



Universidad de Costa Rica
Escuela de Matemáticas
Modelos Lineales y de Supervivencia
CA-0414

Anteproyecto

Modelos Lineales y de Supervivencia

PROFESOR

Luis A. Barboza

ESTUDIANTES

Luis Fernando Amey Apuy - C20470

Javier Hernández Navarro - C13674

Anthony Mauricio Jiménez Navarro - C24067

Gustavo Alberto Amador Fonseca - C20459

I - 2025

1. Planteamiento del Problema

El dengue es una enfermedad viral transmitida por mosquitos que representa una amenaza significativa para la salud pública en regiones tropicales y subtropicales. Según la Organización Mundial de la Salud, se estima que ocurren entre 50 y 100 millones de infecciones por dengue cada año, con una carga considerable en países en desarrollo (Organización Mundial de la Salud, 2024).

En el contexto del Upazila Health Complex, ubicado en Kalai, Jaipurhat, Bangladesh, se ha observado una incidencia notable de casos de dengue. La identificación temprana y precisa de la infección por dengue es crucial para el manejo clínico efectivo y la prevención de complicaciones graves.

Diversos estudios han demostrado que ciertos parámetros hematológicos, como el recuento de glóbulos blancos, plaquetas y niveles de hemoglobina, pueden estar asociados con la presencia y severidad del dengue. Por ejemplo, investigaciones realizadas en Tailandia y Brasil han reportado que los pacientes con dengue presentan leucopenia, trombocitopenia y hemoconcentración durante las fases agudas de la enfermedad (Gonçalves et al., 2012), (Chaloemwong et al., 2018). Estos hallazgos sugieren que los indicadores hematológicos podrían servir como herramientas diagnósticas complementarias en entornos con recursos limitados. En este contexto, surge la siguiente pregunta de investigación:

¿Que asociación existe entre indicadores hematológicos y la presencia del dengue en pacientes del Upazila Health Complex, Kalai, Jaipurhat, Bangladesh durante el 2 de Noviembre del 2023?

La exploración de esta asociación es fundamental para mejorar las estrategias de diagnóstico y manejo clínico del dengue en áreas endémicas. Además, contribuirá al cuerpo de conocimiento existente sobre las manifestaciones hematológicas del dengue, permitiendo una mejor comprensión de su fisiopatología y facilitando la implementación de intervenciones de salud pública más efectivas.

2. Marco Teórico Conceptual

El dengue es una enfermedad viral transmitida por mosquitos del género *Aedes*, cuya presentación clínica puede ir desde formas asintomáticas hasta manifestaciones severas como el síndrome de shock por dengue. La detección temprana y precisa constituye un elemento fundamental para un adecuado manejo clínico y la prevención de complicaciones graves (Organización Mundial de la Salud, 2024). En este contexto, los parámetros hematológicos han cobrado un papel crucial en la evaluación y seguimiento de los pacientes afectados. Alteraciones como la trombocitopenia, la leucopenia y la hemoconcentración se presentan con frecuencia en individuos infectados, reflejando la respuesta fisiopatológica del organismo frente al virus e incluso proporcionando indicios sobre la posible severidad de la enfermedad (Gonçalves et al., 2012).

Entre estas alteraciones, la leucopenia —una disminución del recuento de glóbulos blancos— suele observarse en las fases iniciales de la infección y puede atribuirse a la supresión medular inducida por el virus (Chaloemwong et al., 2018). La trombocitopenia, por su parte, se manifiesta como una reducción significativa en el número de plaquetas y está directamente relacionada con un mayor riesgo de hemorragias. Esta última se ha consolidado como una característica distintiva del dengue, al punto que se ha propuesto su uso como marcador diferencial frente a otras enfermedades febriles (Gonçalves et al., 2012). La hemoconcentración, entendida como el aumento de los niveles de hemoglobina

y hematocrito producto de la fuga plasmática, aparece con mayor frecuencia durante las fases críticas de la enfermedad y se ha vinculado con una progresión hacia cuadros clínicos de mayor gravedad, como el shock hipovolémico (Gonçalves et al., 2012). Adicionalmente, investigaciones recientes han resaltado el valor diagnóstico de la proporción neutrófilo-linfocito (NLR), así como la aparición de linfocitos atípicos, como posibles indicadores del curso clínico del dengue (Chaloemwong et al., 2018).

