Universidad Nacional Mayor de San Marcos
Universidad del Perú. Decana de América

Dirección General de Estudios de Posgrado
Facultad de Ciencias Matemáticas
Unidad de Posgrado

Análisis espacial de la tasa de las variantes del SARS-CoV-2 en Perú, 2021-2022

TESIS

Para optar el Grado Académico de Magíster en Bioestadística

AUTOR

Mariangelica Carolina MONTOYA HUAMAN

ASESOR

Mg. Emma Norma CAMBILLO MOYANO

Lima, Perú

2025



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

https://creativecommons.org/licenses/by-nc-sa/4.0/

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Cambillo, E. (2025). *Análisis espacial de la tasa de las variantes del SARS-CoV-2 en Perú, 2021-2022*. [Tesis de maestría, Universidad Nacional Mayor de San Marcos, Facultad de Ciencias Matemáticas, Unidad de Posgrado]. Repositorio institucional Cybertesis UNMSM.

Metadatos complementarios

Datos de autor	
Nombres y apellidos	Mariangelica Carolina Montoya Huaman
Tipo de documento de identidad	DNI
Número de documento de identidad	47729749
URL de ORCID	https://orcid.org/0009-0005-6662-0004
Datos de asesor	
Nombres y apellidos	Emma Norma Cambillo Moyano
Tipo de documento de identidad	DNI
Número de documento de identidad	15377390
URL de ORCID	https://orcid.org/0000-0003-3173-9425
Datos del jurado	
Presidente del jurado	
Nombres y apellidos	Luis Alberto Navarro Huamani
Tipo de documento	DNI
Número de documento de identidad	06445826
Miembro del jurado 1	
Nombres y apellidos	Misael Erikson Maguiña Palma
Tipo de documento	DNI
Número de documento de identidad	41974596
Miembro del jurado 2	
Nombres y apellidos	Edith Seier Zúñiga
Tipo de documento	DNI
Número de documento de identidad	07941184
Miembro del jurado 3	
Nombres y apellidos	Emma Norma Cambillo Moyano
Tipo de documento	DNI
Número de documento de identidad	15377390

Datos de investigación	
Línea de investigación	A.3.2.3. Bioestadística
Grupo de investigación	Ciencias Matemáticas para las Ciencias de la Vida - CMATVIDA
Agencia de financiamiento	Sin financiamiento.
Ubicación geográfica de la investigación	Universidad Nacional Mayor de San Marcos País: Perú Departamento: Lima Provincia: Lima Distrito: Lima Latitud: -12.05642315 Longitud: -77.08433269016213
Año o rango de años en que se realizó la investigación	Julio 2022 – junio 2024
URL de disciplinas OCDE	Estadísticas, Probabilidad https://purl.org/pe-repo/ocde/ford#1.01.03

UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
Universidad del Perú, DECANA DE AMÉRICA

FACULTAD DE CIENCIAS MATEMÁTICAS
VICEDECANATO DE INVESTIGACIÓN Y POSGRADO
UNIDAD DE POSGRADO

ACTA DE SUSTENTACIÓN DE TESIS PARA OPTAR EL GRADO ACADÉMICO DE MAGÍSTER

Siendo las .18:00 horas del día jueves siete de agosto de dos mil veinticinco, en el Aula 102 de la Facultad de Ciencias Matemáticas de la Universidad Nacional Mayor de San Marcos, el Jurado Examinador de tesis, presidido por el Dr. Luis Alberto Navarro Huamani e integrado por los siguientes miembros: Dr. Misael Erikson Maguiña Palma (Jurado Informante), Dra. Edith Seier Zúñiga (Jurado Externo) y la Mg. Emma Norma Cambillo Moyano como Miembro Asesora; se reunieron para la sustentación de la tesis titulada: «ANÁLISIS ESPACIAL DE LA TASA DE LAS VARIANTES DEL SARS-COV-2 EN PERÚ, 2021-2022», presentada por la Bachiller Mariangelica Carolina Montoya Huaman, para optar el Grado Académico de Magíster en Bioestadística.

Concluida la exposición, los miembros del Jurado de Tesis procedieron a formular sus preguntas que fueron absueltas por la graduanda; acto seguido se procedió a la evaluación correspondiente, según tabla adjunta, habiendo obtenido la Bachiller **MARIANGELICA CAROLINA MONTOYA HUAMAN**, el calificativo de .18. .MUY..BUENO.....

Habiendo sido aprobada la sustentación de la Tesis, el Jurado Examinador recomienda que el Consejo de Facultad apruebe el otorgamiento del **Grado Académico de Magíster en Bioestadística** a la Bachiller **MARIANGELICA CAROLINA MONTOYA HUAMAN.**

Siendo las ..19:10. horas, se levantó la sesión, firmando para constancia la presente Acta:

Dr. Luis Alberto Navarro Huamani
Presidente

Dr. Misael Erikson Maguiña Palma
Miembro

Dra. Edith Seier Zúñiga
Miembro Externo

Mg. Emma Norma Cambillo Moyano
Miembro Asesor

UNIVERSIDAD NACIONAL MAYOR DE
SAN MARCOS
Universidad del Perú, Decana de América

CERTIFICADO DE SIMILITUD

Yo, **Emma Norma Cambillo Moyano** en mi condición de asesor acreditado con la Resolución Decanal N° 001023-2022-D-FCM/UNMSM y Resolución Decanal N° 000297-2025-D-FCM/UNMSM (cambio de título) de fecha 06 de febrero de 2025, de la tesis, cuyo título es: "**ANÁLISIS ESPACIAL DE LA TASA DE LAS VARIANTES DEL SARS-COV-2 EN PERÚ, 2021-2022"**, presentada por la bachiller **Mariangelica Carolina Montoya Huaman**, para optar el grado académico de Magíster en Bioestadística, CERTIFICO que se ha cumplido con lo establecido en la Directiva de Originalidad y de Similitud de Trabajos Académicos, de Investigación y Producción Intelectual. Según la revisión, análisis y evaluación mediante la plataforma PIDS, el documento evaluado cuenta con el **13** % de similitud, nivel **PERMITIDO** para continuar con los trámites correspondientes y para su **publicación en el repositorio institucional.**

Se emite el presente certificado en cumplimiento de lo establecido en las normas vigentes, como uno de los requisitos para la obtención del grado académico de Magíster.

Lima, 09 de abril de 2025

Firma del Asesor

Nombres y Apellidos del asesor: **Emma Norma Cambillo Moyano**

DNI: **15377390**

Dedicado a

La memoria de mi amada abuela María Domitila

Agradecimientos

Deseo expresar mi profundo agradecimiento a la Universidad Nacional Mayor de San Marcos, que me permitió ser alumna de su prestigiosa y renombrada institución, a la Facultad de Ciencias Matemáticas y a los profesores que formaron parte de la plana docente de la maestría de Bioestadística, que con sus enseñanzas me condujeron a la culminación de la maestría; este agradecimiento está dirigido principalmente a mi asesora de tesis Mg. Emma Norma Cambillo Moyano, por su dedicación, su orientación, diligente apoyo y paciencia que me daban el impulso para continuar con la ejecución de esta investigación; sin su apoyo y conocimientos este trabajo no hubiera sido posible.

Expreso también mi sincera gratitud al grupo de investigación CMATVIDA, la experiencia y conocimiento de cada uno de sus miembros, constituyeron una motivación para continuar en el camino de la investigación. A cada miembro del grupo, gracias por compartir sus conocimientos y por las discusiones académicas que permitieron continuar con el desarrollo de mi proyecto.

No podría culminar este agradecimiento sin mencionar a mi amada familia, a mi esposo por alentarme en los momentos de dificultad y por confiar en mí, a mi madre por ser mi ejemplo y por motivarme siempre a alcanzar mis metas y a mis tíos por su cariño y apoyo incondicional.

ANÁLISIS ESPACIAL DE LA TASA DE LAS VARIANTES DEL SARS-COV-2 EN PERÚ, 2021-2022

Disertación presentada a la Universidad Nacional Mayor de San Marcos
para optar el grado de Magíster en Bioestadística
En la Facultad de Ciencias Matemáticas

2025

Bachiller: Mariangelica Carolina Montoya Huaman
Asesora: Emma Cambillo Moyano

ÍNDICE GENERAL

RESUMEN X
ABTRACT XI
CAPÍTULO 1: INTRODUCCIÓN 1
1.1. Situación Problemática 1
1.2. Formulación del Problema 3
Problema General 3
Problemas Específicos 3
1.3. Justificación teórica 3
1.4. Justificación práctica 4
1.5. Objetivos 5
Objetivo general 5
Objetivos específicos 5
CAPÍTULO 2: MARCO TEÓRICO 6
2.1. Antecedentes de investigación 6
Internacionales 6
Nacionales 13
2.2. Bases Teóricas 13
Análisis Espacial 13
Conceptos previos vinculados al Análisis Espacial 14
Análisis exploratorio de datos espaciales 16
Efectos espaciales 22
Matriz de pesos espaciales 25

Indicadores de autocorrelación espacial 27
Indicadores globales 27
Indicadores Locales 31
SARS-CoV-2 35
CAPÍTULO 3: METODOLOGÍA 38
Tipo y Diseño de Investigación 38
Unidad de Análisis 38
Población de Estudio 38
Descripción de la base de datos 38
Proceso de filtración de la base de datos y descripción de las variables ... 39
CAPÍTULO 4: RESULTADOS Y DISCUSIÓN 42
4.1. Análisis, interpretación y discusión de resultados 42
Análisis Exploratorio 42
Análisis Exploratorio de Datos Espaciales 47
CONCLUSIONES 70
RECOMENDACIONES 71
REFERENCIAS BIBLIOGRÁFICAS 73
ANEXOS 85

LISTA DE TABLAS

Tabla 1 18
Tabla 2 41
Tabla 3 45
Tabla 4 51
Tabla 5 54
Tabla 6 58
Tabla 7 63
Tabla 8 67

LISTA DE FIGURAS

Figura 1 15
Figura 2 19
Figura 3 20
Figura 4 21
Figura 5 22
Figura 6 24
Figura 7 26
Figura 8 40
Figura 9 43
Figura 10 44
Figura 11 47
Figura 12 49
Figura 13 53
Figura 14 57
Figura 15 62
Figura 16 66

RESUMEN

La alta circulación del virus SARS-CoV-2 ha dado lugar a mutaciones, ocasionando nuevas variantes, algunas de estas con características distintas, mejores adaptadas a su entorno. La presente investigación tiene por objeto analizar el comportamiento espacial de las variantes de SARS-CoV-2 predominantes en el Perú en el periodo de 01 de enero 2021 hasta el 30 de junio 2022, a partir de la aplicación de metodologías del análisis espacial. Con esta finalidad se realizó un estudio descriptivo, se utilizaron los datos del resultado de Linaje Genómico de secuenciamiento genético del COVID-19 notificados por el Instituto Nacional de Salud (INS). Se identificaron las variantes predominantes y se aplicaron técnicas gráficas y estadísticas del análisis estadístico convencional y del análisis exploratorio de datos espaciales (AEDE), a través del software SPSS v.25 y la herramienta de software libre Geoda v. 1.20 respectivamente. La aplicación de los índices de autocorrelación global de Moran y C de Geary permitieron identificar autocorrelación espacial estadísticamente significativa para las variantes Gamma, Delta y Ómicrom. Los mapas elaborados a partir de los estadísticos Locales permitieron apreciar diferencias en el comportamiento de cada variante del virus a través de las provincias del Perú. Finalmente, considerar el espacio geográfico en el análisis de la dinámica de la evolución de las variantes del virus permitió identificar el predominio de una variante frente a otras en determinadas zonas; además brindó otra perspectiva en comparación con aquellas investigaciones donde no se considera el espacio geográfico identificando que cada variante del SARS-CoV-2 presentó un patrón de distribución diferente a través de las provincias del Perú.

Palabras clave: SARS-CoV-2, análisis espacial, Perú

ABTRACT

The high circulation of the SARS-CoV-2 virus has given rise to mutations, causing new variants, some of these with different characteristics, better adapted to their environment. The purpose of this research is to analyze the spatial behavior of the predominant SARS-CoV-2 variants in Peru in the period from January 1, 2021 to June 30, 2022, based on the application of spatial analysis methodologies. For this purpose, a descriptive study was carried out, using data from the Genomic Lineage result of genetic sequencing of COVID-19 reported by the National Institute of Health (INS), identifying the predominant variants and applying graphical and statistical techniques of conventional statistical analysis and exploratory spatial data analysis (AEDE), using SPSS v.25 software and the free software tool Geoda v. 1.20 respectively.

The application of Moran's and Geary's global autocorrelation indices allowed the identification of statistically significant spatial autocorrelation for the Gamma, Delta and Ómicrom variants. The maps produced from Local statistics made it possible to appreciate differences in the behavior of each variant of the virus across the provinces of Peru. Finally, considering geographical space in the analysis of the dynamics of the evolution of the virus variants made it possible to identify the predominance of one variant over others in certain areas; It also provided another perspective in comparison with those studies where geographical space is not considered, identifying that each variant of SARS-CoV-2 presented a different distribution pattern across the provinces of Peru.

Keywords: SARS-CoV-2, spatial analysis, Peru

CAPÍTULO 1: INTRODUCCIÓN

1.1. Situación Problemática

El SARS-CoV-2 es un virus de la familia de los coronavirus que generalmente atacan al sistema respiratorio pudiendo causar neumonía grave e incluso la muerte. Según la OMS la propagación del virus se da a través de pequeñas partículas líquidas emitidas de la boca o nariz de una persona infectada cuando esta tose, estornuda, habla, canta o respira (Organización Mundial de la Salud [OMS], 2022a).

La alta circulación del virus ha dado lugar a mutaciones, ocasionando nuevas variantes, algunas de estas con características distintas, mejores adaptadas a su entorno, con mayor transmisibilidad y virulencia que el virus original; todo esto representaría un reto para las medidas que buscan mitigar la pandemia, reduciendo la sensibilidad a tratamientos o vacunas (Chen, et al*.,* 2021; Vargas-Herrera et al*.*, 2021; Aleem, et al*.*, 2023). Hasta la fecha se han reportado en el mundo diversas variantes, las cuales han venido siendo estudiadas y vigiladas por la OMS desde enero de 2020.

El primer caso importado de SARS-CoV-2 en Perú fue confirmado el 6 de marzo del 2020 (Plataforma digital única del Estado Peruano, 2023). El gobierno del Perú declaró el estado de emergencia y aislamiento social obligatorio el 16 de marzo del mismo año (Presidencia del Consejo de Ministro, Decreto Supremo N° 044-2020, 2020). A pesar de las medidas tomadas la rápida diseminación del virus ha sido inevitable y con ello la presencia de nuevas variantes de SARS-CoV-2 con diferentes características que el virus original.

Siendo el SARS-CoV-2 un virus ARN, es más susceptible a cambios en el material genético en comparación con aquellos virus de ADN, a pesar de ello la diversidad viral depende también de los procesos ecológicos y demográficos.

(Sanjuán y Domingo-Calap, 2021) En consecuencia es importante resaltar la relevancia de la epidemiologia espacial tomado como primer ejemplo de un avistamiento incipiente de estas disciplinas el estudio del Dr. John Snow sobre la epidemia del cólera en Londres en 1954, que descubrió que la enfermedad se transmitía a través del agua potable al observar la agrupación de muertes alrededor de una bomba de agua pública. Por tanto, es preciso anotar la importancia de la geografía de las enfermedades infecciosas considerándola fundamental para el estudio de su propagación; principalmente en organismos que se transmiten de persona a persona y se diseminan rápidamente a través del territorio (Wilson, 2017).

Mientras el virus siga circulando se reportarán, seguramente, nuevas variaciones del SARS-CoV-2. El Instituto Nacional de Salud (INS), a través de la Plataforma Nacional de Datos Abierto, ha puesto a disposición datos de la secuenciación de los genomas del SARS-CoV-2, constituyendo un gran paso para la investigación de la pandemia.

Por otro lado, PARVIN, et al. (2021), indicaron que la variante Lambda predominó a nivel nacional durante la segunda ola, a pesar de la circulación paralela de otras dos variantes preocupantes: Alfa y Gamma. Asimismo, señalaron que la variante Lambda podría tener características fenotípicas que pueden brindar ventajas frente a las infecciones causadas por otras variantes de preocupación.

Vargas-Herrera, et al. (2021), atribuyeron la expansión y predominio de la variante Lambda en Perú, durante la segunda ola, frente a la variante Gamma por “efecto fundador” y por mutaciones responsables de mayores tasas de infectividad y capacidad de evadir el sistema inmunológico; sin embargo, resaltaron el predominio de la variante Gamma en la cuenca norte del Amazonas adyacente a la frontera con Brasil debido a su ubicación geográfica y mayor seroprevalencia de SARS-CoV-2 en la región (70%).

A nivel internacional Taboada, et al. (2021), caracterizaron la distribución geográfica y temporal de los linajes del SARS-CoV-2 durante el primer año de la pandemia en México y hallaron regiones que exhibieron mayor diversidad de

linajes, mientras que otras regiones presentaron menor diversidad con predominio de un determinado linaje.

No se hallaron investigaciones a nivel nacional que aborden el análisis espacial del SARS-CoV-2. Un enfoque espacial de la expansión de las variantes del SARS-CoV-2, abordado en esta investigación, permite una compresión del predominio de una variante frente a otras considerando el espacio geográfico como un factor importante para el análisis de la dinámica de la evolución de las variantes del virus en el Perú, identificando áreas claves de la crisis sanitaria.

1.2. Formulación del Problema

Problema General

¿Cuál es el patrón en la distribución espacial de la tasa de las variantes del SARS-CoV-2 en las provincias del Perú?

Problemas Específicos

¿Cuál es el patrón en la distribución espacial de la tasa de la variante Lambda del SARS-CoV-2 en las provincias del Perú?

¿Cuál es el patrón en la distribución espacial de la tasa de la variante Gamma del SARS-CoV-2 en las provincias del Perú?

¿Cuál es el patrón en la distribución espacial de la tasa de la variante Delta del SARS-CoV-2 en las provincias del Perú?

¿Cuál es el patrón en la distribución espacial de la tasa de la variante Ómicrom del SARS-CoV-2 en las provincias del Perú?

1.3. Justificación teórica

La producción científica alrededor de este tema ha sido muy vertiginosa. A nivel internacional algunos investigadores han abordado el tema haciendo uso de metodologías de análisis espacial. A nivel nacional el Centro Nacional de Planeamiento Estratégico publicó en el 2020 un informe sobre la distribución

espacial de los casos confirmados y fallecidos por COVID-19 en el Perú, relacionándolos con variables económicas, sociales y ambientales (Centro Nacional de Planeamiento Estratégico [CEPLAN], 2020); sin embargo, dicha investigación no consideró la aplicación de las herramientas del análisis espacial, lo cual se propone en esta investigación para el hallazgo de dependencia y patrones espaciales además del análisis del virus según la ocurrencia de las variantes.

Por otra parte, el estudio del entorno geográfico ha cobrado mayor relevancia a partir del desarrollo de los sistemas de información geográfica (SIG) (Waller, 2004). A pesar que su uso es aun limitado en el campo de la investigación en salud los SIG son una valiosa herramienta para el análisis de diversos eventos en salud.

1.4. Justificación práctica

Las mutaciones del SARS COV-2, en ocasiones, pueden modificar el fenotipo del virus y conllevar al aumento de su transmisibilidad, escape inmunológico o riesgo de reinfección (Taboada, et al., 2021; Aleem, et al., 2023). Evaluar los cambios en la frecuencia de las variantes en las provincias del territorio nacional es importante para determinar el ritmo de la variabilidad y propagación del virus en el país, además de su impacto en la gravedad de la enfermedad y en las medidas de contención.

Analizar la distribución de las variantes del SARS-CoV-2 empleando la estadística espacial, contribuye al entendimiento de la dinámica de la distribución de las variantes del virus a nivel nacional y brinda otra perspectiva en comparación a aquellas investigaciones donde no consideran el espacio geográfico, de este modo ayudar a las autoridades de salud en la elaboración de programas de contención del virus priorizando las provincias más vulnerables del país. Además de ello esta investigación marca un precedente para el empleo de la estadística espacial para la comprensión de brotes de otras enfermedades en donde interese el espacio geográfico.

Por lo expuesto en líneas anteriores en esta investigación se planteó analizar el comportamiento espacial de las variantes de SARS-CoV-2, a partir de la aplicación de metodologías del análisis espacial.

1.5. Objetivos

Objetivo general

Analizar el patrón en la distribución espacial de las variantes del SARS-CoV-2 predominantes en las provincias del Perú.

Objetivos específicos

Analizar el patrón en la distribución espacial de la variante Lambda del SARS-CoV-2 en las provincias del Perú.

