

UNIVERSIDAD POLITÉCNICA ESTATAL DEL CARCHI

DIRECCIÓN DE POSGRADO



MAESTRÍA EN ESTADÍSTICA APLICADA

“Modelo de regresión múltiple para estimar el contagio de COVID-19 en base a la meteorología local de Ecuador, en el año 2020”

Trabajo de titulación previa la obtención del

Título de Magister en Estadística Aplicada

Autora: Fís. Villacís Galarza Barbara Alessia

Tutor: PhD. Castillo Páez Sergio Alberto

Tulcán, 2026

CERTIFICADO DEL TUTOR

Certifico que la maestrante Villacís Galarza Barbara Alessia con el número de cédula 175082466-4 ha elaborado el trabajo de titulación: “Modelo de regresión múltiple para estimar el contagio de COVID-19 en base a la meteorología local de Ecuador”.



f.....

PhD. Castillo Páez Sergio Alberto

TUTOR

Tulcán, 2026

AUTORÍA DE TRABAJO

El presente trabajo de titulación constituye un requisito previo para la obtención del título de Magister en Estadística Aplicada.

Yo, Villacís Galarza Barbara Alessia con cédula de identidad número 175082466-4 declaro: que la investigación es absolutamente original, auténtica, personal y los resultados y conclusiones a los que he llegado son de mi absoluta responsabilidad.



Validar únicamente en FirmaSC.
Firmado electrónicamente por:
**BARBARA ALESSIA
VILLACIS GALARZA**

f.....

Fís. Villacís Galarza Barbara Alessia

AUTORA

Tulcán, 2026

ACTA DE CESIÓN DE DERECHOS DEL TRABAJO DE TITULACIÓN

Yo, Villacís Galarza Barbara Alessia declaro ser autor/a de los criterios emitidos en el trabajo de titulación: “Modelo de regresión múltiple para estimar el contagio de COVID-19 en base a la meteorología local de Ecuador” y eximo expresamente a la Universidad Politécnica Estatal del Carchi y a sus representantes legales de posibles reclamos o acciones legales.



Validar Únicamente en FirmaEC.

Firmado electrónicamente por:

BARBARA ALESSIA

VILLACIS GALARZA

f.....

Fís. Villacís Galarza Barbara Alessia

AUTORA

Tulcán, 2026

AGRADECIMIENTO

Quiero empezar expresando un profundo agradecimiento a Dios, a mi familia, a mi tutor PhD. Sergio Castillo y cada uno de los mentores que hicieron posible este trabajo.

DEDICATORIA

Dedico este trabajo a mi madre, Margarita Galarza, quien me enseñó el significado de la valentía, la fuerza y la perseverancia, no a través de discursos incesantes sino a través de sus actos, sus incansables intentos de mejorar cada día nuestra realidad, su gran compromiso con el trabajo y sus responsabilidades de madre.

ÍNDICE

RESUMEN	xii
ABSTRACT	xiii
CAPÍTULO I.....	1
PROBLEMA	1
1.1. Planteamiento del problema.....	1
1.2. Preguntas de investigación o hipótesis.....	4
1.3. Objetivos de investigación.....	4
1.3.1. Objetivo General.....	4
1.3.2. Objetivos Específicos	5
1.4. Justificación	5
CAPÍTULO II.....	8
FUNDAMENTACIÓN TEÓRICA	8
2.1. Antecedentes de investigación	8
2.2. Marco teórico	11
2.2.1. Variables Meteorológicas y Propagación del SARS-CoV-2.	11
2.2.2. Contaminación Atmosférica y COVID-19.....	12
2.2.3. Factores Socioeconómicos y Transmisión del COVID-19 en Ecuador.....	13
2.2.4. Modelos Estadísticos Aplicados al Estudio del COVID-19.....	13
2.3. Marco legal	17
CAPÍTULO III.....	19
METODOLOGÍA.....	19
3.1. Descripción del área de estudio/grupo de estudio.....	19
3.2. Enfoque y tipo de investigación	20
3.3. Definición y operacionalización de variables	21
3.4. Procedimientos	26
3.5. Consideraciones Bioéticas	51

CAPÍTULO IV	53
RESULTADOS Y DISCUSIÓN	53
Resultados.....	53
4.1. Estadística Descriptiva	53
4.1.1. Resumen De Las Variables Cuantitativas	53
4.1.2. Resumen De Las Covariables Cuantitativas.....	56
4.1.3. Visualización de la dispersión de las variables	57
4.2. Visualización De Tendencias Temporales	59
4.3. Distribución y correlaciones de las variables.....	61
4.3.1. Distribución y correlaciones a nivel provincial.....	61
4.3.2. Correlaciones entre variables a nivel agregado.....	65
4.4. Análisis de Componentes Principales (PCA)	68
4.5. Estimación del modelo de datos de panel	71
4.5.1. Modelo Contagios COVID-19 (marzo 2020 - 2021)	72
4.5.2. Modelo Muertes por COVID-19 (marzo 2020 - 2021)	76
4.6. Validación de supuestos estadísticos	78
4.7. Ajuste funcional del modelo.....	80
DISCUSIÓN.....	81
CONCLUSIONES Y RECOMENDACIONES.....	83
CONCLUSIONES.....	83
RECOMENDACIONES.....	85
REFERENCIAS	86

ÍNDICE DE FIGURAS

Figura 1 Provincias seleccionadas para el estudio	19
Figura 2 Estaciones meteorológicas del INAMHI en Ecuador en 2020 – 2021	28
Figura 3 Diagrama de caja de variables independientes por provincia marzo 2020 - marzo 2021	57
Figura 4 Diagrama de caja de variables epidemiológicas por provincia marzo 2020 - marzo 2021	58
Figura 5 Evolución temporal de la precipitación media por provincia marzo 2020 – marzo 2021	59
Figura 6 Evolución temporal de la temperatura media por provincia marzo 2020 – marzo 2021	60
Figura 7 Evolución temporal de la concentración de PM2.5 por provincia marzo 2020 – marzo 2021	60
Figura 8 Evolución temporal del número de contagios por provincia	61
Figura 9 Evolución temporal del número de muertes.....	61
Figura 10 Distribución de variables meteorológicas, ambientales y epidemiológicas por provincia (marzo 2020 – marzo 2021)	62
Figura 11 QQ Plots de normalidad de las variables por provincia	63
Figura 12 Heatmap de correlaciones entre variables por provincia.....	65
Figura 13 QQ Plots de normalidad de las variables a nivel agregado	65
Figura 14 Heatmap de correlaciones de variables independientes a nivel agregado	67
Figura 15 Heatmap de correlaciones de covariables independientes	68
Figura 16 Proyección de las provincias en el plano factorial (biplot del PCA).....	70

ÍNDICE DE TABLAS

Tabla 1	Comparación entre Modelos de Panel	17
Tabla 2	Operacionalización de variables independientes y dependientes.	22
Tabla 3	Resumen del proceso de depuración y limpieza de datos.....	37
Tabla 4	Resumen de las variables cuantitativas por provincias marzo 2020 – marzo 2021.....	54
Tabla 5	Resumen de las variables cuantitativas a nivel agregado marzo 2020 – marzo 2021.....	55
Tabla 6	Resumen de las covariables cuantitativas proyecciones año 2020.....	56
Tabla 7	Matriz de correlaciones entre variables independientes por provincias.....	64
Tabla 8	Matriz de correlaciones entre variables independientes a nivel agregado (Spearman).....	66
Tabla 9	Matriz de correlaciones entre covariables independientes a nivel agregado (Pearson).....	67
Tabla 10	Proporción y varianza explicada por cada componente del PCA.....	69
Tabla 11	Cargas de las variables en los componentes principales (matriz de rotación)	70
Tabla 12	Resultados de modelos de panel para Contagios por COVID-19	72
Tabla 13	Modelo Pooled OLS por Bloques (sin Driscoll-Kraay)	74
Tabla 14	Modelo por bloques con errores robustos Driscoll-Kraay, variable CTG	76
Tabla 15	Modelo por bloques con errores White-Huber, variable MRT	77
Tabla 16	Resultados del test RESET de Ramsey por bloques y variable dependiente	80

RESUMEN

Este estudio examina cómo los factores meteorológicos y socioeconómicos influyen en la propagación del COVID-19 en el Ecuador, utilizando un modelo de regresión múltiple basado en datos de panel para las provincias de Pichincha, Guayas y Azuay durante el periodo marzo 2020–marzo 2021. Las variables meteorológicas (temperatura, precipitación) y concentración de *PM*_{2.5} se sintetizaron mediante un Análisis de Componentes Principales (PCA), del cual emergieron dos factores dominantes: *PC*₁, que representa condiciones cálidas y secas con mayor carga de contaminación, y *PC*₂ que refleja ambientes húmedos y ventilados. Estos componentes, junto con los indicadores socioeconómicos (densidad poblacional y desigualdad), fueron incorporados en un modelo Pooled OLS con correcciones Driscoll–Kraay y White–Huber (HC3) para garantizar inferencias robustas. Los resultados muestran que las desigualdades sociales y la densidad urbana tienen mayor influencia en la propagación del virus que las condiciones meteorológicas, aunque ambas dimensiones actúan de manera complementaria. El estudio evidencia la utilidad del enfoque de panel para fenómenos epidemiológicos y aporta una base metodológica replicable en análisis ambientales y de salud pública.

Palabras clave: COVID-19, datos de panel, contaminación atmosférica, variables meteorológicas, factores socioeconómicos.

ABSTRACT

This study analyzes the effect of meteorological and socioeconomic variables on the spread of COVID-19 in Ecuador using a multiple regression panel data model for the provinces of Pichincha, Guayas, and Azuay during the period from March 2020 to March 2021. Meteorological variables (temperature, precipitation) and $PM_{2.5}$ concentration were synthesized through a Principal Component Analysis (PCA), yielding two dominant factors: PC_1 , representing warm and dry conditions with higher air pollution, and PC_2 , capturing humid and ventilated environments. These components, together with socioeconomic indicators (population density and inequality), were included in a Pooled OLS model corrected with Driscoll–Kraay and White–Huber (HC3) estimators to ensure robust inference. The findings indicate that social inequality and urban density exert a stronger influence on virus transmission than meteorological factors, although both dimensions interact. The study highlights the usefulness of panel data methods for epidemiological analysis and provides a replicable methodological framework for future environmental and public health research.

Keywords: COVID-19, panel data, air pollution, meteorological variables, socioeconomic factors.

CAPÍTULO I

PROBLEMA

1.1. Planteamiento del problema

A finales de diciembre de 2019 se identificaron en Wuhan, China, los primeros casos de una neumonía atípica causada por el Coronavirus del Síndrome Respiratorio Agudo Severo 2 (SARS-CoV-2), agente etiológico de la enfermedad COVID-19. En marzo de 2020, la Organización Mundial de la Salud (OMS) declaró oficialmente el estado de pandemia debido a la rápida expansión del virus y a su elevada capacidad de transmisión. Desde entonces, el COVID-19 se convirtió en una de las crisis sanitarias más severas del siglo XXI, con profundas consecuencias sanitarias, económicas y sociales a escala mundial. De acuerdo con el panel estadístico oficial de la OMS, hasta 2024 se habían reportado a nivel mundial cientos de millones de casos confirmados y más de siete millones de muertes asociadas al COVID-19, enfermedad con enorme impacto sobre los sistemas de salud pública y la dinámica socioeconómica global (OMS, 2024).

América Latina fue una de las regiones más afectadas por la pandemia debido a las marcadas desigualdades estructurales, la alta informalidad laboral y las limitaciones en la capacidad hospitalaria y epidemiológica de varios países (CEPAL, 2021; ONU, 2022). De hecho, Ecuador destacó como uno de los territorios con mayor impacto sanitario durante las primeras etapas de la emergencia. El primer caso confirmado de COVID-19 en el país fue reportado el 29 de febrero de 2020; sin embargo, pocas semanas después, Ecuador presentó una de las tasas de mortalidad per cápita más elevadas de Sudamérica (Parra y Carrera, 2021). Particularmente, la provincia del Guayas experimentó un colapso sanitario y funerario sin precedentes, situación que fue ampliamente documentada por organismos nacionales e internacionales. Estudios realizados por Espinosa (2020) y reportes del Ministerio de Salud Pública (MSP) reportaron un incremento abrupto de fallecimientos entre marzo y abril de 2020, acompañado de un importante subregistro de casos y defunciones debido a la limitada disponibilidad de pruebas diagnósticas tipo PCR —Reacción en Cadena de la Polimerasa (Polymerase Chain Reaction)— y a las debilidades estructurales del sistema sanitario ecuatoriano.

Las afectaciones derivadas de la pandemia no se distribuyeron de manera homogénea en el territorio nacional: provincias como Guayas, Pichincha y Azuay presentaron dinámicas epidemiológicas bastante diferenciadas, las cuales podrían estar asociadas a factores demográficos, climáticos y socioeconómicos particulares. Guayas, caracterizada por su elevada densidad poblacional y movilidad comercial, concentró gran parte de los contagios y muertes durante los primeros meses de la pandemia (Siddharta et al., 2025; Velásquez-Serra et al., 2020). Por otro lado, Pichincha registró incrementos significativos de casos, aparentemente relacionados con la flexibilización de restricciones y la reactivación de actividades económicas y sociales, sin considerarse epicentro (Ortiz-Prado y Fernández-Naranjo, 2020). Estas diferencias territoriales sugieren que la propagación del SARS-CoV-2 no dependió exclusivamente de mecanismos biológicos inherentes al virus, sino también de condiciones ambientales y estructurales de cada región.

Diversas investigaciones internacionales han señalado que variables meteorológicas y ambientales, como temperatura, humedad relativa, precipitación y contaminación atmosférica pueden influir en la transmisión y persistencia de enfermedades respiratorias infecciosas (Hyrkäs-Palmu et al., 2025; Jiang et al., 2026; Samaila y Muhammad-Kalgo, 2024; Wagatsuma et al., 2026). Además, estudios recientes han identificado asociaciones entre concentraciones elevadas de material particulado fino PM_{2.5} —Particulate Matter, 2.5 micrómetros— y un aumento en la incidencia, severidad y mortalidad por COVID-19, por deterioro de la función respiratoria y exacerbación de enfermedades preexistentes (Rahmadina et al., 2026; Sheppard et al., 2023); investigaciones y revisiones sistemáticas en distintos países muestran que factores climáticos como la temperatura y la precipitación pueden modificar la dinámica de dispersión viral y la supervivencia del SARS-CoV-2 en el ambiente (Mecenas et al., 2020; Morris et al., 2020).

Desafortunadamente, los resultados de investigaciones publicadas a nivel internacional continúan siendo heterogéneos y, en muchos casos, contradictorios, especialmente debido a diferencias metodológicas, escalas espaciales y temporales de análisis, así como a la interacción simultánea de múltiples variables socioeconómicas y ambientales. En Ecuador, aunque existen investigaciones relacionadas con el comportamiento epidemiológico del COVID-19, la mayoría se han enfocado principalmente en análisis descriptivos, proyecciones estadísticas generales o estudios sanitarios locales, sin integrar de manera simultánea variables meteorológicas, contaminación atmosférica y factores socioeconómicos mediante modelos

econométricos robustos que permitan evaluar la heterogeneidad temporal y territorial del fenómeno.

Esta limitación es relevante al considerar variables como densidad poblacional, pobreza, desigualdad económica y estructura etaria, las cuales podrían actuar juntamente con factores ambientales en la letalidad del virus. De acuerdo con datos del Instituto Nacional de Estadística y Censos (INEC), las provincias ecuatorianas presentan marcadas diferencias socioeconómicas y demográficas que podrían modificar significativamente la vulnerabilidad de la población frente a enfermedades emergentes (INEC, 2022b). Por tanto, el análisis aislado de variables epidemiológicas resulta insuficiente para comprender la complejidad del comportamiento espacial y temporal del COVID-19 en el país.

Así, surge la necesidad de desarrollar investigaciones que permitan analizar de forma integrada la relación entre variables meteorológicas, contaminación atmosférica y factores socioeconómicos asociados al esparcimiento del COVID-19 en Ecuador, tomando como referencia investigaciones semejantes en otros países. La aplicación de modelos de regresión múltiple con datos de panel constituye una alternativa metodológica pertinente, debido a que permite evaluar simultáneamente la dimensión temporal y transversal de los datos, controlar la heterogeneidad no observable entre provincias y estimar relaciones estadísticas más consistentes entre las variables explicativas y los niveles de contagio registrados.

A pesar de la relevancia sanitaria y científica del problema, todavía existe un vacío en la literatura ecuatoriana respecto a estudios que integren información climática, ambiental y socioeconómica para modelar el comportamiento del COVID-19 mediante técnicas econométricas de datos de panel. Esta ausencia de evidencia limita la comprensión de los factores asociados a la transmisión del SARS-CoV-2 en contextos locales y dificulta la formulación de estrategias preventivas y políticas públicas adaptadas a las características territoriales del Ecuador.

Por lo expuesto anteriormente, la presente investigación propone un modelo de regresión múltiple con datos de panel para estimar el contagio de COVID-19 en función de variables meteorológicas, contaminación atmosférica y factores socioeconómicos según la meteorología local del Ecuador, tomando las provincias de Guayas, Pichincha y Azuay durante el periodo 2020-2021. Para ello se emplean datos de temperatura obtenidos del Instituto Nacional de Meteorología e Hidrología, precipitación acumulada obtenida de CHIRPs, concentraciones de PM_{2.5} derivadas de productos satelitales del

portal NASA Goddard Earth Sciences Data and Information Services Center mediante la plataforma Giovanni, así como indicadores socioeconómicos provenientes de la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) del INEC. Con ello se busca generar evidencia científica que contribuya a comprender la dinámica territorial del COVID-19 y apoyar el diseño de políticas públicas orientadas a la gestión de futuras emergencias epidemiológicas.

1.2. Preguntas de investigación o hipótesis

El presente estudio plantea las siguientes preguntas de investigación para abordar la problemática descrita:

¿Cómo se comportaron las variables meteorológicas y de contaminación del aire en las provincias seleccionadas durante el periodo de estudio?

¿Qué diferencias existen entre los factores sociodemográficos de Pichincha, Guayas y Azuay que podrían incidir en la propagación del COVID-19?

¿Qué relación existe entre las variables meteorológicas, los factores socioeconómicos y los niveles de contagio de COVID-19 en Ecuador, según un modelo estadístico de datos de panel?

Con base en estas preguntas, se establecieron las siguientes hipótesis:

H₀: No existe una relación significativa entre las variables meteorológicas y sociodemográficas con el número de contagios confirmados por COVID-19 en las provincias estudiadas.

H₁: Existe una relación significativa entre al menos una de las variables meteorológicas o sociodemográficas y el número de contagios confirmados por COVID-19 en las provincias de Pichincha, Guayas y Azuay.

1.3. Objetivos de investigación

1.3.1. Objetivo General

Cuantificar el efecto de las variables meteorológicas y factores socioeconómicos sobre el esparcimiento de COVID-19 en Ecuador, utilizando métodos estadísticos para la estimación de un modelo de regresión múltiple basado en datos de panel que capture la variabilidad temporal e interprovincial.

1.3.2. Objetivos Específicos

1. Caracterizar la evolución de variables climatológicas (temperatura, precipitación) y del contaminante atmosférico PM_{2.5} en las provincias de Pichincha, Guayas y Azuay durante el periodo marzo 2020 – marzo 2021.

2. Analizar los factores sociodemográficos de las tres provincias seleccionadas (Índice de desigualdad, tasa de pobreza, densidad poblacional y proporción de adultos mayores) y su posible influencia en los niveles de contagio y muertes por COVID-19.

3. Estimar y analizar un modelo de regresión de datos de panel que relacione las variables meteorológicas y sociodemográficas con la variabilidad mensual de contagios y muertes por COVID-19 en Ecuador, considerando las diferencias geográficas y temporales.

1.4. Justificación

La pandemia ocasionada por el SARS-CoV-2 representó uno de los mayores desafíos sanitarios, sociales y económicos del siglo XXI, en la cual se evidenció la necesidad de fortalecer la investigación científica orientada a comprender los factores asociados a la transmisión y severidad de enfermedades infecciosas emergentes. Aunque desde el inicio de la pandemia se desarrollaron numerosos estudios epidemiológicos y modelos predictivos, aún persisten limitaciones en la comprensión integral de la interacción entre variables ambientales, meteorológicas y socioeconómicas relacionadas con el comportamiento del COVID-19, especialmente en contextos territoriales específicos como el ecuatoriano.

Diversas investigaciones internacionales han indicado que factores como la temperatura, precipitación, contaminación atmosférica y densidad poblacional pueden influir en la dinámica de transmisión de enfermedades respiratorias. Sin embargo, los resultados que reportan las investigaciones continúan siendo heterogéneos debido a diferencias metodológicas, disponibilidad de datos y características propias de cada región. En Ecuador, los estudios desarrollados sobre COVID-19 se han concentrado principalmente en análisis descriptivos o epidemiológicos generales, y todavía son limitadas las investigaciones que integran simultáneamente variables meteorológicas, contaminación atmosférica y factores socioeconómicos mediante modelos econométricos de datos de panel.

La presente investigación propone un modelo de regresión múltiple con datos de panel para analizar la relación entre variables climáticas, concentración de material

particulado PM2.5 y factores socioeconómicos asociados al contagio de COVID-19 en las provincias de Guayas, Pichincha y Azuay durante el periodo 2020-2021. La aplicación de modelos de datos de panel permite evaluar simultáneamente la variación temporal e interprovincial del fenómeno, controlar la heterogeneidad no observable y obtener estimaciones más robustas respecto a los posibles factores asociados al comportamiento epidemiológico del virus.

Asimismo, esta investigación tiene un gran componente práctico y social, pues el planteamiento de un modelo de panel en el contexto de la pandemia puede constituir un insumo técnico para el análisis epidemiológico y la formulación de estrategias de prevención y respuesta ante futuras emergencias sanitarias. La identificación de posibles relaciones entre factores ambientales, climáticos y socioeconómicos con la propagación del COVID-19 puede contribuir al fortalecimiento de los sistemas de vigilancia epidemiológica y a la toma de decisiones basadas en evidencia científica dentro del ámbito de la salud pública.

Desde el ámbito académico, el estudio aporta al fortalecimiento de la investigación interdisciplinaria entre estadística, epidemiología, ciencias ambientales y análisis de datos, y promueve así el uso de herramientas cuantitativas avanzadas aplicadas al estudio de fenómenos epidemiológicos complejos. Además, contribuye al desarrollo de evidencia científica contextualizada para Ecuador, donde aún existe limitada producción académica relacionada con modelos econométricos aplicados al análisis territorial del COVID-19.

La investigación se enmarca en la línea de investigación de análisis estadístico y modelización de fenómenos socioambientales y epidemiológicos, al integrar técnicas econométricas de datos de panel con variables climáticas, ambientales y sociodemográficas para el estudio de una problemática de salud pública. De igual manera, el estudio guarda relación con los Objetivos de Desarrollo Sostenible (ODS) planteados por la Organización de las Naciones Unidas (Naciones Unidas, 2018) particularmente con el ODS 3 y ODS 13, respectivamente:

- Salud y Bienestar, orientado a garantizar una vida sana y promover el bienestar para todas las personas;
- Acción por el Clima, orientado al análisis de variables de fenómenos meteorológicos y ambientales para permitir un desarrollo sostenible.

Además, cabe resaltar que esta investigación se alinea con el Plan de Creación de Oportunidades 2021-2025 de Ecuador, el cual promueve el fortalecimiento de la educación superior, la investigación científica y la generación de conocimiento aplicado para la solución de problemáticas nacionales. En particular, el Objetivo 7 del mencionado plan establece la necesidad de impulsar investigación de alto impacto orientada al desarrollo científico y tecnológico del país (Secretaría Nacional de Planificación, 2021). Así, la presente investigación busca fortalecer el conocimiento científico sobre la relación entre factores ambientales y epidemiológicos en Ecuador, con evidencia cuantitativa útil para futuras investigaciones y para la gestión de riesgos sanitarios en contextos territoriales específicos.

CAPÍTULO II

FUNDAMENTACIÓN TEÓRICA

2.1. Antecedentes de investigación

El estudio sobre la propagación del SARS-CoV-2 es y seguirá siendo un área de interés en la comunidad científica dado que es crucial conocer el comportamiento de un virus a nivel local (principales provincias de Ecuador) para poder elaborar estrategias de mitigación de este. Desde la aparición del virus en 2019, diversas investigaciones han explorado su relación con factores meteorológicos, contaminación ambiental y condiciones socioeconómicas, tanto a corto como a largo plazo. La crisis epidemiológica generada por el SARS-CoV-2, reconocida como uno de los desafíos de salud pública más profundos en la era moderna, llevó a la adopción de medidas de confinamiento y estrategias de mitigación, las cuales fueron orientadas por estudios epidemiológicos y modelos predictivos basados en datos ambientales y sociodemográficos.