En los últimos años, el desarrollo de modelos predictivos basados en parámetros hematológicos ha sido objeto de creciente interés. Estos modelos, que emplean enfoques ingenieriles y herramientas provenientes de las nuevas tecnologías, buscan mejorar tanto el diagnóstico temprano como el pronóstico clínico de la enfermedad. Diversos trabajos han subrayado la utilidad de integrar múltiples variables clínicas y de laboratorio para lograr una evaluación más precisa y robusta del paciente con sospecha de dengue (Arrubla Hoyos, 2022). Asimismo, se ha explorado la dinámica de diversos marcadores bioquímicos —como las enzimas hepáticas, especialmente las aminotransferasas— en combinación con alteraciones hematológicas, como medio para anticipar una posible evolución hacia formas graves de la enfermedad (Irshad et al., 2023).

Es importante señalar que gran parte de la literatura científica relacionada con el dengue y los parámetros hematológicos se ha generado en contextos regionales específicos. Esto aporta a la necesidad de contar con estudios que consideren las particularidades de otras regiones geográficas. Un ejemplo relevante lo constituye un estudio realizado en Bangladesh, donde se analizó la relación entre alteraciones hematológicas y la infección por dengue, arrojando información valiosa sobre los patrones clínicos observados en esa población (Assaduzzaman et al., 2024). El estudio recopiló datos de casos atendidos en el Upazila Health Complex de Kalai, en Jaipurhat, encontrando que la leucopenia y la trombocitopenia eran hallazgos frecuentes entre los pacientes infectados. Dichos resultados permitieron a los autores proponer el uso de estos parámetros como indicadores diagnósticos útiles, especialmente en contextos con recursos limitados donde el acceso a pruebas moleculares puede estar restringido (Assaduzzaman et al., 2024).

A pesar de los avances logrados, persisten importantes brechas en el conocimiento que justifican la realización de nuevas investigaciones. Una de ellas es la escasez de estudios locales enfocados en analizar la relación entre indicadores hematológicos y la presencia de dengue, particularmente en el contexto específico de Bangladesh y, más concretamente, en el Upazila Health Complex de Kalai. Asimismo, existe una clara necesidad de diseñar y validar modelos predictivos que utilicen datos regionales para mejorar la capacidad diagnóstica en poblaciones locales. A esto se suma la variabilidad observada en los valores de referencia y en la interpretación de los parámetros hematológicos, la cual dificulta la estandarización de criterios clínicos entre diferentes regiones y estudios. Se vuelve fundamental emprender un estudio que analice rigurosamente la asociación entre parámetros hematológicos y la presencia de dengue, utilizando datos empíricos y métodos estadísticos robustos que permitan generar evidencia confiable y contextualizada.

3. Objetivos

3.1. Objetivo General

Analizar la asociación entre indicadores hematológicos y la presencia de infección por dengue en pacientes atendidos en el Upazila Health Complex, Kalai, Jaipurhat, Bangladesh, el 2 de noviembre de 2023.

3.2. Objetivos Específicos

1. **Caracterizar** los valores de recuento de métricas hematológicas en pacientes con diagnóstico confirmado de dengue y compararlos con aquellos sin diagnóstico de dengue, utilizando el conjunto de datos seleccionado.
2. **Determinar** la existencia de diferencias estadísticamente significativas en los parámetros hematológicos mencionados entre los grupos de pacientes con y sin dengue, aplicando pruebas estadísticas apropiadas.
3. **Evaluar** la capacidad predictiva de los indicadores hematológicos para la presencia de dengue mediante el desarrollo de modelos estadísticos, y analizar su desempeño utilizando métricas apropiadas.

