

$$v = \left(\frac{\chi^2}{\chi_{max}^2} \right)^{\frac{1}{2}}$$

donde χ^2 es la estadística Ji cuadrada obtenida de la tabla de contingencia observada, y χ_{max}^2 es el valor máximo posible de Ji cuadrada según la tabla obtenida y el tamaño de muestra. Este coeficiente va de 0 a 1, donde 0 indica que no hay asociación entre las variables y 1 indica que la asociación entre estas es perfecta (Acock y Stavig, 1979).

3.5.5. Análisis de Correspondencias Múltiples

El Análisis de Correspondencias Múltiples (ACM) es una técnica descriptiva multivariante de interdependencia utilizada para la reducción dimensional y percepción de asociaciones entre variables. Consiste en determinar la “correspondencia” entre categorías de variables cualitativas, nominales u ordinales, representadas en un gráfico de dos dimensiones (Biplot) que se construye mediante el cálculo de similitud entre las categorías por medio de una tabla de contingencias multientrada (Hair *et al.*, 1999) y la descomposición de valores singulares (SVD), que produce las coordenadas de cada categoría.

Partiendo de una matriz indicadora \mathbf{Z} , que tiene tantas filas como sujetos y tantas columnas como categorías, en las celdas contiene ceros excepto donde los sujetos tienen la característica; se construye la matriz de Burt $\mathbf{C} = \mathbf{Z}^T \mathbf{Z}$ y se aplica una SVD con el procedimiento:

$$\sqrt{n} \frac{\mathbf{Z}}{Q_n} \mathbf{D}^{-\frac{1}{2}} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T$$

donde n es el tamaño de muestra, \mathbf{Z} es la matriz indicadora, Q es el número de variables categóricas, y \mathbf{D} es la matriz diagonal de la matriz de Burt \mathbf{C} . Con esta descomposición se obtienen los vectores singulares \mathbf{U} para los individuos y \mathbf{V} para las categorías de las variables. Los valores singulares elevados al cuadrado se conocen como la inercia correspondiente a cada dimensión (σ^2), y define la variabilidad explicada por cada dimensión construida (Greenacre y Blasius, 2006).

3.5.6. Validación cruzada por bloques y rendimiento de un clasificador

Esta técnica consiste en dividir el conjunto de datos en k subconjuntos diferentes de igual tamaño aproximadamente (Figura 3). El modelo se ajusta con los $k - 1$ subconjuntos, que sirven como muestra de entrenamiento. Posteriormente, el modelo es validado con el subconjunto resultante, que es conocido como muestra de prueba. Este proceso se repite hasta que todos los subconjuntos hayan servido como muestra de prueba. Por lo tanto, se obtienen k métricas de clasificación, que se conocen como medidas de desempeño de validación cruzada (Berrar, 2019).

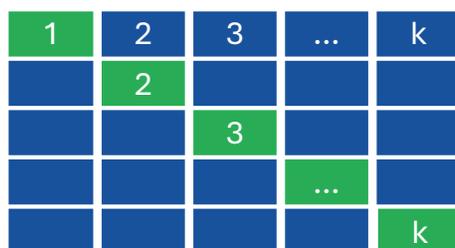


Figura 3. Proceso de validación cruzada por bloques. En color azul se resaltan los bloques que servirán como conjunto de entrenamiento, y en verde aquellos que servirán como conjunto de prueba.

Referencia: Elaboración propia.

Para cada modelo ajustado con conjunto de prueba se puede medir el desempeño de la clasificación a través de una matriz de confusión representada como una tabla de contingencia, donde las filas indican las clases observadas y las columnas las clases predichas por el modelo ajustado (Tabla 4).

Tabla 4. Matriz de confusión.			
		Valores predichos	
		Positivo	Negativo
Valores observados	Positivo	Verdadero Positivo (VP)	Falso Negativo (FN)
	Negativo	Falso Positivo (FP)	Verdadero Negativo (VN)

Referencia: Elaboración propia. Adaptado de “Data Science and Predictive Analytics, Biomedical and Health Applications using R” por Dinov. 2018. Springer. p. 478.

Las métricas de clasificación más usadas son la exactitud o *accuracy*, que se entiende como la proporción de clasificación correcta $\left(\frac{VP+VN}{VP+FN+FP+VN}\right)$; la sensibilidad o *sensitivity* es

la proporción de elementos positivos clasificados correctamente $\left(\frac{VP}{VP+FN}\right)$; por último, la especificidad o *specificity* $\left(\frac{VN}{FP+VN}\right)$ explica la proporción de elementos negativos clasificados correctamente (Dinov, 2018).

3.5.7. Regresión Logística

La regresión logística binaria se aplica cuando se busca establecer una relación entre una variable dependiente dicotómica y (éxito o fracaso) y una o más variables independientes x_1, \dots, x_p . Teniendo el modelo lineal $y_i = X_i^T \beta + e_i$ (donde β son los parámetros del modelo y e_i son los errores), y asumiendo que $y_i \sim \text{Bernoulli}(\pi_i)$, debemos ajustar el modelo de manera que se puedan obtener valores de 0 a 1. Este problema se resuelve al introducir la función de regresión logística:

$$\pi_i = E(y_i) = \frac{e^{(X^T \beta)}}{1 + e^{(X^T \beta)}}$$

donde $X^T \beta$ es el predictor lineal. Esta función se distribuye con forma sigmoide (Figura 4), con lo cual se pueden obtener predicciones como probabilidades en el rango (0,1).

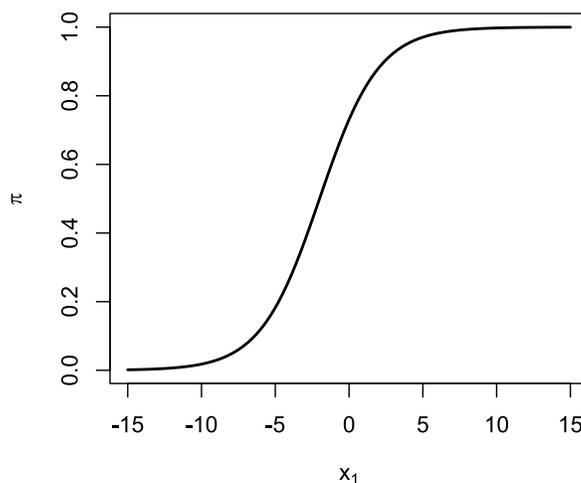


Figura 4. Función de regresión logística (sigmoide).
Referencia: Elaboración propia.

Para estimar los parámetros β se aplica el método de máxima verosimilitud, función que encuentra su máximo a través de métodos iterativos como el método de Newton-Raphson. Una vez estimados los parámetros, se estima el modelo final para explicar la

relación encontrada con respecto a la variable respuesta. Dado que el modelo de regresión logística estima probabilidades, es posible determinar un punto de corte con el fin de clasificar a cada sujeto en uno de los dos posibles valores y . El punto de corte más usado es 0.5, pero puede ajustarse en términos de costos y errores de clasificación (Chung, 2020).

Para probar la significancia de los coeficientes de regresión $\hat{\beta}_j$ se utiliza el estadístico de Wald, bajo las hipótesis:

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0$$

Este estadístico se obtiene elevando al cuadrado el cociente entre $\hat{\beta}_j$ y su error estándar, y tiene una distribución Ji cuadrada con 1 grado de libertad. Al contrastar el valor p asociado al estadístico de Wald contra un valor de significancia α definido por el investigador, se rechaza H_0 siempre que $p < \alpha$, evidenciando que el coeficiente de regresión es significativo (Pardo y Ruiz, 2013). Para interpretar los coeficientes obtenidos, es necesario aplicar la transformación $e^{\hat{\beta}_j}$, obteniendo $odds(y = 1)$, que indican la razón de probabilidad de éxito de la variable dependiente.

3.5.8. Árboles de decisión

Un árbol de decisión es una partición recursiva del espacio de instancias (Figura 5). Un árbol se compone de nodos formados por los atributos, ramas conformadas por los valores que puede tomar cada atributo, y hojas (posibles valores de la variable dependiente).

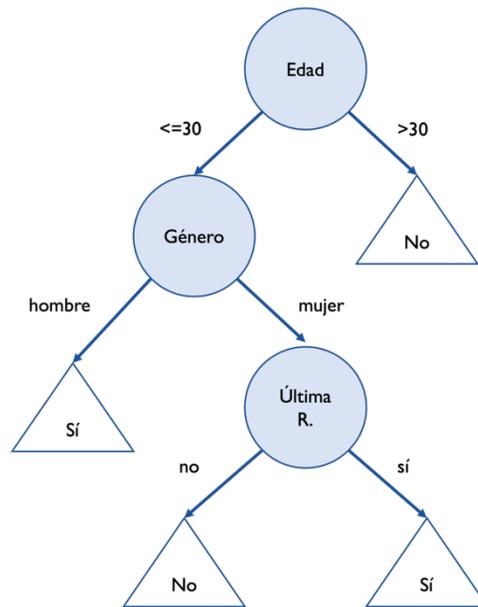


Figura 5. Ejemplo de árbol de decisiones de respuesta directa al correo.
 Referencia: Elaboración propia. Adaptada de “Decision trees” por Rokach, L. y Maimon, O. 2005. Data mining and knowledge discovery handbook. p. 165-192.

El procedimiento consiste en definir, mediante una métrica determinada, el mejor atributo que servirá como nodo raíz. Posteriormente, se divide el espacio en dos o más subconjuntos de acuerdo con una función de los posibles valores que puede tomar dicho atributo. A cada hoja se le asigna la clase con mayor frecuencia en el subconjunto. Este proceso se repite recursivamente hasta que el árbol no se pueda dividir más. Existen distintos algoritmos para crear árboles, los más conocidos son: ID3, C4.5, CART y CHAID (Rokach y Maimon, 2005).

El algoritmo C4.5 divide el conjunto de entrenamiento T mediante la *ratio* de ganancia. Para cada atributo S se obtiene la métrica y se busca aquel que tenga la mejor *ratio*, el cual servirá como nodo raíz. El árbol se construye dividiendo T en tantos subconjuntos como valores v tenga el atributo S , y para cada subconjunto el proceso se repite recursivamente, hasta que el árbol no se pueda dividir más, etiquetando las hojas con la clase C con mayor frecuencia en cada subconjunto (Dai y Ji, 2014).

$$RatioDeGanancia(S,T) = \frac{Ganancia(S,T)}{Información(S,T)}$$

donde

$$Ganancia(S,T) = Entropía(S) - \sum_{v \in values(T_s)} \frac{T_{S,y}}{T_S} Entropía(S_y)$$

$$Información = - \sum_{v \in values(T_s)} \frac{|T_{S,y}|}{|T_S|} \cdot \log \frac{|T_{S,y}|}{|T_S|}$$

$$Entropía = - \sum_{j=1}^C p(S,j) \cdot \log p(S,j)$$

Al obtener el árbol de decisión, se pueden construir reglas de decisión para clasificar a un nuevo individuo.

