

**UNIVERSIDAD NACIONAL AGRARIA**

**LA MOLINA**

**FACULTAD DE INGENIERÍA AGRÍCOLA**



**“IMPACTO DEL ACCESO AL SANEAMIENTO EN LA  
PROPAGACIÓN DEL COVID-19: APLICACIÓN DE MODELOS  
GLOBALES CON GIS”**

**TESIS PARA OPTAR TÍTULO DE  
INGENIERA AGRÍCOLA**

**KAREN IMETTE ARANGUREN TUTAYA**






**LIMA – PERÚ**

**2022**

## Document Information

Analyzed document	TESIS-ARANGUREN_02.01.23.pdf (D154879115)
Submitted	1/4/2023 1:04:00 PM
Submitted by	LIA RAMOS FERNANDEZ
Submitter email	liarf@lamolina.edu.pe
Similarity	1%
Analysis address	liarf.unalm@analysis.arkund.com

## Sources included in the report

<b>SA</b>	<b>GRUPO 6 (1).docx</b> Document GRUPO 6 (1).docx (D86733368)		<b>2</b>
<b>W</b>	URL: <a href="https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at...">https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at...</a> Fetched: 11/12/2020 4:14:23 AM		<b>1</b>
<b>W</b>	URL: <a href="https://www.nature.com/articles/s41579-020-00459-7">https://www.nature.com/articles/s41579-020-00459-7</a> Fetched: 11/3/2020 11:57:59 AM		<b>1</b>
<b>SA</b>	<b>TFM_RojasQuiroz_VersionFinal.pdf</b> Document TFM_RojasQuiroz_VersionFinal.pdf (D80748444)		<b>1</b>
<b>SA</b>	<b>7076-Rimachi Fernandez, Manuel Antonio.pdf</b> Document 7076-Rimachi Fernandez, Manuel Antonio.pdf (D36075025)		<b>1</b>

## Entire Document

La UNALM es titular de los derechos patrimoniales de la presente investigación (Art. 24 – Reglamento de Propiedad Intelectual) UNIVERSIDAD NACIONAL AGRARIA LA MOLINA FACULTAD DE INGENIERÍA AGRÍCOLA “IMPACTO DEL ACCESO AL SANEAMIENTO EN LA PROPAGACIÓN DEL COVID-19: APLICACIÓN DE MODELOS GLOBALES CON GIS” TESIS PARA OPTAR TÍTULO DE INGENIERO AGRÍCOLA KAREN IMETTE ARANGUREN TUTAYA LIMA – PERÚ 2022 DEDICATORIA A mi familia por su apoyo. .

AGRADECIMIENTOS A la Dra. Lía Ramos Fernández, patrocinadora de este trabajo por su guía y tiempo dedicado para la culminación de esta investigación. A mi familia por su apoyo incondicional.

**UNIVERSIDAD NACIONAL AGRARIA LA  
MOLINA**

**FACULTAD DE INGENIERÍA AGRÍCOLA**

**“IMPACTO DEL ACCESO AL SANEAMIENTO EN LA  
PROPAGACIÓN DEL COVID-19: APLICACIÓN DE MODELOS  
GLOBALES CON GIS”**

TESIS PARA OPTAR EL TÍTULO DE:

**INGENIERA AGRÍCOLA**

Presentado por:

**BACH. KAREN IMETTE ARANGUREN TUTAYA**

Sustentado y aprobado por el siguiente jurado:

Mg. Sc. ROSA MARÍA MIGLIO TOLEDO  
Presidente

Dra. LIA RAMOS FERNANDEZ  
Asesor

Dr. EDUARDO ABRAHAM CHÁVARRI VELARDE  
Miembro

Mag. SAÚL MOISÉS TORRES MURGA  
Miembro

Mg. KERLIM ARTURO VERGARA VILLARINO  
Co-Asesor

LIMA – PERÚ

2022

## **DEDICATORIA**

A mi familia por su apoyo.

## **AGRADECIMIENTOS**

A la Dra. Lía Ramos Fernández, patrocinadora de este trabajo por su guía y tiempo dedicado para la culminación de esta investigación.

A mi familia por su apoyo incondicional.

# ÍNDICE GENERAL

RESUMEN .....	x
ABSTRACT .....	xi
<b>I. INTRODUCCIÓN .....</b>	<b>1</b>
1.1. OBJETIVOS.....	2
<b>II. REVISIÓN DE LITERATURA .....</b>	<b>3</b>
2.1. COVID-19.....	3
2.2. SANEAMIENTO .....	4
2.2.1. Panorama .....	5
2.2.2. Relación del saneamiento con el COVID-19.....	6
2.3. MODELO GLOBALES .....	7
2.3.1. Mínimos cuadrados ordinarios (MCO).....	8
2.3.2. Modelos considerando dependencia espacial .....	9
<b>III. METODOLOGÍA.....</b>	<b>12</b>
3.1. ZONA DE ESTUDIO .....	12
3.2. CARACTERIZACIÓN DE INFORMACIÓN.....	17
3.2.1. Servicios de saneamiento y otras variables .....	17
3.2.2. Casos positivos y tasa de incidencia de COVID .....	19
3.3. MODELOS GLOBALES PARA AJUSTAR LA TASA DE INCIDENCIA .....	20
<b>IV. RESULTADOS Y DISCUSIÓN.....</b>	<b>27</b>
4.1. CARACTERIZACIÓN DE LOS SERVICIOS DE SANEAMIENTO .....	27
4.1.1. Agua potable y alcantarillado .....	27
4.1.2. Residuos sólidos .....	32
4.2. CARACTERIZACIÓN DE CASOS POSITIVOS Y TASA DE INCIDENCIA DE COVID-19.....	34
4.2.1. Número de casos positivos .....	34
4.2.2. Tasa de incidencia.....	38
4.3. MODELOS GLOBALES PARA AJUSTAR LA TASA DE INCIDENCIA DE COVID-19 EN TODOS LOS DISTRITOS PERTENECIENTES A LA CUENCA.....	42
4.3.1. Correlación no paramétrica de la tasa de incidencia y servicios de saneamiento .....	42

4.3.2. Ajuste de modelos globales .....	46
<b>V. CONCLUSIONES .....</b>	<b>56</b>
<b>VI. RECOMENDACIONES .....</b>	<b>58</b>
<b>VII. BIBLIOGRAFÍA .....</b>	<b>59</b>
<b>VIII. ANEXOS.....</b>	<b>65</b>

## ÍNDICE DE TABLAS

Tabla 1: Población y la disponibilidad de agua para Lima Metropolitana en comparación con otras ciudades capitales de América Latina .....	6
Tabla 2: Variables utilizadas en la obtención del modelo .....	15
Tabla 3: Estaciones meteorológicas a utilizarse para la interpolación .....	18
Tabla 4: Generación de residuos sólidos por provincia en la cuenca Rímac .....	33
Tabla 5: Números de casos positivos por provincia .....	35
Tabla 6: Número de casos positivos por provincia según sexo .....	36
Tabla 7: Coeficiente de correlación no paramétrica Kendall ( $\tau$ ) y Spearman ( $\rho$ ) entre la tasa de incidencia de COVID-19 y los servicios de saneamiento en la cuenca Rímac .....	44
Tabla 8: Índice de Morán para la tasa de incidencia de COVID-19 .....	46
Tabla 9: Datos estadísticos de la ecuación de regresión para el modelo de mínimos cuadrados ordinarios .....	47
Tabla 10: Resultados del modelo MCO para todos los meses de estudio .....	48
Tabla 11: Propuesta de nivel de impacto de la tasa de incidencia modelada .....	49
Tabla 12: Propuesta de nivel de impacto de la clasificación y el efecto de las variables independientes .....	49
Tabla 13: Parámetros de los modelos de retraso espacial y error espacial .....	54
Tabla 14: Coeficiente de determinación ( $R^2$ ) y criterio de información de Akaike (AIC) para los modelos de mínimos cuadrados ordinarios (OLS), retraso espacial (SLM) y error espacial (SEM) .....	55
Tabla 15: Población y nivel de acceso a agua potable y alcantarillado por distrito en la cuenca Rímac .....	66
Tabla 16: Cobertura de agua potable por provincia en la cuenca Rímac .....	68
Tabla 17: Cobertura por red pública de alcantarillado por provincia en la cuenca Rímac .....	68
Tabla 18: Porcentaje de composición de residuos sólidos por provincia en la cuenca Rímac .....	68



## ÍNDICE DE FIGURAS

Figura 1: Matriz de pesos espaciales .....	9
Figura 2: Ubicación de la cuenca del río Rímac.....	13
Figura 3: Variación diaria de la temperatura promedio según estaciones en la cuenca baja.....	17
Figura 4: Variación diaria de la temperatura promedio según estaciones en la cuenca media .....	18
Figura 5: Variación diaria de la temperatura promedio según estaciones en la cuenca alta .....	18
Figura 6: Diagrama de cajas para la temperatura promedio en las estaciones de la cuenca Rímac.....	19
Figura 7: Estrategia de selección de modelos.....	24
Figura 8: Cobertura de agua potable según la frecuencia por días de la semana por provincia.....	27
Figura 9: Variación espacial de la cobertura de agua potable por red pública por distrito .	28
Figura 10: Cobertura de agua potable en red pública por distrito en las provincias de Huarochirí (a), Callao (b) y Lima (c) .....	29
Figura 11: Distribución espacial del tipo de pago del servicio de agua en la cuenca Rímac.....	30
Figura 12: Variación espacial de la cobertura de alcantarillado por red pública a nivel distrital.....	31
Figura 13: Cobertura de alcantarillado por red pública por provincia: Huarochirí (a), Callao (b) y Lima (c) .....	32
Figura 14: Composición de los residuos sólidos en los distritos de la cuenca Rímac: orgánicos (a.), inorgánicos (b), no aprovechables (c) y peligrosos (d) categorizada por cuantiles.....	34
Figura 15: Evolución mensual del número de casos positivos de COVID-19 por provincia.....	35
Figura 16: Pirámide poblacional de los casos positivos de COVID-19 por provincia.....	37
Figura 17: Tasa de incidencia de Covid-19 por provincia: Huarochirí (a), Callao (b) y Lima (c) .....	39
Figura 18: Distribución espacial de la tasa de incidencia por distritos .....	41
Figura 19: Correlación de Spearman de la tasa de incidencia COVID-19 y los servicios de saneamiento en la cuenca Rímac .....	45

Figura 20: Distribución de las variables altitud y porcentaje de cobertura de la población que tiene acceso a agua por medio de pago a camión cisterna u otro con la distribución por niveles de la tasa de incidencia de COVID-19.....	52
Figura 21: Comparativa entre el nivel de la tasa de incidencia modelada y el efecto de variables independientes para el mes de agosto 2020 .....	53
Figura 22: Distribución espacial de las variables sociodemográficas en la cuenca Rímac .	69
Figura 23: Índice de Morán para la tasa de incidencia del mes de marzo.....	70
Figura 24: Índice de Morán para la tasa de incidencia del mes de abril.....	70
Figura 25: Índice de Morán para la tasa de incidencia del mes de mayo .....	71
Figura 26: Índice de Morán para la tasa de incidencia del mes de junio.....	71
Figura 27: Índice de Morán para la tasa de incidencia del mes de julio.....	72
Figura 28: Índice de Morán para la tasa de incidencia del mes de agosto .....	72
Figura 29: Índice de Morán para la tasa de incidencia del mes de septiembre .....	73
Figura 30: Índice de Morán para la tasa de incidencia del mes de octubre .....	73
Figura 31: Índice de Morán para la tasa de incidencia del mes de noviembre .....	74
Figura 32: Índice de Morán de los residuos del modelo para el mes de marzo: diagrama de dispersión (a), mapa de conglomerados (b) y mapa de significancia (c).....	75
Figura 33: Índice de Morán de los residuos del modelo para el mes de abril: diagrama de dispersión (a), mapa de conglomerados (b) y mapa de significancia (c).....	75
Figura 34: Índice de Morán de los residuos del modelo para el mes de mayo: diagrama de dispersión (a), mapa de conglomerados (b) y mapa de significancia (c).....	76
Figura 35: Índice de Morán de los residuos del modelo para el mes de junio: diagrama de dispersión (a), mapa de conglomerados (b) y mapa de significancia (c).....	76
Figura 36: Índice de Morán de los residuos del modelo para el mes de julio: diagrama de dispersión (a), mapa de conglomerados (b) y mapa de significancia (c).....	77
Figura 37: Índice de Morán de los residuos del modelo para el mes de agosto: diagrama de dispersión (a), mapa de conglomerados (b) y mapa de significancia (c).....	77
Figura 38: Índice de Morán de los residuos del modelo para el mes de septiembre: diagrama de dispersión (a), mapa de conglomerados (b) y mapa de significancia (c).....	78
Figura 39: Índice de Morán de los residuos del modelo para el mes de octubre: diagrama de dispersión (a), mapa de conglomerados (b) y mapa de significancia (c).....	78
Figura 40: Índice de Morán de los residuos del modelo para el mes de noviembre: diagrama de dispersión (a), mapa de conglomerados (b) y mapa de significancia (c).....	79

## ÍNDICE DE ANEXOS

ANEXO 1: ACCESO A LOS SERVICIOS DE SANEAMIENTO POR DISTRITO .....	66
ANEXO 2: CARACTERIZACIÓN DE LAS VARIABLES SOCIODEMOGRÁFICAS .	69
ANEXO 3: TEST DE AUTOCORRELACIÓN DE LA VARIABLE TASA DE INCIDENCIA DE COVID-19.....	70
ANEXO 4: TEST DE ÍNDICE DE MORÁN PARA LOS RESIDUOS DEL MODELO DE MÍNIMOS CUADRADOS ORDINARIOS .....	75

## LISTA DE ABREVIATURAS

<b>Abreviatura</b>	<b>Significado</b>
Ag_nopagan	Porcentajes de viviendas que no pagan el servicio de agua
Ag_nopub	Porcentaje de viviendas no conectadas a la red pública; abastecidas por camión, cisterna, pozo, manantial, río, acequia, lago, laguna u otro.
AgPaga_EPS	Porcentajes de viviendas que pagan el servicio de agua a una empresa prestadora de servicio
Ag_PagaMuniOrg	Porcentajes de viviendas que pagan el servicio de agua a la municipalidad o una organización comunal
Ag_Pagan	Porcentajes de viviendas que sí pagan el servicio de agua
AgPagaotr	Porcentajes de viviendas que pagan el servicio de agua a camiones cisterna u otro similar, pozo o manantial, puquio, rio, acequia, lago, laguna, otros.
Ag_RP18ho	Porcentaje de viviendas que tiene agua todos los días, pero de 10 a 18 horas
Ag_RP24h	Porcentaje de viviendas que tiene agua todos los días, pero de 19 a 24 horas
Ag_RP7dia	Porcentaje de viviendas que tienen agua los 7 días de la semana
Ag_RP9ho	Porcentaje de viviendas que tiene agua todos los días, pero de 1 a 9 horas
Ag_RPdentro	Porcentajes de viviendas conectadas a la red pública dentro de su vivienda
Ag_RPfuera	Porcentajes de viviendas conectadas a la red pública fuera de la vivienda
Ag_RPnotdia	Porcentaje de viviendas que no tienen agua los 7 días de la semana (aquí se encuentran las viviendas de 1 a 6 días)
Ag_RPubl	Porcentaje de viviendas conectadas a la red pública, ya sea dentro o fuera de la vivienda
Altitud	Altitud en metros sobre el nivel del mar

Asegurados	Porcentaje de población con afiliación a algún tipo de seguro de salud
D_NoPub	Porcentaje de viviendas no conectadas a la red pública de desagüe (pozo séptico, pozo ciego, biodigestor, cuerpo abierto, otros)
D_RPdentro	Porcentaje de viviendas conectadas a la red pública de desagüe dentro de la vivienda
D_RPfuera	Porcentaje de viviendas conectadas a la red pública de desagüe fuera de la vivienda, pero dentro de la edificación
D_RPubl	Porcentaje de viviendas conectadas a la red pública de desagüe dentro o fuera de la vivienda
Densi_hab_km2	Densidad poblacional en habitantes por kilómetro cuadrado
Hombres	Porcentaje de hombres
Hrmax	Humedad relativa máxima promedio mensual
Hrmin	Humedad relativa mínima promedio mensual
Hrprom	Humedad relativa media promedio mensual
INEI	Instituto Nacional de Estadística e Informática
Luz_RPubl	Porcentaje de viviendas con alumbrado público
Mas65	Porcentaje de la población con 65 años o más
MCO	Mínimos Cuadrados Ordinarios
Menos14	Porcentaje de la población menor de 14 años
Mercados_total	Cantidad total de mercados
Mujeres	Porcentaje de mujeres
NLEmas15	Porcentaje de la población que no lee menor de 15 años
OLS	<i>Ordinary Least Square</i> (Mínimos cuadrados ordinarios)
OMS	Organización Mundial de la Salud
PEA	Porcentaje de la población que se encuentra en la edad económicamente activa
Quechua	Porcentaje de la población que se identifica quechua
R_ComNoap	Porcentaje de residuos no aprovechables
R_ComOrg	Porcentaje de residuos orgánicos
R_ComPeli	Porcentaje de residuos peligrosos
R_CompInor	Porcentaje de residuos inorgánicos
R_Gndomici	Generación de residuos sólidos domiciliarios per cápita

R_GnMunici	Generación de residuos sólidos municipales per cápita
R_PtoCriti	Cantidad de puntos críticos
SLM	<i>Spatial Lag Model</i> (Modelo de retraso espacial)
SEM	<i>Spatial Error Model</i> (Modelo de error espacial)
Rural	Porcentaje de población rural
Tmax	Temperatura máxima promedio mensual
Tmin	Temperatura mínima promedio mensual
Tprom	Temperatura media promedio mensual
Urbana	Porcentaje de población urbana
Vviento	Velocidad del viento promedio mensual

## RESUMEN

En este estudio, se evaluó el impacto del acceso al saneamiento en la propagación del COVID-19 para el periodo de la primera ola, esto mediante la aplicación de modelos globales con sistemas de información geográfica (GIS). Se analizó la asociación espacial global entre las variables relacionadas al saneamiento, con la tasa de incidencia del COVID-19 en la cuenca del río Rímac, mediante el uso de modelos globales tales como mínimos cuadrados ordinarios (OLS), modelo de retraso espacial (SLM) y el modelo de error espacial (SEM). Se evaluó el grado de asociación entre el nivel de acceso a los servicios de saneamiento, además de incluir variables sociodemográficas y meteorológicas para determinar el poder explicativo de estas. Además, no se tuvo en cuenta la vacunación puesto que esta es un proceso dinámico y sigue en proceso de aplicación.

Los resultados mostraron que las variables altitud y porcentaje de viviendas que se abastecen de agua por medio de camión cisterna u otro, se relacionaron de manera negativa y positiva respectivamente, encontrándose que este modelo se repite en 7 de los 9 meses con valores de  $R^2$  por encima del 0.60. Además, al presentarse autocorrelación espacial, se buscó corregir ello mediante la aplicación de modelos que incorporen este fenómeno, encontrando una mejora en los valores de coeficiente de determinación, por ende, un mejor ajuste.

**Palabras clave:** OLS, COVID-19, modelos globales, saneamiento, SIG.

## **ABSTRACT**

In this study, the impact of access to sanitation on the spread of COVID-19 was evaluated for the period of the first wave, by applying global models with geographic information systems (GIS). The global spatial association between the variables related to sanitation, with the incidence rate of COVID-19 in the Rímac river basin, was analyzed through the use of global models such as ordinary least squares (OLS), spatial delay model (SLM) and the spatial error model (SEM). The degree of association between the level of access to sanitation services was evaluated, in addition to including sociodemographic and meteorological variables to determine their explanatory power. In addition, vaccination was not taken into account since this is a dynamic process and is still in the process of being applied.

The results showed that the variables altitude and percentage of houses that are supplied with water by means of a tanker truck or another, were related in a negative and positive way, respectively, finding that this model is repeated in 7 of the 9 months with values of R<sup>2</sup> for above 0.60. In addition, when presenting spatial autocorrelation, we sought to correct this by applying models that incorporate this phenomenon, finding an improvement in the coefficient of determination values, therefore, a better fit.

**KeyWords:** OLS, COVID-19, global models, sanitation, GIS.



## I. INTRODUCCIÓN

Tras el brote de la enfermedad por un nuevo coronavirus (COVID-19) en China, con una alarmante propagación a escala internacional y un aumento del número de casos, la Organización Mundial de la Salud (OMS) la declaró pandemia mundial, además exhortó a los países a juntar esfuerzos en la lucha contra la emergencia de salud pública actual (Organización Mundial de la salud, 2020).

Ante la problemática de la pandemia por COVID-19 se requiere la participación y colaboración interdisciplinaria para contribuir en la explicación de los factores que podrían estar asociados con la propagación de este virus; que actualmente ha reportado no sólo pérdidas humanas, sino también económicas, sociales y medioambientales. Si bien existen numerosas investigaciones que abordan esta problemática, fuentes como Eichelberger *et al.* (2021), Ekumah *et al.* (2020) y UNICEF (2020) manifiestan que las diferencias existentes en las comunidades en relación al acceso al agua y saneamiento, así como la importancia de las condiciones higiénicas cumplen un rol vital para explicar las posibles causas de la propagación de este virus. Por su parte Aabed & Lashin (2021), realizaron un estudio sobre los factores ambientales que se relacionan con la incidencia del COVID-19, este análisis se realizó en 188 países y se encontró un aumento de la tasa de incidencia en países templados en relación a los tropicales. Adicionalmente Mollalo *et al.* (2020) concluyen en su investigación que existen una relación positiva de factores sociodemográficos como el ingreso familiar medio en la propagación del virus.