Para poder predecir el comportamiento de este virus la formulación óptima vendría de un modelo Newtoniano, es decir, un modelo en el que la física clásica describa exactamente la interacción del virus con su medio circundante y los efectos causados por el mismo. Sin embargo, aunque la física tiene mucho que ofrecer en cuanto a la comprensión de cómo se mueven las partículas por el aire y cómo interactúan con las superficies, predecir la propagación de un virus nuevo y desconocido es una tarea compleja que requiere un enfoque multidisciplinario (Singhal, 2020). Los epidemiólogos, los funcionarios de salud pública y los expertos médicos son los que se especializan en predecir la propagación de enfermedades infecciosas como el coronavirus. Para ello se utiliza modelos matemáticos, análisis de datos y otras herramientas tal que sea posible predecir la propagación del virus y desarrollar estrategias para mitigar su impacto.

Diversos trabajos han puesto énfasis en comprender si las características climáticas del entorno modifican los patrones de circulación del virus. Previas investigaciones (Binaku y Schmeling, 2017; Elminir, 2005; Manotoa, 2022; Srivastava, 2021; Sun et al., 2020; Zheng et al., 2021) han detallado que los contaminantes atmosféricos, los índices de movilidad y las variables meteorológicas pueden influir en la propagación de los distintos coronavirus y en su letalidad. Para hacer frente a la transmisión, es necesario conocer las condiciones meteorológicas que favorecen o

afectan a la mortalidad del virus, así como los contaminantes atmosféricos y otros factores que pueden influir en las muertes causadas por la epidemia. De este modo, un análisis de su asociación puede facilitar a los responsables políticos la adopción de estrategias y la implementación de medidas de seguridad sanitaria para afrontar los impactos producidos por el virus hacia la población con respecto a salud, interacciones y labores sociales, educación, empleo y salud mental.

Existen antecedentes que sugieren que variables como la temperatura, la humedad relativa, la radiación ultravioleta y la velocidad del viento pueden incidir en la transmisibilidad del SARS-CoV-2 (Passerini et al., 2020). En este sentido, estudios han demostrado que la epidemia de SARS en 2002 y la de COVID-19 en 2019 surgieron en períodos fríos y secos en China, lo que sugiere un posible patrón estacional de transmisión viral (Sun et al., 2020). Sin embargo, los hallazgos sobre la relación entre el clima y la propagación del COVID-19 no han sido concluyentes. Investigaciones como las de Islam et al. (2020) encontraron que la temperatura y la humedad relativa estaban inversamente asociadas con la tasa de contagio del virus, sugiriendo que el clima frío y seco favorecía su supervivencia. Por otro lado, Ficetola y Rubolini (2020) identificaron relaciones no lineales entre el crecimiento del COVID-19 y variables climáticas, mientras que otros estudios (Chen et al., 2020) señalaron que la transmisión del virus era óptima a temperaturas de aproximadamente 8°C y niveles de humedad relativa entre el 60% y el 90%.

Así mismo, existen investigaciones que no han encontrado asociaciones significativas entre la temperatura y la incidencia del virus, lo que sugiere que otros factores como el impacto de la contaminación atmosférica pueden desempeñar un papel más determinante en la propagación del SARS-CoV-2. Estudios previos han evidenciado que la exposición a partículas finas ($PM_{2.5}$) está relacionada con un aumento en la incidencia y letalidad del virus, debido a sus efectos en la inflamación del sistema respiratorio y la debilitación del sistema inmunológico (Altamimi y Ahmed, 2020; Rahman et al., 2020). En particular, se ha observado que regiones con altos niveles de contaminación atmosférica presentan tasas de mortalidad más elevadas por COVID-19, lo que refuerza la hipótesis de que la calidad del aire puede ser un factor clave en la evolución de la pandemia (Wu et al., 2020). El material particulado fino ($PM_{2.5}$) es considerado una variable clave para evaluar la propagación del COVID-19, pues se asocia con contaminantes atmosféricos capaces de penetrar profundamente en el sistema respiratorio humano, aumentando la susceptibilidad a enfermedades

infecciosas, tal como lo han demostrado diversos estudios recientes (Houweling et al., 2024; Kephart et al., 2021).

McIntosh y Hirsch (2023) en su recopilación tienen en cuenta los últimos resultados de las investigaciones sobre la transmisión del COVID-19 y las medidas de prevención, y hacen hincapié que el virus se propaga principalmente a través de las gotitas respiratorias y el contacto estrecho con personas infectadas. Por otro lado, la transmisión del Covid-19 no depende solamente del contacto directo de persona a persona, pues se ha investigado ampliamente la relación entre la transmisión de COVID-19 y las variables meteorológicas (Ai et al., 2022; Islam y Noor, 2022; Ogunjo et al., 2022; Zheng et al., 2021), analizando la relación entre factores meteorológicos como la temperatura, la humedad, las precipitaciones, la velocidad del viento, la presión atmosférica y la transmisión de COVID-19. Los resultados de cada investigación son diversos e, incluso, las relaciones halladas pueden resultar contradictorias, mas es importante considerar que todas estas investigaciones se desarrollan en diversas condiciones climáticas, geográficas y demográficas por lo que es trascendental extender la documentación disponible acerca de la propagación y contagio de este virus para la realidad nacional de Ecuador.

Además de las condiciones ambientales, los factores socioeconómicos también han sido identificados como determinantes en la propagación del COVID-19. Investigaciones como las de Coelho et al. (2020), Armijos-Toro et al. (2021) y Shammi et al. (2020) han mostrado que variables como la densidad poblacional, el acceso a servicios de salud, el nivel de pobreza y la urbanización pueden influir significativamente en la dinámica de transmisión del virus. En Ecuador, la crisis sanitaria impactó de manera diferenciada a las distintas provincias, con efectos más severos en aquellas con mayor densidad de población y menor acceso a infraestructura sanitaria (Jaffe et al., 2020). A pesar de la abundante literatura sobre estos factores, sigue existiendo una falta de consenso sobre la magnitud exacta de su impacto en la propagación del virus, en gran parte debido a diferencias metodológicas en los estudios, la variabilidad de las muestras y la dificultad para aislar el efecto de cada variable (Carlson et al., 2020).

Uno de los métodos estadísticos más utilizados en la literatura científica para estudiar la propagación del COVID-19 es el modelo de regresión lineal múltiple. Rath et al. (2020) indican que la principal limitación de la regresión lineal es que explora una relación entre la media de las variables de entrada y las variables de salida, pero dado que la media no es una descripción completa de una única variable, la regresión

lineal no permite comprender claramente las relaciones entre variables. Por lo tanto, se obtiene un modelo más fiable al realizar un análisis de los distintos factores utilizando un modelo de regresión lineal múltiple (Almalki et al., 2022; Gambhir et al., 2020; Khan et al., 2022).

El estudio publicado por Guo et al. (2022) usa datos de panel para analizar la evolución del COVID-19 a través del tiempo y en diferentes regiones, lo que permite capturar la variabilidad tanto temporal como espacial de la pandemia. Este enfoque metodológico combina observaciones longitudinales de múltiples unidades y facilita el control de heterogeneidades no observadas que pueden influir en la propagación del virus. Armijos-Toro et al. (2021) estudian la propagación del COVID-19 con modelo de datos de panel y hallan que los niveles de infección no tienen por qué ser necesariamente similares entre provincias, pues existen factores locales que influyen en el número de personas infectadas por COVID-19. Paralelamente Cirakli et al. (2022) basándose en Guliyev (2020), emplearon un enfoque de datos de panel y su análisis demostró que la metodología de datos de panel facilita la identificación de tendencias a largo plazo, permitiendo evaluar de forma más precisa los efectos diferenciales de las medidas implementadas a lo largo de la pandemia.

En este contexto, el uso de modelos de datos de panel se presenta como una metodología robusta para analizar estos factores de manera conjunta y evaluar su efecto en la propagación del COVID-19 a lo largo del tiempo y en diferentes regiones del Ecuador. Por lo tanto, la presente investigación busca contribuir al conocimiento existente mediante la estimación de un modelo de regresión de datos de panel que permita evaluar el impacto de variables meteorológicas, contaminación atmosférica y factores socioeconómicos en la propagación del COVID-19 en Ecuador durante el período 2020-2021.

2.2. Marco teórico

2.2.1. Variables Meteorológicas y Propagación del SARS-CoV-2.

Las variaciones en las condiciones meteorológicas han demostrado influir en la dinámica de transmisión del SARS-CoV-2. Por ejemplo, Sangkham et al. (2021) reportaron que, en la Región Metropolitana de Bangkok, parámetros como la temperatura, la humedad relativa, la humedad absoluta y la velocidad del viento mostraron asociaciones significativamente positivas con el número diario de casos de COVID-19. Estos hallazgos sugieren que las fluctuaciones en las condiciones climáticas

pueden afectar la supervivencia y la transmisión del virus, posiblemente modificando el comportamiento de las gotas respiratorias y la susceptibilidad del huésped. Esta relación subraya la importancia de comprender la variabilidad meteorológica regional al evaluar el riesgo y la propagación del virus (Sangkham et al., 2021).

En contraste, otros estudios han indicado que las condiciones más cálidas y húmedas podrían suprimir la transmisión viral. Wang et al. (2020) encontraron que el incremento de la temperatura y la humedad relativa se asociaba significativamente con una reducción en el número reproductivo efectivo (R) del COVID-19, estimando que un aumento de un grado Celsius en la temperatura y un incremento de un punto porcentual en la humedad relativa reducían R en 0.0225 y 0.0158, respectivamente. Estos resultados respaldan la hipótesis de que temperaturas y niveles de humedad más altos pueden disminuir la viabilidad del SARS-CoV-2 en el ambiente, mitigando así su propagación. En conjunto, estos estudios destacan una interacción compleja entre los factores meteorológicos y la propagación del COVID-19, sugiriendo que las condiciones ambientales pueden exacerbar o mitigar la transmisión según el contexto regional específico (Wang et al., 2020).

2.2.2. Contaminación Atmosférica y COVID-19.

La relación entre las variables meteorológicas y la concentración de contaminantes atmosféricos es fundamental para comprender la dinámica de la calidad del aire en entornos urbanos. En el distrito de Villa María del Triunfo, durante mayo de 2018, se llevó a cabo un estudio que analizó esta correlación, considerando contaminantes como PM_{10} , $PM_{2.5}$, CO, SO_2 , NO_2 y O_3 , junto con diversos indicadores climáticos, incluyendo temperatura, humedad y dinámica del viento, y los resultados evidenciaron que cuando la temperatura aumenta, los valores de CO tienden a disminuir, mientras que la correlación entre la temperatura y el O_3 fue directamente proporcional y fuerte. Además, la humedad relativa mostró una correlación inversa con PM_{10} , CO y NO_2 , siendo más pronunciada en el caso del CO (RIOS GARCES, 2020).

En la tesis de Valeriano (2021), se investigó la relación entre variables meteorológicas (temperatura, humedad relativa, velocidad y dirección del viento) y la dispersión espaciotemporal de contaminantes atmosféricos (NO_2 , O_3 , SO_2 , PM_{10} y $PM_{2.5}$) en el distrito de Pacocha, provincia de Ilo, durante 2019 y 2020. Los datos se obtuvieron de la estación de monitoreo de calidad del aire de la UNAM – Moquegua. El estudio encontró que las condiciones meteorológicas influyen significativamente en las concentraciones de estos contaminantes, afectando su dispersión y, por ende, la calidad

del aire en la región; el análisis estadístico indicó un coeficiente de correlación negativo asociado a la temperatura y las concentraciones de NO_2 y SO_2 , mientras que la humedad relativa mostró una relación directa con $\text{PM}_{2.5}$ y PM_{10} . Además, la velocidad del viento presentó una correlación inversa con las concentraciones de PM_{10} y $\text{PM}_{2.5}$ indicando que velocidades más altas favorecen la dispersión de estas partículas.

Diversos estudios han analizado cómo la exposición a contaminantes del aire puede influir en la propagación y severidad del virus SARS-CoV-2 (Cao et al., 2021; Chubarova et al., 2021; Tellier et al., 2019). Específicamente, una revisión sistemática en Oyarzún y Valdivia (2021), sugiere que la exposición a $\text{PM}_{2.5}$, PM_{10} , NO_2 y ozono en el caso de la incidencia. Además, durante los confinamientos implementados para frenar la propagación del virus, se observaron reducciones significativas en las concentraciones de ciertos contaminantes atmosféricos. Por ejemplo, Hernández-Paniagua et al. (2021) en el Valle de México, la calidad del aire mejoró notablemente debido a la disminución de emisiones de NO_2 y CO , lo que resultó en beneficios para la salud de la población. n. n.

2.2.3. Factores Socioeconómicos y Transmisión del COVID-19 en Ecuador.

En Armijos-Toro et al. (2021) se estudió la relación entre los diferentes niveles de infección por COVID 19 específicamente en las provincias de Ecuador, con sus características sociodemográficas respectivas, incluyendo diferente información provincial relacionada con la distribución de la población de cada provincia según ingresos, actividad económica, acceso a servicios básicos, educación, entre otros factores. En esta investigación implementaron un modelo de datos de panel que permite identificar los efectos fijos individuales de cada provincia con respecto al número total de infecciones que constan en los registros, y hallaron que los niveles de infección dependen de factores locales: particularmente al analizar Guayas y Pichincha se pudo asociar, por análisis de componentes principales, los contagios con la actividad económica y las infraestructuras de los hogares de cada provincia.

2.2.4. Modelos Estadísticos Aplicados al Estudio del COVID-19.

Los modelos estadísticos han desempeñado un papel esencial en el análisis de la pandemia de COVID-19, proporcionando herramientas para comprender la dinámica de transmisión, evaluar políticas sanitarias y predecir escenarios epidemiológicos. Desde 2020, el uso de técnicas de regresión y panel de datos ha permitido integrar

dimensiones espaciales y temporales, lo que favorece una comprensión más completa de la propagación del virus (Baltagi, 2021; Kumar et al., 2022).

Diversos estudios han empleado modelos de series temporales, regresiones espaciales y paneles dinámicos para explorar la relación entre los factores ambientales, demográficos y socioeconómicos con los indicadores de contagio y mortalidad. Por ejemplo, Ganegoda et al. (2021) y Ahmad et al. (2022) demostraron que las variaciones térmicas y de humedad influyen en la transmisibilidad del SARS-CoV-2, mientras que Gozzi et al. (2021) y Oronce et al. (2020) confirmaron que la desigualdad y la densidad urbana potencian la vulnerabilidad al contagio. En América Latina, estudios recientes han destacado la importancia de incorporar factores estructurales y territoriales —como el acceso desigual a servicios de salud y la contaminación del aire— en la modelización de la pandemia (Castro et al., 2025; Pérez et al., 2023). Estas aproximaciones consolidan el uso de modelos de panel con correcciones robustas como una estrategia idónea para analizar procesos epidemiológicos multivariados con heterogeneidad espacial.

Análisis de Componentes Principales (PCA)

El Análisis de Componentes Principales (PCA) es una técnica estadística multivariante utilizada para reducir la dimensionalidad de un conjunto de variables correlacionadas, transformándolas en un nuevo sistema de variables ortogonales llamadas componentes principales (Jolliffe y Cadima, 2016; Abdi y Williams, 2010). Matemáticamente, el PCA busca una transformación lineal:

$$Z = XW,$$

donde Z es la matriz de componentes principales, X la matriz de variables originales y W la matriz de autovectores o pesos que maximizan la varianza. Cada componente principal se expresa como:

$$Z_k = w_{\{1k\}}X_1 + w_{\{2k\}}X_2 + \dots + w_{\{pk\}}X_p$$

Los autovalores asociados a cada autovector representan la proporción de varianza explicada por cada componente, y su suma equivale a la varianza total del conjunto. En epidemiología ambiental, el PCA permite sintetizar múltiples indicadores meteorológicos (como temperatura, precipitación y concentración de $PM_{2.5}$) en pocos componentes que resumen la estructura subyacente del clima y la contaminación (Hair et al., 2019; Ahmad et al., 2022). Su empleo reduce la multicolinealidad entre variables explicativas, mejorando la eficiencia y estabilidad de los modelos econométricos subsiguientes.

Modelos de Regresión para Datos de Panel

El análisis de datos de panel combina observaciones temporales y transversales, permitiendo estudiar simultáneamente la variación temporal dentro de cada unidad y las diferencias estructurales entre unidades (Baltagi, 2021). Esta dualidad mejora la precisión de las estimaciones, controla la heterogeneidad no observada y aumenta los grados de libertad (Wooldridge, 2021). Los principales enfoques para modelar datos de panel son el modelo Pooled OLS, el modelo de Efectos Fijos (FE) y el modelo de Efectos Aleatorios (RE).

Modelo Pooled (Agrupado)

El modelo Pooled OLS asume que los datos de panel pueden combinarse en una sola muestra transversal, ignorando las diferencias individuales entre unidades o periodos (Gujarati y Porter, 2020). Su forma general es:

$$Y_{it} = \beta_0 + \beta_1 X_{1it} + \dots + \beta_k X_{kit} + \varepsilon_{it}$$

donde Y_{it} representa la variable dependiente de la unidad i en el tiempo t , X_{kit} son las variables explicativas, y ε_{it} es el término de error idiosincrático. Aunque es el más simple, su limitación radica en que no controla la heterogeneidad entre unidades, por lo que suele ser un punto de partida más que un modelo final (Baltagi, 2021).

Modelo de Efectos Fijos (FE)

El modelo de Efectos Fijos (FE) posibilita ajustar por la heterogeneidad no observable existente entre las distintas unidades de análisis, asumiendo que esta es constante en el tiempo y puede capturarse mediante interceptos individuales. Su forma general es:

$$Y_{it} = \alpha_i + \beta_1 X_{1it} + \beta_2 X_{2it} + \dots + \beta_k X_{kit} + u_{it}$$

donde α_i representa el efecto específico de cada unidad, que recoge características invariables en el tiempo. Este modelo se estima eliminando los efectos fijos mediante la transformación within o centrado en la media temporal (Wooldridge, 2021). El modelo de efectos fijos es preferible en situaciones en las que los componentes no medidos podrían estar vinculados a los predictores. Como contrapartida, este enfoque descarta la posibilidad de analizar variables que se mantienen constantes en el tiempo, ya que su falta de variación impide identificar sus coeficientes.

Modelo de Efectos Aleatorios (RE)

El modelo de Efectos Aleatorios (RE) asume que la heterogeneidad entre unidades es aleatoria y no está correlacionada con las variables explicativas (Hsiao, 2022; Baltagi, 2021). Se expresa como:

$$Y_{it} = \beta_0 + \beta_1 X_{1it} + \dots + \beta_k X_{kit} + \mu_i + \varepsilon_{it}$$

donde μ_i representa el componente no observable específico de cada unidad, tratado como un término aleatorio con distribución normal, media nula y una varianza definida σ^2 , y ε_{it} el error idiosincrático. El estimador RE utiliza una combinación ponderada de la variación entre y dentro de las unidades, lo que lo hace más eficiente que el FE cuando se cumple la exogeneidad. La elección entre FE y RE se realiza mediante la prueba de Hausman, que contrasta la correlación entre μ_i y X_{it} (Hausman, 1978; Wooldridge, 2021).

Tabla 1*Comparación entre Modelos de Panel*

Criterio	Pooled OLS	Efectos Fijos (FE)	Efectos Aleatorios (RE)
Supuesto clave	Homogeneidad entre unidades y periodos	Heterogeneidad no observada correlacionada con los regresores	Heterogeneidad no observada no correlacionada con los regresores
Ecuación general	$Y_{it} = \beta_0 + \beta X_{it} + \varepsilon_{it}$	$Y_{it} = \alpha_i + \beta X_{it} + u_{it}$	$Y_{it} = \beta_0 + \beta X_{it} + \mu_i + \varepsilon_{it}$
Método de estimación	Mínimos Cuadrados Ordinarios (OLS)	Mínimos Cuadrados Within (demeaned)	Generalized Least Squares (GLS)
Ventajas	Sencillo, útil como modelo base	Controla heterogeneidad no observable correlacionada	Más eficiente si se cumple exogeneidad
Desventajas	Ignora heterogeneidad individual	No estima variables invariantes	Inconsistente si existe correlación entre μ_i y X_{it}
Prueba de selección	Chow Test (vs FE)	Hausman Test (vs RE)	Breusch–Pagan LM (vs Pooled)

2.3. Marco legal

Por cuanto a la normativa que se encuentra vigente del Ministerio de Salud Pública (MSP), es importante detallar las resoluciones que constan en el Sistema Integrado de Vigilancia Epidemiológica del Ecuador (SIVE). Se especifica que “el Estado garantizará este derecho mediante políticas económicas, sociales, culturales, educativas y ambientales; y, el acceso permanente, oportuno y sin exclusión a

programas, acciones y servicios de promoción y atención integral de salud...” (MSP, 2014).

Se establecen entonces políticas reguladoras, entre las que se encuentra principalmente el mejorar los servicios de cuidado de la salud, de forma preventiva, lo que permite a su vez mejorar la calidad de vida de los ciudadanos. Para este particular es pertinente enunciar el literal b, donde se establece que es imprescindible mantener de forma permanente un perfil epidemiológico y sanitario del país, ya que constituye una herramienta clave para organizar la oferta de servicios de prevención y promoción de la salud en función de las necesidades previstas. Asimismo, el literal c indica que, para enfrentar adecuadamente los riesgos que podrían generar aumentos en la mortalidad o morbilidad, es fundamental reforzar los mecanismos de vigilancia y control epidemiológico. Finalmente, se señala que la detección temprana de enfermedades y patógenos requiere de un decidido impulso a la investigación en el ámbito sanitario, con el fin de reconocer procesos, mecanismos y estrategias que permitan mitigar la posible expansión de futuras epidemias.

El principal objetivo de las políticas y leyes presentadas es el de alertar a la ciudadanía en general y al MSP para que pueda dar una respuesta oportuna en caso de que susciten eventos de alto potencial epidémico. Así sería posible evitar que se desencadene una o varias emergencias de Salud Pública (MSP, 2014). El desarrollo de la presente investigación está en concordancia con las leyes detalladas en este apartado ya que a través del modelo de regresión que se busca obtener será posible identificar y analizar la relación de las variables meteorológicas, climáticas y sociodemográficas con la transmisión de Covid-19 en Ecuador tal que sea posible para los profesionales de la salud actuar oportunamente frente a casos de enfermedades, incluyendo epidemias, que están bajo estrategias de control y eliminación.

CAPÍTULO III

METODOLOGÍA

3.1. Descripción del área de estudio/grupo de estudio

Si bien la mayoría de los países del mundo han reportado efectos cuantitativos del COVID-19 en la población, como nuevos casos y muertes relacionadas con el virus, Ecuador posee registros de dichos efectos a partir del 13 de marzo de 2020 (SNGRE, 2020), pues, aunque el primer contagio confirmado fue reportado oficialmente el 29 de febrero de 2020, la emergencia sanitaria fue declarada por el Gobierno Nacional el 13 de marzo de 2020. Tomando esto en consideración la presente investigación abarca un rango de registros mensuales de Contagios y Muertes por COVID-19 desde el mes de marzo de 2020 hasta marzo de 2021 con un total de 13 observaciones por cada provincia: se seleccionaron estratégicamente 3 provincias del Ecuador, lo que genera un total de 39 observaciones.

Para el desarrollo del análisis se seleccionaron las provincias de Pichincha, Guayas y Azuay como áreas de estudio, con el propósito de examinar el comportamiento de la COVID-19 en distintos contextos territoriales del Ecuador. La Figura 1 muestra la ubicación geográfica de las tres provincias seleccionadas dentro del territorio nacional.

Figura 1

Provincias seleccionadas para el estudio.



Nota. Elaboración propia, el color amarillo representa la provincia de Pichincha, el naranja corresponde a Guayas y el verde a Azuay.

El presente estudio se desarrolla en el contexto ecuatoriano, tomando como área de análisis las provincias de Guayas, Pichincha y Azuay, cuyas capitales Guayaquil, Quito y Cuenca respectivamente, representan centros urbanos clave con realidades geográficas, climáticas y socioeconómicas aparentemente diferenciadas. La selección de estas provincias no fue aleatoria, sino estratégica, al considerar los siguientes criterios:

- Representatividad climática: Guayas se ubica en la región litoral con clima tropical húmedo; Pichincha en la Sierra con clima templado andino, y Azuay presenta características interandinas de menor densidad poblacional.
- Diversidad en concentración poblacional y urbanización: Guayaquil y Quito son las ciudades más pobladas del país, mientras que Cuenca presenta menor densidad poblacional por kilómetro cuadrado, lo cual permite analizar el efecto de la densidad humana sobre la propagación del virus.
- Diferencias en la respuesta sanitaria y gubernamental: Cada ciudad presentó variaciones en la aplicación de políticas de contención, manejo hospitalario y campañas de prevención durante el periodo más crítico de la pandemia.

3.2. Enfoque y tipo de investigación

3.2.1. Enfoque

El presente estudio adopta un enfoque cuantitativo, centrado en el análisis exhaustivo de datos numéricos relacionados con variables meteorológicas, calidad del aire y factores socioeconómicos en el contexto de la pandemia de COVID-19 en Ecuador. Este enfoque permite una evaluación detallada y objetiva de las interrelaciones entre estas variables y la propagación del virus.