Analizar el patrón en la distribución espacial de la variante Gamma del SARS-CoV-2 en las provincias del Perú.

Analizar el patrón en la distribución espacial de la variante Delta del SARS-CoV-2 en las provincias del Perú.

Analizar el patrón en la distribución espacial de la variante Ómicrom del SARS-CoV-2 en las provincias del Perú.

CAPÍTULO 2: MARCO TEÓRICO

2.1. Antecedentes de investigación

Los antecedentes considerados en esta investigación involucran el análisis de datos espaciales aplicados a los contagios de la enfermedad por COVID-19; sin embargo investigaciones halladas hasta el momento no discuten la expansión de las variantes del virus desde la perspectiva del análisis espacial; cabe recalcar que resulta interesante conocer la dinámica espacial de cada variante ya que se ha evidenciado diferencias en la patogenicidad de la enfermedad debido a las mutaciones ocurridas en el material genético del virus (Giovanetti, et al., 2021).

Internacionales

Amdaoud et al. (2021), con el objetivo de identificar los mecanismos de propagación de la pandemia del COVID-19, priorizando la dimensión socioeconómica, realizaron un análisis exploratorio de autocorrelaciones espaciales empleando el Índice global y local de Moran y modelos de regresión espacial aplicados a datos de mortalidad en 125 regiones de 12 países europeos, el análisis exploratorio mostró valores del estadístico I de Moran altamente significativos, indicando fuerte autocorrelación espacial positiva y la identificación de clústeres ; además hallaron que la proporción de personas mayores en la población, el PIB per cápita, la distancia de los objetivos de la Unión Europea (indicador sintético que incorpora la información contenida en un conjunto de indicadores sociales) y la tasa de desempleo se correlacionan positivamente con las altas tasas de mortalidad por COVID-19, por otro lado el número de médicos y camas de hospital y el nivel de confianza social (proporcionada por la Encuesta Social Europea) se correlacionan negativamente con las tasas de mortalidad por COVID-19. Los autores concluyeron que la propagación de la epidemia puede explicarse por una confluencia de factores que describen el contexto socioeconómico de las regiones europeas.

Aral y Bakir (2021), realizaron un Análisis espacio - temporal con el objetivo de determinar los factores que afectan la propagación de la pandemia mediante el análisis espacial de los casos de COVID-19 en Turquía, identificaron las diferencias a nivel provincial de COVID-19 a través del análisis exploratorio de datos. En el análisis de datos espaciales emplearon matrices de ponderación espacial, aplicando la contigüidad de tipo reina, para determinar las interacciones de las localizaciones estudiadas. Se utilizó la autocorrelación espacial global y local mediante el Índice de Moran para determinar la existencia de la asociación espacial de infecciones por COVID-19 y métodos de estimación de modelos de regresión espacial para hallar la relación geográfica de los posibles factores que influyen en el número de casos de COVID-19. Los resultados mostraron que los valores se concentraron en regiones con autocorrelación positiva y se negó la hipótesis de aleatoriedad. El análisis de autocorrelación local identificó conglomerados espaciales. Los modelos de regresión OLS y SLM (el cual proporcionó resultados más efectivos según el AIC) mostró que la densidad de población y la proporción de ancianos son relevantes para explicar el modelo. Además, hallaron que los casos de COVID-19 se ven influenciados por las provincias vecinas. Los autores concluyeron que la ubicación geográfica es un factor importante en la investigación de los factores que determinan la infección por COVID-19.

Das y Sudaryo (2021), en su investigación tuvieron como objetivo identificar las tendencias epidemiológicas y la distribución espacial de la pandemia COVID-19 en cinco municipios de DKI Yakarta en Indonesia. Los autores analizaron las estadísticas de la tasa de mortalidad, tasa de incidencia y tasa de letalidad y la distribución de los casos antes y después de la aplicación de las restricciones sociales a gran escala. Para el análisis espacial emplearon el Índice de Moran Global, se evaluó mediante permutaciones, finalmente, utilizaron el análisis LISA mediante el mapa de clusteres. Los resultados mostraron una tendencia epidemiológica creciente, siendo el centro de Yakarta el municipio más afectado. El Índice de Moran Global obtenido (I = 0.0268 y p valor p = 0.01) mostró una autocorrelación espacial positiva. Una totalidad de 107 aldeas urbanas fueron

significativas (p- valor <0.05). El Índice de Moran local mostró la presencia de clústeres, los puntos críticos (regiones con valores muy altos) se situaron en las áreas periféricas de la ciudad. El estudio concluyó en la efectividad de las restricciones sociales a gran escala sobre la reducción de los contagios en DKI Yakarta.

Munazza et al. (2021), presentaron un artículo cuyo objeto de estudio fue analizar espacialmente las tasas de incidencia de COVID 19 en el año 2020, mediante el análisis de la agrupación espacial e identificación de puntos calientes y clústeres atípicos a escala mundial; para ello recopilaron datos a partir de los informes emitidos por la Organización Mundial de la Salud, calcularon la tasa bruta de incidencia por cada mes para cada país utilizando las estimaciones de la población de cada país del sitio web Worldometer a mediados de julio del 2020. Analizaron la tasa bruta de la distribución mediante mapas como complemento al análisis estadístico espacial. Para analizar la agrupación espacial y la detección de valores atípicos de conglomerados espaciales de la incidencia de la enfermedad utilizaron el Índice Global de Moran y el Índice Local de Moran respectivamente. Las variaciones de los Hotspots (puntos calientes) y Coldspots (puntos fríos) se analizaron mediante el Índice G* de Getis-Ord; las puntuaciones Z y los valores P obtenidos dan cuenta de la agrupación de valores altos rodeados por vecinos con valores altos (puntuaciones Z positivas altas significativas) o la agrupación de valores bajos rodeados por vecinos con valores bajos (puntuaciones negativas Z altas significativas). Los resultados obtenidos del análisis de las tasas brutas grafican el avance de la pandemia desde China al resto del mundo, convirtiéndola a finales de marzo en una pandemia que se había extendido a 146 países. A inicios del segundo trimestre de la pandemia identificaron una elevada tasa bruta en Europa y América (>100 casos/100.000 habitantes), posteriormente se sumó la región del Mediterráneo oriental; este escenario se mantuvo durante el periodo de estudio, los autores atribuyen estas variaciones a diversos factores, principalmente a la tasa de transmisión y a las condiciones propias de cada país. El análisis de puntos calientes evidenció en febrero un único punto caliente importante en el Pacifico Occidental, siendo China

el más significativo (nivel de confianza del 99%), en los siguientes meses el punto caliente se desplazó principalmente hacia Europa y América (nivel de confianza del 95% al 99%).A partir de Julio los países del Mediterráneo oriental también se convirtieron en puntos calientes y África se mantuvo como punto frio durante todo el año. El análisis de conglomerados y valores atípicos a través del I Local de Moran identificó a fines de febrero un clúster de alta incidencia en China en el Pacifico Occidental; situación que cambio al mes siguiente identificando clúster de alta incidencia en Europa sumándosele en los siguientes meses países como Canadá, Arabia Saudí e Irak; a fines del segundo trimestre Sudamérica se identificó como el único conglomerado de alta incidencia, posteriormente , en los meses de julio y noviembre los países de la región Mediterránea Oriental se agruparon en conglomerados de alta incidencia junto con Europa y América, situación que se repite a finales del año para la mayoría de los países en Europa. Por otro lado; el valor de I Moran Global más alto, de 0.32 con una puntuación Z de 14.05, refleja el patrón altamente agrupado de la incidencia de la enfermedad en el mes de diciembre, al contrario del mes de junio con un valor de I de Moran de 0.02 y una puntuación Z de 1.85 que se traduce en una menor agrupación de la enfermedad en dicho mes del año. Los autores concluyen el estudio indicando que los hallazgos representan una oportunidad para el análisis más a fondo desde un enfoque ecologista de la enfermedad.

Parvin et al. (2021), en su estudio realizado en India, plantearon como objetivo identificar los principales puntos críticos del COVID-19, a nivel distrital, considerando como factores decisivos de la pandemia los casos confirmados notificados del COVID-19, la población total, la densidad de población y la llegada de turistas extranjeros a la India, para ello emplearon el Índice de Moran Global para verificar autocorrelación espacial. Para el análisis de clústeres atípicos se utilizó el Índice de Anselin Local Moran. Así mismo, para identificar grupos de locaciones con valores similares estadísticamente significativos, realizaron un análisis de Hot Spot del COVID-19 utilizando el estadístico Getis-Ord G_i^*. Este estadístico por lo general se utiliza como complemento a la información que proporciona el Estadístico Local de Moran, ya que G_i* es capaz de detectar

vecinos con valores extremos. Los resultados de la autocorrelación espacial mostraron que los cuatro factores considerados para el análisis estaban agrupados y los resultados fueron estadísticamente significativos (p valor <0.01). El resultado de la estadística Getis-Ord G_i^* mostró que la población total y los casos de COVID-19 notificados tienen una mayor puntuación Z (>3 y > 0,7295 respectivamente); valores de Z >0 y significativos indican agrupamiento de valores similares elevados alrededor de la unidad de estudio. Concluyeron que las áreas con una alta proporción de población se relacionan con un alto número de casos de COVID-19 en los distritos de la India.

Quinino et al. (2021), realizaron un estudio cuyo objetivo principal fue investigar la evolución espacial y temporal de la incidencia de COVID-19 y los factores asociados a la pandemia en los municipios del estado brasileño de Pernambuco, para ello elaboraron mapas a partir del coeficiente de incidencia suavizado. Estudiaron la correlación entre el coeficiente de incidencia de los casos de COVID-19 y factores socioeconómicos (población, densidad de poblacional, IDH-M, Índice de Gini, número total de establecimientos económicos de todo tipo y tasa de urbanización). Se calculó el índice de Moran global y el Indicador Local de Asociación Espacial para identificar agrupaciones espaciales significativas. Los investigadores identificaron una correlación positiva significativa (5% de significancia) entre el coeficiente de incidencia, la población y la densidad de población, y una correlación negativa entre la tasa de urbanización y el IDH-M para el área metropolitana. En el interior del estado de Pernambuco, la incidencia (361,29 / 100.000 habitantes) se correlacionó con una mayor población, densidad demográfica y tasa de urbanización, y un área más pequeña (p <0,05). La asociación espacial obtenida de la prueba de Moran Global resultó significativa, con el índice de coeficientes suavizados igual a 0,679 (p <0,05). Los autores concluyen que en las zonas de mayor desplazamiento existe mayor riesgo de transmisión de la enfermedad.

Sarkar et al. (2021), con el objetivo de identificar los factores de riesgo demográfico responsables de la transmisión de COVID-19 en Bangladesh

mediante modelos espaciales y estadísticos, realizaron una investigación considerando el número de casos confirmados de COVID-19 por mil habitantes a nivel distrital además de otras nueve variables demográficas explicativas. Los investigadores utilizaron las estadísticas Índice de Moran local univariante y bivariante para el análisis de patrones espaciales, además utilizaron el estadístico G Getis - Ord para identificar el área agregada con mayor precisión. Se emplearon diferentes técnicas de regresión espacial (modelo de retraso espacial y modelo de error espacial) y técnicas no espaciales (modelo clásico). El resultado del Índice de Moran univariante fue 0.261 (a un nivel de significancia de 5%) demostrando, de este modo, autocorrelación espacial positiva de los casos de COVID-19. El estadístico Índice de Moran bivariante determinó que dos parámetros, la densidad de población y número de trabajadores industriales, eran significativos en un intervalo de confianza del 95%; mientas que Índice local de Moran permitió observar los conglomerados significativos de valores similares y disimiles y G Getis – Ord los puntos calientes y fríos. Así mismo el modelo de error espacial resultó ser más eficiente para modelar el cual determinó que la densidad de población influyó en el número de casos confirmados de COVID-19. Se concluyó que la zona central de Bangladesh (Dhaka y distritos adyacentes) es más vulnerable al COVID-19 debido a su mayor densidad de población.

El Deeb (2021), con el objetivo de estudiar la autocorrelación espacial de infecciones de COVID-19 registradas en los 26 distritos administrativos del Líbano, realizó un estudio espacial mediante el Índice univariado de Moran considerando seis modelos diferentes de parametrización, adyacencia (caso I), proximidad (caso II), población (Caso III), densidad de población (caso IV), tasa de pobreza (caso V) y densidad de pobreza (caso VI), además propuso un análisis temporal de la infección. El análisis mediante Índice de Moran determinó autocorrelación espacial positiva para las variables en estudio, a excepción de la tasa de pobreza. El autor concluyó que la tasa de pobreza no es estadísticamente relevante para la propagación espacial de la enfermedad, mientras que la frontera geográfica, la distancia entre los centros de los distritos, el número y la densidad

de residentes y la densidad de la pobreza se agrupan espacialmente con diferente grado de intensidad.

Kim et al. (2020), los investigadores se plantearon como objetivo medir la asociación espacial en los casos de COVID-19 en China Continental en las primeras etapas de la pandemia, para ello utilizaron el estadístico Índice de Moran aplicando seis tipos de modelos; los cuales consideraban la adyacencia de fronteras y las distancias entre dos provincias (modelo 1 y 2), población y densidad de población (modelo 3 y 4), número de médicos y camas de hospitales o centros médicos (modelo 5 y 6). Los resultados mostraron que la mayoría de los modelos, excepto los modelos basados en la atención médica, indicaron una asociación espacial significativa de infecciones por COVID-19. Los autores concluyeron que la asociación espacial fue la clave para la propagación espacial durante las primeras etapas de la pandemia de COVID-19 en China continental.

Kim y Castro (2020), llevaron a cabo un estudio espacio temporal en Corea del Sur. El objetivo del estudio fue evaluar la agrupación, el patrón y duración de la enfermedad a nivel distrital. Utilizaron el Índice Moran global para evaluar la correlación espacial de la incidencia acumulada de COVID-19 por cada 100 000 habitantes a nivel distrital, la significación estadística se evaluó mediante simulación de Monte Carlo. Así mismo realizaron una exploración del espacio-tiempo retrospectivo basado en el modelo de Poisson discreto para analizar la ocurrencia de conglomerados. Los resultados mostraron una fuerte autocorrelación espacial en todo el país, con un Índice de Moran global de 0.784 ($p = 0,0001$). Se identificaron doce grupos espaciotemporales estadísticamente significativos mediante la estadística de exploración de espacio-tiempo. El patrón espacial de los conglomerados cambió disminuyendo en tamaño y duración a través del periodo estudiado. Como conclusión del estudio los autores refirieron que las estrategias de contención de la COVID-19 de Corea del Sur resultó eficaz tanto en la detección temprana como en la mitigación de los contagios.

Nacionales

CEPLAN (2020), presentó un informe con el objetivo de investigar la distribución espacial de la COVID-19 en el Perú, para ello consideraron los casos confirmados de infectados y fallecidos, relacionándolas con variables económicas, sociales y físico- ambientales. El análisis se limitó a la presentación de cartografías y análisis de correlación, de este modo hallaron que existe una correlación positiva (0,61) entre la densidad poblacional y los distritos con mayor número de contagiados. Por otro lado, hallaron una correlación negativa (-0,25) entre la zona con un número alto de casos confirmados por COVID-19 y la altitud, de este modo concluyeron que la propagación del COVID-19 es mayor en las ciudades del Perú con poblaciones grandes.

Como se mencionó al inicio del capítulo, las investigaciones citadas se vinculan con la investigación planteada, ya que involucran el análisis de datos espaciales para una mejor interpretación de la dinámica de la COVID-19, en este caso, a diferencia de las anteriores, se estudiará las variantes del virus para averiguar el patrón de la ocurrencia de estas a través de las provincias del Perú.

A la fecha no se registran investigaciones que aborden desde la perspectiva del análisis espacial las variantes del virus; los antecedentes hallados hasta el momento estudian los contagios y las muertes provocados por el virus; sin embargo, es de conocimiento que las mutaciones del virus producen cambios en la patogenicidad del virus (Giovanetti, et al., 2021); por ello resulta interesante conocer la dinámica espacial de cada variante.

2.2. Bases Teóricas

Análisis Espacial

En palabras de Anselin y Rey (2010) el análisis espacial de datos ocupa la atención de la estadística y la econometría, así como de muchos campos de aplicación.

El análisis estadístico de datos implica la suposición de independencia entre los datos; sin embargo, los valores asignados a una variable que se observan a lo largo de un territorio guardan relación entre las unidades geográficas en estudio, esto implica la necesidad de métodos que tomen en cuenta esta condición.

Conceptos previos vinculados al Análisis Espacial

Sistemas de Información Geográfica. El desarrollo de los sistemas de Información Geográfica, conocidos como GIS por sus siglas en inglés, ha permitido el desarrollo de las investigaciones de análisis de datos espaciales (Anselin y Rey ,2010). En particular el desarrollo de softwares para el análisis de datos espaciales ha permitido la visualización, organización y manipulación de datos espaciales que hacen posibles la aplicación de herramientas del análisis espacial (Anselin y Getis, 1992).

Datos espaciales. Un dato espacial define a un atributo que está vinculado a una ubicación geográfica. Según Olaya (2020) existen diferentes enfoques teóricos que representan la realidad correspondiente a la variable bajo estudio en una capa que recoge dicha información de forma precisa y pueda ser empleada dentro del SIG.

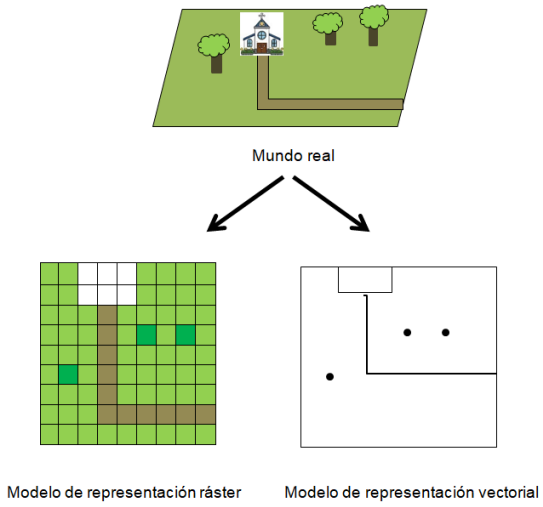
El autor refiere dentro de los modelos de representación o modelos de datos el modelo ráster y modelo vectorial (Figura 1).

El primero de ellos corresponde a una división completa del espacio en unidades mínimas denominadas celdas, su estudio corresponde a la geoestadística. Generalmente son datos inherentes a un espacio continuo. Este tipo de dato es común cuando se trata de la observación de variables ambientales por ejemplo podría medirse las precipitaciones, la elevación de un terreno, las temperaturas e incluso la contaminación atmosférica de áreas previamente definidas. La técnica más común utilizada para este tipo de datos implica la predicción espacial mediante kriging.

Por su lado, el modelo vectorial define al espacio como una serie de elementos geométricos que tienen un valor o atributo asociado y se representan mediante puntos, líneas y polígonos.

Figura 1

Representacion de la realidad según el modelo ráster y el modelo vectorial.



Cabe destacar la importancia de identificar el tipo de dato con el cual se trabaja ya que de ello dependerá la correcta elección de la técnica estadística espacial a emplear. A continuación, se abordará los tipos de datos que corresponden al modelo vectorial, ya que en esta investigación se hace uso de este tipo de datos.

Puntos espaciales: este tipo de datos señala la ubicación geográfica de un evento puntual (coordenadas). Algunas de las técnicas utilizadas para su estudio son el análisis de cuadrantes, análisis de vecinos más cercanos, la función K de Ripley. Debido a que no son materia de esta investigación no se profundizará su estudio.

Líneas: resultan de la unión de varios puntos, pueden representar objetos del espacio tales como ríos, carreteras, líneas ferroviarias, etc.

Polígonos: también denominados datos de área o lattice, este tipo de datos representan datos agregados en una determinada área geográfica, esta área puede ser regular o irregular (cuando se trata de áreas con delimitaciones político

–administrativas). Es importante tener en cuenta la escala a la que se analizan los datos debido a que a mayor área es menor la uniformidad de la población y los fenómenos estudiados; sin embargo, considerar áreas pequeñas conlleva a que el estimador fluctúe aleatoriamente. Estos datos son los más empleados para analizar temas asociados al área de la salud (Andrade, et al., 2007).

Fisher y Wang (2011 citado en Zymanzik, 2014) añaden a esta clasificación un cuarto tipo de dato, los datos de flujo origen – destino; este tipo de datos están asociado dos ubicaciones de puntos o áreas geográficas. Son interesantes para mediciones donde importe el flujo de personas, por ejemplo, entre el hogar y el trabajo o partida y llegada de personas en un aeropuerto.

En este estudio se trabajó con datos de área representadas por polígonos, los polígonos de cada provincia se georeferenciaron con los ubigeos disponibles en la base de datos de las variantes de SARS-CoV-2.