3.3. Origen de los Datos

El presente estudio se fundamenta en el conjunto de datos recopilados por el Dr. Md. Faisal Nahid Pabitra, que documenta las características clínicas y hematológicas de pacientes atendidos en el Upazila Health Complex, Kalai, Jaipurhat, Bangladesh, el 2 de noviembre de 2023. Este conjunto de datos incluye variables como recuento de glóbulos blancos, plaquetas, niveles de hemoglobina y diagnóstico confirmado de dengue, entre otros parámetros relevantes (Assaduzzaman et al., 2024).

3.4. Proceso de Recolección de Datos

La recolección de datos fue realizada por el personal médico del Upazila Health Complex mediante la utilización de formularios estandarizados y protocolos clínicos establecidos. Se emplearon técnicas de laboratorio validadas para la obtención de los parámetros hematológicos, asegurando la precisión y consistencia de las mediciones. La información fue registrada en sistemas electrónicos de salud, garantizando la integridad y trazabilidad de los datos.

3.5. Validez y Confiabilidad

La validez y confiabilidad de los datos utilizados en este estudio están respaldadas por varios factores fundamentales que refuerzan la calidad metodológica. El proceso de recolección de datos fue supervisado por el Dr. Md. Faisal Nahid Pabitra, médico especialista y funcionario de salud del Upazila Health Complex en Kalai, Jaipurhat, Bangladesh, lo cual garantiza una obtención rigurosa, sistemática y conforme a criterios clínicos establecidos. Adicionalmente, la curación de los datos, conceptualización, análisis estadístico y validación de resultados fueron realizados por un equipo especializado. Esto contribuye a la coherencia interna del conjunto de datos y a la precisión en su interpretación. Otro aspecto clave de la confiabilidad radica también en la transparencia y disponibilidad pública de la base de datos. Este acceso abierto permite la replicación de resultados, la verificación independiente por parte de otros investigadores.

3.6. Consideraciones Éticas

El estudio original obtuvo la aprobación del comité de ética correspondiente y se adhirió a las directrices éticas internacionales para la investigación en seres humanos. Se garantizó la confidencialidad de la información de los pacientes mediante la anonimización de los

datos y el uso de identificadores codificados. Toda información que permitiera la identificación del paciente fue eliminada. Los participantes fueron informados sobre los objetivos del estudio y proporcionaron su consentimiento personal o por medio de su guardián legal.

4. Descripción de las Variables del Conjunto de Datos

El conjunto de datos empleado en este estudio, recopilado por Hasan et al. (2024), incluye una serie de variables clínicas y hematológicas relevantes para el análisis de la asociación entre estos parámetros y la presencia de infección por dengue en pacientes atendidos en el Upazila Health Complex, Kalai, Jaipurhat, Bangladesh, el 2 de noviembre de 2023. A continuación, se describen detalladamente las variables incluidas:

- **Edad (Age):** Variable cuantitativa continua que representa la edad del paciente en años. La edad es un factor importante en la susceptibilidad y respuesta inmunológica frente al dengue (Gonçalves et al., 2012).
- **Sexo (Sex):** Variable cualitativa dicotómica que indica el género del paciente (masculino o femenino). Estudios previos han observado diferencias en la presentación clínica del dengue según el sexo (Chaloemwong et al., 2018).
- **Hemoglobina (Hemoglobin):** Variable cuantitativa continua que mide la concentración de hemoglobina en la sangre, expresada en gramos por decilitro (g/dL). Niveles elevados pueden indicar hemoconcentración, mientras que niveles bajos pueden reflejar anemia (Gonçalves et al., 2012).
- **Recuento de Glóbulos Blancos (WBC Count):** Variable cuantitativa continua que indica el número total de leucocitos por microlitro (μL) de sangre. La leucopenia es una característica común en pacientes con dengue (Chaloemwong et al., 2018).
- **Conteo Diferencial de Leucocitos (Differential Count):** Conjunto de variables cuantitativas que representan el porcentaje de diferentes tipos de leucocitos (neutrófilos, linfocitos, monocitos, eosinófilos y basófilos). Alteraciones en estos porcentajes pueden ser indicativas de infecciones virales como el dengue (Gonçalves et al., 2012).
- **Panel de Glóbulos Rojos (RBC Panel):** Conjunto de variables cuantitativas que incluyen el recuento total de eritrocitos, volumen corpuscular medio (MCV), hemoglobina corpuscular media (MCH) y concentración de hemoglobina corpuscular media (MCHC). Estos parámetros ayudan a evaluar la morfología y funcionalidad de los glóbulos rojos (Gonçalves et al., 2012).
- **Recuento de Plaquetas (Platelet Count):** Variable cuantitativa continua que mide el número de plaquetas por microlitro (μL) de sangre. La trombocitopenia es una manifestación hematológica frecuente en el dengue y se asocia con un mayor riesgo de hemorragias (Chaloemwong et al., 2018).
- **Amplitud de Distribución Plaquetaria (PDW):** Variable cuantitativa continua que indica la variabilidad en el tamaño de las plaquetas, expresada en porcentaje. Valores elevados pueden reflejar una producción aumentada de plaquetas jóvenes, mientras que valores bajos pueden indicar una producción deficiente (Gonçalves et al., 2012).