3.5.9. Análisis de varianza de una vía con medidas repetidas

Se define como la evaluación de los efectos de un factor en la variable respuesta cuando las mediciones se realizan con los mismos sujetos, es decir, la característica principal es que las mediciones no son independientes entre sí. La hipótesis que se prueba es la siguiente:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_j$$

H_1 : Al menos una μ_j es diferente.

En la práctica, consiste en determinar las diferencias en las medias entre más de dos grupos, a través del estadístico F que se distribuye con $\frac{j-1}{(j-1)(n-1)}$ grados de libertad, donde j es el número de grupos y n es el tamaño de la muestra (Pardo y San Martín, 2010). Este estadístico de prueba se calcula mediante la tabla de análisis de varianza (Tabla 5) y al obtener el valor p asociado, éste se contrasta con un nivel de significancia α definido por el investigador. Si $p < \alpha$, se rechaza H_0 , demostrando que al menos una de las medias es estadísticamente diferente.

Tabla 5. Tabla de análisis de varianza (ANOVA) de una vía con medidas repetidas.

Fuente de Variación	Suma de Cuadrados	Grados de libertad	Cuadrados medios	F
Grupos	SCTG	j-1	CMG = SCG / j-1	F = CMG/CME
Error	SCE	(j-1) (n-1)	CME = SCE / [(j-1) (n-1)]	
Total	SCT			

Referencia: Elaboración propia. Adaptado de “Análisis y diseño de experimentos” por Gutiérrez y De la Vara. 2008. McGraw-Hill Interamericana. p. 69.

Para realizar el análisis de varianza es necesario que el conjunto de datos cumpla con los supuestos de normalidad y homogeneidad de varianzas.

Supuesto de normalidad (Prueba de Shapiro-Wilk)

Esta prueba se utiliza para determinar si una muestra se comporta como una distribución normal en muestras pequeñas (menores a 50 casos), y es clave en el análisis previo a un ANOVA. Parte de la hipótesis:

H_0 : Los datos siguen una distribución normal

H_1 : Los datos no siguen una distribución normal

El estadístico de prueba a calcular es

$$W = \frac{\sum_{i=1}^n \alpha y_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

donde Y_i son los elementos de la muestra ordenados por tamaño, \bar{Y} es la media de la muestra y α son los coeficientes de cada dato ordenado y_i , obtenido mediante los valores esperados y las covarianzas de los estadísticos de orden (Flores y Flores, 2021). El valor p asociado se contrasta con un valor de significancia α definido por el investigador, y se rechaza H_0 cuando $p < \alpha$, sin embargo, en la práctica se busca que $p > \alpha$, para asumir como verdadera la hipótesis nula, indicando que la muestra sigue una distribución normal.

Supuesto de homogeneidad de Varianzas (Prueba de Mauchly)

También llamada prueba de esfericidad, se aplica para determinar si las varianzas de las p medidas repetidas son estadísticamente iguales. Consiste en comparar las hipótesis:

H_0 : La matriz de covarianzas de las medidas repetidas es esférica.

H_1 : La matriz de covarianzas de las medidas repetidas no es esférica.

El estadístico de prueba es

$$W = \frac{(p - 1)^{p-1} |CSC^T|}{(trCSC^T)^{p-1}}$$

donde p es el número de medidas repetidas, C es una matriz cuyas filas son los contrastes ortogonales normalizados de las p medidas repetidas y S es la matriz de varianzas y covarianzas de los residuos intra-sujeto (Kuehl, 2001). El valor p asociado al estadístico W se compara con un nivel de significancia α definido por el investigador y, así como en la prueba de Shapiro-Wilk, se busca que $p > \alpha$ para no rechazar H_0 , y verificar que la matriz de varianzas y covarianzas de las medidas repetidas es esférica y, por consiguiente, cumplir con el supuesto de homogeneidad de varianzas.

3.6. Justificación de la elección de estadísticos

La elección de las técnicas mencionadas se debió a la naturaleza de los datos. Para utilizar la prueba Ji cuadrada, prueba exacta de Fisher y el estadístico V de Cramer, el requerimiento principal es que las variables a utilizar sean de tipo cualitativo. También se aplicó un análisis de correspondencias múltiples debido a que todas las variables son de tipo cualitativo. Para aplicar un modelo de regresión logística binaria, la condición es que la variable dependiente sea de tipo cualitativo binario; en este caso se cumple, dado que todas las variables independientes son de tipo cualitativo, así como la dependiente, que además solo cuenta con dos posibles valores: ambulatorio y hospitalizado. De la misma forma, se usó el algoritmo C4.5 debido a que trabaja con variables cualitativas, aunque esta técnica puede soportar atributos numéricos. Finalmente, se aplicó ANOVA de medidas repetidas porque

las métricas de clasificación obtenidas son continuas y cumplen con el supuesto de normalidad (prueba de Shapiro-Wilk) y homogeneidad de varianzas (Prueba de Mauchly), y las métricas son obtenidas con el mismo bloque y al mismo tiempo.

4. Resultados

4.1. Características de los casos de dengue

Durante el período de estudio, se registraron 57,177 casos relacionados con dengue. 14.23% fueron durante el año 2020 (8,137 casos), 6.29% durante el año 2021 (3,595 casos), 10.25% durante 2022 (5,863 casos) y 69.23% durante 2023 (39,582 casos; Figura 6).

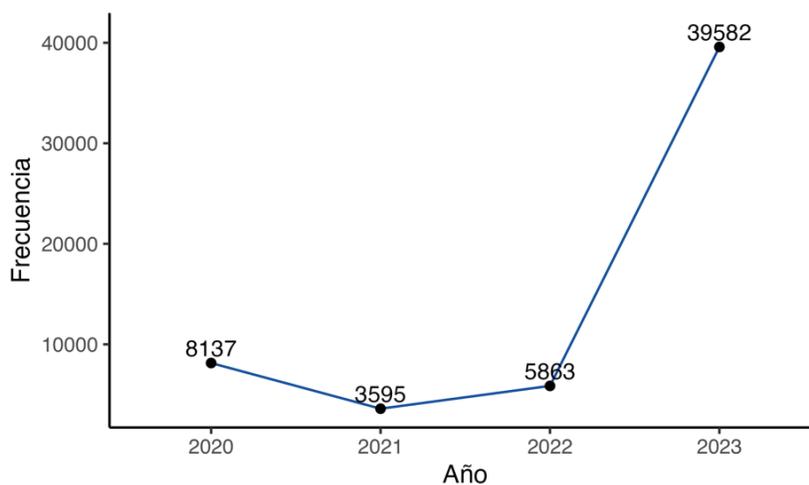


Figura 6. Distribución de todos los casos de dengue por año.
Referencia: Elaboración propia.

La edad promedio de los casos registrados fue de 30 años, con una desviación estándar de 19 años. De los casos registrados en todo el período, 14,511 fueron casos confirmados (25.38%), 14,548 fueron descartados (25.44%) y 28,118 (49.18%) se mantuvieron como probables casos de dengue (Figura 7).

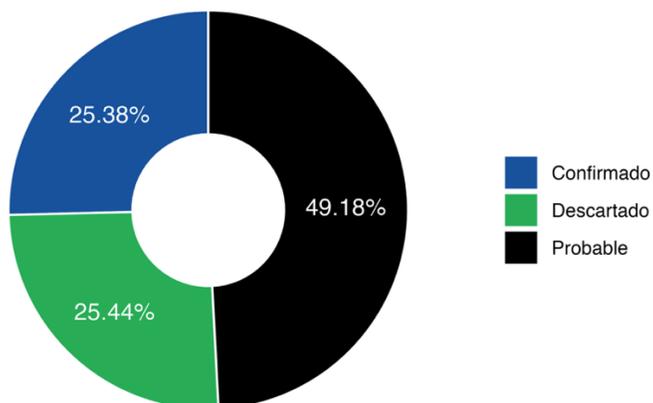


Figura 7. Distribución de todos los casos por estatus.
Referencia: Elaboración propia.

La proporción de casos confirmados durante 2020 es de $\hat{p}_{2020}=0.182$, es decir, 1,480 casos; para 2021 se confirmaron 1,009 casos, es decir, una proporción igual a $\hat{p}_{2021}=0.281$; durante 2022 se registraron 1,740 casos confirmados ($\hat{p}_{2022}=0.297$); para 2023 se registraron 10,282 casos confirmados, es decir, una proporción igual a $\hat{p}_{2023}=0.260$ (Tabla 6).

Tabla 6. Casos totales de dengue y casos confirmados por año.

Año	Casos	Casos positivos	Proporción
2020	8137	1480	0.182
2021	3595	1009	0.281
2022	5863	1740	0.297
2023	39582	10282	0.260
Total	57177	14511	

Referencia: Elaboración propia.

De 2020 y 2021, se encontró una diferencia en la proporción de casos confirmados igual a 0.099 (Tabla 7). Las hipótesis fueron: $H_0: p_{2020} = p_{2021}$ contra $H_1: p_{2020} \neq p_{2021}$. El estadístico de prueba fue igual a $\chi^2=144.970$ y un valor $p<0.001$, por lo que se rechaza la hipótesis nula, indicando diferencias estadísticamente significativas entre las proporciones de 2020 y 2021.

De los años 2020 y 2022, la diferencia en la proporción de casos confirmados de dengue fue de 0.115 (Tabla 7). Las hipótesis fueron: $H_0: p_{2020} = p_{2022}$ contra $H_1: p_{2020} \neq p_{2022}$. La prueba Ji cuadrada obtuvo un estadístico igual a $\chi^2=253.340$ con un valor $p<0.001$, por lo que es evidencia para rechazar la hipótesis nula, y afirmar que existen diferencias estadísticamente significativas en las proporciones de casos confirmados de 2020 y 2022.

Para los años 2020 y 2023, la diferencia en la proporción de casos confirmados fue igual a 0.078 (Tabla 7). Las hipótesis fueron $H_0: p_{2020} = p_{2023}$ contra $H_1: p_{2020} \neq p_{2023}$. Se encontró un estadístico de prueba igual a $\chi^2=219.990$ con un valor p asociado igual a $p<0.001$, por lo que existe evidencia para rechazar H_0 , e indicar que existen diferencias estadísticamente significativas en la proporción de casos confirmados de dengue de estos años.