Análisis exploratorio de datos espaciales

Las observaciones georreferenciadas requieren de herramientas apropiadas para su análisis, el análisis exploratorio de datos espaciales, denominado AEDE por sus siglas, constituye la primera etapa de todo análisis que implique datos georreferenciados.

La naturaleza de los datos espaciales puede invalidar las interpretaciones de los métodos del análisis exploratorio de datos tradicional, los cuales se basan en la suposición de independencia de los datos, ello ha llevado al desarrollo de métodos especializados para el análisis exploratorio de datos espaciales (Anselin, 1996).

Según Anselin (1994, citado en Anselin, 1996), el AEDE comprende un conjunto de técnicas que permiten “describir y visualizar distribuciones espaciales, identificar ubicaciones atípicas (valores atípicos espaciales), descubrir patrones de asociación espacial (agrupaciones espaciales) y sugerir diferentes regímenes

espaciales y otras formas de inestabilidad espacial o no estacionariedad espacial".

Dentro de las principales técnicas del AEDE se puede diferenciar entre técnicas de representación de distribuciones espaciales y las técnicas de visualización del fenómeno de asociación espacial, que puede ser global, local o multivariante, además de estos métodos existe un grupo de técnicas que permiten visualizar el efecto de heterogeneidad espacial (Chasco,2003).

Tabla 1

Técnicas de visualización del análisis exploratorio de datos espaciales (AEDE)

Visualización de distribuciones espaciales	Univariante	-Diagrama/Mapa de cajas
	Multivariante	-Diagrama dispersión-caja
Visualización de autocorrelación espacial	Global	-Gráfico de retardo espacial - Diagrama/mapa de dispersión de Moran
	Local	- Puntos atípicos en el diagrama de dispersión de Moran - Mapas LISA - Diagrama de caja LISA
	Multivariante	-Diagrama de dispersión multivariante de Moran
Heterogeneidad Espacial		- Mapa del histograma de Frecuencias - Diagrama de Dispersión

Fuente: Chasco (2003).

Técnicas de visualización de distribuciones espaciales. Entre estas figuran los mapas de cajas, las cuales permiten visualizar los valores atípicos superiores e inferiores mediante la extensión de un mapa de coropletas con la identificación de estos valores en un diagrama de cajas. Por otro lado, el diagrama de dispersión/caja permite ilustrar dos variables en un diagrama de dispersión, de igual forma se representan los diagramas de cajas correspondiente a cada variable (Chasco, 2003).

Técnicas de visualización de la autocorrelación espacial. Entre las técnicas que permiten observar el fenómeno de autocorrelación espacial a nivel global están el gráfico de retardo espacial y los gráficos de dispersión de Moran (Chasco, 2003).

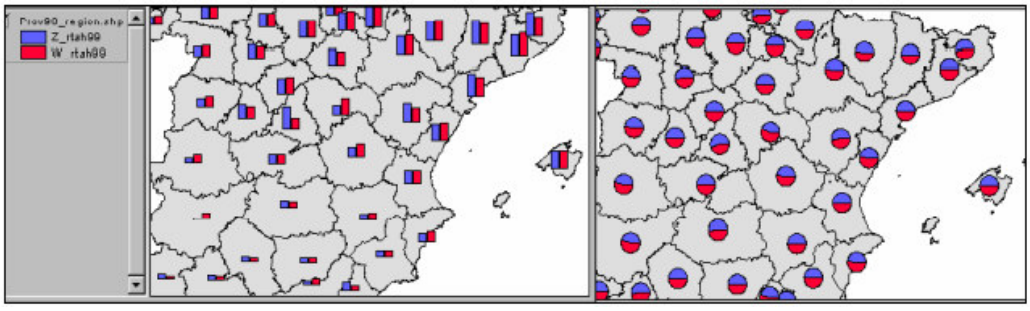
Chasco (2003) considera dentro de esta categoría a los mapas de contigüidades espaciales; sin embargo, esta herramienta se considerará para efectos de esta investigación como un paso previo para determinar los vecinos espaciales de las unidades geográficas.

Los gráficos de retardo espacial permiten visualizar la similitud o diferencia entre el valor observado de una ubicación y su retardo espacial para una matriz de ponderaciones espaciales dada, “la asociación entre una variable y su rezago espacial se puede visualizar por medio de los llamados gráficos circulares de retraso espacial y gráficos de barras de retraso espacial” (Anselin, 1996).

Chasco (2003) presenta un ejemplo ilustrativo de este tipo de gráficos utilizando como variable la renta disponible per cápita de una selección de provincias españolas (Figura 2).

Figura 2

Gráfico de retardo espacial del PBI per cápita en provincias españolas.



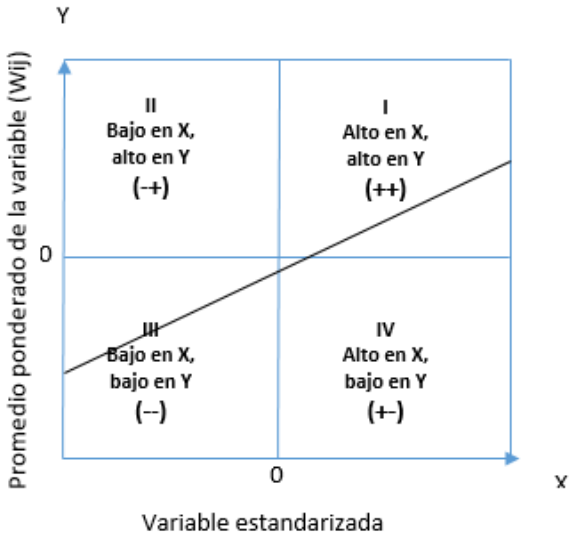
Fuente. Chasco (2003)

Así mismo, se puede emplear el gráfico de dispersión de Moran para ilustrar el efecto de autocorrelación. Según Celemín (2009) y Chasco (2010), el gráfico de dispersión de Moran es una técnica del AEDE que permite visualizar las unidades espaciales. Su configuración consta de cuatro cuadrantes como muestra la Figura 3, en el eje de abscisas aparecen los valores estandarizados de la variable (Yi) y en el eje de ordenadas su retardo espacial ($W_{ij}Y_j$).En el cuadrante I y III le corresponde la situación de dependencia espacial positiva, se ubican las unidades espaciales con valores superiores o inferiores a la media rodeados por unidades vecinas con valores similares (alto-alto o bajo-bajo). En contraparte los cuadrantes II y IV registran a las categorías de dependencia espacial negativa, las cuales corresponden a unidades espaciales con vecinos que registran valores opuestos (alto-bajo o bajo alto).

En la Figura 4 se ilustra la aplicación del grafico de dispersión de Moran a través de un ejemplo propuesto por Chasco (2003) aplicado al PBI per cápita a nivel provincial en España.

Figura 3

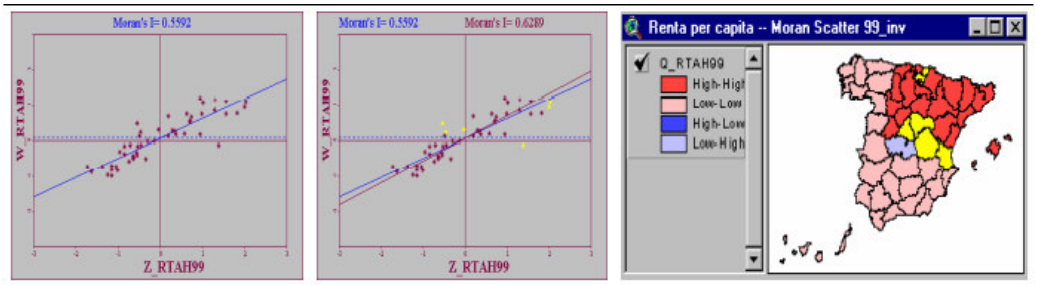
Diagrama de dispersión de Moran.



Fuente. Chasco (2003)

Figura 4

Diagrama y mapa de dispersión de Moran del PBI per cápita a nivel provincial en España.



Fuente. Chasco (2003)

En el gráfico de dispersión de Moran se puede también identificar puntos atípicos permitiendo la visualización de autocorrelación espacial a nivel local. Por otro lado, los Mapas LISA permite visualizar las ubicaciones de los puntos del diagrama de dispersión de Moran, donde cada cuadrante identificado es representado por un color que permite evaluar la asociación espacial local (Anselin, 1996). Así mismo el diagrama de caja LISA permite visualizar la distribución del I local de Moran (Chasco, 2003).

Por otro lado, el diagrama de dispersión multivariante de Moran es útil cuando el interés es visualizar el efecto de autocorrelación espacial de dos indicadores (Chasco, 2003).

Técnicas de visualización de heterogeneidad espacial. Chasco (2003) presentan dos técnicas del AEDE para representar la ausencia de estabilidad de la variable en el espacio, estos son el mapa del histograma de frecuencias y el diagrama de dispersión.

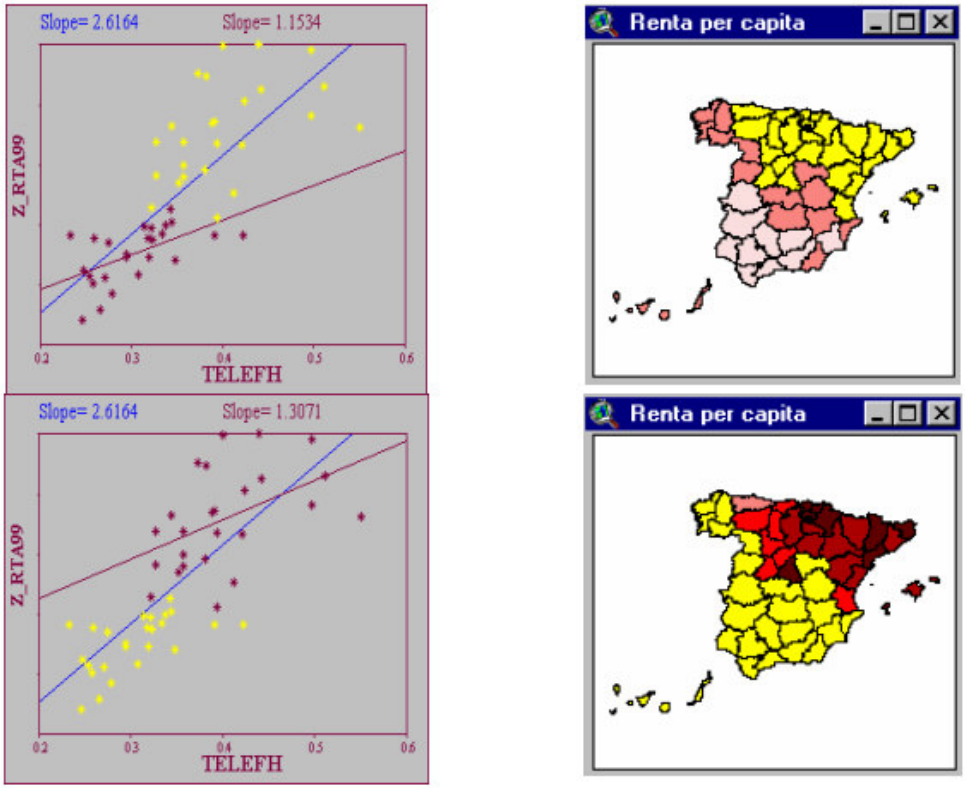
El primero de ellos es un gráfico dinámico que permiten seleccionar las unidades espaciales que corresponden a intervalos de un histograma anexado a un mapa, el cual permite elegir el número de intervalos más conveniente para comprensión del comportamiento de una variable. Por otro lado, el diagrama de dispersión permite observar las variaciones del valor del coeficiente estimado de la variable

independiente de un modelo de regresión lineal, seleccionando un número determinado de unidades espaciales con respecto al total de las unidades bajo estudio (Chasco, 2003).

Chasco (2003) ilustra este último con la variable número de teléfonos por habitante con y la renta per cápita en las provincias españolas (Figura 5).

Figura 5

Diagrama de dispersión de la variable teléfonos por habitante vs renta disponible por habitante de las provincias españolas en 1999.



Fuente. Chasco (2003)

Efectos espaciales

Los efectos espaciales son características propias de los datos georreferenciados, por lo que requieren de técnicas estadísticas espaciales propias del AEDE. Los efectos espaciales citados en la literatura son la

dependencia espacial o autocorrelación espacial y la heterogeneidad espacial (Acevedo y Velásquez, 2008).

***Heterogeneidad espacial*.** Son las variaciones de los efectos de dependencia espacial (concepto que se desarrolla en los párrafos siguientes) o bien de la relación entre las variables bajo estudio (Anselin y Rey ,2010).

Es decir, la heterogeneidad espacial comprende por un lado a la inestabilidad estructural de los datos, debido a que las unidades espaciales estudiadas poseen diferentes características en cuanto a tamaño de área, composición demográfica, social entre otros factores; y por otro la heterocedasticidad, que proviene de los errores de medición de las variables (Acevedo y Velásquez,2008).

***Autocorrelación espacial o dependencia espacial*.** Según Getis (2010) a finales de la década de 1950 en la universidad de Washintong surge el concepto de autocorrelación espacial por Michael F. Dacey, cabe mencionar que anterior a ello diversos investigadores sentaron las bases sobre el principio de proximidad, este concepto se puede resumir en los aportes del geógrafo Tobler, quien formuló la primera ley de la geografía: "todo está relacionado con todo lo demás, pero las cosas cercanas están más relacionadas que las cosas distantes".

Hasta mediados de la década de los sesenta se usaban términos como dependencia espacial, asociación espacial, interacción espacial o interdependencia espacial. El término de autocorrelación espacial se formaliza en un marco estadístico con los aportes de Cliff y Ord en 1969, investigadores de la Universidad de Bristol en Gran Bretaña (Celemin, 2009; Getis, 2010).
Siguiendo a Acevedo y Velásquez (2008) señalan que la autocorrelación espacial puede considerarse como un fenómeno y una técnica estadística mientras que la dependencia espacial es la explicación teórica de la misma; sin embargo, también hay referencias de otros autores como Chasco (2003) que consideran ambos términos como intercambiables. En esta investigación se adopta esta última posición.

A partir de estas aclaraciones conceptuales se puede ahondar en el concepto de autocorrelación espacial. En palabras de Anselin (1988), la autocorrelación espacial es la falta de independencia que a menudo está presente en el análisis de datos transversales.

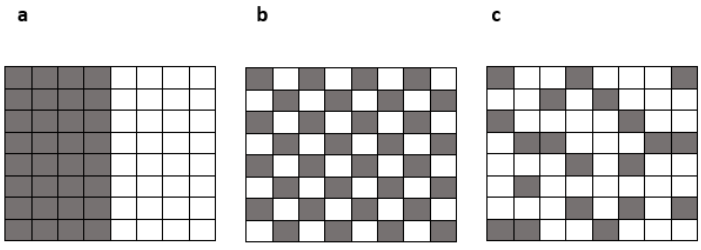
Según autores como Chasco (2003) y Acevedo y Velásquez (2008) la autocorrelación refleja el grado de concentración o dispersión de los valores de una variable en unidades geográficas próximas.

Este efecto puede ser de signo positivo, negativo o nulo. En el primer caso las unidades geográficas están próximas a zonas con valores similares, en la segunda situación las unidades geográficas están próximas a otras zonas con valores disimiles, el efecto es nulo cuando la variable se distribuye de manera aleatoria (Chasco, 2010).

La Figura 6 ilustra los diferentes patrones según el tipo de autocorrelación observado sobre el espacio.

Figura 6

Patrones de autocorrelación espacial, (a) Autocorrelación espacial positiva, (b) Autocorrelación espacial negativa, (c) patrón aleatorio.



La autocorrelación espacial establece relaciones multidireccionales de las variables a lo largo de las unidades espaciales, las cuales poseen diferentes características, para resolver este problema es necesario tener presente el concepto de vecindad geográfica el cual está representado mediante las fronteras (si las unidades geográficas están representadas por polígonos), o

distancias (si las unidades bajo estudio están representadas por puntos). El concepto de vecindad espacial se formaliza mediante la matriz de pesos espaciales (W) (Chasco, 2010).

Matriz de pesos espaciales

La matriz de pesos espaciales (W) permite expresar las interdependencias espaciales, esta matriz es simétrica y está formada por tantas filas y columnas como número de unidades geográficas estudiadas. A cada fila se le asigna valores según el criterio de vecindad adoptado para relacionar las unidades geográficas. La forma más común para construir esta matriz de contigüidades es la notación binaria; asignándole 0 a las unidades espaciales que no son vecinas y 1 a las unidades espaciales que sí lo son, según sea el criterio de vecindad escogido por el investigador (Acevedo y Velásquez,2008; Chasco,2010).

Sean $Y_1, Y_2, \dots, Y_k$ variables aleatorias y sea S la ubicación de la unidad geográfica bajo observación. La matriz generada es la siguiente:

$$\left(\begin{array}{cccc|c} y_1(1) & y_2(1) & \dots & y_k(1) & s(1) \\ y_1(2) & y_2(2) & \dots & y_k(2) & s(2) \\ \vdots & \vdots & \vdots & & \vdots \\ y_1(n) & y_2(n) & \dots & y_k(n) & s(n) \end{array}\right)$$

La cual puede ser representada de la siguiente forma:

$$\{y_1(i),, y_2(i) \dots y_k(i) | s(i)\}_{i=1\dots n}$$

Donde,

y_k es el valor de datos de la variable $Y_k(k = 1, \dots, K)$,

(i) observación de un punto, área, polígono o zona espacial en particular. Para cada caso, $i = 1, \dots, n$,

$s(i)$ representa la ubicación del objeto espacial.

Además, una unidad geográfica no es vecina consigo misma, por lo que la diagonal de la matriz quedará formada por ceros, de esta forma:

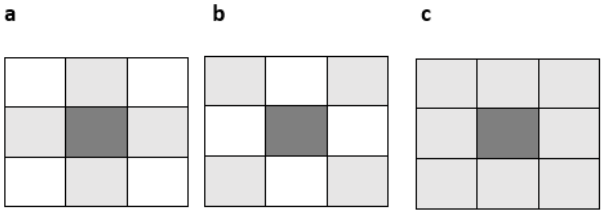
$wij = 0, \ \forall \ i = j$

Como se señaló anteriormente existen diferentes formas de establecer el criterio de vecindad entre las unidades espaciales (fronteras o distancias), de esta forma queda definida por el investigador con base a su conocimiento sobre las unidades geográficas y el fenómeno estudiado. El criterio de adyacencia es la más empleada y es la que se aplicará en esta investigación.

Estas fronteras están determinadas por diversos criterios, para simplificar su explicación se plantea su similitud a los movimientos de las fichas en un tablero de ajedrez. De este modo se tiene, contigüidad de tipo torre, según las unidades compartan un lado en común con la región de interés; contigüidad de alfil, cuando las unidades comparten un vértice en común; y contigüidad tipo reina para aquellas regiones con un lado y vértice en común (Acevedo y Velásquez, 2008; Chasco ,2003), en la Figura 7 se muestra de manera gráfica la disposición de las unidades vecinas según el criterio escogido.

Figura 7

Contigüidad espacial basada en diferentes criterios de vecindad.



(a) Contigüidad de torre, (b) contigüidad de alfil, (c) contigüidad de reina.
Fuente. Chasco (2003).

Usualmente a la matriz W se le realiza una operación de estandarización de forma que cada elemento i de una fila j sea dividido por la suma de pesos de dicha fila (Chasco, 2003).

$$\frac{w_{ij}}{\sum_{j\epsilon j} w_{ij}}$$

Siendo j el grupo de regiones relacionadas con i.

Cabe señalar que la suma resultante de cada fila de la matriz estandarizada es igual a la unidad, por tanto, la suma de todos los pesos espaciales resultará en el tamaño de muestra (Chasco, 2003).

Indicadores de autocorrelación espacial

Además de la visualización del comportamiento de una variable es necesario contrastar el grado y significación del efecto espacial (Chasco, 2010). Existen varias herramientas propias del AEDE e indicadores estadísticos que se pueden emplear en el análisis en función de los objetivos de la investigación.

Para medir la autocorrelación espacial se emplean distintos tipos de estadísticos o índices que posibilitan establecer una relación de la dependencia entre las ubicaciones y los valores o atributos de las variables

Indicadores globales

El concepto de autocorrelación espacial tomó importancia a partir de los estudios realizados por diversos autores a inicios de la década de 1950 como Moran y Geary (Bivand y Wond, 2018). Los mismos que formularon dos de los indicadores más usados para el cálculo de la autocorrelación espacial. La aplicación de estos índices se extiende hasta la actualidad. La forma general de estos índices está dada por (Getis, 1991):

$$\Gamma_{ij} = \sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij} Y_{ij}$$

Dónde:

Γ, es una medida de autocorrelación para n observaciones georeferenciadas,

n, es el total de lugares del mapa,

Wij , una matriz de pesos espaciales,

Yij , es una medida de la distancia de los valores de las unidades i y j .