Tal como manifiesta León (2020), existe la necesidad de agilizar el conocimiento del comportamiento del COVID-19, por lo que se exige disponer de soluciones informáticas adecuadas que permitan incrementar los esfuerzos para lograrlo. Con esto, se confirma la importancia de los conocimientos matemáticos que permiten estimar comportamientos, a través de modelos, para tener un panorama más amplio de lo que acontece con la enfermedad (Yojcon & Ruiz, 2020). Agregando a ello Velasco *et al.* (2021) en su estudio con datos de

141 países encontró satisfactoriamente, con el uso del modelo de mínimos cuadrados ordinarios, los factores que podrían influir en la cantidad de casos y muertes por COVID-19.

### **1.1. OBJETIVOS**

Evaluar la relación entre el nivel de atención de servicios de saneamiento y recolección de residuos sólidos, con la propagación del COVID-19 en la cuenca del río Rímac, mediante la aplicación de modelos globales con GIS.

Esto se consiguió cumpliendo los siguientes objetivos específicos:

- Realizar un diagnóstico del nivel de atención de los servicios de saneamiento y recolección de residuos sólidos en la cuenca del río Rímac.
- Caracterizar las variables asociadas al COVID-19, tales como la cantidad de casos positivos y la tasa de incidencia de manera mensual.
- Obtener la correlación paramétrica y no paramétrica de las variables del COVID 19, con relación a los servicios de saneamiento.
- Calibrar el modelo de mínimos cuadrados ordinarios, de retraso y error espacial para ajustar la tasa de incidencia de COVID-19.

## II. REVISIÓN DE LITERATURA

### 2.1. COVID-19

Los coronavirus son una extensa familia de virus que pueden causar enfermedades tanto en animales como en humanos. En los humanos, se sabe que varios coronavirus causan infecciones respiratorias que pueden ir desde el resfriado común hasta enfermedades más graves como el síndrome respiratorio de Oriente Medio (MERS) y el síndrome respiratorio agudo severo (SARS) (OMS, 2020).

Tal como da a conocer Aragón-Nogales *et al.* (2019) los coronavirus están muy extendidos y pueden infectar a seres humanos, mamíferos y aves, provocando trastornos respiratorios, intestinales, hepáticos y neurológicos. Según el Comité Internacional de Taxonomía de Virus, los coronavirus pertenecen al orden Nidovirales, familia Coronaviridae, subfamilia Coronavirinae, esta última consta de cuatro géneros Alphacoronavirus, Betacoronavirus, Gammacoronavirus y Deltacoronavirus.

La enfermedad por coronavirus 2019, denominada COVID-19, es una enfermedad infecciosa respiratoria emergente causada por el coronavirus del síndrome respiratorio agudo grave de tipo 2, también conocido como SARS-CoV-2 por sus siglas en inglés *Severe acute respiratory síndrome associated coronavirus*, la cual se detectó por primera vez a principios de diciembre de 2019 en la ciudad de Wuhan, China (Pan *et al.*, 2020).

Las señales más comunes comprenden la fiebre, tos y sensación de falta de aire, a pesar de que también se ha observado que puede causar una enorme variedad de síntomas, como la anosmia o la ageusia, o incluso pasar inadvertida al ser asintomática (Ministerio de Sanidad, 2020).

Como lo hace notar Rahimi *et al.* (2021) existen dos características primordiales del COVID-19 que han hecho que este virus se extienda de las fronteras geográficas, la característica clínica y la transmisión rápida de persona a persona. La primera alude a que la enfermedad responde poco o casi nada a los tratamientos convencionales; por ende, el tratamiento de aquellos que se encuentran en condiciones críticas es un desafío. La segunda, implica la transmisión rápida por el medio ambiente, que es también vital en el desarrollo global del COVID-19, factores tales como el aire, aguas residuales, el agua, los factores meteorológicos (velocidad del viento, la humedad absoluta, la luz solar, la presión atmosférica, etc.) y los fómites son las causas comúnmente conocidas del aspecto ambiental de COVID-19.

## **2.2. SANEAMIENTO**

El Decreto Legislativo 1280 que aprueba la Ley Marco de la Gestión y Prestación de los Servicios de Saneamiento define que los servicios de saneamiento están conformados por sistemas y procesos, de acuerdo con el siguiente detalle:

- a. Servicio de agua potable:
  - Sistema de producción, que comprende los procesos de: captación, almacenamiento y conducción de agua cruda; tratamiento y conducción de agua tratada, mediante cualquier tecnología.
  - Sistema de distribución, que comprende los procesos de: almacenamiento, distribución, entrega y medición al usuario mediante cualquier tecnología.
- b. Servicio de alcantarillado sanitario, que comprende los procesos de: recolección, impulsión y conducción de aguas residuales hasta el punto de entrega para su tratamiento.
- c. Servicio de tratamiento de aguas residuales para disposición final o reúso, que comprende los procesos de mejora de la calidad del agua residual proveniente del servicio de alcantarillado mediante procesos físicos, químicos, biológicos u otros, y los componentes necesarios para la disposición final o reúso.
- d. Servicio de disposición sanitarias de excretas, que comprende los procesos para la

disposición final del agua residual y la disposición sanitaria de excretas a nivel intradomiciliario, con o sin arrastre hidráulico.

### **2.2.1. Panorama**

El saneamiento deficiente y la falta de una calidad adecuada del agua han sido responsables de muchas enfermedades y muertes en todo el mundo. Se estima que 1,6 millones de muertes y 105 millones de años de vida ajustados por discapacidad (AVAD) son atribuibles a agua, saneamiento e higiene (WASH) inadecuado, incluidas solo enfermedades que podrían cuantificarse, lo que representa el 2,8 por ciento del total de muertes y el 3,9 por ciento del total de AVAD en 2016. De esas, 829 000 muertes se deben a enfermedades diarreicas. El 60 por ciento de la carga general de diarrea, el 13 por ciento de la carga de las infecciones respiratorias agudas, el 16 por ciento de la carga de la desnutrición proteico-energética, el 43 por ciento de la carga de la esquistosomiasis, el 80 por ciento de la carga de la malaria y el 100 por ciento de ambas cargas de las infecciones por helmintos transmitidos por el suelo y la carga del tracoma se atribuyen a un WASH inadecuado (Prüss-Ustün *et al.*, 2019).

En el caso del Perú, según un reporte del Instituto Nacional de Estadística e Informática (INEI) del 2017, la pobreza extrema afectó al 3,8 por ciento de la población del país, equivalente a 1 millón 205 mil personas, el 54,1 por ciento de los pobres se localiza en el área urbana y el 45,9 por ciento en el área rural. Según la región natural, el 46,2 por ciento de pobres están en la sierra, el 36,9 por ciento en la costa y el 16,9 por ciento en la selva. Asimismo, se indica que el 80,4 por ciento de viviendas se provee de agua por red pública, ya sea dentro o fuera de la vivienda. En el área urbana, este servicio cubre 83,2 por ciento; mientras que en el área rural al 71,1 por ciento de las viviendas. El 71,3 por ciento de las viviendas contaba con servicio de saneamiento conectado a red pública, 12,0 puntos porcentuales más que en el 2012. En el área urbana 87,2 por ciento de las viviendas contaba con este tipo de servicio de saneamiento, en cambio en el área rural solo el 20,0 por ciento disponía de este servicio; todo lo cual determina la presencia de enfermedades asociadas a la carencia de estos servicios (Cabezas, 2018).

La ciudad de Lima Metropolitana utiliza principalmente el agua proveniente de tres ríos: Rímac, Chillón y Lurín, los que conforman la cuenca de Chillón-Rímac-Lurín. La ciudad capital alberga aproximadamente a 10 millones de personas y en ella se genera por lo menos

el 50 por ciento del PBI nacional, básicamente en los sectores de la industria y los servicios. La columna vertebral y fuente más importante de agua para la ciudad en la cuenca es el río Rímac, que comprende el 69 por ciento de la oferta de agua superficial y tiene el mayor caudal y un régimen de relativa regulación por parte de las empresas SEDAPAL y EDEGEL3, y su sistema de lagunas e interconexiones con la cuenca del Mantaro en la parte alta (AQUAFONDO *et al.*, 2016).

En la Tabla 1 se muestra, la disponibilidad de agua para Lima Metropolitana es menor en comparación con otras ciudades capitales de América Latina.

**Tabla 1: Población y la disponibilidad de agua para Lima Metropolitana en comparación con otras ciudades capitales de América Latina**

Ciudad	Población (millones)	Precipitación (mm/año)	Capacidad de producción (m <sup>3</sup> /s)	Reserva (millones m <sup>3</sup> )	Reserva per capita (millones m <sup>3</sup> )	Dotación (lit/hab/año)	Agua no facturada (%)	Precio del Agua (US\$/m <sup>3</sup> )
Sao Paulo	20	1,500	90	2,073	105	222	38	1.03
México	22	1,158	31	984	44	347	40	4.23 -9.01
Buenos Aires	14	1,146	19	584	43	613	37	0.50
<b>Lima</b>	<b>9</b>	<b>9</b>	<b>20</b>	<b>330</b>	<b>37</b>	<b>248</b>	<b>29</b>	<b>0.74 - 2.22</b>
Bogotá	7	800	25	800	117	181	35	1.67
Santiago	5	384	24	900	166	302	29	1.63

FUENTE: AQUAFONDO *et al.* (2016)

### 2.2.2. Relación del saneamiento con el COVID-19

Saravia Matus *et al.* (2020) en su informe mencionan que, América Latina y el Caribe enfrentan múltiples desafíos relacionados con la gestión del agua. Estos van desde abordar las desigualdades en el acceso y calidad del agua potable y saneamiento, mejorar la eficiencia y productividad en el uso del agua, reducir la contaminación y sobreexplotación del recurso hídrico, hasta preservar las condiciones ambientales de ríos y lagos. Así mismo Corburn *et al.* (2020) enfatizan que los asentamientos informales del hemisferio sur, se encuentran menos preparados para afrontar la pandemia de COVID-19 ya que las necesidades básicas como servicio de agua, saneamiento, drenaje, recolección de residuos y vivienda segura y adecuada ya son escasas o inexistentes. Sumado a ello están las limitaciones de espacio, la violencia y el hacinamiento en los barrios marginales que hacen que el distanciamiento físico y medidas como el confinamiento no sean prácticas y adecuadas lo que resulta en una propagación rápida de la infección.

De la misma forma Castro y Perez (2009) señalan que la carencia de agua potable en cantidad, calidad, continuidad y a un costo adecuado, así como la disposición inadecuada de excretas y de residuos sólidos, además de perjudicar la calidad de vida y las condiciones de producción afectan la integridad de las cuencas hidrográficas en general y de las fuentes de agua en particular. Dicho deterioro, unido a la carencia de agua potable crean ambientes insalubres que propician las enfermedades y disminuyen la productividad de la población.

Teniendo en cuenta lo que menciona Gwenzi (2021), existirían tres posibles factores que expliquen la transmisión del COVID-19 por medio de la ruta fecal-oral, estos serían por el agua contaminada con escaso o nulo tratamiento, el consumo de alimentos marinos crudos y de fuentes de agua contaminada y el sistema de reúso de aguas para la agricultura en alimentos que se consumen sin necesidad de cocción. También plantea la posibilidad significativa de exposición humana al SARS-CoV-2 a través de bioaerosoles y aerosoles de aguas residuales. A ello agrega que en los países en desarrollo esta transmisión podría ser alta debido a la deficiente infraestructura para la potabilización del agua, las pésimas prácticas de higiene y seguridad con los alimentos y sumado a ello el colapso del sistema de salud, por lo que plantea que factores como un adecuado suministro de agua, disposición de excretas y en general un sistema de saneamiento adecuado serían los principales factores a atacar en la lucha contra este virus.

### 2.3. MODELO GLOBALES

Tal como lo definen Bacchini *et al.* (2010), el modelo de regresión múltiple viene a ser una extensión de la regresión lineal simple pues toma en cuenta más de una variable independiente. Esto es, en lugar de usar sólo una variable independiente  $x$  para explicar la variación en  $y$ , es posible utilizar de manera simultánea varias variables independientes.

Este modelo sigue la siguiente forma:

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_px_p + \epsilon$$

En el modelo de regresión múltiple:  $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ , son los parámetros y el término de error  $\epsilon$  es una variable aleatoria. Examinando con cuidado este modelo vemos que  $y$  es una función lineal de  $x_1, x_2, \dots, x_p$  (la parte de  $\beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_px_p$ ) más el término

del error  $\epsilon$ . Este último término se refiere a la variabilidad en  $y$  que no puede ser explicada por el efecto lineal de las  $p$  variables independientes (Anderson *et al.*, 2012).

### 2.3.1. Mínimos cuadrados ordinarios (MCO)

Como mencionan Gujarati y Porter (2010), el método de mínimos cuadrados ordinarios se le atribuye al matemático de origen alemán Carl Friedrich Gauss, además que este método presenta propiedades estadísticas muy atractivas que lo han convertido en uno de los más eficaces y populares del análisis de regresión.

En Anderson *et al.* (2012), manifiestan que el método de mínimos cuadrados (MCO) es un procedimiento en el que se usan los datos muestrales para encontrar la ecuación de regresión estimada. Es así que para una ecuación de regresión simple estimada como la siguiente:

$$\hat{y}_i = b_0 + b_1 x_i$$

Donde:

$\hat{y}_i$  = valor estimado de la variable dependiente en la observación  $i$ ésima

$b_0$  = intersección de la recta de regresión estimada en el eje  $y$

$b_1$  = pendiente de la recta de regresión estimada

$x_i$  = variable independiente en la observación  $i$ ésima

En este método se usan los datos muestrales para obtener los valores de  $b_0$  y  $b_1$  que minimicen la suma de los cuadrados de las desviaciones (diferencias) entre los valores observados de la variable dependiente  $y_i$  y los valores estimados de la variable dependiente  $\hat{y}_i$ . El criterio que se emplea en el método de mínimos cuadrados se basa en la expresión:

$$\min \sum (y_i - \hat{y}_i)^2$$

Donde:

$y_i$  = valor observado de la variable dependiente en la observación  $i$ ésima

$\hat{y}_i$  = valor estimado de la variable dependiente en la observación  $i$ ésima

En el estudio de Mollalo *et al.* (2020) utilizan el MCO, pues es un tipo de modelo de regresión global que examina las relaciones no espaciales entre el conjunto de variables de control y respuesta con el supuesto fundamental de homogeneidad y no variabilidad espacial.



$$y_i = \beta_0 + x_i\beta + \varepsilon_1$$

Donde  $i$  y  $y_i$  son los parámetros de incidencia de COVID-19,  $\beta_0$  es la intersección,  $x_i$  es el vector de variables seleccionadas,  $\beta$  es el vector de coeficientes de regresión y  $\varepsilon_i$  es un error aleatorio. La función fundamental de MCO es optimizar los coeficientes de regresión ( $\beta$ ) reduciendo la suma de los errores de predicción al cuadrado (Sannigrahi *et al.*, 2020).

También este modelo exige el cumplimiento de supuestos como lo dan a conocer Gujarati & Porter (2010) tales como el modelo debe ser lineal en los parámetros, el número de observaciones debe ser mayor que el número de parámetros que se desea estimar, no debe haber colinealidad exacta entre las variables X, el valor medio de la perturbación es igual a cero, la homoscedasticidad del término del error y la no autocorrelación de los mismos.

### 2.3.2. Modelos considerando dependencia espacial

Los modelos de dependencia espacial, también denominados modelos espaciales dinámicos, podrían definirse como aquellos modelos de regresión lineal que consideran explícitamente la existencia del efecto espacial de dependencia o autocorrelación (Yrigoyen, 2003).

En Herrera *et al.* (2010) se expresa que la matriz de pesos espaciales, comúnmente identificada por  $W$ , es un elemento muy característico en modelos econométricos que utilizan datos georeferenciados. Esta matriz permite introducir dependencia entre las unidades. Suponiendo que el tamaño muestral es igual a  $N$ , la matriz  $W$  de orden  $N \times N$ , puede representarse como se observa en la Figura 1:

$$W = \begin{bmatrix} 0 & w_{1,2} & \cdots & w_{1,j} & \cdots & w_{1,N} \\ w_{2,1} & 0 & \cdots & w_{2,j} & \cdots & w_{2,N} \\ \vdots & \vdots & \ddots & \cdots & \cdots & \cdots \\ w_{i,1} & w_{i,2} & \vdots & 0 & \cdots & w_{i,N} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \cdots \\ w_{N,1} & w_{N,2} & \vdots & w_{N,j} & \vdots & 0 \end{bmatrix}$$

**Figura 1: Matriz de pesos espaciales**

FUENTE: Herrera *et al.* (2010)

Donde las filas y columnas identifican a las observaciones de corte transversal, siendo N el tamaño del conjunto de datos y  $w_{i,j}$  ( $i,j= 1,2,\dots,N$ ) las ponderaciones o pesos que aproximan la relación entre dos unidades  $i$ (filas) y  $j$ (columnas). La diagonal principal está formada por ceros, estableciendo que ninguna observación puede estar relacionada consigo misma, es decir, no ser vecinos de ellos mismos (Herrera *et al.*, 2010).

**a. Modelo de retraso espacial (SLM)**

Así como se menciona en Anselin (2007), en el modelo de regresión lineal estándar, la dependencia espacial se puede incorporar de dos formas distintas: como un regresor adicional en la forma de una variable dependiente espacialmente rezagada ( $Wy$ ), o en la estructura de error ( $E [\varepsilon_i \varepsilon_j] \neq 0$ ). El primero se denomina modelo de rezago espacial y es apropiado cuando el foco de interés es la evaluación de la existencia y la fuerza de la interacción espacial.

Esto se interpreta como una dependencia espacial sustantiva en el sentido de estar directamente relacionado con un modelo espacial.

Formalmente, un modelo de rezago espacial, o un modelo mixto, regresivo y autorregresivo espacial, se expresa como:

$$y = \rho Wy + X\beta + \varepsilon$$

donde  $\rho$  es un coeficiente autorregresivo espacial,  $\varepsilon$  es un vector de términos de error, y la otra notación es como en el modelo de mínimos cuadrados ordinarios.

**b. Modelo de error espacial (SEM)**

En Fischer y Wang (2011), se menciona que otra forma de dependencia espacial ocurre cuando la dependencia funciona a través del proceso de error, en el que los errores de diferentes áreas pueden mostrar covarianza espacial.

Según Yrigoyen (2003), el modelo del error espacial es un caso particular de modelos con perturbación aleatoria no esférica, que no satisface los supuestos de homocedasticidad (varianza constante) y no autocorrelación. La dependencia espacial en la perturbación aleatoria puede adoptar formas diversas. La especificación de este modelo es el MBRL (modelo básico de regresión lineal

múltiple) con perturbación autorregresiva espacial:

$$y = X\beta + u$$
$$u = \lambda Wu + \varepsilon$$

Donde:

X: matriz de observaciones de las variables explicativas (N, K)

$\beta$ : vector de coeficientes de la regresión (K,1)

u: vector de términos de la perturbación aleatoria (N,1)

$\lambda$ : coeficiente autorregresivo

Wu: retardo espacial de la perturbación aleatoria

$\varepsilon$ : perturbación aleatoria esférica, con media nula y matriz de varianzas y covarianzas  $\sigma^2 I$  (homoscedasticidad y no autocorrelación).

### **III. METODOLOGÍA**

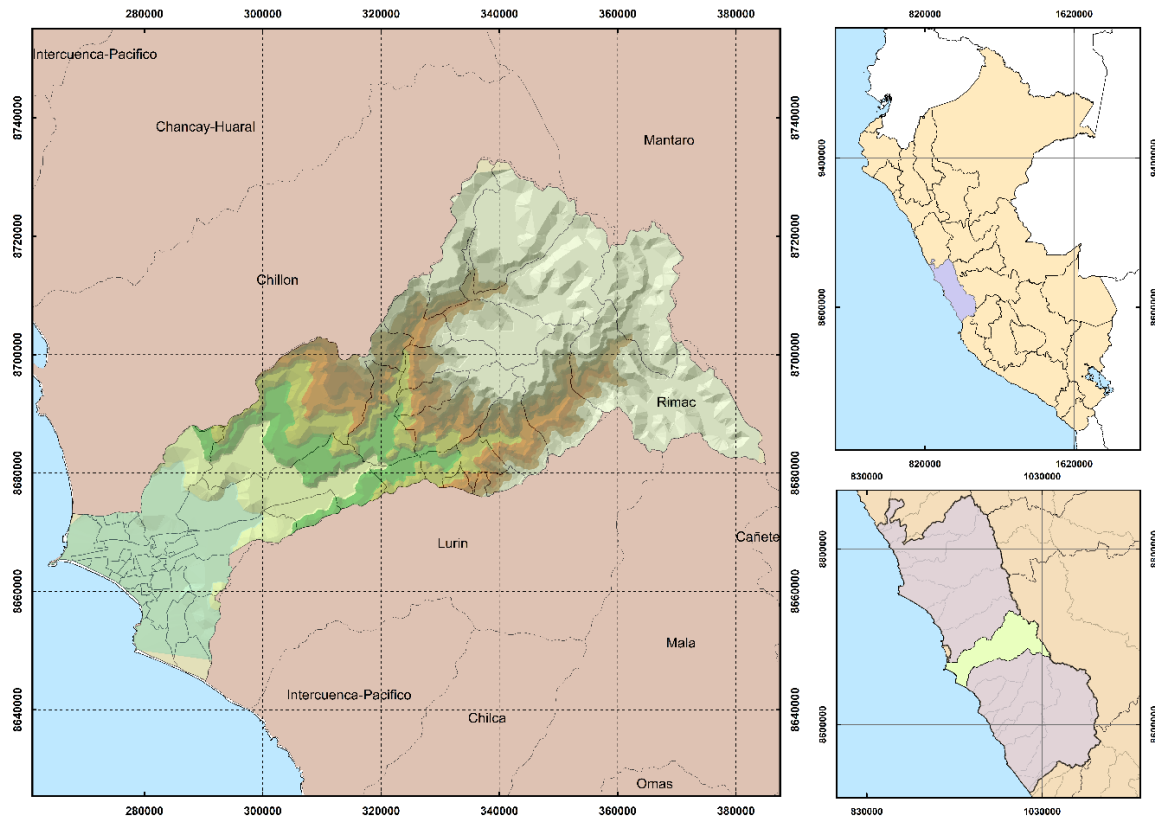
#### **3.1. ZONA DE ESTUDIO**

La cuenca del río Rímac se encuentra inscrita entre los paralelos 11°30' y 12°15' de latitud sur, 76° y 77° de longitud oeste. Su área es 3 240,6 km<sup>2</sup>. Se ubica políticamente en la jurisdicción del departamento de Lima y en menor proporción en el departamento de Junín, enmarcándose en las provincias de Lima, Huarochirí y Yauli, frente a la costa del Perú (Observatorio del Agua Chillón Rímac Lurín, 2019).