Para los años 2021 y 2022, las hipótesis contrastadas fueron $H_0: p_{2021} = p_{2022}$ contra $H_1: p_{2021} \neq p_{2022}$. La diferencia obtenida entre las proporciones fue de 0.016 (Tabla 7). La prueba Ji cuadrada obtuvo un estadístico igual a $\chi^2=2.727$ y un valor $p=0.099$, con lo que no se rechaza H_0 , es decir, no se encontraron diferencias estadísticamente significativas entre las proporciones de casos confirmados de dengue de 2021 y 2022.

De los años 2021 y 2023, se encontró una diferencia en las proporciones de -0.021 (Tabla 7). Las hipótesis fueron $H_0: p_{2021} = p_{2023}$ contra $H_1: p_{2021} \neq p_{2023}$. La prueba Ji cuadrada obtuvo un estadístico $\chi^2=7.349$ y un valor $p=0.007$, por lo que se tiene evidencia a favor de la hipótesis alterna, encontrando diferencias estadísticamente significativas entre las proporciones de casos confirmados de dengue de 2021 y 2023.

Por último, para los años 2022 y 2023, la diferencia entre las proporciones fue de -0.037 (Tabla 7). Las hipótesis que se contrastaron fueron $H_0: p_{2022} = p_{2023}$ contra $H_1: p_{2022} \neq p_{2023}$. Con la prueba Ji cuadrada, el estadístico resultó en $\chi^2=35.765$, con un valor p asociado igual a $p<0.001$, por lo que se rechaza la hipótesis nula, indicando que existen diferencias estadísticamente significativas entre la proporción de casos confirmados de dengue para los años 2022 y 2023.

Tabla 7. Prueba de proporciones de casos confirmados por año.			
Años	Diferencia	χ^2	p
2020-2021	0.099	144.970	<0.001
2020-2022	0.115	253.340	<0.001
2020-2023	0.078	219.990	<0.001
2021-2022	0.016	2.727	0.099
2021-2023	-0.021	7.349	0.007
2022-2023	-0.037	35.765	<0.001

Nota: Nivel de significancia $\alpha = 0.05$.
Referencia: Elaboración propia.

Del total de casos registrados en 2020, 49.22% fueron mujeres y 50.78% fueron hombres; durante el año 2021, 47.32% de los casos fueron del sexo femenino y 52.68% del sexo masculino; en 2022 el porcentaje de casos femeninos registrados fue del 45.51%, mientras

que el porcentaje de los casos masculinos registrados fue de 54.49%; en 2023, del total de casos registrados, 45.35% fueron mujeres y 54.65% fueron hombres (Figura 8).

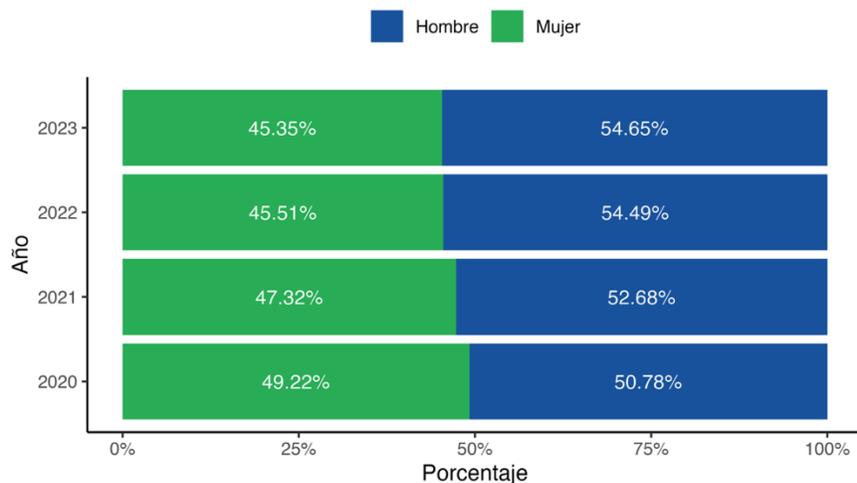


Figura 8. Distribución de todos los casos de dengue por año y sexo.

Referencia: Elaboración propia.

De los 14,511 casos confirmados, 4,273 casos fueron diagnosticados por una prueba PCR. 8.12% del total de casos tuvo el serotipo DENV-1 (347 casos), 63.52% fue diagnosticado con el serotipo DENV-2 (2,714 casos), 22.51% fue diagnosticado con el serotipo DENV-3 (962 casos) y 5.85% presentó el serotipo DENV-4 (250 casos).

4.2. Tipo de atención médica recibida y características clínicas

Se encontró que 2,246 (52.56%) casos confirmados por prueba PCR tuvieron un tipo de atención médica ambulatoria, mientras que 2,027 (47.44%) pacientes fueron hospitalizados. Se contrastaron las hipótesis:

H_0 : El tipo de atención médica recibida es estadísticamente independiente de la característica clínica x_j .

H_1 : El tipo de atención médica recibida es estadísticamente dependiente de la característica clínica x_j .

Donde x_j son las características clínicas del paciente: sexo, edad, paciente hemorrágico, presencia de diabetes, hipertensión, enfermedad UCL péptica, enfermedad

renal, inmunosupresión, cirrosis hepática, embarazo, resultado de la prueba PCR y unidad médica de registro.

El 54.36% de los casos ambulatorios fueron hombres, contra 45.64% de casos del sexo femenino, en tanto que el 56.44% de los casos hospitalizados fueron hombres y 43.56% fueron mujeres (Tabla 8); con la prueba Ji cuadrada se obtuvo un valor $p=0.183$, lo que evidencia que no se rechaza la hipótesis nula, y se deduce que el sexo y el tipo de atención recibida por el paciente son variables estadísticamente independientes.

Tabla 8. Atención médica recibida por sexo.						
Sexo	Ambulatorio <i>f</i> (%)		Hospitalizado <i>f</i> (%)		p_{χ^2}	v
Hombre	1221	54.36	1144	56.44	0.183	0.021
Mujer	1025	45.64	883	43.56		
Total	2246	100	2027	100		
Nota: Nivel de significancia $\alpha=0.05$. Referencia: Elaboración propia.						

De igual manera, la mayoría de los casos confirmados de dengue se encontraron en los rangos de edad de 0 a 20 años (33.26% en los casos ambulatorios y 48.30% en los casos hospitalizados) y de 21 a 40 años (40.47% en los pacientes ambulatorios y 28.27% en los pacientes hospitalizados); el resultado de la prueba Ji cuadrada obtuvo un valor $p<0.001$ y un coeficiente de asociación $v=0.183$, con lo que se rechaza la hipótesis nula, y se determina que la edad de los casos confirmados está asociada estadísticamente y a un nivel débil a moderado con el tipo de atención recibida (Tabla 9).

Tabla 9. Atención médica recibida por grupos de edad.						
Edad	Ambulatorio		Hospitalizado		p_{χ^2}	v
	f	(%)	f	(%)		
0 a 20 años	747	33.26	979	48.30	<0.001	0.183
21 a 40 años	909	40.47	573	28.27		
41 a 60 años	451	20.08	301	14.85		
60 a 80 años	132	5.88	142	7.01		
Más de 80 años	7	0.31	32	1.58		
Total	2246	100	2027	100		

Nota: Nivel de significancia $\alpha=0.05$.
Referencia: Elaboración propia.

También, 0.71% de los pacientes ambulatorios fueron pacientes hemorrágicos, contra un 0.89% de pacientes hospitalizados hemorrágicos; con un valor $p=0.636$ asociado a la prueba Ji cuadrada, se evidencia que los pacientes hemorrágicos y el tipo de atención son estadísticamente independientes, ya que no se rechaza la hipótesis nula (Tabla 10).

Tabla 10. Atención médica recibida por paciente hemorrágico.						
Paciente hemorrágico	Ambulatorio		Hospitalizado		p_{χ^2}	v
	f	(%)	f	(%)		
No	2230	99.29	2009	99.11	0.636	0.010
Sí	16	0.71	18	0.89		
Total	2230	99.29	2009	99.11		

Nota: Nivel de significancia $\alpha=0.05$.
Referencia: Elaboración propia.

Por otro lado, 2.45% de los pacientes con tipo de atención ambulatoria padecen de diabetes, contra 5.03% de pacientes hospitalizados con la misma enfermedad; de la prueba Ji cuadrada se obtuvo un valor $p<0.001$, y un coeficiente $v=0.069$, con lo cual existe evidencia suficiente para rechazar la hipótesis nula, y afirmar que la presencia de diabetes y el tipo de atención recibida son variables estadísticamente dependientes, aunque la asociación entre ellas es débil (Tabla 11).

Tabla 11. Atención médica recibida por presencia de diabetes.						
Diabetes	Ambulatorio		Hospitalizado		p_{χ^2}	v
	f	(%)	f	(%)		
No	2191	97.55	1925	94.97	<0.001	0.069
Sí	55	2.45	102	5.03		
Total	2246	100.00	2027	100.00		

Nota: Nivel de significancia $\alpha=0.05$.
Referencia: Elaboración propia.

Siguiendo esta línea, 1.96% de los pacientes ambulatorios padece de hipertensión, mientras que 4.93% de los pacientes hospitalizados también lo padece (Tabla 12); la dependencia entre la presencia de hipertensión y el tipo de atención médica resultó estadísticamente significativa dados los resultados de la prueba Ji cuadrada y del coeficiente V de Cramer ($p<0.001$, $v=0.082$), aunque la asociación también es débil.

Tabla 12. Atención médica recibida por presencia de hipertensión.						
Hipertensión	Ambulatorio		Hospitalizado		p_{χ^2}	v
	f	(%)	f	(%)		
No	2202	98.04	1927	95.07	<0.001	0.082
Sí	44	1.96	100	4.93		
Total	2246	100.00	2027	100.00		

Nota: Nivel de significancia $\alpha=0.05$.
Referencia: Elaboración propia.

En contraste, se muestra que 0.04% de los pacientes con atención ambulatoria sufre de úlcera péptica, así como el 0.10% de los pacientes hospitalizados (Tabla 13); pero esta asociación no resultó significativa, pues no se rechaza la hipótesis nula según lo obtenido por la prueba exacta de Fisher ($p=0.607$).

Tabla 13. Atención médica recibida por presencia de enfermedad UCL péptica.						
Enfermedad UCL péptica	Ambulatorio		Hospitalizado		p_{Fisher}	ν
	f (%)	f (%)	f (%)	f (%)		
No	2245	99.96	2025	99.90	0.607	0.010
Sí	1	0.04	2	0.10		
Total	2246	100.00	2027	100.00		

Nota: Nivel de significancia $\alpha=0.05$.
Referencia: Elaboración propia.