3.2.2. Tipo de Investigación

Se trata de una investigación no experimental, ya que el análisis se basa exclusivamente en datos previamente recopilados, sin modificar las condiciones bajo las cuales esas variables fueron originalmente registradas y en las bases de datos del INEC y ENEMDU. Además, es un estudio longitudinal que abarca el periodo comprendido entre marzo de 2020 y marzo de 2021, lo cual permite analizar la evolución temporal de la pandemia en Ecuador. El carácter explicativo de la investigación se sustenta en el uso de modelos de regresión de datos de panel, que permiten identificar y cuantificar la relación causal entre las condiciones ambientales y la propagación del COVID-19 a lo largo del tiempo y entre diferentes regiones del país.

3.3. Definición y operacionalización de variables

A continuación, se presenta la definición conceptual y operativa de las variables consideradas en el presente estudio, distinguiendo entre las variables explicativas (independientes) y las variables de respuesta (dependientes). Asimismo, se detallan los indicadores y procedimientos utilizados para su medición, con el fin de captar su aporte específico dentro del marco analítico propuesto y su relevancia en el contexto del Ecuador durante la pandemia de COVID-19.

Las variables dependientes analizadas son el número mensual de contagios por COVID-19 y el número mensual de muertes por COVID-19, ambas cuantificadas por provincia y mes. Estas fueron modeladas en función de las variables independientes, dentro de un marco longitudinal y comparativo que permite evaluar su evolución y asociación durante el período comprendido entre marzo de 2020 y marzo de 2021.

Tabla 2

Operacionalización de variables independientes y dependientes.

Expectativa		Examinar los factores que contribuyen o inhiben el carácter contagioso del virus SARS-CoV-2.				
Población		La población de estudio corresponde a los casos confirmados de Covid-19 en Ecuador.				
Muestra		La muestra de estudio corresponde a los casos confirmados de Covid-19 en Ecuador, particularmente en las provincias de Azuay, Guayas, Pichincha.				
Variable	Definición conceptual	Dimensión o unidad de análisis	Indicador	Ítem o ratio	Fuente del Ítem	Técnica e Instrumentos estadísticos
CTG	Número de casos positivos confirmados de COVID-19 en una provincia	Provincia / mes	Incidencia mensual de casos confirmados	Número de contagios mensuales	MSP / SNGRE / ECUACOVID Robalino (2021)	Regresión de datos de panel

MRT	Número de muertes confirmadas por COVID-19 en una provincia	Provincia / mes	Número de muertes	Número de muertes mensuales	MSP / SNGRE / ECUACOVID Robalino (2021)	Regresión de datos de panel
Variable	Definición conceptual	Dimensión o unidad de análisis	Indicador	ítem o ratio	Fuente del Ítem	Técnica e Instrumentos estadísticos
Precipitación	Cantidad promedio de lluvia caída en una provincia durante un mes, medida en milímetros	Provincia / mensual	Precipitación media mensual	Promedio mensual en mm	INAMHI	Regresión panel / PCA
PM2.5	Concentración promedio de material particulado fino en suspensión en el aire	Provincia / mensual	Concentración mensual promedio de PM2.5	µg/m³ o equivalente satelital (kg/m³)	NASA Giovanni / INAMHI	Regresión panel / PCA

TMP	Nivel de desigualdad económica expresado a través de medidas de ingreso	Provincia / mensual	Temperatura media mensual	Promedio en °C	INAMHI	Regresión panel / PCA
-----	---	------------------------	------------------------------	----------------	--------	--------------------------

Variable	Definición conceptual	Dimensión o unidad de análisis	Indicador	Ítem o ratio	Fuente del Ítem	Técnica e Instrumentos estadísticos
Coeficiente de Gini	Medida de desigualdad en la distribución del ingreso	Provincia/anual	Gini provincial	Valor entre 0 y 1	ENEMDU – INEC (2019)	Regresión panel
Tasa de pobreza	Proporción de la población que no cubre sus necesidades básicas	Provincia/anual	Porcentaje de personas en situación de pobreza multidimensional	Porcentaje (%)	ENEMDU – INEC (2019)	Regresión panel
Densidad poblacional	Número de habitantes por kilómetro cuadrado	Provincia/anual	Habitantes por km ²	Valor numérico	Proyecciones INEC (2020)	Regresión panel
Proporción de adultos ≥60	Porcentaje de población mayor a 60 años como indicador de vulnerabilidad	Provincia/anual	Porcentaje de adultos mayores	Porcentaje (%)	INEC (ENEMDU 2019)	Regresión panel

3.4. Procedimientos

Selección de variables para el modelo de regresión.

La unidad de análisis de la presente investigación corresponde a observaciones mensuales para cada provincia, abarcando el periodo comprendido entre marzo de 2020 y marzo de 2021. Este intervalo permite capturar tanto la fase inicial de expansión del virus como los momentos de mayor impacto epidemiológico, antes del inicio masivo de campañas de vacunación.

Para cada provincia, se recopilaron y organizaron datos correspondientes a:

- **Indicadores epidemiológicos:** número mensual de contagios de COVID-19 confirmados, número de muertes por COVID-19.
- **Variables meteorológicas:** temperatura y precipitación mensual media.
- **Contaminantes atmosféricos:** concentración superficial mensual promedio de PM2.5.
- **Indicadores socioeconómicos:** coeficiente de Gini anual, tasa de pobreza multidimensional anual, densidad poblacional anual y proporción de adultos mayores anual.

Se elige tres provincias de Ecuador al existir disponibilidad de datos para las variables involucradas en este estudio como son variables meteorológicas, socioeconómicas y epidemiológicas. Así las provincias que se analizan en el presente estudio son: Pichincha, Azuay, Guayas. Es posible que se produzcan sesgos debido a la disparidad de datos por el contagio de COVID-19 en diferentes regiones (es decir, hay ciudades donde se llevaron registros de forma continua y constante mientras en otras hubo ausencia de registros pese a que si existieron contagios).

Teniendo esto en consideración es necesario analizar tanto cuantitativa como cualitativamente los datos, para lograr que el efecto de patrones por región sea mínimo, (prácticamente despreciable) por cuanto concierne a meteorología local, misma que obedece en gran medida la geografía del sector, según la Primera Ley Geográfica de Tobler (1970) como se cita en Andrades (2015): en los datos espaciales es bastante común que todo esté relacionado con todo, pero las cosas que están próximas entre sí, están más relacionadas que las que se hallan distantes.

Existen limitaciones en la disponibilidad y calidad de datos oficiales relacionados con la pandemia de COVID-19 en Ecuador. Por ejemplo, las bases de datos del Ministerio de Salud Pública solo reportan registros del COVID-19 actualizados hasta julio de 2021, lo cual

implica un posible subregistro de contagios y fallecimientos, especialmente en regiones con acceso limitado a servicios de salud (INEC, 2021, 2022a; Trujillo y Merino, 2024). No obstante, el presente estudio se apoya en un diseño metodológico de tipo longitudinal-comparativo, a través de modelos de datos de panel balanceados, que permiten capturar la variabilidad temporal y las diferencias estructurales entre unidades de análisis (Wooldridge, 2010). El uso de este enfoque permite evaluar la influencia de variables meteorológicas, contaminantes atmosféricos y factores sociodemográficos sobre la evolución de los contagios por COVID-19 con riqueza en la interpretación de los resultados y la validez interna del modelo econométrico propuesto (Baltagi, 2008; Hsiao, 2014).

Fuentes de datos

Datos Meteorológicos

En esta investigación se incorporaron variables meteorológicas como la precipitación (medida en milímetros) y la temperatura media mensual (en grados Celsius), con el propósito de analizar su posible relación con la propagación del COVID-19 en Ecuador. Estos datos fueron proporcionados directamente por el Instituto Nacional de Meteorología e Hidrología (INAMHI), tras una solicitud formal realizada mediante correo institucional. El INAMHI cuenta con una red de estaciones meteorológicas automáticas y convencionales distribuidas a nivel nacional, que permiten registrar variables climáticas con criterios de calidad establecidos según los estándares de la Organización Meteorológica Mundial (WMO, 2021). Gracias a dicha cobertura, los datos entregados garantizan representatividad espacial y consistencia temporal (FONAG y EPMAPS, 2021).

Para caracterizar la disponibilidad y distribución espacial de la información meteorológica utilizada en el estudio, se revisó la red de estaciones administrada por el Instituto Nacional de Meteorología e Hidrología (INAMHI). La Figura 2 presenta la distribución de las estaciones meteorológicas disponibles en Ecuador durante el período 2020–2021, así como su estado de transmisión al momento de la consulta.

Figura 2

Estaciones meteorológicas del INAMHI en Ecuador en 2020 – 2021



Nota. Distribución de las estaciones meteorológicas administradas por el INAMHI y su estado de transmisión. Adaptado de *Visor de estaciones meteorológicas*, por Instituto Nacional de Meteorología e Hidrología [INAMHI], [2022], [<https://inamhi.gob.ec/info/>].

Temperatura

Los registros de temperatura fueron proporcionados como promedios mensuales por estación, con un número variable de observaciones por mes (N. Obs). En las provincias de Guayas y Pichincha, que cuentan con más de una estación operativa, se calculó la temperatura media provincial mensual aplicando la siguiente expresión:

$$T_{m,p} = \frac{1}{n_p} \sum_{i=1}^{n_p} T_{m,i},$$

Ecuación 1 Promedio mensual TMP por estación

donde $T_{m,p}$ es la temperatura media del mes m en la provincia p ; $T_{m,i}$ la temperatura media mensual reportada por la estación i ; n_p el número de estaciones activas con datos válidos en esa provincia.

Precipitación

Los registros de precipitación fueron proporcionados en formato diario con coordenadas de latitud y longitud, como se establece en las recomendaciones de cálculo de la precipitación total diaria establecidas en la *Guide to Instruments and Methods of Observation* (WMO-No.8) (WMO, 2021); además, de acuerdo con los procedimientos técnicos descritos en los Anuarios Meteorológicos de FONAG y EPMAPS (2021) la precipitación mensual se calcula como la suma de las precipitaciones diarias registradas durante los días

válidos del mes, promediando los valores de las estaciones disponibles. Así, en este estudio se implementó dicho procedimiento en dos etapas:

1. Agregación diaria provincial: se obtuvo el promedio diario de precipitación considerando todas las estaciones activas en cada provincia

$$P_{d,p} = \frac{1}{n_{d,p}} \sum_{i=1}^{n_{d,p}} P_{d,i},$$

Ecuación 2 Promedio diario PRE por estación

donde $P_{d,i}$ corresponde a la precipitación diaria registrada por la estación i ; $n_{d,p}$ es el número de estaciones con datos válidos el día d .

2. Agregación mensual y control de cobertura: se calculó la precipitación mensual media promediando los valores diarios

$$P_{m,p} = \frac{1}{D_m} \sum_{d=1}^{D_m} P_{d,p},$$

Ecuación 3 Promedio mensual PRE por estación

donde D_m representa el número de días con datos válidos dentro del mes m . se conservaron los meses, pues tuvieron una cobertura del 100% de días con datos válidos.

Variables de Contaminación Atmosférica ($PM_{2.5}$)

Para representar la concentración de material particulado fino $PM_{2.5}$, se utilizaron datos de reanálisis satelital, los cuales fueron obtenidos de Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), desarrollados por “NASA Global Modeling and Assimilation Office” (GMAO). Este conjunto de datos proporciona estimaciones globales mensuales de aerosoles y gases traza desde el año 1980 en adelante, con resolución espacial de aproximadamente $0.5^\circ \times 0.625^\circ$ (Buchard et al., 2017). Los datos fueron descargados mediante el portal de acceso de la Goddard Earth Sciences Data and Information Services Center (GES DISC) y el interfaz interactivo Giovanni: facilitan la obtención de las variables necesarias en formato NetCDF.

Las variables se extrajeron del paquete M2TMNXAER, que brinda promedios mensuales como cuadrícula espacial 2D, de un solo nivel con resolución horizontal completa. Las variables seleccionadas para el cálculo de $PM_{2.5}$ fueron DUSMASS (Dust Surface Mass Concentration - $PM_{2.5}$), SO4SMASS (SO₄ Surface Mass Concentration), OCSMASS (Organic Carbon Surface Mass Concentration), BCSMASS (Black Carbon Surface Mass

Concentration) Y SSSMASS (Sea Salt Surface Mass Concentration) de acuerdo con la formulación de Buchard et al. (2017):

$$PM_{2.5} = DUSMASS25 + OCSMASS + BCSMASS + SSSMASS25 + SO4MASS * \frac{132.14}{96.06}$$

Ecuación 4 Formulación para cálculo de PM2.5

donde el factor de $\frac{132.14}{96.06}$ es recomendado por Buchard et al. (2017) para convertir el ion sulfato (SO_4^{2-}) a sulfato de amonio ($(NH_4)_2SO_4$), asumiendo una neutralización completa.

Posteriormente, los valores se convirtieron a microgramos por metro cúbico ($\mu g/m^3$) para facilitar la interpretación y comparación con estudios epidemiológicos previos. Con este procesamiento se generó una base de datos mensual - desde enero 2020 hasta diciembre 2021 - con las concentraciones estimadas de $PM_{2.5}$ ($\mu g/m^3$) para las tres provincias de estudio y fue empleada como variable ambiental de contaminación atmosférica en los modelos de regresión de panel planteados en el presente estudio.

Cambio de soporte espacial

Para delimitar espacialmente las unidades de análisis del presente estudio como son Azuay, Guayas y Pichincha se utilizó información geográfica vectorial en formato shapefile (.shp) para extraer información satelital específica. El archivo shapefile fue proporcionado formalmente por el Ministerio de Gobierno del Ecuador, en el marco del Decreto Ejecutivo N.º 1046 (9 de mayo de 2020) y conforme a lo dispuesto en el artículo 13, literal h), de la Ley para la Fijación de Límites Territoriales Internos, la cual faculta a dicha institución a mantener actualizada la división político-administrativa del país (Secretaría Técnica del Comité Nacional de Límites Internos, 2020).

El shapefile fue descargado desde el enlace compartido en el oficio oficial y contiene geometrías tipo MULTIPOLYGON representando a las provincias del Ecuador; en la comunicación oficial se especificó que la información se encontraba referida al sistema

- WGS 84 / UTM zona 17 Sur (EPSG:32717).

Por tal motivo, este sistema fue asignado manualmente en el entorno estadístico R mediante la función `st_crs()` del paquete `sf`. Posteriormente, las geometrías fueron transformadas al sistema geográfico

- WGS 84 (EPSG:4326).

Este proceso de conversión fue un requisito indispensable para poder usarlo en la extracción de información de variables de M2TMNXAER, las cuales corresponden a datos

satelitales en coordenadas de latitud y longitud. Una vez transformadas las geometrías, se definieron áreas rectangulares (bounding boxes) que abarcan completamente los límites geográficos de cada provincia. Estas cajas envolventes se construyeron a partir de las coordenadas mínimas y máximas de latitud y longitud extraídas de cada polígono provincial.

Por otro lado, el conjunto de datos de concentración de material particulado fino ($PM_{2.5}$) derivado del producto satelital MERRA-2 posee una resolución espacial de $0.5^\circ \times 0.625^\circ$, lo cual corresponde aproximadamente a celdas de entre 50 y 70 km de lado en el territorio ecuatoriano. Dado que el objetivo del estudio es obtener estimaciones comparables entre las provincias de Guayas, Pichincha y Azuay, fue necesario realizar un cambio de soporte espacial que permitiera transformar los valores de celdas regulares del modelo a unidades administrativas provinciales. Para ello, se desarrolló un procedimiento automatizado en R utilizando las librerías `ncdf4`, `terra`, `sf`, `geodata` y `exactextractr`.

En primer lugar, se procesaron los archivos NetCDF (.nc4) correspondientes a los meses del período 2020–2021. Cada archivo fue inspeccionado mediante funciones de lectura de metadatos para identificar las variables de interés incluidas en el módulo de aerosoles de MERRA-2: `DUSMASS25`, `SO4SMASS`, `OCSMASS`, `BCSMASS` y `SSSMASS`, que representan respectivamente las masas superficiales de polvo fino, sulfatos, carbono orgánico, carbono negro y sales marinas. Luego, las capas raster resultantes fueron recortadas espacialmente a los límites provinciales obtenidos desde la base GADM (Geodata) mediante la función `geodata::gadm()`. Estos límites fueron transformados al formato de geometrías espaciales (`sf`) y filtrados para conservar únicamente las provincias de Guayas, Pichincha y Azuay. El proceso de extracción provincial se realizó con la función `exact_extract()` del paquete `exactextractr`, la cual calcula estadísticas ponderadas por la fracción de superposición entre cada celda raster y el polígono administrativo correspondiente.

Factores Sociodemográficos

Las variables sociodemográficas consideradas en este estudio fueron: coeficiente de Gini, tasa de pobreza, proporción de adultos mayores y densidad poblacional. Estos factores son ampliamente reconocidos en la literatura como determinantes sociales de la salud, ya que influyen en la movilidad poblacional, el acceso a servicios sanitarios y la vulnerabilidad social frente a brotes epidémicos (Comunian et al., 2020; Houweling et al., 2024; Rocklöv y Sjödin, 2021). Los datos se obtuvieron del Instituto Nacional de Estadística y Censos (INEC) del Ecuador, principalmente de la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) correspondiente al año 2019, pues en 2020 no se pudo levantar información estadística compatible con la agregación de muestra que requiere la ENEMDU anual (INEC,

2024). Estos indicadores se complementaron con los boletines estadísticos provinciales y el portal oficial “Ecuador en cifras”, particularmente las proyecciones poblacionales que el INEC realiza anualmente para Ecuador: proyección de la población para el año 2020 por provincias, esto es, valores estimados para el año 2020, para así garantizar la coherencia temporal con las variables epidemiológicas y climáticas analizadas en el modelo a implementar, correspondientes al período marzo 2020 – marzo 2021.

Coeficiente de Gini

Valores extraídos de la ENEMDU 2019, el índice de Gini (G) expresa matemáticamente el nivel de concentración del ingreso; sus valores se sitúan entre 0, que simboliza distribución completamente equitativa, y 1, que denota la mayor desigualdad posible. El INEC calcula este indicador a partir de los microdatos de la ENEMDU, siguiendo la metodología del Banco Mundial, con base en la curva de Lorenz. Según la documentación de metodología de INEC (2019) su formulación es:

$$G = 1 - \sum_{i=1}^n (Y_i + Y_{i-1})(X_i - X_{i-1}),$$

Ecuación 5 Fórmula para Coeficiente de Gini

donde X_i representa la proporción acumulada de la población ordenada por ingreso, y Y_i la proporción acumulada del ingreso total correspondiente.

Tasa de pobreza

Los valores de la tasa de pobreza (P) también fueron extraídos de la ENEMDU 2019, y representa el porcentaje de personas cuyo ingreso per cápita es inferior al costo de la canasta básica familiar definida por el INEC. Según la metodología institucional hallada en la documentación de metodología, se calcula como (INEC, 2019):

$$P = \frac{N_p}{N_t} \times 100$$

Ecuación 6 Fórmula para Tasa de Pobreza

Proporción de adultos mayores

La proporción de adultos mayores (A) fue calculada en la presente investigación a partir de las proyecciones de población por edad simple del INEC (2020), aplicando la metodología oficial empleada por el Instituto para este indicador.

La variable se definió como el porcentaje de personas de 65 años o más respecto al total de la población provincial:

$$A = \frac{P_{65+}}{P_T} \times 100,$$

Ecuación 7 Fórmula para Proporción de Adultos Mayores

Donde P_{65+} representa la población de 65 años o más, y P_T la población total proyectada para el año 2020. Los cálculos se realizaron a partir del archivo “Tabulado_provincial_edad_simple_1990–2035_rev2024.csv” extraído de la página web oficial del INEC, este archivo formato CSV contiene las proyecciones intercensales por provincia y edad publicadas por el INEC de forma anual, por lo que se tomó el dato correspondiente al año 2020 para Azuay, Guayas y Pichincha.

Densidad poblacional

La densidad poblacional (D) también fue calculada netamente para esta investigación, utilizando los datos de dos archivos mencionados previamente: las proyecciones poblacionales provinciales del INEC (2020) y la superficie territorial oficial de cada provincia. El área total de cada provincia en estudio fue calculada a partir de los datos geográficos oficiales del GADM correspondiente a Ecuador. Los límites administrativos provinciales fueron descargados mediante la función `geodata::gadm()` en R y convertidos a formato simple features (sf) para su procesamiento. Posteriormente, se estimó el área de cada provincia en kilómetros cuadrados (km²) empleando las funciones del paquete sf: la función `st_area()` de la librería sf calcula el área de cada polígono provincial en metros cuadrados, y posteriormente se convirtió este valor a kilómetros con un cálculo aritmético sencillo. Los resultados fueron utilizados con las proyecciones poblacionales del INEC (2020) para el cálculo de la densidad poblacional, cuya formulación es (INEC, 2020):

$$D = \frac{P_T}{S},$$

Ecuación 8 Fórmula para Densidad Poblacional

Donde P_T es la población total proyectada para 2020, y S el área provincial en kilómetros cuadrados (km²).

Variables Epidemiológicas

Para evaluar la dinámica del COVID-19 en las provincias de interés —Azuay, Guayas y Pichincha— se utilizaron datos mensuales de casos confirmados y defunciones por COVID-19. Estos indicadores permiten medir respectivamente la incidencia y la letalidad de la enfermedad en cada territorio, constituyéndose en métricas esenciales para evaluar el comportamiento epidemiológico del virus SARS-CoV-2 en función de factores ambientales y

estructurales (Dong, Du, y Gardner, 2020). Los datos fueron extraídos del proyecto Ecuacovid (Robalino, 2020), el cual recopila información de los informes oficiales del Servicio Nacional de Gestión de Riesgos y Emergencias (SNGRE), el Ministerio de Salud Pública del Ecuador (MSP) y el Registro Civil. Esta fuente ha sido utilizada en diversos estudios académicos como son Castro et al. (2025) y Gutierrez et al. (2021) por su transparencia, trazabilidad y consistencia con los reportes oficiales.

Los datos se presentan en formato tabular con información por provincia y mes. Para este estudio se consideró el período comprendido entre marzo de 2020 y marzo de 2021, coincidente con la fase inicial de la pandemia en Ecuador. Las variables analizadas fueron:

- Contagios mensuales nuevos: corresponde al número de casos confirmados de COVID-19 reportados en cada mes para cada provincia

$$C_{i,t} = \sum_{d=1}^{n_t} c_{i,d}$$

- Defunciones mensuales nuevas: corresponde al número de fallecimientos por COVID-19 reportados en cada mes para cada provincia

$$M_{i,t} = \sum_{d=1}^{n_t} m_{i,d}$$

donde $c_{i,d}$ y $m_{i,d}$ son los casos de contagios y muertes diarias de la provincia i en el día d del mes t , respectivamente; n_t es el número de días reportados en el mes t .

Depuración, limpieza e integración de los conjuntos de datos

El proceso de depuración y limpieza de los datos se efectuó con el propósito de garantizar la calidad, consistencia y coherencia temporal entre las distintas fuentes que componen la base de datos principal del estudio, como se recomienda en un análisis multivariante con variables atmosféricas y sociodemográficas realizado por Peña Benitez y Bauz (2022). Cada conjunto fue sometido a un proceso de validación independiente antes de su integración en una estructura de panel provincial-mensual. A continuación, se detallan los procedimientos aplicados a cada tipo de variable.

Variables meteorológicas (temperatura y precipitación)

Inicialmente las bases de datos facilitadas por el INAMHI consistían en:

- Promedios mensuales de temperatura media (°C) desde enero 2020 hasta diciembre 2021 para las provincias de Guayas, Pichincha y Azuay por separado.

- Registros diarios de precipitación (mm) desde el 01 de enero 2020 hasta el 31 de diciembre 2021 para las provincias de Guayas, Pichincha y Azuay por separado.

El proceso de depuración inició con la revisión de la cobertura mensual reportada por cada estación. Siguiendo los criterios del Manual de Observación Meteorológica de la Organización Meteorológica Mundial (WMO-No.8, 2018), se consideró como dato confiable aquel cuya cobertura mensual fuera igual o superior al 75% de observaciones válidas. En el caso de temperatura y precipitación mantuvieron una cobertura del 100%, por lo que no se eliminaron ni interpolaron observaciones. Luego, se calcularon los promedios mensuales por provincia, considerando el número de estaciones disponibles. Finalmente se filtraron los valores mensuales solo en el período de interés: marzo 2020 – marzo 2021.

Variables de contaminación atmosférica ($PM_{2.5}$)

Los datos del archivo M2TMNXAER, en formato netCDF, extraídos mediante la plataforma Earthdata correspondían a promedios mensuales expresados en kilogramos por metro cúbico (kg/m^3) de aerosoles de superficie en formato NetCDF (.nc4), con resolución espacial de $0.5^\circ \times 0.625^\circ$, y abarcaron el período marzo 2020 – marzo 2021. Para su adecuación al presente análisis se realizó un proceso de extracción, depuración y limpieza en R, utilizando las librerías ncdf4, raster, terra, sf, exactextractr, dplyr, tidyr, lubridate y purrr. El flujo metodológico aplicado fue el siguiente:

1. Inspección de archivos NetCDF y selección de variables relevantes.

Mediante la función `inspect_nc_vars()`, se verificó la estructura interna de los archivos y se identificaron las variables de masa superficial reconocidas por la NASA como los principales componentes de $PM_{2.5}$ en la reanálisis MERRA-2:

- DUSMASS25 (polvo mineral $2.5 \mu m$)
- SO4SMASS (sulfatos)
- OCSMASS (carbono orgánico)
- BCSMASS (carbono negro)
- SSSMASS (sal marina).

