



Departamento de Posgrados

Maestría en Estadística Aplicada

Análisis estadístico de la mortalidad fetal: evaluación de características maternas, edad gestacional y principales causas de defunción en Ecuador

Trabajo de titulación previo a la obtención del grado de Magíster
en Estadística Aplicada

Autora:

Eduarda Micaela Dávila Rodríguez

Director:

David Christopher Siddons

Cuenca – Ecuador

2025

DEDICATORIA

A Locky, quien es y será siempre la historia de amor
más mágica de mi vida. Gracias por mover la llamita,
siempre.

A Gorda, por regalarme cada día lleno de amor y
compañía. Que sean muchos más.

A Patricio y Margarita, porque el cielo siempre es
bonito gracias a ustedes.

AGRADECIMIENTO

A mis padres, Patricio y Martha, porque gracias a ellos estoy donde estoy. Espero algún día devolverles todo lo que me han dado, y mucho más.

A Juan David, quien me acompaña en cada paso, permitiéndome ser y recordándome siempre que la vida también se escribe con ternura.

A los docentes de la maestría, por compartir no solo su conocimiento, sino también su pasión por la investigación y el rigor científico.

RESUMEN

El estudio analiza la mortalidad fetal en Ecuador entre 2019 y 2023, identificando factores clínicos, sociales y territoriales asociados. Se aplicaron métodos estadísticos y de aprendizaje automático —como regresión, VAR, Random Forest, K-means y KNN— para evaluar relaciones, predecir causas y clasificar perfiles de riesgo. Más del 50% de las defunciones se codificaron como “no especificadas”, lo que limita la inferencia por causa y subraya la necesidad de mejorar la codificación en INEC/MSP. Los resultados evidencian que la edad materna, los controles prenatales y las desigualdades en el acceso a la salud influyen significativamente en las defunciones fetales. Se concluye que el fortalecimiento de los registros clínicos y la equidad sanitaria son esenciales para reducir muertes fetales evitables.

Palabras clave: mortalidad fetal, análisis estadístico, aprendizaje automático, salud materna, Ecuador, predicción de riesgo, salud pública.

ABSTRACT

This study analyzes fetal mortality in Ecuador from 2019 to 2023, identifying clinical, social, and territorial factors associated with it. Statistical and machine learning methods—such as regression, VAR, Random Forest, K-means, and KNN—were applied to evaluate relationships, predict causes, and classify risk profiles. Over 50% of fetal deaths were coded as “unspecified,” which limits cause-specific inference and highlights the need to improve coding practices at INEC/MSP. Results show that maternal age, prenatal care, and inequalities in healthcare access significantly influence fetal deaths. Strengthening clinical records and promoting equitable healthcare are essential to reduce preventable fetal mortality.

Keywords: fetal mortality, statistical analysis, machine learning, maternal health, Ecuador, risk prediction, public health.



David Siddons
Director de Tesis

Índice de Contenidos

RESUMEN	IV
ABSTRACT.....	IV
ÍNDICE DE FIGURAS.....	VII
ÍNDICE DE TABLAS	VIII
1. INTRODUCCIÓN.....	1
2. OBJETIVOS	3
2.1. OBJETIVO GENERAL.....	3
3. LABORATORIO: ENFOQUE DESCRIPTIVO E INFERENCIAL	3
3.1. OBJETIVOS ESPECÍFICOS	3
3.2. MÉTODO	3
3.2.1. <i>Descripción de los datos</i>	3
3.3. DISCUSIÓN Y RESULTADOS.....	5
3.3.1. <i>Estadística Descriptiva</i>	5
3.3.2. <i>Relación entre la semana de gestación y las 5 principales causas de defunción fetales</i>	10
3.3.3. <i>Características maternas asociadas a un mayor número de defunciones fetales</i>	12
3.3.4. <i>Regresión lineal múltiple</i>	14
3.4. CONCLUSIONES DEL LABORATORIO.....	15
4. LABORATORIO: ENFOQUE MACHINE LEARNING	15
4.1. OBJETIVOS ESPECÍFICOS	15
4.2. MÉTODO	15
4.2.1. <i>Selección de los datos</i>	15
4.2.2. <i>Selección de técnicas</i>	17
4.2.3. <i>Descripción de cada técnica aplicada</i>	18
4.3. DISCUSIÓN Y RESULTADOS.....	18
4.3.1. <i>Modelo VAR</i>	19
4.3.2. <i>Random Forest</i>	23
4.3.3. <i>Algoritmo K-means</i>	25
4.3.4. <i>Discusión</i>	27
4.4. CONCLUSIONES DEL LABORATORIO.....	27
5. LABORATORIO: ENFOQUE TOMA DE DECISIONES.....	28
5.1. OBJETIVOS ESPECÍFICOS	28
5.2. MÉTODO	28
5.2.1. <i>Selección de los datos</i>	28
5.2.2. <i>Selección de técnicas</i>	28

5.2.3.	<i>Descripción de cada técnica aplicada</i>	29
5.3.	DISCUSIÓN Y RESULTADOS.....	29
5.3.1.	<i>Clústeres</i>	29
5.3.2.	<i>K-Nearest Neighbors (KNN)</i>	30
5.3.3.	<i>Discusión</i>	32
5.4.	CONCLUSIONES DEL LABORATORIO.....	34
6.	LABORATORIO: ENFOQUE ROBUSTEZ Y VALIDACIÓN	34
6.1.	OBJETIVOS.....	34
6.2.	MÉTODO	34
6.2.1.	<i>Validación de modelos</i>	34
6.3.	DISCUSIÓN Y RESULTADO.....	35
6.3.1.	<i>Regresión lineal múltiple</i>	35
6.3.2.	<i>Modelo de Autorregresión Vectorial (VAR)</i>	37
6.3.3.	<i>Árbol de decisión</i>	41
6.3.4.	<i>K-means</i>	42
6.3.5.	<i>K-Nearest Neighbors (KNN)</i>	43
6.3.6.	<i>Discusión</i>	45
6.4.	CONCLUSIONES DEL LABORATORIO.....	45
6.5.	LIMITACIONES DEL ESTUDIO	46
7.	CONCLUSIÓN GENERAL	46
8.	BIBLIOGRAFÍA	47

Índice de figuras

Figura 1	Distribución del promedio de controles prenatales por rango de semanas de gestación.....	6
Figura 2	Frecuencia de la variable nivel instructivo	7
Figura 3	Frecuencia de la variable área	7
Figura 4	Frecuencia de la variable etnia.....	8
Figura 5	Boxplot con jitter plot de la variable control prenatal.....	8
Figura 6	Boxplot con jitter plot de la variable hijo nacido muerto.....	9
Figura 7	Evolución temporal de las variables control prenatal, edad de la madre e hijo nacido muerto	9
Figura 8	Gráfico de las cinco principales causas de defunciones fetales	10
Figura 9	Matriz de Correlación de variables numéricas	12
Figura 10	Gráfico de hijos nacidos muertos por etnia	13
Figura 11	Gráfico de hijos nacidos muertos por área.....	13
Figura 12	Gráfico de hijos nacidos muertos por nivel instructivo.....	14
Figura 13	Peso de todas las variables del dataset con respecto a la variable Causa Fetal.....	16
Figura 14	Series de tiempo de la región Amazonía, Costa y Sierra del año 2019 al 2023.....	19
Figura 15	Predicciones para la región Sierra comparado con los valores reales	20
Figura 16	Predicciones para la región Costa comparado con los valores reales.....	21
Figura 17	Predicciones para la región Amazonía comparado con los valores reales	22
Figura 18	Distribución de Clústeres.....	27
Figura 19	Clasificación por edad materna	30
Figura 20	Clasificación por semana de gestación.....	30
Figura 21	Gráfica de QQ del modelo de regresión múltiple.....	36
Figura 22	ACF de los residuos del modelo de regresión lineal múltiple	36
Figura 23	Gráfica de QQ del modelo de VAR Sierra	38
Figura 24	ACF de los residuos del modelo de VAR Sierra	38
Figura 25	Gráfica de QQ del modelo de VAR Costa	39
Figura 26	ACF de los residuos del modelo de VAR Costa	40
Figura 27	Gráfica de QQ del modelo de VAR Amazonía.....	41
Figura 28	ACF de los residuos del modelo de VAR Amazonía.....	41
Figura 29	Gráfica de QQ del modelo K-means.....	43
Figura 30	ACF de los residuos del modelo K-means	43
Figura 31	Gráfica de QQ del modelo KNN	44
Figura 32	ACF de los residuos del modelo KNN	45

Índice de tablas

Tabla 1	Diccionario de datos	4
Tabla 2	Estadística descriptiva de las variables de interés	6
Tabla 3	Prueba de Post Hoc de Dunn	11
Tabla 4	Análisis de colinealidad	15
Tabla 5	RMSE de las variables analizadas en la región Sierra	20
Tabla 6	RMSE de las variables analizadas en la región Costa	22
Tabla 7	RMSE de las variables analizadas en la región Amazonía	23
Tabla 8	Matriz de confusión para Random Forest con baja precisión (accuracy)	23
Tabla 9	Matriz de confusión para Random Forest	24
Tabla 10	Índice de Silhouette por Clúster	25
Tabla 11	Clasificación de los casos en clústeres	26
Tabla 12	Centroides de clústeres según características maternas y gestacionales	26
Tabla 13	Matriz de confusión para el conjunto de entrenamiento	31
Tabla 14	Matriz de confusión para el conjunto de prueba	31
Tabla 15	Evaluación de supuestos para el modelo de regresión lineal múltiple	35
Tabla 16	ANOVA para el modelo de regresión lineal múltiple	36
Tabla 17	Evaluación de supuestos para el modelo VAR Sierra	37
Tabla 18	Evaluación de supuestos para el modelo VAR Costa	39
Tabla 19	Evaluación de supuestos para el modelo VAR Amazonía	40
Tabla 20	Análisis de errores del modelo de árbol de decisión	42
Tabla 21	Evaluación de supuestos para el modelo K-means	42
Tabla 22	ANOVA para el modelo K-means	43
Tabla 23	Evaluación de supuestos para el modelo KNN	44
Tabla 24	ANOVA para el modelo de KNN	45

ANÁLISIS ESTADÍSTICO DE LA MORTALIDAD FETAL: EVALUACIÓN DE CARACTERÍSTICAS MATERNAS, EDAD GESTACIONAL Y PRINCIPALES CAUSAS DE DEFUNCIÓN EN ECUADOR

1. Introducción

La mortalidad fetal constituye un problema de salud pública constante y sensible, cuya complejidad exige un abordaje integral desde la estadística aplicada, la epidemiología y las ciencias sociales. Este proyecto adquiere relevancia al buscar comprender los factores que inciden en las muertes fetales ocurridas en Ecuador entre 2019 y 2023, mediante la aplicación de métodos estadísticos y de aprendizaje automático permitiendo identificar patrones, causas y perfiles de riesgo. En un contexto donde las desigualdades en salud siguen marcando la realidad nacional, disponer de información confiable y modelos predictivos robustos puede contribuir a mejorar las estrategias de prevención y la planificación sanitaria.

La importancia del presente trabajo radica en la necesidad de contar con evidencia cuantitativa sobre la mortalidad fetal, un indicador sensible del bienestar materno-infantil y de la calidad del sistema de salud. En Ecuador, a pesar de los avances en el control prenatal y la atención obstétrica, la tasa de defunciones fetales presenta variaciones regionales significativas (INEC, 2023). Comprender los factores asociados —como la edad materna, las semanas de gestación, el número de controles prenatales o las condiciones sociales— resulta crucial para orientar políticas públicas y acciones médicas oportunas.

Los datos analizados en este estudio provienen de la base de datos oficial de defunciones fetales del Instituto Nacional de Estadística y Censos (INEC), descargada directamente desde su portal institucional. El conjunto de datos abarca 7080 registros correspondientes al período 2019–2023, lo que permite un análisis representativo de las tendencias y características maternas, gestacionales y socioeconómicas relacionadas con la mortalidad fetal en el país.

A nivel global, la Organización Mundial de la Salud (2024) estima que cada año se producen alrededor de 1,9 millones de muertes fetales, el 98% en países de ingresos bajos y medios. Estas cifras evidencian la necesidad de fortalecer los sistemas de vigilancia epidemiológica y la investigación aplicada que permita identificar patrones de riesgo específicos en cada contexto.

Diversas investigaciones recientes confirman la importancia de abordar la mortalidad fetal desde un enfoque interdisciplinario que combine factores clínicos, sociales y tecnológicos. En una revisión sistemática, Mukherjee et al. (2024) identificaron que los principales determinantes de la muerte fetal en África subsahariana incluyen la edad materna avanzada, la hipertensión, la baja escolaridad y el acceso limitado a servicios de salud, evidenciando la interacción entre condiciones clínicas y desigualdades estructurales. Xue et al. (2023) demostraron que las disparidades geográficas y ambientales, como la exposición al humo de biomasa, incrementan el riesgo de mortinatos en países de ingresos medios y bajos. Gwako

et al., (2021) hallaron que la baja calidad y la escasa utilización de los servicios prenatales se asocian directamente con mayores tasas de muerte fetal, lo que refuerza la necesidad de fortalecer la atención gestacional. Desde la perspectiva tecnológica, Cersonsky (2023) aplicó modelos de machine learning para predecir el riesgo de muerte fetal con mayor precisión que los métodos tradicionales. Finalmente, de Mucio et al. (2023), en un estudio multicéntrico en América Latina, evidenciaron que complicaciones como la preeclampsia, la diabetes y las desigualdades en la atención médica siguen siendo los principales factores asociados a la mortalidad fetal en contextos de bajos recursos.

Estos estudios coinciden en que el fenómeno no puede comprenderse únicamente desde lo clínico, sino que requiere integrar variables sociales, económicas y territoriales para explicar la heterogeneidad de los resultados.

El contexto ecuatoriano se caracteriza por marcadas desigualdades sociales y territoriales que condicionan la salud materno-infantil. En zonas rurales y amazónicas, la escasez de centros obstétricos y las brechas culturales entre pacientes y personal sanitario dificultan el acceso oportuno al control prenatal (MSP, 2023). Factores como el nivel educativo, la situación conyugal y la etnia influyen en las prácticas de autocuidado y en la detección temprana de complicaciones del embarazo. El área urbana, aunque concentra más registros, no necesariamente refleja una mejor calidad de atención, sino una mayor capacidad de registro institucional.

La fortaleza de esta investigación radica en su enfoque integrador, que combina estadística descriptiva, inferencial y modelos de aprendizaje automático (VAR, Random Forest, K-Means y KNN) para abordar un fenómeno complejo desde múltiples perspectivas. En un país donde los datos administrativos presentan limitaciones, el uso de técnicas robustas de modelado y validación permite analizar tendencias y generar herramientas predictivas útiles para la toma de decisiones. Este trabajo se alinea con los Objetivos de Desarrollo Sostenible (ODS 3.1 y 3.2), orientados a reducir la mortalidad materna y neonatal mediante sistemas de salud más equitativos y basados en evidencia. (Naciones Unidas, 2015)

A pesar de los esfuerzos del sistema sanitario ecuatoriano, la mortalidad fetal sigue siendo un evento subregistrado y con causas poco claras. La ausencia de un registro unificado y la falta de estandarización en los certificados de defunción limitan la capacidad de análisis y de formulación de estrategias efectivas. El problema central que aborda este proyecto es la identificación de factores determinantes y patrones asociados a la mortalidad fetal en Ecuador, a partir de información oficial entre 2019 y 2023. Se busca responder a la pregunta:

¿Cómo influyen las características maternas, la edad gestacional y las condiciones socioeconómicas en la probabilidad y causa de defunción fetal?