También se observa que el 0.09% de los pacientes atendidos de forma ambulatoria padece de alguna enfermedad renal, y 1.23% de los pacientes que fueron hospitalizados también sufren de ésta (Tabla 14); en la prueba Ji cuadrada se obtuvo un valor $p < 0.001$, así como un coeficiente $\nu = 0.072$, por lo que se rechaza la hipótesis nula y se afirma que la enfermedad renal y el tipo de paciente son variables estadísticamente dependientes, con un nivel de asociación débil.

Tabla 14. Atención médica recibida por presencia de enfermedad renal.						
Enfermedad renal	Ambulatorio		Hospitalizado		p_{χ^2}	ν
	f (%)	f (%)	f (%)	f (%)		
No	2244	99.91	2002	98.77	<0.001	0.072
Sí	2	0.09	25	1.23		
Total	2246	100.00	2027	100.00		

Nota: Nivel de significancia $\alpha=0.05$.
Referencia: Elaboración propia.

El 0.13% de los pacientes ambulatorios registró inmunosupresión, así como 0.20% de los pacientes hospitalizados (Tabla 15); no obstante, esta asociación no resultó significativa dado el valor p de la prueba exacta de Fisher ($p=0.715$).

Tabla 15. Atención médica recibida por presencia de inmunosupresión.						
Inmunosupresión	Ambulatorio		Hospitalizado		p_{Fisher}	ν
	f (%)	f (%)	f (%)	f (%)		
No	2243	99.87	2023	99.80	0.715	0.008
Sí	3	0.13	4	0.20		
Total	2246	100.00	2027	100.00		

Nota: Nivel de significancia $\alpha=0.05$.
Referencia: Elaboración propia.

Asimismo, sólo 0.04% de los pacientes con tipo de atención ambulatoria presentaron cirrosis hepática, contra 0.30% de pacientes hospitalizados que también la presentaron (Tabla 16); a pesar de esta diferencia, la asociación entre padecer cirrosis hepática y el tipo de atención recibida no resultó significativa según la prueba exacta de Fisher, puesto que no se rechazó la hipótesis nula ($p=0.058$).

Tabla 16. Atención médica recibida por presencia de cirrosis hepática.						
Cirrosis hepática	Ambulatorio		Hospitalizado		p_{Fisher}	v
	f	(%)	f	(%)		
No	2245	99.96	2021	99.70	0.058	0.031
Sí	1	0.04	6	0.30		
Total	2246	100.00	2027	100.00		

Nota: Nivel de significancia $\alpha=0.05$.
Referencia: Elaboración propia.

Por el contrario, 1.74% de las pacientes confirmadas de dengue con tratamiento ambulatorio se encontraban embarazadas al momento de su diagnóstico, contra un 4.88% de las pacientes hospitalizadas que también presentaron embarazo (Tabla 17). Se obtuvo un valor $p<0.001$ en la prueba Ji cuadrada y un coeficiente de asociación $v=0.089$, con lo que se evidencia el rechazo de la hipótesis nula, y se define que la presencia de embarazo en pacientes confirmadas de dengue y el tipo de atención médica recibida son variables estadísticamente dependientes.

Tabla 17. Atención médica recibida por presencia de embarazo.						
Embarazo	Ambulatorio		Hospitalizado		p_{χ^2}	v
	f	(%)	f	(%)		
No	2207	98.26	1928	95.12	<0.001	0.089
Sí	39	1.74	99	4.88		
Total	2246	100.00	2027	100.00		

Nota: Nivel de significancia $\alpha=0.05$.
Referencia: Elaboración propia.

Con respecto a la prueba PCR, se observa que la mayoría de los casos confirmados tuvieron el serotipo DENV-2 (60.86% en los pacientes ambulatorios y 66.45% en los pacientes hospitalizados), seguido del serotipo DENV-3 con 23.46% del total de casos ambulatorios y 21.46% en los casos hospitalizados (Tabla 18); con un valor $p < 0.001$ según la prueba Ji cuadrada y un coeficiente de asociación $v = 0.064$, se rechaza la hipótesis nula y se verifica que existe una asociación estadísticamente significativa, pero débil, entre el serotipo detectado por la prueba PCR y el tipo de atención médica recibida.

Tabla 18. Atención médica recibida por resultado de la prueba PCR.						
Resultado de la prueba PCR	Ambulatorio		Hospitalizado		p_{χ^2}	v
	f	(%)	f	(%)		
DENV-1	208	9.26	139	6.86	<0.001	0.064
DENV-2	1367	60.86	1347	66.45		
DENV-3	527	23.46	435	21.46		
DENV-4	144	6.41	106	5.23		
Total	2246	100.00	2027	100.00		

Nota: Nivel de significancia $\alpha = 0.05$.
Referencia: Elaboración propia.

Finalmente, se encontró que la mayoría de los casos confirmados de dengue fueron registrados en el IMSS (64.96% en pacientes ambulatorios y 44.15% en pacientes hospitalizados) y en la SSA (27.60% en los casos ambulatorios y 44.60% en los casos hospitalizados), por lo que, con un valor $p < 0.001$ en la prueba de asociación Ji cuadrada y un coeficiente de asociación $v = 0.218$, se rechaza la hipótesis nula, con lo que se comprueba que la unidad médica de registro y el tipo de atención médica son estadísticamente dependientes, y esta asociación es de débil a moderada (Tabla 19).

Tabla 19. Atención médica recibida por unidad médica de registro del paciente.

Unidad médica de registro	Ambulatorio		Hospitalizado		p_{χ^2}	v
	f (%)		f (%)			
Otra	30	1.34	80	3.95	<0.001	0.218
IMSS	1459	64.96	895	44.15		
IMSS-BIENESTAR	42	1.87	59	2.91		
ISSSTE	95	4.23	89	4.39		
SSA	620	27.60	904	44.60		
Total	2246	100.00	2027	100.00		

Nota: Nivel de significancia $\alpha=0.05$.
Referencia: Elaboración propia.

4.3. Análisis de Correspondencias Múltiples

El análisis de correspondencias múltiples (ACM) explica un 15.71% del total de la varianza de los pacientes confirmados en las primeras dos dimensiones. Las categorías con mayor contribución a la dimensión 1 (9.09% de varianza explicada) son la presencia de diabetes (Diab), hipertensión (Hip) y enfermedad renal (Ren), las edades de 41 años en adelante (61-80, 80+ y 41-60) y si el paciente sufrió dengue hemorrágico (PH); por su parte, las categorías con mayor peso en la dimensión 2 (6.62%) son el tipo de atención médica (Ambulatorio y Hospitalizado), las unidades médicas SSA e IMSS, y las edades de 20 a 60 años (0-20, 21-40 y 41-60; Figura 9).

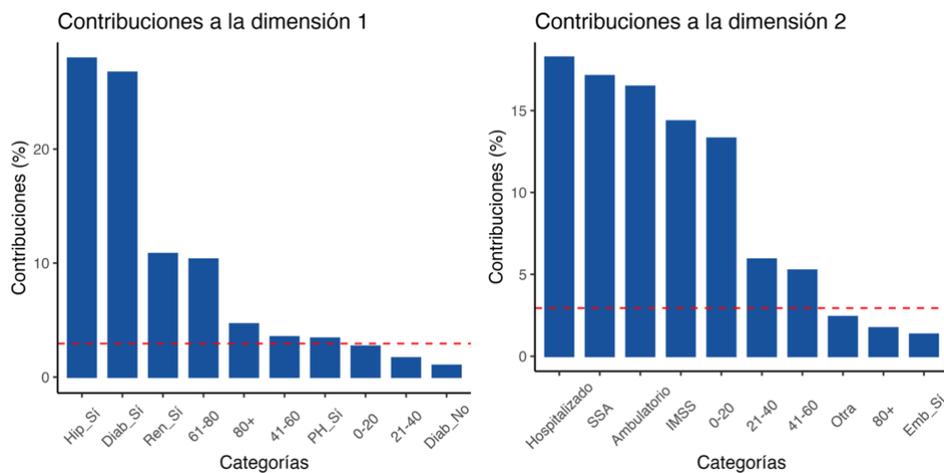


Figura 9. Contribuciones de las categorías a las dimensiones.

Referencia: Elaboración propia.

A nivel multivariado, se observan pocos grupos de categorías relacionadas en el Biplot (Figura 10). En el primer cuadrante del Biplot se identifica que la presencia de diabetes (Diab) se corresponde con la presencia de hipertensión (Hip) y padecer dengue hemorrágico (PH), así como los pacientes de 80 años y más (80+) corresponden a padecer enfermedad UCL péptica (UCL). También se observa que la mayoría de las categorías se encuentran cerca del origen, y según el color del gradiente, la mayor parte de estas tiene una contribución muy baja a las dos dimensiones. Con respecto al tipo de atención médica recibida, se identifica que los pacientes hospitalizados se encuentran relacionados con los registros de la SSA, con los pacientes de 20 años o menos (0-20), el serotipo DENV-1 y la presencia de embarazo (Emb). Por otro lado, los pacientes ambulatorios se asociaron con grupos de edad de 20 a 60 años (20-40, 41-60), registrados en el IMSS y con serotipo DENV-4. También se identifica que la enfermedad renal tiene la contribución más alta en las dos dimensiones, sin embargo, no muestra correspondencia con otras variables.

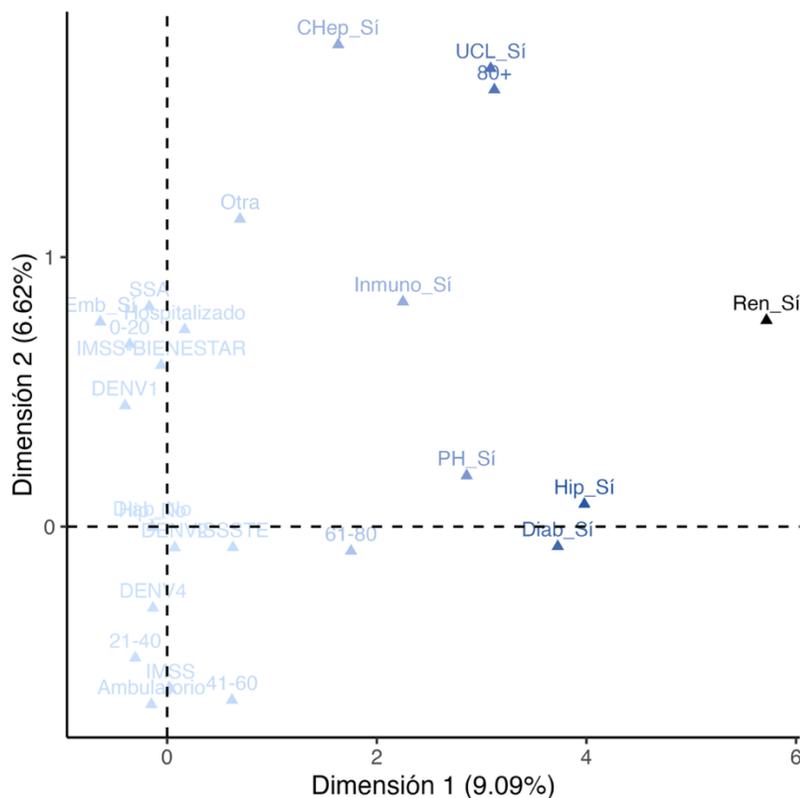


Figura 10. ACM para casos confirmados de dengue. Se visualizan 25 categorías con mayor contribución en las dos dimensiones.