Estadístico Índice de moran (I de Moran). El estadístico I de Moran, fue establecido por Moran en 1950, para identificar el grado de dependencia espacial de una variable en determinadas zonas (Chasco, 2010; Celemín,2009).

De acuerdo a Chasco (2010), este estadístico es expresado en función de una variable (Y) en desviaciones a la media, observadas en las unidades geográficas (i, j), y los elementos de una matriz de pesos espaciales W y está dada por la siguiente expresión:

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad , i \neq j$$

Donde

w_{ij}: es el elemento de la matriz de pesos espaciales W de las unidades geográficas (i, j),

S_0=$\sum_i \sum_j W_{ij}$: es la suma de los elementos de la matriz de pesos espaciales,

$\bar{y}$: es el valor promedio de la variable Y,

n : número total de unidades en estudio,

Cuando esta matriz está estandarizada por filas se obtiene lo siguiente:

$$I = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} W *_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Siendo $W * ij$ el elemento de la matriz de pesos espaciales estandarizada por filas.

Siguiendo a Chasco (2010) el I de Moran es una adaptación del coeficiente de correlación de Pearson con la diferencia que se considera en el cálculo de los valores una matriz de pesos espaciales. El I de Moran resulta en valores numéricos entre -1 (dispersión perfecta) y 1 (autocorrelación perfecta).

Al calcular el producto de la diferencia de las observaciones de dos zonas con la media general $\bar{y}$ se determina hasta dónde varían las observaciones

conjuntamente. Si tanto y_i como y_j están al mismo lado de la media, este producto es positivo; si por el contrario, están ubicados a lados diferentes de la media, el producto es negativo y el tamaño absoluto del valor resultante depende de qué tan cercanos sean los valores observados a la media general. (Acevedo y Velásquez, 2008, p23).

El estadístico I de Moran interpretado como pendiente de una línea de regresión es lo que permite visualizar la relación lineal entre una variable y su retardo espacial en forma de diagrama de dispersión (Chasco, 2010).

A continuación de la estimación de la pendiente, es necesario saber en qué grado es estadísticamente significativo este valor (Chasco, 2010).

La hipótesis nula plantea la aleatoriedad, de modo que, si la significancia resulta mayor al p-valor se rechaza la hipótesis de aleatoriedad lo cual permite afirmar que la configuración espacial no se produce de manera aleatoria. (Celemín 2009; Chasco, 2010). En este sentido, la media teórica del I de Moran es el cociente:

$$E(I) = -\frac{1}{(n-1)}$$

Tal como se muestra el valor esperado es negativo y varía de acuerdo al número de unidades observadas. Se puede observar que esta media tiende a cero cuando el tamaño de la muestra es grande (Chasco, 2010).

Por otro lado, la varianza y la desviación normal estándar quedan definidas bajo la siguiente forma (Bivand y Wond, 2018):

$$Var(I) = E(I^2) - [E(I)]^2$$

$$Z(I) = \frac{I - E(I)}{\sqrt{Var(I)}}$$

Siguiendo a Celemín (2009) otra alternativa para contrastar la aleatoriedad involucra un proceso conocido como aleatorización a través del método de permutación de los datos observados en las unidades espaciales, es decir se

intercambian los datos de las ubicaciones aleatoriamente obteniendo valores distintos de autocorrelación que son comparados con el valor obtenido.

Estadístico* C *de Geary. C de Geary es una medida menos utilizada que el I de Moran. Este se basa en las desviaciones en las respuestas de cada observación entre sí y está dado por (Bivand y Wond, 2018):

$$C = \frac{n-1}{2s_0}\frac{\sum_i \sum_j w_{ij}(x_i - x_j)^2}{\sum_i (x_i - \bar{x})^2}$$

Donde,

n, es el número de observaciones para la variable x,

$\bar{x}$, es la media de la variable x ,

w_{ij} , es el peso basado en la distancia dado a cada par (i,j), y está dado por $w_{ij} = {1}\Big/{\sqrt{(p-k)^2 + (q-l)^2}}$, donde (p,q) es la ubicación de la coordenada de la $i\ésima$ observación y (k,l) es la posición de la $j\ésima$ observación,

S_0 , es la suma de los pesos y está dado por $S_0 = \sum_i \sum_j w_{ij}$

El valor de C de Geary se encuentra entre 0 y 2. Los valores cercanos a 0 muestran un aumento positivo de autocorrelación espacial, valores cercanos a 2 indican incremento de la autocorrelación espacial negativa, bajo la hipótesis nula de aleatoriedad espacial con media y varianza constante el valor esperado del estadístico es el siguiente (Waller y Gotway, 2004):

$$E(C) = 1$$

La varianza y la desviación normal estándar, bajo el supuesto de normalidad, están definidas por (Cliff y Ord, 1981, citado en Bivand y Wong, 2018):

$$\text{Var} = \frac{(2S_1 + S_2)(n-1) - 4S_0^2}{2(n+1)S_0^2}$$

Donde,

S_0 , es la suma de los pesos y está dado por $S_0 = \sum_i \sum_j w_{ij}$

$S_1 = \frac{1}{2} \sum_{(2)} (wij + wji)^2$

$S_2 = \sum_{i=1}^{n} (\sum_{j=1}^{n} wij + \sum_{j=1}^{n} wji)^2$

$$Z(C) = \frac{E(C) - C}{\sqrt{Var(C)}}$$

Para ambos estadísticos se prefiere la derivación analítica es más fácil de interpretar, pero puede ser inconveniente ya que muchas veces el análisis de datos espaciales incumple los criterios subyacentes a una distribución normal. Anselin (2020) nombra a estos p valores resultantes como pseudo p valores y su cálculo está dado por la siguiente expresión:

$$p = \frac{R + 1}{M + 1}$$

Siendo R el número de permutaciones igual o más extremo que el estadístico observado y M el número de permutaciones. En esta investigación se utiliza el enfoque de permutaciones, con 999 permutaciones para el cálculo de los p valores.

Indicadores Locales

Las estadísticas globales permiten medir la magnitud de la autocorrelación espacial de la variable analizada a través de un área completa de estudio; sin embargo, surgió la necesidad de profundizar en el conocimiento de dependencias espaciales locales, que no se visualizan con las estadísticas globales mediante el uso de medidas que permitan detectar este fenómeno en cada una de las unidades espaciales que compone la totalidad del área en estudio. Según Ord (2006), las primeras investigaciones se remontan a finales de la década de los 70 con la monografía de Getis y Boots, quienes estaban interesados en el análisis de procesos espaciales puntuales; sin embargo fue hasta principios de los 90, según el propio Ord, que Getis se contactó con él para discutir sobre medidas locales de autocorrelación espacial que el autor había desarrollado; fue de este

modo que en 1992 Getis y Ord plantearon el uso del estadístico G_i para identificar el comportamiento local de la variable bajo estudio (Getis y Ord, 1992; Ord, 2006).Posteriormente, en 1995 Anselin desarrolla las versiones locales de estadísticos globales formulados por Moran y Geary bajo la denominación de estadísticos LISA(Chasco, 2003; Ord, 2006). Estos estadísticos fueron desarrollados por Anselin con el objetivo de identificar focos locales de no estacionariedad o puntos calientes (como G_i de Getis y Ord) y, además de ello, identificar valores atípicos, cabe señalar que la suma de todos los valores locales es proporcional al valor del estadístico global de Moran y Geary (Anselin, 1995).

A continuación, se detalla cada uno de los estadísticos mencionados.

Estadístico local de Getis y Ord $\mathrm{G_i}$**.** Este estadístico se calcula para cada entidad de un conjunto de datos. Mide la relación de la suma de los valores en ubicaciones vecinas dentro de un radio de distancia d u orden de continuidad definida con respecto al punto original y se compara con la suma de todos los valores del conjunto de datos. De este modo el estadístico queda definido como (Anselin y Bao, 1997):

$$G_i(d) = \frac{\sum_{j=1}^{N} W_{ij}(d) y_j}{\sum_{j=1}^{N} y_j} \quad ; para\ j \neq i$$

Dónde:

$W_{ij}\ (d)$: son los elementos de una matriz simétrica y binaria (0-1), dado $W_{ij} = 1$ para j unidades vinculadas i en un radio d, y W_{ij} = 0 para el resto de unidades, incluida la relación del punto i consigo mismo;

y_j : son los valores situados en un radio d del punto i (excepto y_i)

Los valores positivos y significativos de la estadística G están relacionados con el agrupamiento de valores altos, mientras los valores negativos significativos se relación con la agrupación de valores bajos. Por el contrario, un valor positivo del I de Moran denota agrupamiento de los valores de la variable, ya sean altos o bajos y un valor negativo indica disimilitud (Anselin y Rey ,2010).

Getis y Ord formularon otra versión de este estadístico, denotada como G_i^* Getis-Ord, en la que se incorpora el valor de la unidad central en estudio, de este modo omite la condición de $i \neq j$ planteada en la ecuación que define G_i (Bivand y Wong, 2018). En palabras de Ord (2006) G_i^* es más apropiado para usar en la detección de puntos calientes, en tato G_i es más pertinente para hallar la dependencia espacial local en el caso de los residuos de regresión. En esta investigación el G_i^* se empleará para el análisis aplicando el criterio de adyacencia de tipo reina para la construcción de pesos espaciales; cabe mencionar que se realizaron pruebas con ambos estadísticos y no se identificaron diferencias sustanciales.

Índice Local de Moran. El I de Moran local fue desarrollado por Anselin bajo la denominación de Indicadores locales de asociación espacial, LISA. Este indicador permite calcular el aporte de cada observación local al valor global, identificando agrupaciones de valores similares (conglomerados) estadísticamente significativas; de este modo analiza no solo el grado de asociación espacial entre las unidades en estudio si no también la heterogeneidad resultante de la contribución de cada unidad espacial. (Chasco, 2010; Siabato y Guzmán-Manrique, 2019).

El estadístico queda definido por la siguiente ecuación (Bivand y Wong, 2018):

$$I_i = \frac{z_i \sum_{j=1}^{n} w_{ij} z_j}{m_2}$$

Siendo

z_i , $x_i - \bar{x}$

w_{ij} , la matriz de pesos espaciales

z_j , $x_i - \bar{x}$

m_2 , $n^{-1} \sum_{i=1}^{n} Z_i^2$.

El valor del I de Moran local puede tomar valores positivos como consecuencia de agrupamiento de valores similares en torno a la unidad espacial i; o bien puede adoptar valores negativos como indicativo de agrupación de valores muy diferentes (Chasco, 2010).

Como se señaló la contribución de cada unidad espacial aporta a la conformación del indicador global de Moran (Bivand y Wong, 2018):

$$I = \frac{\sum_{i=1}^{n} I_i}{S_0}$$

El proceso de inferencia, se basa en el enfoque analítico y computacional mediante permutaciones para conocer el aporte de cada indicador local $\mathrm{I_i}$ como máximo contribuyente al valor del indicador del I de Moran global en términos de nivel de significancia; según el propio Anselin el enfoque computacional se prefiere debido a que el enfoque analítico requiere de un gran número de vecinos lo cual no se cumple en el análisis de autocorrelación local (GeoDa Sofware, 2021). La representación de los valores significativos $\mathrm{I_i}$ se lleva a cabo a través de los mapas LISA de Clusters, mediante la distinción por colores que permite conocer en torno a cuál de los puntos del diagrama existe mayor concentración espacial significativa de datos similares o diferentes de una variable (Chasco, 2010).

Como indica Ord (2006) LISA es muy similar al estadístico G de Getis y Ord; sin embargo la presencia de z_i en LISA establece las diferencias en los resultados obtenidos entre ambos índices, de este modo cuando los valores de z_i son cercanos a 0, el estadístico G de Getis y Ord es capaz de detectar vecinos con valores extremos, mientras que LISA no ; a pesar de ello el uso de LISA es eficiente para detectar situaciones en las que z_i es similar a sus vecinos de aquellos en los que está fuertemente diferenciado. Por tanto, ambos estadísticos ofrecen información complementaria.

Estadístico C ***local de Geary.*** La versión local de este estadístico fue formulada por Anselin en 1995 (Chasco, 2003). La versión local del estadístico

de Geary queda definido bajo la siguiente formula (Anselin, 1995, Siabato y Guzmán-Manrique, 2019):

$$c_i = \frac{1}{\sum_{i=1}^{n}(x_{i-}\bar{x})^2/_n} \sum_{j=1}^{n} w_{ij}(x_i - x_j)^2$$

Según Anselin (2019, como se citó en Siabato y Guzmán-Manrique, 2019) el primer término de la expresión es, al igual que el I de Moran, un factor de normalización. Sin embargo, considerando la naturaleza del término $(xi - xj)$ y al no emplear la media en la operación, no se evidencian diferencias significativas si se trabaja o no con los valores originales (escala original) o con los valores estandarizados.

Al igual que los estadísticos globales, la inferencia se basa en un enfoque de permutaciones sin embargo; a diferencia de las estadísticas globales, para las estadísticas locales se calculan permutaciones condicionales, es decir todas las observaciones se reasignan aleatoriamente excepto x_i , que se mantiene fijo, es decir no se emplea en la permutación (Anselin, 1995; Anselin, 2018; Anselin, 2020). En esta investigación se empleó el enfoque de permutaciones con 999 permutaciones.

SARS-CoV-2

Taxonomía. Los coronavirus pertenecen a la subfamilia Coronavirinae en la familia Coronaviridae y el orden Nidovirales. Son virus envueltos, con un genoma de ARN monocatenario de sentido positivo (Mousavizadeh, et al., 2021; Fehr y Perlman, 2015). Causan principalmente infecciones del tracto respiratorio y gastrointestinal y se clasifican genéticamente en cuatro géneros principales: Alphacoronavirus, Betacoronavirus, Gammacoronavirus y Deltacoronavirus. (Giovanetti, et al., 2021). El SARS-CoV-2 pertenece al género de los Betacoronavirus (Wu, et al., 2020).

Estructura. Las principales proteínas estructurales son la de membrana (M), la de envoltura (E), la nucleocápside (N) y la espicular (S) (Fehr y Perlman, 2015).

El SARS-CoV-2 ingresa a las células hospederas mediante la unión de la proteína S a los receptores ACE2 (Fehr, y Perlman, 2015; Ghafouri-Fard, et al., 2020). El dominio de unión al receptor (RBD) de la proteína Spike (S) de superficie, es la que reconoce el receptor ACE2, se une a él y media la penetración del virus en la célula huésped (Wu, et al.,2020).

Replicación. El SARS-CoV-2 se une al receptor ACE2 (encima convertidora de angiotensina 2) mediante la proteína espicular (proteína S), de esta forma e virus inicia el proceso de entrada e infección a la célula, para completar este proceso la proteína debe descomponerse en unidades proteínicas más pequeñas con ayuda de una enzima proteasa, llamada TMPRSS2, de esta forma completa el proceso de infección a la célula (Mousavizadeh, et al., 2021).
Una vez dentro de la célula hospedera el virus inicia el proceso de replicación y transcripción del genoma en las membranas citoplasmáticas, este conlleva procesos coordinados de síntesis de RNA mediados por un enorme complejo de proteínas codificado por el gen replicasa 20-kb (Mousavizadeh, et al., 2021).

Luego del proceso de replicación y síntesis de ARN subgenómico, se lleva a cabo la traducción de proteínas estructurales del virus, S, E y M y se insertan en el retículo endoplásmico (ER). Estas proteínas en la membrana celular del retículo endoplasmático-Golgi y se incorpora el ARN genómico. “Allí, los genomas virales encapsulados por la proteína N, formando viriones maduros brotan del interior de las membranas celulares” (Mousavizadeh, et al., 2021; Fehr y Perlman, 2015).

Patogénesis. El SARS-CoV-2 es el agente causal del Síndrome Respiratorio Agudo (SARS), responsable de un gran número de muertes alrededor del mundo.

De acuerdo a la OMS (2022b), “el virus puede propagarse desde la boca o nariz de una persona infectada en pequeñas partículas líquidas cuando tose, estornuda, habla, canta o respira. Estas partículas van desde gotículas respiratorias más grandes hasta los aerosoles más pequeños”.

Las personas contagiadas, en su mayoría, experimentan síntomas respiratorios leves o moderados; sin embargo, en algunos casos las personas, especialmente

aquellos grupos de personas de edad avanzada y que padecen de comorbilidades desarrollarán cuadros graves de la enfermedad e incluso morir (OMS, 2022b).

Linajes y variantes del SARS-CoV-2. Al igual que otros virus de ARN el SARS-CoV-2, es susceptible a mutaciones que provocan la aparición de múltiples cambios en el material genético del virus (Centros para el Control y la Prevención de Enfermedades [CDC], 2023). Para identificar estos cambios es necesario realizar la secuenciación genómica del virus, este proceso permite monitorizar los cambios en el ARN (material genético) del SARS-CoV-2 (CDC, 2022).

Estos cambios en el genoma del virus contribuyen al surgimiento de grupos relacionados genéticamente con un origen en común, los cuales se clasifican en linajes. La designación de estos linajes en variantes corresponde a una similitud de características compartidas entre los virus; la OMS ha convenido nombrar a las variantes del SARS-CoV-2 mediante el uso de letras griegas a fin de facilitar la comunicación por parte de la comunidad científica (CDC, 2023). Inicialmente OMS designó como Variantes de Interés (VOI, por sus siglas en inglés) y Variantes de Preocupación (VOC, por sus siglas en inglés) según la capacidad del virus para escapar a la acción del sistema inmunitario, de interferir en la eficacia de terapia y/o diagnóstico o por el incremento de una transmisión significativa en distintos países (OMS, 2022c).

CAPÍTULO 3: METODOLOGÍA

Tipo y Diseño de Investigación

Se trata de un estudio con enfoque cuantitativo debido a que se basa en la medición de la tasa de las variantes del SARS-CoV-2 y a partir de ello se obtuvo los índices de autocorrelación local y global.

Por otro lado, se aborda un diseño de investigación de tipo observacional .El objetivo fundamental fue caracterizar el patrón espacial de las variantes del SARS-CoV-2 a nivel provincial, al tener una base de datos que proporciona información de residencia a nivel provincia del individuo del cual se extrajo la muestra, se trabajó con datos agregados calculando la tasa de las variantes por provincias del Perú, por tanto se desarrolló un estudio con un diseño de tipo ecológico, los cuales están incluidos dentro de la clasificación de estudios de tipo descriptivo (McBride, et al., 2019).

Unidad de Análisis

La unidad de análisis está representada por una provincia del Perú.

Población de Estudio

Todas las provincias del Perú donde se han registrado resultados de Linaje Genómico de Pruebas Moleculares para COVID-19. No se va a considerar muestra.

Descripción de la base de datos

Se utilizaron los datos del resultado de Linaje Genómico de Pruebas Moleculares para COVID-19 notificados por el Instituto Nacional de Salud (INS), procedentes de la vigilancia genómica a nivel nacional, la información corresponde a los casos acumulados notificados desde el inicio de la pandemia hasta el 13 de octubre del 2022, disponible en: https://www.datosabiertos.gob.pe/user/3165/og_group_ref/6114?sort_by=changed

Proceso de filtración de la base de datos y descripción de las variables

Se analizaron los casos que comprenden el periodo de 01 de enero 2021 hasta el 30 de junio 2022, cada observación corresponde a un resultado de secuenciamiento genómico, no a un individuo per se. La base de datos inicial estaba conformada por 2700 genomas secuenciados, se Excluyen los genomas secuenciados del 2020 y los del mes de julio del 2022 en adelante debido a que a partir de esta fecha se secuenció la mayor cantidad de genomas; además dentro de este periodo se puede observar el comportamiento de las cuatro variantes de SAR CoV-2 que son objeto de esta investigación por periodos trimestrales. Cabe precisar que el secuenciamiento genómico por parte del INS se inició en mayo del 2021(Instituto Nacional de Salud [INS], 2021).

Posteriormente se excluyeron genomas secuenciados correspondientes a otras variantes (Iota, Alfa, Zeta, Mu, Recombinantes y otros), debido a que se identificaron en menor cantidad de muestras de pacientes (Figura 10), además de las pruebas moleculares secuenciadas incongruentes con la denominación provincia y sin datos. Consecuentemente la muestra final quedó conformada por 19252 secuencias del genoma del SARS-CoV-2, distribuidas en las 196 provincias de Perú. El proceso de selección de registros se muestra en la Figura 8 y la descripción de las variables en la Tabla 2.

Figura 8

Proceso de selección de los datos correspondientes a los genomas secuenciados durante enero 2021- junio 2022.