El estudio se limita a los registros nacionales de defunciones fetales proporcionados por el INEC. Se analizan variables maternas (edad, nivel educativo, etnia), gestacionales (semanas de gestación, número de controles prenatales) y contextuales (área geográfica, tipo de

establecimiento). Su alcance es exploratorio, analítico y predictivo: busca identificar asociaciones, generar modelos explicativos y construir perfiles de riesgo que orienten futuras políticas de salud pública. No se pretende establecer causalidad directa, sino evidenciar patrones estadísticos significativos y proponer líneas de acción.

Finalmente, la estructura general del documento se compone de varios apartados. En primer lugar, se presenta la introducción, donde se expone el contexto, la justificación y los objetivos del estudio. Posteriormente, se desarrolla el objetivo general del proyecto y se incluyen cuatro secciones correspondientes a los laboratorios realizados, cada una con sus objetivos específicos, descripción metodológica, resultados, discusión, conclusiones parciales y limitaciones. Por último, se incorpora una conclusión general que sintetiza los principales hallazgos y limitaciones del estudio, junto con la bibliografía empleada.

2. Objetivos

2.1. Objetivo general

Estudiar la mortalidad fetal desde una perspectiva multifactorial, evaluando características maternas, edad gestacional y causas principales de defunción fetal, con el objetivo de comprender su impacto y posibles patrones.

3. Laboratorio: Enfoque Descriptivo e Inferencial

3.1. Objetivos específicos

- Evaluar la relación entre la semana de gestación y las 5 principales causas de defunción fetal.
- Identificar las características maternas asociadas a un mayor número de defunciones fetales.

3.2. Método

3.2.1. Descripción de los datos

La base de datos de defunciones fetales (EDF) descargada de la página web del INEC, está compuesta por 7080 observaciones, correspondientes al total de defunciones inscritas entre los años 2019 y 2023. Cuenta con 14 variables, como se muestra en la Tabla 1.

Tabla 1*Diccionario de datos*

Variable	Descripción	Tipo	Escala de medición
anio	Año de ocurrencia	Categórica	2019 a 2023
provincia	Provincia de fallecimiento	Categórica	23 provincias
canton	Cantón de fallecimiento	Categórica	138 cantones
area	Área de fallecimiento	Categórica	rural / urbano
sexo	Sexo	Categórica	femenino / masculino
sem_gest	Periodo de gestación	Numérica	[12 - 42]
producto_emb	Producto del embarazo	Categórica	4 tipos de productos de embarazo
asis_por	Asistido por	Categórica	7 tipos de asistencia
establecimiento	Lugar de ocurrencia	Categórica	7 tipos de establecimiento
causa_fetal	Causa que ocasionó la defunción fetal	Categórica	136 causas
edad_mad	Edad de la madre	Numérica	[12 - 52]
hij_viv	Cuántos hijos nacidos vivos	Numérica	[0 – 13]
hijo_nacm	Cuántos hijos nacieron muertos (previamente)	Numérica	[1 - 6]
control_pren	Cuántos controles prenatales recibió en este embarazo	Numérica	[0 - 25]
etnia	Autoidentificación étnica de la madre	Categórica	9 etnias
nivel_instructivo	Nivel de instrucción alcanzado	Categórica	9 tipos de nivel de instrucción

Se realizó una limpieza de datos que incluyó renombrar columnas para facilitar la interpretación, filtrar valores atípicos según el contexto, convertir tipos de datos, eliminar datos faltantes, y verificar que no existieran observaciones duplicadas. Las incongruencias fueron eliminadas, ya que representaban menos del 1% de la base.

Adicionalmente, se agruparon las semanas de gestación en cuatro categorías: "Extremadamente pre-término" (<28 semanas), "Pre-término" (28-36 semanas), "Término" (37-42 semanas) y "Post-término" (>42 semanas).

3.2.1.1 Estadística descriptiva

Se calcularon estadísticos descriptivos como media, mediana y desviación estándar para comprender características principales de las variables de interés, que son: "sem_gest, causa_fetal, hijo_nacm, edad_mad, control_pren, etnia, area, nivel_instructivo".

3.2.1.2 Análisis exploratorio

Se identificaron y manejaron los valores atípicos en variables numéricas mediante el cálculo del rango intercuartílico. El análisis de distribución de las variables categóricas mediante gráficos de barras y la evaluación de la relación entre semanas de gestación y causas principales de defunción fetal, utilizando pruebas estadísticas como Kruskal-Wallis y el post hoc de Dunn.

3.2.1.3 Análisis de correlación

Como los datos no siguen una distribución normal, se calculó una matriz de correlación con el método de Spearman para explorar relaciones entre variables numéricas.

3.2.1.4 Modelo

Se ajustó un modelo lineal para evaluar la influencia de variables maternas en el número de hijos nacidos muertos y se comprobó la multicolinealidad de las variables independientes utilizando el factor de inflación de la varianza (VIF).

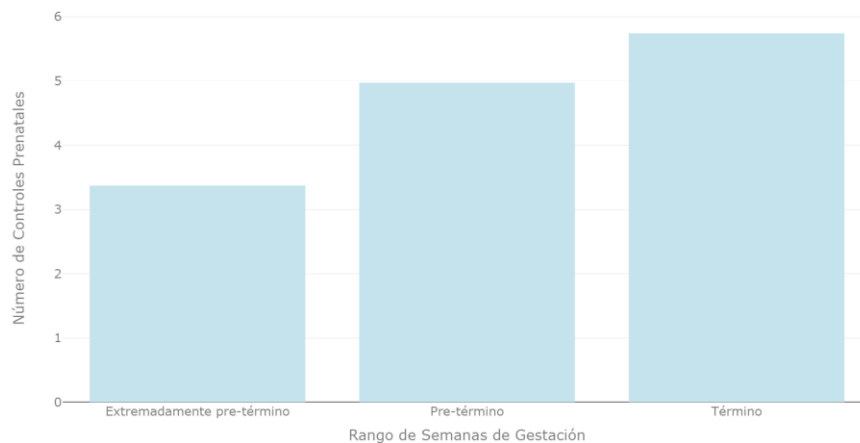
3.3. Discusión y resultados

3.3.1. Estadística Descriptiva

La limpieza de los datos permitió trabajar con un conjunto sin valores faltantes y con valores atípicos manejados adecuadamente.

Figura 1

Distribución del promedio de controles prenatales por rango de semanas de gestación



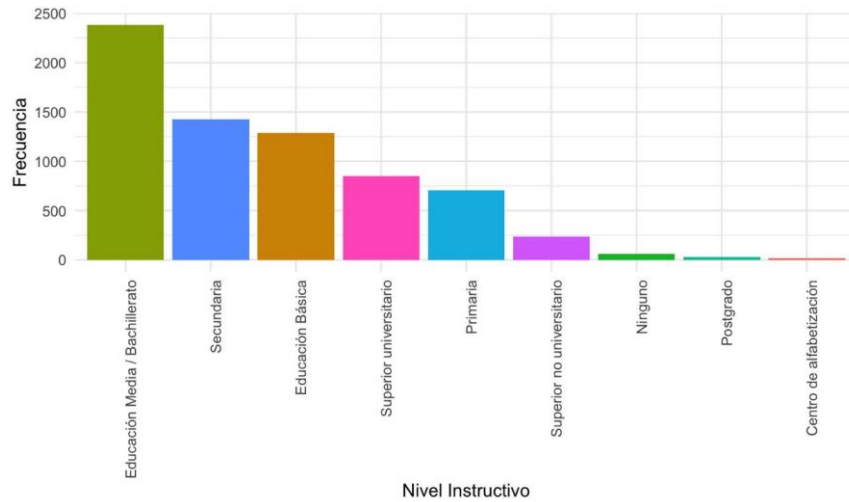
La Figura 1 presenta la distribución del promedio de controles prenatales por rango de semanas de gestación, con mayor concentración en “Término”, seguida de “Pre-término” y “Extremadamente pre-término”. Las gestaciones que alcanzan el término acumulan más controles porque transcurren más semanas y el esquema habitual de seguimiento incrementa la frecuencia hacia el final del embarazo. Por ello, la diferencia en el conteo total responde principalmente a la duración gestacional y no constituye evidencia directa sobre la calidad o equidad de la atención. Para una comparación más informativa entre grupos, se requieren tasas ajustadas por semanas o un estimador de controles esperados según edad gestacional.

Tabla 2

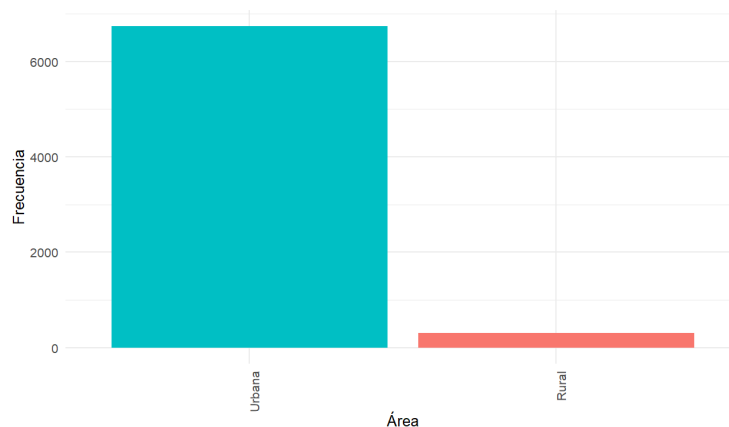
Estadística descriptiva de las variables de interés

Variable	Media	Mediana	Desviación Estándar
hijo_nacm	1,0852	1	0,3406
edad_mad	27,2434	26	7,2736
control_pren	4,5333	5	2,7856

En la Tabla 2 se presentan los estadísticos descriptivos de las variables de interés para el objetivo 1. Estos valores indican que, aunque la variabilidad en el número de hijos nacidos muertos es baja, las variables edad de la madre y controles prenatales tienen una mayor dispersión, sugiriendo diferencias significativas entre las observaciones individuales.

Figura 2*Frecuencia de la variable nivel instructivo*

La Figura 2 muestra la frecuencia de la variable nivel instructivo, destacando que la mayoría de las observaciones se concentran en la categoría Educación Media/Bachillerato, con casi 2,500 registros. A esta le siguen las categorías Secundaria, Educación Básica, Superior Universitario y Primaria, reflejando una mayor proporción en niveles educativos intermedios y avanzados. En contraste, las categorías con menor frecuencia incluyen Ninguno, Postgrado y Centro de alfabetización.

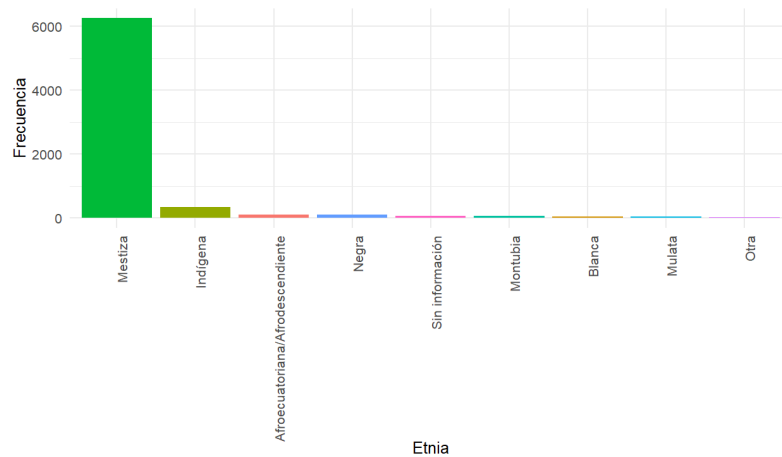
Figura 3*Frecuencia de la variable área*

La Figura 3 muestra la frecuencia de registros por área de residencia, con una mayor concentración en el entorno urbano (más de 6.000 observaciones) frente al rural. Este patrón describe cómo se registran los eventos y no equivale necesariamente a un riesgo más alto en

las ciudades. La diferencia puede deberse a mejor captación y notificación en ámbitos urbanos y a menor captación en zonas rurales, lo que puede sesgar el perfil observado.

Figura 4

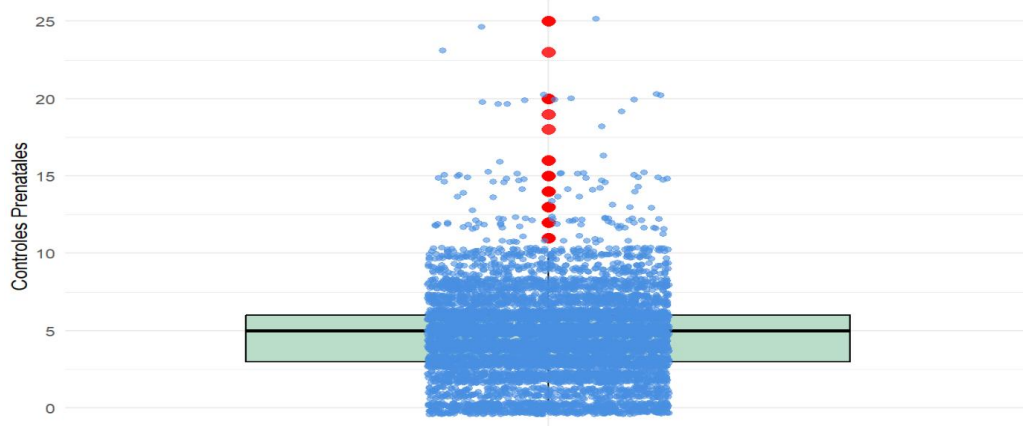
Frecuencia de la variable etnia



La Figura 4 representa la frecuencia de la variable etnia, en la cual se observa que la mayoría de las observaciones corresponden a la categoría Mestiza, con más de 6,000 casos registrados. Este predominio es coherente con la composición demográfica del Ecuador, donde la población es mayoritariamente mestiza, especialmente en el caso de las mujeres. Por el contrario, las demás categorías —Indígena, Afroecuatoriana/Afrodescendiente, Negra, Blanca y otras— presentan frecuencias significativamente menores, reflejando una distribución altamente desigual, que sería inusual si las proporciones fueran más equitativas entre grupos étnicos.

Figura 5

Boxplot con jitter plot de la variable control prenatal

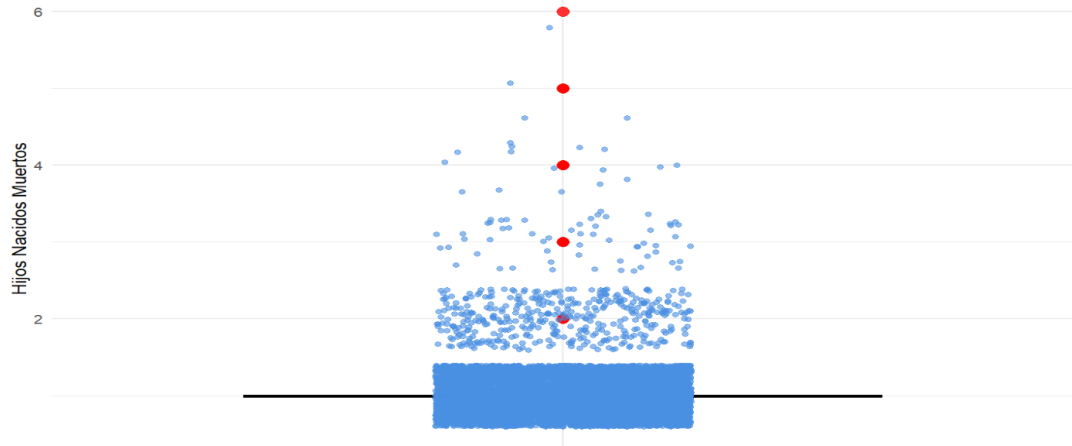


La Figura 5 presenta la distribución de los controles prenatales, donde se observa una gran concentración de datos alrededor de los valores más comunes (entre 3 y 6), con

algunos valores extremos que representan casos con un número significativamente mayor de controles.