Referencia: Elaboración propia.

4.4. Modelo predictivo

4.4.1. Selección del mejor modelo

Se ajustaron los modelos predictivos: RL-TV, AD-TV, RL-VS y AD-VS en 10 bloques de validación cruzada con una semilla igual a 1234.

Exactitud

Se determinó la prueba de comparaciones para la métrica de desempeño exactitud, mediante las hipótesis de los supuestos:

H_0 : Las métricas de exactitud en el modelo j siguen una distribución normal.

H_1 : Las métricas de exactitud en el modelo j no siguen una distribución normal.

H_0 : La matriz de covarianzas de la métrica de exactitud por modelo es esférica.

H_1 : La matriz de covarianzas de la métrica de exactitud por modelo es esférica.

Para todos los modelos ajustados se cumple la normalidad de las métricas de exactitud según las pruebas de Shapiro-Wilk (todos los $p > \alpha = 0.05$), mientras que la prueba de Mauchly obtuvo $p = 0.187$, por lo que no se rechaza H_0 indicando que el supuesto de homocedasticidad se cumple. Por lo tanto, se utilizó la prueba ANOVA de medidas repetidas para encontrar diferencias significativas entre las medias de las exactitudes por modelo (Tabla 20).

Tabla 20. Prueba a utilizar para comparar la exactitud por modelo según la verificación de los supuestos.

Modelo	Supuesto de normalidad $p_{Shapiro-Wilk}$	Supuesto de homocedasticidad $p_{Mauchly}$	Prueba a utilizar
RL-TV	0.514	0.187	ANOVA de medidas repetidas
AD-TV	0.574		
RL-VS	0.867		
AD-VS	0.427		

Nota: Nivel de significancia $\alpha = 0.05$.
Referencia: Elaboración propia.

Las hipótesis del modelo ANOVA fueron:

$$H_0: \mu_{RLTV} = \mu_{ADTV} = \mu_{RLVS} = \mu_{ADVS}$$

H_1 : Al menos una μ_j es diferente

El modelo que obtuvo la mejor exactitud promedio fue RL-VS (0.621±0.030), pero el análisis ANOVA obtuvo un estadístico $F=0.235$ con un valor p asociado $p=0.871$, por lo que no se rechaza la hipótesis nula, es decir, no existen diferencias significativas entre la exactitud promedio de los modelos (Tabla 21).

Tabla 21. Comparación de exactitud promedio por modelo según la prueba ANOVA.

Modelo				ANOVA de medidas repetidas
RL-TV Media (±DE)	AD-TV Media (±DE)	RL-VS Media (±DE)	AD-VS Media (±DE)	
0.617 (±0.031)	0.617 (±0.030)	0.621 (±0.030)	0.619 (±0.035)	$F=0.235, p=0.871$

Nota: Nivel de significancia $\alpha=0.05$. Se resalta en negritas el modelo con la mejor exactitud promedio. RL: Regresión Logística, AD: Árboles de Decisión, TV: Todas las Variables, VS: Variables Seleccionadas, DE: desviación estándar.
Referencia: Elaboración propia.

Sensibilidad

Para comparar la sensibilidad de los modelos se probaron las siguientes hipótesis sobre los supuestos:

H_0 : Las métricas de sensibilidad en el modelo j siguen una distribución normal.

H_1 : Las métricas de sensibilidad en el modelo j no siguen una distribución normal.

H_0 : La matriz de covarianzas de la métrica de sensibilidad por modelo es esférica.

H_1 : La matriz de covarianzas de la métrica de sensibilidad por modelo es esférica.

Se observa que todos los modelos cumplieron el supuesto de normalidad según la prueba de Shapiro-Wilk (todos los valores $p > \alpha = 0.05$), así como el supuesto de homocedasticidad de acuerdo con la prueba de Mauchly (valor $p=0.461$), por ello se aplicó

un ANOVA de medidas repetidas para encontrar diferencias significativas entre la sensibilidad promedio por modelo (Tabla 22).

Tabla 22. Prueba a utilizar para comparar la sensibilidad por modelo según la verificación de los supuestos.

Modelo	Supuesto de normalidad $p_{Shapiro-Wilk}$	Supuesto de homocedasticidad $p_{Mauchly}$	Prueba a utilizar
RL-TV	0.839	0.461	ANOVA de medidas repetidas
AD-TV	0.951		
RL-VS	0.975		
AD-VS	0.858		

Nota: Nivel de significancia $\alpha=0.05$.
Referencia: Elaboración propia.

El modelo que obtuvo mayor sensibilidad promedio fue AD-VS (0.592 ± 0.074), aunque también obtuvo la mayor desviación estándar; con el análisis ANOVA se probaron las siguientes hipótesis:

$$H_0: \mu_{RLTV} = \mu_{ADTV} = \mu_{RLVS} = \mu_{ADVS}$$

$$H_1: \text{Al menos una } \mu_j \text{ es diferente}$$

Se obtuvo un estadístico $F=6.020$ y un valor $p=0.003$, por lo que se rechaza la hipótesis nula y con ello, se asevera que existen diferencias significativas en la sensibilidad promedio según el modelo ajustado (Tabla 23).

Tabla 23. Comparación de sensibilidad promedio por modelo según la prueba ANOVA.

Modelo				ANOVA de medidas repetidas
RL-TV Media (\pm DE)	AD-TV Media (\pm DE)	RL-VS Media (\pm DE)	AD-VS Media (\pm DE)	
0.526 (± 0.055)	0.538 (± 0.053)	0.531 (± 0.046)	0.592 (± 0.074)	$F=6.020, p=0.003$

Nota: Nivel de significancia $\alpha=0.05$. Se resalta en negritas el modelo con la mejor sensibilidad promedio. RL: Regresión Logística, AD: Árboles de Decisión, TV: Todas las Variables, VS: Variables Seleccionadas, DE: desviación estándar.
Referencia: Elaboración propia.

En la Figura 11 se observa que el modelo AD-VS, que fue el mejor promedio de sensibilidad, es el que se diferencia significativamente del resto de modelos ajustados.

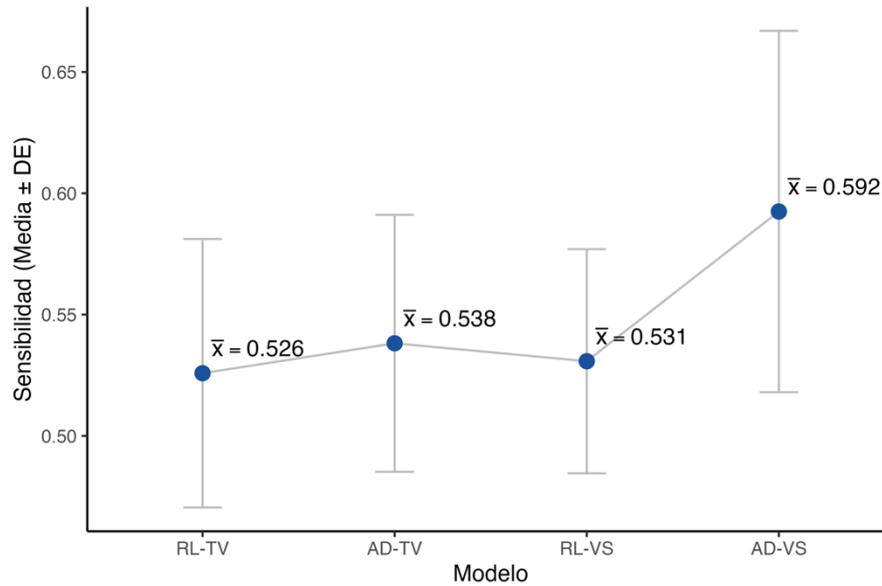


Figura 11. Comparación de sensibilidad promedio según modelos ajustados.
Referencia: Elaboración propia.

Especificidad

Por último, para verificar los supuestos de especificidad de los modelos, se probaron las hipótesis:

H_0 : Las métricas de especificidad en el modelo j siguen una distribución normal.

H_1 : Las métricas de especificidad en el modelo j no siguen una distribución normal.

H_0 : La matriz de covarianzas de la métrica de especificidad por modelo es esférica.

H_1 : La matriz de covarianzas de la métrica de especificidad por modelo es esférica.

Los supuestos de normalidad para las métricas de especificidad se cumplieron (todos los $p > \alpha = 0.05$) así como el supuesto de homocedasticidad de Mauchly (ya que $p = 0.585$, no se rechaza H_0); entonces, la prueba que se utilizó fue ANOVA de medidas repetidas (Tabla 24).

Tabla 24. Prueba a utilizar para comparar la especificidad por modelo según la verificación de los supuestos.

Modelo	Supuesto de normalidad <i>p</i> _{Shapiro-Wilk}	Supuesto de homocedasticidad <i>p</i> _{Mauchly}	Prueba a utilizar
RL-TV	0.741	0.585	ANOVA de medidas repetidas
AD-TV	0.522		
RL-VS	0.765		
AD-VS	0.847		

Nota: Nivel de significancia $\alpha=0.05$.
Referencia: Elaboración propia.

Con la prueba ANOVA se encontró un estadístico $F=4.887$ y un valor $p=0.008$, es decir, existe evidencia suficiente para afirmar que se encontraron diferencias estadísticamente significativas en la especificidad promedio según el modelo ajustado (Tabla 25). Se observa que el modelo con mayor promedio de especificidad es RL-VS (0.702 ± 0.070 ; Figura 12).

Tabla 25. Comparación de especificidad promedio por modelo según la prueba ANOVA.