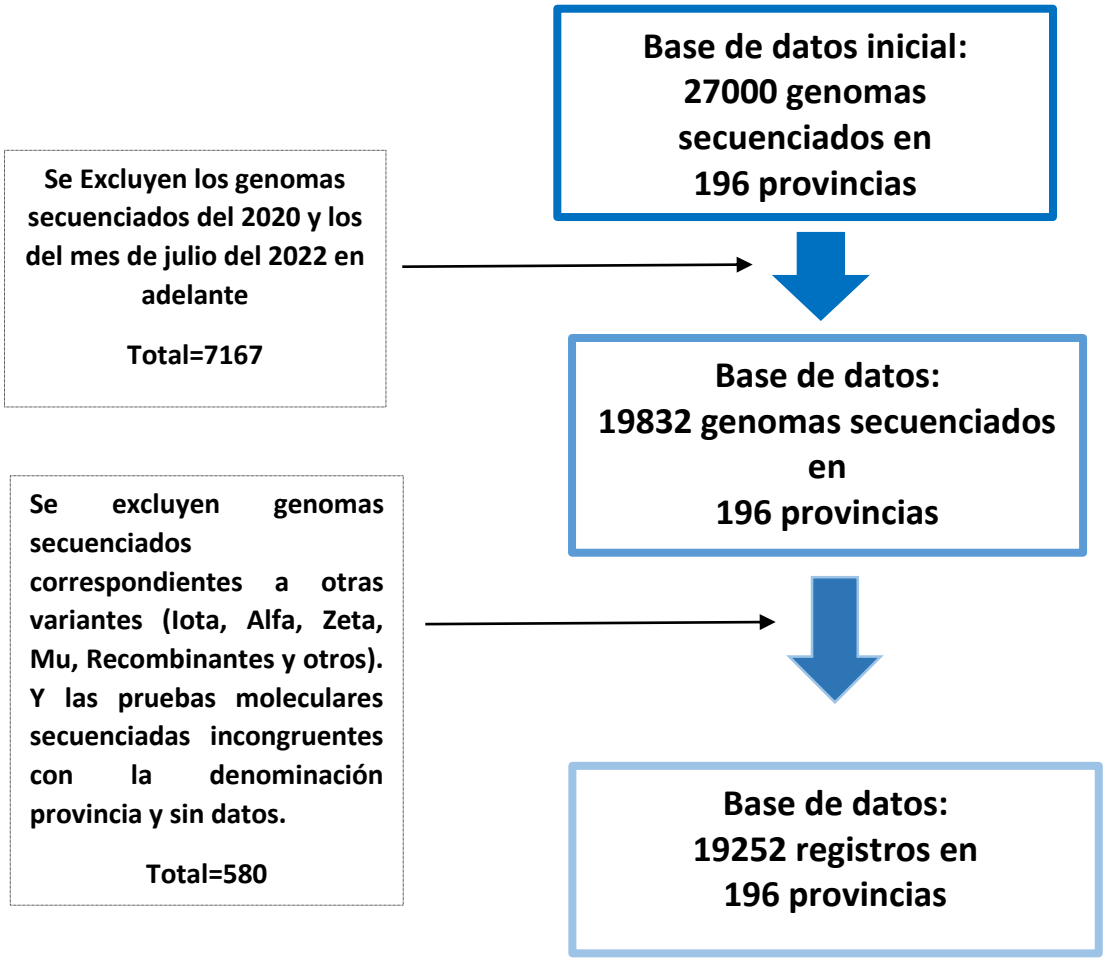


Tabla 2

Descripción de las variables de la base de datos considera para el análisis.

Variable	Tipo de Datos	Descripción
Fecha _muestra	Fecha	Fecha de toma muestra
Edad	Cuantitativa	Edad del paciente (en años)
Sexo	Cualitativa	Sexo del paciente
Ubigeo_paciente	Cualitativa	Ubigeo del paciente, correspondiente al ubigeo de RENIEC
Departamento_paciente	Cualitativa	Departamento de residencia del paciente
Provincia_paciente	Cualitativa	Provincia de residencia del paciente
Distrito_paciente	Cualitativa	Distrito de residencia del paciente
Resultado_linaje	Cualitativa	Resultado de linaje genómico de SARS-COV-2
Resultado_variante	Cualitativa	Variante de SARS-COV-2 (Delta, Gamma, Lambda, Omicrom)

CAPÍTULO 4: RESULTADOS Y DISCUSIÓN

4.1. Análisis, interpretación y discusión de resultados

Para determinar la dinámica de la propagación de las variantes en las provincias del país, durante el periodo de 01 de enero 2021 hasta el 30 de junio 2022, se aplicó técnicas gráficas y estadísticas del análisis estadístico convencional y del análisis exploratorio de datos espaciales (AEDE), a través del software SPSS v.25 y la herramienta de software libre Geoda v. 1.20 respectivamente.

El INS realiza la vigilancia epidemiológica del virus del SARS-CoV-2 en el Perú para detectar variantes del virus que circulan en el país a través del secuenciamiento genómico en muestras provenientes de laboratorios públicos y privados que realizan la detección molecular del virus. La elevada tasa de mutaciones del SARS-CoV-2 es la causante de la amplia diversidad genómica del virus, generando nuevos linajes, estos se ubican dentro de variantes y son el resultado de la acumulación de varias mutaciones en el genoma del virus. Dado que los datos disponibles consisten en el registro de los linajes, primero se agrupó los linajes de acuerdo a la variante asignada, posteriormente el cálculo de la tasa de las variantes se realizó dividiendo el total de cada variante sobre total de secuencias reportadas a lo largo del periodo de estudio en cada provincia.

$$TI = \frac{Total\ de\ genomas\ secuenciados\ de\ cada\ variante\ del\ SARS\ CoV-2}{Total\ de\ genomas\ secuenciados\ en\ la\ provincia}\ x100$$

Análisis Exploratorio

El análisis exploratorio de la distribución evidencia baja frecuencia de algunas variantes, esto corresponde con la tendencia general observada a nivel mundial (Global Initiative on Sharing All Influenza Data [GISAID], sin fecha), reportándose también en menor número en otros países. Se puede notar que estas variantes

de baja frecuencia circulan por un periodo de tiempo limitado en comparación con variantes de mayor relevancia cuya permanencia es más duradera (Figura 9), lo que sugiere que las mutaciones en el genoma de estas variantes podrían reducir su capacidad para competir con otras variantes del virus; sin embargo, no se debería descartar la pérdida de información de estas variantes dada su baja frecuencia. En cuanto a las variantes identificadas con mayor frecuencia en el país durante el periodo de estudio se identificó cuatro variantes principalmente, representando estas el 97.10% de todas las variantes caracterizadas en el país, durante el periodo de estudio, y son las variantes objeto de esta investigación, la variante Gamma (9.02 %), la variante Lambda (18.40%), la variante Delta (27.77%), y variante Ómicron (41.91%) (Figura 10).

Figura 9

Circulación de las variantes (inicio-término) en Perú desde marzo de 2020 hasta junio de 2022.

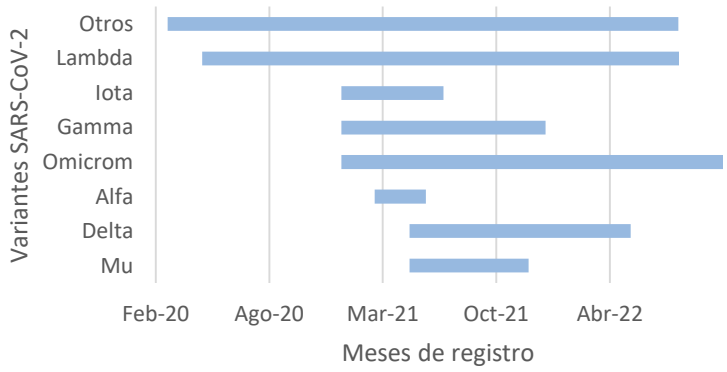
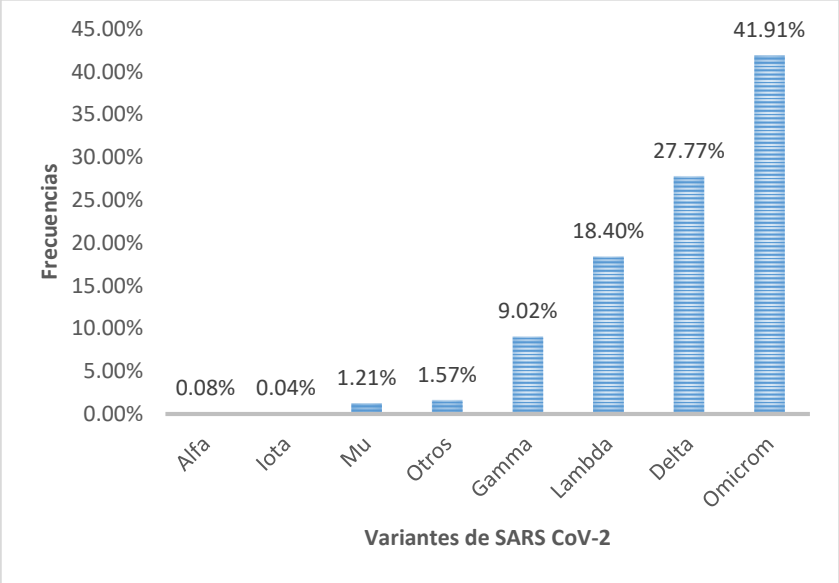


Figura 10

Frecuencias de las variantes registradas en Perú desde marzo de 2020 hasta junio de 2022.



Datos demográficos asociados a las secuencias analizadas. De los 19252 casos de Covid 19 analizados, el 54.11% corresponden a sujetos del sexo femenino y el 45.89% a sujetos del sexo masculino. La edad mediana de las personas cuyas muestras fueron analizadas es de 40.71 años (DS 17.93); la mínima de 1 año y máximo 100 años (Tabla 3). Estos datos concuerdan con el informe elaborados por el CDC - Perú al 03 de junio del 2022, el cual da a conocer que la mayoria de casos confirmados corresponden a la poblacion adulta de 30-59 años y las menores cifras corresponden a niños de 0 - 11 años y adolescentes de 12-17 años (MINISERIO DE SALUD [MINSA],2022).

El rango de edad de 30 a 59 años presentó el mayor número de casos para todas las variantes en estudio, lo cual concuerda con lo reportado por el CDC Perú, que informa que el mayor porcentaje de casos se presenta en los adultos que acumulan el 56% de los casos (MINSA,2022).

Tabla 3

Características demográficas de los sujetos, cuyas muestras han sido analizadas, durante el periodo de 01 de enero 2021 hasta el 30 de junio 2022.

	Variante			
	Delta (n=5504)	Gamma (n=1789)	Lambda (n=3648)	Omicron (n=8311)
Edad				
Niño (0 – 11 años)	206 (3.7%)	28 (1.6%)	62 (1.7%)	204 (2.5%)
Adolescente (12 – 17 años)	343 (6.2%)	72 (4.0%)	196 (5.4%)	444 (5.3%)
Joven (18 – 29 años)	1374 (25.0%)	450 (25.2%)	832 (22.8%)	1634 (19.7%)
Adulto (30 – 59 años)	2770 (50.3%)	1008 (56.3%)	2114 (57.9%)	4386 (52.8%)
Adulto mayor (60 a más años)	811 (14.7%)	231 (12.9%)	444 (12.2%)	3129 (19.8%)
Sexo				
Femenino	2876 (52.3%)	909 (50.8%)	1845 (50.6%)	4788 (57.6%)
Masculino	2628 (47.7%)	880 (49.2%)	1803 (49.4%)	3523 (42.4%)

Al analizar la distribución de las variantes se observan tendencias interesantes de la distribución de las variantes durante el periodo de estudio. El diagrama de cajas de la Figura 11 indica que existe alta variabilidad a nivel provincial de las tasas de las variantes de SARS-Cov2, además, con respecto a la variante Gamma y Lambda se observa la presencia de valores atípicos con los valores más altos para las provincias de Oyón, Sechura, Condorcanqui, Maynas, San Martin, Bellavista, Requena, Alto Amazonas, El Dorado, Uctubamba, Junín Y Lamas.

Por otro lado, la variante Lambda representa la mayor cantidad de las secuencias genómicas durante el primer semestre del año 2021 y alrededor de julio de ese mismo año comenzó a declinar. Estos primeros 6 meses corresponden al periodo de la segunda ola de la pandemia, este periodo se caracterizó por el elevado número de muertes, incluso mayor que la primera ola, Desde el inicio de la pandemia hasta junio del 2021 se habían registrado 190 425 defunciones, la

mayor proporción (50,9%) ocurrió durante el primer semestre del 2021(MINSA, 2021).

Cabe destacar que la variante Lambda es de particular interés, se detectó por primera vez en Perú en mayo de 2020 y luego comenzó a aumentar su frecuencia dramáticamente para convertirse en el linaje dominante en el primer semestre del 2021.

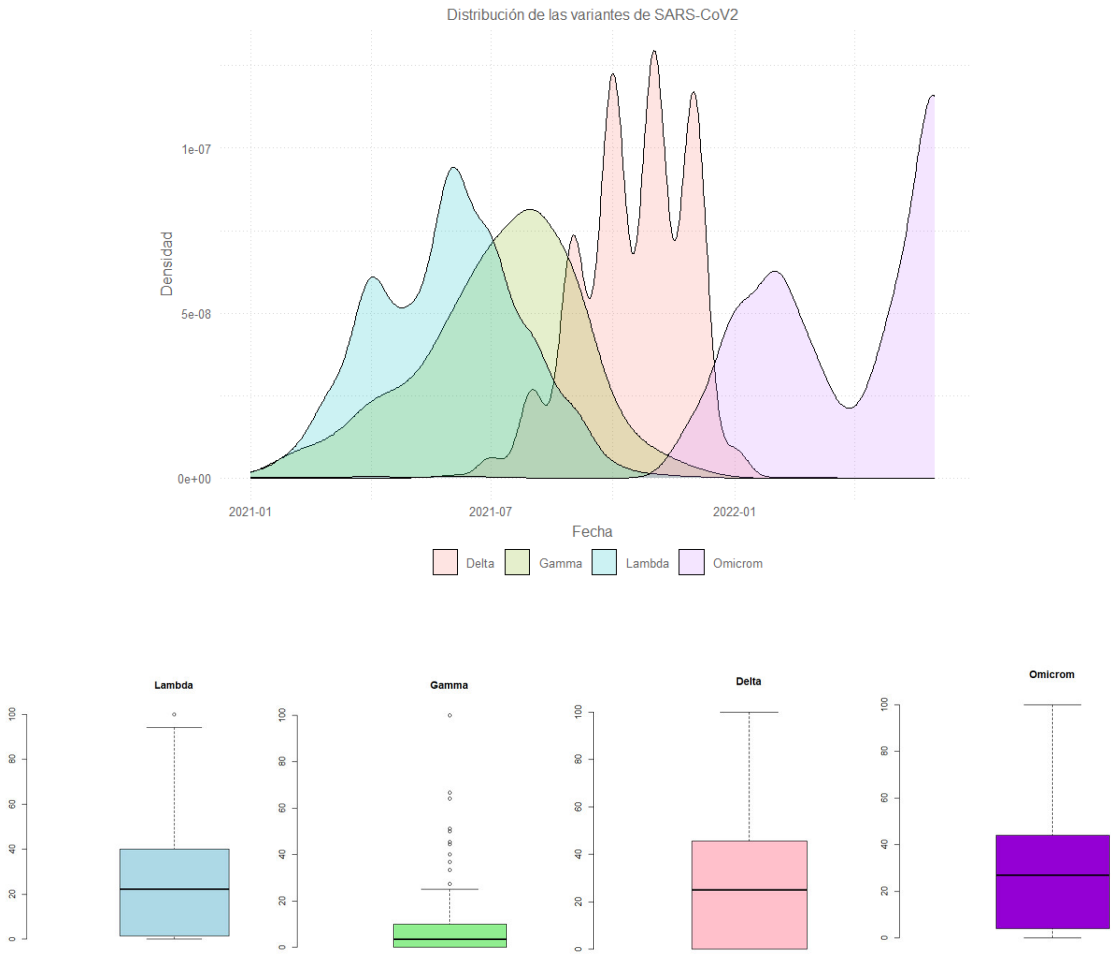
La variante Gamma fue reportada durante todo el año 2021, circulando con otras variantes, sin embargo, la frecuencia de esta variante fue menor en comparación a la variante Lambda, la cual fue la predominante durante este periodo.

La variante Delta, que no se identificó en muchas muestras durante los primeros meses, comenzó a propagarse a finales de setiembre de ese mismo año marcando el inicio de la tercera ola y luego disminuyó drásticamente a partir de enero del 2022. Este periodo se caracterizó por el incremento del ritmo de vacunación en la población (Plataforma digital única del Estado Peruano, 2021); por el contrario, en ciertos países de Europa y en los Estados Unidos de Norteamérica ocurrió un aumento dramático de los casos de COVID-19 y de las hospitalizaciones, la gran mayoría a causa de la variante Delta (Lizaraso Caparó y Pareja Cruz, 2021).

La variante Ómicrom reemplazó a Delta, ascendiendo rápidamente a finales del año 2021 hasta convertirse en la variante predominante hasta la fecha. La variante Ómicrom se propaga con mayor facilidad que las variantes anteriores del virus que causa el COVID-19, incluida la variante Delta; sin embargo, los casos de infección por la variante Ómicrom suelen ser menos letales que las infecciones causadas por las variantes anteriores (CDC, 2021), lo cual se condice con las características de la tercera ola, que se caracterizó por el elevado número de casos (mayores que la primera y segunda ola), pero menor número de muertes reportadas.

Figura 11

Distribución temporal y por cuartiles de las variantes más frecuentes en Perú entre enero de 2020 y junio de 2022.



Análisis Exploratorio de Datos Espaciales

***Representación cartográfica de la información*.** El mapeo de las variantes facilitó la comparación visual de la distribución de estas en las provincias del Perú. La variante Lambda predominó en las provincias de la costa y sierra del Perú mientras que Gamma, principalmente, en las provincias de la selva lo que puede sugerir que el predominio de una variante en una determinada zona limita la propagación de otra, lo cual se ve reflejado en la representación de los mapas (Figura 12 a y b). Cabe señalar que ambas variantes circularon durante el mismo

periodo de tiempo y fueron las predominantes durante la segunda ola de la pandemia.

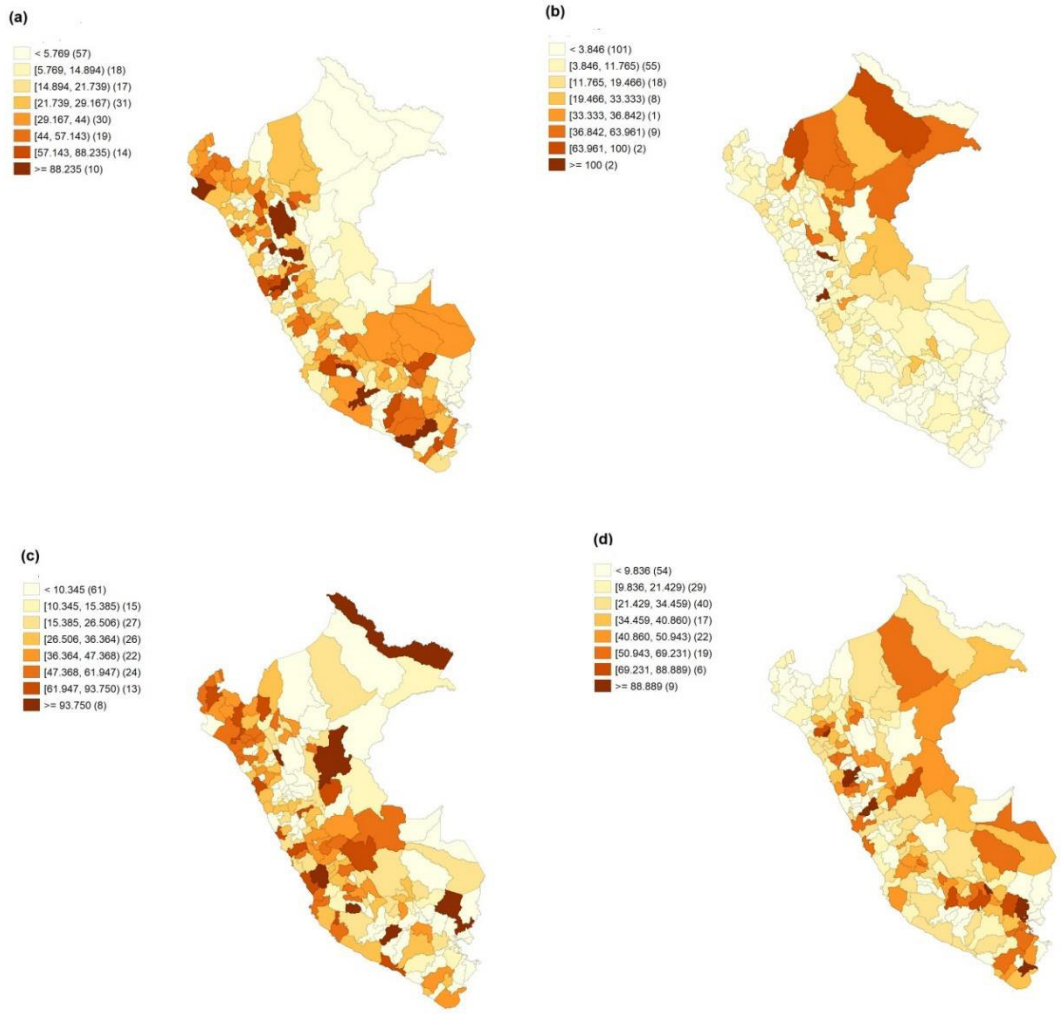
Como se señaló anteriormente Lambda y Gamma circularon en el periodo donde se reportó el mayor número de fallecimientos durante la pandemia; sin embargo, en las provincias que corresponden a la selva del Perú, donde imperó Gamma, las muertes reportadas fueron menores en comparación con las regiones de la costa y sierra, donde predominó Lambda (MINSA, 2023). Esto podría sugerir que Lambda presenta mutaciones que le confieren mayor letalidad que Gamma.