Figura 6

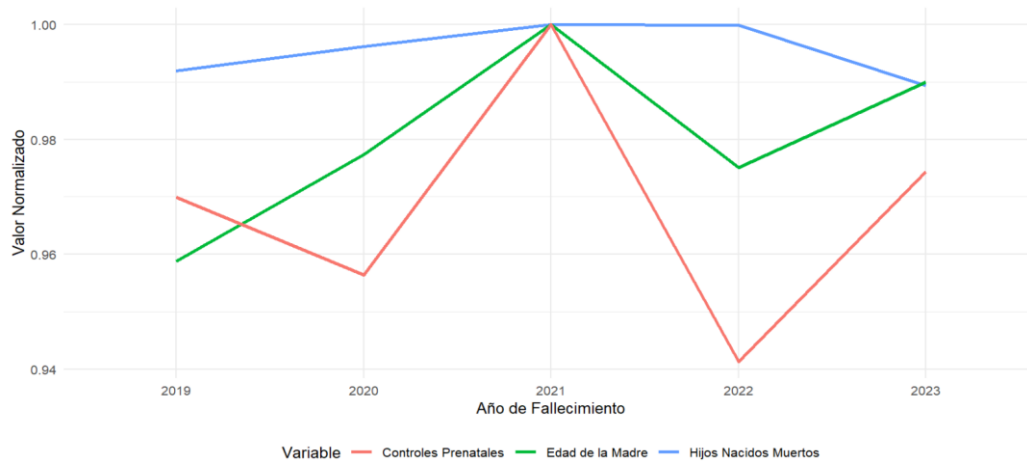
Boxplot con jitter plot de la variable hijo nacido muerto



La Figura 6 muestra la distribución de hijos nacidos muertos. Se observa una alta concentración en valores bajos, principalmente en 1, mientras que algunos valores extremos reflejan casos con un número considerablemente mayor de hijos nacidos muertos.

Figura 7

Evolución temporal de las variables control prenatal, edad de la madre e hijo nacido muerto

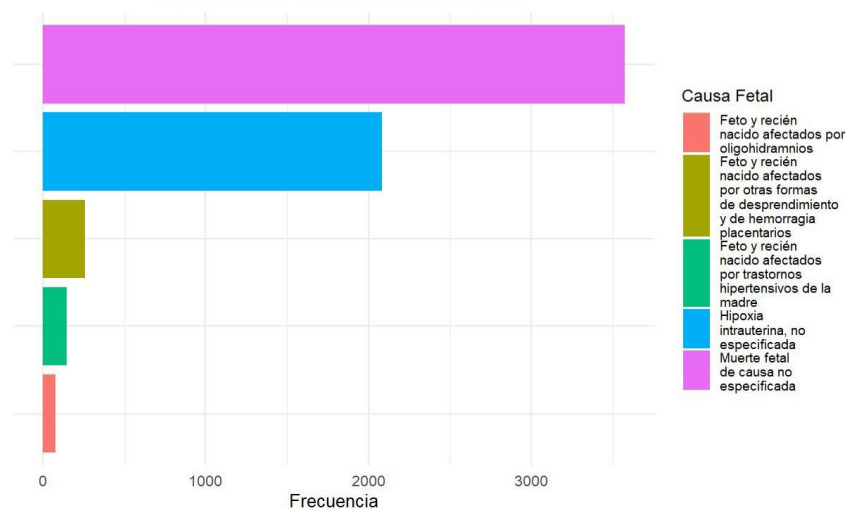


La Figura 7 muestra la evolución temporal de variables importantes, normalizadas para facilitar la comparación. Se observa que los valores de controles prenatales, edad de la madre y número de hijos nacidos muertos presentan variaciones consistentes a lo largo del tiempo, destacando un aumento en 2021 seguido de una estabilización en los años posteriores.

3.3.2. Relación entre la semana de gestación y las 5 principales causas de defunción fetales

Figura 8

Gráfico de las cinco principales causas de defunciones fetales



La Figura 8 muestra las principales causas de defunciones fetales, destacando Muerte fetal de causa no especificada como la más frecuente, seguida por Hipoxia intrauterina y Trastornos hipertensivos de la madre, entre otras.

Se aplicó la prueba de normalidad de Kolmogorov-Smirnov a la variable `sem_gest`, y al obtener un p-valor menor a 0,05 ($< 2,2e-16$), se rechazó la hipótesis nula de normalidad. Esto indica que la variable no sigue una distribución normal, por lo que se recomienda utilizar métodos no paramétricos para su análisis.

Posteriormente, se aplicó el test de Kruskal-Wallis, obteniendo un chi-cuadrado de 98,69 con 4 grados de libertad y un p-valor de $2,2e-16$, lo que indica diferencias estadísticamente significativas en las distribuciones de las semanas de gestación entre los grupos definidos por las causas fetales. Esto sugiere que al menos una de las causas tiene una distribución distinta respecto a las semanas de gestación.

Para identificar qué grupos presentan diferencias específicas, se realizó una prueba Post Hoc, ajustando los p-valores para comparaciones múltiples.

Tabla 3*Prueba de Post Hoc de Dunn*

Comparación	Z	P.unadj	P.adj
Feto y recién nacido afectados por oligohidramnios - Feto y recién nacido afectados por otras formas de desprendimiento y de hemorragia placentarios	-3,0600826	2,212760e-03	2,212760e-02
Feto y recién nacido afectados por oligohidramnios - Feto y recién nacido afectados por trastornos hipertensivos de la madre	-1,6588911	9,713774e-02	9,713774e-01
Feto y recién nacido afectados por otras formas de desprendimiento y de hemorragia placentarios - Feto y recién nacido afectados por trastornos hipertensivos de la madre	1,5635199	1,179304e-01	1,000000e+00
Feto y recién nacido afectados por oligohidramnios - Hipoxia intrauterina, no especificada	0,8006893	4,233116e-01	1,000000e+00
Feto y recién nacido afectados por otras formas de desprendimiento y de hemorragia placentarios - Hipoxia intrauterina, no especificada	7,4197871	1,173087e-13	1,173087e-12
Feto y recién nacido afectados por trastornos hipertensivos de la madre Hipoxia intrauterina, no especificada	3,7871809	1,523662e-04	1,523662e-03
Feto y recién nacido afectados por oligohidramnios - Muerte fetal de causa no especificada	-1,1214861	2,620810e-01	1,000000e+00
Feto y recién nacido afectados por otras formas de desprendimiento y de hemorragia placentarios - Muerte fetal de causa no especificada	4,1566883	3,228941e-05	3,228941e-04
Feto y recién nacido afectados por trastornos hipertensivos de la	1,2344780	2,170248e-01	1,000000e+00

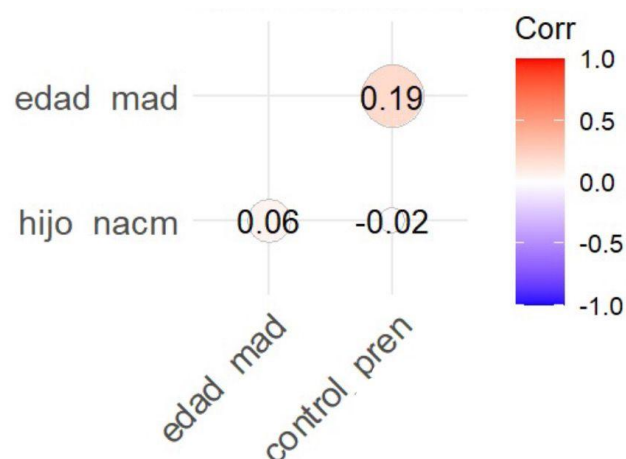
madre - Muerte fetal de causa no especificada			
Hipoxia intrauterina, no especificada - Muerte fetal de causa no especificada	-8,0044409	1,200106e-15	1,200106e-14

En la Tabla 3 se muestra el análisis post hoc con la prueba de Dunn, ajustada por Bonferroni, revelando diferencias significativas en las semanas de gestación entre varias causas fetales, como entre "Muerte fetal de causa no especificada" y otras categorías, así como entre "Hipoxia intrauterina" y "Hemorragias placentarias". Sin embargo, no se observan diferencias significativas en algunos pares, como "Oligohidramnios" y "Trastornos hipertensivos de la madre".

3.3.3. Características maternas asociadas a un mayor número de defunciones fetales

Figura 9

Matriz de Correlación de variables numéricas



La Figura 9 muestra la matriz de correlación entre las variables edad de la madre, número de hijos nacidos muertos y controles prenatales, destacando una correlación positiva leve entre edad de la madre y controles prenatales ($r = 0,19$). Sin embargo, los valores de correlación inferiores a 0,2 indican una relación débil o casi nula entre las demás variables, lo que sugiere que no existe una relación significativa entre ellas.

Dado que los datos no siguen una distribución normal, se aplicó el test de Kruskal-Wallis, cuyos resultados indican que no existen diferencias significativas entre las etnias ni entre los niveles de instrucción respecto al número de hijos nacidos muertos, ya que los p-valores 0,3169 y 0,621 respectivamente son mayores que 0,05. Sin embargo, se encontró una diferencia significativa entre el área de

residencia de la madre (urbana vs rural), con un p-value de 0,01348, lo que sugiere que el área influye en el número de hijos nacidos muertos. Los gráficos realizados muestran que la categoría "Mestiza" predomina en las defunciones fetales, con más de 6000 observaciones, mientras que las categorías "Afroecuatoriana/Afrodescendiente", "Blanca" e "Indígena" tienen frecuencias menores, y las categorías "Montubia", "Mulata", "Negra" y "Sin información" presentan frecuencias aún más bajas.

Figura 10

Gráfico de hijos nacidos muertos por etnia

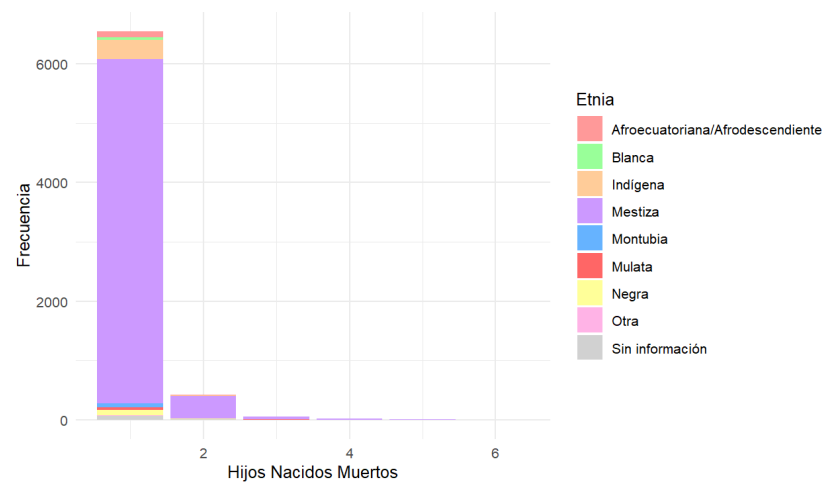
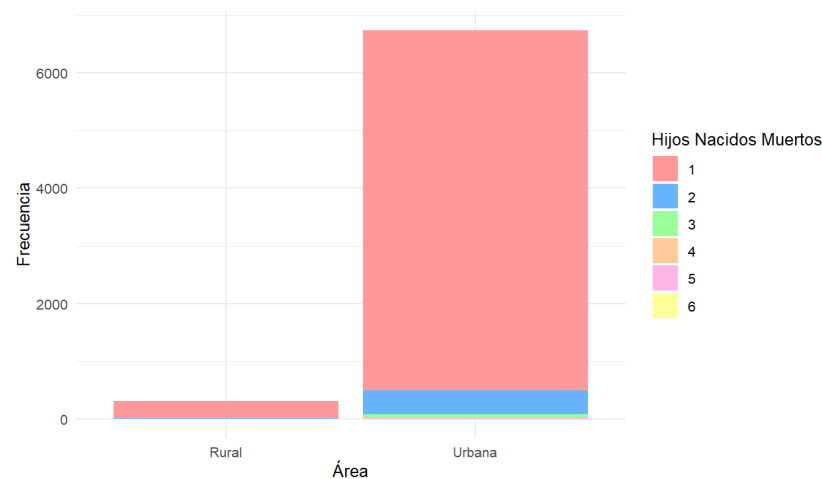


Figura 11

Gráfico de hijos nacidos muertos por área

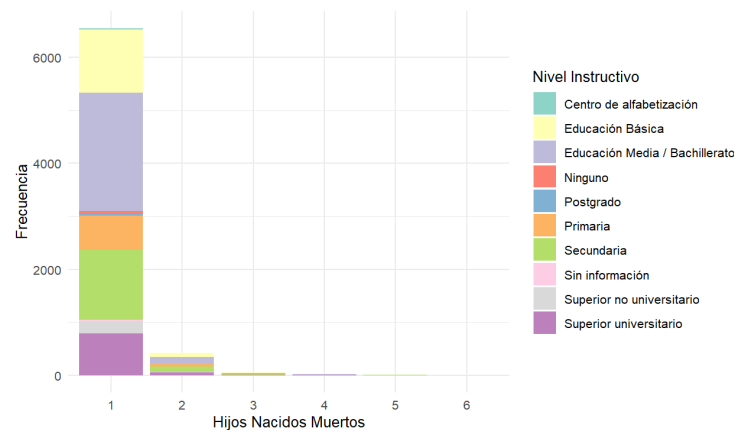


La Figura 11 muestra la distribución de los hijos nacidos muertos según el área (rural o urbana). La gran mayoría de las observaciones corresponden al área urbana, representando más de 6000 registros, mientras que en el área rural las frecuencias son mucho menores. Las categorías de "Hijos Nacidos Muertos" (de 1

a 6) se distribuyen de manera más uniforme en el área rural, mientras que, en el área urbana, la mayoría de los casos corresponden a la categoría "1".

Figura 12

Gráfico de hijos nacidos muertos por nivel instructivo



La Figura 12 muestra conteos por nivel educativo materno. La mayor presencia en Educación Básica y Media/Bachillerato es coherente con su mayor tamaño poblacional y no implica un riesgo más alto en esos grupos. Para evaluar diferencias de riesgo entre niveles se requieren proporciones por estrato o tasas por 1.000 nacidos vivos que controlen el tamaño de cada grupo.

3.3.4. Regresión lineal múltiple

El modelo de regresión lineal múltiple, con `hijo_nacm` como variable dependiente y `edad_mad`, `control_pren` y `area` como variables independientes, mostró que tanto la edad de la madre, los controles prenatales y el área urbana son factores significativos para predecir los hijos nacidos muertos. La fórmula utilizada fue:

$$\text{hijo_nacm} = \beta_0 + \beta_1 \cdot \text{edad_mad} + \beta_2 \cdot \text{control_pren} + \beta_3 \cdot \text{etnia} + \beta_4 \cdot \text{area} + \beta_5 \cdot \text{nivel_instructivo}$$

El modelo tiene un R-cuadrado ajustado de 0,0047, lo que indica una baja capacidad explicativa, aunque el F-statistic (12,04, $p = 7,374e-08$) sugiere que el modelo es globalmente significativo.