2. Lectura y procesamiento de variables.

Para cada archivo mensual, se aplicó la función `extract_pm25_merra()`, que:

- Lee las cinco variables mencionadas mediante `terra::rast()`.
- Calcula el $PM_{2.5}$ total estimado como la suma ponderada de las masas superficiales, de acuerdo con la metodología de MERRA-2.

- Los cálculos se realizaron sobre grillas mensuales y se limitaron espacialmente a l territorio ecuatoriano.

3. Agregación espacial por provincia.

Se extrajeron los promedios espaciales mediante la función `exact_extract()` del paquete `exactextractr`. Esta función calcula medias ponderadas por la proporción de área de celda que interseca cada polígono provincial.

4. Conversión temporal y control de calidad.

Consistencia temporal (ausencia de saltos bruscos o picos anómalos) mediante inspección visual con `ggplot2`, confirmando que las variaciones intermensuales fueron coherentes con los patrones esperados de concentración de aerosoles en el Ecuador.

5. Cobertura y Validación Final.

No se identificaron valores faltantes ni inconsistencias en la serie mensual para el período marzo 2020 – marzo 2021, por lo que la cobertura final fue del 100%. Los datos fueron exportados en formato `.csv` bajo el nombre `PM25_2020_2021`, el cual contiene para cada provincia los promedios mensuales de $PM_{2.5}$ y de cada componente individual (polvo, sulfatos, carbono orgánico, carbono negro y sal marina).

Variables socioeconómicas y demográficas (INEC / ENEMDU)

De la ENEMDU se extrajeron indicadores provinciales de pobreza multidimensional, densidad poblacional, proporción de adultos mayores y tasa de urbanización, mientras que las proyecciones de población permitieron estimar la estructura demográfica. Dado que estos indicadores poseen periodicidad anual, se realizó una homogeneización temporal a través de repetición de valores para cada mes por provincia, con el fin de generar series mensuales consistentes para el período analizado.

Variables epidemiológicas (COVID-19)

Las variables epidemiológicas, correspondientes a casos confirmados y muertes por COVID-19 fueron descargadas en formato `.csv` del repositorio de datos abiertos de Robalino (2021): los registros existentes fueron filtrados para las provincias de Guayas, Pichincha y Azuay, y se agregaron de forma mensual, desde marzo 2020 hasta marzo 2021 a fin de mantener compatibilidad temporal con las demás variables. Se efectuó un control de calidad para detectar duplicados, inconsistencias entre casos y muertes (por ejemplo, meses con muertes sin casos reportados) y valores negativos originados por actualizaciones retroactivas. No existieron registros vacíos ni inconsistentes.

El proceso de depuración permitió homogeneizar las distintas fuentes de información y conservar únicamente los registros correspondientes al período y unidades espaciales definidos para el estudio. Como resultado, las bases meteorológicas, epidemiológicas, socioeconómicas y ambientales fueron transformadas a una estructura provincial y mensual comparable. La Tabla 3 resume los registros iniciales, los datos eliminados durante el proceso de limpieza y el número de observaciones finalmente incorporadas al análisis.

Tabla 3

Resumen del proceso de depuración y limpieza de datos.

Variable	Registros iniciales	Registros eliminados	Observaciones mensuales finales	Observación
Temperatura	72	33	39	72 registros mensuales, se filtró solo por fecha de interés: se eliminan los meses fuera del rango temporal en estudio.
Precipitación	929408	593600	39	929408 registros diarios, se eliminaron registros fuera del tiempo de estudio y promedió mensualmente.
Coefficiente de Gini, Tasa de pobreza, Densidad poblacional, Proporción de adultos mayores	3	0	39	Se repitieron los registros por provincia, en todos los meses del tiempo en estudio.

$PM_{2.5}$	13 515 840	0	39	13 515 840 registros mensuales por grillas: se filtraron solo las grillas de interés y se obtuvo un promedio para las grillas que componen cada provincia. Se aplicó la fórmula para obtener el $PM_{2.5}$ como una suma ponderada.
Contagio	72	33	39	Se eliminan los meses fuera del rango temporal en estudio.
Muerte	72	33	39	

Nota. Elaboración propia a partir de datos del Instituto Nacional de Meteorología e Hidrología (INAMHI), GES DISC–NASA, Robalino (2021) e Instituto Nacional de Estadística y Censos (INEC).

Integración de los conjuntos de datos

Una vez los conjuntos de datos individuales fueron depurados y limpiados, se integraron en una estructura de panel provincial-mensual, definida por las claves provincia y fecha, abarcando el período marzo 2020 – marzo 2021. La integración se realizó mediante uniones internas (inner joins) con coincidencia exacta entre fechas y provincias en todas las fuentes. Se verificó la consistencia de unidades de medida, la homogeneización de nombres de variables y la ausencia de duplicados. Como resultado se tuvo un panel balanceado que permitió analizar la evolución temporal y las diferencias espaciales de las variables meteorológicas, atmosféricas, epidemiológicas y socioeconómicas en las tres provincias seleccionadas.

Análisis descriptivo

Antes de proceder con la estimación del modelo econométrico, se llevó a cabo un análisis descriptivo detallado con el propósito de examinar el comportamiento de las variables consideradas y verificar su idoneidad para los análisis posteriores. En una primera etapa, se obtuvieron estadísticas básicas para las variables cuantitativas, incluyendo medidas de tendencia central (como la media), de dispersión (como la desviación estándar) y valores límite (mínimos y máximos). Este conjunto de indicadores se calculó para las variables

meteorológicas, ambientales y epidemiológicas: temperatura (TMP), precipitación (PRE) y concentración de material particulado fino (PM_{2.5}), número de contagios (CTG), número de muertes (MRT) a nivel provincial y también sobre la muestra agregada: total de 39 observaciones.

Se estimaron estadísticas descriptivas básicas para las variables incluidas en el estudio (CTG, MRT, DEN, PMA, GIN, POB, PRE, TMP y $PM_{2.5}$), a nivel provincial y también a nivel agregado. Además, se generaron visualizaciones exploratorias específicas como histogramas, diagramas de caja (boxplots) y series temporales provinciales. Los histogramas permitieron observar la forma de la distribución (asimetría, curtosis) y detectar valores atípicos tal como se defiende en Rongala (2024), los boxplots facilitaron la comparación entre provincias y la identificación clara de valores extremos (Shreffler y Huecker, 2023) y las series temporales mostraron la evolución mensual de cada variable en las tres provincias. Esta triangulación gráfica-numérica fortalece la comprensión inicial de los datos y puede llegar a ser conveniente incluso antes de proceder a realizar alguna inferencia, como indican Hehman y Xie (2021).

Pruebas de Normalidad

Dado que el conjunto de datos corresponde a un panel provincial-mensual (3 provincias $\times$ 13 meses = 39 observaciones), se aplicaron pruebas de normalidad por provincia y también sobre la muestra agregada. Se utilizó el test de Shapiro–Wilk complementado con gráficos Q–Q, histogramas y boxplots para cada provincia y variable. Este enfoque es coherente con buenas prácticas en exploración de datos (Ou et al., 2020), pues la prueba de Shapiro–Wilk se considera como uno de los métodos más robustos y sensibles para evaluar la normalidad de una variable continua en muestras pequeñas o moderadas ($n < 2000$), condición típica en paneles con pocos cortes transversales y periodos cortos (Razali y Wah, 2021; Singh et al., 2022).

El estadístico de Shapiro–Wilk se define como:

$$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2},$$

donde $x_{(i)}$ son los valores ordenados de menor a mayor, $\bar{x}$ es la media muestral, a_i son coeficientes calculados a partir de las covarianzas esperadas de los valores ordenados en una muestra normal y n es el tamaño de la muestra.

El estadístico W toma valores entre 0 y 1; mientras más próximo esté a 1, mayor es la evidencia de que la distribución de los datos se aproxima a la normal. La hipótesis nula y alternativa son:

H_0 : Los datos provienen de una distribución normal

H_1 : Los datos no provienen de una distribución normal

En aquellos casos en que la prueba indicó $p \leq 0.05$ se concluyó que la suposición de normalidad no se cumplía. En consecuencia, para el análisis de asociaciones se adoptó un criterio adaptativo: si ambas variables en una provincia resultaron normales y lineales, se aplicó Pearson; en los demás casos, Spearman (El-Hashash et al., 2022; Schober y Schwarte, 2018). En el análisis de la muestra agregada, dado el reducido tamaño de muestra y la naturaleza de las variables (ambientales, epidemiológicas), se adoptó Spearman como estrategia metodológica principal para garantizar robustez (Schober y Schwarte, 2018).

Correlaciones Entre Variables Independientes

Según Wooldridge (2021), en la fase exploratoria de un estudio econométrico o epidemiológico con datos de panel, la matriz de correlaciones es una herramienta descriptiva previa al análisis de regresión y debe centrarse en las variables independientes y covariables, se utiliza para:

- Detectar relaciones lineales entre variables explicativas.
- Identificar multicolinealidad que puede distorsionar los coeficientes en modelos de regresión múltiple.
- Si se plantea este análisis con la variable dependiente sirve solo para interpretación exploratoria, esto es, explorar asociaciones preliminares entre la variable dependiente y las independientes antes de ajustar el modelo.

Una vez verificada la distribución de las variables por provincia y globalmente, se procedió al análisis de correlaciones bivariadas. Se generaron todas las combinaciones posibles de pares de variables independientes PRE, TMP, $PM_{2.5}$ y por otro lado las covariables DEN, PMA, GIN, POB. Con el propósito de examinar la estructura relacional entre las variables explicativas, se elaboraron matrices de correlaciones tanto a nivel agregado (39 observaciones) como de forma desagregada por provincia (Azuay, Guayas y Pichincha), como respuesta a las recomendaciones metodológicas en estudios de panel ambiental–epidemiológico (Hehman et al., 2021; Wooldridge, 2021), que sugieren analizar la heterogeneidad espacial antes de la modelización formal.

Las correlaciones se estimaron bajo un esquema adaptativo: cuando ambas variables presentaron distribución normal según la prueba de Shapiro–Wilk, se utilizó el coeficiente de

Pearson; en caso contrario, se aplicó Spearman. Para cada par se reportó el tipo de método, el coeficiente de correlación y el valor p, lo cual permite identificar asociaciones monotónicas o lineales entre factores sociodemográficos, meteorológicos y de contaminación. Este procedimiento está respaldado en la literatura como buena práctica para estudios de panel con heterogeneidad espacial (Hehman et al., 2021). Las matrices resultantes se graficaron con ggplot mediante mapas de calor (“heatmaps”) y gráficos de dispersión, para poder observar la dirección y magnitud de las relaciones. Esta representación gráfica, complementada con histogramas y diagramas de caja, favorece la interpretación exploratoria y facilita la identificación de patrones espaciales o dependencias no lineales (Rongala, 2024; Shreffler y Huecker, 2023; Hehman y Xie, 2021).

Diagnóstico De Multicolinealidad

La presencia de multicolinealidad entre las variables explicativas puede distorsionar las estimaciones de los coeficientes en los modelos de regresión múltiple (Gujarati y Porter, 2020), por lo cual, se aplicó el factor de inflación de la varianza (VIF) a las variables independientes principales (TMP, PRE y $PM_{2.5}$) y, por separado, a las covariables estructurales (DEN, GIN, PMA, POB).

$$VIF_j = \frac{1}{1-R_j^2},$$

Ecuación 9 Fórmula para cálculo de VIF

donde R_j^2 es el coeficiente de determinación al regredir la variable X_j contra las demás.

El primer diagnóstico se realizó sobre las variables meteorológicas y de contaminación atmosférica mediante la función `vif()` de la librería `car` en R, con un modelo lineal base de la forma:

$$CTG_{it} = \beta_0 + \beta_1 TMP_{it} + \beta_2 PRE_{it} + \beta_3 PM_{2.5it} + \varepsilon_{it}$$

Ecuación 10 Modelo de regresión OLS para TMP, PRE y $PM_{2.5}$

Para diagnosticar la colinealidad entre las variables explicativas se siguió el criterio propuesto por Kutner et al. (2005), según el cual valores del VIF superiores a 5 son indicativos de colinealidad moderada y mayores a 10 de colinealidad severa. Este umbral ha sido ampliamente respaldado por la literatura reciente, por ejemplo, Jundt y Hu (2025) en *Frontiers in Marine Science* quienes señalan que un $VIF \geq 10$ representa una señal crítica de multicolinealidad. En concordancia con estos lineamientos, y dado que las pruebas de diagnóstico revelaron valores elevados del VIF, se procedió a aplicar un Análisis de Componentes Principales (PCA) sobre las variables con alta correlación, a fin de reducir la dimensionalidad y mitigar los efectos de la colinealidad en el modelo.

De manera similar se realizó otro diagnóstico de multicolinealidad sobre las covariables DEN, GIN, PMA y POB. En este caso se optó por eliminar las variables con mayor dependencia lineal y conservar aquellas que aportaban mayor estabilidad teórica y empírica al modelo. Esta decisión metodológica se fundamenta en las recomendaciones de Gujarati y Porter (2020) y Wooldridge (2021), quienes señalan que la colinealidad severa puede inflar las varianzas de los estimadores, distorsionar los signos de los coeficientes y reducir la capacidad explicativa del modelo. Asimismo, Jolliffe (2002) enfatiza que la eliminación de predictores redundantes suele ser preferible a la transformación mediante métodos multivariantes cuando las variables poseen un significado teórico bien definido.

En consecuencia, no se aplicó el análisis de componentes principales (PCA) a las covariables estructurales (DEN, GIN, PMA, POB), debido a que estas variables representan características socioeconómicas y demográficas fijas por provincia, las cuales tienen interpretaciones sustantivas directas —por ejemplo, la densidad poblacional o la desigualdad económica— que se perderían al transformarlas en componentes abstractos. Según Hair et al. (2022) y Baltagi (2021), la aplicación de PCA es apropiada cuando las variables son altamente correlacionadas y representan dimensiones de un mismo fenómeno continuo (como temperatura y contaminación), pero no cuando constituyen indicadores estructurales que deben conservar su significado físico o social dentro del modelo.

Análisis de Componentes Principales.

Con el fin de resolver los problemas de multicolinealidad detectados entre las variables meteorológicas y de contaminación (temperatura, precipitación y concentración de $PM_{2.5}$), se aplicó un Análisis de Componentes Principales (PCA). Esta técnica pertenece a los métodos multivariantes de reducción de dimensionalidad y permite transformar un conjunto de variables correlacionadas en un número menor de componentes linealmente independientes (Jolliffe y Cadima, 2016). Su aplicación es ampliamente recomendada en estudios ambientales y epidemiológicos cuando los predictores muestran alta interdependencia (Hair et al., 2022; Ganegoda et al., 2021).

Procedimiento computacional

El análisis computacional fue realizado en el entorno R, utilizando la función `prcomp()` del paquete `base stats`. Esta función se encarga de ejecutar internamente una descomposición en valores singulares (SVD por sus siglas en inglés) de la matriz de datos estandarizada, lo que equivale a diagonalizar la matriz de covarianzas o correlaciones entre las variables. De esta manera se obtiene una nueva base ortogonal de variables no correlacionadas que reciben el nombre de “componentes principales” (PCs). Cada componente se interpreta como una combinación lineal de las variables originales, donde los autovalores indican la proporción de varianza explicada y los autovectores las cargas de cada variable en el componente. Antes de aplicar el PCA, se realizó una normalización de las variables con el argumento `scale (TRUE)` en la propia función `prcomp()`. Esto se realizó para ganatizar que todas las magnitudes (temperatura en °C, precipitación en mm, $PM_{2.5}$ en $\mu g/m^3$) tuvieran igual peso en la descomposición, tal como recomiendan Jolliffe y Cadima (2016) y Abdi y Williams (2010).

Selección e incorporación de componentes

De acuerdo con la literatura, la selección de componentes se basa en el criterio de Kaiser–Guttman (autovalores > 1) y en el porcentaje acumulado de varianza explicada (Jolliffe, 2002). En este caso, se conservaron los dos primeros componentes (PC_1 y PC_2). Estos dos componentes fueron añadidos a la base de datos como nuevas variables (PC_1 , PC_2) para ser utilizadas posteriormente como predictores en los modelos de regresión de panel. Este procedimiento se justifica porque permite mantener la información esencial de los factores meteorológicos y de contaminación, pero se evita la redundancia lineal entre ellos. En términos econométricos, los componentes principales no están correlacionados entre sí, por lo que mejoran la estabilidad numérica y la precisión de los estimadores (Gujarati y Porter, 2020; Wooldridge, 2021).

El uso de PCA previo al modelado es una práctica ampliamente respaldada en estudios que integran variables ambientales y epidemiológicas de alta correlación, pues garantiza robustez estadística del modelo final (Ganegoda et al., 2021; WHO, 2021). Además, permite mantener la interpretación física del fenómeno, ya que los componentes representan combinaciones ponderadas de temperatura, precipitación y contaminación que reflejan condiciones atmosféricas conjuntas, más que variables aisladas. En síntesis, el PCA se aplicó como una etapa de preprocesamiento metodológico orientado a depurar la estructura interna de las variables explicativas antes de su incorporación en los modelos de datos de panel.

Modelado Estadístico Con Datos De Panel.

Para estimar la dinámica provincial-mensual del número de contagios (CTG) y número de muertes por COVID-19 (MRT) en las tres provincias analizadas (Azuay, Guayas y Pichincha) durante el período marzo 2020–marzo 2021, se emplearon modelos de datos de panel balanceado ($N = 3$, $T = 13$) mediante la función `plm()` del paquete `plm` en R. Se evaluaron tres especificaciones alternativas:

Modelo agrupado – Pooled OLS

El modelo pooled trata todas las observaciones como pertenecientes a una sola entidad agregada y estima un único intercepto asumiendo homogeneidad entre unidades. Se estimó matemáticamente mediante:

$$CTG_{it} = \beta_{0C} + \beta_{1C}PC1_{it} + \beta_{2C}PC2_{it} + \beta_{3C}GIN_i + \beta_{4C}DEN_i + \varepsilon_{it},$$

Ecuación 11 Modelo Pooled Para Contagios de COVID-19

$$MRT_{it} = \beta_{0M} + \beta_{1M}PC1_{it} + \beta_{2M}PC2_{it} + \beta_{3M}GIN_i + \beta_{4M}DEN_i + \varepsilon_{it},$$

Ecuación 12 Modelo Pooled Para Muertes por COVID-19

donde i indica provincia y t el mes. En este modelo se asume que no existen efectos específicos de la provincia que incidan sistemáticamente sobre los contagios, es decir, $\alpha_i = 0$ para todas las provincias i .

Dado el reducido número de unidades espaciales (3 provincias) y la ausencia de variación temporal en algunas covariables estructurales, el modelo pooled sirve como referencia de parsimonia antes de introducir heterogeneidad espacial. Estudios recientes en epidemiología de panel han recomendado su evaluación cuando la heterogeneidad cruzada es limitada (Smith et al., 2020; Nguyen et al., 2021). Así mismo, se aplicaron errores estándar robustos tipo HC3 para controlar heterocedasticidad (Long y Ervin, 2000).

Modelo De Efectos Fijos

Este modelo controla la heterogeneidad no observable y constante en el tiempo entre provincias, que podría sesgar los coeficientes si se omitiera. La transformación “within” descuenta la media de cada provincia, elimina α_i y estima solo el efecto temporal de los predictores (Baltagi, 2021). La ecuación que describe el planteamiento de este modelo es:

$$CTG_{it} = \alpha_{iC} + \beta_{1C}PC1_{it} + \beta_{2C}PC2_{it} + \varepsilon_{it},$$

Ecuación 13 Modelo FE Para Contagios de COVID-19

$$MRT_{it} = \alpha_{iM} + \beta_{1M}PC1_{it} + \beta_{2M}PC2_{it} + \varepsilon_{it},$$

Ecuación 14 Modelo FE Para Muertes por COVID-19

en este modelo se omiten las variables que no varían en el tiempo, pues por la propia transformación “within” que resta la media provincial: como las covariables DEN y GIN son constantes en el tiempo se eliminan en dicha transformación.

Cuando se espera que factores estructurales no observados (infraestructura de salud, políticas locales, clima base) puedan correlacionarse con los regresores, el modelo FE proporciona estimadores consistentes (Wooldridge, 2021). En estudios panel recientes de enfermedades infecciosas se recomienda FE cuando las unidades geográficas tienen características fijas que pueden afectar la propagación (Lee et al., 2022).

Modelo de Efectos Aleatorios

Este modelo incorpora la variabilidad entre provincias como parte del término de error, mientras asume que el término del efecto aleatorio provincial u_i no está correlacionado con los predictores; supone que los efectos individuales son aleatorios e independientes de los regresores y se encarga de capturar la variabilidad entre provincias a través de un

componente de error compuesto (Baltagi, 2021). Su modelo se expresa matemáticamente con la siguiente expresión:

$$CTG_{it} = \beta_{0C} + \beta_{1C}PC1_{it} + \beta_{2C}PC2_{it} + \beta_{3C}GIN_i + \beta_{4C}DEN_i + u_i + \varepsilon_{it},$$

Ecuación 15 Modelo RE Para Contagios de COVID-19

$$MRT_{it} = \beta_{0M} + \beta_{1M}PC1_{it} + \beta_{2M}PC2_{it} + \beta_{3M}GIN_i + \beta_{4M}DEN_i + u_i + \varepsilon_{it},$$

Ecuación 16 Modelo RE Para Muertes por COVID-19

El modelo RE permite estimar efectos de variables que no varían en el tiempo (como GIN o PMA) y es más eficiente si el supuesto de independencia del efecto individual se cumple (Greene, 2020). En trabajos recientes sobre panel de contaminación y salud se ha usado efectos aleatorios RE con precaución en los casos en que el número de unidades espaciales sea reducido (Martínez et al., 2023).

Selección Del Modelo Estadístico.

Una vez estimadas las tres especificaciones del modelo de datos de panel —agrupado (Pooled OLS), efectos fijos (FE) y efectos aleatorios (RE)—, se aplicaron contrastes secuenciales para determinar empíricamente la estructura más adecuada según la naturaleza de los datos y los supuestos de cada modelo.

a) Prueba F de significancia de efectos individuales

El primer contraste evaluó si existen diferencias sistemáticas entre las unidades de corte transversal (provincias) que justifiquen un modelo de efectos fijos en lugar del modelo agrupado. Este test, propuesto por Baltagi (2021), contrasta la hipótesis nula:

$$H_0: \alpha_i = \alpha, \forall i$$

$$H_1: \alpha_i \neq \alpha \text{ para algún } i$$

El estadístico F compara la variabilidad entre provincias con la variabilidad dentro de cada provincia. En la práctica, se aplicó este test mediante la función `pFtest()` del paquete `plm` (Croissant y Millo, 2008). Un valor p inferior al nivel de significancia del 0.05 indica que los interceptos provinciales difieren significativamente, justificando el uso del modelo de efectos fijos.

b) Prueba de Hausman

Para decidir entre efectos fijos y efectos aleatorios, se aplicó la prueba de Hausman (Hausman, 1978), la cual compara dos estimadores de los mismos parámetros:

β_{RE} es eficiente bajo la hipótesis de exogeneidad de los efectos aleatorios, mientras que β_{FE} es consistente aun cuando exista correlación entre los efectos individuales y los regresores. El estadístico de prueba sigue una distribución Chi Cuadrado χ^2_k con k grados de libertad, calculado como:

$$H = (\hat{\beta}_{FE} - \hat{\beta}_{RE})' [Var(\hat{\beta}_{FE}) - Var(\hat{\beta}_{RE})]^{-1} (\hat{\beta}_{FE} - \hat{\beta}_{RE})$$

Ecuación 17 Definición Estadístico de Prueba H (Hausman)

En el entorno de R, se usó la función `phtest()`, que calcula automáticamente esta diferencia y su matriz de varianzas conjunta, proporcionando el valor p correspondiente. Si el resultado es $p < 0.05$, se rechaza la hipótesis de exogeneidad y se concluye que el modelo de efectos fijos es más apropiado. En caso contrario, se prefiere el modelo de efectos aleatorios por su mayor eficiencia.

c) Prueba de Breusch–Pagan Lagrange Multiplier

Se evaluó si los efectos aleatorios son significativos respecto al modelo agrupado. La prueba Lagrange Multiplier LM de Breusch y Pagan (1980), implementada con `plmtest()` contrasta las siguientes hipótesis:

$$H_0: \sigma_u^2 = 0$$

$$H_1: \sigma_u^2 > 0$$

En síntesis, la hipótesis nula indica que el modelo de datos agrupado es el adecuado para describir el comportamiento de los datos, mientras que la hipótesis alternativa indica la posible presencia de efectos aleatorios para explicar el comportamiento de los datos. Este contraste es apropiado para muestras pequeñas ($N < 10$) y tiempos cortos ($T \leq 15$), ya que no requiere normalidad estricta (Baltagi, 2021; Greene, 2020). Un valor $p < 0.05$ sugiere que el modelo de efectos aleatorios explica mejor la variabilidad total que el modelo agrupado.