El río Rímac tiene su inicio a una altura de 3 400 m.s.n.m. por la confluencia del río Blanco con el Yauliyaco. Este último nace en Ticlio, ubicado en el distrito de Chicla provincia de Huarochirí a 4 820 m.s.n.m. de altura, y el río Blanco es alimentado por el nevado Tatajaico a 5000 m.s.n.m. de altura (Tumialán De la Cruz, 2016).

Dentro de las cuencas de la costa peruana, la del río Rímac abastece de agua a la mayoría de actividades que se realizan en la zona. Además del uso principal para agua potable y de generación de energía (se ubican 5 de las centrales hidroeléctricas más importantes del país), sus aguas satisfacen la demanda de uso agrícola de las áreas cultivables del valle (Fundación Futuro Latinoamericano (FFLA), 2015), además posee 11 ríos, los cuales confluyen en las dos subcuencas principales para formar el río Rímac (Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI), 2014).

La zona de estudio abarca la cuenca del río Rímac, la misma que comprende 54 distritos, los cuales se distribuyen en 3 provincias., la provincia Constitucional del Callao tiene 5, la provincia de Lima posee 30 y por último la provincia de Huarochirí alberga a 19 distritos. En la Figura 2 se aprecia la cuenca del río Rímac que abarca los distritos de estudio.



**Figura 2: Ubicación de la cuenca del río Rímac**

Para el estudio se llevó a cabo la selección de los distritos dentro de la cuenca; se excluyeron aquellos con un porcentaje de territorio total por debajo del 5 por ciento y no se tomaron en cuenta, quedando así 54 de los 71 distritos iniciales.

Para la selección de variables, se consultó la bibliografía existente para recopilar las que abarcan el sector saneamiento, tales como el acceso a agua potable y al servicio de alcantarillado, puesto que la relación entre estas y el COVID-19 se enmarca en satisfacer un suministro oportuno y correcto del recurso hídrico además de poder llevar a cabo prácticas de higiene como el lavado de manos para prevenir la propagación de enfermedades (Neal, 2020).

En tal sentido, Ekumah *et al.* (2020) realizaron el estudio mediante la regresión logística ordenada, encontrando la relación entre el acceso desigual al agua y saneamiento con el riesgo de aumentar la propagación del COVID-19 en África subsahariana. Además Silva *et al.* (2021) en su investigación, encontraron que la mayoría de casos de COVID-19 y

fallecidos por este virus expresados como la tasa de incidencia y tasa de mortalidad respectivamente, se distribuyeron entre los municipios con las tasas más bajas de acceso a agua potable o un sistema de alcantarillado. Por consiguiente, el incluir las variables en relación al acceso al agua potable y saneamiento es importante por las razones antes expuestas.

Además, dentro de las variables de saneamiento se tomaron en cuenta las que pertenecen al grupo de residuos sólidos, puesto que la contaminación de las superficies debido a coronavirus de origen humano puede permanecer hasta por 9 días a temperatura ambiente y ser una fuente de transmisión viral (Kampf *et al.*, 2020). Adicionalmente en los países en vías de desarrollo los residuos sólidos generalmente son vertidos a botaderos que no cuentan con los debidos protocolos de seguridad que una pandemia amerita y están al libre acceso de personas que realizan actividades de reciclaje sin equipos de protección (Nzediegwu y Chang, 2020).

Si bien el presente trabajo se enfocó en las variables referidas al saneamiento, se consideró recomendable la inclusión de dos grupos más de variables debido a su importancia; las meteorológicas y las referidas al aspecto sociodemográfico en virtud de complementan el estudio, por lo mencionado a continuación:

Las variables meteorológicas se seleccionaron a causa de la relación encontrada en estudios anteriores como es el caso de Paraskevis *et al.* (2021), que realizaron una revisión bibliográfica entre 59 estudios a nivel mundial encontrando que el impacto de las variables meteorológicas con la propagación del COVID-19 parece probable aunque se necesitan más investigaciones.

En lo que respecta a la selección de las variables sociodemográficas como población, densidad, personas con más de 65 años, porcentaje de mujeres y hombres etc., se han considerado por causa de que estudios anteriores avalan su uso, tal es el caso chileno analizado por Rivera-Córdova (2021), se concluye que existe una asociación entre el número de casos confirmados de coronavirus con una mayor densidad poblacional o el que plantea Jabłońska *et al.* (2021), en la cual la población urbana está relacionada con el alto índice de

tasa de mortalidad por COVID-19 en países europeos.

El presente estudio se realizó en función a información secundaria disponible en los portales nacionales. Al respecto, la información de las variables de acceso al agua potable, saneamiento y sociodemográficas fueron obtenidas de la base de datos Censos Nacionales 2017: XII de Población, VII Vivienda y III de Comunidades Indígenas, del Instituto Nacional de Estadística e Informática (INEI) realizado el año 2017, las cuales se observan en la Tabla 2. En relación a las variables relacionadas a la generación de residuos sólidos se consiguieron del portal “Seguimiento e información sobre gestión de residuos sólidos” del Ministerio del Ambiente. Los datos de número de casos positivos, así como las defunciones atribuidas al COVID-19 se obtuvieron de la plataforma: “Datos Abiertos Minsa y Gestión del Conocimiento en COVID-19” del Ministerio de Salud. Se seleccionaron las variables meteorológicas de la base de datos del Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI), tales como temperatura, humedad relativa, velocidad del viento y precipitación.

En la Tabla 2 se detalla la descripción de las variables que se utilizaron en el estudio.

**Tabla 2: Variables utilizadas en la obtención del modelo**

Tipo de variable	Descripción
Abastecimiento de agua potable	Procedencia del agua por red pública
	Procedencia del agua por camión cisterna u otro similar, pozo o manantial, puquio, río, acequia, lago, laguna y otros
	Red pública dentro de la vivienda
	Red pública está fuera de la vivienda, pero dentro de la edificación
	Red pública es pilón o pileta de uso público
	Disponen de agua por red pública menos de 7 horas
	Disponen de agua por red pública entre 7 y 18 horas
	Disponen de agua por red pública más de 18 horas
Entidad que suministra el servicio de agua potable	Pagan por el servicio de agua a una empresa prestadora de servicios
	Pagan por el servicio de agua a la municipalidad y organización comunal
	Pagan por el servicio de agua a camión cisterna u otros particulares

«Continuación»

Condición de pago por el servicio de agua	Sí pagan por el servicio de agua No pagan por el servicio de agua
Servicio de saneamiento	Servicio higiénico conectado a red pública de desagüe dentro de la vivienda Servicio higiénico conectado con red pública de desagüe fuera de la vivienda, pero dentro de la edificación Servicio higiénico conectado a letrina, pozo ciego, río, acequia, canal o similar u otro
Residuos sólidos	Generación domiciliaria Generación municipal Composición de los residuos orgánica Composición de los residuos no aprovechable Composición de los residuos peligrosa Composición de los residuos inorgánica Cantidad de puntos críticos
Población	Población total Población de mujeres Población de hombres Densidad de población Cantidad de mercados (espacio público) Población económicamente activa Población con algún sistema de salud Población menor de 14 años Población mayor de 65 años Población que no lee menor de 15 años Población que se identifica quechua Población urbana Población rural Altura sobre el nivel del mar
Meteorológicas	Temperatura máxima Temperatura promedio Temperatura mínima Humedad relativa máxima Humedad relativa promedio Humedad relativa mínimo



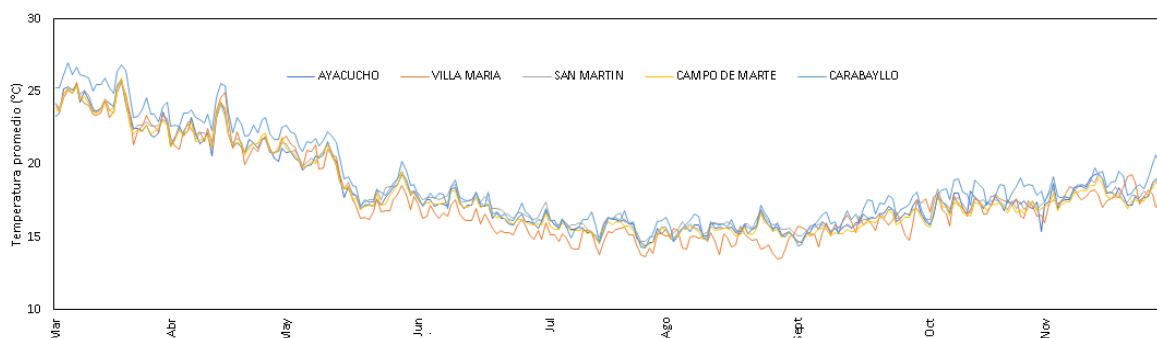
## 3.2. CARACTERIZACIÓN DE INFORMACIÓN

### 3.2.1. Servicios de saneamiento y otras variables

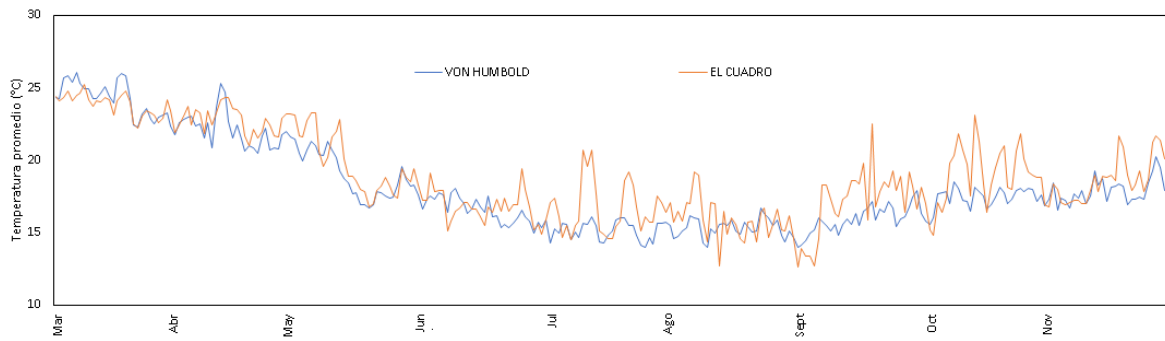
Se realizó el reconocimiento de las variables de saneamiento como el acceso a agua potable y el acceso al servicio de alcantarillado, disponibilidad del servicio en cuanto a cantidad de horas y frecuencia, empresa o entidad encargada de suministrar, etc. Además de datos meteorológicos y las variables sociodemográficas a nivel de distrito. Este reconocimiento permitió un diagnóstico de la situación actual de la propagación de la pandemia dentro la cuenca, así como la identificación de factores que puedan o no estar influyendo en ello.

Para las variables meteorológicas primero se efectuó un análisis visual que permitía detectar e identificar la inconsistencia de la información tanto de la variable temperatura, humedad relativa y velocidad del viento para así poder encontrar el período o los períodos en los cuales los datos fueron dudosos o faltantes, con este método fue posible también observar “picos” que son valores muy altos o muy bajos, “saltos” y/o “tendencias”, los cuales se comprobaron si son fenómenos naturales que efectivamente han ocurrido o fueron producidos por errores sistemáticos.

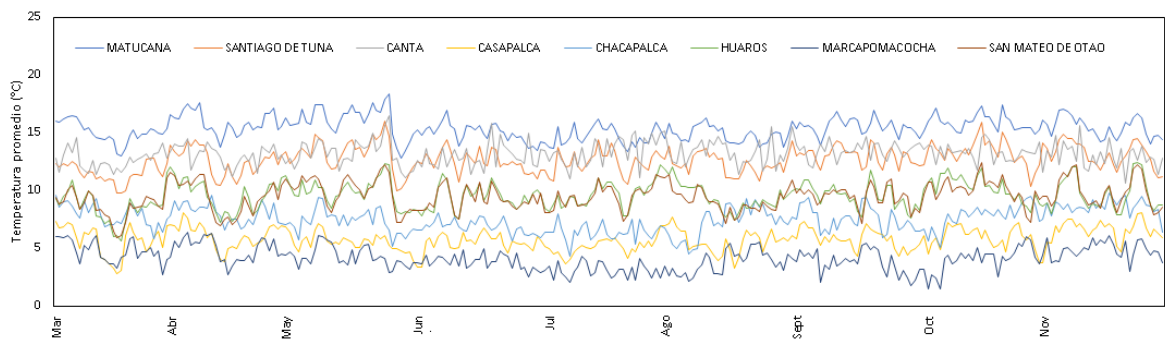
El periodo comprende desde el 6 de marzo al 30 de noviembre del 2020, para lo que se dividió en tres grupos: Cuenca baja (Ayacucho, Villa María, San Martín Campo de Marte y Carabayllo), cuenca media (Von Humboldt y El Cuadro) y cuenca alta (Matucana, Santiago de Tuna, Canta, San Mateo de Otao, Huaros, Marcapomacocha, Casapalca y Chacapalca) (ver Figura 3, Figura 4 y Figura 5).



**Figura 3: Variación diaria de la temperatura promedio según estaciones en la cuenca baja**



**Figura 4: Variación diaria de la temperatura promedio según estaciones en la cuenca media**



**Figura 5: Variación diaria de la temperatura promedio según estaciones en la cuenca alta**

Se trabajó con las siguientes variables meteorológicas: temperatura, humedad relativa y velocidad del viento. Esta información se obtuvo de las estaciones meteorológicas identificadas dentro y fuera de la cuenca del río Rímac, las mismas que se aprecian en la Tabla 3.

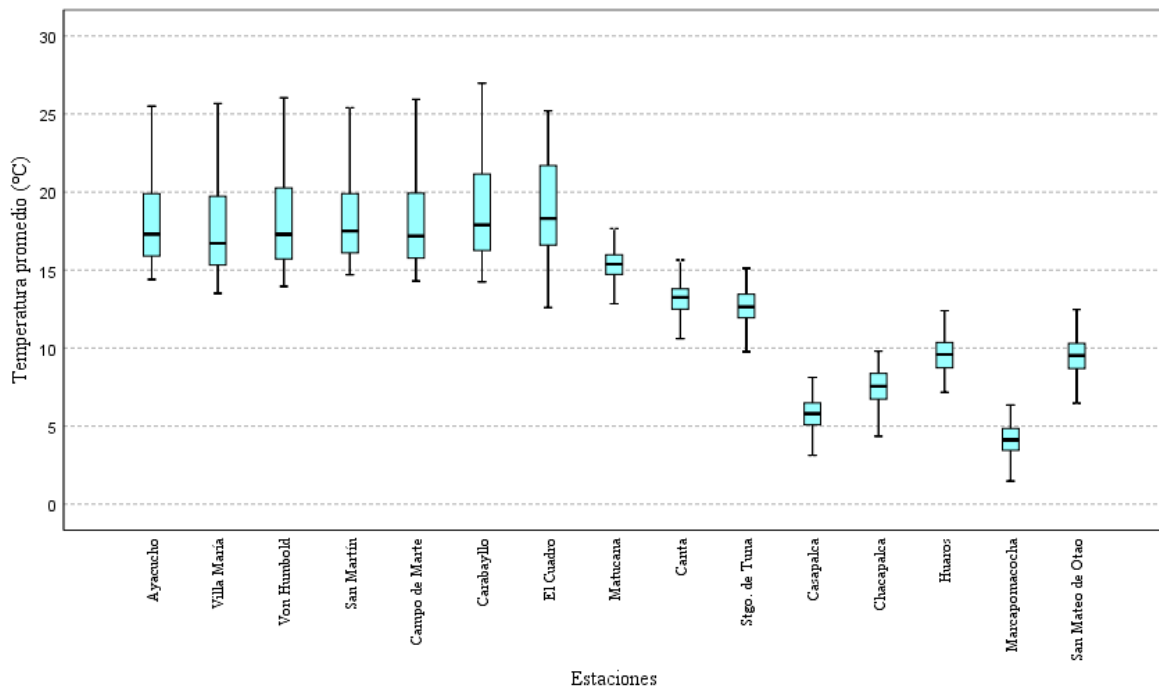
**Tabla 3: Estaciones meteorológicas a utilizarse para la interpolación**

N	NOMBRE	ESTE	NORTE
1	Ayacucho	283 434,80	8 658 280,64
2	Campo de marte	277 594,00	8 664 813,00
3	Canta	322 668,88	8 731 416,98
4	Carabayllo	278 497 ,59	8 683 450,49
5	Casapalca	365 554,40	8 713 177,02
6	Chacapalca	417 652,85	8 702 882,10
7	El Cuadro	304 016,00	8 673 912,00
8	Huaros	329 066,44	8 739 329,79
9	Marcapomacocha	355 431,76	8 738 970,98
10	Matucana	349 898,00	8 690 864,00

«Continuación»

11	San Martín de Porres	272 463,30	8 669 263,00
12	San Mateo de Otao	329 674,00	8 689 892,00
13	Santiago de Tuna	334 054,00	8 674 855,00
14	Villa María	291 083,00	8 654 309,00
15	Von Humboldt	288 915,00	8 663 607,00

La información faltante o con errores se rellenó con el método de completación por regresiones múltiples. Para la interpolación de las variables se utilizó el software ArcGIS Desktop 10.8. Además, se hizo el análisis exploratorio de los datos para encontrar el método con mejor ajuste, encontrándose que el método *IDW* (Interpolación con la distancia inversa ponderada) para las variables humedad y velocidad del viento mientras que se usó el método Kriging bayesiano empírico para la variable temperatura lo que permitió encontrar los valores de cada variable meteorológicas dentro de los distritos de estudios en la cuenca.



**Figura 6: Diagrama de cajas para la temperatura promedio en las estaciones de la cuenca Rímac**

### 3.2.2. Casos positivos y tasa de incidencia de COVID

Se realizó el reconocimiento de las variables como número de casos positivos, así como la tasa de incidencia según distritos. Se obtuvieron de la plataforma: “Datos Abiertos Minsa y

Gestión del Conocimiento en COVID-19” del Ministerio de Salud según distrito y provincia, por lo que se procesó la información y se extrajo de manera mensual; además de obtener la pirámide poblacional para el total de casos por provincia. También se calcularon los valores de tasa de incidencia por distrito de manera mensual, esto se logró sumando el número de casos positivos, dividiendo este valor entre la población del distrito y multiplicado por 100 mil según bibliografía.

### **3.3. MODELOS GLOBALES PARA AJUSTAR LA TASA DE INCIDENCIA**

#### **3.3.1. Correlación no paramétrica de la tasa de incidencia y servicios de saneamiento**

Se realizaron las correlaciones de las variables asociadas al saneamiento con los datos asociados al COVID-19, utilizando la estadística paramétrica y no paramétrica. Acorde a la estadística paramétrica, tal como se menciona en Santabárbara (2019), para una interpretación adecuada de la relación entre dos variables, lo ideal sería disponer de una medida de la relación entre ellas que no esté influenciada por las unidades en que vienen medidas. Una forma de conseguir este objetivo es dividir la covarianza (medida de la variabilidad conjunta) por el producto de las desviaciones típicas de las variables, ya que así se obtiene un coeficiente adimensional denominado coeficiente de correlación lineal de Pearson que se simboliza con  $r$  y cuya expresión matemática es:

$$r = \frac{S_{xy}}{S_x \times S_y}$$

Donde  $S_{xy}$  es la covarianza entre X y Y,  $S_x$  es la desviación típica de X, y  $S_y$  la de Y.

Con relación a la estadística no paramétrica se hizo uso del coeficiente de correlación de Spearman. El cual es una medida de asociación lineal que utiliza los rangos, números de orden, de cada grupo de sujetos y compara dichos rangos. Este coeficiente es muy útil cuando el número de pares de sujetos ( $n$ ) que se desea asociar es pequeño (menor de 30). Aparte de permitir conocer el grado de asociación entre ambas variables, con Rho de Spearman es posible determinar la dependencia o independencia de dos variables aleatorias (Mondragon, 2000).

La fórmula de este coeficiente es la siguiente:

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

Siendo: n, la cantidad de sujetos que se clasifican;  $x_i$ , el rango de sujetos u con respecto a una variable,  $y_i$ , el rango de sujetos i con respecto a una segunda variable y  $d_i$  la diferencia de  $x_i - y_i$

Otra prueba no paramétrica fue el cálculo del coeficiente Tau ( $\tau$ ) de Kendall. El coeficiente tau ( $\tau$ ) de Kendall está basado más en los intervalos jerarquizados de las observaciones que en los propios datos, esto hace que la distribución de  $\tau$  sea independiente de la que presentan las variables X y Y, siempre y cuando los datos representados por estas 2 variables sean independientes y continuas. Este coeficiente es más utilizado por algunos investigadores que el de Spearman, pero es más difícil de calcular, aunque presenta la ventaja que el  $\tau$  tiende más rápido a la distribución normal que el de Spearman, especialmente, en el caso de la certeza de  $H_0$  (Guillen *et al.*, 2014).