Tabla 4*Análisis de colinealidad*

	GVIF	Df	GVIF^{^(1/2*Df)}
edad_mad	1,457880	1	1,207427
control_pren	1,086023	1	1,042124
etnia	1,118715	8	1,007036
area	1,022365	1	1,011121
nivel_instructivo	1,628493	9	1,027462

Se realizó el análisis de colinealidad, donde la Tabla 4, muestra que todas las variables tienen un **GVIF** menor a 5, indicando que no hay colinealidad significativa en el modelo.

3.4. Conclusiones del laboratorio

A lo largo de este estudio, se observaron asociaciones entre la distribución de la edad gestacional entre las principales causas, con contraste marcado entre hipoxia intrauterina y diagnósticos no determinados. En los perfiles maternos, mayor edad y menor acumulado de controles prenatales se asociaron con mayor frecuencia relativa de defunciones en la base analizada. Por área de residencia, la concentración de registros en entornos urbanos describe un patrón de notificación, no necesariamente de riesgo.

4. Laboratorio: Enfoque Machine Learning

4.1. Objetivos específicos

- Evaluar modelos VAR multivariados para predecir mensualmente variables maternas asociadas a defunciones fetales, diferenciando por región geográfica en el periodo 2019-2023.
- Desarrollar un modelo supervisado con Random Forest para clasificar la causa probable de defunción fetal a partir de variables maternas, gestacionales y socioeconómicas, evaluando su precisión por categoría.
- Aplicar K-Means para identificar perfiles de riesgo mediante el agrupamiento no supervisado de variables numéricas relacionadas con la mortalidad fetal.

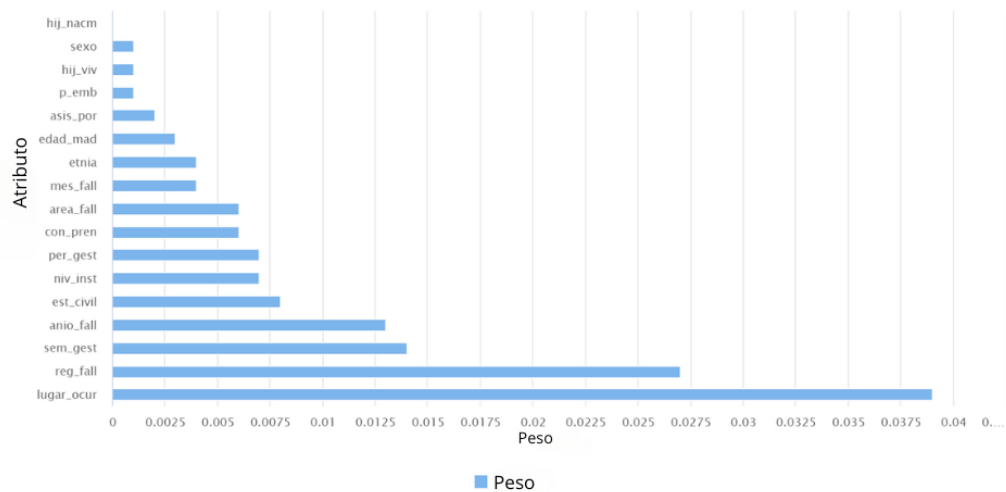
4.2. Método

4.2.1. Selección de los datos

La Figura 13 ilustra el peso relativo de cada variable seleccionada en el dataset con respecto a la variable Causa fetal. Las variables fueron elegidas debido a su relevancia predictiva, considerando su impacto en la clasificación de las defunciones fetales.

Figura 13

Peso de todas las variables del dataset con respecto a la variable Causa Fetal



Las variables seleccionadas para el modelado fueron:

- **Edad materna:** Relacionada con los riesgos de mortalidad fetal, influenciada por contextos sociales.
- **Nivel educativo:** Impacta el acceso y calidad de los controles prenatales, afectando los desenlaces fetales.
- **Control prenatal:** Refleja el nivel de atención médica recibido, influyendo en la mortalidad fetal.
- **Semanas de gestación:** Clasificada en pretérmino, término y post-término, esencial para el análisis de mortalidad fetal.
- **Lugar de ocurrencia:** Agrupado en establecimientos públicos y privados, ya que afecta la probabilidad de defunción fetal.
- **Región geográfica:** Segmentada en Costa, Sierra y Amazonía, debido a las diferencias en infraestructura de salud.

La imputación de valores faltantes en variables clave se realizó mediante técnicas de bootstrap, lo que garantizó la consistencia y fiabilidad de las series temporales necesarias para los modelos de predicción.

Como parte del proceso de preparación de los datos, para el modelo de vectores regresivos se crearon nuevas variables derivadas de características sociales y clínicas, las cuales permitieron una segmentación más precisa de los datos.

Estas incluyen:

- Proporción de partos en establecimientos públicos.
- Proporción de mujeres casadas.

- Proporción de partos a término.

4.2.2. Selección de técnicas

Las técnicas fueron seleccionadas por su capacidad para abordar los objetivos específicos del estudio, que incluyen predicción, clasificación y segmentación.

4.2.2.1. VAR (Vectores Autorregresivos)

Se utilizó VAR multivariado para modelar las relaciones temporales entre las variables maternas asociadas a las defunciones fetales, diferenciando los datos por región geográfica (Sierra, Costa y Amazonía) en el período 2019-2023.

Justificación: Ideal para analizar series temporales multivariadas, ya que permite incluir múltiples variables interrelacionadas, proporcionando una visión más completa de las dinámicas regionales.

4.2.2.2. Árbol de decisión

Se aplicó el algoritmo Random Forest, un modelo supervisado que utiliza múltiples árboles de decisión para clasificar las causas de defunción fetal en función de variables como edad materna, nivel educativo, control prenatal y semanas de gestación.

Justificación: Robusto y eficaz para manejar grandes volúmenes de datos, especialmente en clasificación de causas complejas, siendo crucial para problemas donde las relaciones entre variables son no lineales.

4.2.2.3. K-means

Se empleó K-means, una técnica de agrupamiento no supervisado, para identificar perfiles de riesgo relacionados con la mortalidad fetal. Segmentando los datos en clústeres basados en variables como la edad materna y las semanas de gestación.

Justificación: Explora patrones latentes en datos sin etiquetar y descubrir grupos de casos similares, facilitando la identificación de perfiles de riesgo asociados a la mortalidad fetal.

4.2.3. Descripción de cada técnica aplicada

4.2.3.1. VAR (Vectores Autorregresivos)

El modelo VAR multivariado fue aplicado para analizar las interacciones temporales entre las variables maternas asociadas a la mortalidad fetal, con un enfoque específico por región geográfica. La elección del lag óptimo fue determinada mediante el criterio AIC, y se dividió el conjunto de datos en entrenamiento (2019-2022) y prueba (2023).

- **Evaluación:** Se calcularon predicciones para el año 2023, comparando los valores reales con los valores predichos y evaluando el desempeño del modelo con métricas como RMSE.

4.2.3.2. Random Forest

El algoritmo Random Forest fue entrenado con las variables más relevantes para clasificar las causas de la defunción fetal. El modelo utilizó 19 árboles para reducir el sobreajuste y mejorar la generalización. Las métricas de evaluación incluyeron precisión, recall y matriz de confusión, para medir el desempeño del modelo en diferentes clases de causas de defunción.

4.2.3.3. K-means

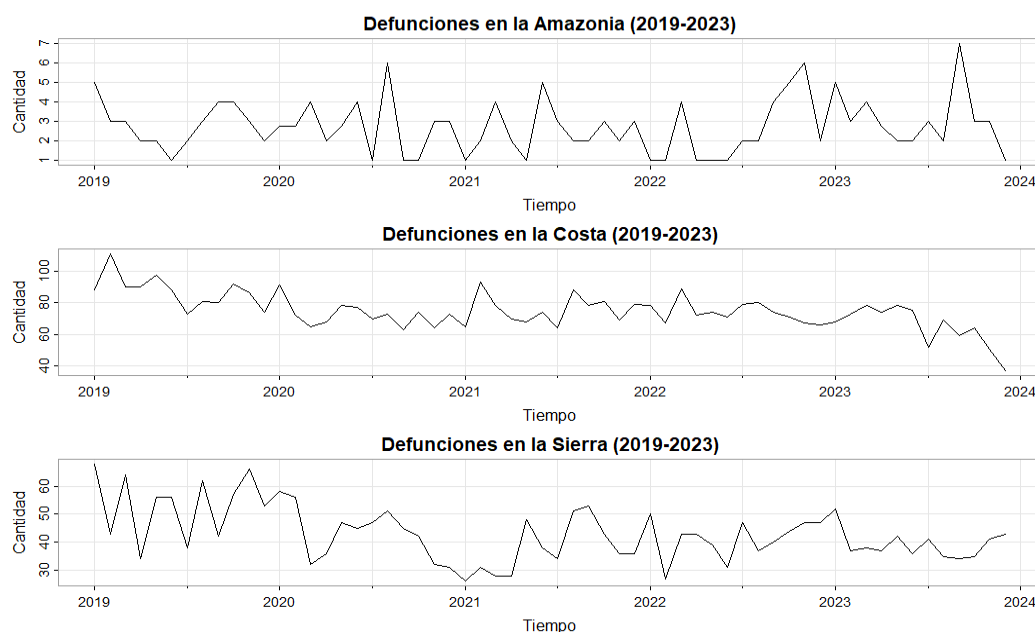
La técnica K-means fue utilizada para segmentar el conjunto de datos en 4 clústeres, basándose en variables como edad materna, semanas de gestación y número de hijos vivos. La segmentación fue evaluada mediante el índice Silhouette, identificando también perfiles diferenciados de riesgo.

4.3. Discusión y resultados

La Figura 14 muestra la evolución de las defunciones fetales en las regiones (2019-2023). En la Amazonía, hay alta variabilidad, mientras que, en la Costa, las defunciones son más estables. Por último, en la Sierra, aunque las defunciones son consistentes, se observan fluctuaciones. Teniendo en cuenta que los valores faltantes fueron imputados con bootstrap para mantener la consistencia en el análisis.

Figura 14

Serie de tiempo de la región Amazonía, Costa y Sierra del año 2019 al 2023



4.3.1. Modelo VAR

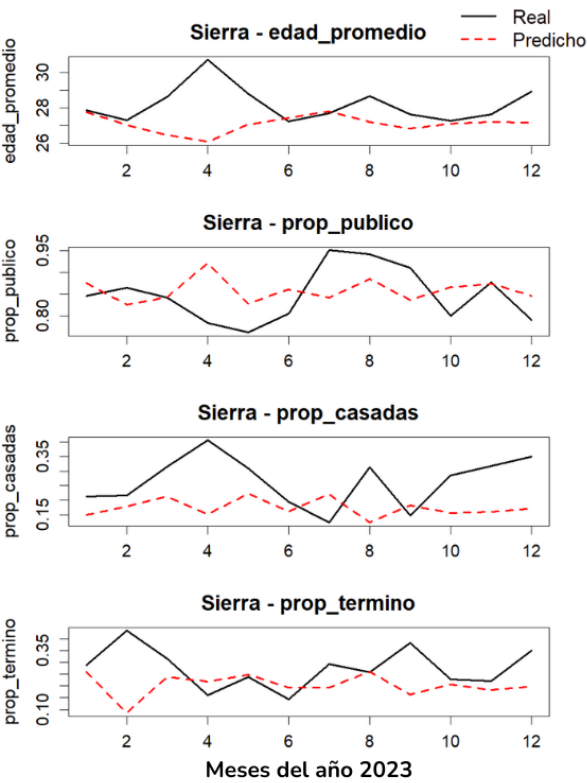
El modelo VAR fue utilizado para predecir variables maternas asociadas a las defunciones fetales en las regiones, utilizando como entrenamiento los años 2019 al 2022 y un lag de 6. Su desempeño fue heterogéneo según la región y la variable analizada. Mientras que para variables como la proporción de partos en establecimientos públicos, proporción de mujeres casadas y proporción de partos a término mostró ajustes razonables (RMSE bajos), en otras como la edad promedio materna presentó dificultades importantes, especialmente en la Sierra y Amazonía, donde los errores fueron elevados. Esto sugiere que el modelo es útil para series con mayor estabilidad temporal, pero menos fiable en variables con mayor fluctuación interanual.

4.3.1.1. VAR Sierra

La Figura 15 muestra las predicciones para 2023 en la Sierra, donde el modelo tiene desajustes en la edad promedio, lo que indica dificultades para capturar su variabilidad. En cambio, las predicciones para proporción de partos públicos, casadas y a término son más precisas, lo que sugiere que el modelo funciona mejor con estas variables, aunque la edad materna sigue siendo más difícil de predecir.

Figura 15

Predicciones para la región Sierra comparado con los valores reales



La Tabla 5 muestra los RMSE para las variables en la región Sierra. La edad promedio tiene el RMSE más alto (28,1), indicando mayor incertidumbre en su predicción. En cambio, las variables proporción pública (0,852), proporción casadas (0,287) y proporción término (0,305) tienen RMSE bajos, sugiriendo que el modelo predice estas variables con mayor precisión debido a su consistencia en la región.

Tabla 5

RMSE de las variables analizadas en la región Sierra

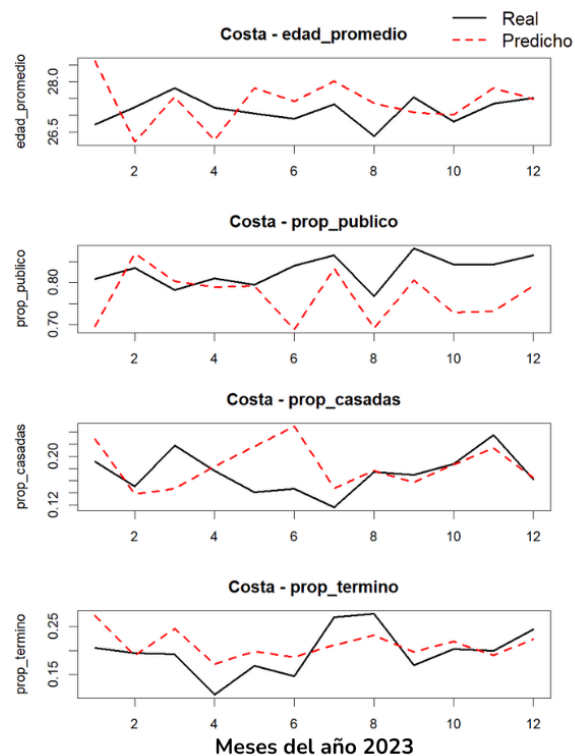
Variable	RMSE
Edad promedio	28,1
Proporción pública	0,852
Proporción casadas	0,287
Proporción término	0,305

4.3.1.2. VAR Costa

La Figura 16 muestra las predicciones y los valores reales de las variables maternas en la Costa para 2023. El modelo VAR tiene un buen desempeño, especialmente para proporción casadas y proporción término, con predicciones muy cercanas a los valores reales. Sin embargo, para edad promedio y proporción pública, aunque las predicciones son cercanas, se observan fluctuaciones en algunos meses, sugiriendo incertidumbre en estas variables.