- **Diagnóstico de Dengue (Dengue Diagnosis):** Variable cualitativa dicotómica que indica la presencia o ausencia de infección por dengue, confirmada mediante pruebas serológicas como la detección del antígeno NS1 o anticuerpos IgM/IgG (As-saduzzaman et al., 2024).

5. Análisis Exploratorio de Datos

Previo al desarrollo del análisis, se realizó una depuración de las variables disponibles, considerando únicamente aquellas que guardan una relación directa con el objetivo de estudio. En este sentido, se excluyeron variables como la edad y el sexo.

5.1. Análisis de correlación entre variables

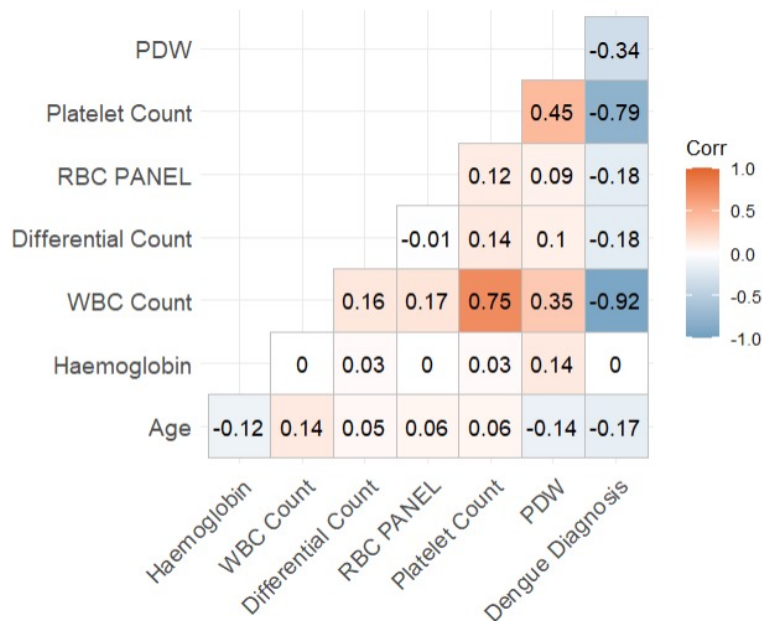


Figura 1: Matriz de correlación entre variables.

Fuente: Elaboración propia.

Para esto se considera en la Figura 1 que muestra la matriz de correlación de Pearson calculada entre las variables numéricas, incluyendo variables continuas y binarias codificadas como 0 y 1. El análisis de correlación muestra una relación fuerte ($r = 0,75$) entre *WBC Count* y *Platelet Count*, lo que sugiere posible colinealidad que se debe considerar al escoger el modelo.