Las pruebas estadísticas se desarrollaron con el software IBM SPSS Statistics 25, en el cual se pueden efectuar las correlaciones tanto paramétricas como no paramétricas, así como encontrar los estadísticos descriptivos tales como la media, la moda, el valor máximo y el mínimo, la desviación estándar, la varianza, la curtosis y la asimetría.

### **3.3.2. Ajuste de modelos globales**

Los datos obtenidos se recopilaron en una base de datos dentro del entorno GIS para vincular las variables explicativas con el *shapefile* de los límites administrativos de los distritos de la cuenca de estudio.

Para la calibración del modelo global de mínimos cuadrados ordinarios se utilizó el software ArcGIS Desktop 10.8 mediante el apartado de la opción “Herramientas de estadísticas espaciales” (Spatial statistics tools) y el ítem “Modelado de relaciones espaciales” (Modeling spatial relationships) con la opción Ordinary Least Squares.

Para la obtención de los modelos globales que tienen en cuenta la autocorrelación espacial se hizo uso del software de código abierto: GeoDa. Anselin *et al.* (2006) mencionan que el

diseño de GeoDa consiste en un entorno interactivo que combina mapas con gráficos estadísticos, utilizando la tecnología de ventanas vinculadas dinámicamente. Está orientado al análisis de datos discretos geoespaciales, es decir, objetos caracterizados por su ubicación en el espacio como puntos (coordenadas de puntos) o polígonos (coordenadas de límites de polígonos). En términos generales, la funcionalidad se puede clasificar en seis categorías: Manipulación de datos espaciales y utilidades: entrada, salida y conversión de datos; transformación de datos: transformaciones de variables y creación de nuevas variables; mapeo: mapas de coropletas, cartograma y animación de mapas gráficos estadísticos; autocorrelación espacial: estadísticas de autocorrelación espacial global y local, con inferencia y visualización y regresión espacial: diagnóstico y estimación de máxima verosimilitud de modelos de regresión espacial lineal.

Mediante el uso del software GeoDa se procedió a realizar la calibración de los modelos globales como el de error espacial y de retraso espacial, para lo cual se considerará:

- Variables sanitarias (abastecimiento de agua, servicio de saneamiento y generación de residuos)
- Variables meteorológicas
- Variables sociodemográficas
- Tasa de incidencia del COVID-19

Si bien la tasa de incidencia y las variables meteorológicas varían en tiempo y espacio según la información brindada por el Ministerio de Salud y el SENAMHI, mientras que las variables relacionadas al saneamiento, conservan el mismo valor a lo largo del periodo de estudio de acuerdo a los datos obtenidos del Censo Nacional 2017: XII de Población, VII Vivienda y III de Comunidades Indígenas, del Instituto Nacional de Estadística e Informática (INEI) realizado el año 2017; sin embargo esto permitió evaluar si estas últimas se asocian positiva o negativamente con la propagación del COVID-19 con el mismo nivel de importancia en cada mes de estudio.

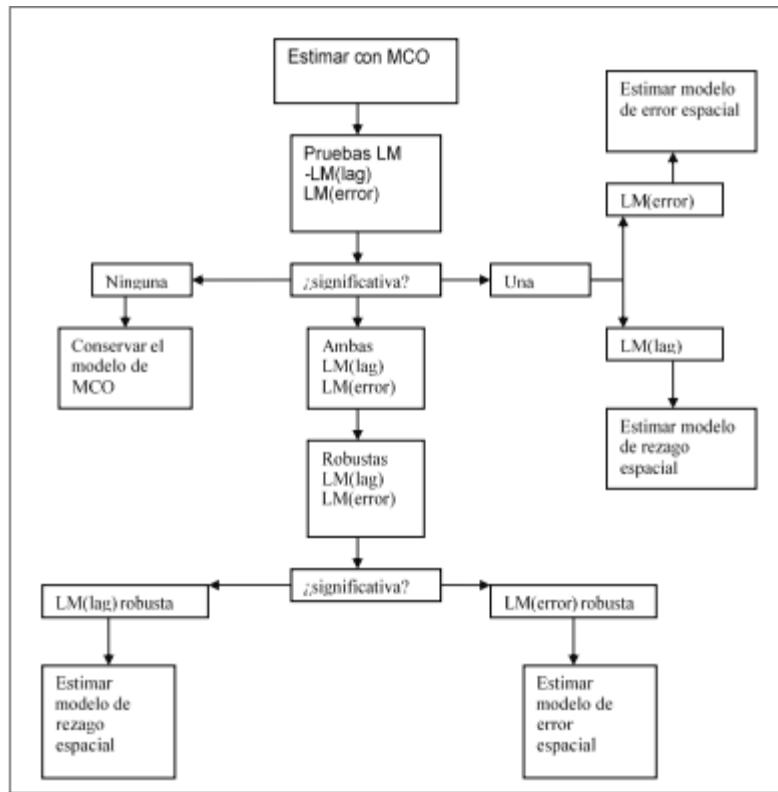
Antes de comenzar el modelado se organizaron las variables y se usó la transformación logarítmica de la variable dependiente. Con la ayuda de la herramienta “Regresión

exploratoria” (Exploratory regression) dentro de la interfaz de ArcGIS Desktop se prueban todas las posibles combinaciones de variables regresoras o independientes que se indican como datos de entrada para explicar mediante un modelo de mínimos cuadrados ordinarios la variable dependiente de acuerdo a las condiciones dadas. Sólo se eligen aquellas que cumplan parcial o totalmente los criterios establecidos para luego modelar de manera independiente cada posible combinación elegida y poder descartar su uso.

Una vez obtenidas las probables combinaciones de variables se inició con el modelo global de mínimos cuadrados ordinarios (MCO) propiamente dicho, el mismo que se construyó con las variables antes mencionadas y respetando los supuestos y condiciones propias del modelo las cuales son:

- Los coeficientes de las variables tengan el signo esperado
- Los coeficientes de las variables sean estadísticamente significativos
- No exista redundancia entre las variables explicativas
- Los residuos del modelo tengan una distribución normal
- Los residuos del modelo no estén espacialmente correlacionados
- Se tenga un fuerte  $R^2$  ajustado

Luego de aplicar la prueba de índice de Morán a los residuos del modelo de mínimos cuadrados ordinarios, de no cumplir el supuesto de que los mismos no estén espacialmente correlacionados se procederá como en la Figura 7 en donde se aprecia la estrategia para la selección de los modelos de retraso o error espacial:



**Figura 7: Estrategia de selección de modelos**

FUENTE: Mendoza (2016)

Para realizar el modelo de retraso espacial o error espacial fue necesario seguir los pasos descritos en la Figura 7, esto significó usar el test del multiplicador de Lagrange tanto en el coeficiente de la matriz de autocorrelación espacial en el caso del modelo de retraso espacial, así como en el coeficiente autorregresivo del error para el caso del modelo de error espacial. Todo ello para encontrar cuál es el mejor modelo que se ajusta a la naturaleza de los datos. Esta obtención de los modelos que consideran la autocorrelación de los datos se realizó con el software GeoDa.

Se seleccionaron los modelos que presentaron los mejores ajustes con aquellas variables que explicaron los casos positivos a nivel de distrito. Para lograr esto se tomó en cuenta el valor del coeficiente de determinación ( $R^2$ ) ajustado, tal como menciona Martínez (2005), quien afirma que  $R^2$  ajustado tiene la propiedad de ser neutral frente a la introducción de variables adicionales.

López (2017) menciona que el coeficiente de determinación es la proporción de la varianza



total de la variable explicada por la regresión. Este coeficiente también llamado  $R^2$ , refleja la bondad del ajuste de un modelo a la variable que pretender explicar, además se sabe que el resultado del coeficiente de determinación oscila entre 0 y 1. Cuanto más cerca de 1 se sitúe su valor, mayor será el ajuste del modelo a la variable que estamos intentando explicar, mientras que el coeficiente de determinación ajustado ( $R^2$  ajustado) es la medida que define el porcentaje explicado por la varianza de la regresión en relación con la varianza de la variable explicada. Es decir, lo mismo que el  $R^2$ , pero con una diferencia: El coeficiente de determinación ajustado penaliza la inclusión de variables.

Los criterios de información comúnmente usados en la literatura son: Criterio de información de Akaike (AIC) y criterio de información Bayesiano (BIC). La selección basada en criterios de información busca determinar la significancia de la inclusión de algunos parámetros en el modelo (una entrada o una neurona), aplicando, a su vez, un factor de penalización a modelos con muchos parámetros, por lo que se puede considerar, beneficia la parsimonia (Sánchez, 2012).

Martínez *et al.* (2009) explican que el criterio de información de Akaike proporciona un método simple y objetivo que selecciona el modelo más adecuado para caracterizar los datos experimentales. Este criterio, que se enmarca en el campo de la teoría de la información, se define como:

$$AIC = -2 \log(\mathcal{L}(\hat{\theta})) + 2k$$

Donde  $\log(\mathcal{L}(\hat{\theta}))$  es el logaritmo de la máxima verosimilitud, que permite determinar los valores de los parámetros libres de un modelo estadístico, y  $K$  es el número de parámetros libres del modelo. Esta expresión proporciona una estimación de la distancia entre el modelo y el mecanismo que realmente genera los datos observados, que es desconocido y en algunos casos imposible de caracterizar.

AIC, penaliza modelos según su desviación de los datos reales, mientras que BIC penaliza modelos según el número de parámetros, por lo cual este último modelo tenderá a elegir modelos con pocos parámetros. Para ambos criterios el mejor modelo será aquel que exhiba el valor más pequeño entre todos los modelos evaluados, siendo este el que muestra el mejor ajuste a los datos (Sánchez, 2012).

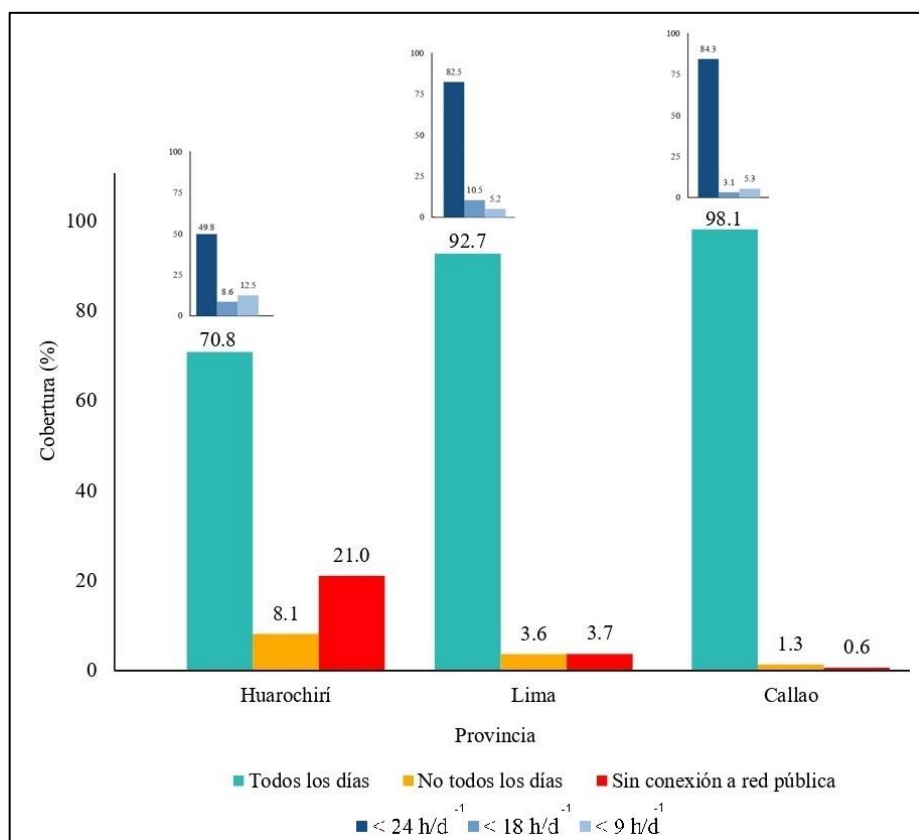
Para la selección del mejor modelo se emplearon los criterios de  $R^2$  ajustado y el criterio de información de Akaike. Los cuales pueden ser empleados para la toma de decisiones en función a disminuir la propagación y el nivel de contagios.

## IV. RESULTADOS Y DISCUSIÓN

### 4.1. CARACTERIZACIÓN DE LOS SERVICIOS DE SANEAMIENTO

#### 4.1.1. Agua potable y alcantarillado

En la Figura 8 se distingue que la cobertura de agua potable por red pública con una disponibilidad diaria es de modo diferente entre la provincia de Huarochirí con un 70,8 por ciento mientras que para el Callao y Lima se tienen más del 90 por ciento. Adicional a ello, el porcentaje de viviendas que no cuentan con el servicio de agua potable todos los días es de un 8 por ciento en promedio en la provincia de Huarochirí mientras que para Callao y Lima este porcentaje disminuye a menos del 4 por ciento.

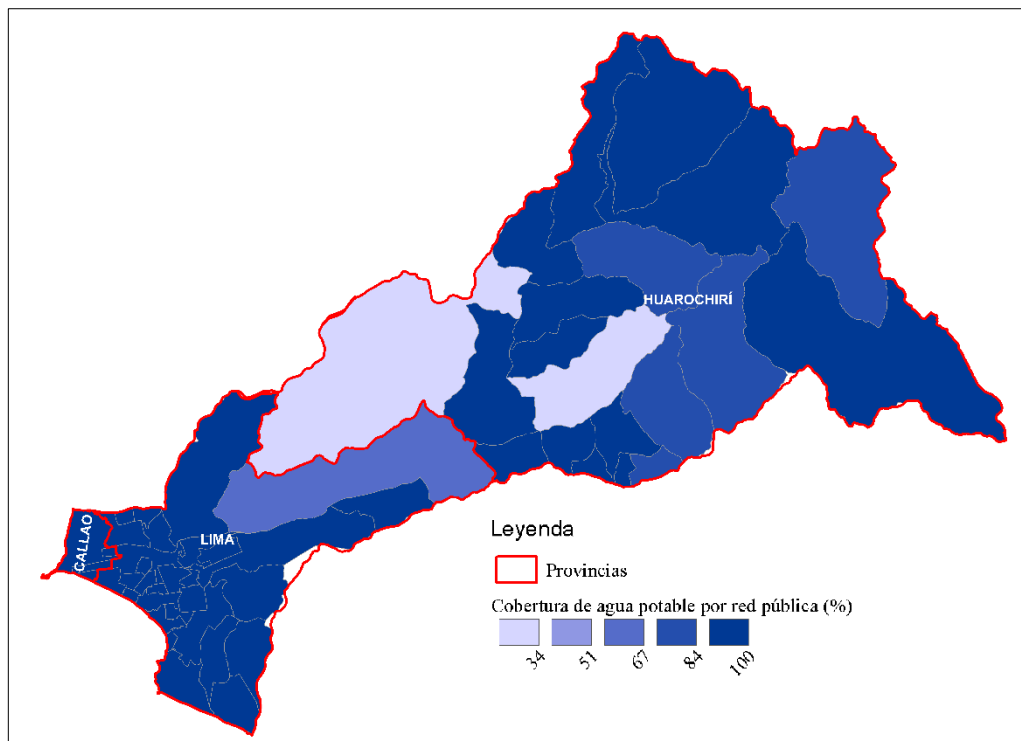


**Figura 8: Cobertura de agua potable según la frecuencia por días de la semana por provincia**

Es posible observar también que la provincia de Huarochirí presenta menos de un 50% de disponibilidad del servicio de agua con una frecuencia entre 18 y 24 horas, mientras que Lima y Callao presentan un 80% en relación a la misma frecuencia de servicio horario.

En la Tabla 16 (Anexo 1) se aprecia con mayor detalle el porcentaje de cobertura por red pública de agua potable (dentro o fuera de la vivienda), en la provincia de Huarochirí presenta en promedio un 79 por ciento en contraste con Lima y Callao que poseen más del 96 por ciento.

En la Figura 9 se aprecia la variación espacial de los porcentajes de cobertura de agua potable por red pública por distritos. Es evidente la presencia de dos distritos con valores por debajo del 34 por ciento y un distrito entre 34 y 51 por ciento, ubicados en la parte media de la cuenca. El resto de distritos con cobertura mayor del 51 por ciento.

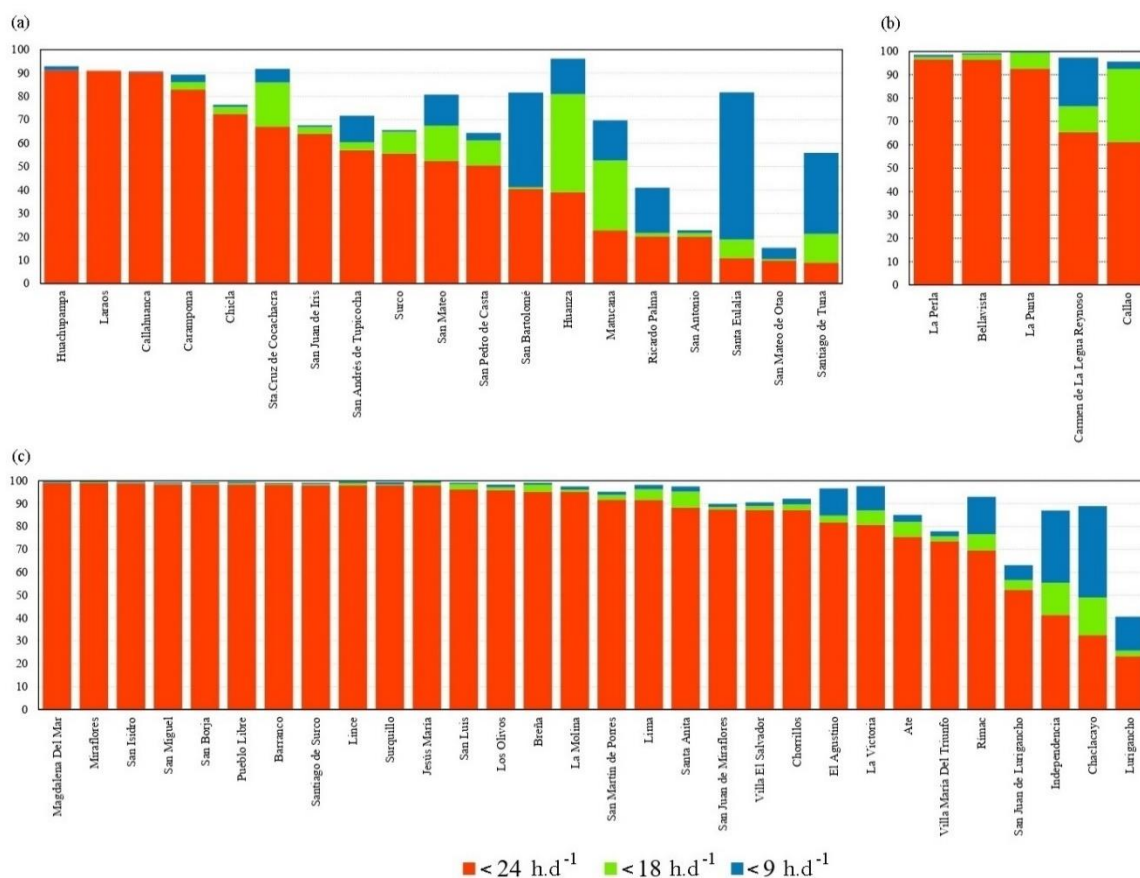


**Figura 9: Variación espacial de la cobertura de agua potable por red pública por distrito**

A nivel de distrito es posible percibir en la Figura 10 que para la provincia de Huarochirí se tiene que más de la mitad de los distritos no alcanzan una cobertura ni del 50 por ciento de

agua potable entre 18 y 24 horas del día.

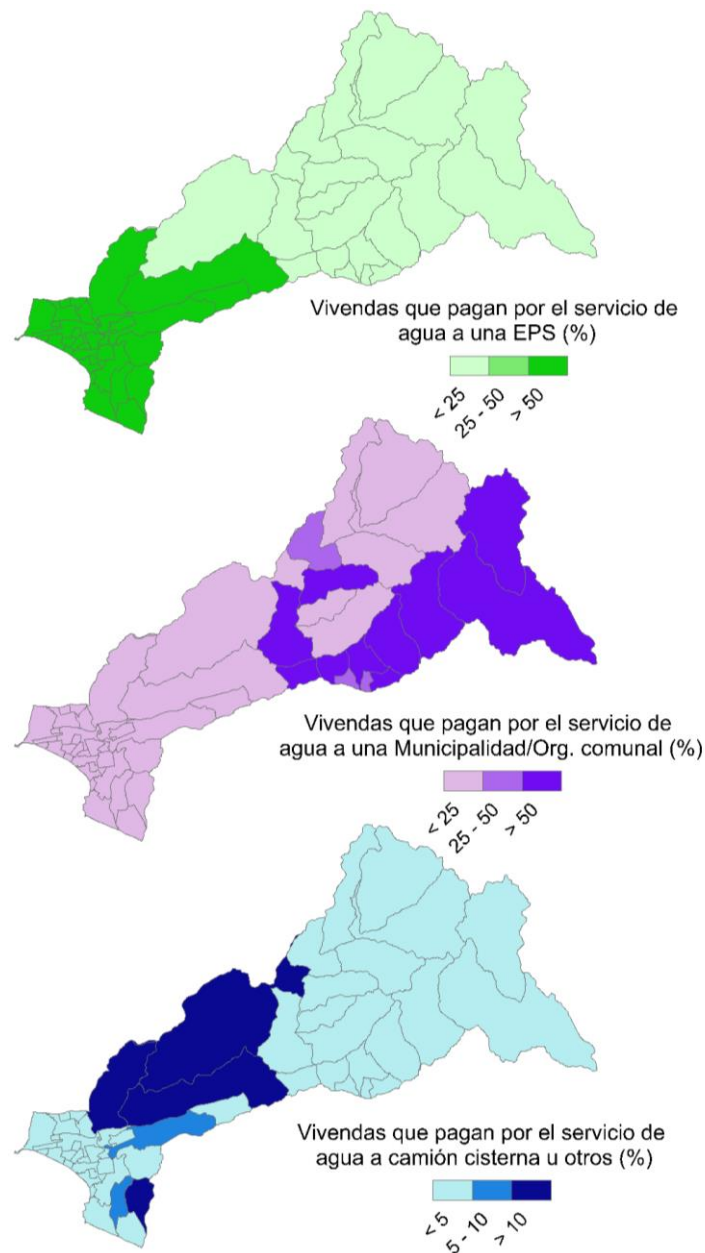
En la provincia del Callao los distritos superan el 60 por ciento de cobertura de agua potable entre 18 y 24 horas al día. Así mismo, en la provincia de Lima sólo 4 distritos no superan el 70 por ciento de cobertura de agua entre 18 y 24 horas.