Modelo				ANOVA de medidas repetidas
RL-TV Media (\pm DE)	AD-TV Media (\pm DE)	RL-VS Media (\pm DE)	AD-VS Media (\pm DE)	
0.699 (± 0.077)	0.687 (± 0.055)	0.702 (± 0.070)	0.643 (± 0.078)	$F=4.887, p=0.008$

Nota: Nivel de significancia $\alpha=0.05$. Se resalta en negritas el modelo con la mejor especificidad promedio.
RL: Regresión Logística, AD: Árboles de Decisión, TV: Todas las Variables, VS: Variables Seleccionadas, DE: desviación estándar.
Referencia: Elaboración propia.

Con lo anterior se determinó que el modelo óptimo es RL-VS, debido a que sus métricas de desempeño cuentan con menor variabilidad, mayor promedio de exactitud y especificidad, y es un modelo simple, pues solo se incluyen las variables clínicas con asociación significativa al tipo de atención médica. Se seleccionó el bloque 1 (Tabla 26) dado que obtuvo las mayores métricas de rendimiento en el conjunto de prueba (exactitud=0.668, sensibilidad=0.532 y especificidad=0.791).

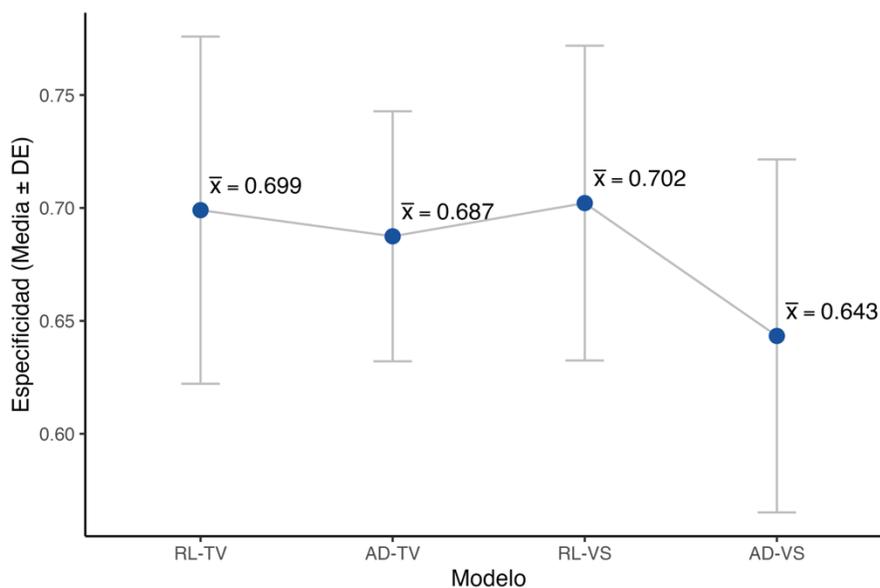


Figura 12. Comparación de especificidad promedio según modelos ajustados.

Referencia: Elaboración propia.

Tabla 26. Métricas por cada bloque de validación cruzada del modelo RL-VS.

Métrica	Bloque									
	1	2	3	4	5	6	7	8	9	10
Precisión	0.668	0.632	0.641	0.565	0.600	0.591	0.635	0.642	0.609	0.626
Sensibilidad	0.532	0.552	0.530	0.493	0.505	0.616	0.571	0.460	0.488	0.562
Especificidad	0.791	0.705	0.741	0.631	0.684	0.569	0.692	0.804	0.719	0.684

Nota: Se resalta en negritas el bloque elegido.

Referencia: Elaboración propia.

4.4.2. Representación del modelo predictivo de tipo de atención médica

El modelo final estimado de regresión logística se expresa en la siguiente ecuación:

$$P(\widehat{\text{Hospitalizado}}) = \frac{e^{\widehat{\text{logit}}}}{1 + e^{\widehat{\text{logit}}}}$$

Donde

$$\begin{aligned} \widehat{\text{logit}} = & 0.827 - 0.728 (\text{Ed} = 21 - 40) - 0.692 (\text{Ed} = 41 - 60) - 0.402 (\text{Ed} = 61 - 80) \\ & + 0.641 (\text{Ed} = 80 +) + 0.567 (\text{PCR} = \text{DENV2}) + 0.290 (\text{PCR} = \text{DENV3}) \\ & + 0.329 (\text{PCR} = \text{DENV4}) + 0.523 (\text{Diab} = \text{Sí}) + 0.778 (\text{Hip} = \text{Sí}) \\ & + 3.127 (\text{Ren} = \text{Sí}) + 1.203 (\text{Emb} = \text{Sí}) - 1.424 (\text{UM} = \text{IMSS}) \\ & - 0.548 (\text{UM} = \text{IMSS} - \text{BIENESTAR}) - 0.968 (\text{UM} = \text{ISSSTE}) \\ & - 0.669 (\text{UM} = \text{SSA}) \end{aligned}$$

En la Tabla 27 y Figura 13 se destacan los coeficientes que representan el modelo estimado final, así como el error estándar de los mismos, la prueba de significancia asociada y su valor p , y también los *odds ratio* de cada categoría con respecto a las categorías de referencia.

Las hipótesis fueron:

$$H_0: \beta_p = 0$$

$$H_1: \beta_p \neq 0$$

donde β_p son los coeficientes de regresión obtenidos.

De igual forma, se observa que la edad del paciente infectado por dengue fue una variable que predice significativamente el tipo de atención: si el paciente se encuentra entre 21 y 40 años, los *odds* de ser hospitalizados son 0.483 veces mayor ($OR=0.483$, $p<0.001$, se rechaza H_0); si el paciente tiene entre 41 y 60 años, los *odds* de ser hospitalizado aumentan 0.5 veces con respecto a las edades de 20 años o menos ($OR=0.500$, $p<0.001$, se rechaza H_0); por último, si el paciente tiene entre 61 y 80 años, el riesgo de hospitalización aumenta 0.669 veces con respecto a las edades de 20 años o menos ($OR=0.669$, $p=0.007$, se rechaza H_0).

En relación con el resultado obtenido a las pruebas PCR, si el paciente es detectado con el serotipo DENV-2, los *odds* de hospitalización aumentan 1.763 veces con respecto al serotipo DENV-1 ($OR=1.763$, $p=0.001$, se rechaza H_0). Además, si el paciente es diagnosticado con el serotipo DENV-3, los *odds* de ser hospitalizado aumentan en 1.336 veces ($OR=1.336$, $p=0.038$, se rechaza H_0).

También se observa que la presencia de diabetes en el paciente confirmado de dengue fue un factor significativo: si el paciente sufre de diabetes, el riesgo de ser hospitalizado es 1.688 veces mayor ($OR=1.688$, $p=0.013$, se rechaza H_0).

Asimismo, si el paciente padece hipertensión, los *odds* de que el paciente sea hospitalizado son 2.178 veces mayor ($OR=2.178$, $p<0.001$, se rechaza H_0).

Se identificó que la presencia de enfermedad renal en el paciente con dengue fue un determinante muy significativo en el riesgo de hospitalización, ya que si se cuenta con ella el riesgo de hospitalización es hasta 22.815 veces mayor con respecto a no contar la enfermedad ($OR=22.815$, $p=0.002$, se rechaza H_0).

La presencia de embarazo al momento del diagnóstico resultó un factor significativo. Si las pacientes confirmadas por dengue se encuentran embarazadas, el riesgo de hospitalización es 3.331 veces mayor ($OR=3.331$, $p<0.001$, se rechaza H_0).

Finalmente, si el paciente fue confirmado de dengue en el IMSS ($OR=0.241$, $p<0.001$, se rechaza H_0), el ISSSTE ($OR=0.380$, $p<0.001$, se rechaza H_0) o la SSA ($OR=0.512$, $p<0.001$, se rechaza H_0), los *odds* de hospitalización del paciente aumentan en 0.241, 0.380 y 0.512 veces, respectivamente, con respecto a ser registrado en otra institución médica.

Con el modelo de regresión se tiene una capacidad de predecir el tipo de atención médica en un paciente confirmado de dengue del 66.8%, con una predicción correcta del riesgo de hospitalización del 53.2%, y una predicción correcta del riesgo de tratamiento ambulatorio del 79.1% (Tabla 26). Además, se puede establecer una predicción de un caso nuevo. Por ejemplo, si se tiene un paciente registrado en la SSA, en edad de 0 a 20 años, con presencia de diabetes, y se le diagnostica dengue con el serotipo DENV-2, la probabilidad de que éste sea hospitalizado es igual a 0.777, por lo que el paciente se clasificaría con el tipo de atención médica “hospitalizado”.

$$P(\widehat{\text{Hospitalizado}}) = \frac{e^{\widehat{\text{logit}}}}{1 + e^{\widehat{\text{logit}}}} = 0.777$$

Tabla 27. Modelo Final: Regresión logística de tipo de atención médica recibida.

Características	β	Error estándar	z	p	OR	I.C. 95% para β
Intercepto	0.827	0.250	3.308	<0.001	2.286	(0.346,1.328)
Edad – 0 a 20 años					1	
Edad – 21 a 40 años	-0.728	0.080	-9.114	<0.001	0.483	(-0.885,-0.572)
Edad – 41 a 60 años	-0.692	0.100	-6.894	<0.001	0.500	(-0.89,-0.496)
Edad – 61 a 80 años	-0.402	0.151	-2.669	0.008	0.669	(-0.697,-0.107)
Edad – Más de 80 años	0.641	0.451	1.423	0.155	1.898	(-0.197,1.596)
PCR – DENV-1					1	
PCR – DENV-2	0.567	0.128	4.419	<0.001	1.763	(0.317,0.820)
PCR – DENV-3	0.290	0.140	2.071	0.038	1.336	(0.016,0.565)
PCR – DENV-4	0.329	0.186	1.771	0.077	1.390	(-0.035,0.694)
Diabetes – No					1	
Diabetes – Sí	0.523	0.211	2.481	0.013	1.688	(0.112,0.940)
Hipertensión – No					1	
Hipertensión – Sí	0.778	0.225	3.457	<0.001	2.178	(0.341,1.225)
Enfermedad Renal – No					1	
Enfermedad Renal – Sí	3.127	1.034	3.024	0.002	22.815	(1.528,6.025)
Embarazo – No					1	
Embarazo – Sí	1.203	0.215	5.605	<0.001	3.331	(0.791,1.634)
U.M. – Otra					1	
U.M. – IMSS	-1.424	0.229	-6.222	<0.001	0.241	(-1.887,-0.986)
U.M. – IMSS-BIENESTAR	-0.548	0.313	-1.751	0.080	0.578	(-1.167,0.062)
U.M. – ISSSTE	-0.968	0.278	-3.486	<0.001	0.380	(-1.521,-0.431)
U.M. – SSA	-0.669	0.231	-2.893	0.004	0.512	(-1.136,-0.226)

Nota: OR= Odds ratio. Se resaltan en negritas los valores p significativos. Nivel de significancia $\alpha=0.05$.
Referencia: Elaboración propia.