También se observa una correlación moderada ($r = 0,45$) entre *Platelet Count* y *PDW*, indicando posible relación entre cantidad y variabilidad plaquetaria. *Hemoglobin* presenta correlaciones cercanas a cero, aunque podría ser relevante en el análisis. Las variables binarias (*Differential Count* y *RBC PANEL*) muestran correlaciones bajas pero asociaciones positivas con *WBC Count* y *Platelet Count*, lo que sugiere utilidad clínica. Además, la variable *Dengue Diagnosis* muestra correlaciones negativas moderadas con *WBC Count* ($r = -0,35$) y *PDW* ($r = -0,34$), lo que podría reflejar patrones hematológicos característicos en pacientes infectados. Estas asociaciones también podrían justificar su inclusión en modelos predictivos.

5.2. Análisis de variables binarias

Dengue Diagnosis

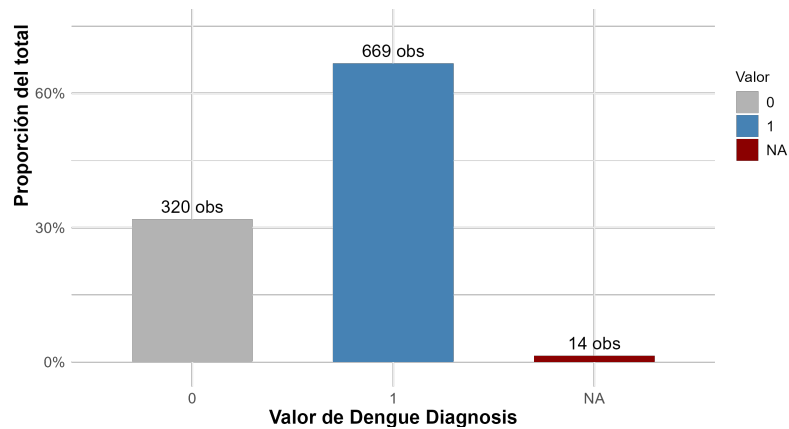


Figura 2: Distribución de la variable Dengue Diagnosis.

Fuente: Elaboración propia.

Como se observa en la Figura 2, 669 observaciones (66.7%) corresponden a casos positivos, 320 (32%) a negativos y 14 (1.4%) presentan valores faltantes (NA), proporción poco significativa. El sesgo hacia casos positivos sugiere que la muestra se enfoca en pacientes sintomáticos, lo cual es esperable. Sin embargo, este desbalance debe considerarse en el ajuste de modelos, ya que puede afectar su rendimiento e interpretación. Los valores faltantes deberán tratarse mediante limpieza o imputación, según el enfoque metodológico adoptado.

Differential Count

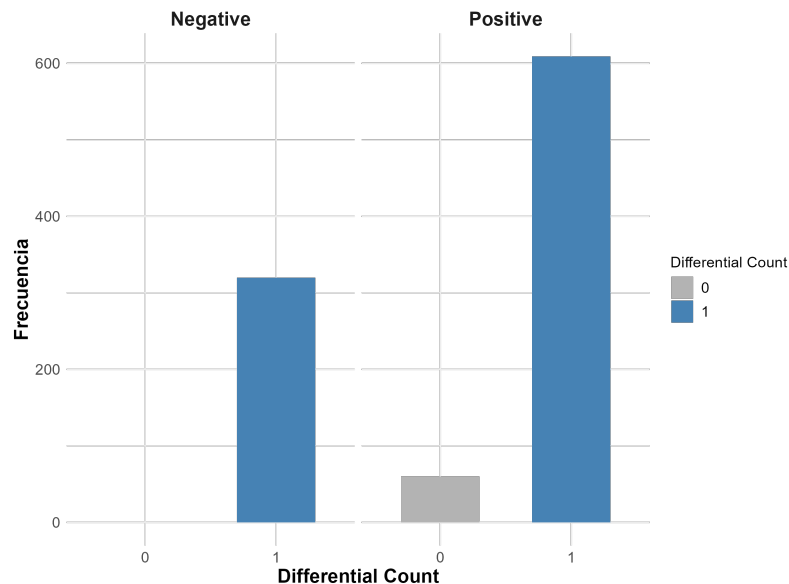


Figura 3: Distribución de Differential Count según diagnóstico de dengue.