Figura 16

Predicciones para la región Costa comparado con los valores reales



La Tabla 6 muestra que el modelo VAR tiene un buen desempeño en la región Costa, con RMSE bajos en las variables analizadas. La edad promedio presenta un RMSE de 0,839, indicando una ligera incertidumbre en su predicción. En cambio, las variables proporción pública (0,082), proporción casadas (0,045) y proporción término (0,041) tienen RMSE muy bajos, reflejando una alta precisión del modelo en predecir estas variables.

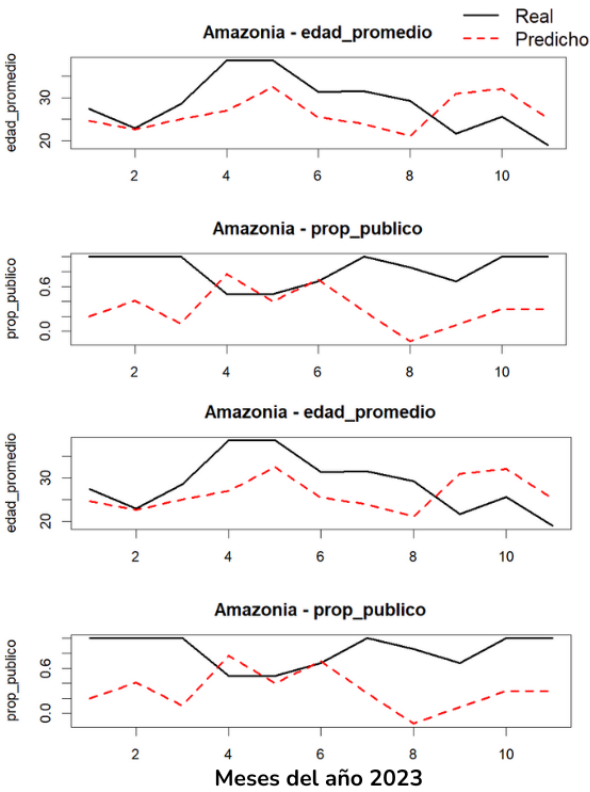
Tabla 6
RMSE de las variables analizadas en la región Costa

Variable	RMSE
Edad promedio	0,839
Proporción pública	0,082
Proporción casadas	0,045
Proporción término	0,041

4.3.1.3. VAR Amazonía

La Figura 17 muestra las predicciones y valores reales para las variables en la Amazonía en 2023. El modelo VAR presenta variabilidad en las predicciones, especialmente en la edad, donde se desvían de los valores reales. Sin embargo, las predicciones para proporción pública, proporción casadas y proporción término siguen más de cerca los valores reales, aunque con algunos desajustes.

Figura 17
Predicciones para la región Amazonía comparado con los valores reales



La Tabla 7 muestra los RMSE para las variables en la Amazonía entre 2019 y 2023. La edad promedio tiene el mayor RMSE (6,808), indicando incertidumbre en su predicción debido a la variabilidad en la edad materna en la región. En cambio, las variables proporción pública (0,656), proporción casadas (0,607) y proporción término (0,932) presentan RMSE más bajos, reflejando un mejor desempeño, aunque con algunas fluctuaciones.

Tabla 7

RMSE de las variables analizadas en la región Amazonía

Variable	RMSE
Edad promedio	6,808
Proporción pública	0,656
Proporción casadas	0,607
Proporción término	0,932

4.3.2. Random Forest

Random Forest es una técnica de aprendizaje supervisado que combina múltiples árboles de decisión para hacer predicciones más precisas. En este estudio, se utilizó para identificar patrones asociados a la mortalidad fetal en Ecuador, mejorando la toma de decisiones en salud materno-infantil. El modelo fue entrenado con un 80% de los datos para entrenamiento y 20% para prueba, usando 19 árboles, configurados con el criterio de information gain con diferente profundidad para su evaluación.

Tabla 8

Matriz de confusión para Random Forest con baja precisión (accuracy)

Accuracy: 74,68%

	Causas no especifica das Verdadera s	Factores Maternos Verdadera s	Restricción del Crecimiento Verdaderas	Malformacio nes Congénitas Verdaderas	Precisión por clase
pred. Causas no	3245	445	829	14	71,59%

especificadas					
pred.	25	735	33	0	92,69%
Factores Maternos					
pred.	322	120	1290	5	74,27%
Restricción del Crecimiento					
pred.	0	0	0	17	100,00%
Malformaciones Congénitas					
Recuperación por clase	90,34%	56,54%	59,94%	47,22%	

Tabla 9

Matriz de confusión para Random Forest

Accuracy: 86,40%

	Causas no especificadas Verdaderas	Factores Maternos Verdaderas	Restricción del Crecimiento Verdaderas	Malformaciones Congénitas Verdaderas	Precisión por clase
pred. Causas no especificadas	3308	188	359	7	85,66%
pred. Factores Maternos	31	1021	28	0	94,54%
pred. Restricción del Crecimiento	253	91	1765	6	83,45%

pred.	0	0	0	23	100,00
Malformaciones Congénitas					%
Recuperación por clase	92,09%	78,54	82,02%	63,89%	
		%			

La Tabla 9 muestra que el modelo Random Forest con 19 árboles alcanzó una precisión del 86,40%, destacando en la clasificación de malformaciones congénitas con una precisión del 100%. En comparación, el modelo con 10 árboles (Tabla 8) obtuvo una precisión de 74,68%, con un rendimiento inferior, especialmente en la clasificación de causas no especificadas y factores maternos, donde el recall fue más bajo. Esto sugiere que una mayor profundidad de los árboles mejora la capacidad de generalización y el desempeño en la clasificación de causas complejas.

La clase Malformaciones congénitas tiene bajo soporte (muy pocos casos), lo que reduce la sensibilidad y hace más inestable su estimación entre configuraciones. La exactitud global está influida por clases mayoritarias; por eso, la interpretación por clase debe considerar el desbalance.

4.3.3. Algoritmo K-means

Se utilizó K-Means para agrupar los casos de defunción fetal en 4 clústeres basados en variables maternas y gestacionales, identificando perfiles de riesgo asociados a la mortalidad fetal. Permitiendo detectar patrones complejos, lo que facilita el diseño de estrategias preventivas en salud materno-infantil.

El número de clústeres, fue seleccionado utilizando Silhouette, evaluando la calidad de la segmentación. Se determinó que 4 clústeres proporcionaban un índice promedio de 0,400, lo que indica una separación moderada entre los clústeres. Esta información se presenta en la Tabla 10.

Tabla 10

Índice de Silhouette por Clúster

Clúster	Silhouette Width
Clúster 0	0,397
Clúster 1	0,439
Clúster 2	0,399
Clúster 3	0,353

La Tabla 11 muestra la clasificación de los 7080 casos de defunción fetal en 4 clústeres. Indicando diferentes perfiles de riesgo basados en características maternas y gestacionales. Sugiriendo que algunos perfiles son más comunes que otros, reflejando patrones prevalentes en la mortalidad fetal, asociados a factores como edad materna y semanas de gestación.

Tabla 11

Clasificación de los casos en clústeres

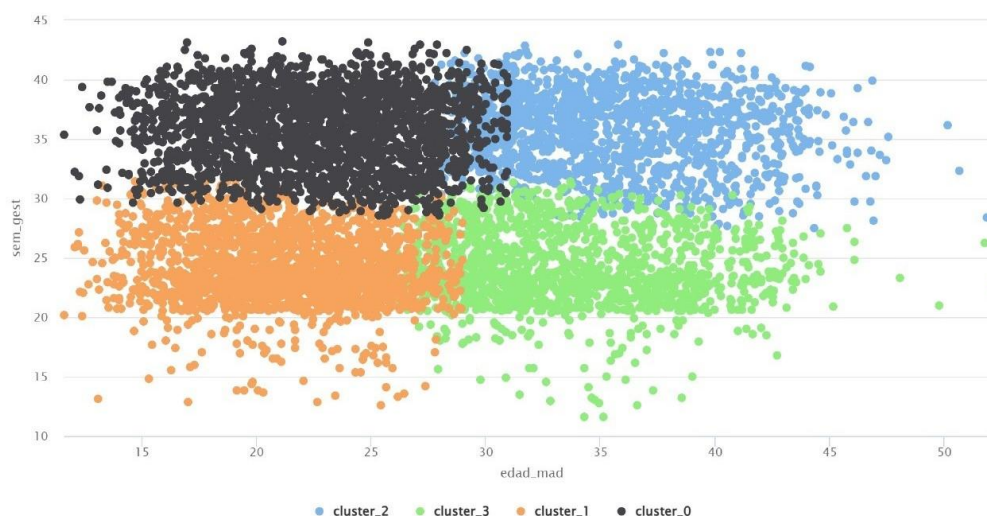
Clúster	Casos
Clúster 0	2,228 casos
Clúster 1	1,955 casos
Clúster 2	1,460 casos
Clúster 3	1,437 casos

La Tabla 12 muestra los centroides de los 4 clústeres identificados, cada uno con un perfil de riesgo distinto. Por su parte, la Figura 18 ilustra gráficamente esta segmentación, donde el Clúster 0 agrupa a madres jóvenes con gestaciones avanzadas, el Clúster 1 a madres muy jóvenes con gestaciones cortas, el Clúster 2 a madres mayores con gestaciones largas, y el Clúster 3 a madres de edad media con gestaciones cortas.

Tabla 12

Centroides de clústeres según características maternas y gestacionales

Clúster	Promedio (edad_mad)	Promedio (hijo_nacm)	Promedio (hij_viv)	Promedio (sem_gest)
Clúster 0	22,783	1,052	0,809	35,683
Clúster 1	21,503	1,117	0,698	24,213
Clúster 2	35,658	1,068	2,087	35,736
Clúster 3	33,476	1,175	1,704	24,135

Figura 18*Distribución de Clústeres*

4.3.4. Discusión

Los resultados de las tres técnicas (VAR, Random Forest y K-Means) evidencian un desempeño mixto entre técnicas y regiones. En el caso del VAR, hubo buen ajuste en algunas series (p. ej., proporción de casadas y partos a término en Costa) y ajuste limitado en otras (p. ej., edad materna promedio en Sierra/Amazonía); por ello, las conclusiones se formulan por variable y región. Esta heterogeneidad es coherente con la variabilidad de las series y con diferencias en captación/calidad del registro entre áreas.

La alta proporción de “no especificadas” (>50%) sugiere limitaciones en la codificación de causa fetal y variabilidad en la calidad del registro. Esto distorsiona la distribución aparente de causas e impide estimar con precisión el peso relativo de categorías específicas (p. ej., hipoxia, malformaciones).

4.4. Conclusiones del laboratorio

Como conclusión, este estudio muestra que VAR, Random Forest y K-Means aportan valor de manera complementaria, con fortalezas diferenciadas según variable y región. Los modelos caracterizaron factores como la edad materna extrema y las gestaciones pretérmino entre los perfiles más vinculados a las defunciones analizadas (sin inferir causalidad). El VAR fue útil para captar dinámicas temporales, especialmente en Costa, aunque con ajuste heterogéneo (alto en algunas series y moderado en otras, como la edad materna promedio en Sierra/Amazonía). Random Forest destacó en la clasificación (p. ej., malformaciones congénitas y factores maternos), con dificultades en las causas no especificadas. K-Means permitió segmentar los datos en perfiles de riesgo, con sensibilidad a la normalización de variables.

5. Laboratorio: Enfoque Toma de Decisiones

5.1. Objetivos específicos

- Clasificar la mortalidad fetal por perfil de riesgo usando K-NN.
- Explicar el valor de los perfiles de riesgo para personalizar la atención prenatal y reducir la mortalidad fetal.

5.2. Método

5.2.1. Selección de los datos

Para esta investigación, los clústeres que fueron generados en la Figura 18 del Laboratorio 2, fueron utilizados como base para estructurar la clasificación de riesgo de mortalidad fetal. Estos, fueron agrupados según las variables maternas y gestacionales, identificando perfiles de riesgo asociados a la mortalidad fetal.

5.2.2. Selección de técnicas

Después de que se clasificó en estos grupos, estos se juntaron en una nueva variable denominada riesgo, para dividir las observaciones en conjuntos de entrenamiento y prueba, para aplicar KNN.

Las técnicas fueron seleccionadas por su capacidad para abordar los objetivos específicos del estudio.

5.2.2.1. K-NN

Se utilizó el modelo de K-Nearest Neighbors (KNN) para predecir el riesgo de mortalidad fetal a partir de características como número de hijos vivos, número de hijos nacidos muertos y semanas de gestación. Si bien se utilizó la edad materna para clasificar previamente, y es una variable conocida en el riesgo de mortalidad fetal, en este caso se observó que la semana de gestación es un predictor más fuerte. Edad materna y semanas de gestación están altamente correlacionadas, por lo que, al considerarlas por separado, la edad materna es menos relevante que las semanas de gestación. Se tomó la decisión de excluir la variable de la edad de la madre del modelo de KNN.

Justificación: Ideal por su simplicidad, flexibilidad y capacidad de manejar datos multivariantes sin requerir una distribución de los datos. KNN es un modelo basado en la distancia, por lo que tiene la ventaja de clasificar las observaciones de mortalidad fetal de manera clara y directa.

5.2.3. Descripción de cada técnica aplicada

5.2.3.1. Descripción de K-Nearest Neighbors (KNN)

El conjunto de datos se dividió en 80% para entrenamiento y 20% para prueba. Al aplicar el modelo de KNN, se utilizó el parámetro $k = 5$, siendo este, el número de vecinos más cercanos considerados para determinar la clase de un nuevo punto de datos. Este valor fue escogido para equilibrar la precisión y simplicidad del modelo, evitando el sobreajuste.

Para cada observación en el conjunto de prueba, el modelo asigna una clase (riesgo "Alto", "Bajo", "Moderado") según la mayoría de las clases de los 5 vecinos más cercanos en el espacio de características.

Una vez que se entrenó el modelo, se evaluó el mismo con la matriz de confusión, permitiendo comparar las clases predichas con las clases reales del conjunto de prueba. Proporcionando métricas de desempeño como precisión y sensibilidad.

5.3. Discusión y resultados

5.3.1. Clústeres

Los clústeres se clasificaron de la siguiente manera:

1. Clasificación por edad materna, como se muestra en la Figura 19:
 - *Muy Alto*, para madres de más de 40 años.
 - *Alto*, para madres de 35 a 40 años.
 - *Bajo*, para madres de menos de 35 años.
2. Clasificación por semanas de gestación, como se muestra en la Figura 20:
 - *Muy Alto*, para embarazos con menos de 28 semanas de gestación.
 - *Alto*, para embarazos entre 28 y 30 semanas de gestación.
 - *Moderado*, para embarazos entre 30 y 37 semanas de gestación.
 - *Bajo*, para embarazos mayores de 37 semanas de gestación.