**Figura 10: Cobertura de agua potable en red pública por distrito en las provincias de Huarochirí (a), Callao (b) y Lima (c)**

Por otra parte, en lo que respecta al tipo de pago por el servicio de agua por una empresa prestadora de servicio (EPS), se tiene que los distritos de la provincia de Huarochirí tienen una cobertura por debajo del 25 por ciento, en contraste con los distritos de las provincias de Lima y Callao que superan el 50 por ciento. Además, los porcentajes de las viviendas que pagan a una Municipalidad u a una Organización comunal no superan el 25 por ciento en la parte baja de la cuenca mientras que en la parte alta más de la mitad de los distritos de la provincia de Huarochirí tienen por encima del 50 por ciento siendo abastecido por una de estas entidades.

Finalmente, en relación al porcentaje de viviendas que pagan por el servicio de agua a un camión cisterna u otro medio se tiene que hay 4 distritos en toda la cuenca con una cobertura mayor al 10 por ciento, ubicados en la parte media que comprende distritos de las provincias de Lima y Huarochirí. Mientras que toda la parte baja y alta de la cuenca tiene un porcentaje inferior al 5 por ciento.

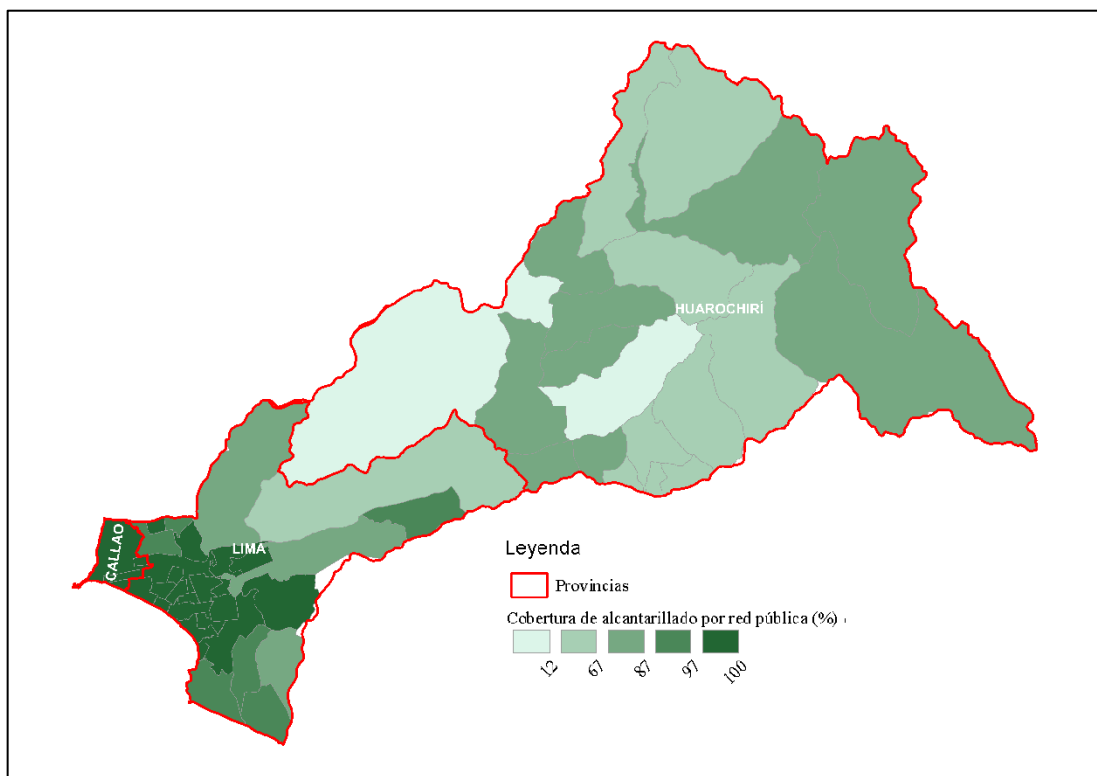


**Figura 11: Distribución espacial del tipo de pago del servicio de agua en la cuenca Rímac**

De la misma manera ocurre en relación a la cobertura por red pública de alcantarillado

(dentro o fuera de la vivienda), puesto que la provincia de Huarochirí presenta en promedio un 62,4 por ciento de cobertura en contraste con Lima y la provincia del Callao que poseen más del 95 por ciento como se visualiza en la Tabla 17 (Anexo 1).

En la Figura 12 se exhibe que los distritos de la provincia de Huarochirí tienen el menor porcentaje de cobertura de alcantarillado por red pública a diferencia de Lima y Callao. Además, nótese que los distritos ubicados al este y sur de la provincia de Lima son aquellos que presentan los valores más bajos de cobertura con menos del 87 por ciento.

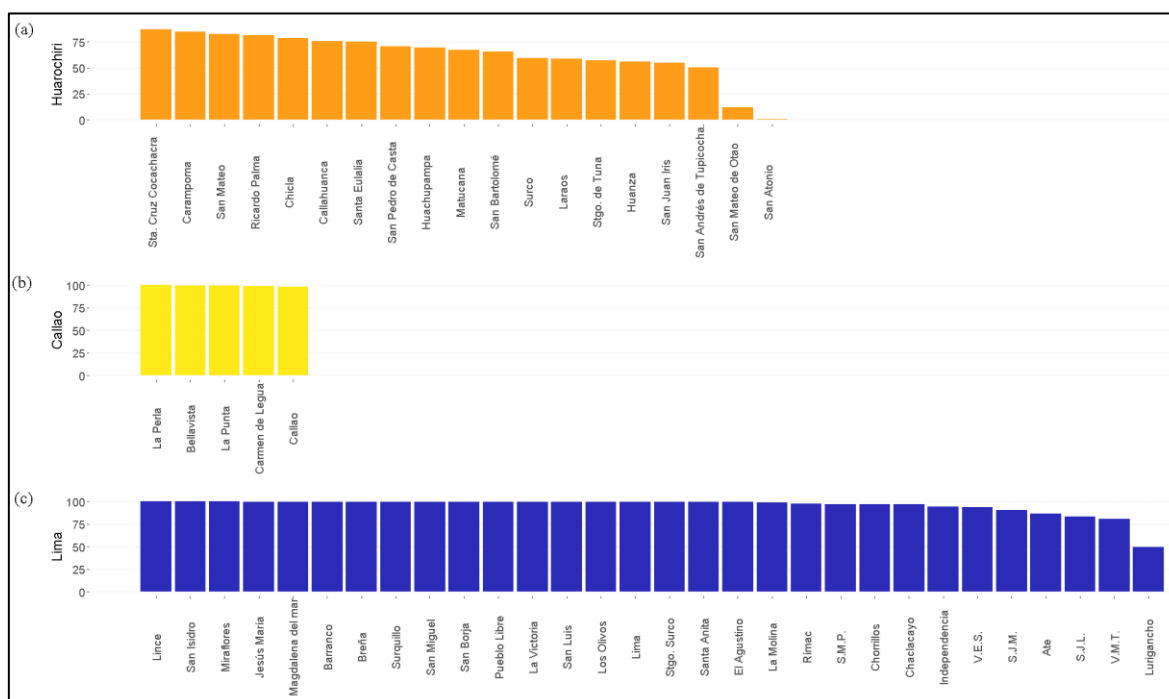


**Figura 12: Variación espacial de la cobertura de alcantarillado por red pública a nivel distrital**

De igual forma se aprecia en la Figura 13 que dentro de la provincia de Lima el distrito de Lurigancho es el que presenta el porcentaje más bajo de cobertura con un 49.9 por ciento, ubicándose muy por debajo de la media de sus vecinos.

Para el caso de la provincia del Callao conviene subrayar que los 5 distritos abarcados presentan valores muy altos en cuanto a cobertura por red pública sin embargo para la

provincia de Huarochirí los distritos de San Mateo de Otayo y San Antonio presentan valores menores al 15 por ciento, lo que los posiciona como los 2 distritos más críticos en relación a la cobertura por red pública de alcantarillado de toda el área de estudio, valores que se pueden apreciar en la Tabla 16 (Anexo 1).



**Figura 13: Cobertura de alcantarillado por red pública por provincia: Huarochirí (a), Callao (b) y Lima (c)**

#### 4.1.2. Residuos sólidos

En cuanto a los residuos sólidos se tiene la generación municipal en toneladas, así como la caracterización según el Ministerio del Ambiente que divide en 4 categorías que se detallan a continuación:

- Orgánicos: Residuos de alimentos, residuos de maleza y poda y otros orgánicos
- Inorgánicos: Papel, cartón, vidrios, aceite, plástico, latas, metales, otros metales
- No aprovechables: Textiles, caucho, bolsas plásticas, tecnopor, residuos inertes, restos de medicamentos, otros residuos no categorizados
- Peligrosos: Residuos sanitarios, pilas, otros

En la Tabla 4 se detalla la caracterización de los residuos sólidos a nivel de provincia, por ejemplo, cabe resaltar la heterogeneidad en cuanto la cantidad que se genera dentro de los



distritos de la provincia de Lima que asciende a 4 997,78 toneladas por día en comparación con las 19,67 toneladas que se generan en los distritos de la provincia de Huarochirí, esta última representa menos del 1 por ciento de lo que se genera en Lima. Si bien esto se da por la cantidad de población que albergan, cuando se observa la generación por habitante también existen diferencia como por ejemplo un habitante en un distrito de la provincia de Huarochirí genera en promedio 0,44 kg. d<sup>-1</sup> de residuo frente a los 0,71 kg. d<sup>-1</sup> que genera un habitante en la provincia de Lima.

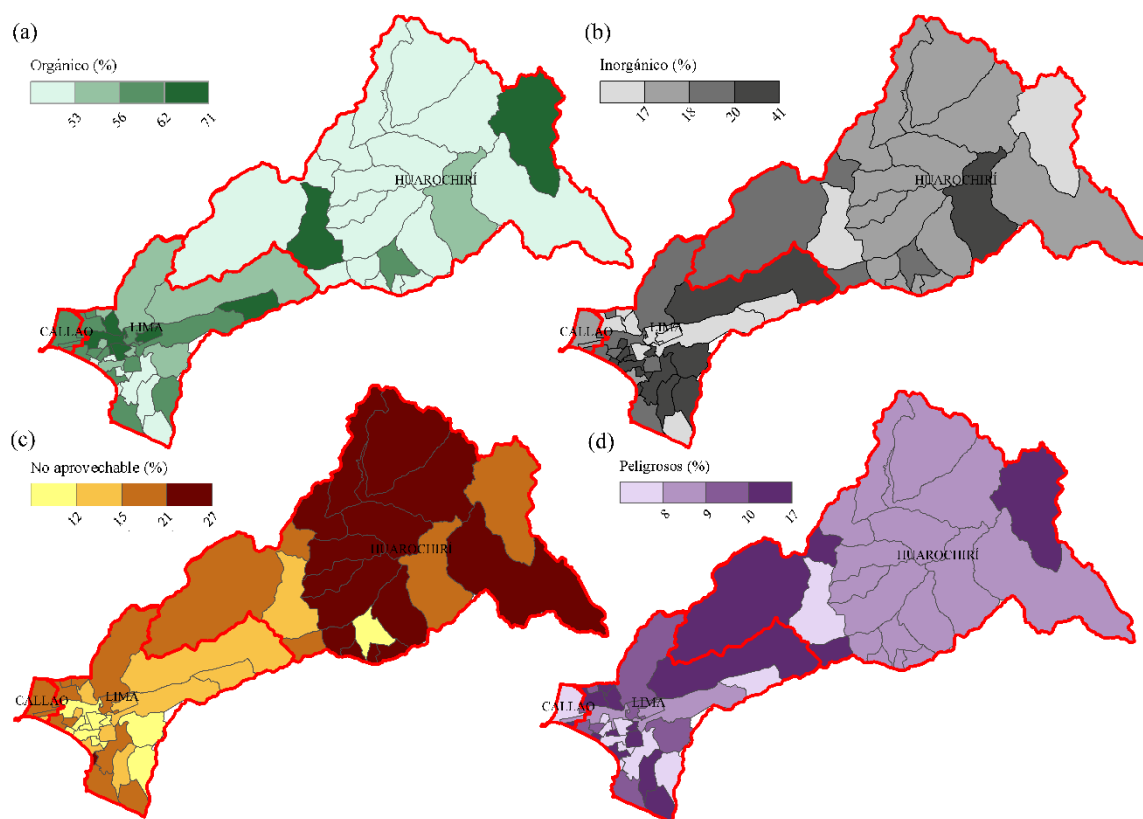
**Tabla 4: Generación de residuos sólidos por provincia en la cuenca Rímac**

<b>Provincia</b>	<b>Población total</b>	<b>Generación residuos (kg . d<sup>-1</sup> . hab<sup>-1</sup>)</b>	<b>Generación total de residuos (t . d<sup>-1</sup>)</b>
Huarochirí	45 223	0,44	19,67
Callao	633 597	0,61	388,30
Lima	7 020 743	0,71	4997,78

En la Figura 14 es posible observar la variación en la composición de los residuos sólidos en cada uno de los distritos de la cuenca del río Rímac. Para los residuos orgánicos se observa que, en la provincia de Huarochirí, alrededor del 80 por ciento de los distritos tienen <53 por ciento de generación, mientras que para los distritos de las provincias de Lima y Callao varían entre 53 a 71 por ciento. Respecto a la fracción inorgánica se tiene que para la parte media-alta de la cuenca (provincia de Huarochirí) sólo hay un distrito con más del 20 por ciento de la composición.

En relación a la composición no aprovechable es evidente el contraste entre las provincias de Lima y Callao frente a Huarochirí. Los primeros presentan valores <21 por ciento en su generación llegando hasta menos del 12 por ciento, mientras que alrededor del 70 por ciento de los distritos de Huarochirí tienen en su composición entre 21 y 27 por ciento.

Finalmente, para la porción peligrosa de los residuos, se tienen los valores más bajos no >17 por ciento. Logrando apreciar un patrón diferenciador para los distritos de la parte alta de Huarochirí que producen <8 por ciento.



**Figura 14: Composición de los residuos sólidos en los distritos de la cuenca Rímac: orgánicos (a.), inorgánicos (b), no aprovechables (c) y peligrosos (d) categorizada por cuantiles**

## **4.2. CARACTERIZACIÓN DE CASOS POSITIVOS Y TASA DE INCIDENCIA DE COVID-19**

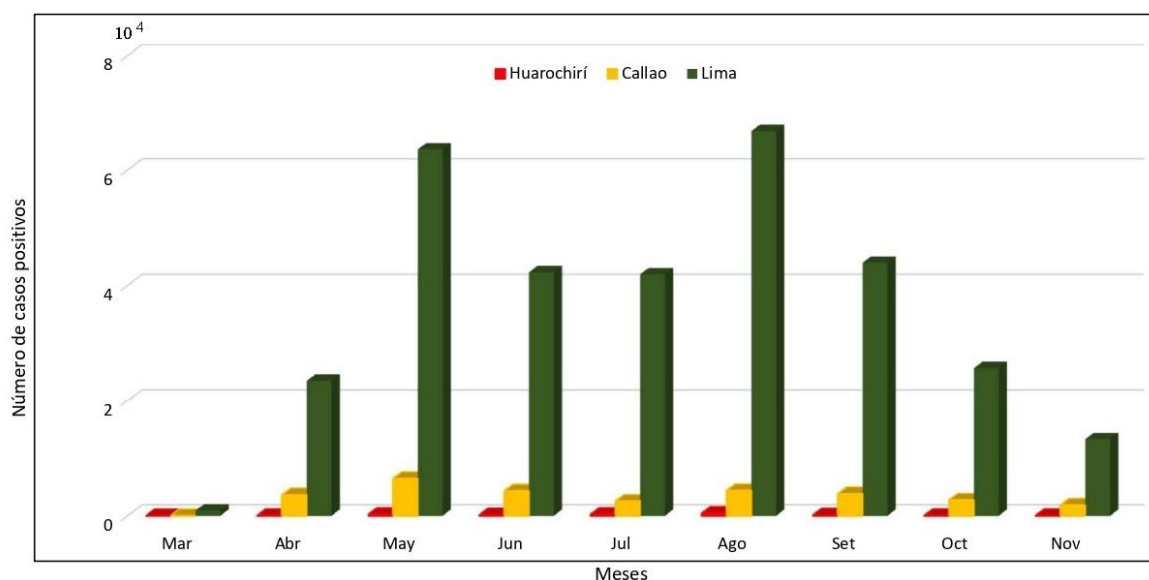
### **4.2.1. Número de casos positivos**

La evolución del número de casos positivos confirmados presenta grandes diferencias entre las provincias de la parte baja de la cuenca (provincias del Callao y Lima) frente a la parte media y alta de la zona de estudio (provincia de Huarochirí). Se aprecia en la Tabla 5 que en total hay 353 662 casos positivos confirmados al 30 de noviembre del 2020, en la cual Lima posee el 91,19 por ciento de estos con 321 841 casos, seguido de la provincia constitucional del Callao con 29 713 casos que representan el 8,41 por ciento y finalmente la provincia de Huarochirí con 1397 casos que equivale al 0,40 por ciento.

**Tabla 5: Números de casos positivos por provincia**

Provincia	Número de casos positivos	%
Lima	322 527	91,19
Callao	29 737	8,41
Huarocharí	1 398	0,40
Total	353 662	100,00

En la Figura 15 se aprecia la evolución de los casos positivos por provincia durante los meses de estudio. Aquí es posible observar los dos grandes picos de contagios ocurridos en la provincia de Lima en los meses de mayo y agosto, además de notar la enorme diferencia que existe en el nivel de contagio entre las tres provincias en estudio, pudiéndose atribuir a ello muchos factores tales como los económicos, sociales, meteorológicos e incluso la disponibilidad de pruebas para detectar el virus.



**Figura 15: Evolución mensual del número de casos positivos de COVID-19 por provincia**

Para el caso de la distribución de los casos positivos por sexo por provincia se tiene que el porcentaje de hombres tanto en Huarochirí, Callao y Lima es superior en el porcentaje de mujeres presentando valores de 52,15, 56,10 y 52,53 por ciento, respectivamente, tal como se aprecia en la Tabla 6:

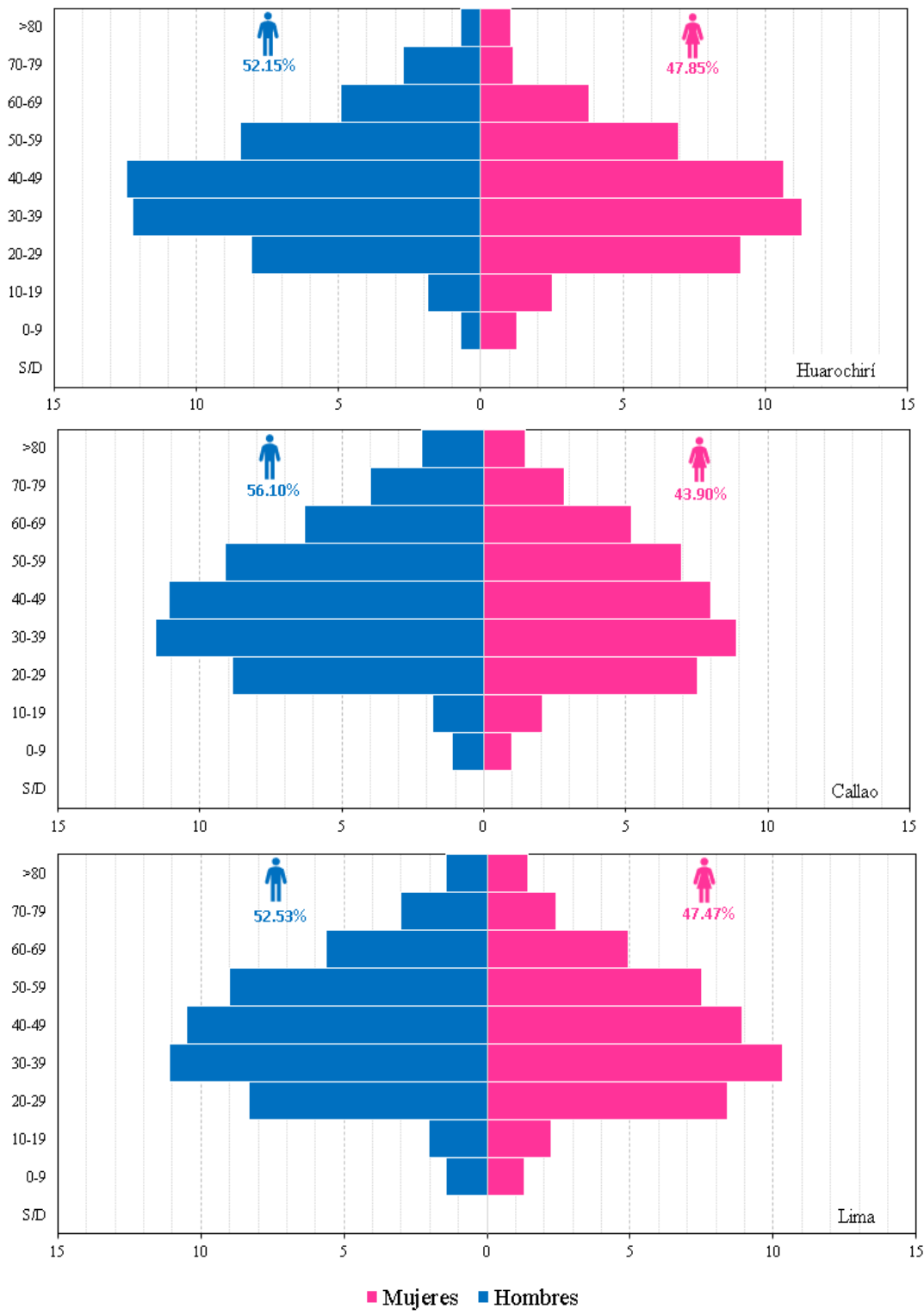
**Tabla 6: Número de casos positivos por provincia según sexo**

Edad	Provincia		
	Huarocharí	Callao	Lima
Hombres	729 (52,15) *	16 681 (56,10) *	169 423 (52,53) *
Mujeres	669 (47,85) *	13 056 (43,90) *	153 104 (47,47) *

( ) \* : Porcentaje (%)

Para el caso de la pirámide poblacional de los casos positivos por provincia se tiene que en el rango de edad entre 30 y 49 años están los mayores porcentajes de contagio en las tres provincias. En ese sentido en el caso de las mujeres se aprecia en la Figura 16 que para las tres provincias el mayor número de contagios se encuentra entre el rango de 30 a 39 años mientras que para el caso de los hombres en las provincias de Lima y Callao están dentro del mismo rango anterior, sin embargo, para la provincia de Huarocharí se encuentra entre los 40 a 49 años.