Fuente: Elaboración propia.

Como se observa en la Figura 3, en ambos grupos (diagnóstico positivo y negativo) se presenta una mayor frecuencia de pacientes con **Differential Count = 1**, lo que indica

que las alteraciones en el conteo diferencial de leucocitos son comunes en la muestra. En el grupo positivo, la concentración en el valor 1 es más marcada, mientras que en el negativo predomina el valor 0. Esta diferencia sugiere una posible asociación entre el conteo diferencial alterado y el diagnóstico de dengue, aunque debe interpretarse con cautela debido al desbalance en el número de casos.

RBC PANEL

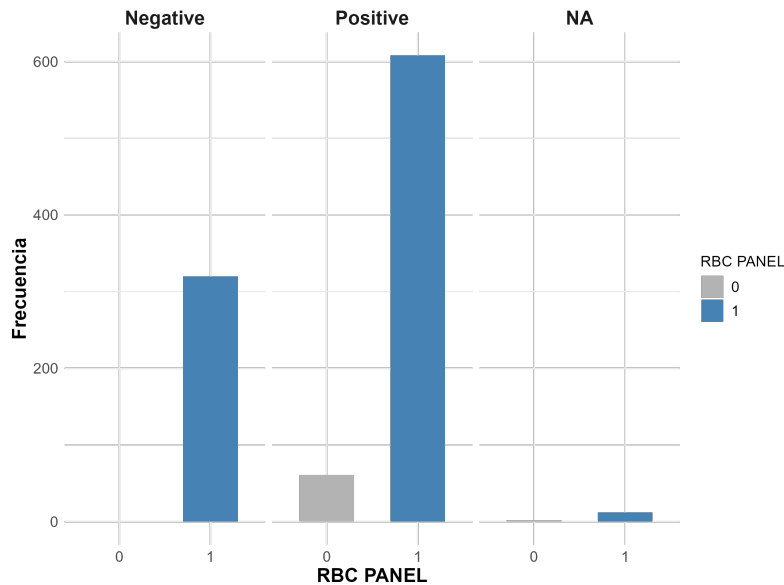


Figura 4: Distribución de Differential Count según diagnóstico de dengue.

Fuente: Elaboración propia.

En la distribución de la variable RBC PANEL según el diagnóstico de dengue, se observa que la mayoría de los pacientes presentan alteraciones en el panel eritrocitario ('RBC PANEL = 1'), tanto en el grupo con dengue como en el grupo sin diagnóstico.

Aunque el grupo positivo concentra una mayor cantidad de casos con 'RBC PANEL = 1', esta tendencia debe interpretarse con cautela debido al desequilibrio en el número de observaciones entre los grupos. En términos generales, el gráfico sugiere que las alteraciones en el panel de glóbulos rojos son frecuentes en la muestra, pero no muestra una diferenciación visual clara entre los estados diagnósticos, pero no se descartaría su uso.

5.3. Análisis de variables numéricas

5.3.1. Exploración mediante estadísticos descriptivos

Con el objetivo de identificar patrones relevantes entre variables hematológicas y el diagnóstico de dengue, se presentan los estadísticos descriptivos agrupados según el resultado de la variable final Dengue Diagnosis. Estas comparaciones permiten explorar si existen diferencias sustanciales entre pacientes con y sin dengue.

Variable Hemoglobina

En promedio, los niveles de hemoglobina son similares entre pacientes con y sin diagnóstico de dengue (13.71 vs 13.72 g/dL), con medianas prácticamente iguales y una ligera mayor dispersión en el grupo sin dengue. No se observaron diferencias relevantes en hemoglobina entre los grupos.

Cuadro 1: Resumen estadístico por grupo diagnóstico de Hemoglobin

Diagnóstico de Dengue	Haemoglobin media	Haemoglobin sd	Haemoglobin median
0	13.72	1.53	13.80
1	13.71	1.47	13.70
NA	13.14	1.28	13.65

Variable WBC Count

Los pacientes con diagnóstico positivo de dengue presentan un recuento promedio de glóbulos blancos notablemente menor (2851.91 células/ μ L) en comparación con quienes no tienen dengue (7448.40 células/ μ L). Esta diferencia también se refleja en las medianas (2900 vs 7500) y en una menor dispersión. Estas observaciones sugieren que el recuento de leucocitos podría ser útil para discriminar entre casos positivos y negativos de dengue.