Validación del modelo.

Una vez seleccionada la estructura del modelo de panel más adecuada mediante las pruebas F, LM y Hausman, fue necesario verificar los supuestos clásicos del modelo lineal: homocedasticidad, independencia temporal, normalidad de los residuos, ausencia de multicolinealidad e independencia transversal. La validación de supuestos constituye una etapa fundamental en la modelación con datos de panel, ya que su incumplimiento puede generar sesgos en los coeficientes estimados, inflar los errores estándar o conducir a conclusiones no robustas (Baltagi, 2021; Wooldridge, 2021), es decir, cuando alguno de estos

se viola, los errores estándar convencionales pueden resultar sesgados, lo que afecta la fiabilidad de los intervalos de confianza y los contrastes de hipótesis.

a) Homocedasticidad de los errores

El supuesto de homocedasticidad requiere que la varianza del término de error sea constante entre provincias y a lo largo del tiempo:

$$Var(\varepsilon_{it}) = \sigma^2 \quad \forall i, t$$

La presencia de heterocedasticidad ($Var(\varepsilon_{it}) \neq \sigma^2$) puede producir estimadores insesgados pero ineficientes, lo que afecta directamente la validez de los tests t y F (Greene, 2020). Para su diagnóstico se aplicó la Prueba de Breusch–Pagan (bptest) con corrección de Student, adaptada al contexto de panel y la Prueba de Wald modificada para heterocedasticidad grupal. En caso de detectarse varianza no constante, se emplearon errores estándar robustos tipo White (HC3) o Driscoll–Kraay, estos últimos recomendados para paneles con N pequeño y T moderado, pues corrigen simultáneamente heterocedasticidad y dependencia temporal (Hoechle, 2007; Baltagi, 2021).

b) Autocorrelación serial de los errores

La autocorrelación implica dependencia entre los errores de distintos periodos para la misma unidad:

$$Cov(\varepsilon_{it}, \varepsilon_{i,t-1}) \neq 0$$

La presencia de esta correlación vulnera el supuesto de independencia y puede subestimar la varianza de los coeficientes. Se aplicaron las pruebas de Durbin–Watson para paneles (pdwtest) y la Prueba de Wooldridge (pwartest), que contrasta la hipótesis nula de ausencia de autocorrelación de primer orden. En los casos en los que se detectó correlación serial, los errores se estimaron mediante el estimador robusto Driscoll–Kraay, que además es consistente ante dependencia transversal (Driscoll y Kraay, 1998; Hoechle, 2007).

c) Normalidad de los residuos

Aunque la normalidad de los errores no es un requisito para la consistencia del estimador de mínimos cuadrados, sí resulta importante en muestras pequeñas para la validez de los intervalos de confianza y las pruebas t. Es así como el diagnóstico se realizó mediante la Prueba de normalidad de Shapiro – Wilk (adecuada para muestras de tamaño reducido) y como complemento se realizaron Gráficos Q–Q (Quantile–Quantile) de los residuos.

Se plantea la hipótesis nula H_0 de que los residuos se distribuyen normalmente. En caso de rechazar la hipótesis nula, se evaluó la posibilidad de aplicar transformaciones logarítmicas a la variable dependiente o robustecer los errores estándar. Este tipo de

comprobaciones se ha documentado como una buena práctica en modelos epidemiológicos de corto panel (Millo y Piras, 2012; Liu et al., 2021).

d) Ausencia de multicolinealidad

Dado que la inspección del Factor de Inflación de la Varianza (VIF) al plantear un modelo global arrojó valores elevados (mayores a 10) se evidenció una colinealidad estructural entre variables meteorológicas (PC1, PC2) y socioeconómicas (DEN, GIN), por lo que se optó por una estrategia de estimación por bloques de variables. Conforme a la literatura sobre análisis de regresión por bloques que permite evaluar contribuciones incrementales de conjuntos de predictores diferenciados (Sage Research Methods) y al enfoque que propone agrupar predictores altamente correlacionados en “bloques” para mejorar la estabilidad estimativa (Tsao, 2018), se establecieron dos modelos separados: uno meteorológico y otro socioeconómico. Esta separación facilita la interpretación clara de cada dominio explicativo, reduce la distorsión de coeficientes debida a colinealidad (Johnston, 2018) y mantiene la coherencia metodológica en el contexto panel.

Por tanto, los modelos finales (bloques meteorológico y socioeconómico) se estimaron bajo la especificación Pooled OLS con errores estándar robustos Driscoll–Kraay tipo HC1, garantizando inferencias consistentes incluso ante posibles correlaciones residuales entre predictores o unidades espaciales. Esta estrategia metodológica permitió asegurar la ausencia de multicolinealidad significativa dentro de cada bloque y mantener la validez estadística de los coeficientes estimados. Los modelos fueron así estimados mediante las siguientes expresiones, sea $i = 1, \dots, N$ la provincia y $t = 1, \dots, T$ el mes.:

Modelos para Contagios de COVID-19

$$CTG_{it} = \alpha^{(M)} + \beta_1 PC1_{it} + \beta_2 PC2_{it} + \varepsilon_{it}^{(M)}$$

Ecuación 18 *Bloque meteorológico (Pooled OLS)*

$$CTG_{it} = \alpha^{(S)} + \gamma_1 DEN_i + \gamma_2 GIN_i + \varepsilon_{it}^{(S)}$$

Ecuación 19 *Bloque socioeconómico (Pooled OLS)*

Modelos para Muertes por COVID-19

$$MRT_{it} = \delta^{(M)} + \lambda_1 PC1_{it} + \lambda_2 PC2_{it} + \zeta_{it}^{(M)}$$

Ecuación 20 *Bloque meteorológico (Pooled OLS)*

$$MRT_{it} = \delta^{(S)} + \eta_1 DEN_i + \eta_2 GIN_i + \zeta_{it}^{(S)}$$

Ecuación 21 *Bloque socioeconómico (Pooled OLS)*

e) Independencia transversal

En modelos de panel, se asume que los errores de una unidad de corte transversal (una provincia) son independientes de los errores de otra. Sin embargo, en estudios ambientales o epidemiológicos es frecuente que existan efectos comunes no observados — como políticas nacionales, shocks climáticos o movilidad interprovincial— que generen dependencia transversal entre unidades. Si este supuesto se viola, los errores estándar convencionales resultan inconsistentes (Pesaran, 2004; Sarafidis y Wansbeek, 2012).

Para diagnosticar este tipo de dependencia se aplicó la prueba de Pesaran (2004), basada en el promedio de las correlaciones de los residuos entre unidades. Su estadístico se define como:

$$CD = \sqrt{\frac{-2T}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \hat{\rho}_{ij},$$

donde $\hat{\rho}_{ij}$ es la correlación de los residuos estimados entre las unidades i y j , N el número de entidades independientes que integran la estructura transversal del conjunto de datos y T el número de periodos.

Bajo la hipótesis nula $H_0: Cov(\varepsilon_{it}, \varepsilon_{jt}) = 0 \forall i \neq j$, los errores son independientes entre provincias. Si el valor $p < 0.05$, se rechaza la independencia y se concluye la presencia de dependencia transversal; la literatura recomienda mantener el modelo estructural estimado, pero ajustar la matriz de varianzas mediante errores estándar Driscoll–Kraay, que son robustos ante este tipo de correlación cruzada (Wang et al., 2024).

Ajuste Funcional Del Modelo de datos de panel.

La evaluación de la especificación funcional tiene como objetivo verificar si la forma funcional seleccionada (relación lineal entre las variables explicativas y la variable dependiente) es adecuada para describir los datos observados. Como reporta Wooldridge (2019) y Gujarati y Porter (2020), una forma funcional incorrecta puede generar sesgo por omisión o inclusión incorrecta de términos, afectando la consistencia y eficiencia de los estimadores. En modelos de datos de panel, una práctica común consiste en aplicar la prueba RESET (Regression Equation Specification Error Test) de Ramsey (1969), adaptando el modelo a una versión equivalente en regresión lineal para detectar posibles errores de especificación. El test evalúa si términos no lineales o combinaciones polinómicas de los predictores explican significativamente la variable dependiente, lo que indicaría una forma funcional incorrecta.

Este Test plantea un modelo extendido que incluye potencias de los valores ajustados $\hat{y}_{it}$, en el cual los parámetros $\lambda_1, \lambda_2, \dots, \lambda_p$ son los coeficientes de los términos adicionales que

recogen posibles no linealidades o errores de especificación. Si el modelo lineal inicial fue correctamente especificado, los valores ajustados $\hat{y}_{it}$ no contienen información adicional más allá de las combinaciones lineales de los regresores y en ese caso, los coeficientes de los términos no lineales ($\hat{y}_{it}^2, \hat{y}_{it}^3, \dots$) deberían ser estadísticamente nulos.

$$H_0: \lambda_1 = \lambda_2 = \dots = 0$$

El test RESET se implementa como un contraste F o χ^2 , comparando el modelo restringido (sin los términos λ), contra el modelo no restringido (con los términos $\hat{y}^2, \hat{y}^3$).

$$F = \frac{(SSR_R - SSR_U)/m}{SSR_U/(n-k-m-1)},$$

donde SSR_R es la suma de residuos del modelo restringido (original), SSR_U la suma de residuos del modelo extendido, m el número de restricciones (número de λ), n el tamaño muestral y k el número de regresores originales.

Si F resulta estadísticamente significativo, se rechaza H_0 y se concluye que el modelo necesita un ajuste funcional.

3.5. Consideraciones Bioéticas

La presente investigación se realizó exclusivamente mediante el uso de fuentes secundarias de datos, obtenidas de bases públicas y oficiales, como el Ministerio de Salud Pública del Ecuador (MSP), el Instituto Nacional de Estadística y Censos (INEC), y productos satelitales de acceso libre. No se trabajó con muestras biológicas, información genética, ni se interactuó directamente con seres humanos, comunidades o su conocimiento tradicional.

Por esta razón, no fue necesario obtener consentimiento informado, ya que los datos empleados no permiten identificar directa ni indirectamente a individuos, y su uso cumple con las condiciones de anonimato, acceso abierto y propósito académico. No se vulnera la privacidad, integridad ni autonomía de ningún grupo humano.

Sin embargo, se han tenido en cuenta los principios bioéticos que rigen toda investigación:

- Beneficencia: Se procura contribuir al conocimiento científico con aplicaciones potenciales en salud pública, orientadas a prevenir enfermedades respiratorias como el COVID-19.
- Precaución y responsabilidad: Se han tomado precauciones al interpretar los datos ambientales y epidemiológicos, señalando sus limitaciones y evitando extrapolaciones indebidas.

- Justicia: Se ha garantizado el acceso equitativo a la información utilizada, trabajando con bases públicas y transparentes.
- Autonomía: No se afecta la libertad ni autodeterminación de personas o comunidades, ya que no se recopilaron datos personales ni se realizaron intervenciones.

En cumplimiento con los lineamientos éticos institucionales y nacionales, esta investigación respeta los estándares bioéticos vigentes, al centrarse exclusivamente en análisis de datos secundarios anonimizados y de uso público.

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

Resultados

Fase 1: Caracterización de la evolución de variables

4.1. Estadística Descriptiva.

4.1.1. Resumen De Las Variables Cuantitativas

Los indicadores estadísticos que caracterizan a las variables meteorológicas y ambientales consideradas en el estudio —temperatura media (TMP), precipitación mensual (PRE) y concentración de material particulado fino ($PM_{2.5}$)—, a nivel provincial se encuentran organizados en la Tabla 4 y en la Tabla 5 a nivel agregado para las provincias de Azuay, Guayas y Pichincha durante el período marzo 2020 – marzo 2021. Estos indicadores permiten identificar la variabilidad climática y atmosférica de cada región, así como su posible relación con la dinámica de contagios y muertes por COVID-19.

Se evidencia diferencias climáticas consistentes con la ubicación geográfica de cada provincia. Guayas, perteneciente a la región Costa, registra la temperatura media más alta ($25.42\text{ }^{\circ}\text{C}$) y también la mayor dispersión ($DE = 0.94\text{ }^{\circ}\text{C}$), evidenciando condiciones térmicas propias de un clima tropical húmedo. En contraste, Azuay y Pichincha, situadas en la región interandina, presentan temperaturas medias de $13.48\text{ }^{\circ}\text{C}$ y $13.95\text{ }^{\circ}\text{C}$, respectivamente, con menor variabilidad interna ($DE \approx 0.58\text{ }^{\circ}\text{C}$), lo cual refleja la estabilidad térmica asociada a su altitud promedio superior a los 2 000 msnm.

Los indicadores estadísticos utilizados para caracterizar las variables meteorológicas, ambientales y epidemiológicas consideradas en el estudio fueron la media, el mínimo, el máximo y la desviación estándar. La Tabla 4 presenta los estadísticos descriptivos de temperatura media (TMP), precipitación mensual (PRE), concentración de material particulado fino ($PM_{2.5}$), contagios (CTG) y muertes (MRT) para las provincias de Azuay, Guayas y Pichincha durante el período comprendido entre marzo de 2020 y marzo de 2021. Estos resultados permiten comparar la variabilidad climática, ambiental y epidemiológica entre las tres provincias analizadas.

Tabla 4*Resumen de las variables cuantitativas por provincias marzo 2020 – marzo 2021*

Variable	Mínimo	Máximo	Media	Desviación Estándar
Provincia de Azuay				
TMP (°C)	12.1	14.40	13.47692	0.5804508
PRE (mm)	1.091737	4.728433	2.31297	1.24038
$PM_{2.5}$ ($\mu g/m^3$)	6.99576	13.5094	10.41174	1.569316
CTG	78	5666	1950.07	1645.443
MRT	2	45	26.00	14.4164
Provincia de Guayas				
TMP (°C)	24.2	26.85	25.42308	0.941765
PRE (mm)	0.1456892	13.393952	3.360801	4.19645
$PM_{2.5}$ ($\mu g/m^3$)	16.349987	33.47219	25.561956	4.91086
CTG	951	44017	6397.308	11645.23
MRT	23	1025	165.00	271.209
Provincia de Pichincha				
TMP (°C)	12.9	14.75	13.95385	0.5695252
PRE (mm)	1.4832992	14.483845	5.733334	4.03450
$PM_{2.5}$ ($\mu g/m^3$)	6.075967	13.1431	8.520424	2.45071
CTG	219	50787	14765.077	16501.32
MRT	5	522	191.846	135.44

Nota. TMP = temperatura media (°C); PRE = precipitación mensual (mm); PM_{2.5} = concentración de material particulado fino (µg/m³); CTG = número de contagios por COVID-19; MRT = número de muertes por COVID-19. Cálculos propios a partir de las bases de datos utilizadas en el estudio.

Respecto a la concentración promedio de PM_{2.5}, Guayas presenta el valor más elevado (25.56 µg/m³), seguido por Azuay (10.41 µg/m³) y Pichincha (8.52 µg/m³). Este patrón evidencia una mayor carga de contaminantes atmosféricos en la región costera, influenciada por la actividad industrial y vehicular de Guayaquil (Buchard et al., 2017; Castro et al., 2025). Las provincias serranas, por su altitud y condiciones de ventilación, mantienen concentraciones más bajas, lo que sugiere la presencia de fuentes locales de emisión y transporte regional de aerosoles.

Con el propósito de complementar la descripción provincial, se calcularon también los estadísticos descriptivos considerando de manera conjunta las observaciones de Azuay, Guayas y Pichincha. La Tabla 5 presenta el resumen agregado de las variables meteorológicas, ambientales y epidemiológicas para el período comprendido entre marzo de 2020 y marzo de 2021. Esta síntesis permite identificar el rango global de variación y la dispersión de las variables consideradas antes de examinar sus asociaciones.

Tabla 5

Resumen de las variables cuantitativas a nivel agregado marzo 2020 – marzo 2021

Variable	Mínimo	Máximo	Media	Desviación Estándar
TMP (°C)	12.1	26.85	17.61795	5.638209
PRE (mm)	0.1456892	14.48385	3.802368	3.645318
PM _{2.5} (µg/m³)	6.075967	33.4721	14.83137	8.365987
CTG	78	50787	7704.154	12595.02
MRT	2	1025	127.6154	185.7659

Nota. Los estadísticos corresponden al conjunto de observaciones de Azuay, Guayas y Pichincha. TMP = temperatura media (°C); PRE = precipitación mensual (mm); PM_{2.5} = concentración de material particulado fino (µg/m³); CTG = número de contagios por COVID-

19; MRT = número de muertes por COVID-19. Cálculos propios a partir de datos del INAMHI, MSP y GES DISC–NASA.

Los resultados agregados evidencian un comportamiento coherente con la diversidad geográfica del Ecuador, en particular los indicadores epidemiológicos CTG y MRT agregados presentan desviaciones estándar altas (12595 y 185, respectivamente), lo que confirma la fuerte dispersión espacial y temporal del impacto del COVID-19 en Ecuador. Estas magnitudes reflejan la concentración de casos en provincias densamente pobladas y los picos de contagios registrados durante las olas epidémicas de 2020. El resumen nacional aporta evidencia empírica que sustenta la necesidad de un enfoque multivariado y longitudinal: los elevados niveles de dispersión observados entre unidades espaciales y temporales confirman la heterogeneidad estructural del fenómeno, una condición esencial para justificar la estimación de modelos de panel.

4.1.2. Resumen De Las Covariables Cuantitativas

En la Tabla 6 se presentan los valores de las covariables estructurales y sociodemográficas consideradas en el análisis: coeficiente de Gini (GIN), tasa de pobreza multidimensional (POB), proporción de adultos mayores (PMA) y densidad poblacional (DEN). Estas variables, de naturaleza anual, fueron integradas a la estructura de datos de panel mediante replicación mensual, con el fin de mantener coherencia temporal con las variables meteorológicas y epidemiológicas analizadas en el modelo.

Tabla 6

Resumen de las covariables cuantitativas proyecciones año 2020

Variable	Azuay	Guayas	Pichincha
Coeficiente de Gini	0.433	0.418	0.473
Tasa de pobreza multidimensional (%)	0.247	0.332	0.113
Proporción de adultos mayores (%)	0.0917453	0.0679385	0.0746151
Densidad poblacional (hab/km ²)	100.6120	287.2589	337.7009

En conjunto, las covariables analizadas evidencian una heterogeneidad estructural que condiciona las diferencias observadas en los niveles de contagios y mortalidad por COVID-19 entre provincias. Las disparidades en desigualdad, pobreza, envejecimiento y densidad poblacional configuran contextos de riesgo diferenciados que se reflejan posteriormente en los modelos de panel.

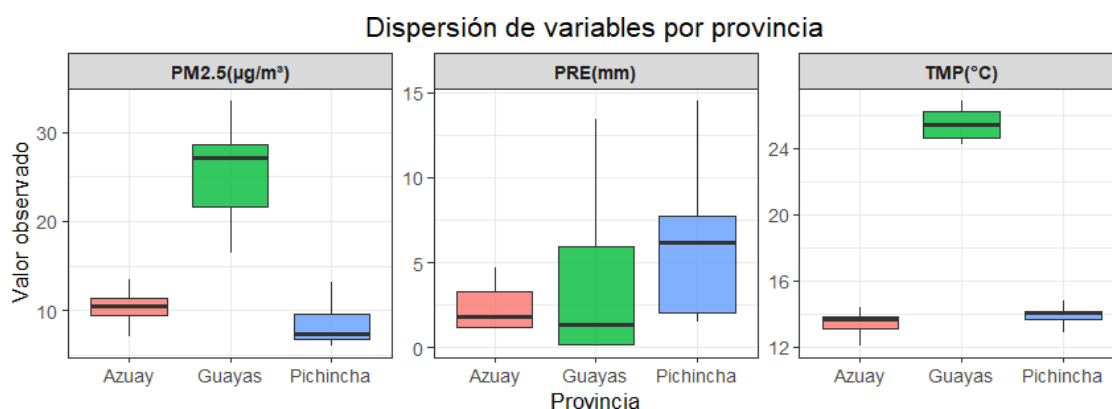
4.1.3. Visualización de la dispersión de las variables

Variables Independientes

Con el objetivo de complementar el análisis descriptivo, se elaboraron diagramas de caja para representar la distribución y variabilidad de las variables meteorológicas y ambientales en Azuay, Guayas y Pichincha. La Figura 3 presenta la distribución provincial de la temperatura media (TMP), la precipitación mensual (PRE) y la concentración de material particulado fino ($PM_{2.5}$) durante el período comprendido entre marzo de 2020 y marzo de 2021. Este tipo de representación permite identificar diferencias en la mediana, amplitud intercuartílica, dispersión y posibles valores atípicos entre provincias.

Figura 3

Diagrama de caja de variables independientes por provincia marzo 2020 - marzo 2021



Nota. TMP = temperatura media (°C); PRE = precipitación mensual (mm); $PM_{2.5}$ = concentración de material particulado fino ($\mu g/m^3$). Elaboración propia a partir de los datos utilizados en el estudio.

La temperatura muestra una separación clara entre provincias, este comportamiento evidencia una alta variabilidad interprovincial, así como baja variabilidad intra-provincial; en cuanto a la precipitación, se observa la más alta dispersión, especialmente en Pichincha, cuya

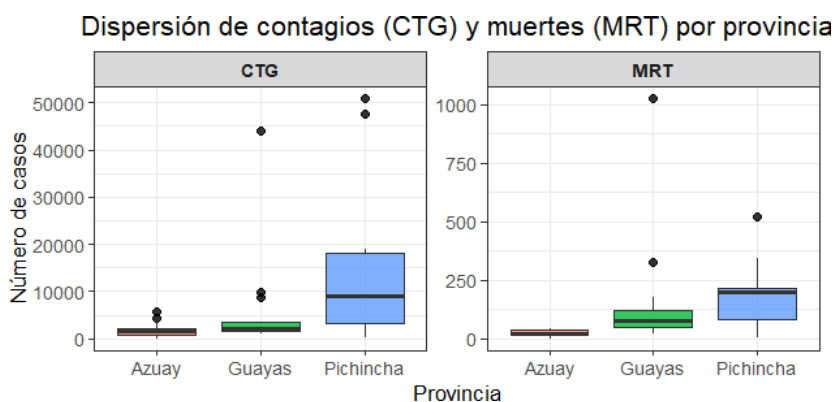
amplitud intercuartílica refleja la alternancia entre meses secos y lluviosos característica de su régimen bimodal; respecto a la concentración de $PM_{2.5}$, Guayas vuelve a destacar por su mayor variabilidad y rango intercuartílico. Esto puede indicar una mayor exposición a contaminantes atmosféricos en la región costera, asociada a la densidad vehicular e industrial de Guayaquil, mientras que las provincias serranas presentan menores niveles y una variabilidad más acotada.

Variables Dependientes

De manera complementaria, se examinó la distribución de las variables epidemiológicas utilizadas en el estudio. La Figura 4 presenta los diagramas de caja correspondientes al número de contagios (CTG) y muertes (MRT) por COVID-19 en Azuay, Guayas y Pichincha durante el período marzo de 2020–marzo de 2021. La representación permite comparar la dispersión, tendencia central y presencia de valores atípicos entre las tres provincias.

Figura 4

Diagrama de caja de variables epidemiológicas por provincia marzo 2020 - marzo 2021



Nota. CTG = número de contagios por COVID-19; MRT = número de muertes por COVID-19. Elaboración propia a partir de los datos epidemiológicos utilizados en el estudio.

Los boxplots de contagios revelan una dispersión amplia en Guayas y Pichincha, con presencia de outliers que corresponden a los picos de transmisión durante los primeros meses de la pandemia; para las muertes, la dispersión mantiene el mismo patrón espacial: los valores más elevados y variables corresponden a Pichincha y Guayas, con casos atípicos de mortalidad extrema, mientras que Azuay conserva un comportamiento más estable y de menor magnitud. Esta heterogeneidad interprovincial sugiere que los factores estructurales y ambientales pueden haber influido de forma diferenciada en la evolución de la pandemia.

4.2. Visualización De Tendencias Temporales

Las Figuras 5, 6 y 7 muestran la evolución mensual de las variables meteorológicas y ambientales durante el período marzo 2020 – marzo 2021. En general, se observa que la temperatura mantuvo una tendencia estable en las tres provincias, con valores constantes alrededor de 25 °C en Guayas y de 14 °C en Azuay y Pichincha, lo que refleja la estabilidad térmica típica de cada zona climática. La precipitación, en cambio, presentó una marcada estacionalidad, con máximos en marzo–abril y febrero–marzo, correspondientes a las épocas lluviosas tanto en Sierra y Costa, seguidas por un descenso sostenido hasta septiembre y una posterior recuperación hacia el final del periodo analizado. En cuanto al PM_{2.5}, Guayas mostró las concentraciones más elevadas y variables, con un pico entre mayo y julio de 2020 ($\approx 33 \mu\text{g}/\text{m}^3$), coincidiendo con la reducción del régimen lluvioso y condiciones de mayor estabilidad atmosférica que limitan la dispersión de contaminantes; Azuay y Pichincha mantuvieron niveles más bajos y estables, lo que evidencia un ambiente más limpio y menor exposición al material particulado.