Por su parte, Delta y Ómicrom muestran un comportamiento más aleatorio en el espacio; ambas variantes se observaron en las provincias de la costa, sierra central y parte de la selva (Figura 12 c y d).

Al observar de manera conjunta los cuatro mapas, se puede notar que las áreas sombreadas de los mapas coropléticos se complementan con respecto a las variantes que circularon de manera conjunta, Lambda - Gamma , Delta – Ómicrom; esto podría explicarse por las diferencias fenotípicas del virus que afectan su patogenicidad (capacidad de infectar) y virulencia (grado de daño que puede causar un microorganismo) y le pueden conferir más o menos resistencia a ciertas condiciones ambientales como la altitud o el tipo de clima que permiten mejor adaptabilidad de algunas variantes frente a otras en determinadas zonas o en individuos con determinadas características.

Figura 12

Representación geográfica de las tasas de las variantes de SARS-CoV-2 por provincia en Perú. a) Lambda b) Gamma c) Delta d) Ómicrom. (Datos obtenidos desde enero 2021 hasta junio2022).



Análisis de autocorrelación espacial Global. Para medir la autocorrelación global se empleó el Indice Global de Moran y el Indice C de Geary. Los valores que toman ambos estadísticos difieren entre sí. I de Moran toma valores de -1 a 1 (donde -1 expresa autocorrelación negativa perfecta y 1 autocorrelación positiva perfecta). En contraste con I de Moran, los valores bajos de C de Geary denotan autocorelación positiva y los valores altos indican autocorrelación negativa (Waller y Gotway, 2004), en tal sentido C de Geary asumes valores que varían de 0 a 2 (donde los valores cercanos a 0 están positivamente autocorrelacionados y los valores cercanos a 2 están negativamente autocorrelacionados). Al observar los resultados de esta investigación, los valores de ambos índices permiten llegar a las mismas conclusiones a nivel global.

Los patrones de distribución espacial, descritos en los mapas coropléticos, sugieren zonas con ciertas agrupaciones de valores similares lo que podría proponer autocorrelación espacial global. El cálculo del I de Moran y el C de Geary confirman ligera autocorrelación espacial global positiva para la variante Delta (I Moran= 0.096, p valor ≤0.05; C Geary=0.894, p valor ≤0.05), Gamma (I Moran= 0.204, p valor ≤0.01; C Geary=0.813, p valor ≤0.01) y Ómicrom (I Moran = 0.157, p valor ≤0.01; C Geary=0.839, p valor ≤0.01). Por otra parte, la autocorrelación de las tasas de la variante Lambda no fue significativa en las provincias del Perú (Tabla 4); sin embargo, se analizarán las correlaciones a nivel local para identificar la presencia de clústeres espaciales o valores atípicos a nivel local.

Tabla 4

Comparación del Índice de Moran y C de Geary de las variantes de SARS-Cov2 entre enero de 2021 y junio de 2022.

	I de Moran		**C de Geary**	
	Estadístico	Significancia	Estadístico	Significancia
Gamma	0.204	0.001	0.813	0.008
Lambda	0.022	0.260	0.964	0.247
Delta	0.096	0.014	0.894	0.016
Ómicrom	0.157	0.001	0.839	0.004

***Análisis de autocorrelación espacial Local*.** Para complementar el análisis global se utilizan los indicadores locales de asociación espacial, los mapas permitieron apreciar diferencias en el comportamiento de cada variante del virus a través de las provincias del Perú.

El I de Moran Local para la variante Lambda muestra algunas zonas de alto riesgo (cluster alto – alto) que se ubican en ciertas provincias del norte y centro. Cabe mencionar que Padilla, et al. (2021), Indicaron el dominio de la variante Lambda a nivel nacional durante la segunda ola señalando que podría deberse a características fenotípicas que la colocan en una posición de ventaja frente a la infección causada por otras variantes que circularon paralelamente; sin embargo esta investigación demuestra que el dominio de la variante Lambda no fue uniforme a lo largo de territorio nacional, evidenciando preponderancia de ciertas zonas en los mapas de clusteres. Por otro lado, es notable el predominio de zonas de bajo riesgo (cluster bajo- bajo) principalmente en provincias del oriente y norte las cuales se expresan en la formación de un cluster de gran tamaño formado por seis provincias, además de tres provincias a lo largo de la sierra norte y la sierra sur del Perú. En tanto las provincias catalogadas con valores atípicos (alto-bajo y bajo-alto) están distribuidas en las provincias de costa y sierra (Figura 13).

El mapa de C de Geary Local revela, además clústeres que no fueron detectados por el I de Moran Local. Se puede apreciar clusteres de alto riesgo en provincias como La Mar (Ayacucho), Calca, La Convención y Urubamba (Cuzco), Ayabaca, Huancabamba, Morropón, Sullana (Piura) y Contralmirante Villar (Tumbes). Además C de Geary revela también otros clústeres de bajo riesgo como Chanchamayo (Junín), Ferreñafe (Lambayeque) y Chepén (La libertad); los detalles se pueden apreciar en la Tabla 5.

El G^* de Getis y Ord identifica puntos calientes y fríos en las mismas provincias que I de Moran; además de las provincias Yunguy y Moho (Puno), como puntos fríos, que en el análisis con I de Moran no se identificó; cabe resaltar que G^* de Getis y Ord identifica como punto caliente provincias que I de Moran identifica como cluster bajo-alto (Camaná, Huanca Santos, Dos de Mayo, Huaycabamba, Bolivar y Pataz) y como puntos fríos a las provincias que I de Moran categoriza como cluster Alto- Bajo (Marañón) esto se aprecia con más detalle en la Tabla 5, esto es a causa de que G^* de Getis y Ord no toma en cuenta la unidad central de análisis si no el promedio de sus vecinos por tanto al estar rodeada por provincias con valores bajos la unidad central resultara como un punto frio, de la misma forma las unidades rodeadas por vecinos con valores altos resultaran en puntos calientes.

Figura 13

Mapa clúster de las tasas de la variante Lambda en las provincias del Perú de enero 2021 –junio 2022. Índice Local de Moran, Índice C *de Geary y* G^* *Getis y Ord.*

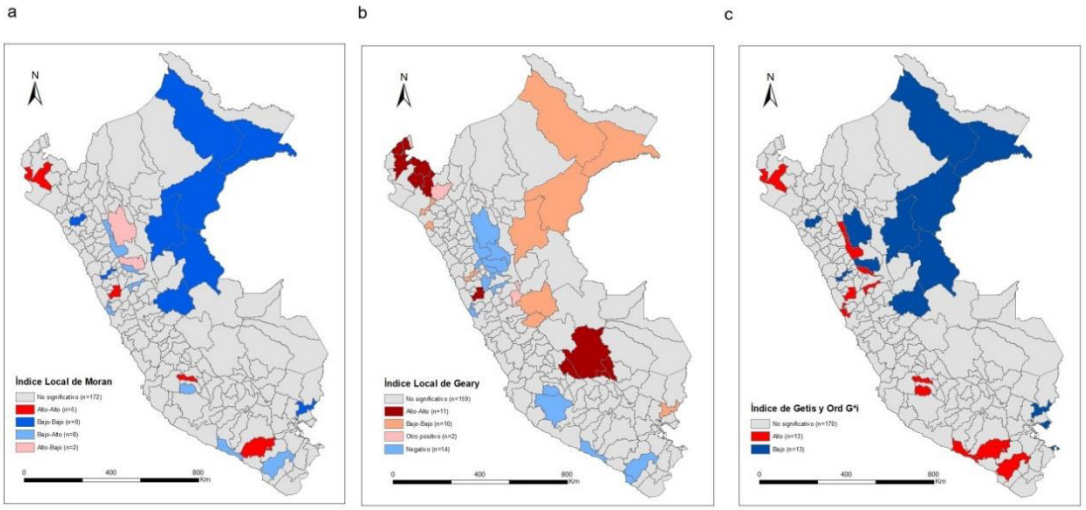


Tabla 5

Clústeres identificados con el Índice Local de Moran, Índice C de Geary y G^* Getis y Ord de las tasas de la variante Lambda en las provincias del Perú de enero2021 –junio 2022.

Departamento	Provincia	I Moran	C Geary	G* Getis y Ord
Ancash	Recuay			
	Yungay			
Arequipa	Arequipa			
	Camaná			
Ayacucho	Cangallo			
	Huanca Santos			
	La Mar			
Cajamarca	San Miguel			
Cusco	Calca			
	La Convención			
	Urubamba			
Huánuco	Dos de Mayo			
	Huaycabamba			
	Marañón			
	Mariscal Nieto			
Junín	Chanchamayo			
Lambayeque	Ferreñafe			
La Libertad	Chepén			
	Bolívar			
	Pataz			
Lima	Barranca			
Loreto	Maynas			
	Mariscal Ramón Castilla			

	Requena			
	Ucayali			
Moquegua	Mariscal Nieto			
Pasco	Oxapampa			
Piura	Ayabaca			
	Huancabamba			
	Morropón			
	Sullana			
	Paita			
	Piura			
Tumbes	Contralmirante Villar			
	Tumbes			
Pasco	Oxapampa			
Puno	San Antonio de Putina			
	Moho			
	Yunguyo			
San Martin	Mariscal Cáceres			
Ucayali	Coronel Portillo			

Nota. Los colores indican la formación de clústeres Alto- Alto; Bajo- Bajo; Bajo –Alto; Alto-Bajo.

El mapa de I de Moran Local de la variante Gamma destaca por zonas de alto riesgo en su mayoría en la zona de la amazonia, los resultados concuerdan con Vargas-Herrera, et al. (2021), quienes resaltaron el predominio de la variante Gamma en la cuenca norte del Amazonas, los autores atribuyen esta situación a la ubicación geográfica (frontera con Brasil) y mayor seroprevalencia (porcentaje de personas con anticuerpos) de SARS-CoV-2 en la región, la cual representa el 70%. Las zonas de bajo riesgo predominan en el norte y parte de las provincias del centro y sur. Las provincias agrupadas por valores atípicos se ubican en

menor número, dos provincias catalogadas como alto-bajo y siete provincias como bajo-alto destacando entre estas últimas la provincia de Putumayo en Loreto (Figura 14), es curiosa esta situación ya que indicaría que esta provincia presenta menor número de casos en comparación con las provincias vecinas, es relevante resaltar que esta provincia colinda con el sur de Colombia, país en el cual predominaron otras variantes por encima de la variante Gamma a mediados de agosto del 2021 (GISAID, sin fecha).

Cabe señalar que con el análisis de diagrama de cajas (Figura 11) se detectó un importante número de provincia con valores atípicos que podrían estar generando resultados engañosos ya que I de Moran Local es sensible al efecto de valores atípicos en la muestra pues este indicador se construye a partir del rezago espacial (promedio de los vecino), de este modo la presencia de un vecino con valor muy alto o muy bajo puede conducir a la identificación errada de zonas con valores atípicos bajos – altos o altos- bajos, por ello es preciso la comparación con los otros dos estadísticos.

C de Geary Local se asemeja al I de Moran Local con respecto a las provincias ubicadas en la amazonia y centro-norte del país (Figura 14); sin embargo, identifica otros puntos calientes en Urubamba (Cusco), Puerto Inca (Huánuco), Canta (Lima), Maynas (Loreto), Bellavista y El Dorado (ambas en San Martin) (Tabla 6). Destaca nuevamente la provincia de Putumayo, C de Geary identifica ausencia de autocorrelación local en esta provincia como I de Moran , pero no hay formas de clasificarla como clústeres bajo-alto o alto –bajo, esto es debido a que los valores locales de las diferencias cuadradas , calculada con C de Geary , no presentan signo por tanto de este modo se podría deducir que I de Moran ofrece mayor información para analizar esta provincia; sin embargo Loreto se presenta bajo el mismo escenario con C de Geary Local, esta provincia resultó como un cluster de alto riesgo con I de Moran; pero es conveniente considerer que los valores de I de Moran se ven afectados por el promedio de los valores de la provincias vecinas y en este caso C de Geary podría sugerir que la provincia

de Loreto no necesariamente sería una zona de alto riesgo pero podría llegar a serlo, posteriormente, por la infuencia de sus vecinos.

El G^* de Getis y Ord identifica también un gran conglomerado como punto caliente en provincias del oriente y se extiende a algunas provincias del centro. Los puntos fríos están ubicados en ciertas provincias del norte, centro y en el sur del país principalmente, evidenciando un patrón más aleatorio en estas zonas; además distingue algunas provincias, no identificadas antes como puntos fríos, Chucuito, Yunguyo (Puno) y Tacna (Tacna). Nuevamente ciertas provincias clasificadas con I de Moran como disimiles se clasifican como puntos fríos (Cayoma) y calientes (Bongara, Antonio Raymondi, Marañón, Putumayo, Ucayali, Daniel Alcides Carrión y Pasco), se pueden ver mayores detalles en la Tabla 6.

Figura 14

Mapa clúster de las tasas de la variante Gamma en las provincias del Perú de enero 2021 –junio 2022. Índice Local de Moran, Índice C de Geary y G^* Getis y Ord.

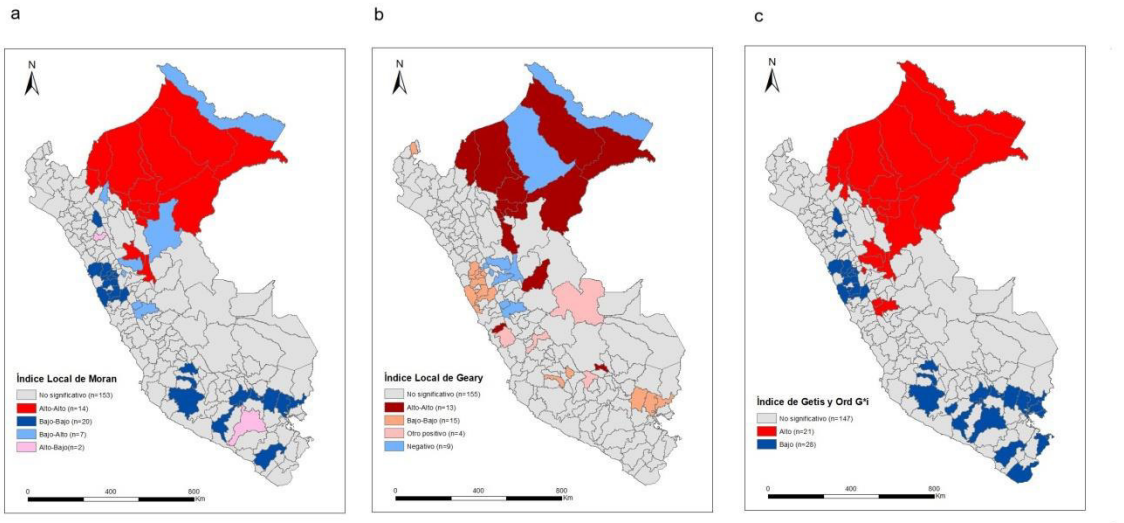


Tabla 6

Clústeres identificados con el Índice Local de Moran, Índice C de Geary y G^* Getis y Ord de las tasas de la variante Gamma en las provincias del Perú de enero 2021 –junio 2022.

Departamento	Provincia	I Moran	C Geary	G* Getis y Ord
Amazonas	Bagua			
	Bongara			
	Condorcanqui			
Ancash	Antonio Raymondi			
	Huaraz			
	Asunción			
	Bolognesi			
	Carhuaz			
	Huarmey			
	Huaylas			
	Recuay			
	Santa			
	Yungay			
Apurímac	Chincheros			
Arequipa	Caylloma			
	Condesuyos			
Ayacucho	Huamanga			
	Lucanas			
	Páucar de Sara Sara			
	Víctor Fajardo			
Cajamarca	Cajabamba			
	Celendín			
Cusco	Canas			

	Chumbivilcas			
	Urubamba			
Huánuco	Marañón			
	Leoncio Prado			
	Puerto Inca			
Lima	Barranca			
	Canta			
Loreto	Alto Amazonas			
	Loreto			
	Mariscal Ramón Castilla			
	Maynas			
	Requena			
	Putumayo			
	Ucayali			
	Datem del Marañón			
Moquegua	Mariscal Nieto			
Pasco	Daniel Alcides Carrión			
	Pasco			
Puno	Azángaro			
	Chucuito			
	Huancané			
	Melgar			
	Moho			
	San Antonio de Putina			
	Yunguyo			
San Martin	Bellavista			
	El Dorado			

	Moyobamba	■		■
	Lamas	■	■	■
	Picota	■		■
	San Martin	■	■	■
	Tocache	■		■
Tacna	Tacna			■
Tumbes	Tumbes		■	

Nota. Los colores indican la formación de clústeres ■ Alto- Alto; ■ Bajo- Bajo; ■ Bajo –Alto; ■ Alto-Bajo.

Es interesante comparar el comportamiento de Lambda y Gamma debido a que ambas variantes circularon durante el mismo periodo de tiempo. Por su parte Lambda, identificada por primera vez en Perú en diciembre del 2020 predominó en la costa y sierra del Perú en la primera mitad del año 2021 (Vargas-Herrera, et al. ,2022); sin embargo, el análisis de autocorrelación espacial global evidenció que el comportamiento de esta variante no fue similar a lo largo de las provincias que conforman esta zona, lo cual se confirmó con el análisis de clusteres.

Por otro lado, la variante Gamma dominó el norte de la amazonia peruana (Vargas-Herrera, et al. ,2022), en tal sentido este análisis confirmó que, en esta zona, en efecto, formaron un gran conglomerado de alto riesgo (provincias con altas tasas de contagios por Gamma rodeadas de provincias vecinas con tasas similares).

Como señala Álvarez-Antonio, et al. (2021), luego de la primera ola de contagios la población del norte de la amazonia peruana experimento altas tasas de seroprevalencia, lo que podría indicar la presencia de mutaciones en Gamma que le confieran ciertas características al escape inmunológico.

Los resultados para Delta del I de Moran Local, C de Geary Local y G^* de Getis y Ord se pueden observar en la Figura 15. I de Moran Local hace notar que la agrupación de puntos calientes ocurre en las provincias del norte y centro del

país. Estas provincias presentan elevados casos de Delta al igual que sus vecinos. Por su parte las zonas de bajo riesgo configuran un patrón más disperso a lo largo de las provincias. Las zonas discordantes; bajo-alto se ubican en ciertas provincias de la sierra central, una en el oriente y una en la sierra sur en Puno. Las zonas categorizadas como alto-bajo se identifican en una provincia del oriente, provincias del norte y provincias del sur del Perú.

C de Geary Local muestra un patrón similar, sin embargo, resalta la ausencia de Uctubamba (Amazonas) en el Oriente como punto significativo. Otros clústeres de alto riesgo identificados son la provincia de Huamanga (Ayacucho), Chota (Cajamarca), Satipo, Tarma (Ambos en Junín), Huaral (Lima), Pasco (Pasco), Huancabamba, Talara (Ambos en Piura) y Contralmirante Villar (Tumbes), por otro lado, un cluster de bajo riesgo, no identificado con I de Moran, en Abancay (Apurímac), ver Tabla 7.