Figura 19
Clasificación por edad materna

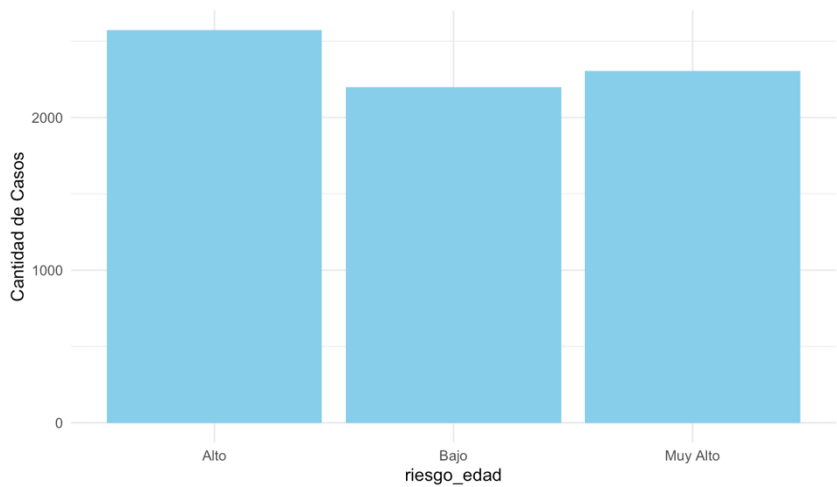
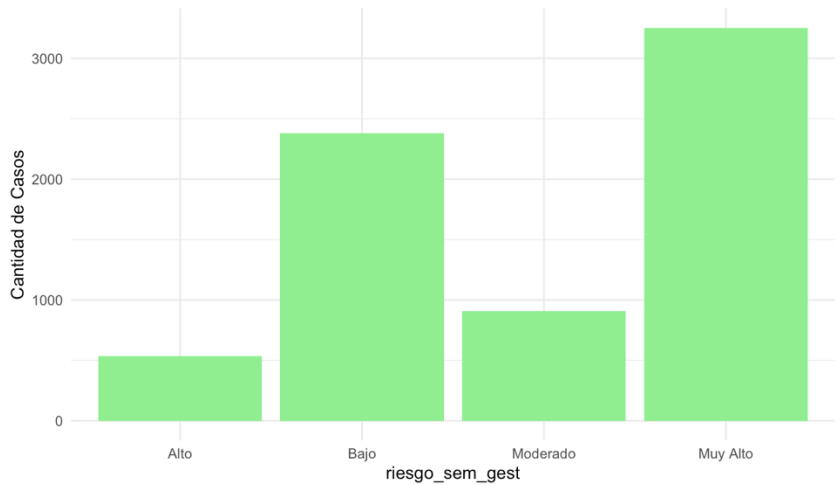


Figura 20
Clasificación por semana de gestación



5.3.2. K-Nearest Neighbors (KNN)

Los resultados obtenidos al aplicar el modelo de KNN, tanto el conjunto de entrenamiento como el de prueba mostraron un rendimiento destacado en la clasificación de la mayoría de las clases, mientras que se identificaron también áreas de mejora en otras. A continuación, se muestra la matriz de confusión para el conjunto de entrenamiento en la Tabla 13.

Tabla 13*Matriz de confusión para el conjunto de entrenamiento*

Clase	Sensibilidad	Especificidad	Valor	Valor	Precisión Global
			Predictivo	Predictivo	
			Positivo	Negativo	
Bajo	0,1000	1,0000	1,0000	0,9936	84,07%
Alto	0,9865	0,9569	0,8469	0,9963	99,35%
Muy Alto	0,9943	0,9312	0,8270	0,9980	99,43%
Moderado	0,0000	1,0000	NA	1,0000	100,00%

La precisión global del conjunto de entrenamiento fue de 84,07%, indicando que el modelo mostró un buen rendimiento general en la clasificación de los datos de entrenamiento según sus clases. El valor Kappa fue de 0,7861, mostrando una buena concordancia entre las clases predichas y las reales.

De la misma forma, los resultados del conjunto de prueba, es decir el 20% del conjunto de datos, muestran cómo el modelo generaliza a nuevos datos. A continuación, se presenta la matriz de confusión en la Tabla 14.

Tabla 14*Matriz de confusión para el conjunto de prueba*

Clase	Sensibilidad	Especificidad	Valor	Valor	Precisión Global
			Predictivo	Predictivo	
			Positivo	Negativo	
Bajo	0,0000	1,0000	NA	0,9926	83,12%
Alto	0,9968	0,9552	0,8399	0,9927	99,25%
Muy Alto	0,9886	0,9332	0,8302	0,9960	96,27%
Moderado	0,0000	1,0000	NA	0,9982	100,00%

La precisión global del conjunto de prueba fue de 83,12%, mostrando una ligera disminución en comparación con el modelo de entrenamiento, pero se mantiene con un valor alto, indicando que el modelo generaliza bien. Asimismo, el modelo Kappa fue de 0,7746, lo que también indica buena concordancia entre las predicciones y las clases reales.

Las clases se desempeñaron de la siguiente manera:

- **Bajo:** La sensibilidad para esta clase es de 0,0000 en el conjunto de prueba, indicando que el modelo no logró identificar correctamente los casos de riesgo.

- **Alto:** La sensibilidad para la clase alto fue de 0,9968 en el conjunto de prueba, indicando una predicción muy precisa para esta clase.
- **Muy Alto:** La sensibilidad para esta clase fue de 0,9886, indicando una buena capacidad para identificar correctamente esta clase.
- **Moderado:** La sensibilidad para esta clase fue de 0,0000, indicando que el modelo no identificó correctamente estos casos en el conjunto de prueba.

En KNN, las clases Bajo y Moderado muestran sensibilidad = 0 en prueba debido a su muy baja prevalencia, por lo que el algoritmo favorece las clases mayoritarias en la vecindad. Este desbalance explica la caída de sensibilidad en minorías, a pesar de una precisión global alta.

La importancia de contar con modelos predictivos capaces de detectar perfiles de riesgo minoritarios cobra aún más relevancia en contextos críticos como la pandemia por COVID-19. En un estudio multicéntrico, Gholami et al. (2023) hallaron aumentos significativos en complicaciones obstétricas como preeclampsia, diabetes gestacional, partos prematuros y cesáreas, incluso entre mujeres previamente sanas. Estos hallazgos refuerzan la utilidad de modelos como KNN para anticipar riesgos materno-fetales y guiar decisiones clínicas bajo escenarios de presión sanitaria.ese a una precisión global alta.

5.3.3. Discusión

El modelo de KNN mostró una precisión global de 84,07% en el conjunto de entrenamiento y de 83,12 en el conjunto de prueba, con un buen desempeño en las clases “Alto” y “Muy Alto”. El modelo tiene una alta especificidad en todas las clases, lo que revela que cuando predice una clase, lo hace con certeza. Sin embargo, en las clases de “Bajo” y “Moderado” sugiere que el modelo podría beneficiarse de ajustes en el muestreo, dado que presenta dificultades con sensibilidad nula en ambas clases. Esto se debe principalmente a la baja prevalencia de estas clases en los datos. Resaltando la necesidad de mejorar la calidad de los registros clínicos para obtener predicciones más precisas y optimizar las estrategias preventivas.

La clasificación del riesgo de mortalidad fetal en perfiles como “Alto”, “Bajo” y “Moderado” a partir de características maternas como edad y semanas de gestación tienen un valor significativo al momento de personalizar la atención prenatal. Una vez que se identificaron a las mujeres embarazadas en riesgo alto, tales como con menos de 28 semanas de gestación y mayor a 40 años, se pueden implementar estrategias de prevención o intervención temprana que ayuden con las

complicaciones graves. De la misma manera las otras clases pueden recibir apoyo y seguimiento ajustándose a sus necesidades específicas.

En un estudio regional sobre mortalidad materna durante la pandemia de COVID-19 en Latinoamérica, se identificó un patrón preocupante: las gestantes más vulnerables (jóvenes, sin educación formal, bajo atención pública) presentaron las mayores tasas de mortalidad. Este hallazgo pone de relieve la influencia de factores sociales estructurales y refuerza la necesidad de priorizar recursos en perfiles de alto riesgo, como los que este modelo identifica Sotomayor Valdiviezo et al. (2023).

En un estudio nacional reciente, Araujo Zarate et al. (2024) aplicaron modelos de machine learning a más de 560.000 registros de nacimientos y defunciones fetales en Colombia. Sus resultados confirmaron que variables como la edad materna, el nivel educativo, el régimen subsidiado, la zona de residencia y el peso al nacer fueron predictoras clave de mortinato, siendo el modelo KNN el más preciso.

Estos hallazgos concuerdan con lo reportado por Huang et al. (2008), quienes encontraron que la edad materna avanzada (≥ 35 años) se asocia con un mayor riesgo de mortinato, con razones de riesgo que van de 1.2 a 4.5 en estudios de cohortes y casos y controles, incluso después de ajustar por factores confusores.

Los perfiles de riesgo permiten priorizar recursos y a su vez, atención médica, lo que puede llevar a una reducción de la mortalidad fetal. Dado que, al enfocar los esfuerzos en las mujeres embarazadas con un mayor riesgo de defunción fetal, se optimizan los recursos disponibles y se optimizan las probabilidades de tener resultados positivos en los embarazos más complicados.

Con base en los resultados de la clasificación de riesgo, una decisión clave sería dar prioridad a las intervenciones y recursos de salud para mujeres embarazadas que hayan sido clasificadas como riesgo muy alto o alto, incluyendo:

- Aumentar el monitoreo médico, como ultrasonidos con más frecuencia, control de presión arterial y pruebas de riesgo de parto prematuro.
- Educación dirigida a madres, explicando riesgos asociados con cada perfil y dando importancia a la detección temprana.
- Dar asesoramiento y acompañamiento psicológico, especialmente a mujeres con embarazos de alto riesgo, mejorando la adherencia al tratamiento.

Con base en los resultados de la clasificación de riesgo, una decisión clave sería dar prioridad a las intervenciones y recursos de salud para mujeres embarazadas que hayan sido clasificadas como riesgo muy alto o alto, incluyendo:

Estas recomendaciones se alinean con las directrices de la Organización Mundial de la Salud para una experiencia de parto positiva, las cuales subrayan la importancia del acompañamiento emocional, el trato respetuoso, la toma de decisiones informadas y la atención clínica oportuna, especialmente para mujeres con embarazos de alto riesgo. Una atención perinatal integral centrada en la persona es clave para reducir complicaciones evitables y mejorar los desenlaces del embarazo (Organización Mundial de la Salud, 2018).

5.4. Conclusiones del laboratorio

El modelo de KNN, mostró una precisión general de 84,07% en el conjunto de entrenamiento y un 83,12% en el conjunto de prueba. Clasificando bien las clases “Alto” y “Muy Alto”. Sin embargo, presenta dificultades en “Bajo” y “Moderado” debido a su baja prevalencia. Clasificar adecuadamente los perfiles de riesgo es esencial para personalizar la atención médica, priorizando detecciones tempranas en casos de alto riesgo, optimizando recursos que ayuden a la reducción de la mortalidad fetal.

6. Laboratorio: Enfoque Robustez y Validación

6.1. Objetivos

- Validar los supuestos estadísticos de los modelos utilizados en los análisis previos.

6.2. Método

6.2.1. Validación de modelos

Para la validación, se consideraron los modelos construidos en laboratorios anteriores:

- Regresión lineal múltiple
- Modelo de Autorregresión Vectorial (VAR)
- Árbol de decisión
- K-means
- K-Nearest Neighbors

En cada modelo, se evaluó los siguientes supuestos teniendo en cuenta sus restricciones: normalidad, homocedasticidad, ortogonalidad, independencia, aleatoriedad y determinación. Asimismo, se contrastó la robustez de los mismos, mediante comparaciones de errores.

Para el caso del modelo del árbol de decisión, se realizó a partir de la matriz de confusión presentada en el segundo laboratorio, se evaluaron los siguientes supuestos: MAE, MSE, RMSE, MAPE y correlación.

6.3. Discusión y resultado

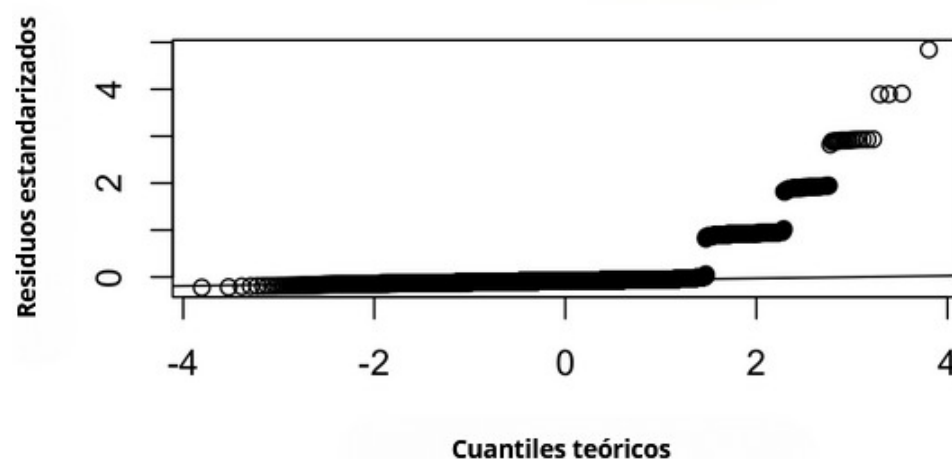
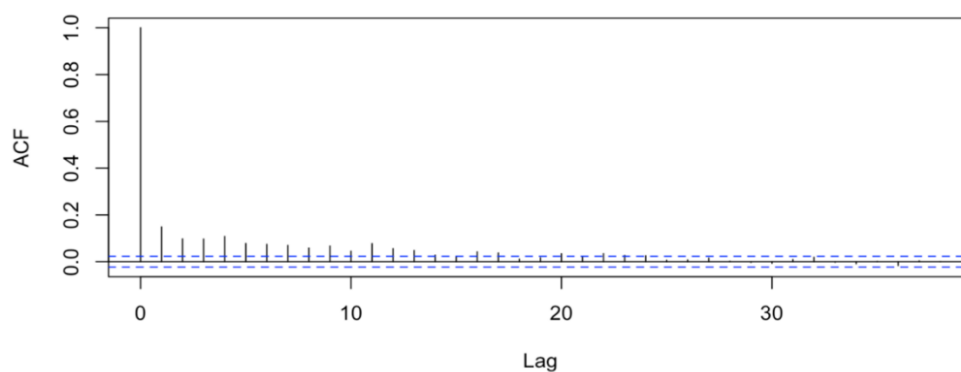
6.3.1. Regresión lineal múltiple

En la Tabla 15 se observa que el modelo de regresión lineal múltiple incumple la normalidad, aunque mantiene homocedasticidad. El estadístico Durbin-Watson cercano a 2 indica replicabilidad y autocorrelación positiva. La Figura 21 confirma la falta de normalidad y la Figura 22 muestra lags significativos. Según la Tabla 16, edad materna, control prenatal y área resultaron variables significativas.