Figura 5

Evolución temporal de la precipitación media por provincia marzo 2020 – marzo 2021

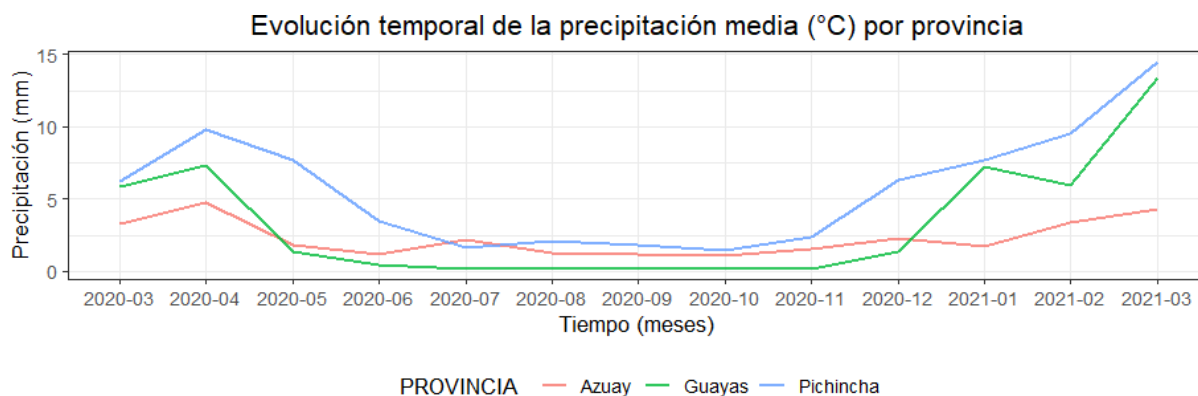


Figura 6

Evolución temporal de la temperatura media por provincia marzo 2020 – marzo 2021

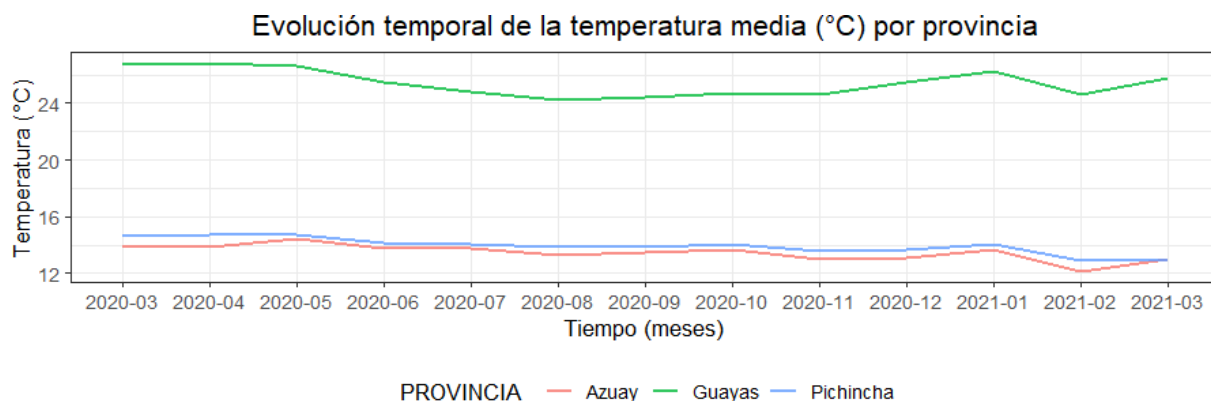
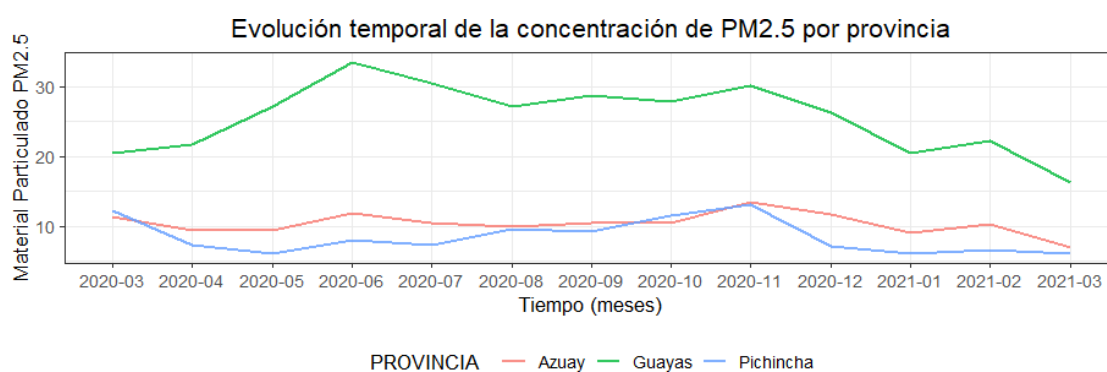


Figura 7

Evolución temporal de la concentración de PM2.5 por provincia marzo 2020 – marzo 2021



La Figura 8 y Figura 9 ilustran la evolución temporal de los contagios y muertes por COVID-19. Se identifican dos fases críticas: la primera entre abril y junio de 2020, donde Guayas registró los picos iniciales de contagios y mortalidad (correspondientes a la primera ola epidémica y al colapso sanitario documentado en esa provincia), y la segunda entre diciembre de 2020 y febrero de 2021, donde Pichincha presenta un incremento sostenido hasta alcanzar el máximo de casos y muertes en enero de 2021. Por otro lado, Azuay mantuvo valores reducidos y sin fluctuaciones extremas durante todo el periodo, reflejando una menor incidencia del virus. Estas diferencias espaciales y temporales evidencian una clara heterogeneidad interprovincial en la dinámica de la pandemia, probablemente influenciada por factores de densidad poblacional, movilidad urbana y vulnerabilidad estructural.

Figura 8

Evolución temporal del número de contagios por provincia

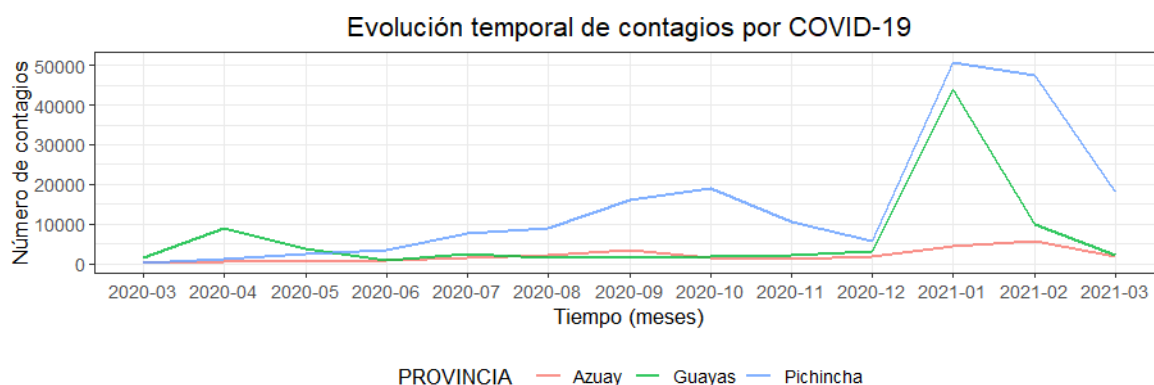
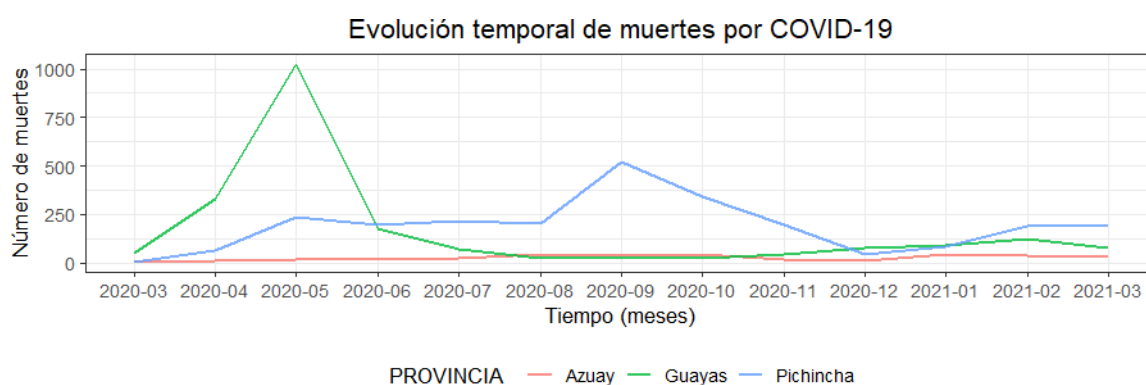


Figura 9

Evolución temporal del número de muertes



Fase 2: Análisis de influencia entre variables

4.3. Distribución y correlaciones de las variables

4.3.1. Distribución y correlaciones a nivel provincial

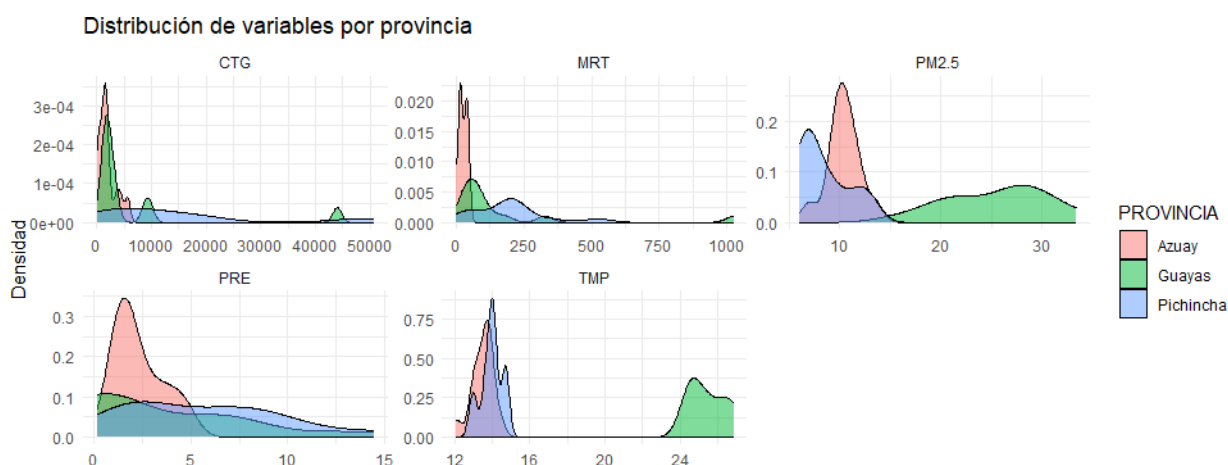
En la Figura 10 se observa la distribución de las variables principales por provincia. Los gráficos de densidad trazados permiten notar la heterogeneidad espacial y temporal de los datos. Pichincha presenta la mayor dispersión en el número de contagios y muertes, mientras que Azuay exhibe valores más concentrados y de menor magnitud, lo cual refleja su menor tamaño poblacional y densidad urbana. Guayas concentra las temperaturas más elevadas (entre 24 °C y 27 °C) y las mayores concentraciones de $PM_{2.5}$, coherentes con su condición costera e industrial; Azuay y Pichincha muestran distribuciones más estrechas de temperatura y concentraciones reducidas de contaminantes. La precipitación muestra fuerte

asimetría derecha en las tres provincias, destacando la variabilidad estacional y la presencia de meses con lluvias intensas, especialmente en Pichincha.

Figura 10

Distribución de variables meteorológicas, ambientales y epidemiológicas por provincia

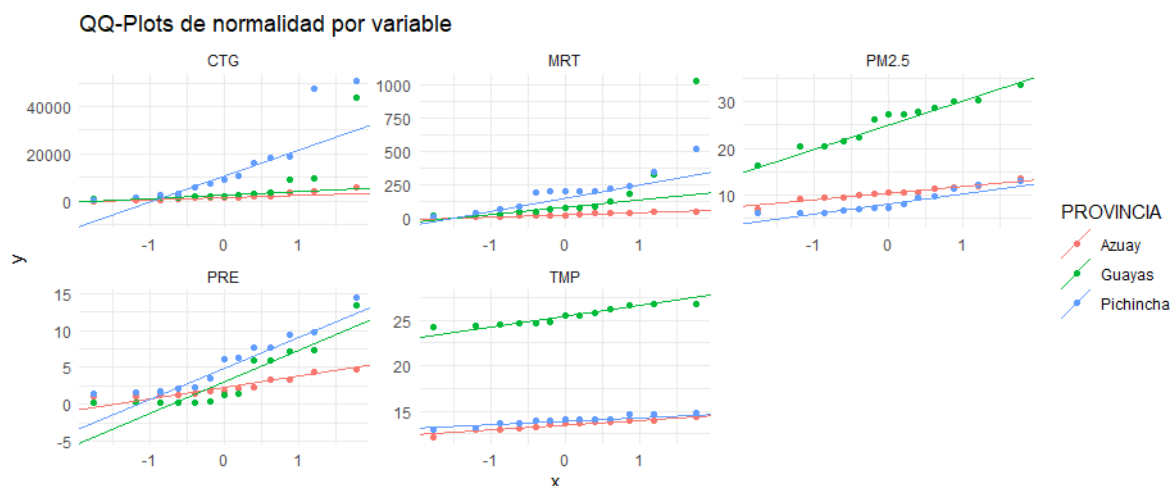
(marzo 2020 – marzo 2021)



Para poder estudiar las correlaciones entre variables explicativas fue necesario primero testear la normalidad. La Figura 11 muestra los QQ-plot que permiten visualizar la adecuación de las variables a la distribución normal: es una herramienta exploratoria previa a la selección del coeficiente de correlación (Pearson o Spearman). En términos generales las variables TMP, PRE y $PM_{2.5}$ presentan una alineación relativamente cercana a la diagonal teórica en las tres provincias, lo que sugiere una distribución aproximadamente normal; las variables CTG y MRT exhiben desviaciones marcadas respecto a la línea de referencia, evidenciando asimetrías positivas y la presencia de valores extremos que reflejan picos epidémicos mensuales.

Figura 11

QQ Plots de normalidad de las variables por provincia



Se probó normalidad numéricamente con el Test Shapiro-Wilk, para los pares de variables en la cual al menos una era de distribución no paramétrica se testeó correlación con el método “Spearman”, para aquellos pares normales se aplicó el método de Spearman. Los resultados se muestran la Tabla 7: presenta los coeficientes de correlación bivariada entre las variables meteorológicas y $PM_{2.5}$ para las tres provincias analizadas. En general, se evidencia una relación negativa entre la precipitación y el material $PM_{2.5}$, más pronunciada en Guayas y en menor medida en Pichincha, donde las lluvias contribuyen a la reducción de contaminantes. La correlación positiva entre temperatura y $PM_{2.5}$ en Guayas indica que las condiciones cálidas favorecen la acumulación de partículas. Azuay no muestra asociaciones estadísticamente significativas, reflejando una mayor estabilidad climática y menor variabilidad térmica.

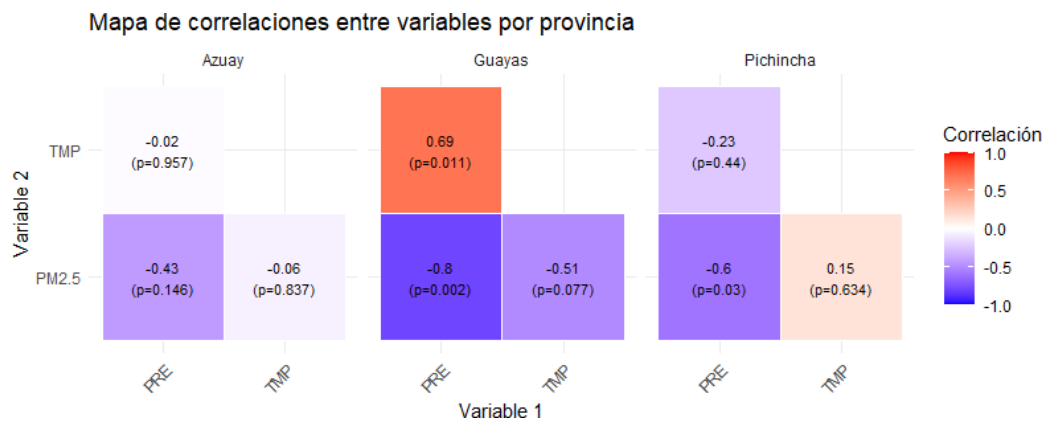
Tabla 7*Matriz de correlaciones entre variables independientes por provincias*

Provincia	Variable 1	Variable 2	Método	Correlación	p-value
Azuay	PRE	TMP	Spearman	-0.017	0.957
Azuay	PRE	PM _{2.5}	Spearman	-0.429	0.146
Azuay	TMP	PM _{2.5}	Pearson	-0.063	0.837
Guayas	PRE	TMP	Spearman	0.692	0.011
Guayas	PRE	PM _{2.5}	Spearman	-0.802	0.002
Guayas	TMP	PM _{2.5}	Pearson	-0.507	0.077
Pichincha	PRE	TMP	Pearson	-0.235	0.440
Pichincha	PRE	PM _{2.5}	Pearson	-0.601	0.030
Pichincha	TMP	PM _{2.5}	Pearson	0.146	0.634

Nota. Correlaciones bivariadas entre variables independientes por provincia. Se emplearon coeficientes de Spearman o Pearson según la distribución de los datos. Se puede apreciar visualmente estos resultados en la Figura 12.

Figura 12

Heatmap de correlaciones entre variables por provincia

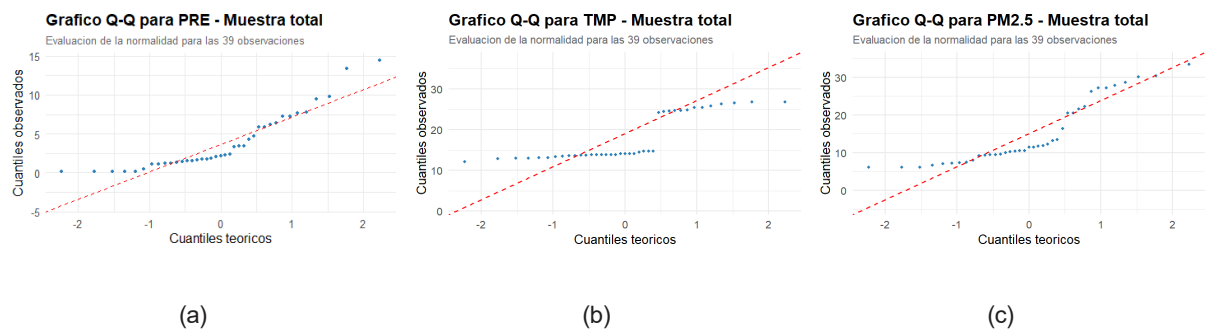


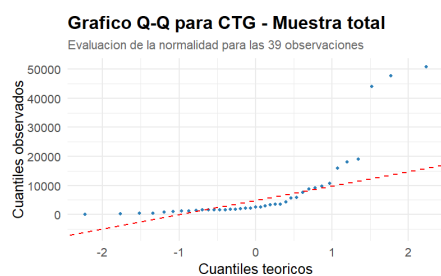
4.3.2. Correlaciones entre variables a nivel agregado

Los gráficos Q-Q presentados en la Figura 13 son una herramienta visual exploratoria. En las representaciones a), b) y c) los puntos se desvían de la línea diagonal teórica, especialmente en los extremos, lo que evidencia la presencia de asimetrías y posibles valores atípicos, resultado esperable dado el comportamiento estacional y la variabilidad intermensual de las condiciones climáticas y de calidad del aire. Las representaciones d) y e) muestran la distribución de los contagios y muertes por COVID-19 en la muestra total: en ambos casos se observa una desviación clara respecto a la línea diagonal, con asimetría positiva y presencia de valores extremos en los cuantiles superiores, atribuible a los picos epidémicos durante los meses críticos de la pandemia.

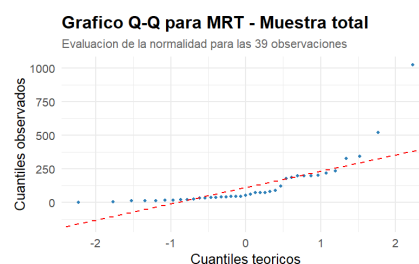
Figura 13

QQ Plots de normalidad de las variables a nivel agregado





(d)



(e)

La Tabla 8 muestra las correlaciones entre las variables meteorológicas y la concentración de material particulado $PM_{2.5}$ considerando la muestra total. La prueba de Shapiro–Wilk indicó que en las distribuciones de normalidad en las distribuciones, se aplicó el coeficiente de Spearman, adecuado para relaciones monótonas no lineales. Los resultados evidencian una correlación negativa moderada entre PRE y $PM_{2.5}$ (-0.568), lo que sugiere que mayores niveles de precipitación tienden a reducir la concentración de partículas en el aire por procesos de lavado atmosférico. Se observa una correlación positiva entre TMP y $PM_{2.5}$ (0.606), indicando que las condiciones más cálidas pueden favorecer la acumulación de aerosoles o limitar su dispersión. La correlación entre PRE y TMP (-0.081) revela independencia estadística entre ambas variables. En conjunto, los resultados confirman relaciones físicas coherentes con la dinámica atmosférica tropical del Ecuador.

Tabla 8

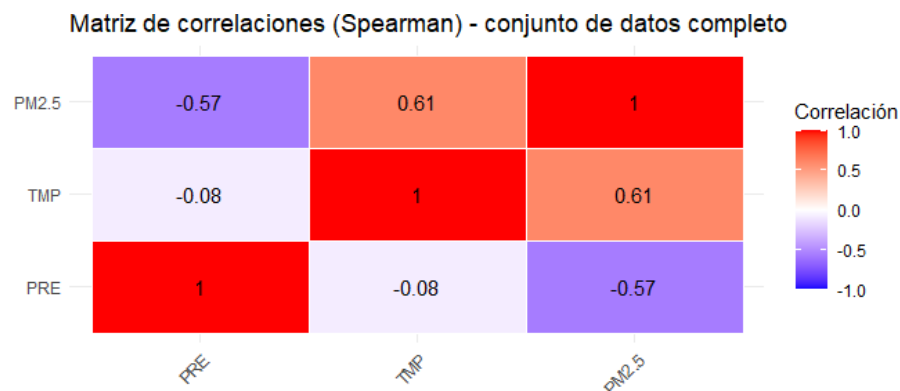
Matriz de correlaciones entre variables independientes a nivel agregado (Spearman)

	PRE	TMP	$PM_{2.5}$
PRE	1.000	-0.081	-0.568
TMP	-0.081	1.000	0.606
$PM_{2.5}$	-0.568	0.606	1.000

Nota. Se puede apreciar visualmente estos resultados en la Figura 14.

Figura 14

Heatmap de correlaciones de variables independientes a nivel agregado



La Tabla 9 presenta las correlaciones entre las variables sociodemográficas: densidad poblacional (DEN), índice de desigualdad de Gini (GIN), proporción de adultos mayores (PMA) y población total (POB). Los resultados evidencian una alta correlación negativa entre GIN y POB (-0.991) (provincias con mayor población tienen una distribución del ingreso más equitativa). Asimismo, se observa una correlación negativa fuerte entre DEN y PMA (-0.887), indicando que las zonas con mayor densidad poblacional registran menor proporción de adultos mayores, un patrón coherente con la concentración urbana de población joven y laboralmente activa. Las demás correlaciones son moderadas o débiles, lo que sugiere que no existe colinealidad severa entre los predictores. Por estas razones se eliminaron las covariables de PMA y POB para los análisis posteriores.

Tabla 9

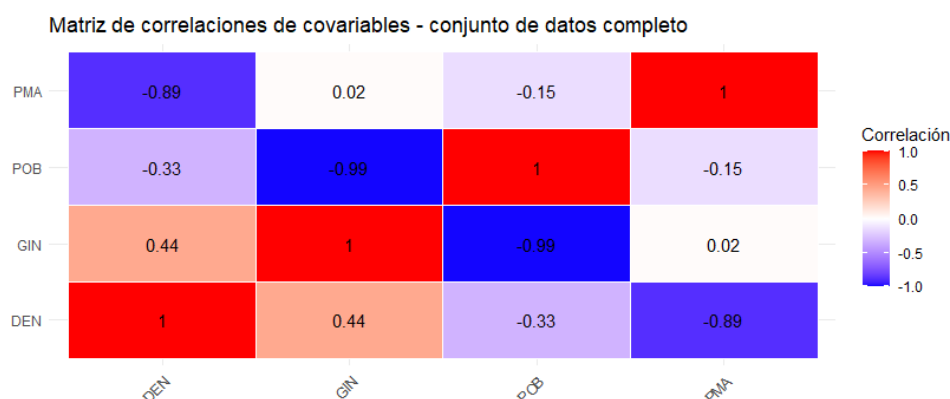
Matriz de correlaciones entre covariables independientes a nivel agregado (Pearson)

	DEN	GIN	PMA	POB
DEN	1.000	0.444	-0.887	-0.325
GIN	0.444	1.000	0.018	-0.991
PMA	-0.887	0.018	1.000	-0.146
POB	-0.325	-0.991	-0.146	1.000

Nota. Se puede apreciar visualmente estos resultados en la Figura 15.