Con G^* de Getis y Ord se aprecian, además, las provincias de Chucuito, Yunguyo, Moyobamba y Moho (provincias ubicadas en Puno) como puntos fríos de concentración de la variante Delta ; por otro lado, se detectan como puntos calientes algunas provincias que con I de Moran resultan en clusteres de valores atípicos bajo- alto, Uctubamba (Amazonas), Castro Virreyna (Huancavelica), Concepción (Junín), Huarochirí (Lima) y Sandia (Puno) y como puntos fríos, provincias detectadas por I de Moran como clústeres alto – bajo, Huari (Ancash),La Unión (Arequipa),Espinar (Cusco), Bolívar (La Libertad) y Mariscal Nieto (Moquegua) (Tabla 7).

Figura 15

Mapa clúster de las tasas de la variante Delta en las provincias del Perú de enero2021 –junio 2022. Índice Local de Moran, Índice C *de Geary y* G* *Getis y Ord.*

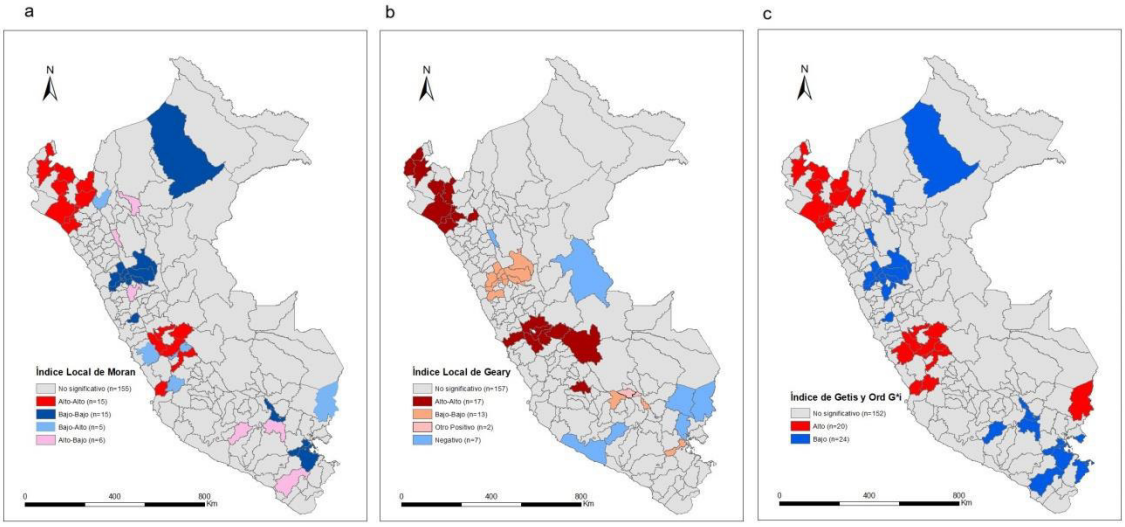


Tabla 7

Clústeres identificados con el Índice Local de Moran, Índice C de Geary y G^* Getis y Ord de las tasas de la variante Delta en las provincias del Perú de enero 2021 –junio 2022.

Departamento	Provincia	I Moran	C Geary	G* Getis y Ord
Amazonas	Uctubamba			
Ancash	Huaraz			
	Carlos Fermín Fitzcarrald			
	Huari			
	Huaylas			
	Mariscal Luzuriaga			
	Pomabamba			
	Sihuas			
	Yungay			
Apurímac	Abancay			
Arequipa	La Unión			
Ayacucho	Huamanga			
Cajamarca	Chota			
	Jaén			
	San Ignacio			
Cusco	Acomayo			
	Canas			
	Espinar			
Huancavelica	Castrovirreyna			
Huánuco	Huacaybamba			
	Marañón			
Ica	Chincha			
Junín	Huancayo			
	Chanchamayo			

	Concepción			
	Jauja			
	Junín			
	Satipo			
	Tarma			
	Yauli			
La Libertad	Bolívar			
Lambayeque	Chiclayo			
	Ferreñafe			
	Lambayeque			
Lima	Cajatambo			
	Huaral			
	Huarochirí			
Loreto	Loreto			
Moquegua	Mariscal Nieto			
Pasco	Pasco			
Piura	Ayabaca			
	Huancabamba			
	Morropón			
	Sullana			
	Talara			
Puno	Chucuito			
	Yunguyo			
	Moyobamba			
	Moho			
	Puno			
	Sandia			
	San Román			
San Martin	Moyobamba			
	Tocache			

Tumbes	Contralmirante Villar			
	Tumbes			

Nota. Los colores indican la formación de clústeres Alto- Alto; Bajo- Bajo; Bajo –Alto; Alto-Bajo.

Finalmente, Ómicrom advierte clusteres de menor tamaño, las zonas de alto riesgo se ubican dispersas en el norte, centro y sur. Destaca un conglomerado de bajo riesgo en provincias del norte del país. Los valores disimiles bajo – alto y alto – bajo son menores ubicándose en una provincia del norte y sur del país (Figura 16).

Los resultados del Índice de Geary indican como conglomerados de alto riesgo, además, a las provincias de Caravelí (Apurímac), Acobamba (Huancavelica) y Palpa (Ica). Otros clústeres de bajo riesgo no identificadas con I de Moran se ubican en las provincias de Uctubamba (Amazonas), Chanchamayo, Tarma (ambos en Junín), Ascope (La Libertad), Huancabamba, Sechura (ambos en Piura), Lamas y Picota (Ambas en San Martin) y Tumbes (Tumbes), como se puede apreciar en la Tabla 8.

Por su parte G^* de Getis y Ord, identifica provincias en Puno como puntos fríos, que no fueron identificadas con el I de Moran, estas son las provincias de Yunguyo y Moho (Tabla 8).

Si bien Delta y Ómicrom co-circularon por un corto tiempo, ya que Ómicrom desplazó a Delta en el auge de la tercera ola, ambas variantes cobraron un papel importante durante este periodo de contagios. Lo que sugiere un mayor escape inmunológico por parte de Ómicrom. Por el contrario, en ciertos países de Europa y en los Estados Unidos de Norteamérica la variante Delta fue la causante de la mayoría de casos (Lizaraso Caparó y Pareja Cruz, 2021).

Figura 16

Mapa clúster de las tasas de la variante Ómicrom en las provincias del Perú de enero2021 –junio 2022. Índice Local de Moran, Índice C *de Geary y* G* *Getis y Ord.*

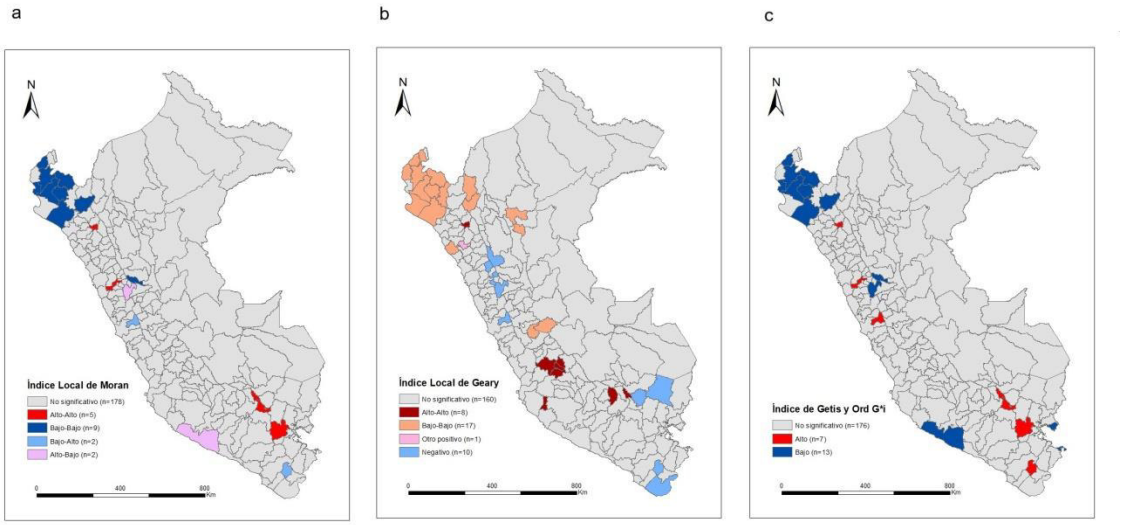


Tabla 8

Clústeres identificados con el Índice Local de Moran, Índice C de Geary y G^* Getis y Ord de las tasas de la variante Ómicrom en las provincias del Perú de enero 2021 –junio 2022.

Departamento	Provincia	I Moran	C Geary	G* Getis y Ord
Amazonas	Bagua			
	Uctubamba			
Ancash	Huari			
	Yungay			
Apurímac	Cotabamba			
Arequipa	Caravelí			
Cajamarca	Hualgayoc			
	Jaén			
Cusco	Acomayo			
	Canas			
Huancavelica	Acobamba			
	Angaraes			
	Churcampa			
	Huancavelica			
Huánuco	Huacaybamba			
Ica	Palpa			
Junín	Chanchamayo			
	Tarma			
La Libertad	Ascope			
Lambayeque	Lambayeque			
Lima	Ayón			
Piura	Ayabaca			
	Huancabamba			
	Morropón			
	Paita			

	Piura			
	Sechura			
	Sullana			
	Talara			
Puno	Lampa			
	Moho			
	Yunguyo			
San Martin	Lamas			
	Picota			
Tacna	Candarave			
Tumbes	Contralmirante Villar			
	Tumbes			

Nota. Los colores indican la formación de clústeres Alto- Alto; Bajo- Bajo; Bajo –Alto; Alto-Bajo.

El I de Moran y el C de Geary a nivel global permitieron llegar a las mismas conclusiones; sin embargo, el grado en que se correlacionan las observaciones localmente entre las provincias adyacentes presentaron ciertas diferencias en el comportamiento de las variantes del SARS-CoV-2.

Según Anselin (GeoDa Sofware, 2021) es imposible intentar comparar los resultados entre I de Moran y el C de Geary locales ya que miden diferentes criterios. Mientras que I de Moran es una prueba de correlaciones; C de Geary mide el cuadrado de las distancias entre los atributos.

Al comparar las figuras entre I Local de Moran y G^* de Getis y Ord se pueden apreciar mayores similitudes entre ambos índices; sin embargo para G^* de Getis y Ord un valor positivo señala agrupación de valores altos y un valor negativo una agrupación de valores bajos, en tanto que I Local de Moran señala como valores positivos la agrupación de valores similares (ya sean bajos o altos), y como valores negativos la agrupación de valores diferentes; dado que G^* de Getis y Ord no es una estadística de productos cruzados no mide valores atípicos (Anselin,1995).

Por otro lado, G^* de Getis y Ord identificó puntos fríos, para las cuatro variantes, en las provincias del sur del Perú que I de Moran no identificó, es importante señalar que la mayoría de estas provincias se ubican en Puno, región donde se reportaron menor número de secuencias genómicas. Por otro lado, cabe destacar que en los casos en donde la diferencia entre la unidad de estudio y la media general de los datos es cercano a 0, G^* de Getis y Ord señala los vecinos extremos mientras que I Local de Moran no (Ord, 2006), esto podría ser una probable explicación de las diferencias en estas provincias del sur del país ya que como se observa en los mapas coropléticos se puede notar la tenue coloración de las zonas vecinas que rodean estas provincias (Figura 12).
En tal sentido, se puede deducir que tanto I Local de Moran, C de Geary y G^* de Getis y Ord ofrecen información complementaria sobre el patrón de la distribución de las variantes del SAR-CoV-2 en las provincias del Perú, Por lo tanto no se puede determinar la conveniencia de un estadístico sobre otro. Debido a que

tanto el I de Moran, C de Geary y G de Getis y Ord presentan ventajas y limitaciones es conveniente el análisis de más de una técnica que permita complementar las interpretaciones.

CONCLUSIONES

En esta investigación se planteó analizar el comportamiento espacial de las variantes de SARS-CoV-2. La aplicación de los índices de autocorrelación global de Moran y C de Geary permitieron llegar a las mismas conclusiones identificando autocorrelación espacial estadísticamente significativa para las variantes Gamma, Delta y Ómicrom.

Por otro lado, los mapas elaborados a partir de los estadísticos Locales permitieron apreciar diferencias en el comportamiento de cada variante del virus a través de las provincias del Perú mostrando el predominio de una variante frente a otras en determinadas zonas al considerar el espacio geográfico como un factor importante para el análisis de la dinámica de la evolución de las variantes del virus; además brindó otra perspectiva en comparación con aquellas investigaciones donde no se considera el espacio geográfico identificando que cada variante del SARS-CoV-2 presentó un patrón de distribución diferente.

Como muestran los resultados la variante Lambda, a pesar de ser la variante dominante durante la segunda ola, muestra un patrón de conglomerados de alto riesgo dispersos en las provincias; sin embargo, es notable el predominio de zonas de bajo riesgo principalmente en provincias del oriente del país.

El comportamiento de Gamma es complementario al de Lambda, ya que ambas variantes circularon más o menos durante el mismo periodo, destacando un gran conglomerado de alto riesgo en la zona del oriente.

Los resultados de Delta indican que la agrupación de zonas de alto riesgo ocurre en las provincias del norte y centro del país.

Finalmente, Ómicrom, presenta un patrón de comportamiento más disperso que las variantes antes mencionadas, las zonas de alto riesgo se ubican dispersas en el norte, centro y sur de las provincias del Perú.

Cabe recalcar que la evaluación de los cambios en el patrón de distribución de las variantes a lo largo de las provincias del Perú, permitió determinar el ritmo de la variabilidad y propagación del virus lo cual servirá para la implementación de medidas de contención en caso de nuevos brotes, e incluso, servirá de referencia para la evaluación de otras afecciones que impliquen el factor geográfico en el desarrollo de la enfermedad.

Por otro lado, una de las limitaciones del estudio es la disposición de los datos secundarios. La base de datos disponible no era susceptible de ser analizada por técnicas exploratorias para patrones de puntos espaciales ya que la data disponible no presenta información sobre las coordenadas geográficas de los eventos; sin embargo, se disponía de la información de la provincia de dónde provenía la muestra analizada. Lo anterior conlleva a analizar la data a nivel de cuadrantes (en este caso delimitadas políticamente a nivel provincial); esta disposición de la data supone una pérdida considerable de la información espacial al considerar índices unidimensionales, además cualquier patrón espacial detectado puede depender del tamaño de la provincia al contrario de los métodos de estimación de la intensidad de un patrón de puntos y los métodos de distancia; sin embargo el análisis de datos de tipo área permiten tener una noción global de las localizaciones con un número alto o bajo de eventos por área (Fischer,2001).

RECOMENDACIONES

Es conveniente que investigaciones posteriores incluyan otras variables de investigación con el fin de asociarlas a los contagios de COVID-19 de acuerdo al tipo de Variante del virus, a través del planteamiento de modelos espaciales de regresión para la comprensión de los patrones descritos en esta investigación.

Cabe recalcar la importancia de la elaboración de programas de contención del virus priorizando las provincias más vulnerables del país en caso de que la enfermedad rebrote con la misma magnitud que se evidenció durante la pandemia.

Se advierte a los investigadores considerar el empleo de la estadística espacial para la comprensión enfermedades en donde el espacio geográfico juegue un papel importante en el desarrollo de la enfermedad.

REFERENCIAS BIBLIOGRÁFICAS

Acevedo, I., y Velásquez, H. (2008). Algunos conceptos de la econometría espacial y el análisis exploratorio de datos espaciales. *Ecos De Economía*, *12* (27), 9–34.

Aleem, A., Akbar Samad, A. B., y Slenker, A. K. (2023). *Emerging Variants of SARS-CoV-2 And Novel Therapeutics Against Coronavirus (COVID-19).* StatPearls Publishing. https://pubmed.ncbi.nlm.nih.gov/34033342/

Álvarez-Antonio, C., Meza-Sánchez, G., Calampa, C., Casanova, W., Carey, C., Alava, F., Rodriguez-Ferrucci, H. & Quispe, A. M. (2021). Seroprevalence of anti-SARS-CoV-2 antibodies in Iquitos, Peru in July and August, 2020: a population-based study. *The Lancet Global Health*, 9(7). https://doi.org/10.1016/S2214-109X(21)00173-X

Amdaoud, M., Arcuri, G. y Levratto, N. (2021). Are regions equal in adversity? A spatial analysis of spread and dynamics of COVID-19 in Europe. *The European Journal Health Economics*, *22* (4), 629–642. https://doi.org/10.1007/s10198-021-01280-6

Andrade, A., L., Vieira, A., M., Barcellos, C., Lisboa, E., P., Acosta, L., M., Mattos, M., Varnieri, M., R., Carvalho M., S., Atanaka dos Santos, M., A., Cruz, O., G., Souza dos Santos, R., Flores, R., dos Santos, S., M., Almeida e Silva, S., Souza, W., V. (2007). Introdução à Estatística Espacial para a Saúde

Pública. *Ministério da Saúde. Fundação Oswaldo Cruz. https://ares.unasus.gov.br/acervo/html/ARES/1199/1/livro_3.pdf*

Anselin, L. (1988). Lagrange Multiplier Test Diagnostics for Spatial Dependence and Spatial Heterogeneity. *Geographical Analysis*, 20 (1), 1-17. https://doi.org/10.1111/j.1538-4632.1988.tb00159.x

Anselin, L. (1995), Local Indicators of Spatial Association—LISA. *Geographical Analysis*, 27 (2), 93-115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Anselin, L. (1996). Interactive Techniques and Exploratory Spatial Data Analysis. *Regional Research Institute Working Papers*. https://researchrepository.wvu.edu/rri_pubs/200

Anselin, L. (2018). A Local Indicator of Multivariate Spatial Association: Extending Geary's c. *Geographical Analysis, 51 (2), 133-150.* https://doi.org/10.1111/gean.12164

Anselin, L. (09 de octubre de 2020). Global Spatial Autocorrelation (1). Visualizing Spatial Autocorrelation. GeoDa. https://geodacenter.github.io/workbook/5a_global_auto/lab5a.html#permutation-inference

Anselin, L., y Bao, S. (1997). Exploratory Spatial Data Analysis Linking SpaceStat and ArcView. *Regional Research Institute Working Papers*. https://doi.org/10.1007/978-3-662-03499-6_3

Anselin, L., y Getis, A. (1992). Spatial statistical analysis and geographic information systems. En Anselin, L., y Rey, S. J. (Eds.). *Perspectives on spatial data analysis* (pp. 35-47). Springer. 10.1007/978-3-642-01976-0

Anselin, L., y Rey, S. J. (Eds.). (2010). *Perspectives on spatial data analysis in Perspectives on spatial data analysis*. Springer. 10.1007/978-3-642-01976-0

Aral, N., y Bakir, H. (2021). Spatiotemporal Analysis of COVID-19 in Turkey. *Sustainable cities and society*, *76*. https://doi.org/10.1016/j.scs.2021.103421

Bivand, R.S., y Wong, D.W.S. (2018). Comparing implementations of global and local indicators of spatial association.*TEST*, 27, 716–748 https://doi.org/10.1007/s11749-018-0599-x

Centro Nacional de Planeamiento Estratégico (2020). *Análisis espacial de la situación e impactos en los servicios de salud y educación en el contexto de la pandemia por la COVI-19.* Presidencia del Consejo de Ministros. https://www.gob.pe/institucion/ceplan/informes-publicaciones/1450687-analisis-espacial-de-la-situacion-e-impactos-en-los-servicios-de-salud-y-educacion-en-el-contexto-de-la-pandemia-por-la-covid-19

Celemín, J. P. (2009). Autocorrelación espacial e indicadores locales de asociación espacial: Importancia, estructura y aplicación. *Revista Universitaria de Geografía*, *18* (1), 11-31. https://www.redalyc.org/articulo.oa?id=383239099001

Centros para el Control y la Prevención de Enfermedades. (2021,11 de agosto). *Variantes del virus.* https://espanol.cdc.gov/coronavirus/2019-ncov/variants/index.html#:~:text=La%20variante%20%C3%B3micron%20se%20propaga,el%20virus%20a%20otras%20personas.