Además Jin *et al.* (2020) y Mukherjee & Pahan (2021), concluyen que no es posible afirmar que los hombres y mujeres poseen la misma prevalencia o una diferencia significativa en las tasas de contagio, sin embargo sí encontraron que fueron precisamente los hombres los que poseen más riesgo obtener resultados críticos y desencadenar en muerte.

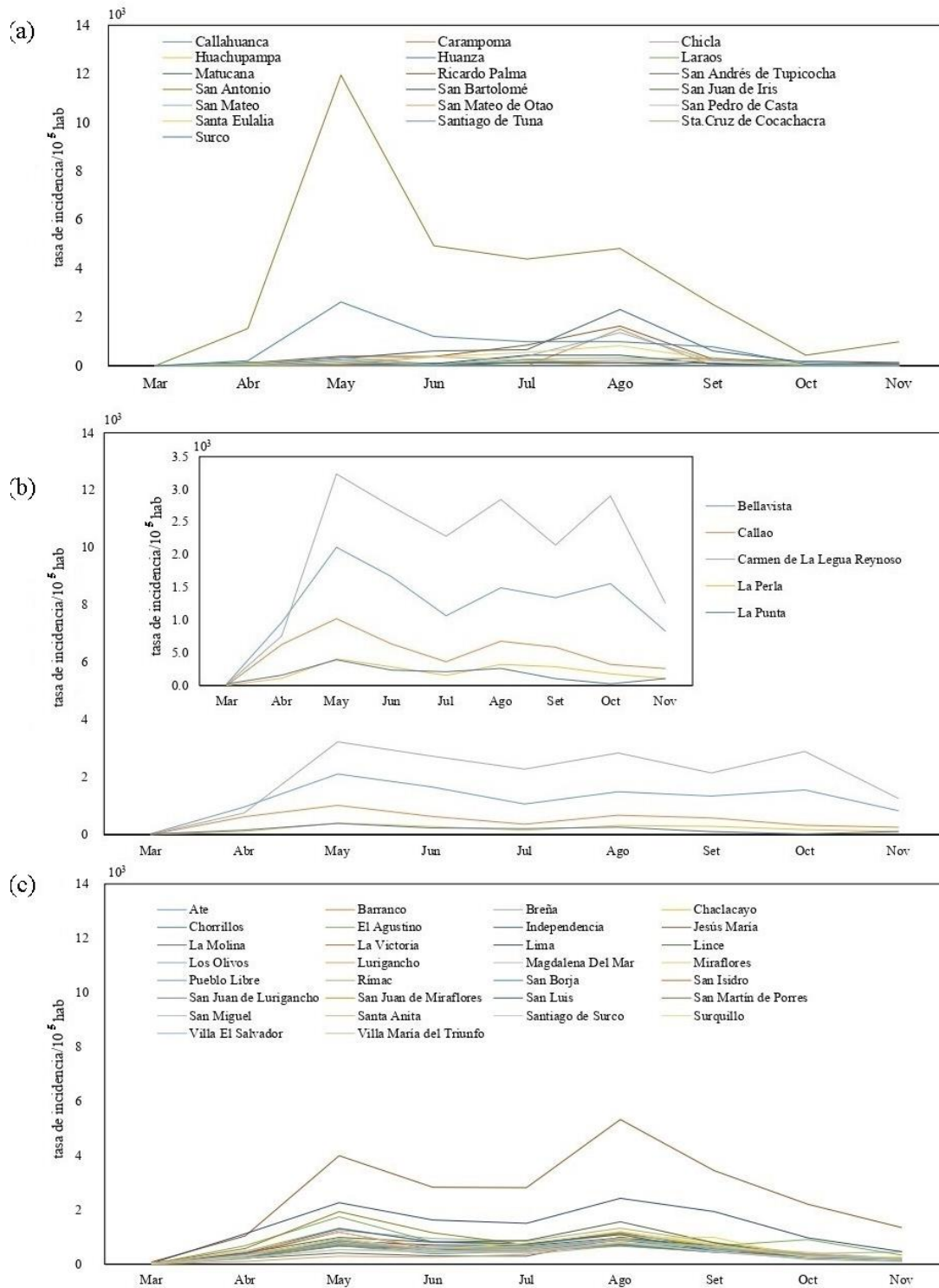


**Figura 16: Pirámide poblacional de los casos positivos de COVID-19 por provincia**

#### **4.2.2. Tasa de incidencia**

En cuanto a las tasas de incidencia a lo largo del periodo de estudio se tiene que los mayores picos se dan en los meses de mayo y agosto, coincidiendo con las fases 1 y 3 de la reanudación económica de actividades. Es así que el 03 de mayo la Presidencia del Consejo de Ministros emite el Decreto Supremo N° 080-2020-PCM en donde se aprueba la reanudación de actividades económicas en forma gradual y progresiva, iniciándose así la Fase 1 que contempla las actividades vinculadas a la minería e industria, construcción, servicios y turismo y comercio. Adicionalmente el 30 de junio del 2020 se emite el Decreto Supremo N° 117-2020-PCM donde se da la aprobación e implementación de la Fase 3 de la reanudación de actividades entre las que destacan el comercio al por mayor y menor, los restaurantes en general (con aforo al 40%), el transporte aéreo (sólo vuelos nacionales) y el centro de entretenimiento y atractivos turísticos con capacidad limitada. Estas reactivaciones podrían estar relacionadas con el aumento en la tasa de incidencia puesto que conllevaron a un mayor flujo de personas, aglomeración en lugares públicos, medidas sanitarias incipientes traduciéndose en una propagación acelerada del virus.

Para el caso de la tasa de incidencia para los distritos de la provincia del Callao, los mayores valores se pueden observar en la Figura 17, en mayo el distrito de Carmen de La Legua y Reynoso presenta la mayor tasa de incidencia con 3 233,90 casos por cada 100 000 habitantes, mientras que en el distrito de La Punta se tiene 391,75 casos por cada 100 000 habitantes.



**Figura 17: Tasa de incidencia de Covid-19 por provincia: Huarochirí (a), Callao (b) y Lima (c)**

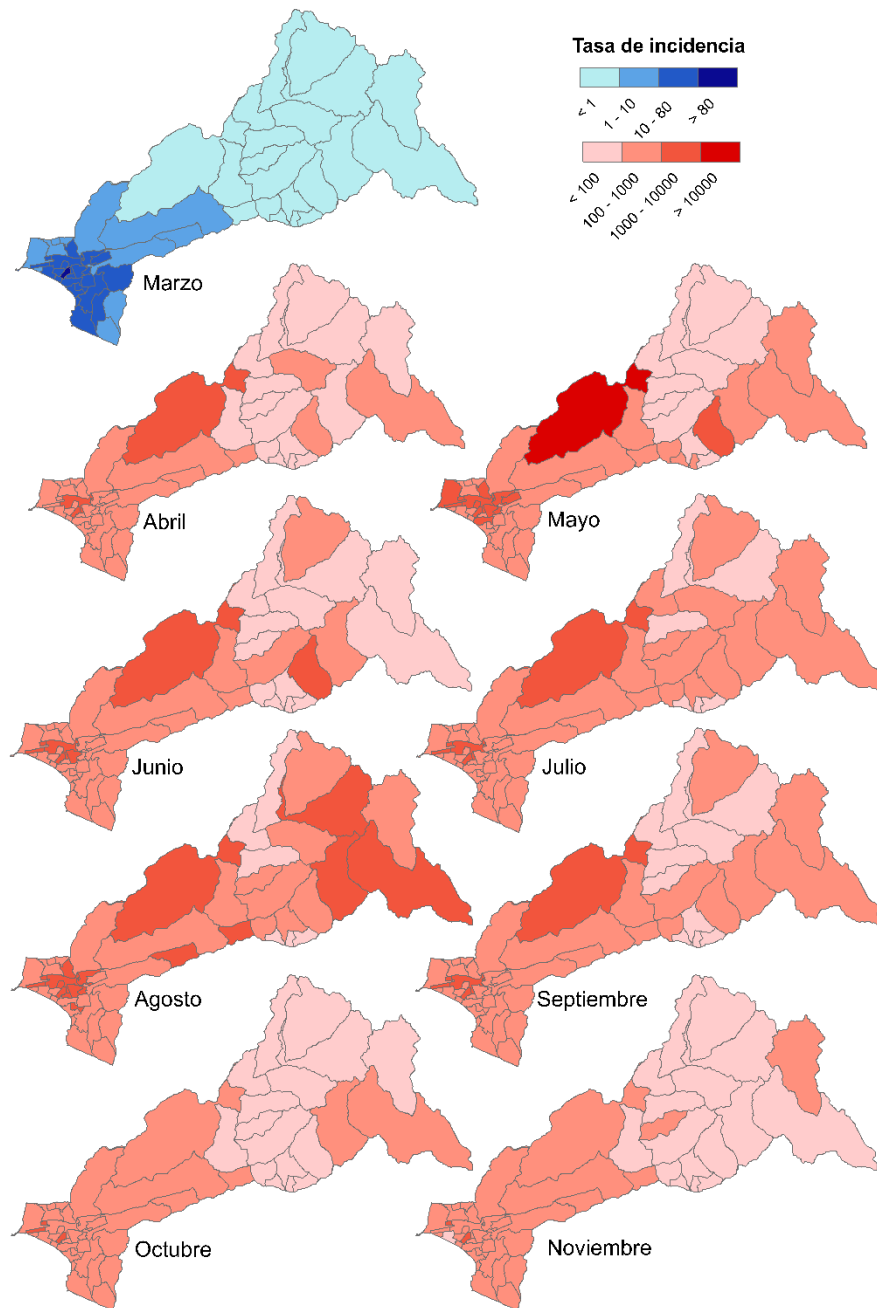
En cuanto a los distritos ubicados en la provincia de Lima se tiene que Jesús María presenta en mayo la más alta tasa de incidencia con 3 996,87 casos positivos por cada 100 000 habitantes seguida por Lima con 2 265,31 casos por cada 100 000 habitantes, adicionalmente se nota el segundo pico en el mes de agosto en estos mismos distritos con 5 323,85 y 2 425,92 casos por cada 100 000 habitantes respectivamente.

Agregando a lo anterior, para los distritos de la provincia de Huarochirí se puede observar en la Figura 17 que en el mes de mayo el distrito de San Antonio presenta una tasa de incidencia de 11 951,75 casos por cada 100 000 habitantes seguida por Surco con 2 629,71 casos por cada 100 000 habitantes.

Igualmente, para el mes de agosto el distrito de San Antonio es el que presenta el más alto valor con 4 824,56, sin embargo, esta vez es seguido por el distrito de Matucana con 2 316,41 casos por cada 100 000 habitantes.

En la Figura 18 se aprecia la distribución espacial de la tasa de incidencia por distritos en los meses de estudio, cabe señalar que el mes de marzo es aquel donde se presentan las menores tasa de incidencia comparando con los otros meses de estudio. Además, durante el periodo abril – septiembre se tuvieron las más altas tasas comenzando a descender a partir del mes de octubre. Siendo el mes de agosto el que posee las más altas tasas de incidencia para todos los distritos.





**Figura 18: Distribución espacial de la tasa de incidencia por distritos**

Se destaca que el distrito de San Antonio en la provincia de Huarochirí es el que posee el más alto valor de tasa de incidencia durante varios meses de estudio, cuestión que puede verse influenciada por la cercanía con el distrito de San Juan de Lurigancho que es el más poblado de la capital.

### **4.3. MODELOS GLOBALES PARA AJUSTAR LA TASA DE INCIDENCIA DE COVID-19 EN TODOS LOS DISTRITOS PERTENECIENTES A LA CUENCA**

#### **4.3.1. Correlación no paramétrica de la tasa de incidencia y servicios de saneamiento**

En la Tabla 7 se muestran los resultados obtenidos de la correlación de Kendall ( $\tau$ ) y Spearman ( $\rho$ ) entre la tasa de incidencia y las variables de saneamiento, los valores con los símbolos \* y \*\* denotan el nivel de significancia del cinco y uno por ciento respectivamente. Aquí es posible apreciar una correlación lineal significativa positiva entre los porcentajes de cobertura de agua potable por red pública (Ag\_RPubl), porcentaje de viviendas que pagan por el servicio de agua (Ag\_Pagan) y porcentaje de viviendas que pagan por el servicio de agua a una empresa prestadora de servicios (Ag\_PagaEPS) con la tasa de incidencia de COVID-19 durante todos los meses de estudio, con el mayor valor en marzo con 0,85, 0,88 y 0,92 respectivamente como coeficiente de correlación de Spearman. A diferencia de Rocha da Silva *et al.* (2020), que encontraron que la tasa de incidencia mostró una correlación significativamente negativa con el índice de servicio total de agua en Brasil, lo que explican que se podría deber a que en áreas económicamente pobres los problemas de saneamiento son severos y estarían influyendo en el aumento de la tasa de incidencia. Por otro lado, para el caso peruano la relación positiva puede ser debido a varios factores como: el acceso al agua potable por red pública está relacionado a las áreas urbanas con mayor densidad de población y congestión de espacios públicos, así mismos es en estas zonas las que poseen el mayor número de casos positivos.

Para el caso de acceso al agua potable por red pública dentro de la vivienda también existe una correlación significativa positiva en gran parte del periodo de estudio, a excepciones de los meses de julio y agosto donde no es significativa la correlación y podría deberse a que son los meses en los que la población ha salido del estado de cuarentena y las medidas de control e higiene están siendo implementadas gradualmente.

Siguiendo con el análisis de correlaciones relacionadas al acceso al agua potable se encontró que para la variable porcentaje de viviendas con servicio de agua potable por red pública brindado los 7 días a la semana se tiene una correlación positiva significativa con la tasa de incidencia de COVID-19 durante casi todo el periodo de estudio menos julio y agosto, mientras que las variables como porcentaje de viviendas con servicio de agua potable por

red pública brindado no todos los días, porcentaje de viviendas con el agua por red pública de 9, 18 y 24 horas, no son significativas.

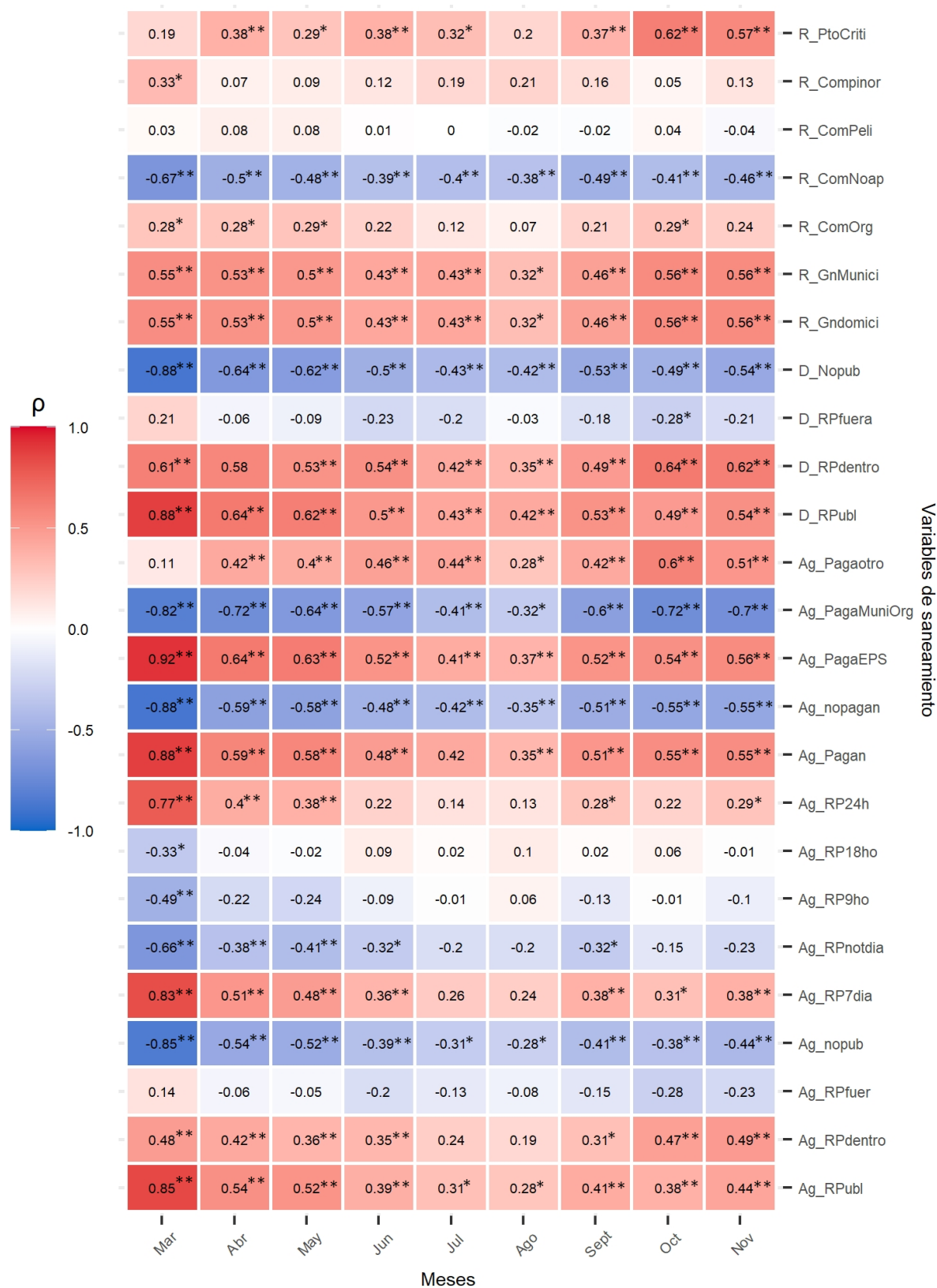
Para el caso de las correlaciones en relación a las variables de acceso al servicio de alcantarillado, se ha encontrado que el porcentaje de viviendas con acceso a red de alcantarillado por red pública, tiene una correlación positiva significativa con la tasa de incidencia de COVID-19 durante todos los meses de estudio, siendo en marzo el mayor valor con 0,88. Además, la variable porcentaje de viviendas con red pública de alcantarillado dentro de la vivienda también tiene una correlación positiva significativa durante todo el periodo de estudio con el máximo valor de 0,64 en el mes de octubre. Lo que podría deberse como mencionan Gormley *et al.* (2020) a que los sistemas de conexiones de alcantarillado podrían ser considerados como una vía de transmisión potencial para COVID-19 debido a que existe una carga viral importante dentro del sistema combinado con el potencial de transmisión aérea debido a la aerosolización del virus.