Figura 15

Heatmap de correlaciones de covariables independientes



Los análisis de correlación revelan patrones de asociación diferenciados entre los grupos de variables. Las variables meteorológicas presentan dependencias moderadas coherentes con procesos físicos atmosféricos, mientras que las covariables sociodemográficas exhiben relaciones estructurales internas vinculadas con la distribución poblacional y la desigualdad. Esta diferenciación justifica el tratamiento en bloques independientes (meteorológico y socioeconómico) en la modelización posterior, y para reducir la multicolinealidad se aplica Análisis de Componentes Principales (PCA) para sintetizar la información de estas variables sin pérdida significativa de variabilidad; en el caso de las covariables se eliminaron aquellas que estaban altamente correlacionadas.

4.4. Análisis de Componentes Principales (PCA)

El análisis de componentes principales (PCA) se aplicó como estrategia de reducción de dimensionalidad sobre el conjunto de variables meteorológicas y ambientales (TMP, PRE Y $PM_{2.5}$), tal que se pueda sintetizar la información atmosférica sin pérdida sustancial de variabilidad. Los resultados del PCA de la Tabla 10 muestran que los dos primeros componentes principales explican conjuntamente el 98.49 % de la varianza total del conjunto de datos, con proporciones individuales del 66.39 % para PC_1 y 32.10 % para PC_2 . En consecuencia, la dimensionalidad del bloque meteorológico se redujo de tres a dos componentes sin pérdida significativa de información estadística (Jolliffe y Cadima, 2016).

Tabla 10*Proporción y varianza explicada por cada componente del PCA*

Componente	Desviación estándar	Proporción de varianza	Proporción acumulada
PC1	1.4113	0.6639	0.6639
PC2	0.9813	0.3210	0.9849
PC3	0.2129	0.0151	1.0000

En la Tabla 11 se observa que el Componente Principal 1 (PC_1) presenta cargas positivas altas para la temperatura (0.646) y para el $PM_{2.5}$ (0.700), y una carga negativa moderada para la precipitación (-0.305). Este patrón sugiere que PC_1 representa condiciones cálidas y secas asociadas a mayor concentración de material particulado, lo cual caracteriza escenarios atmosféricos de acumulación de contaminantes; mientras tanto, el Componente Principal 2 (PC_2) se encuentra dominado por la precipitación (0.918) y en menor medida por la temperatura (0.394), esto refleja condiciones húmedas o pluviales, que favorecen procesos de dispersión de aerosoles.

PC_1 y PC_2 resumen dos ejes físicos fundamentales del comportamiento meteorológico: el primero vinculado al calentamiento y estabilidad del aire, y el segundo a la dinámica hídrica y ventilación atmosférica. En síntesis, el PCA permitió reducir la colinealidad entre las variables meteorológicas y obtener factores latentes ortogonales que capturan los principales gradientes climáticos del periodo analizado. Por ello, los componentes PC_1 y PC_2 fueron seleccionados como predictores representativos del bloque meteorológico en los modelos de datos de panel, garantizando parsimonia y robustez en la estimación de los efectos sobre las variables epidemiológicas.

Tabla 11

Cargas de las variables en los componentes principales (matriz de rotación)

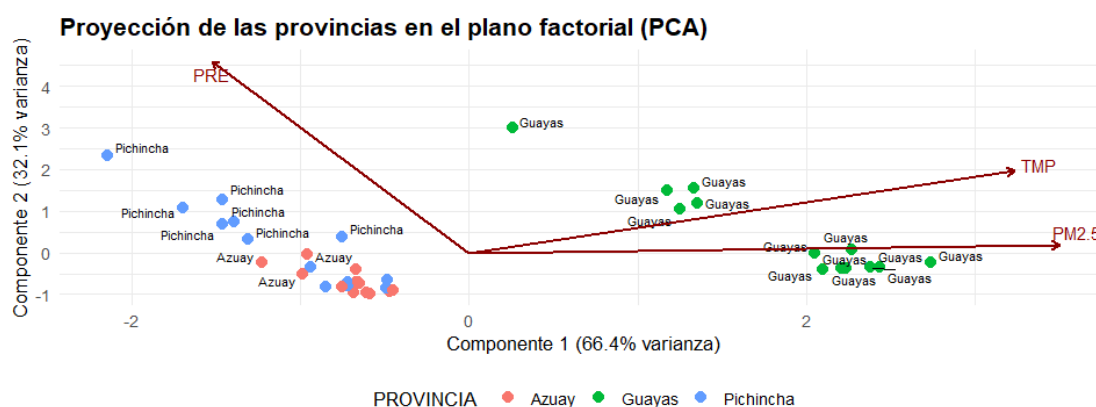
Variable	PC1	PC2	PC3
TMP	0.6461	0.3938	0.6539
PRE	-0.3046	0.9185	-0.2522
PM _{2.5}	0.6999	0.0362	-0.7133

Nota. El primer componente (PC1) explica el 66.4% de la varianza total y refleja el efecto combinado de la temperatura y la concentración de PM_{2.5}. El segundo componente (PC2) captura un 32.1% adicional, principalmente asociado con la precipitación. En conjunto, los dos primeros componentes explican el 98.5% de la variabilidad de los datos meteorológicos.

La Figura 16 muestra la proyección de las provincias en el plano definido por los dos primeros componentes principales. Se distingue un claro gradiente espacial: Guayas se asocia a condiciones cálidas, secas y con mayor carga de PM_{2.5}, mientras que Pichincha se ubica en el eje de mayor precipitación y Azuay mantiene una posición intermedia. La orientación ortogonal de las variables confirma que los ejes PC₁ y PC₂ capturan dimensiones atmosféricas complementarias, lo que respalda su inclusión como predictores sintéticos del bloque meteorológico en el modelo de panel.

Figura 16

Proyección de las provincias en el plano factorial (biplot del PCA)



Fuente: Elaboración propia con datos de TMP, PRE y PM2.5 (2020–2021)

Fase 3: Estimación y análisis de modelo de regresión

4.5. Estimación del modelo de datos de panel

En esta sección se presenta la estimación del modelo de datos de panel, cuyo objetivo es medir cómo influyen las variables meteorológicas y socioeconómicas en la expansión del COVID-19 en las provincias estudiadas. Para ello, se evaluaron tres especificaciones econométricas: el modelo Pooled OLS, el modelo de efectos fijos (FE) y el de efectos aleatorios (RE). Con el objetivo de identificar la estructura más adecuada para los datos se aplicaron las pruebas de selección de modelo (Chow, Breusch–Pagan, Hausman) y los test de diagnóstico de supuestos para determinar la conveniencia de incorporar errores robustos de Driscoll–Kraay. Finalmente, se estimaron los modelos por bloques meteorológico y socioeconómico, lo que permitió una interpretación diferenciada de los determinantes climáticos y estructurales del contagio y la mortalidad por COVID-19.

4.5.1. Modelo Contagios COVID-19 (marzo 2020 - 2021)

Tabla 12

Resultados de modelos de panel para Contagios por COVID-19

Modelo POOLED OLS				
Variable	Estimación	Error Estándar	t-valor	p-valor
(Intercept)	332046.60	270186.89	1.2290	0.2275
PC1	-14872.25	10019.96	-1.4843	0.1470
PC2	-4589.78	5936.85	-0.7731	0.4448
DEN	188.19	119.61	1.5733	0.1249
GIN	-838044.18	676682.25	-1.2385	0.2240
$R^2 = 0.2873$; R^2 ajustado = 0.2035; $F(4, 34) = 3.426$, $p = 0.0186$.				
Modelo Efectos Fijos FE				
Variable	Estimación	Error Estándar	t-valor	p-valor
PC1	-14872.30	10020.00	-1.4843	0.1470
PC2	-4589.80	5936.90	-0.7731	0.4448
$R^2 = 0.1281$; R^2 ajustado = 0.0255; $F(2, 34) = 2.497$, $p = 0.0973$.				
Modelo De Efectos Aleatorios RE				
Variable	Estimación	Error Estándar	t-valor	p-valor
(Intercept)	332046.60	270186.89	1.2290	0.2191
PC1	-14872.25	10019.96	-1.4843	0.1377
PC2	-4589.78	5936.85	-0.7731	0.4395
DEN	188.19	119.61	1.5733	0.1157
GIN	-838044.18	676682.25	-1.2385	0.2155
$R^2 = 0.2873$; R^2 ajustado = 0.2035; $\chi^2(4) = 13.7056$, $p = 0.0089$.				

La Tabla 12 muestra la comparación de los tres modelos: el modelo Pooled OLS presenta el mejor ajuste global, con un $R^2_{ajustado} = 0.2035$ y una significancia conjunta del modelo ($p = 0.0186$), lo que sugiere una capacidad explicativa moderada de las variables meteorológicas y socioeconómicas. En contraste, el modelo de efectos fijos (FE) reduce su poder explicativo ($R^2_{ajustado} = 0.0255$) y no alcanza significancia global ($p = 0.097$), lo que indica que las variaciones intra-provinciales no aportan información adicional relevante. El modelo de efectos aleatorios (RE) mantiene resultados similares al Pooled, con igual $R^2_{ajustado}$ y significancia global ($p < 0.01$), sugiriendo que la variabilidad entre provincias puede modelarse como aleatoria sin afectar la consistencia de los estimadores.

En cuanto a los coeficientes individuales, ninguna variable resulta estadísticamente significativa al 5 %, aunque los signos son coherentes con la teoría: PC1 y PC2 muestran efectos negativos sobre los contagios (menor transmisión bajo condiciones cálidas o húmedas), mientras que DEN y GIN presentan signos positivos, indicando que la densidad poblacional y la desigualdad tienden a incrementar la propagación. Los resultados justifican mantener la especificación Pooled o RE, por lo que es necesaria la validación mediante pruebas de selección formal (Chow, Breusch–Pagan y Hausman).

Pruebas de selección del modelo de panel para contagios por COVID-19

Para determinar el modelo de datos de panel más adecuado, se aplicaron las pruebas:

- F de efectos fijos (Chow test), Lagrange Multiplier de Breusch–Pagan (LM test) y Hausman. El test F para efectos individuales ($pFtest$) arrojó un resultado no definido ($F = NaN$; $p = NA$), lo que indica ausencia de variabilidad significativa entre unidades o insuficiencia de grados de libertad para estimar efectos fijos. En consecuencia, no se justifica el uso del modelo de efectos fijos frente al modelo agrupado (Pooled OLS).
- El test LM de Breusch–Pagan ($\chi^2 = 1.625$; $p = 0.2024$) no rechaza la hipótesis nula de homogeneidad entre individuos, es decir que no existen efectos aleatorios significativos, y por tanto, el modelo Pooled OLS resulta estadísticamente equivalente al de efectos aleatorios.
- El test de Hausman ($\chi^2 = 3 \times 10^{-14}$; $p = 1$) muestra que no existen diferencias sistemáticas entre los estimadores de efectos fijos y aleatorios, por lo que ambos resultan consistentes, aunque el modelo RE es menos eficiente debido a la falta de variabilidad individual significativa.

Los tres contrastes apuntan hacia la preferencia del modelo Pooled OLS como especificación adecuada para los contagios por COVID-19, al no evidenciarse efectos individuales relevantes entre provincias.

Modelo Pooled OLS por Bloques

Dada la naturaleza heterogénea de las variables explicativas, el modelo se estimó de manera diferenciada en dos bloques conceptuales:

(1) bloque meteorológico, que integra las condiciones atmosféricas sintetizadas en los componentes principales (PC₁, PC₂),

(2) bloque socioeconómico, que agrupa factores estructurales como la densidad poblacional (DEN) y el coeficiente de Gini (GIN).

Tabla 13

Modelo Pooled OLS por Bloques (sin Driscoll-Kraay)

Bloque Meteorológico				
Variable	Estimate	Std. Error	t-value	p-value
(Intercept)	7704.20	1893.20	4.0695	0.0002***
PC1	-2002.70	1359.00	-1.4737	0.1493
PC2	4350.50	1954.50	2.2259	0.0324*

$R^2 = 0.1653$, R^2 ajustado = 0.1189, $F(2, 36) = 3.563$, $p = 0.0387$.

Bloque Socioeconómico				
Variable	Estimate	Std. Error	t-value	p-value
(Intercept)	-53970.79	37854.01	-1.4258	0.1626
DEN	33.579	20.502	1.6379	0.1102
GIN	121345.09	90060.84	1.3474	0.1863

$R^2 = 0.1826$, R^2 ajustado = 0.1372, $F(2, 36) = 4.021$, $p = 0.0265$.

En el bloque meteorológico de la Tabla 13, el modelo Pooled OLS muestra un ajuste moderado, los componentes derivados del PCA explican aproximadamente el 12 % de la variación en los contagios. El componente PC_1 presenta un coeficiente negativo, aunque no significativo, lo que sugiere que condiciones cálidas y secas podrían asociarse con una ligera disminución en los casos. Por su parte, PC_2 es positivo y significativo, puede indicar que combinaciones de precipitación y temperatura moderadas tienden a favorecer un incremento en los contagios. Estos resultados respaldan la existencia de efectos atmosféricos diferenciados sobre la propagación viral, aunque de magnitud limitada. En el bloque socioeconómico, el modelo explica una proporción similar de la variabilidad, pero con signos y magnitudes más coherentes con la literatura (Pani et al., 2022). Los coeficientes positivos de DEN y GIN reflejan que mayor densidad poblacional y desigualdad social tienden a elevar los niveles de contagio, lo cual es consistente con estudios recientes que vinculan la concentración urbana y la vulnerabilidad estructural con la expansión del COVID-19 (Ganegoda et al., 2021; Pani et al., 2022).

Ambos modelos confirman que, si bien las condiciones meteorológicas aportan información relevante sobre la dinámica estacional de los contagios, los factores socioeconómicos tienen un peso relativamente mayor y explican de forma más estable las diferencias interprovinciales en la incidencia de la enfermedad.

Tabla 14*Modelo por bloques con errores robustos Driscoll–Kraay, variable CTG*

Variable	Bloque Meteorológico	Bloque Socioeconómico
PC1	-2002.74** (975.06)	
PC2	4350.51 (2781.01)	
DEN		33.579** (16.50)
GIN		121345.10** (64361.31)
Constante	7704.15*** (2654.21)	-53970.79** (27992.31)
R ²	0.165	0.183
R ² ajustado	0.119	0.137
F(2, 36)	3.563**	4.021**

Nota. Estimaciones de los modelos Pooled OLS corregidos con errores robustos de Driscoll–Kraay. *p<0.1; **p<0.05; ***p<0.01.

4.5.2. Modelo Muertes por COVID-19 (marzo 2020 - 2021)

En la estimación inicial del modelo de mortalidad MRT se aplicaron también las tres especificaciones clásicas del análisis de datos de panel: modelo Pooled OLS, modelo de efectos fijos (FE) y modelo de efectos aleatorios (RE). Para determinar la estructura más adecuada, se realizaron las pruebas de contraste estándar:

- El test F para efectos individuales arrojó un valor $F = -\text{Inf}$, p-value = NA, lo que indica que no existen efectos individuales significativos; por tanto, el modelo de efectos fijos no resulta apropiado frente al modelo agrupado.
- El test de Breusch–Pagan (LM), con un valor $\chi^2 = 1.625$, $p = 0.2024$, sugiere que no hay evidencia de varianza entre unidades, descartando la conveniencia del modelo de efectos aleatorios.
- El test de Hausman ($\chi^2 = 5.11 \times 10^{-14}$, $p = 1$) confirma la consistencia del modelo Pooled respecto a las alternativas FE y RE, indicando que no existe correlación sistemática entre los efectos individuales y las variables explicativas.

En consecuencia, el modelo Pooled OLS se considera la especificación más adecuada para analizar la mortalidad, dado que no se detectan efectos no observados significativos entre provincias.

En la Tabla 15 se observa los resultados del modelo Pooled OLS, corregido con errores robustos de White–Huber (HC3), muestran que las variables meteorológicas y socioeconómicas explican parcialmente la variabilidad de la mortalidad por COVID-19. En el bloque meteorológico, los coeficientes de los componentes principales PC_1 y PC_2 no resultaron estadísticamente significativos, lo cual indica que las condiciones atmosféricas sintetizadas en estos factores no tuvieron un efecto directo sobre la mortalidad. El bloque socioeconómico evidencia una relación más relevante: aunque las variables densidad poblacional (DEN) y desigualdad (GIN) no presentan significancia individual, el modelo conjunto es significativo ($F = 3.355$; $p < 0.05$), con un R^2 ajustado de 0.110, lo que sugiere que las diferencias estructurales entre provincias pueden contribuir parcialmente a explicar las variaciones en la mortalidad.

Estos resultados apuntan a que los factores socioeconómicos presentan un mayor peso explicativo sobre la mortalidad que los meteorológicos, aunque la magnitud del efecto global sigue siendo moderada.

Tabla 15

Modelo por bloques con errores White–Huber, variable MRT

Variable	Bloque Meteorológico	Bloque Socioeconómico
PC1	19.382 (15.457)	
PC2	14.370 (24.349)	
DEN		0.730
GIN		-181.514
Constante	127.615*** (37.919)	31.135
Observaciones	39	39
R^2	0.027	0.157
R^2 ajustado	-0.027	0.110
$F(2, 36)$	0.508	3.355**

Nota. Estimaciones Pooled OLS con corrección robusta de White–Huber aplicadas a la variable dependiente MRT. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

4.6. Validación de supuestos estadísticos

Variable CTG

La validación de los supuestos estadísticos se efectuó de manera independiente para los bloques meteorológico y socioeconómico, con el fin de garantizar la consistencia y robustez de las estimaciones obtenidas bajo el modelo Pooled OLS.

- Normalidad de los residuos

La prueba de Shapiro–Wilk arrojó $W = 0.7767$ ($p = 2.88 \times 10^{-6}$) para el bloque meteorológico y $W = 0.6982$ ($p = 1.18 \times 10^{-7}$) para el bloque socioeconómico, indicando rechazo de la hipótesis nula de normalidad en ambos casos. Dado que los residuos no se distribuyen normalmente, se optó por aplicar correcciones robustas de varianza (Driscoll–Kraay y White–Huber) para obtener estimaciones consistentes ante violaciones de este supuesto.

- Heterocedasticidad

La prueba Breusch–Pagan evidenció heterocedasticidad en el bloque meteorológico ($BP = 8.256$, $p = 0.0161 < 0.05$), mientras que en el bloque socioeconómico no se detectó evidencia significativa ($BP = 3.595$, $p = 0.1657 > 0.05$). Esto sugiere que las varianzas de los errores no son constantes en el modelo meteorológico, por lo que se justificó el uso de errores robustos en la estimación final.

- Autocorrelación

El contraste Breusch–Godfrey/Wooldridge reveló autocorrelación de primer orden en ambos bloques: en el Bloque meteorológico: $\chi^2 = 25.256$, $p = 0.0214$; en el Bloque socioeconómico: $\chi^2 = 30.739$, $p = 0.0037$. La presencia de autocorrelación serial indica dependencia temporal en los residuos, lo que refuerza la elección de estimadores robustos a la autocorrelación y heterocedasticidad, como Driscoll–Kraay.

- Dependencia transversal

El test de Pesaran CD mostró evidencia significativa de dependencia contemporánea entre las unidades de sección cruzada en ambos bloques ($z = 4.25$, $p = 2.14 \times 10^{-5}$ para el bloque meteorológico y $z = 4.46$, $p = 8.03 \times 10^{-6}$ para el socioeconómico). Esto implica que los residuos de las provincias no son independientes, lo cual respalda el uso de estimadores robustos que corrigen tanto autocorrelación como dependencia transversal.

Los resultados de las pruebas de diagnóstico confirman que los supuestos clásicos de homocedasticidad, independencia y normalidad no se cumplen plenamente en los modelos de contagios. Sin embargo, la aplicación de correcciones robustas de varianza (Driscoll–

Kraay) garantiza la validez inferencial de los coeficientes estimados, haciendo que los modelos sean robustos y confiables para la interpretación econométrica.

Variable MRT

Al igual que en el caso de los modelos de contagios, se evaluaron los principales supuestos estadísticos del modelo Pooled OLS aplicado a la variable de mortalidad (MRT), diferenciando los bloques meteorológico y socioeconómico. Los resultados permitieron determinar la validez de las estimaciones y la necesidad de aplicar correcciones robustas en caso de incumplimiento.

- Normalidad de los residuos

La prueba de Shapiro–Wilk mostró valores $W = 0.65$ ($p = 1.98 \times 10^{-8}$) para el bloque meteorológico y $W = 0.62$ ($p = 8.77 \times 10^{-9}$) para el bloque socioeconómico, esto es, un rechazo de la hipótesis nula de normalidad en ambos casos. Esto indica que los residuos no siguen una distribución normal; sin embargo, debido al tamaño muestral moderado y al uso de estimadores robustos (White–Huber), los coeficientes estimados permanecen consistentes y válidos para inferencia

- Heterocedasticidad

Los resultados de la prueba Breusch–Pagan no mostraron evidencia de heterocedasticidad: en el Bloque meteorológico $BP = 2.55$, $p = 0.28$, en el Bloque socioeconómico: $BP = 2.37$, $p = 0.31$. Así, se acepta la hipótesis de homocedasticidad, lo que sugiere que las varianzas de los errores son constantes y no afectan la eficiencia de los estimadores.

- Autocorrelación

El test Breusch–Godfrey/Wooldridge evidenció ausencia de autocorrelación serial en ambos modelos: en el Bloque meteorológico: $\chi^2 = 11.732$, $p = 0.5498$; en el Bloque socioeconómico: $\chi^2 = 11.93$, $p = 0.54$. Ambos valores p mayores a 0.05 confirman que los errores no presentan correlación temporal, cumpliéndose el supuesto de independencia de los residuos a lo largo del tiempo.

- Dependencia transversal

El contraste Pesaran CD arrojó $z = 0.48$ ($p = 0.63$) para el bloque meteorológico y $z = 0.43$ ($p = 0.67$) para el bloque socioeconómico, esto indica ausencia de correlación contemporánea entre las provincias.

4.7. Ajuste funcional del modelo

En la Tabla 16 se muestran los resultados de la prueba RESET de Ramsey aplicada a los modelos Pooled OLS por bloques (meteorológico y socioeconómico) para las variables dependientes CTG (contagios) y MRT (mortalidad). En todos los casos, los valores p obtenidos son superiores al umbral de 0.05, lo que indica que no se rechaza la hipótesis nula del test de Ramsey. Por tanto, los modelos lineales propuestos presentan una adecuada forma funcional, sin evidencia de especificación incorrecta o necesidad de incluir términos no lineales o interacciones. De acuerdo con Wooldridge (2021), este resultado respalda la consistencia del modelo en su forma lineal, y según Gujarati y Porter (2020), la ausencia de errores de especificación garantiza que las inferencias derivadas de los coeficientes estimados sean válidas y no sesgadas.

Tabla 16

Resultados del test RESET de Ramsey por bloques y variable dependiente

Variable dependiente	Bloque analizado	Estadístico F	p-valor	Conclusión
CTG	Meteorológico (PC ₁ , PC ₂)	1.4449	0.242	No se rechaza H ₀ : especificación funcional correcta.
CTG	Socioeconómico (DEN, GIN)	0.0000	1.000	No se rechaza H ₀ : especificación funcional correcta.
MRT	Meteorológico (PC ₁ , PC ₂)	0.1424	0.965	No se rechaza H ₀ : modelo correctamente especificado.
MRT	Socioeconómico (DEN, GIN)	0.0000	1.000	No se rechaza H ₀ : modelo correctamente especificado.

DISCUSIÓN

Los resultados de la presente investigación permiten integrar de forma coherente los hallazgos derivados de los bloques meteorológico y socioeconómico, contrastando su influencia conjunta sobre los contagios y muertes por COVID-19 en las provincias de Azuay, Guayas y Pichincha durante el periodo 2020–2021. El análisis basado en modelos de datos de panel con estimaciones Pooled OLS robustas, evidenció diferencias notables en el grado de asociación entre las condiciones ambientales, los factores estructurales y los indicadores epidemiológicos, propósito que perseguían los Objetivos Específicos 1, 2 y 3 de esta investigación.

Efecto de las variables meteorológicas

El bloque meteorológico —compuesto por los componentes principales PC_1 y PC_2 derivados de la temperatura, la precipitación y la concentración de $PM_{2.5}$ — mostró efectos moderados pero consistentes sobre los contagios. En particular, el PC_2 evidenció significancia estadística ($p < 0.05$) bajo corrección de Driscoll–Kraay, sugiriendo que los contrastes térmicos y la variabilidad pluviométrica condicionan la dispersión del virus. Este resultado coincide con investigaciones previas que señalan la sensibilidad del SARS-CoV-2 a los cambios de humedad y temperatura, así como al transporte de aerosoles finos en entornos urbanos (Ganegoda et al., 2021; Bashir et al., 2020). No obstante, la magnitud del efecto fue reducida, lo que concuerda con estudios que advierten que las variables meteorológicas actúan como moduladores indirectos, más que como determinantes primarios del contagio (Ahmad et al., 2022).