Tabla 15

Evaluación de supuestos para el modelo de regresión lineal múltiple

Supuesto	Prueba utilizada	Estadístico / Valor	Valor p	Interpretación
Normalidad	Kolmogorov-Smirnov	D = 0,44129	2,2e-16	Se rechaza la normalidad de los residuos
Homocedasticidad	Levene's	F (99, 6948) = 0,8537	0,8486	No se rechaza la homogeneidad de las varianzas
Ortogonalidad	Durbin Watson	- DW = 1,7017	2,2e-16	Es replicable y evidencia de autocorrelación positiva en los residuos

Figura 21*Gráfica de QQ del modelo de regresión múltiple***Figura 22***ACF de los residuos del modelo de regresión lineal múltiple***Tabla 16***ANOVA para el modelo de regresión lineal múltiple*

Variable	gl (Df)	Suma de cuadrados (Sum Sq)	Media cuadrática (Mean Sq)	F value	Pr (>F)	Signif.
Edad materna	1	2,8	2,8492	24,654	7,02e-07	***
Control prenatal	1	0,6	0,6457	5,587	0,0181	*
Etnia	8	1,2	0,1493	1,292	0,2426	n.s.
Área	1	0,6	0,5872	5,081	0,0242	*
Nivel instructivo	9	0,4	0,0415	0,359	0,9542	n.s.
Residuales	7027	812,1	0,1156	-	-	-

6.3.2. Modelo de Autorregresión Vectorial (VAR)

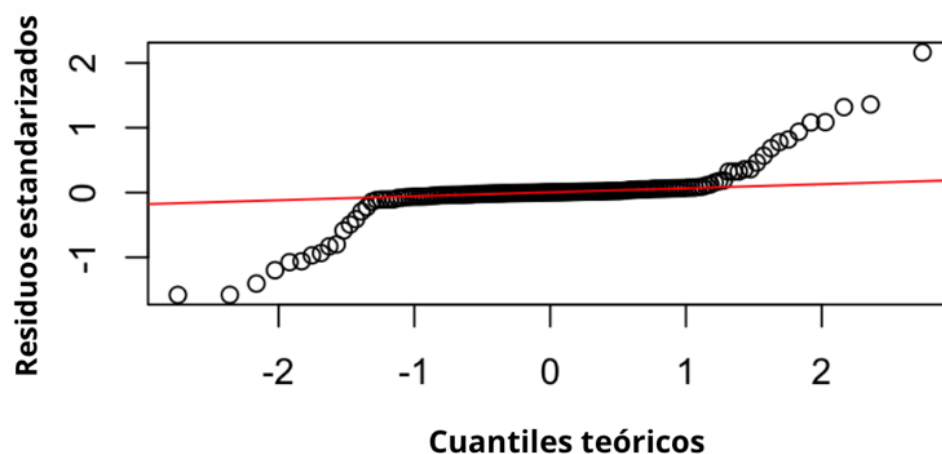
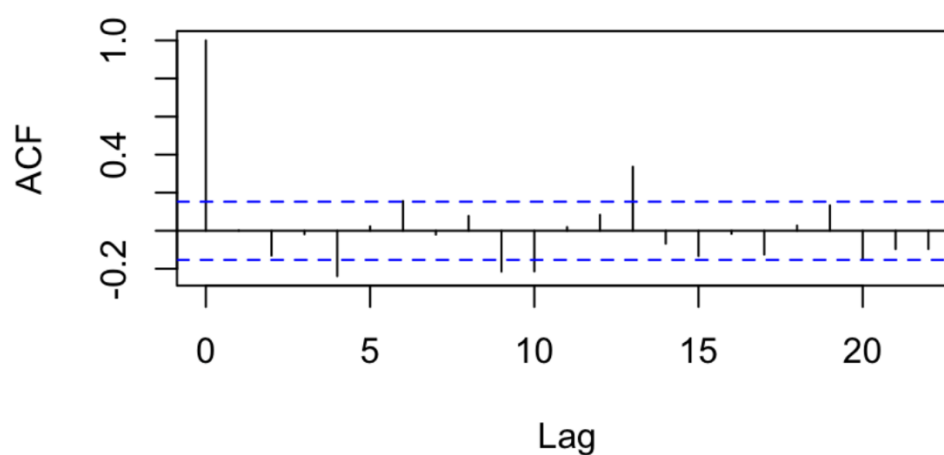
6.3.2.1. VAR Sierra

En la Tabla 17 se muestra que el modelo VAR Sierra no cumple la normalidad, aunque mantiene homocedasticidad, independencia y aleatoriedad. Durbin-Watson cercano a 2 indica replicabilidad. La Figura 23 confirma la falta de normalidad y la Figura 24 muestra autocorrelación. Como se muestra en la Tabla 5 del Laboratorio: Enfoque Machine Learning, el modelo predice con mayor precisión la proporción de mujeres casadas y de partos a término.

Tabla 17

Evaluación de supuestos para el modelo VAR Sierra

Supuesto	Prueba utilizada	Estadístico / Valor	Valor p	Interpretación
Normalidad	Shapiro – Wilk	W = 0,6763	2,2e-16	Se rechaza la normalidad de los residuos
Homocedasticidad	Levene's	F(47,116) = 0,5092	0,995	No se rechaza la homogeneidad de las varianzas
Ortogonalidad	Durbin - Watson	DW = 1,8348	0,3053	Es replicable
Independencia	Fisher's Exact	-	0,9849	Se cumple independencia
Aleatoriedad	Box-Ljung	X ² = 0,00032 df = 1	0,9856	No se rechaza aleatoriedad

Figura 23*Gráfica de QQ del modelo de VAR Sierra***Figura 24***ACF de los residuos del modelo de VAR Sierra***6.3.2.2. VAR Costa**

En la Tabla 18 se observa que el modelo VAR Costa no cumple con la normalidad, mientras conserva homocedasticidad, independencia y aleatoriedad, con un Durbin-Watson cercano a 2. La Figura 25 confirma la desviación de la normalidad y la Figura 26 muestra menor autocorrelación que el modelo anterior. Como se muestra en la Tabla 6 del Laboratorio: Enfoque Machine Learning, el modelo predice con alta precisión las proporciones de partos en establecimientos públicos, mujeres casadas y partos a término.

Tabla 18*Evaluación de supuestos para el modelo VAR Costa*

Supuesto	Prueba utilizada	Estadístico / Valor	Valor p	Interpretación
Normalidad	Shapiro - Wilk	W = 0,6833	2,2e-16	Se rechaza la normalidad de los residuos
Homocedasticidad	Levene's	F(47,116) = 0,5712	0,9843	No se rechaza la homogeneidad de las varianzas
Ortogonalidad	Durbin - Watson	DW = 1,8819	0,3324	Es replicable
Independencia	Fisher's Exact	-	0,5116	Se cumple independencia
Aleatoriedad	Box-Ljung	X ² = 0,5263 df = 1	0,4682	No se rechaza aleatoriedad

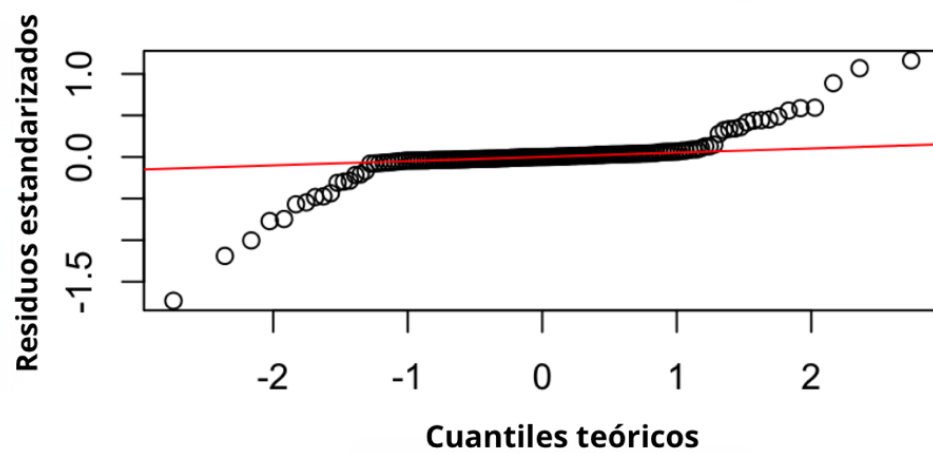
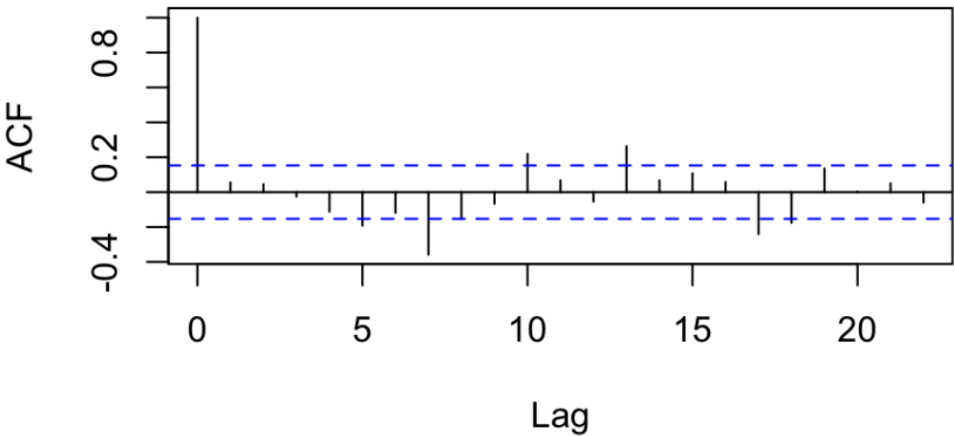
Figura 25*Gráfica de QQ del modelo de VAR Costa*

Figura 26
ACF de los residuos del modelo de VAR Costa



6.3.2.3. VAR Amazonía

En la Tabla 19 se indica que el modelo VAR Amazonía no cumple la normalidad, pero mantiene homocedasticidad, independencia y aleatoriedad. El Durbin-Watson cercano a 2 refleja replicabilidad. La Figura 27 confirma la falta de normalidad y la Figura 28 muestra menor autocorrelación. Como se muestra en la Tabla 7 del Laboratorio: Enfoque Machine Learning, muestra como las predicciones de este modelo son menos precisas que en otras regiones, especialmente para la edad promedio de las madres.

Tabla 19
Evaluación de supuestos para el modelo VAR Amazonía

Supuesto	Prueba utilizada	Estadístico / Valor	Valor p	Interpretación
Normalidad	Shapiro - Wilk	W = 0,7259	1,626e-15	Se rechaza la normalidad de los residuos
Homocedasticidad	Levene's	F(44,107) = 0,3572	0,9999	No se rechaza la homogeneidad de las varianzas
Ortogonalidad	Durbin - Watson	DW = 2,0566	0,4810	Es replicable, aunque se evidencian ligeras fisuras experimentales

Independencia	Fisher's Exact	-	0,3505	Se cumple independencia
Aleatoriedad	Box-Ljung	$X^2 = 0,3931$ df = 1	0,5307	No se rechaza aleatoriedad

Figura 27
Gráfica de QQ del modelo de VAR Amazonía

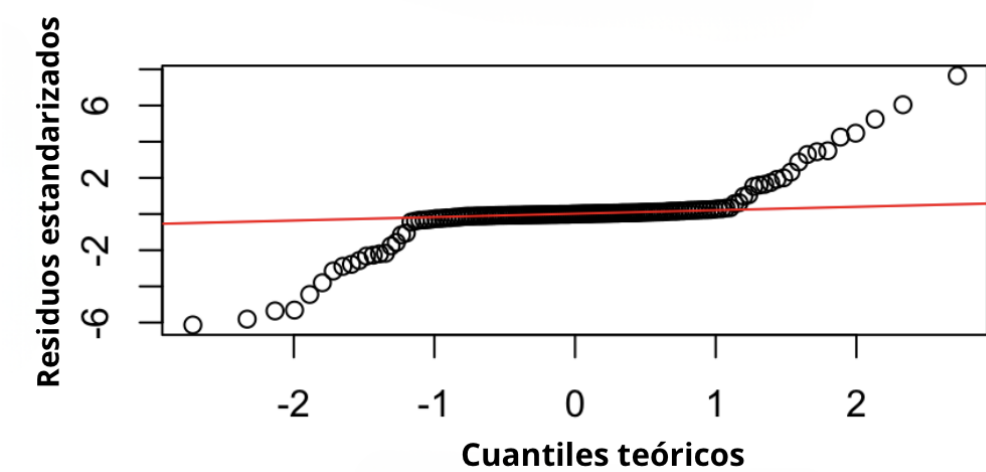
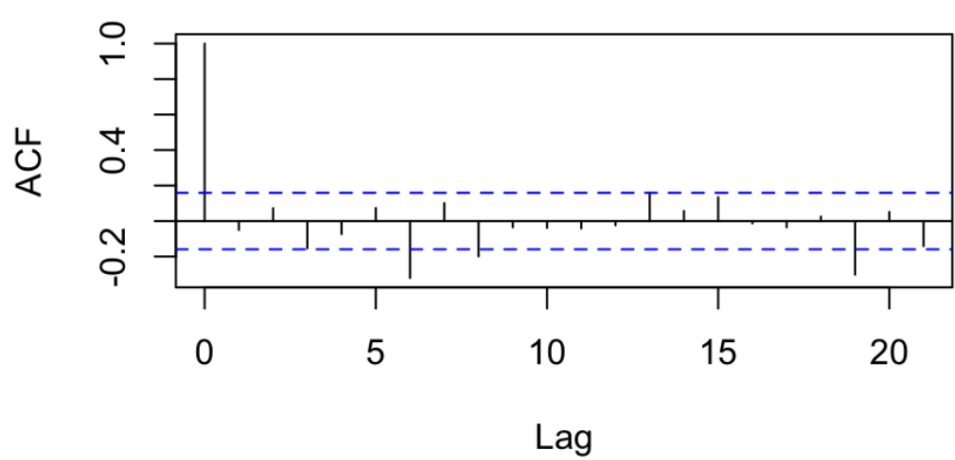


Figura 28
ACF de los residuos del modelo de VAR Amazonía



6.3.3. Árbol de decisión

En la Tabla 20 se muestran los errores del modelo de árbol de decisión, donde las métricas indican un desempeño aceptable y una correlación moderada (0,742) entre valores observados y predichos.

Tabla 20*Análisis de errores del modelo de árbol de decisión*

Métrica	Valor
MAE	0,329 +/- 0,183
MSE	0,142 +/- 0,144
RMSE	0,376 +/- 0,000
MAPE	0,3188 +/- 0,1831
Correlación	0,742

6.3.4. K-means

En la Tabla 21 se observa que el modelo K-means incumple normalidad, independencia y aleatoriedad, aunque mantiene homocedasticidad y un Durbin-Watson replicable. La Figura 29 confirma la falta de normalidad y la Figura 30 evidencia autocorrelación en varios lags. Según la Tabla 22, el factor cluster fue significativo ($p = 0,0035$), validando diferencias entre grupos.