En el caso de las correlaciones con las variables relacionadas a los residuos sólidos, se encontró que existe una asociación positiva significativa en la generación municipal y la cantidad de puntos críticos por distritos con la tasa de incidencia de COVID-19, que estaría relacionado tal como lo manifiestan Klemeš *et al.* (2020), que sostienen que la pandemia del COVID-19 ha generado importantes desafíos en el manejo de residuos sólidos urbanos (RSU) y residuos médicos peligrosos por lo que no se ha podido hacer frente a los residuos plásticos existentes de manera satisfactoria, además que el inminente aumento en la cantidad generada de residuos amenaza con colapsar los sistemas de gestión de residuos existentes, al igual que la capacidad de atención médica. Al respecto podemos mencionar que, en los municipios de los distritos netamente urbanos y más densos es donde se da la mayor generación de residuos sólidos. También cabe resaltar que al dar como resultado una asociación positiva con la cantidad de puntos críticos podemos inferir que cuanto mayor es la cantidad de puntos encontrados mayor es la posibilidad de que las condiciones higiénicas de los pobladores no sean las adecuadas y se estén generando focos de infección en esas zonas. La Tabla 7 resume los resultados de las pruebas de correlación no paramétrica de Kendall y Spearman:

**Tabla 7: Coeficiente de correlación no paramétrica Kendall ( $\tau$ ) y Spearman ( $\rho$ ) entre la tasa de incidencia de COVID-19 y los servicios de saneamiento en la cuenca Rímac**

Variable	Mar		Abr		May		Jun		Jul		Ago		Sept		Oct		Nov	
	$\tau$	$\rho$	$\tau$	$\rho$	$\tau$	$\rho$	$\tau$	$\rho$	$\tau$	$\rho$	$\tau$	$\rho$	$\tau$	$\rho$	$\tau$	$\rho$	$\tau$	$\rho$
Ag_RPubl	0.68 **	0.85 **	0.37 **	0.54 **	0.37 **	0.52 **	0.26 **	0.39 **	0.23 *	0.31 *	0.19 *	0.28 *	0.29 **	0.41 **	0.23 *	0.38 **	0.30 **	0.44 **
Ag_RPdentro	0.29 **	0.48 **	0.30 **	0.42 **	0.24 *	0.36 **	0.26 **	0.35 **	0.155	0.237	0.121	0.190	0.20 *	0.31 *	0.34 **	0.47 **	0.36 **	0.49 **
Ag_RPfuer	0.128	0.137	-0.041	-0.062	-0.022	-0.047	-0.114	-0.200	-0.076	-0.127	-0.044	-0.084	-0.078	-0.148	-0.20 *	-0.28 *	-0.166	-0.234
Ag_NoPub	-0.68 **	-0.85 **	-0.37 **	-0.54 **	-0.37 **	-0.52 **	-0.26 **	-0.39 **	-0.23 *	-0.31 *	-0.19 *	-0.28 *	-0.29 **	-0.41 **	-0.23 *	-0.38 **	-0.30 **	-0.44 **
Ag_RP7dia	0.65 **	0.83 **	0.35 **	0.51 **	0.33 **	0.48 **	0.24 *	0.36 **	0.20 *	0.265	0.168	0.243	0.26 **	0.38 **	0.181	0.31 *	0.26 **	0.38 **
Ag_RPnotdia	-0.49 **	-0.66 **	-0.26 **	-0.38 **	-0.25 **	-0.41 **	-0.20 *	-0.32 *	-0.134	-0.205	-0.135	-0.199	-0.21 *	-0.32 *	-0.095	-0.154	-0.156	-0.231
Ag_RP9ho	-0.37 **	-0.49 **	-0.147	-0.222	-0.155	-0.238	-0.051	-0.085	-0.013	-0.008	0.032	0.060	-0.092	-0.134	0.019	-0.009	-0.057	-0.102
Ag_RP18ho	-0.25 *	-0.33 *	-0.034	-0.036	-0.025	-0.018	0.052	0.090	0.013	0.025	0.068	0.097	0.009	0.025	0.064	0.063	-0.003	-0.010
Ag_RP24h	0.59 **	0.77 **	0.27 **	0.40 **	0.25 **	0.38 **	0.128	0.218	0.103	0.144	0.086	0.130	0.183	0.28 *	0.118	0.222	0.19 *	0.29 *
Ag_Pagan	0.70 **	0.88 **	0.39 **	0.59 **	0.42 **	0.58 **	0.33 **	0.48 **	0.29 **	0.42 **	0.24 *	0.35 **	0.36 **	0.51 **	0.37 **	0.55 **	0.38 **	0.55 **
Ag_Nopagan	-0.70 **	-0.88 **	-0.39 **	-0.59 **	-0.42 **	-0.58 **	-0.33 **	-0.48 **	-0.29 **	-0.42 **	-0.24 *	-0.35 **	-0.36 **	-0.51 **	-0.37 **	-0.55 **	-0.38 **	-0.55 **
Ag_PagaEPS	0.78 **	0.92 **	0.47 **	0.64 **	0.48 **	0.63 **	0.38 **	0.52 **	0.31 **	0.41 **	0.27 **	0.37 **	0.37 **	0.52 **	0.37 **	0.54 **	0.41 **	0.56 **
Ag_PagaMuniOrg	-0.65 **	-0.82 **	-0.54 **	-0.72 **	-0.45 **	-0.64 **	-0.42 **	-0.57 **	-0.30 **	-0.41 **	-0.24 *	-0.32 *	-0.43 **	-0.60 **	-0.53 **	-0.72 **	-0.50 **	-0.70 **
Ag_Pagaotro	0.008	0.114	0.29 **	0.42 **	0.27 **	0.40 **	0.31 **	0.46 **	0.29 **	0.44 **	0.177	0.28 *	0.29 **	0.42 **	0.44 **	0.60 **	0.36 **	0.51 **
D_RPubl	0.72 **	0.88 **	0.47 **	0.64 **	0.48 **	0.62 **	0.36 **	0.50 **	0.31 **	0.43 **	0.29 **	0.42 **	0.39 **	0.53 **	0.33 **	0.49 **	0.39 **	0.54 **
D_RPdentro	0.38 **	0.61 **	0.431	0.579	0.37 **	0.53 **	0.39 **	0.54 **	0.28 **	0.42 **	0.24 *	0.35 **	0.33 **	0.49 **	0.47 **	0.64 **	0.47 **	0.62 **
D_RPfuer	0.21 *	0.209	-0.049	-0.064	-0.054	-0.088	-0.145	-0.232	-0.120	-0.200	-0.012	-0.034	-0.105	-0.180	-0.20 *	-0.28 *	-0.142	-0.205
D_NoPub	-0.72 **	-0.88 **	-0.47 **	-0.64 **	-0.48 **	-0.62 **	-0.36 **	-0.50 **	-0.31 **	-0.43 **	-0.29 **	-0.42 **	-0.39 **	-0.53 **	-0.33 **	-0.49 **	-0.39 **	-0.54 **
R_GnMunici	0.39 **	0.55 **	0.36 **	0.53 **	0.33 **	0.50 **	0.28 **	0.43 **	0.28 **	0.43 **	0.21 *	0.32 *	0.29 **	0.46 **	0.38 **	0.56 **	0.39 **	0.56 **
R_ComOrg	0.39 **	0.55 **	0.36 **	0.53 **	0.33 **	0.50 **	0.28 **	0.43 **	0.28 **	0.43 **	0.21 *	0.32 *	0.29 **	0.46 **	0.38 **	0.56 **	0.39 **	0.56 **
R_ComNoap	0.21 *	0.28 *	0.22 *	0.28 *	0.21 *	.286*	0.154	0.221	0.100	0.119	0.059	0.071	0.157	0.214	0.22 *	0.29 *	0.177	0.242
R_ComPeli	-0.53 **	-0.67 **	-0.37 **	-0.50 **	-0.35 **	-0.48 **	-0.27 **	-0.39 **	-0.28 **	-0.40 **	-0.27 **	-0.38 **	-0.35 **	-0.49 **	-0.29 **	-0.41 **	-0.32 **	-0.46 **
R_Compignor	0.036	0.033	0.077	0.076	0.063	0.079	0.020	0.007	0.009	0.003	-0.002	-0.020	0.007	-0.017	0.048	0.041	0.016	-0.039
R_PtoCriti	0.22 *	0.33 *	0.066	0.068	0.077	0.093	0.100	0.120	0.133	0.187	0.154	0.211	0.138	0.165	0.062	0.053	0.117	0.128

Nota: \*\*, \* representa el 5% y 10% de nivel de significancia.



**Figura 19: Correlación de Spearman de la tasa de incidencia COVID-19 y los servicios de saneamiento en la cuenca Rímac**

Nota : \*\*. La correlación es significativa en el nivel 0.01 (bilateral).

\*. La correlación es significativa en el nivel 0.05 (bilateral).

### 4.3.2. Ajuste de modelos globales

#### a. Índice de Morán de la tasa de incidencia de COVID-19

En la Tabla 8 se visualizan los resultados de la prueba de autocorrelación espacial (test de Índice de Morán) para la tasa de incidencia de COVID-19. Se encontró que los valores que corresponden a marzo y abril, están por encima de los 0,30, siendo este un valor positivo, lo sugiere que las tasas de incidencia están agrupadas y además son estadísticamente significativos. Por otro lado, en mayo, junio, julio, agosto y noviembre, el valor del índice de Morán es cercano a 0 y no son significativos por lo que se acepta la hipótesis nula, la cual plantea que, para esos meses, existe una distribución aleatoria de la tasa de incidencia de COVID-19. Sin embargo, en septiembre y octubre, se tiene valores significativos estadísticamente, pero  $<0,2$ , lo que refleja la presencia de una autocorrelación positiva, es decir, que las tasas de incidencia forman conglomerados en la cuenca Rímac.

**Tabla 8: Índice de Morán para la tasa de incidencia de COVID-19**

Mes	Índice de Morán
Marzo	0,62 ***
Abril	0,35 ***
Mayo	-0,02
Junio	0,06
Julio	0,04
Agosto	0,07
Septiembre	0,20 ***
Octubre	0,18 ***
Noviembre	0,13

\*\*\* Representan el 1%

**b. Mínimos cuadrados ordinarios (MCO)**

En la Tabla 9 se presenta los resultados del ajuste de las variables al modelo MCO, obteniéndose dos variables explicativas, la altitud y el porcentaje de cobertura de agua por camión cisterna (Ag\_pagotr), variables con baja multicolinealidad (<30) entre sí; y cuyo valor del intercepto y los coeficientes son estadísticamente significativos y se cumple el supuesto de la normalidad de los errores del modelo. Siendo que los valores de R<sup>2</sup> del modelo de mínimos cuadrados ordinarios varían entre el 43 y 66 por ciento (ver Tabla 9).

**Tabla 9: Datos estadísticos de la ecuación de regresión para el modelo de mínimos cuadrados ordinarios**

Detalle	Mar	Abr	May	Jun	Jul	Ago	Set	Oct	Nov
Intercepto	1,14	2,49	2,99	2,81	2,83	3,03	2,82	2,50	2,31
Variable 1: Altitud	-0.0004	-0.0006	-0.0006	-0.0006	-0.0004	-0.0004	-0.0006	-0.00063	-0.0006
Variable 2: Ag_pagotr	-0.0043	0.0328	0.0356	0.0309	0.0254	0.0246	0.0300	0.0284	0.0300
Número de datos	54	54	54	54	54	54	54	54	54
R <sup>2</sup>	0,64	0,65	0,61	0,57	0,49	0,43	0,57	0,65	0,66
R <sup>2</sup> adj.	0,63	0,64	0,59	0,56	0,47	0,40	0,56	0,63	0,65
Criterio de información de Akaike - AIC	55,03	109,17	117,85	111,48	104,73	118,92	114,39	107,11	92,38
Criterio de información de Schwarz-SWC	60,99	115,14	123,82	117,45	110,70	124,89	120,36	113,07	98,35
Jarque Bera	1,41	5,75	5,00	1,00	4,96	5,81	2,63	4,31	8,63
Prueba de Breusch-Pagan	2,99	8,79	28,88	23,25	44,77	53,10	34,76	7,35	16,77
Multicolinealidad	2,18	2,18	2,18	2,18	2,18	2,18	2,18	2,18	2,18
I Moran (error)	0,35	0,13	0,16	0,05	-0,11	0,27	0,09	0,21	0,05
LM-Lag	27,46	3,01	1,46	0,05	5,03	2,56	0,00	5,52	0,07
Robust LM-Lag	16,94	1,23	0,70	0,59	10,67	12,44	4,39	0,74	0,13
LM-Error	13,94	1,90	3,04	0,33	1,34	8,22	0,90	4,83	0,24
Robust LM-Error	3,41	0,12	2,29	0,88	6,98	18,10	5,30	0,05	0,31

Nivel de significancia 1% 5%

Ag\_pagotr: % de viviendas que pagan por el servicio de agua potable a camión cisterna u otro

En cuanto a las dos variables seleccionadas, estas presentaron significancia durante todos los meses de estudio a excepción de marzo, puesto que la variable porcentaje de cobertura de agua por camión cisterna u otros ( $Ag\_pago_{tr}$ ) no fue significativa. En relación al test de Jarque Bera que plantea como hipótesis nula la normalidad de los residuos, el cual resulta clave para poder aplicar las pruebas de hipótesis en los modelos de regresión lineal. Se cumplió durante casi todos los meses de marzo a octubre, sin embargo, para el mes de noviembre fue significativo por lo que se rechazó la hipótesis nula.

Para el análisis de homoscedasticidad de los residuos (test de Breusch-Pagan), resultó que para abril a noviembre se rechazó la hipótesis nula de homoscedasticidad de la varianza de los residuos, que esto podría deberse a la presencia de autocorrelación de la variable dependiente o que existen variables que no se han considerado en el modelo.

Además, el test que evidencia la autocorrelación espacial de los residuos es el Índice de Morán, el cual tiene como hipótesis nula la distribución aleatoria de los residuos. Para esta prueba se tiene que los meses de marzo, mayo, agosto y octubre presentan el problema de la autocorrelación de los residuos del modelo. Debido a estos resultados, indica que podría solucionarse este efecto con la aplicación de modelos que puedan medir este fenómeno. En la Tabla 10 se aprecia las ecuaciones para cada mes:

**Tabla 10: Resultados del modelo MCO para todos los meses de estudio**

Mes	Fórmula del modelo	R <sup>2</sup> ajustado
Mar	$T_{inci} = -0,78 \times Altitud - 0,07 \times Ag_{pago_{tr}}$	0,63
Abr	$T_{inci} = -0,79 \times Altitud + 0,32 \times Ag_{pago_{tr}}$	0,64
May	$T_{inci} = -0,76 \times Altitud + 0,34 \times Ag_{pago_{tr}}$	0,59
Jun	$T_{inci} = -0,73 \times Altitud + 0,33 \times Ag_{pago_{tr}}$	0,56
Jul	$T_{inci} = -0,67 \times Altitud + 0,32 \times Ag_{pago_{tr}}$	0,47
Ago	$T_{inci} = -0,63 \times Altitud + 0,28 \times Ag_{pago_{tr}}$	0,41
Sept	$T_{inci} = -0,74 \times Altitud + 0,31 \times Ag_{pago_{tr}}$	0,56
Oct	$T_{inci} = -0,79 \times Altitud + 0,29 \times Ag_{pago_{tr}}$	0,63
Nov	$T_{inci} = -0,79 \times Altitud + 0,34 \times Ag_{pago_{tr}}$	0,65

Nota: Variables estandarizadas  $(X - \bar{X})/\sigma$



En la Tabla 11 se presenta una propuesta para comparar por niveles entre la variable dependiente (Tasa de incidencia). Esta clasificación se efectuó al marcar los valores que diferencien las tres partes de la cuenca (zona baja, media y alta), y se obtuvo aquellos que muestran la variación entre estas zonas.

**Tabla 11: Propuesta de nivel de impacto de la tasa de incidencia modelada**

Nivel	Variable
	<i>Tasa de incidencia (Casos positivos x Hab<sup>-1</sup> x 100 000)</i>
Bajo	< 100
Medio	100 – 500
Alto	500 - 1100
Muy Alto	1100 - 4500

Además, en la Tabla 12 se coloca la clasificación de las variables independientes que afectan en nivel de impacto en la tasa de incidencia.

**Tabla 12: Propuesta de nivel de impacto de la clasificación y el efecto de las variables independientes**

Nivel	Variables	
	<i>Cobertura de agua por camión cisterna u otro (%)</i>	<i>Altitud (m.s.n.m.)</i>
Bajo	< 2	< 200
Medio	2 - 15	200 – 700
Alto	15 - 30	700 - 2000
Muy alto	30 - 75	> 2000

En la Figura 20, se observa la comparativa entre el nivel de tasa de incidencia modelada y el efecto conjunto de las variables independientes del modelo. Es así, que se tiene tres zonas en la cuenca de estudio con diferentes niveles de impacto, así como, dos periodos definidos: abril, octubre y noviembre, que corresponden al inicio y fin de la denominada primera ola de la pandemia), mientras que de mayo a septiembre es donde se presentan los más altos niveles de contagio.

En ese sentido, se tiene que para la primera zona o parte baja (provincias de Lima y Callao) se tiene un impacto en la tasa de incidencia en el rango de "medio" a "muy alto".

El impacto "medio" se da para el primer periodo y corresponde a un efecto "bajo" de las variables independientes (coberturas inferiores al 2 por ciento de agua por camión cisterna u otro y una altitud por debajo de los 200 metros sobre el nivel del mar), mientras que, el impacto entre el rango "alto" a "muy alto" se da en el segundo periodo.

Cabe resaltar que aquellos distritos que poseen valores del tipo "medio" en la cobertura de agua por camión cisterna (distritos de Chorrillos, San Martín de Porres, Independencia, Villa María del Triunfo, Villa El Salvador y San Juan de Miraflores), son los mismos que presentan tasas más elevadas llegando a un nivel de impacto "muy alto".

Esto evidencia que para aquellos distritos netamente clasificados como urbanos y que están ubicados en las provincias de Lima y Callao el porcentaje de viviendas que tienen cobertura de agua por camión cisterna u otro es muy relevante, pues son distritos con altas densidades y temperaturas más elevadas en contraste con la parte alta de la cuenca. Lo que conllevaría a que la población si bien posee agua potable por red pública con altos valores, el hecho de que exista población que se abastece por otro medio toma interés en gran medida.

Con respecto a la zona media de la cuenca de estudio (distritos de las provincias de Lima y Huarochirí), es posible evidenciar que el impacto varía en toda la amplitud de rango, es decir, desde un nivel "bajo y "medio" para la provincia de Huarochirí, y al contrario desde el nivel "medio" y "muy alto" para la provincia de Lima.

Aquí también es preciso señalar, que hay distritos donde el porcentaje de cobertura de agua por camión cisterna varía entre el 2 y 15 por ciento e incluso el distrito de Lurigancho sobrepasa 30 por ciento, y son estos los que poseen un mayor impacto llegando incluso al nivel "muy alto".

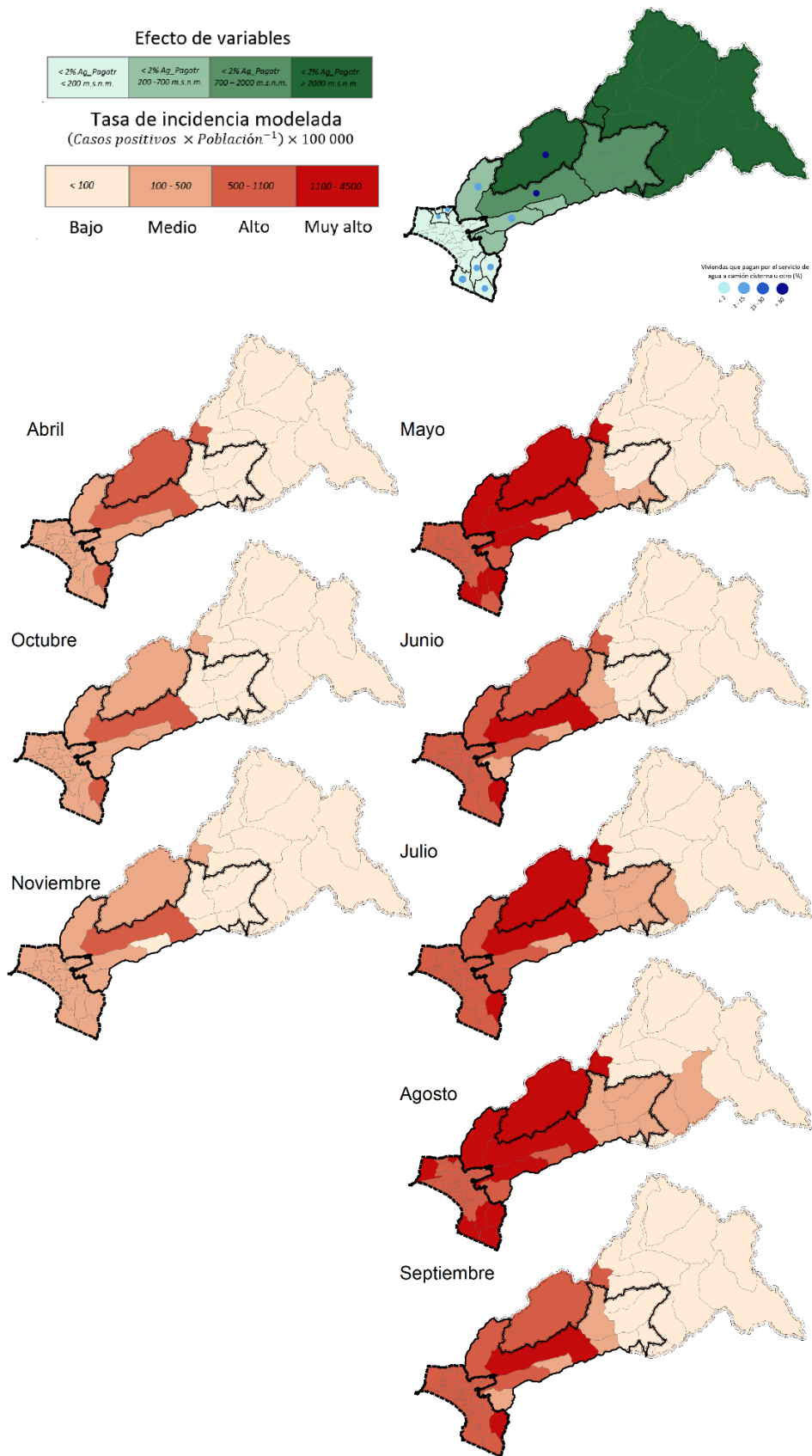
En esta zona el impacto varía mucho puesto que son distritos de transición entre la zona baja con la mayoría de población urbana y la zona alta con gran porcentaje de población rural, y

esto tiene implicancias tanto en las formas de vida como en la movilidad de las personas. Finalmente, a diferencia de la zona baja, en la parte alta de la cuenca se tiene un impacto clasificado como “bajo” y “medio” en su mayoría a excepción del distrito de San Antonio cuyo impacto llega al nivel “muy alto”.