Cuadro 2: Resumen estadístico por grupo diagnóstico de WBC Count

Diagnóstico de Dengue	WBC Count mean	WBC Count sd	WBC Count median
0	7448.40	1477.83	7500
1	2851.91	494.03	2900
NA	3064.29	495.53	3150

Variable Platelet Count

Los pacientes con diagnóstico de dengue presentan un recuento de plaquetas mucho menor (media: 65,862; mediana: 64,000) en comparación con los pacientes sin dengue (media: 216,246; mediana: 180,000). Además, la dispersión es más alta en el grupo sin dengue. Estas diferencias refuerzan el potencial del recuento plaquetario como variable útil para discriminar casos positivos.

Cuadro 3: Resumen estadístico por grupo diagnóstico de WBC Count

Diagnóstico de Dengue	Platelet Count mean	Platelet Count sd	Platelet Count median
0	216246.52	78897.00	180000
1	65862.07	38769.77	64000
NA	76815.38	31072.12	83000

Variable PDW

El valor de PDW, que representa la variabilidad en el tamaño de las plaquetas, muestra diferencias claras entre los grupos. En los pacientes sin dengue, la media y mediana son considerablemente más altas (30.29 y 35, respectivamente), mientras que en los pacientes con diagnóstico positivo se reducen a 19.34 y 17. Esto sugiere una menor diversidad en el tamaño plaquetario en presencia de la enfermedad. Estas diferencias podrían ser relevantes para entender la dinámica hematológica asociada al dengue.

Cuadro 4: Resumen estadístico por grupo diagnóstico de WBC Count

Diagnóstico de Dengue	Platelet Count mean	Platelet Count sd	Platelet Count median
0	30.29	16.13	35
1	19.34	12.53	17
NA	15.32	3.24	16

5.3.2. Gráficos de caja

Se crean diagramas de caja en la Figura 5 para las variables hematológicas numéricas, visualizando su distribución entre pacientes con y sin diagnóstico positivo de dengue (Dengue Diagnosis) y como confirmación visual de la de análisis anterior. Los pacientes con dengue presentan menores valores de WBC, recuento de plaquetas y PDW. La hemoglobina no muestra diferencias relevantes entre grupos

Se identifican valores atípicos, especialmente en el recuento de plaquetas en pacientes sin dengue (superiores a 400,000 células/ μ L), y en PDW dentro del grupo positivo. Aunque estos valores no alteran las medianas, podrían deberse a respuestas fisiológicas particulares o errores de medición.