Tabla 21*Evaluación de supuestos para el modelo K-means*

Supuesto	Prueba utilizada	Estadístico / Valor	Valor p	Interpretación
Normalidad	Kolmogorov-Smirnov	D = 0,1386	2,2e-16	Se rechaza la normalidad de los residuos
Homocedasticidad	Levene's	F(3,7076) = 0,4297	0,7317	No se rechaza la homogeneidad de las varianzas
Ortogonalidad	Durbin Watson	- DW = 1,6125	2,2e-16	Es replicable
Independencia	Chi-squared	X ² = 3781,5 df = 3	2,2e-16	No se cumple independencia
Aleatoriedad	Box-Ljung	X ² = 270,34 df = 1	2,2e-16	No se cumple aleatoriedad

Figura 29
Gráfica de QQ del modelo K-means

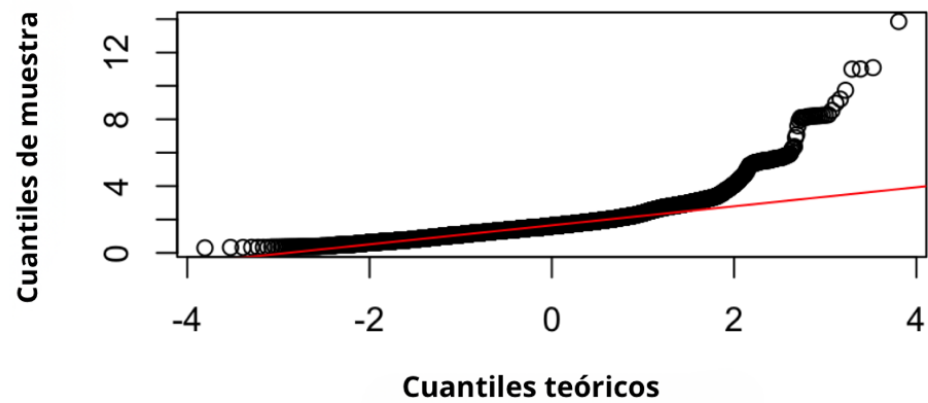


Figura 30
ACF de los residuos del modelo K-means

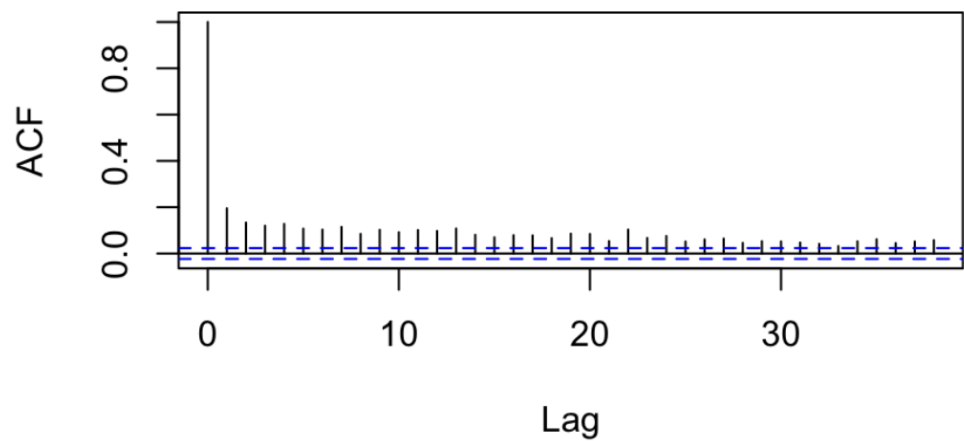


Tabla 22
ANOVA para el modelo K-means

Variable	gl (Df)	Suma de cuadrados (Sum Sq)	Media cuadrática (Mean Sq)	F value	Pr (>F)	Signif.
Clúster	3	11	3,752	4,534	0,0035	**
Residuales	7076	5855	0,827	-	-	-

6.3.5. K-Nearest Neighbors (KNN)

En la Tabla 23 se observa que el modelo KNN incumple normalidad, independencia y aleatoriedad, aunque mantiene homocedasticidad y un Durbin-Watson replicable. La Figura 31 confirma la falta de normalidad, mientras que la Figura 32 muestra pocos lags

significativos, sugiriendo menor autocorrelación. Además, la Tabla 24 evidencia que el factor riesgo fue altamente significativo ($p < 0,001$).

Tabla 23

Evaluación de supuestos para el modelo KNN

Supuesto	Prueba utilizada	Estadístico / Valor	Valor p	Interpretación
Normalidad	Kolmogorov-Smirnov	D = 0,1386	2,2e-16	Se rechaza la normalidad de los residuos
Homocedasticidad	Levene's	F(3,7076) = 0,4297	0,7317	No se rechaza la homogeneidad de las varianzas
Ortogonalidad	Durbin - Watson	DW = 1,6125	2,2e-16	Es replicable
Independencia	Chi-squared	X ² = 3781,5 df = 3	2,2e-16	No se cumple independencia
Aleatoriedad	Box-Ljung	X ² = 270,34 df = 1	2,2e-16	No se cumple aleatoriedad

Figura 31

Gráfica de QQ del modelo KNN

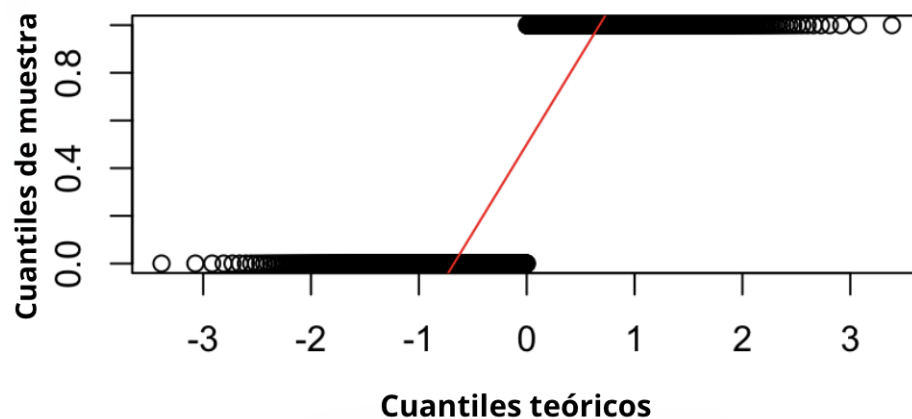


Figura 32
ACF de los residuos del modelo KNN

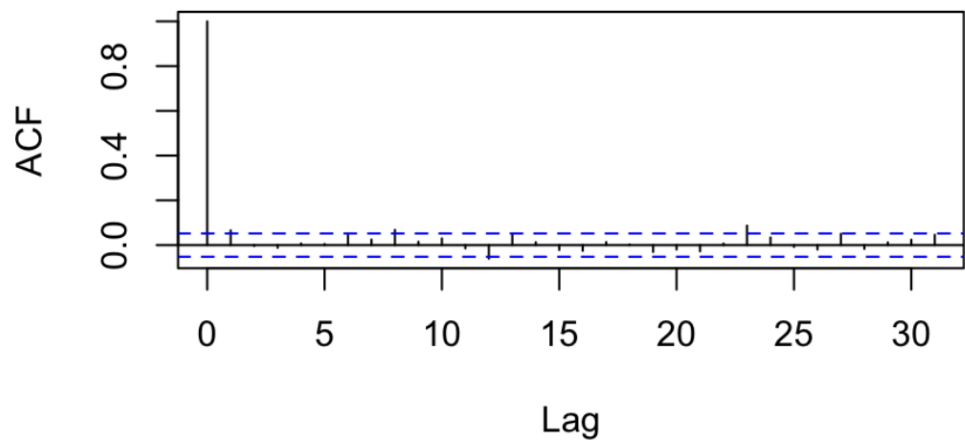


Tabla 24
ANOVA para el modelo de KNN

Variable	gl (Df)	Suma de cuadrados (Sum Sq)	Media cuadrática (Mean Sq)	F value	Pr (>F)	Signif.
Riesgo	11	19,2	1,7417	7,304	3,04e-12	***
Residuales	1400	333,8	0,2385	-	-	-

6.3.6. **Discusión**

Los resultados evidencian que la normalidad fue el supuesto más incumplido en todos los modelos. El VAR destacó al mantener homocedasticidad, independencia y aleatoriedad en las tres regiones, con valores de Durbin-Watson cercanos a 2 y errores RMSE bajos y consistentes, lo que respalda su solidez. En regresión lineal múltiple, el ANOVA mostró que edad materna, control prenatal y área fueron significativas. En K-means y KNN, aunque se incumplieron independencia y aleatoriedad, los factores clúster y riesgo resultaron significativos, aportando valor interpretativo. El árbol de decisión mostró desempeño aceptable, pero con menor rigor estadístico.

6.4. **Conclusiones del laboratorio**

Tras la validación de supuestos, se observa que los resultados de los laboratorios previos presentan limitaciones en su validez, ya que el incumplimiento frecuente de la normalidad y, en algunos casos, de independencia y aleatoriedad (K-means y KNN) limita la confianza en la generalización de los hallazgos. No obstante, modelos como el VAR mostraron consistencia en homocedasticidad, independencia y aleatoriedad, y errores

bajos en varias series, con incertidumbre mayor en otras, lo que respalda su utilidad como insumo donde muestra mejor ajuste y sugiere cautela en las restantes. En cambio, los resultados de regresión lineal múltiple, Random Forest y KNN aportan información relevante, pero deben interpretarse con cautela y complementarse con métricas comparables, ajustes por región/variable y, de ser posible, validación externa.

6.5. Limitaciones del estudio

Este estudio utiliza registros administrativos con posible subregistro y sesgos de notificación, especialmente en zonas rurales y en la categoría de causa “no especificada” (>50%). Estos factores pueden distorsionar distribuciones aparentes, subestimar causas específicas y sobrevalorar patrones urbanos. La calidad y completitud del dato varía entre regiones y años. Además, la base carece de variables clínicas clave (p. ej., paridad, comorbilidades, IMC, tabaquismo, inicio y adecuación del control prenatal), lo que limita el ajuste por confusión y favorece asociaciones residuales. En los modelos de Machine Learning se observaron clases desbalanceadas, afectando la sensibilidad de clases poco frecuentes y la estabilidad de métricas por clase. Por tanto, las inferencias deben interpretarse con cautela, priorizando tasas y proporciones ajustadas y resultados por variable/región.

7. Conclusión general

El presente estudio analizó la mortalidad fetal en Ecuador desde un enfoque estadístico e interdisciplinario, combinando métodos descriptivos, inferenciales y de aprendizaje automático aplicados a registros del INEC (2019–2023). La mortalidad fetal persiste como un problema de salud pública relevante, relacionado con factores clínicos, sociales y territoriales. La edad materna extrema, el número insuficiente de controles prenatales y las desigualdades en el acceso a la atención obstétrica se identificaron como variables asociadas a mayor frecuencia de defunciones.

A nivel mundial, UNICEF estima que cada día mueren 7.200 bebés antes de nacer, lo que equivale a cerca de 2 millones de mortinatos al año. El 84% de estos casos ocurren en países de ingresos bajos y medianos, y se consideran en su mayoría prevenibles con atención de calidad durante el embarazo y el parto (UNICEF, 2025). Esta carga global pone en perspectiva la relevancia de estudios nacionales como el presente, orientados a identificar patrones locales y diseñar intervenciones más efectivas.

Un hallazgo transversal es que más del 50% de los casos se codificaron como “no especificados”, lo que limita la inferencia por causa y puede distorsionar la distribución aparente de etiologías. Este patrón apunta a mejoras necesarias en la codificación y

calidad del registro (INEC/MSP) para fortalecer la interpretación etiológica y la comparabilidad interregional.

A nivel metodológico, los modelos implementados mostraron fortalezas diferenciadas. El VAR capturó dinámicas temporales regionales; su precisión fue mayor en Costa y variable en Sierra/Amazonía, con series de buen ajuste y otras con error más elevado (p. ej., edad materna promedio). Random Forest logró una clasificación eficiente de causas (especialmente malformaciones congénitas), mientras que K-means permitió segmentar perfiles de riesgo según características maternas y gestacionales. KNN alcanzó precisión >80% al clasificar niveles de riesgo “alto” y “muy alto”, apoyando su utilidad operativa.

Las principales limitaciones se vinculan con heterogeneidad y subregistro (especialmente en Amazonía), falta de estandarización de variables contextuales y la alta proporción de “no especificadas”, lo que reduce potencia explicativa y estabilidad inter-región. Se recomienda fortalecer sistemas de información y protocolos de registro/codificación para mejorar la generalización y la validez por causa.

En conjunto, este trabajo muestra que la aplicación de herramientas estadísticas y computacionales puede contribuir a mejorar la vigilancia epidemiológica y orientar políticas públicas basadas en evidencia. Se recomienda fortalecer los sistemas de información en salud, promover el acceso equitativo a controles prenatales y fomentar el uso de modelos predictivos como apoyo a la gestión sanitaria y a la prevención de muertes fetales evitables en el país.

8. Bibliografía

- Araujo Zarate, P., Martínez Lobo, D., & Contreras Chávez, J. (2024). Modelos predictivos en la clasificación de nacidos vivos y mortinatos: Un estudio comparativo entre técnicas de machine learning y regresión logística en función de variables sociodemográficas y clínicas. *Revista Ciencias Biomédicas*, 13(4), 175-189. <https://doi.org/10.32997/rcb-2024-4940>
- Cersonsky, T. E. K. (2023). *Identifying risk of stillbirth using machine learning*.
- Gholami, R., Borumandnia, N., Kalhori, E., Taheri, M., & Khodakarami, N. (2023). The impact of covid-19 pandemic on pregnancy outcome. *BMC Pregnancy and Childbirth*, 23(1), 811. <https://doi.org/10.1186/s12884-023-06098-z>
- Gwako, G. N., Were, F., Obimbo, M. M., Kinuthia, J., Gachuno, O. W., & Gichangi, P. B. (2021). Association between utilization and quality of antenatal care with stillbirths in four tertiary hospitals in a low-income urban setting. *Acta Obstetrica et Gynecologica Scandinavica*, 100(4), 676-683. <https://doi.org/10.1111/aogs.13956>

- Huang, L., Sauve, R., Birkett, N., Fergusson, D., & Van Walraven, C. (2008). Maternal age and risk of stillbirth: A systematic review. *Canadian Medical Association Journal*, 178(2), 165-172. <https://doi.org/10.1503/cmaj.070150>
- INEC. (2023). Banco de Datos Abiertos. <https://aplicaciones3.ecuadorencifras.gob.ec/BIINEC-war/index.xhtml?jsessionid=+R5SLr5J3qdLgv4tcVBEqEH.undefined>
- Mucio, B., Sosa, C., Colomar, M., Mainero, L., Cruz, C. M., Chévez, L. M., Lopez, R., Carrillo, G., Rizo, U., Saint Hillaire, E. E., Arriaga, W. E., Guadalupe Flores, R. M., Ochoa, C., Gonzalez, F., Castro, R., Stefan, A., Moreno, A., Metelus, S., Souza, R. T., ... Serruya, S. J. (2023). The burden of stillbirths in low resource settings in Latin America: Evidence from a network using an electronic surveillance system. *PLOS ONE*, 18(12), e0296002. <https://doi.org/10.1371/journal.pone.0296002>
- Mukherjee, A., Di Stefano, L., Blencowe, H., & Mee, P. (2024). Determinants of stillbirths in sub-Saharan Africa: A systematic review. *BJOG: An International Journal of Obstetrics & Gynaecology*, 131(2), 140-150. <https://doi.org/10.1111/1471-0528.17562>
- Naciones Unidas. (2015, septiembre 25). *Objetivos de Desarrollo Sostenible*. <https://www.un.org/sustainabledevelopment/es/objetivos-de-desarrollo-sostenible/>
- Organización Mundial de la Salud. (2018, febrero 4). *Recomendaciones de la OMS Para los cuidados durante el parto, para una experiencia de parto positiva*.
- Organización Mundial de la Salud. (2024). *Muerte prenatal*. https://www.who.int/es/health-topics/stillbirth#tab=tab_1
- Sotomayor Valdiviezo, K. R., Valera Vásquez, E. L., Zavala Guillén, D. Á., Vela-Ruiz, J. M., Desposorio-Robles, J., & Alatriza Gutierrez De Bambarén, M. D. S. (2023). Mortalidad en gestantes con COVID - 19 durante el periodo 2020-2021 en Latinoamérica. *Revista de Obstetricia y Ginecología de Venezuela*, 83(01), 76-85. <https://doi.org/10.51288/00830111>
- UNICEF. (2025, marzo). *Stillbirth*. <https://data.unicef.org/topic/child-survival/stillbirths/>
- Xue, T., Li, J., Tong, M., Fan, X., Li, P., Wang, R., Li, Y., Zheng, Y., Li, J., Guan, T., & Zhu, T. (2023). Stillbirths attributable to open fires and their geographic disparities in non-Western countries. *Environmental Pollution*, 334, 122170. <https://doi.org/10.1016/j.envpol.2023.122170>