Centros para el Control y Prevención de Enfermedades. (2022, 2 de diciembre). *¿Qué es la vigilancia genómica?* https://espanol.cdc.gov/coronavirus/2019-ncov/variants/genomic-surveillance.html

Centros para el Control y Prevención de Enfermedades. (2023, 1 de setiembre). *Clasificaciones y definiciones de las variantes del SARS-CoV-2*. https://espanol.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html

Chasco, Y., C. (2003). *Econometría espacial aplicada a la predicción-extrapolación de datos microterritoriales* [Tesis de Doctorado, Universidad Autónoma de Madrid]. http://www.madrid.org/bvirtual/BVCM005618.pdf

Chasco, Y. C. (2010). Detección de clusters y otras estructuras regionales y urbanas con técnicas de econometría espacial. *Ciudad y Territorio Estudios Territoriales, 42*(165-166), 497-512. https://recyt.fecyt.es/index.php/CyTET/article/view/76013

Chen, A. T., Altschuler, K., Zhan, S. H., Chan, Y. A., y Deverman, B. E. (2021). COVID-19 CG enables SARS-CoV-2 mutation and lineage tracking by

locations and dates of interest. *eLife*, *10*. https://doi.org/10.7554/eLife.63409

Das, R. K., y Sudaryo, M. K. (2021). Epidemiological Patterns and Spatial Distribution of COVID-19 Cases in DKI Jakarta (March–December 2020). *Jurnal Kesehatan Masyarakat Nasional (National Public Health Journal),16* (5), 17-22. https://doi.org/10.21109/kesmas.v0i0.5074

Presidencia del Consejo de Ministros. (2020). Decreto Supremo Nº 044-2020-PCM. *Decreto Supremo que declara Estado de Emergencia Nacional por las graves circunstancias que afectan la vida de la Nación a consecuencia del brote del COVID-19*. Diario Oficial El Peruano 15 de marzo de 2020. https://www.gob.pe/institucion/pcm/normas-legales/460472-044-2020-pcm

El Deeb, O. (2021). Spatial Autocorrelation and the Dynamics of the Mean Center of COVID-19 Infections in Lebanon. *Frontiers in Applied Mathematics and Statistics*, 6, 1-15. https://doi.org/10.3389/fams.2020.620064

Fehr, A. R., y Perlman, S. (2015). Coronaviruses: an overview of their replication and pathogenesis. En Maier, H., Bickerton, E., Britton, P. (Eds.), *Coronaviruses: Methods and Protocols, Methods in Molecular Biology* (pp. 1–23). Humana Press. https://doi.org/10.1007/978-1-4939-2438-7_1

Fischer M.M. (2001). Spatial Analysis in Geography. En Smelser, N. J. y Baltes P. B. (Eds.), *International Encyclopedia of the Social & Behavioral*

Sciences (pp. 14752-14758). Pergamum. https://doi.org/10.1016/B0-08-043076-7/02489-X

GeoDa Sofware. (2021, 29 de marzo). *Week 6b: Other local spatial autocorrelation statistics* (Introduction to Spatial Data Science) [Video]. YouTube. https://www.youtube.com/watch?v=h7J26lDheFQ&t=387s

Getis, A. (1991). Spatial Interaction and Spatial Autocorrelation: ACross-Product Approach. En Anselin, L., y Rey, S. J. (Eds.). *Perspectives on spatial data analysis in Perspectives on spatial data analysis*. (pp.23-33). Springer. 10.1007/978-3-642-01976-0

Getis, A. (2010). Spatial Autocorrelation. En Fischer, M., y Getis, A. (Eds). *Handbook of Applied Spatial Analysis* (pp. 255–278). Springer. https://doi.org/10.1007/978-3-642-03647-7_14

Getis, A., Ord, J. K. (1992). The Analysis of Spatial Association by Use of Distance Statistics. En Anselin, L., y Rey, S. J. (Eds.). *Perspectives on spatial data analysis in Perspectives on spatial data analysis*. (pp.127-145). Springer. 10.1007/978-3-642-01976-0

Ghafouri-Fard, S., Noroozi, R., Omrani, M. D., Branicki, W., Pośpiech, E., Sayad, A., Pyrc, K., Łabaj, P. P., Vafaee, R., Taheri, M., y Sanak, M. (2020). Angiotensin converting enzyme: A review on expression profile and its association with human disorders with special focus on SARS-CoV-2 infection. *Vascular Pharmacology, 130*. https://doi.org/10.1016/j.vph.2020.106680

Giovanetti, M., Benedetti, F., Campisi, G., Ciccozzi, A., Fabris, S., Ceccarelli, G., Tambone, V., Caruso, A., Angeletti, S., Zella, D., y Ciccozzi, M. (2021). Evolution patterns of SARS-CoV-2: Snapshot on its genome variants. *Biochemical and biophysical research communications*, *538*, 88–91. https://doi.org/10.1016/j.bbrc.2020.10.102

Global Initiative on Sharing All Influenza Data. *Tracking of hCoV-19 Variants*. Obtenido el 24 de octubre de 2022 de https://gisaid.org/hcov-19-variants-dashboard/

Instituto Nacional de Salud (noviembre de 2021). *INS llega a los primeros 10 000 genomas de SARS-CoV-2 secuenciados en el Perú* [Nota de prensa]. https://www.gob.pe/institucion/minsa/noticias/563162-ins-llega-a-los-primeros-10-000-genomas-de-sars-cov-2-secuenciados-en-el-peru

Kang, D., Choi, H., Kim, J. y Choi J. (2020). Spatial epidemic dynamics of the COVID-19 outbreak in China. *International Journal of Infectious Diseases. 94*, 96-102. https://doi.org/10.1016/j.ijid.2020.03.076

Kim S., y Castro M. (2020). Spatiotemporal pattern of COVID-19 and government response in South Korea (as of May 31, 2020). *International Journal of Infectious Diseases, 98*, 328-333. https://doi.org/10.1016/j.ijid.2020.07.004

Lizaraso Caparó, F., y Pareja Cruz, A. (2021). La variante delta y el cambio de jugada en el tablero de la salud pública. *Horizonte Médico* (Lima), 21(3). https://doi.org/10.24265/horizmed.2021.v21n3.01

McBride, K.A., Ogbo, F., Page, A. (2019). Epidemiology. En: Liamputtong, P. (Eds). *Handbook of Research Methods in Health Social Sciences*. (pp. 559–579). Springer, https://doi.org/10.1007/978-981-10-5251-4_91

Ministerio de Salud. (2021). *Boletín Epidemiológico del Perú 2021.* Informe elaborado por el Centro Nacional de Epidemiologia, Prevención y Control de Enfermedades. CDC MINSA – Centro Nacional de Epidemiología, Prevención y Control de Enfermedades. https://www.dge.gob.pe/epipublic/uploads/boletin/boletin_202124_23_145452.pdf

Ministerio de Salud. (2022). *Situación actual COVID19.Peru 2020-2022.* Informe elaborado por el Centro Nacional de Epidemiologia, Prevención y Control de Enfermedades. CDC MINSA – Centro Nacional de Epidemiología, Prevención y Control de Enfermedades. https://www.dge.gob.pe/portalnuevo/salas-situacionales/enfermedades-transmisibles/covid-19/situacion-del-covid-19-en-el-peru/

Ministerio de Salud. (2023). *Mortalidad COVID-19.* Informe elaborado por el Centro Nacional de Epidemiologia, Prevención y Control de Enfermedades. CDC MINSA – Centro Nacional de Epidemiología, Prevención y Control de Enfermedades. https://www.dge.gob.pe/portalnuevo/informacion-publica/mortalidad-covid-19/

Mousavizadeh, L., y Ghasemi, S. (2021). Genotype and phenotype of COVID-19: Their roles in pathogenesis. *Journal of microbiology, immunology, and infection*, *54* (2), 159–163. https://doi.org/10.1016/j.jmii.2020.03.022

Munazza, F., Sana, A., Ibtisam, B., y Saba, A. (2021). Geospatial Clustering and Hot Spot Detection of COVID-19 Incidence in 2020: A Global

Analysis. *International Journal of Geospatial and Environmental Research*, 8(1). https://dc.uwm.edu/ijger/vol8/iss1/4

Olaya, V. (2020). Sistemas de Información Geográfica. https://geoinnova.org/libro/sistemas-informacion-geografica/

Ord, J. K. (2006). Spatial Autocorrelation: A Statistician's Reflections. En Anselin, L., y Rey, S. J. (Eds.). *Perspectives on spatial data analysis en in Perspectives on spatial data analysis*. (pp. 165-180). Springer. https://doi.org/10.1007/978-3-642-01976-0

Organización Mundial de la Salud . (2022a). *Coronavirus*. https://www.who.int/es/health-topics/coronavirus#tab=tab_1

Organización Mundial de la Salud . (2022b).*Preguntas y respuestas sobre la transmision de COVID-19*. https://www.who.int/es/news-room/questions-and-answers/item/coronavirus-disease-covid-19-how-is-it-transmitted

Organización Mundial de la Salud . (2022c).*Seguimiento de las Variants del SARS-CoV-2*. https://www.who.int/es/activities/tracking-SARS-CoV-2-variants

Padilla-Rojas, C., Jimenez-Vasquez, V., Hurtado, V., Mestanza, O., Molina, I. S., Barcena, L., Morales Ruiz, S., Acedo, S., Lizarraga, W., Bailon, H., Cáceres, O., Galarza, M., Rojas-Serrano, N., Vargas-Herrera, N., Lope-Pari, P., Huayra, J., y Solari, L. (2021). Genomic analysis reveals a rapid spread and predominance of lambda (C.37) SARS-COV-2 lineage in Peru

despite circulation of variants of concern. *Journal of Medical Virology*. *93* (12), 6845–6849. https://doi.org/10.1002/jmv.27261

Parvin F., Ali, S. A. Nashmi, S. N. I., y Ahmad, A. (2021). Spatial prediction and mapping of the COVID-19 hotspot in India using geostatistical technique. *Spatial Information Research* , *29* (4), 479–494. https://doi.org/10.1007/s41324-020-00375-1

Plataforma digital única del Estado Peruano. (2021). *Ministro Cevallos: "Tenemos 534 casos de la variante Delta y ha pasado a ser la variante predominante" [Nota de prensa]*. https://www.gob.pe/institucion/minsa/noticias/527733-ministro-cevallos-tenemos-534-casos-de-la-variante-delta-y-ha-pasado-a-ser-la-variante-predominante

Plataforma digital única del Estado Peruano. (2023). *Coronavirus en el Perú: casos confirmados [Nota de prensa]*. https://www.gob.pe/8662-ministerio-de-salud-coronavirus-en-el-peru

Quinino, L. R. de M., Vasconcellos, F. H. M., de Oliveira, I. S. D., Aguiar, L. R., Ramos, Y. T. de M., y de Bastiane, F. (2021). Aspectos espacial e temporal e fatores associados à interiorização da COVID-19 em Pernambuco, Brasil. *Ciencia & saude coletiva*, *26* (6), 2171–2182. https://doi.org/10.1590/1413-81232021266.00642021

Sanjuán, R., & Domingo-Calap, P. (2021). Genetic Diversity and Evolution of Viral Populations. *Encyclopedia of Virology*, 4 (1), 53–61. https://doi.org/10.1016/B978-0-12-809633-8.20958-8

Sarkar, S. K., Ekram, K. M. M., y Das, P. C. (2021). Spatial modeling of COVID-19 transmission in Bangladesh. *Spatial Information Research*, *29* (5), 715-726. https://doi.org/10.1007/s41324-021-00387-5

Siabato, W., y Guzmán-Manrique, J. (2019). La autocorrelación espacial y el desarrollo de la geografía cuantitativa. *Cuadernos de Geografía: Revista Colombiana de Geografía*, *28* (1), 1-22. https://doi.org/10.15446/rcdg.v28n1.76919

Taboada, B., Zárate, S., Iša, P., Boukadida, C., Vazquez-Perez, J. A., Muñoz-Medina, J. E., Ramírez-González, J. E., Comas-García, A., Grajales-Muñiz, C., Rincón-Rubio, A., Matías-Florentino, M., Sanchez-Flores, A., Mendieta-Condado, E., Verleyen, J., Barrera-Badillo, G., Hernández-Rivas, L., Mejía-Nepomuceno, F., Martínez-Orozco, J. A., Becerril-Vargas, E., … Arias, C. F., (2021). Genetic Analysis of SARS-CoV-2 Variants in Mexico during the First Year of the COVID-19 Pandemic. *Viruses, 13* (11). https://doi.org/10.3390/v13112161

Vargas-Herrera, N., Araujo-Castillo, R. V., Mestanza, O., Galarza, M., Rojas-Serrano, N., y Solari-Zerpa, L. (2021). SARS-CoV-2 Lambda and Gamma variants competition in Peru, a country with high seroprevalence. *The Lancet. 6*. https://doi.org/10.1016/j.lana.2021.100112

Waller, L. A., y Gotway, C. A. (2004). Applied spatial statistics for public health data (1ra ed.). John Wiley & Sons. DOI:10.1002/0471662682

Wilson M. E. (2017). Geography of Infectious Diseases. En Cohen, J. Powderly, W. G., y Opal, S. M. (Eds.), *Infectious Diseases* (pp. 938–947). Elsevier. https://doi.org/10.1016/B978-0-7020-6285-8.00106-4

Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., Hu, Y., Tao, Z. W., Tian, J. H., Pei, Y. Y., Yuan, M. L., Zhang, Y. L., Dai, F. H., Liu, Y., Wang, Q. M., Zheng, J. J., Xu, L., Holmes, E. C., y Zhang, Y. Z. (2020). A new coronavirus associated with human respiratory disease in China. *Nature,* 579 (7798),265-269. https://doi.org/10.1038/s41586-020-2008-3

Symanzik, J. (2014). Exploratory Spatial Data Analysis. En Fischer, M., Nijkamp, P. (Eds.). *Handbook of Regional Science*. Springer. (pp. 1295-1310). https://doi.org/10.1007/978-3-642-23430-9_76

ANEXOS

Anexo 1.

Matriz de consistencia del proyecto de investigación Análisis espacial de la tasa de las variantes del SARS-Cov-2 en Peru,2021-2022.

Problema	Objetivos	Hipótesis
General	**General**	**General**
¿Cuál es el patrón en la distribución espacial de la tasa de las variantes del SARS-CoV-2 en las provincias del Perú?	Analizar el patrón en la distribución espacial de las variantes del SARS-CoV-2 predominantes en las provincias del Perú.	No aplica
Específicos	**Específicos**	**Específicas**
- ¿Cuál es el patrón en la distribución espacial de la tasa de la variante Gamma del SARS-CoV-2 en las provincias del Perú?	- Analizar el patrón en la distribución espacial de la variante Gamma del SARS-CoV-2 en las provincias del Perú.	No aplica

- ¿Cuál es el patrón en la distribución espacial de la tasa de la variante Delta del SARS-CoV-2 en las provincias del Perú?	- Analizar el patrón en la distribución espacial de la variante Delta del SARS-CoV-2 en las provincias del Perú.
- ¿Cuál es el patrón en la distribución espacial de la tasa de la variante Lambda del SARS-CoV-2 en las provincias del Perú?	- Analizar el patrón en la distribución espacial de la variante Lambda del SARS-CoV-2 en las provincias del Perú.
- ¿Cuál es el patrón en la distribución espacial de la tasa de la variante Ómicrom del SARS-CoV-2 en las provincias del Perú?	- Analizar el patrón en la distribución espacial de la variante Ómicrom del SARS-CoV-2 en las provincias del Perú.

Anexo 2.

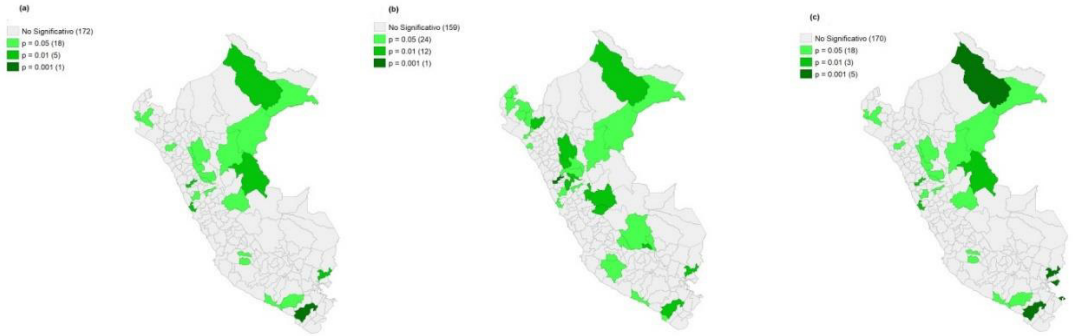
Diseños de estudio.



Nota. De “Epidemiology”, de McBride, K.A., Ogbo, F., Page, A., 2019. En: Liamputtong, P. (Eds). Handbook of Research Methods in Health Social Sciences, pag. 564, Springer, https://doi.org/10.1007/978-981-10-5251-4_91

Anexo 2.

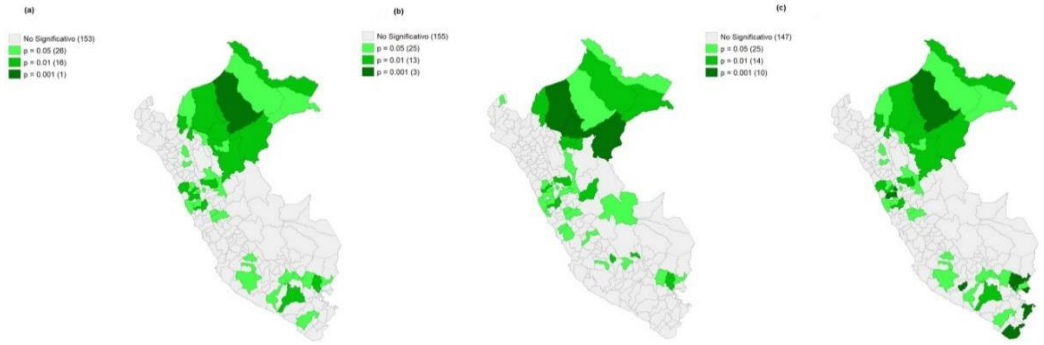
Mapa de significancia de las tasas de la variante Lambda en las provincias del Perú de enero 2021 –junio 2022



Nota. Se aplicó 999 permutaciones. Índice Local de Moran (a), Índice C de Geary (b) y G^* Getis y Ord (c).

Anexo 3.

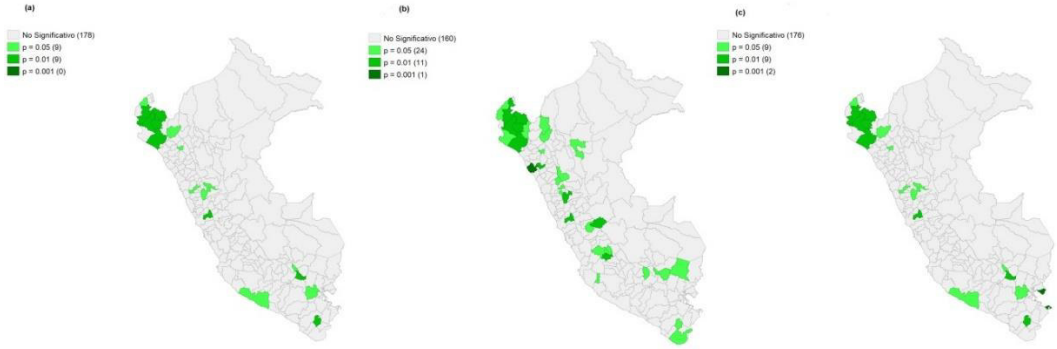
Mapa de significancia de las tasas de la variante Gamma en las provincias del Perú de enero 2021 –junio 2022



Nota. Se aplicó 999 permutaciones. Índice Local de Moran (a), Índice C de Geary (b) y G^* Getis y Ord (c).

Anexo 4.

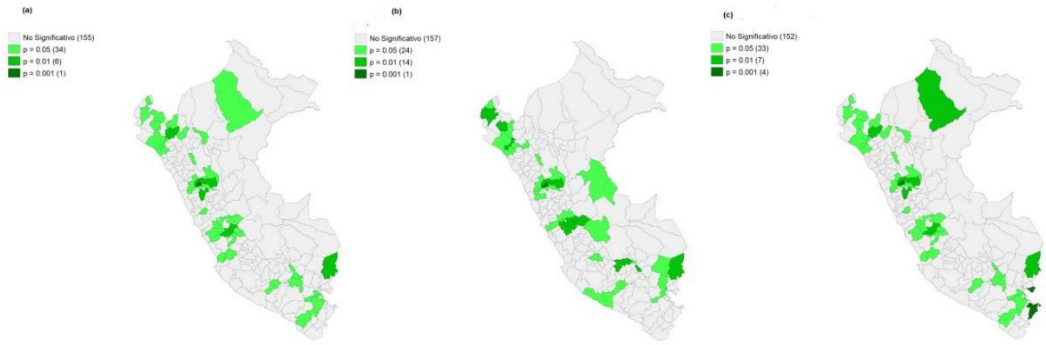
Mapa de significancia de las tasas de la variante Delta en las provincias del Perú de enero 2021 –junio 2022



Nota. Se aplicó 999 permutaciones. Índice Local de Moran (a), Índice C de Geary (b) y G^* Getis y Ord (c).

Anexo 5.

Mapa de significancia de las tasas de la variante Ómicrom en las provincias del Perú de enero 2021 –junio 2022



Nota. Se aplicó 999 permutaciones. Índice Local de Moran (a), Índice C de Geary (b) y G^* Getis y Ord (c).