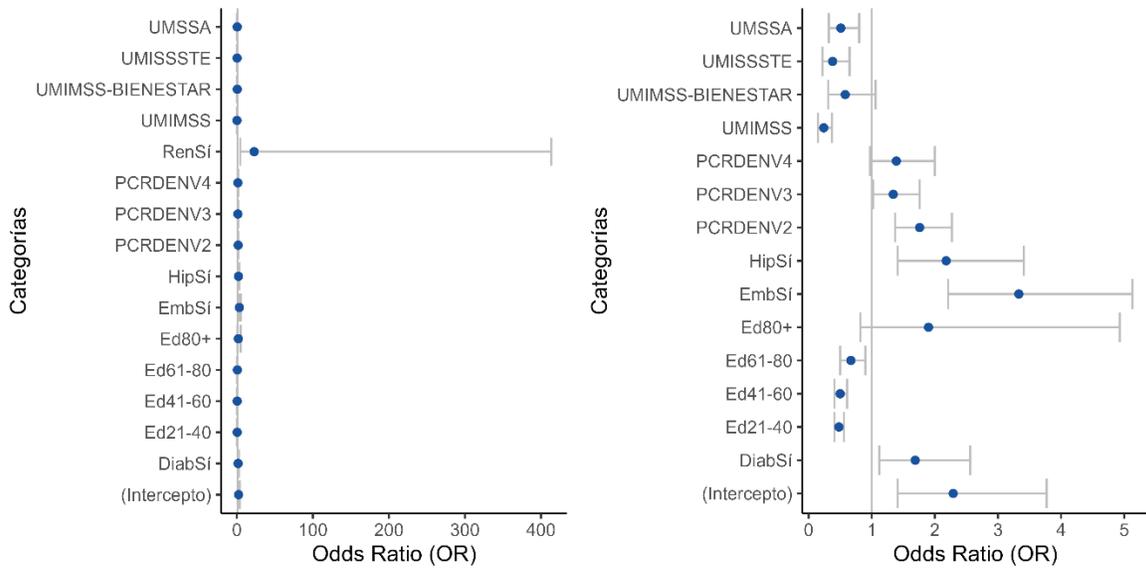


Figura 13. Odds ratio del modelo elegido. A la izquierda, se muestran todas las categorías incluidas en el modelo; a la derecha, se quitó la categoría Ren=Sí para una mejor visualización.

Referencia: Elaboración propia.

5. Discusión

En este trabajo se construyó un modelo para predecir el tipo de atención médica recibida por un paciente confirmado de dengue en el estado de Veracruz. Para ello se tomaron datos abiertos reportados por la Secretaría de Salud en México durante el periodo 2020-2023.

Dentro de los hallazgos se presentaron tres distintos estatus al diagnóstico de los casos de dengue. La distribución de los casos probables es similar a la distribución de casos confirmados y descartados juntos, es decir, que muchos casos registrados quedan en espera de diagnóstico, y este proceso puede demorar. Así mismo, el número de casos confirmados se asemeja al número de casos descartados.

En la investigación se encontraron diferencias significativas en las proporciones de casos confirmados por año; pese a ello, con respecto al año 2023 se identificó una diferencia negativa con respecto a años anteriores, lo que indica que la proporción de casos confirmados disminuyó. Esto no indica una disminución de casos, pues 2023 fue el año con más casos relacionados con dengue, pero a nivel relativo, la proporción de casos confirmados ha disminuido. Aunque este resultado es importante, no se debe bajar la guardia sobre el número total de casos registrados.

Las características principales que se encontraron en los pacientes confirmados de dengue fueron la presencia de diabetes, hipertensión, enfermedad renal, el serotipo DENV-2, y en algunos casos la presencia de embarazo. Además de ello, la mayor parte de los casos confirmados se encontraron en edades relativamente tempranas (40 años o menos).

Se encontró que la presencia de enfermedades crónicas como la diabetes, la hipertensión y la enfermedad renal en pacientes confirmados de dengue son determinantes para definir el tipo de atención médica que se les dará. En particular, la enfermedad renal en pacientes confirmados de dengue resultó ser un factor de riesgo crítico, y esto se observó desde el análisis de correspondencias, por lo que se recomienda realizar un monitoreo continuo en pacientes con enfermedad renal que han sido infectados por dengue, aun cuando no cuenten con signos de alarma. Estos resultados coinciden con los obtenidos por Domínguez y Valdés (2024), que describen estas enfermedades como factores de riesgo.

El segundo resultado importante del estudio fue la identificación del serotipo. Se encontró que los serotipos DENV-2 y DENV-3 aumentan significativamente el riesgo de hospitalización, en especial el primero mencionado. Aunque no se encontraron estudios que trataran este tema en la revisión de la literatura, el trabajo más cercano es el de Annan *et al.* (2023), que asocia los serotipos a la gravedad del dengue en pacientes embarazadas. También es crucial mencionar que el serotipo DENV-2 se detectó con mayor presencia en el estado de Veracruz.

En esta región, se encontró que la probabilidad de hospitalización disminuye en edades mayores de 20 años y hasta los 80 años, resultados que difieren de los obtenidos por Luque *et al.* (2023), que encontró como factor de riesgo de mortalidad a los adultos mayores de 65 años y personas de 19 a 35 años en Perú. Dado que el modelo del presente estudio tomó como nivel de referencia a los pacientes confirmados por dengue de 20 años y menos, se encontró que el riesgo de hospitalización es mayor en personas jóvenes.

Otro hallazgo significativo, y que coincide también con Annan *et al.* (2023) es la influencia del embarazo en el agravamiento de la enfermedad por dengue. En esta investigación, se determinó que la presencia de embarazo en mujeres aumenta significativamente la probabilidad de que la paciente sea hospitalizada. Se destaca el resultado porque, dado que la mayoría de las unidades médicas como IMSS cuentan con las consultas prenatales, es posible adaptar medidas de prevención tales como fumigación o implementación de larvicidas en el entorno de residencia de las pacientes, con el objetivo de evitar que ellas sean infectadas y, por ende, evitar la hospitalización y gravedad de la enfermedad en mujeres con esta condición.

Se menciona que, en el caso de las pacientes embarazadas, el riesgo de hospitalización no se debe considerar como un nivel exacto, puesto que no todos los pacientes registrados en el estado de Veracruz son mujeres, y no todas las mujeres se encuentran embarazadas; sin embargo, se definió importante dejarlo en el análisis debido a que la presencia de esta condición especial sí afecta la hospitalización según las fuentes consultadas mencionadas anteriormente.

Con respecto al modelo, se encontró un porcentaje moderado de predicción del tipo de atención médica recibida por un paciente confirmado de dengue en el estado de Veracruz.

Por ello se toma conciencia de que, cuando se trata del área médica, se requiere una mayor potencia en la predicción de un modelo, porque de ello depende el correcto tratamiento y la mejora de los pacientes contagiados.

Con respecto a las propuestas de análisis estadístico, fue adecuado tomar los casos confirmados mediante prueba PCR, ya que esto favoreció a que las clases de tipo de atención médica fueran balanceadas. Una desventaja que se encontró al realizar el análisis fue que, aunque las clases de tipo de atención se balancearon, no fue así con las características clínicas. Existió mucho desbalance en la distribución de los pacientes con enfermedades crónicas y condiciones especiales como el embarazo, por lo que al realizar el modelo se encontraron riesgos de hospitalización con intervalos de confianza muy grandes y asimétricos, como la presencia de enfermedad renal, la edad y la presencia de embarazo.

Un punto que destacar es que la implementación de la regresión logística facilitó la interpretación de la influencia de las características clínicas en el tipo de atención médica recibida, ya que la predicción de la probabilidad se describió en términos de una función “sencilla”, los signos indican el aumento o disminución del riesgo de hospitalización, y se pudieron identificar los niveles de las variables con mayor significancia.

6. Conclusiones

Con esta investigación, se concluye que se cumplió la hipótesis de investigación, puesto que se encontró que las enfermedades crónicas promueven la hospitalización de los pacientes con dengue en el estado de Veracruz, en particular la presencia de diabetes, hipertensión y enfermedad renal. Además, se encontró que las condiciones especiales como el embarazo y el serotipo diagnosticado también influyen en el tipo de atención médica recibida. En particular, se comprobó que la proporción de casos confirmados ha sido diferente entre los últimos cuatro años. El modelo ajustado predice de forma aceptable la probabilidad de hospitalización, y permite una interpretación sencilla de la relación entre las variables clínicas y el tipo de atención médica. Se invita a las instituciones de salud a tomar medidas de cuidado previas a que los pacientes presenten signos de alarma y sean hospitalizados. Del mismo modo, se reconoce el buen uso de la metodología estadística realizado en este trabajo.

7. Limitaciones y trabajos futuros

Una limitación importante fue que no se contó con el historial clínico completo del paciente, es decir, el tiempo desde que comenzó con los síntomas, la descripción de estos y el tratamiento que se le aplicó; lamentablemente, es complicado acceder a esta información, pero es posible realizando una solicitud al portal de transparencia del Gobierno de México. Por ello surge la propuesta de incluir esta información como complemento al modelo de regresión logística ajustado, con el fin de capturar toda la variabilidad posible, más allá de la obtenida usando los datos abiertos.

La segunda limitación fue que los casos abiertos consideran únicamente los registrados por los centros de salud, es decir, el análisis no contempla el total de casos reales en la región.

Un trabajo futuro es analizar si la gravedad y el tipo de atención médica se comporta igual en el resto de los estados de México. En el trabajo se tomó únicamente el estado de Veracruz debido al aumento exponencial en casos confirmados, pero podría compararse contra otras regiones con comportamientos similares. Se recomendaría construir el modelo considerando también la capacidad de atención hospitalaria que existe en cada estado.

Otra posible contribución será la aplicación de modelos estadísticos más avanzados, como los basados en técnicas de remuestreo o usando conocimiento *a priori* con ayuda de expertos.