En el caso de la mortalidad (MRT), las variables atmosféricas no presentaron significancia estadística, lo cual sugiere que las condiciones climáticas no influyeron directamente en la letalidad del virus, sino de manera indirecta a través de la carga viral ambiental o la exposición prolongada a contaminantes (Coccia, 2021).

Efecto de las variables socioeconómicas

El bloque socioeconómico, integrado por la densidad poblacional (DEN) y el índice de desigualdad (GINI), presentó asociaciones más sólidas con los indicadores epidemiológicos. En el modelo de contagios, ambos coeficientes fueron significativos tras la corrección Driscoll–Kraay, lo cual puede estar en concordancia con que las áreas con mayor concentración urbana y desigualdad social registraron tasas más elevadas de infección. De hecho, este hallazgo respalda la evidencia empírica que asocia la vulnerabilidad social con la propagación del COVID-19, debido a las limitaciones en el distanciamiento físico, la movilidad laboral informal y la exposición ocupacional (Ahmed et al., 2020; Gozzi et al., 2021).

Para la mortalidad, aunque los coeficientes no alcanzaron significancia estadística, el signo esperado de GINI fue positivo, confirmando la tendencia observada en la literatura internacional: la desigualdad y la pobreza aumentan la exposición a comorbilidades y reducen la capacidad de respuesta sanitaria (Oronce et al., 2020; Bambra et al., 2020).

Efecto global

En conjunto, los modelos sugieren que la dinámica de contagios estuvo más influenciada por factores estructurales que por condiciones meteorológicas, aunque ambas dimensiones interactúan. La exposición a contaminación atmosférica y la densidad urbana elevada generan entornos de mayor riesgo, donde la dispersión viral se ve potenciada por la concentración de partículas finas y la movilidad humana, particularmente en Guayas durante los primeros meses de la pandemia. En cambio, la mortalidad reflejó la capacidad adaptativa de los sistemas de salud y las diferencias socioeconómicas, más que los factores ambientales.

CONCLUSIONES Y RECOMENDACIONES

CONCLUSIONES

El presente estudio tuvo como objetivo general cuantificar el efecto de las variables meteorológicas y socioeconómicas sobre el esparcimiento del COVID-19 en Ecuador, mediante la aplicación de un modelo de regresión múltiple basado en datos de panel que capturara simultáneamente la variabilidad temporal e interprovincial durante el periodo marzo 2020 – marzo 2021. El análisis descriptivo evidenció coherencia climática de los datos con la ubicación geográfica de las provincias. Guayas presentó la mayor temperatura media (≈ 26 °C) y concentración de $PM_{2.5}$, mientras que Pichincha mostró valores inferiores de temperatura y mayores precipitaciones, coherentes con su localización andina.

La estructura espacial y temporal identificada permitió sintetizar la información mediante Análisis de Componentes Principales (PCA), del cual emergieron dos componentes representativos de las condiciones atmosféricas conjuntas (PC_1 y PC_2), que explicaron más del 98 % de la varianza total del bloque meteorológico. PC_1 (66% de varianza) puede interpretarse como un indicador sintético de calor y contaminación acumulada, cuya variación refleja la interacción entre radiación solar, estabilidad atmosférica y emisiones urbanas; PC_2 refleja la influencia de los regímenes de lluvia y procesos de ventilación sobre la dispersión o acumulación de aerosoles en el aire, típicos de la región interandina y del Austro.

La exploración de los indicadores de desigualdad, densidad poblacional, proporción de adultos mayores y pobreza multidimensional evidenció disparidades estructurales entre provincias, destacando Guayas por su mayor densidad urbana y desigualdad social. Las correlaciones entre variables confirmaron la necesidad de modelar estos factores de manera separada del bloque meteorológico, evitando colinealidad y preservando la interpretación económica del modelo.

Tras comparar las especificaciones Pooled OLS, efectos fijos y efectos aleatorios, las pruebas de selección (Chow, Breusch–Pagan y Hausman) demostraron que no existen efectos individuales significativos entre provincias, por lo que el modelo Pooled fue el más eficiente. Los diagnósticos de supuestos estadísticos revelaron heterocedasticidad, autocorrelación y dependencia transversal en los modelos de contagios, y un mejor comportamiento de los residuos en los modelos de mortalidad. Por ello, se aplicaron errores robustos Driscoll–Kraay para CTG y White–Huber (HC3) para MRT.

El análisis evidenció una relación estadísticamente significativa entre los componentes atmosféricos y el nivel de contagios. En particular, PC_2 (variabilidad de temperatura y

precipitación) mostró un efecto positivo sobre el número de casos, mientras que PC_1 (condiciones cálidas y secas) presentó una relación negativa. Esto sugiere que los patrones climáticos estacionales y la ventilación atmosférica influyen en la dispersión del virus, aunque su impacto es moderado y dependiente del contexto urbano. En los modelos de mortalidad, los efectos meteorológicos no fueron significativos, indicando que el clima incide más en la transmisión que en la letalidad, coincidiendo con lo reportado por Ganegoda et al. (2021) y Coccia (2021).

Las variables densidad poblacional y desigualdad social fueron las de mayor peso explicativo. Su relación positiva y significativa con los contagios confirma que la aglomeración urbana y la inequidad estructural amplifican la transmisión del virus, coherente con la evidencia empírica internacional (Ahmed et al., 2020; Gozzi et al., 2021). Aunque los efectos sobre la mortalidad no resultaron significativos, la dirección de los coeficientes sugiere que la desigualdad puede agravar la vulnerabilidad y reducir la capacidad de respuesta sanitaria en provincias con menor equidad social. En términos estadísticos, el bloque socioeconómico explicó una proporción ligeramente mayor de la variabilidad total que el bloque meteorológico, reforzando su rol determinante en la dinámica epidemiológica.

Los resultados del test RESET de Ramsey ($p > 0.05$ en todos los casos) confirmaron la adecuada especificación funcional lineal del modelo. Así también, la validación de supuestos evidenció que las correcciones robustas adoptadas aseguran la fiabilidad de los coeficientes estimados. De acuerdo con Wooldridge (2021), esto garantiza que los parámetros obtenidos son insesgados, consistentes y eficientes, dentro de las restricciones propias del panel analizado.

RECOMENDACIONES

Entre las principales limitaciones del estudio se reconoce el tamaño reducido del panel (tres provincias, 13 períodos mensuales), lo que restringe la potencia estadística y la posibilidad de capturar efectos espaciales complejos. Asimismo, las series meteorológicas y de $PM_{2.5}$ corresponden a promedios provinciales, por lo que no reflejan variaciones intraurbanas significativas. Por esto, para investigaciones futuras se recomienda ampliar la escala temporal y espacial, incorporar indicadores sanitarios (capacidad hospitalaria, vacunación, movilidad) y aplicar modelos dinámicos de panel o espaciales (SAR, SEM, SDM) que permitan capturar interdependencias geográficas y efectos rezagados en el tiempo.

Dado que la mortalidad no se explicó plenamente por los factores incluidos, futuras investigaciones deberían integrar indicadores como la capacidad hospitalaria, tasas de vacunación, edad media de contagio, movilidad poblacional y acceso a servicios médicos, que podrían explicar mejor la severidad de los brotes.

Los resultados sugieren que la desigualdad y densidad urbana son determinantes cruciales en la propagación del COVID-19. Por tanto, el modelo puede adaptarse como instrumento de diagnóstico y planificación preventiva ante futuros eventos sanitarios, ayudando a focalizar intervenciones en territorios con alta vulnerabilidad estructural.

REFERENCIAS

- Ai, H., Nie, R., & Wang, X. (2022). Evaluation of the effects of meteorological factors on COVID-19 prevalence by the distributed lag nonlinear model. *Journal of Translational Medicine*, 20(1), Article 170. <https://doi.org/10.1186/s12967-022-03371-1>
- Almalki, A., Gokaraju, B., Acquaah, Y., & Turlapaty, A. (2022). Regression analysis for COVID-19 infections and deaths based on food access and health issues. *Healthcare*, 10(2), Article 324. <https://doi.org/10.3390/healthcare10020324>
- Andrades Grassi, J. E., & López Hernández, J. Y. (2015). Caracterización de los procesos espaciales y temporales y sus interrelaciones en estaciones de precipitación mensual en la zona centro occidental de la República Bolivariana de Venezuela. *GeoFocus. Revista Internacional de Ciencia y Tecnología de la Información Geográfica*, (16), 151–176. <https://www.geofocus.org/index.php/geofocus/article/view/375>
- Armijos-Toro, L. M., Castillo-Páez, S., & Medina-Vásquez, P. (2021). Association of sociodemographic factors with the evolution of COVID-19 infections: Ecuadorian case study. In M. Botto-Tobar, H. Cruz, & A. Díaz Cadena (Eds.), *Artificial intelligence, computer and software engineering advances* (pp. 71–81). Springer. https://doi.org/10.1007/978-3-030-68083-1_6
- Binaku, K., & Schmeling, M. (2017). Multivariate statistical analyses of air pollutants and meteorology in Chicago during summers 2010–2012. *Air Quality, Atmosphere & Health*, 10(10), 1227–1236. <https://doi.org/10.1007/s11869-017-0507-7>
- Castro, P., & De los Reyes, J. C. (2025). Autopsy of SARS-CoV-2 spread dynamics in Ecuador using data assimilation techniques: A tale of two provinces. *Mathematical Biosciences and Engineering*, 22(10), 2686–2719. <https://doi.org/10.3934/mbe.2025099>

- Cirakli, U., Dogan, I., & Gozlu, M. (2022). The relationship between COVID-19 cases and COVID-19 testing: A panel data analysis on OECD countries. *Journal of the Knowledge Economy*, 13(3), 1737–1750. <https://doi.org/10.1007/s13132-021-00792-z>
- Comisión Económica para América Latina y el Caribe. (2021, 4 de marzo). *Pandemia provoca aumento en los niveles de pobreza sin precedentes en las últimas décadas e impacta fuertemente en la desigualdad y el empleo*. <https://www.cepal.org/es/comunicados/pandemia-provoca-aumento-niveles-pobreza-sin-precedentes-ultimas-decadas-impacta>
- El-Hashash, E. F., & Shiekh, R. H. A. (2022). A comparison of the Pearson, Spearman rank and Kendall tau correlation coefficients using quantitative variables. *Asian Journal of Probability and Statistics*, 20(3), 36–48. <https://doi.org/10.9734/ajpas/2022/v20i3425>
- Elminir, H. K. (2005). Dependence of urban air pollutants on meteorology. *Science of the Total Environment*, 350(1–3), 225–237. <https://doi.org/10.1016/j.scitotenv.2005.01.043>
- Fondo para la Protección del Agua (FONAG) & Empresa Pública Metropolitana de Agua Potable y Saneamiento (EPMAPS). (2021). *Anuario hidrometeorológico 2020: Red integrada de monitoreo hidrometeorológico FONAG-EPMAPS* (N.º 14). <https://www.fonag.org.ec/web/wp-content/uploads/2024/06/Anuario-hidroclimatico-2020.pdf>
- Gambhir, E., Jain, R., Gupta, A., & Tomer, U. (2020). Regression analysis of COVID-19 using machine learning algorithms. In *2020 International Conference on Smart Electronics and Communication (ICOSEC)* (pp. 65–71). IEEE. <https://doi.org/10.1109/ICOSEC49089.2020.9215356>
- Gómez Manotoa, O. X. (2022). *The effect of air pollution, meteorology and human mobility on the mortality of COVID-19 in Ecuador* [Tesis de maestría, Dalarna University]. DiVA. <https://du.diva-portal.org/smash/record.jsf?pid=diva2%3A1639521>

- Guliyev, H. (2020). Determining the spatial effects of COVID-19 using the spatial panel data model. *Spatial Statistics*, 38, Article 100443.
<https://doi.org/10.1016/j.spasta.2020.100443>
- Guo, Y., Li, B., Duan, T., Yao, N., Wang, H., Yang, Y., Yan, S., Sun, M., Wang, L., Yao, Y., Sun, Y., Jia, J., & Liu, S. (2022). A panel regression analysis for the COVID-19 epidemic in the United States. *PLOS ONE*, 17(8), e0273344.
<https://doi.org/10.1371/journal.pone.0273344>
- Gutierrez, B., Márquez, S., Prado-Vivar, B., Becerra-Wong, M., Guadalupe, J. J., da Silva Candido, D., Fernandez-Cadena, J. C., Morey-Leon, G., Armas-Gonzalez, R., Andrade-Molina, D. M., Bruno, A., de Mora, D., Olmedo, M., Portugal, D., Gonzalez, M., Orlando, A., Drexler, J. F., Moreira-Soto, A., Sander, A. L., ... Cárdenas, P. (2021). Genomic epidemiology of SARS-CoV-2 transmission lineages in Ecuador. *Virus Evolution*, 7(2), veab051. <https://doi.org/10.1093/ve/veab051>
- Hehman, E., & Xie, S. Y. (2021). Doing better data visualization. *Advances in Methods and Practices in Psychological Science*, 4(4).
<https://doi.org/10.1177/25152459211045334>
- Hyrkäs-Palmu, H., Hugg, T. T., Jaakkola, J. J. K., & Ikäheimo, T. M. (2025). The influence of weather and urban environment characteristics on upper respiratory tract infections: A systematic review. *Frontiers in Public Health*, 13, Article 1487125.
<https://doi.org/10.3389/fpubh.2025.1487125>
- Instituto Nacional de Estadística y Censos. (2021). *Estadísticas vitales: Registro estadístico de defunciones generales de 2020*.
https://www.ecuadorencifras.gob.ec/documentos/web-inec/Poblacion_y_Demografia/Defunciones_Generales_2020/2021-06-10_Principales_resultados_EDG_2020_final.pdf
- Instituto Nacional de Estadística y Censos. (2022). *Estadísticas vitales: Registro estadístico de defunciones generales de 2021*.

https://www.ecuadorencifras.gob.ec/documentos/web-inec/Poblacion_y_Demografia/Defunciones_Generales_2021/Principales_resultados_EDG_2021_v2.pdf

Instituto Nacional de Estadística y Censos. (2024a). *Principales resultados de la Encuesta Nacional de Empleo, Desempleo y Subempleo – Anual 2024*.

https://www.ecuadorencifras.gob.ec/documentos/web-inec/EMPLEO/2024/anual/Principales-resultados_de_Mercado_Laboral_y_Pobreza_Anual_2024_final.pdf

Instituto Nacional de Estadística y Censos. (2024b). *Resultados nacionales definitivos:*

Censo 2022. https://www.censoecuador.gob.ec/wp-content/uploads/2024/05/Presentacion_Nacional_2da_entrega.pdf

Islam, M. M., & Noor, F. M. (2022). Correlation between COVID-19 and weather variables: A meta-analysis. *Heliyon*, 8(8), e10333. <https://doi.org/10.1016/j.heliyon.2022.e10333>

Jiang, F., Niu, Y., Shi, X., Yang, Y., Bao, J., Yuan, M., & Jia, X. (2026). Association between air pollution and acute upper respiratory infection morbidity: Effects of modification under ambient temperature and relative humidity. *Environmental Sciences Europe*, 38, Article 112. <https://doi.org/10.1186/s12302-026-01385-w>

Jundt, E., & Hu, X. (2025). Statistical models for the estimation of pH and aragonite saturation state in the Northwestern Gulf of Mexico. *Frontiers in Marine Science*, 12, Article 1621280. <https://doi.org/10.3389/fmars.2025.1621280>

Khan, M. A., Khan, R., Algarni, F., Kumar, I., Choudhary, A., & Srivastava, A. (2022). Performance evaluation of regression models for COVID-19: A statistical and predictive perspective. *Ain Shams Engineering Journal*, 13(2), Article 101574. <https://doi.org/10.1016/j.asej.2021.08.016>

McIntosh, K. (2023). COVID-19: Epidemiology, virology, and prevention. In M. S. Hirsch (Ed.), *UpToDate*. Wolters Kluwer. Recuperado el [05/02/2025], de <https://www.uptodate.com/contents/covid-19-epidemiology-virology-and-prevention>

- Mecenas, P., da Rosa Moreira Bastos, R. T., Rosário Vallinoto, A. C., & Normando, D. (2020). Effects of temperature and humidity on the spread of COVID-19: A systematic review. *PLOS ONE*, 15(9), e0238339. <https://doi.org/10.1371/journal.pone.0238339>
- Ministerio de Salud Pública del Ecuador. (2013). *Normas del Sistema Integrado de Vigilancia Epidemiológica: Norma técnica*. https://aplicaciones.msp.gob.ec/salud/archivosdigitales/documentosDirecciones/dnn/archivos/norma_sive.pdf
- Morris, D. H., Yinda, K. C., Gamble, A., Rossine, F. W., Huang, Q., Bushmaker, T., Fischer, R. J., Matson, M. J., van Doremalen, N., Vikesland, P. J., Marr, L. C., Munster, V. J., & Lloyd-Smith, J. O. (2020). *The effect of temperature and humidity on the stability of SARS-CoV-2 and other enveloped viruses* [Preprint]. bioRxiv. <https://doi.org/10.1101/2020.10.16.341883>
- Naciones Unidas. (2021, 4 de marzo). *América Latina es la región en desarrollo más afectada del mundo por la pandemia*. <https://news.un.org/es/story/2021/03/1489112>
- Núñez-Cifuentes, I. S., & Núñez-Freile, B. F. (2025). Guayas y Guayaquil: El epicentro de la pandemia por COVID-19 en el año 2020 en Sudamérica. Un análisis del exceso de mortalidad. *Revista Médica-Científica CAMbios HECAM*, 24(1), e947. <https://doi.org/10.36015/cambios.v24.n1.2025.947>
- Ogunjo, S., Olaniyan, O., Olusegun, C. F., Kayode, F., Okoh, D., & Jenkins, G. (2022). The role of meteorological variables and aerosols in the transmission of COVID-19 during Harmattan season. *GeoHealth*, 6(2), e2021GH000521. <https://doi.org/10.1029/2021GH000521>
- Organización Mundial de la Salud. (2024). *Panel de control de COVID-19 de la OMS*. Recuperado el [02/02/2025], de <https://data.who.int/dashboards/covid19/cases>
- Ortiz-Prado, E., & Fernández-Naranjo, R. (2020). Impacto de la COVID-19 en el Ecuador: De los datos inexactos a las muertes en exceso. *Revista Ecuatoriana de Neurología*, 29(2), 8–11. <https://doi.org/10.46997/revecuatneuro129200008>

- Ou, F. S., Le-Rademacher, J. G., Ballman, K. V., Adjei, A. A., & Mandrekar, S. J. (2020). Guidelines for statistical reporting in medical journals. *Journal of Thoracic Oncology*, 15(11), 1722–1726. <https://doi.org/10.1016/j.jtho.2020.08.019>
- Oyarzún G., M., & Valdivia C., G. (2021). Impactos en la salud de la contaminación del aire. *Revista Chilena de Enfermedades Respiratorias*, 37(2), 103–106. <https://doi.org/10.4067/S0717-73482021000200103>
- Parra, M., & Carrera, E. (2021). Evolución de la COVID-19 en Ecuador. *Investigación y Desarrollo*, 13(1), 27–40. <https://doi.org/10.31243/id.v13.2020.1002>
- Peña Benitez, J. J. (2022). *Modelización estadística multivariante para identificar variables atmosféricas y sociodemográficas, asociados a la COVID-19 en la ciudad de Guayaquil* [Tesis de maestría, Escuela Superior Politécnica del Litoral]. DSpace en ESPOL. <http://www.dspace.espol.edu.ec/handle/123456789/56499>
- Rahmadina, F., Ahmad, R. A., & Ramadana, A. L. (2026). Association of particulate matter (PM_{2.5}) with COVID-19 infection and mortality in low- and middle-income Asian countries: A systematic review and meta-analysis. *Journal of Preventive Medicine and Public Health*, 59(1), 12–24. <https://doi.org/10.3961/jpmph.25.499>
- Rath, S., Tripathy, A., & Tripathy, A. R. (2020). Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 14(5), 1467–1474. <https://doi.org/10.1016/j.dsx.2020.07.045>
- Samaila, B., & Muhammad Kalgo, Z. (2024). Impact of atmospheric parameters on human respiratory tract infections (RTIs). *International Journal of Environment, Engineering and Education*, 6(2), 57–72. <https://doi.org/10.55151/ijeedu.v6i2.137>
- Sangkham, S., Thongtip, S., & Vongruang, P. (2021). Influence of air pollution and meteorological factors on the spread of COVID-19 in the Bangkok Metropolitan Region and air quality during the outbreak. *Environmental Research*, 197, Article 111104. <https://doi.org/10.1016/j.envres.2021.111104>

- Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia & Analgesia*, 126(5), 1763–1768.
<https://doi.org/10.1213/ANE.0000000000002864>
- Secretaría Técnica del Comité Nacional de Límites Internos. (2020). *Informe de rendición de cuentas 2019*. Ministerio de Gobierno. <https://www.ministeriodegobierno.gob.ec/wp-content/uploads/2020/11/Informe-de-Rendici%C3%B3n-de-Cuentas-2019.pdf>
- Sheppard, N., Carroll, M., Gao, C., & Lane, T. (2023). Particulate matter air pollution and COVID-19 infection, severity, and mortality: A systematic review and meta-analysis. *Science of the Total Environment*, 880, Article 163272.
<https://doi.org/10.1016/j.scitotenv.2023.163272>
- Shreffler, J., & Huecker, M. R. (2023). Exploratory data analysis: Frequencies, descriptive statistics, histograms, and boxplots. In *StatPearls*. StatPearls Publishing.
<https://www.ncbi.nlm.nih.gov/books/NBK557570/>
- Singhal, T. (2020). A review of coronavirus disease-2019 (COVID-19). *Indian Journal of Pediatrics*, 87(4), 281–286. <https://doi.org/10.1007/s12098-020-03263-6>
- Servicio Nacional de Gestión de Riesgos y Emergencias. (2020). *Situación nacional por COVID-19 (coronavirus)*.
- Srivastava, A. (2021). COVID-19 and air pollution and meteorology—An intricate relationship: A review. *Chemosphere*, 263, Article 128297.
<https://doi.org/10.1016/j.chemosphere.2020.128297>
- Sun, Z., Thilakavathy, K., Kumar, S. S., He, G., & Liu, S. V. (2020). Potential factors influencing repeated SARS outbreaks in China. *International Journal of Environmental Research and Public Health*, 17(5), Article 1633.
<https://doi.org/10.3390/ijerph17051633>
- Thongpan, I., Vongpunsawad, S., & Poovorawan, Y. (2020). Respiratory syncytial virus infection trend is associated with meteorological factors. *Scientific Reports*, 10(1), Article 10931. <https://doi.org/10.1038/s41598-020-67969-5>

- Trujillo Montalvo, P. (2024). Ecuador: Social inequality, violence and pandemics. *International Journal of Health Science*, 4(11), 2–8.
<https://doi.org/10.22533/at.ed.1594112423011>
- Velásquez Serra, G. C., Mera Flores, L. N., Preciado Velásquez, O. E., & Barrera Reyes, C. G. (2020). Determinantes sociales de la salud y la virulencia del COVID-19 en Guayaquil. *Pro Sciences: Revista de Producción, Ciencias e Investigación*, 4(37), 111–128. <https://doi.org/10.29018/issn.2588-1000vol4iss37.2020pp111-128>
- Wagatsuma, K., Feurer, D., Yu, W., Xu, R., Riffe, T., Kniffka, M. S., Acosta, E., Armstrong, B., Mistry, M., Lowe, R., Royé, D., Hashizume, M., Tobias, A., Vicedo-Cabrera, A. M., Madaniyazi, L., Ng, C. F. S., Íñiguez, C., Ragettli, M. S., Lavigne, E., ... Sera, F. (2026). The joint impact of temperature, humidity, and air pollution on COVID-19 incidence: A multi-country time-series study in 439 cities. *Environment International*, 208, Article 110090. <https://doi.org/10.1016/j.envint.2026.110090>
- Wang, J., Tang, K., Feng, K., & Lv, W. (2020). *High temperature and high humidity reduce the transmission of COVID-19* [Preprint]. SSRN. <https://doi.org/10.2139/ssrn.3551767>
- World Meteorological Organization. (2021). *Guide to instruments and methods of observation (WMO-No. 8): Volume IV, space-based observations*.
<https://community.wmo.int/site/knowledge-hub/programmes-and-initiatives/instruments-and-methods-of-observation-programme-imop/guide-instruments-and-methods-of-observation-wmo-no-8>
- Zheng, H.-L., Guo, Z.-L., Wang, M.-L., Yang, C., An, S.-Y., & Wu, W. (2021). Effects of climate variables on the transmission of COVID-19: A systematic review of 62 ecological studies. *Environmental Science and Pollution Research*, 28(39), 54299–54316. <https://doi.org/10.1007/s11356-021-15929-5>