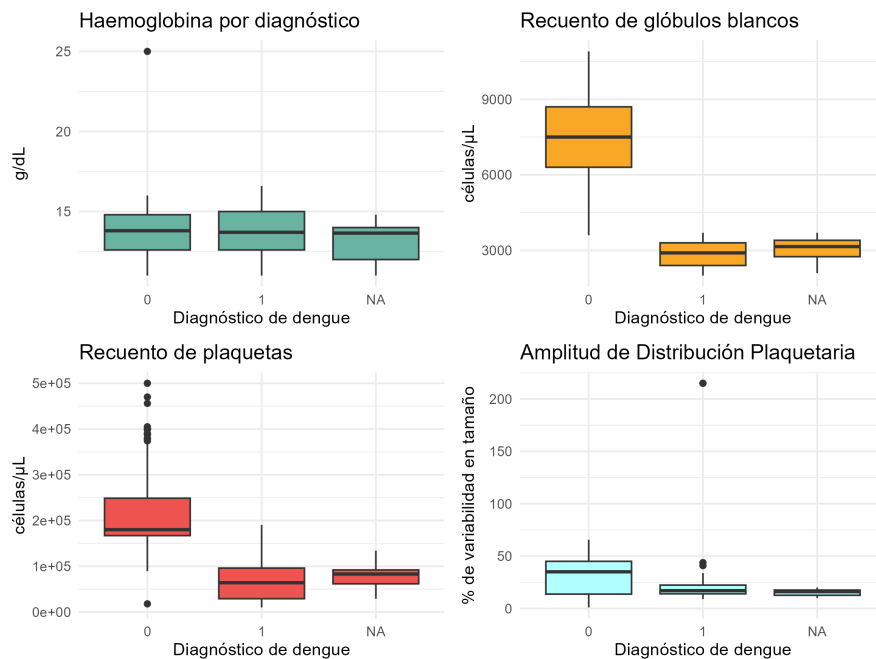


Figura 5: Gráficos de caja de las variables Hemoglobin y WBC Count

Fuente: Elaboración propia.

5.3.3. Conclusión de Análisis Exploratorio

A partir de las tablas y gráficos presentados, se observa que variables como el recuento de glóbulos blancos, el recuento de plaquetas y la amplitud de distribución plaquetaria muestran diferencias claras entre los grupos de diagnóstico, lo que sugiere su potencial utilidad en la etapa de modelado. En cambio, variables como la hemoglobina no presentan

patrones diferenciales marcados, por lo que podrían tener un menor valor discriminativo como primer acercamiento.

En cuanto a las variables binarias, se identificaron patrones diferenciales principalmente en ‘Differential Count’, cuya proporción de casos alterados fue mayor entre los pacientes con diagnóstico positivo. En contraste, variables como ‘RBC PANEL’ mostraron distribuciones más similares entre grupos. Estas observaciones respaldan la inclusión de algunas variables binarias como posibles predictores categóricos, aunque su utilidad deberá evaluarse rigurosamente en la etapa de modelado.

Dado que algunas variables muestran alta correlación, será importante aplicar estrategias que consideren la multicolinealidad. Las observaciones extraídas en esta etapa exploratoria servirán como base para la definición de un modelo. En particular, la selección de variables y el tratamiento de posibles colinealidades serán aspectos clave para garantizar la validez del enfoque estadístico.

Referencias

- Arrubla Hoyos, W. d. J. (2022). Conceptualización del diagnóstico del Dengue desde una perspectiva de la ingeniería y las nuevas tecnologías. *Journal of Computer and Electronic Sciences: Theory and Application*, 3(1), 1-8. <https://doi.org/10.17981/cesta.03.01.2022.01>
- Assaduzzaman, M., Islam, O., Nirob, M. A. S., Mim, M. M. H., & Mahmud, A. (2024). A benchmark dataset for analyzing hematological responses to dengue fever in Bangladesh. *Data in Brief*, 57, 111030. <https://doi.org/10.1016/j.dib.2024.111030>
- Chaloemwong, J., Tantiworawit, A., Rattanathammethee, T., Hantrakool, S., Chai-Adisaksopha, C., & Rattarittamrong, E. (2018). Useful clinical features and hematological parameters for the diagnosis of dengue infection in patients with acute febrile illness: A retrospective study. *BMC Hematology*, 18(1), 20. <https://doi.org/10.1186/s12878-018-0116-1>
- Gonçalves, R. P., de Oliveira, J. M., de Oliveira, J. S., & de Almeida, R. M. (2012). Dengue: Profile of hematological and biochemical dynamics. *Revista Brasileira de Hematologia e Hemoterapia*, 34(1), 36-41. <https://doi.org/10.5581/1516-8484.20120011>
- Irshad, M., Sattar, A., Nadeem, I., & Ishfaq, I. S. (2023). Evaluation of biochemical and hematological dynamics as markers for early detection of severe dengue. *International Journal of Clinical and Pathology*, 10(2), 45-52. <https://medicopublication.com/index.php/ijcpath/article/view/20937>
- Organizacion Mundial de la Salud. (2024, marzo). Dengue y dengue grave [Consultado el 1 de junio de 2025]. <https://www.who.int/es/news-room/fact-sheets/detail/dengue-and-severe-dengue>