Referencias

- Acock, A. y Stavig, G. (1979). A measure of association for nonparametric statistics. *Social Forces*, 57 (4), 1381-1386.
- Annan, E., Nguyen, U., Treviño, J., Yaacob, W., Mangla, S., Pathak, A., Nandy, R. y Haque, U. (2023). Moderation effects of serotype on dengue severity across pregnancy status in Mexico. *Infectious Diseases*, 23 (147), 1-9.
- Appice, A., Gel, Y., Iliev, I., Lyubchich, V. y Malerba, D. (2020). A multi-stage machine learning approach to predict dengue incidence: a case study in Mexico. *Ieee Access*, 8 (1), 52713-52725.
- Benedum, C., Shea, K., Jenkins, H., Kim, L. y Markuzon, N. (2020). Weekly dengue forecasts in Iquitos, Peru; San Juan, Puerto Rico; and Singapore. *Plos Neglected Tropical Diseases*, 14 (10), 1-26.
- Berrar, D. (2019). Cross validation. *Encyclopedia of Bioinformatics and Computational Biology*, 1, 542-545.
- Chung, M. (2020). Introduction to logistic regression. *arXiv preprint arXiv:2008.13567*.
- Conde-Gutiérrez, R., Colorado, D., Márquez-Nolasco, A. y Gonzalez-Flores, P. (2024). Parallel prediction of dengue cases with different risks in Mexico using an artificial neural network model considering meteorological data. *International Journal of Biometeorology*, 68 (1), 1043-1060.
- Dai, W. y Ji, W. (2014). A MapReduce Implementation of C4.5 Decision Tree Algorithm. *International Journal of Database Theory and Application*, 7 (1), 49-60.
- Dinov, I. (2018). *Data Science and Predictive Analytics, Biomedical and Health Applications using R*. Springer.
- Domínguez, A. y Valdés, L. (2024). Pacientes con dengue y signos de alarma durante la epidemia de 2022 en Santiago de Cuba. *MEDISAN*, 28 (6), e4523.
- Ferdousi, T., Cohnstaedt, L. y Scoglio, C. (2021). A windowed correlation-based feature selection method to improve time series prediction of dengue fever cases. *IEEE Access*, 9 (1), 141210-141222.
- Flores, C. y Flores, K. (2021). Pruebas para comprobar la normalidad de datos en procesos productivos: Anderson-Darling, Ryan-Joiner, Shapiro-Wilk y Kolmogórov-Smirnov. *Revista de Ciencias Sociales y Humanísticas*, 23 (2), 83-106.
- Freund, J., Miller, I. y Miller, M. (2000). *Estadística matemática con aplicaciones*, 6ª ed. Prentice Hall.
- Gobierno de México (2024). *Protocolo de Atención Médica para Casos de Infección por Virus Dengue*. Gobierno de México: <https://www.gob.mx/promosalud/documentos/protocolo-de-atencion-medica-para-casos-de-infeccion-por-virus-dengue>.
- Greenacre, M. y Blasius, J. (Eds.). (2006). *Multiple Correspondence Analysis and Related Methods*, 1ª ed. Chapman and Hall/CRC.
- Gutiérrez, H. y De la Vara, R. (2008). *Análisis y diseño de experimentos*. Mc Graw-Hill Interamericana.
- Hair, J., Anderson, R., Tatham, R. y Black, W. (1999). *Análisis Multivariante*, 5ª ed. Prentice Hall.
- Karasinghe, N., Peiris, S., Jayathilaka, R. y Dharmasena, T. (2024). Forecasting weekly dengue incidence in Sri Lanka: Modified Autoregressive Integrated Moving Average modeling approach. *Plos One*, 19 (3), 1-16.
- Kitchenham, B., Budgen, D. y Brereton, P. (2015). *Evidence-based software engineering and systematic reviews*. CRC Press.
- Kuehl, R. (2001). *Diseño de experimentos, principios estadísticos de diseño y análisis de investigación*. Thomson Learning.
- Liu, Y., Saul, S., Rao, A., Robinson, M., Agudelo, O., Sanz, A., Verghese, M., Solis, D., Sibai, M., Huang, C., Sahoo, M., Gelvez, R., Bueno, N., Estupiñan, M., Villar, L., Rojas, E., Rosso, F., Donato, M., Pinsky, B., Einav, S. y Khatri, P. (2022). An 8-gene machine learning model improves clinical prediction of severe dengue progression. *Genome Medicine*, 14 (33), 1-15.
- Luque, N., Cilloniz, C., Pons, M., Donaires, F., Albornoz, R., Mendocilla-Risco, M. y Espinoza, M. (2023). Características clínicas y epidemiológicas de las muertes por dengue durante un brote en el norte del Perú. *Revista Peruana de Medicina Experimental y Salud Pública*, 40 (1), 67-72.
- Martínez-Torres, E. (1995). Dengue y dengue hemorrágico: aspectos clínicos. *Salud pública de México*, 37 (1), 29-44.
- Mehta, C. y Patel, N. (1983). A Network Algorithm for Performing Fisher's Exact Test in $r \times c$ Contingency Tables. *Journal of the American Statistical Association*, 78 (382), 427-434.
- Molina, M. (2021). Una historia de té y números. La prueba exacta de Fisher. *Revista Electrónica AnestesiaR*, 13 (10), 4.

- Organización de las Naciones Unidas (2023). *Objetivos de Desarrollo Sostenible - 3. Salud y Bienestar*. Organización de las Naciones Unidas. <https://www.un.org/sustainabledevelopment/es/health/>.
- Organización Mundial de la Salud (2023, 21 de diciembre). *Dengue - Situación Mundial*. Organización Mundial de la Salud. <https://www.who.int/es/emergencies/disease-outbreak-news/item/2023-DON498>.
- Organización Mundial de la Salud (2024, 23 de abril). *Dengue and severe dengue*. Organización Mundial de la Salud: <https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue>.
- Organización Panamericana de la Salud (2020). *Algoritmos para el manejo clínico del dengue*. Organización Panamericana de la Salud. <https://www.paho.org/es/documentos/algoritmos-para-manejo-clinico-casos-dengue>.
- Organización Panamericana de la Salud (2024, 07 de octubre). *Alerta Epidemiológica, Aumento de casos de dengue en la Región de las Américas*. Organización Panamericana de la Salud. <https://www.paho.org/es/documentos/alerta-epidemiologica-aumento-casos-dengue-region-americas-7-octubre-2024>.
- Organización Panamericana de la Salud (2025, 07 de febrero). *Alerta Epidemiológica, Riesgo de brotes de dengue por la mayor circulación de DENV-3 en la región de las Américas*. Organización Panamericana de la Salud. <https://www.paho.org/sites/default/files/2025-02/2025-feb-7-phe-epi-alerta-dengue-es-final.pdf>.
- Pardo, A. y Ruiz, M. (2013). *Análisis de datos en ciencias sociales y de la salud III*. Editorial Síntesis.
- Pardo, A. y San Martín, R. (2010). *Análisis de datos en ciencias sociales y de la salud II*. Editorial Síntesis.
- Pizarro, D. (2009). Dengue, dengue hemorrágico. *Acta Pediátrica Costarricense*, 21 (1), 08-17.
- Quevedo, F. (2011). La prueba de ji-cuadrado. *Estadística Aplicada a la Investigación en Salud*, 11 (12).
- Rodríguez, S., Méndez, S., Guerrero, S., Ayón, J., García, M., Santos, G. (2024). Tendencias epidemiológicas del dengue en Latinoamérica: 2019-2024. *Medicina Interna de México*, 40 (5), 287-294.
- Rokach, L. y Maimon, O. (2005). Decision trees. *Data mining and knowledge discovery handbook*, 165-192.
- R Core Team (2025). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Schaum, A., Bernal-Jaquez, R. y Sánchez-González, G. (2022). Model-Based Monitoring of Dengue Spreading. *IEEE Access*, 10 (1), 126892-126898.
- Sebrango, C., Sánchez, L., Shkedy, Z., Sistachs, V., Ruff, C. y Matheu, A. (2020). Model uncertainty in the comparison of two single dengue outbreaks. *Revista Investigación Operacional*, 41 (2), 344-351.
- Secretaría de Salud (s. f.). *Panorama epidemiológico de dengue 2024*. Gobierno de México. <https://www.gob.mx/salud/documentos/panorama-epidemiologico-de-dengue-2024>.
- Secretaría de Salud (2024). *Panorama epidemiológico de dengue 2023*. Secretaría de Salud. <https://www.gob.mx/salud/documentos/panorama-epidemiologico-de-dengue-2023>.
- Secretaría de Salud (2025). *Datos Abiertos Bases Históricas de Enfermedades Transmitidas por Vector*. Secretaría de Salud: <https://www.gob.mx/salud/documentos/datos-abiertos-bases-historicas-de-enfermedades-transmitidas-por-vector>.
- Sheldon, M., Fillyaw, M. y Thompson, W. (1996). The use and interpretation of the Friedman test in the analysis of ordinal-scale data in repeated measures designs. *Physiotherapy Research International*, 1 (4), 221-228.
- Singh, P. y Chaturvedi, H. (2024). Socio-ecological predictors of dengue in high incidence area of Delhi, India. *Scientific Reports*, 14 (1), 17005.
- Solís-Navarro, M., Vargas-De-León, C., Gúzman-Martínez, M. y Corzo-Gómez, J. (2022). A bayesian prediction spatial model for confirmed dengue cases in the state of Chiapas, Mexico. *Journal of Tropical Medicine*, 2022 (1), 1971786.
- Stahle, L. y Wold, S. (1989). Analysis of variance (ANOVA). *Chemometrics and intelligent laboratory systems*, 6 (4), 259-272.
- Tamayo, O., García, T., Escobar, N., González, D., Castro, O. (2020). Caracterización clinicoepidemiológica de pacientes con dengue ingresados en el Instituto de Medicina Tropical Pedro Kourí. *MEDISAN*, 24 (4), 654.
- Tebes, G., Peppino, D., Becker, P. y Olsina, L. (2020). Proceso para revisión sistemática de literatura y mapeo sistemático. *SADIO Electronic Journal of Informatics and Operations Research*, 19 (2), 94-118.
- Vargas, A., Bustos, E., Salas, A., Ruvalcaba, J. y Imbert Palafox, J. (2021). Infección por Dengue, un problema de salud pública en México. *Journal of Negative & no Positive Results*, 6 (2), 293-306.

- Wang, X., Li, T., Shu, Y., Zhang, J., Shan, X., Li, D., Ma, D., Long, S., Pan, Y., Chen, J., Liu, P. y Sun, Q. (2022). Clinical characteristics and risk factors for severe dengue fever in Xishuangbanna, during the dengue outbreak in 2019. *Frontiers in microbiology*, 13 (1), 1-9.
- Wayne, D. (1991). *Bioestadística. Base para el análisis de las ciencias de la salud*. Limusa Wiley. 533-534.

“Lis de Veracruz: Arte, Ciencia, Luz”

www.uv.mx