Este nivel de impacto en esta parte alta podría deberse a que, en contraste con las altas densidades y el gran porcentaje de población urbana de la parte baja, aquí en los distritos de la provincia de Huarochirí existe población rural y dispersa, y las condiciones climáticas u otros factores estarían actuando para disminuir el impacto.

Estos resultados en relación a la altitud se asemejan a lo obtenido por Arias-Reyes *et al.* (2020) quienes analizaron niveles de contagio en la región tibetana de China comparada con Bolivia y Ecuador, encontrando una disminución del número de casos a mayor altitud, lo cual podría deberse a factores ambientales (calidad del aire) o fisiológicos. Además, Accinelli & Leon-Abarca (2020) , indican que para nuestro país por cada 500 m de ascenso, la tasa de incidencia y de mortalidad, se reduce un 22 y 40 por ciento respectivamente.

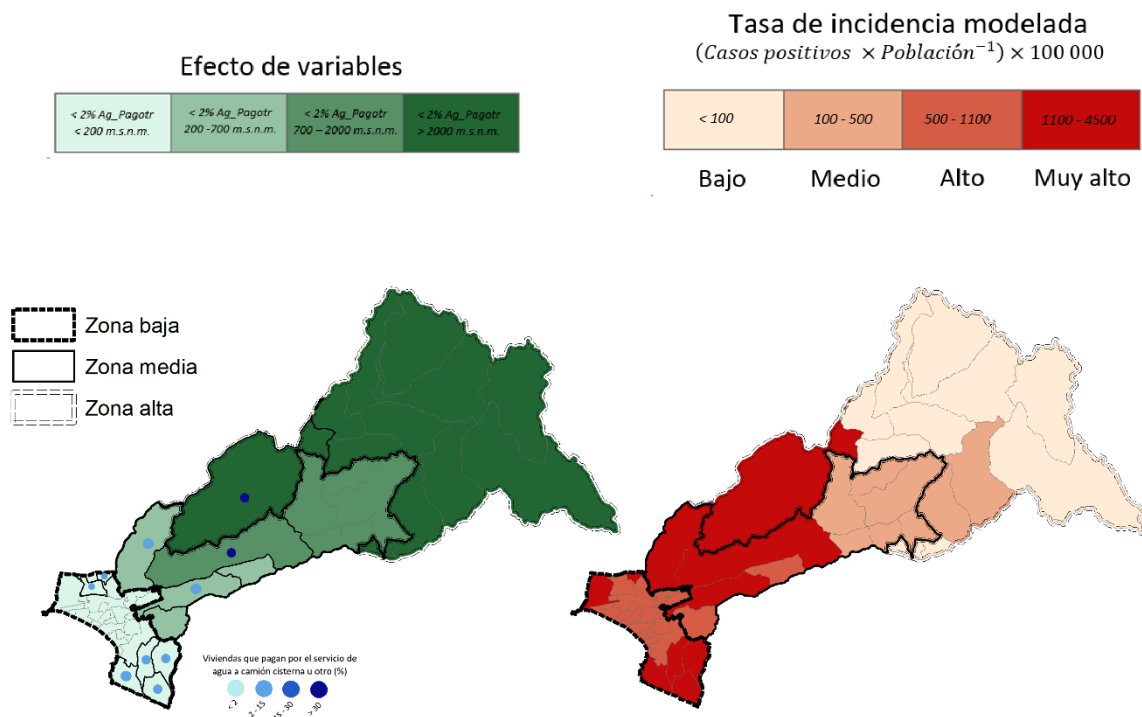
Esto pone en evidencia que la tasa de incidencia está siendo afectada por la variable cobertura de agua por camión cisterna u otro de manera conjunta con la variable altitud, aunque no de manera uniforme en toda la cuenca, la primera variable presenta un impacto positivo mientras que la segunda la afecta de manera negativa. Además, existe tres zonas claramente diferenciadas. Mientras que la parte baja posee un impacto entre el rango de “medio” y “muy alto”, en contraste con la parte alta que tiene entre “bajo” y “medio”.



**Figura 20: Distribución de las variables altitud y porcentaje de cobertura de la población que tiene acceso a agua por medio de pago a camión cisterna u otro con la distribución por niveles de la tasa de incidencia de COVID-19**

En la Figura 21, se aprecia el efecto de las variables en la tasa de incidencia modelada para el mes de agosto 2020, es así, que aquellos distritos con valores más elevados de cobertura de agua potable por camión cisterna son los que tiene un impacto “muy alto”, además de, que en la parte alta se tiene un nivel “bajo” en contraste con la parte alta, esto debido a la variable altitud.

El efecto de las variables independientes no indica exclusivamente que estas influyen de manera determinante en la tasa de incidencia, puesto que, podrían ser el reflejo indirecto de otras variables no consideradas en el estudio tales como: formas de vida, movilidad de las personas, calidad del aire, preexistencias de comorbilidades, etc.



**Figura 21: Comparativa entre el nivel de la tasa de incidencia modelada y el efecto de variables independientes para el mes de agosto 2020**

### c. Modelos considerando dependencia espacial

En la Tabla 13 se presentan los resultados del ajuste de los modelos con dependencia espacial, en meses que se incumple el supuesto de no autocorrelación espacial de los errores (I Morán). Es por ello que se realizó el análisis de los test de Lagrange (LM y Robust LM) para identificar la naturaleza del fenómeno de autocorrelación.

En el mes de marzo se verificó que ambos test (LM-Lag y Robust LM-Lag) fueron significativos, por lo que se aplicó el modelo de retraso espacial, resultando el coeficiente autorregresivo Rho ( $\rho$ ) significativo, por lo tanto, mostró una mejora del ajuste ( $R^2$ ) de 64 según mínimos cuadrados ordinarios a 79 en este modelo. Para mayo y agosto, el coeficiente autorregresivo Lambda ( $\lambda$ ) fue significativo, es por ello que, para estos meses, el modelo de error espacial captura el efecto de autocorrelación en los errores, y mejora el  $R^2$ , de 0,61 y 0,43 a 0,64 y 0,60 respectivamente. Por otro lado, en octubre, ambos modelos cumplieron con la significancia de los coeficientes, aunque el valor de  $R^2$  para el modelo de retraso espacial con 0.71 fue el mejor ajuste.

**Tabla 13: Parámetros de los modelos de retraso espacial y error espacial**

Mes	Modelo	Variables					Bondad de ajuste	
		Intercepto	Altitud	Ag_pagotr	$\rho$	$\lambda$	Breusch-Pagan	$R^2$
Mar	SLM	0,46	-0,0002	-0,0051	0,60	-	6,89	0,79
May	SEM	2,97	-0,0006	0,0403	-	0,36	25,04	0,64
Ago	SEM	3,27	-0,0008	0,0420	-	0,71	25,25	0,60
Oct	SEM	2,42	-0,0006	0,0287	-	0,47	2,76	0,70
	SLM	1,21	-0,0033	0,0222	0,50	-	4,90	0,71

Nivel de significancia: 1% 5%

Ag\_pagotr: % de viviendas que pagan por el servicio de agua potable a camión cisterna u otro

$\rho$ : coeficiente de retardo autorregresivo para SLM

$\lambda$ : coeficiente de retardo autorregresivo para SEM

En la Tabla 14 se presenta una comparación entre la bondad de ajuste entre los modelos, siendo los parámetros utilizados el coeficiente de determinación ( $R^2$ ) y el criterio de información de Akaike (AIC), evidenciando como resultado que para marzo el modelo de retraso espacial con  $R^2$  igual a 0,79 y un valor de AIC de 33,98 fue el mejor. Seguido por octubre con  $R^2$  y AIC de 0,71 y 101,56 respectivamente.

**Tabla 14: Coeficiente de determinación ( $R^2$ ) y criterio de información de Akaike (AIC) para los modelos de mínimos cuadrados ordinarios (OLS), retraso espacial (SLM) y error espacial (SEM)**

Mes	$R^2$			AIC		
	OLS	SLM	SEM	OLS	SLM	SEM
Marzo	0,64	0,79	-	55,03	33,98	-
Abril	0,65	-	-	109,17	-	-
Mayo	0,61	-	0,64	117,85	-	114,43
Junio	0,57	-	-	111,48	-	-
Julio	0,49	-	-	104,73	-	-
Agosto	0,43	-	0,60	118,92	-	106,99
Setiembre	0,57	-	-	114,39	-	-
Octubre	0,65	0,71	0,70	107,11	101,56	101,47
Noviembre	0,66	-	-	92,38	-	-

Al ser un análisis temporal de la evolución de la tasa de incidencia es posible observar una variabilidad en el ajuste del modelo ( $R^2$ ) a lo largo de los meses. Esto podría estar sujeto a otros factores como la dinámica del clima, relacionándose con mayor o menor movimiento de personas en las diferentes estaciones del año, la aglomeración de espacios públicos, las festividades, o cualquier otro agente que el modelo no ha tenido en cuenta.

El mejor ajuste es para octubre, coincidiendo con el periodo de implementación de la Fase 4 de la reanudación de actividades (D.S. N° 157-2020-PCM - 25/10/2020) a partir de la entrada en vigencia de las que destacan el comercio tiendas en general (con un aforo del 60 por ciento), restaurantes y servicios afines (aforo al 50 por ciento), el transporte aéreo internacional, servicios de arte, entretenimiento y esparcimiento.

A pesar de la aplicación de modelos que pudieron medir la autocorrelación espacial, es probable que las variables candidatas presenten una relación o efecto en la tasa de incidencia que pueda variar en el espacio, puesto que se identificó no solo una diferencia entre la parte baja y alta de la cuenca, sino también dentro de los distritos que la conforman. Por lo que se podría asumir que la dependencia espacial no sería el único efecto geográfico presente y se necesitan realizar modelos que puedan modelar este fenómeno.

## V. CONCLUSIONES

1. Los distritos de las provincias de Lima y Callao tienen una cobertura de más del 90 por ciento de agua por red pública mientras que en Huarochirí sólo un 70,8 por ciento. Además, la disponibilidad horaria ( $18 - 24 \text{ h.d}^{-1}$ ) de este servicio tiene un comportamiento parecido puesto que en la parte baja alcanza un 80 por ciento en contraste con menos del 50 por ciento para la parte alta. Respecto al servicio de alcantarillado por red pública se tiene los valores más altos en la parte baja de la cuenca, notándose que los distritos del Callao son los que poseen alrededor del 100 por ciento de cobertura, en contraposición a la provincia de Huarochirí que presenta coberturas de hasta sólo el 12 por ciento.
2. La relación entre el nivel de atención de los servicios de saneamiento y residuos sólidos se da en diferente medida tanto en la parte baja y media-alta de la cuenca. Tal es así, que en las provincias de Lima y Callao (parte baja) el nivel de cobertura de los servicios de saneamiento tiene valores altos, en cuanto al servicio de agua potable y alcantarillado, así también, es en estas dos provincias en las que se da la mayor generación de residuos por habitante. Presentando, además la mayor cantidad de casos positivos de COVID-19 así como valores altos de tasas de incidencia.  
En contraste con la provincia de Huarochirí (parte media y alta de la cuenca) en donde los servicios de saneamiento (agua potable y alcantarillado) presentan valores bajos de cobertura, además de estar manejados por la Municipalidad u organización comunal y no por una empresa prestadora de servicios de saneamiento, y se da la menor generación de residuos sólidos por habitante, además de presentar el menor porcentaje de casos positivos de toda la zona de estudio así como los valores más bajos de tasa de incidencia.
3. Los casos positivos por provincia durante los meses de estudio (primera ola de la pandemia), presentan dos grandes picos de contagios ocurridos en mayo y agosto del 2020. Adicionalmente, se evidencia la enorme diferencia que existe en el nivel de contagio entre las tres provincias de estudio, puesto que, del total de casos positivos



confirmados, Lima, Callao y Huarochirí poseen el 91, 8 y 1 por ciento respectivamente. Además, que para la distribución de los casos positivos por sexo en toda la cuenca de estudio se tuvo que más del 52 por ciento de los contagiados fueron hombres.

4. De abril a octubre, la tasa de incidencia de COVID-19 se ajustó a un modelo de mínimos cuadrados ordinarios, a partir de las variables altitud y el porcentaje de cobertura de agua por camión cisterna u otros, con un  $R^2$  que varió entre 0,43 y 0,65. Teniendo la variable altitud una influencia negativa, pues a medida que esta aumenta en valor la tasa de incidencia disminuye; por el contrario, la otra variable tiene una influencia positiva, es así, a mayor porcentaje de viviendas abastecidas de agua por camión cisterna u otro, mayor valor de la tasa de incidencia e impacto.
5. En relación al impacto en la tasa de incidencia de las variables modeladas, se tiene que la variable cobertura de agua por camión cisterna u otro impacta de manera conjunta con la variable altitud, aunque no de manera uniforme en toda la cuenca, existiendo tres zonas claramente diferenciadas (zona baja, media y alta). Mientras que la parte baja posee un impacto entre el rango de “medio” y “muy alto”, la parte media posee impactos variables de “bajo” a “muy alto” y la parte alta entre “bajo” y “medio”.
6. El fenómeno de autocorrelación espacial o dependencia espacial (relación entre los datos de los distritos cercanos) puede darse por la variable dependiente (tasa de incidencia). Tal es el caso de marzo y octubre, cuya aplicación del modelo de retraso espacial corrige este fenómeno al incluir en su desarrollo el efecto de la tasa de incidencia de los distritos vecinos, lo que da como resultado la combinación de variables de altitud y porcentaje de cobertura de agua por camión cisterna u otros con un  $R^2$  de 0,64 y 0,71 respectivamente. Por otro lado, el mencionado fenómeno puede también darse por alguna variable independiente no considerada en el modelo, lo que se traduce en que la dependencia espacial se ve reflejada en el término del error. Por ello, en mayo y agosto la tasa de incidencia de COVID-19 se ajustó al modelo de error espacial, con  $R^2$  de 0,64 y 0,60 respectivamente.

## **VI. RECOMENDACIONES**

1. Obtener información de campo referido a calidad del agua potable en las viviendas, población con comorbilidades, población migrante, flujo de personas, etc., mediante el muestreo de agua, uso de encuestas, visitas, llamadas telefónicas, entre otros, información que podría influir en la propagación del COVID-19 y sus demás variantes.
2. El nivel de acceso a un sistema continuo de abastecimiento de agua potable o la falta de este, es un indicador indirecto del nivel socioeconómico de la población, en tal sentido, esta variable se debería tener en consideración para futuras investigaciones para describir el comportamiento de la pandemia.
3. Evaluar el impacto del número de personas vacunadas en la tasa de incidencia y mortalidad del COVID-19 y sus variantes como variable explicativa.
4. Mejorar la cobertura del sistema de abastecimiento de agua, variable que influye directamente en la tasa de incidencia de COVID-19, sobre todo a la población de los distritos de San Juan de Lurigancho y San Antonio que tienen más del 10 por ciento de cobertura de agua por medio de camiones cisterna u otro.
5. Evaluar la tasa de incidencia y mortalidad por COVID-19 y sus variantes empleando nuevas herramientas informáticas para el procesamiento de datos como big data, machine learning, modelos multiescala, redes neuronales, random forest, entre otros.
6. Cuantificar el impacto de la variación de las variables cobertura de agua por camión cisterna u otro así como altitud en la tasa de incidencia, teniendo en cuenta escenarios críticos.

## VII. BIBLIOGRAFÍA

- Aabed, K. & Lashin, M.M.A. (2021). An analytical study of the factors that influence COVID-19 spread. *Saudi Journal of Biological Sciences*, 28(2), 1177–1195. <https://doi.org/10.1016/j.sjbs.2020.11.067>
- Accinelli, R.A. & Leon-Abarca, J.A. (2020). En la altura la COVID-19 es menos frecuente: la experiencia del Perú. *Archivos De Bronconeumologia*, 56(11), 760. <https://doi.org/10.1016/J.ARBRES.2020.06.015>
- Anderson, D., Sweeney, D. & Williams, T. (2012). *Estadística para negocios y economía*.
- Anselin, L. (2007). Spatial econometrics in RSUE: Retrospect and prospect. *Regional Science and Urban Economics*, 37(4), 450–456. <https://doi.org/10.1016/j.regsciurbeco.2006.11.009>
- Anselin, L., Syabri, I. & Kho, Y. (2006). *GeoDa: An Introduction to Spatial Data Analysis*. 38, 5–22.
- AQUAFONDO, Cooperación Alemana para el Desarrollo - GIZ, IPROGA, I. de P. para la G. del A.- & Zegarra, E. (2016). Estudio de riesgos hídricos y vulnerabilidad del sector privado en Lima Metropolitana y Callao en un contexto de cambio climático. In *Journal of Materials Processing Technology* (Vol. 1, Issue 1). <http://dx.doi.org/10.1016/j.cirp.2016.06.001><http://dx.doi.org/10.1016/j.powtec.2016.12.055><https://doi.org/10.1016/j.ijfatigue.2019.02.006><https://doi.org/10.1016/j.matlet.2019.04.024><https://doi.org/10.1016/j.matlet.2019.127252><http://dx.doi.org/10.1016/j.matlet.2019.127252>
- Aragón-Nogales, R., Vargas-Almanza, I. & Miranda-Novales, M.G. (2019). COVID-19 por SARS-CoV-2: la nueva emergencia de salud. *Rev Mex Pediatr*, 86(6), 213–218. <https://doi.org/10.35366/91871>
- Arias-Reyes, C., Zubieta-DeUrioste, N., Poma-Machicao, L., Aliaga-Raduan, F., Carvajal-Rodriguez, F., Dutschmann, M., Schneider-Gasser, E.M., Zubieta-Calleja, G. & Soliz, J. (2020). Does the pathogenesis of SARS-CoV-2 virus decrease at high-altitude? *Respiratory Physiology and Neurobiology*, 277. <https://doi.org/10.1016/J.RESP.2020.103443>

- Bacchini, R.D., Vázquez, L.V. & García Fronti, J.I. (2010). Introducción a la probabilidad y estadística. In W. Mendenhall, R. J. Beaver, & B. M. Beaver (Eds.), & J. H. Romo Muñoz & A. E. García Hernández (Trans.), *Cengage Learning* (13th ed.).
- Cabezas, C. (2018). Enfermedades infecciosas relacionadas con el agua en el Perú. *Revista Peruana de Medicina Experimental y Salud Pública*, 35(2), 309–316. <https://doi.org/10.17843/rpmesp.2018.352.3761.309>
- Castro, R. & Perez, R. (2009). *Saneamiento rural y salud: Guía para acciones a nivel local*. <https://iris.paho.org/handle/10665.2/52823>
- Corburn, J., Vlahov, D., Mberu, B., Riley, L., Caiaffa, W.T., Rashid, S.F., Ko, A., Patel, S., Jukur, S., Martínez-Herrera, E., Jayasinghe, S., Agarwal, S., Nguendo-Yongsi, B., Weru, J., Ouma, S., Edmundo, K., Oni, T. & Ayad, H. (2020). Slum Health: Arresting COVID-19 and Improving Well-Being in Urban Informal Settlements. *Journal of Urban Health* 2020 97:3, 97(3), 348–357. <https://doi.org/10.1007/S11524-020-00438-6>
- Eichelberger, L., Dev, S., Howe, T., Barnes, D.L., Bortz, E., Briggs, B.R., Cochran, P., Dotson, A.D., Drown, D.M., Hahn, M.B., Mattos, K. & Aggarwal, S. (2021). Implications of inadequate water and sanitation infrastructure for community spread of COVID-19 in remote Alaskan communities. *Science of the Total Environment*, 776, 145842. <https://doi.org/10.1016/j.scitotenv.2021.145842>
- Ekumah, B., Armah, F.A., Yawson, D.O., Quansah, R., Nyieku, F.E., Owusu, S.A., Odoi, J. O., & Afitiri, A. R. (2020). Disparate on-site access to water, sanitation, and food storage heighten the risk of COVID-19 spread in Sub-Saharan Africa. *Environmental Research*, 189, 109936. <https://doi.org/10.1016/j.envres.2020.109936>
- Fischer, M.M., & Wang, J. (2011). Spatial Data Analysis: Models, Methods and Techniques. *Springer Briefs in Regional Science*, 37(1), 82pp. <https://doi.org/10.1007/978-3-642-21720-3>
- Fundación Futuro Latinoamericano (FFLA). (2015). *Proceso de conformación del Consejo de Recursos Hídricos de la Cuenca Interrregional Chillón-Rímac y Lurín, Perú. Un experiencia de gobernanza*. [www.ffla.net](http://www.ffla.net)
- Gormley, M., Aspray, T.J., & Kelly, D.A. (2020). COVID-19: mitigating transmission via wastewater plumbing systems. *The Lancet Global Health*, 8(5), e643. [https://doi.org/10.1016/S2214-109X\(20\)30112-1](https://doi.org/10.1016/S2214-109X(20)30112-1)
- Guillen, A., Lugo Serrato, O. & Aguilar Garnica, J. (2014). Correlación No-Paramétrica y su Aplicación en la Investigaciones Científica Non-Parametric Correlation and Its

- Application in Scientific Research. In *Daena: International Journal of Good Conscience* (Vol. 9, Issue 2).
- Gujarati, D.N. & Porter, D.C. (2010). *Econometría* (J. Mares Chacón & E. C. Zuñiga Gutiérrez (Eds.); P. Carril Villarreal (Trans.); 5th ed.). McGraw-Hill.
- Gwenzi, W. (2021). Leaving no stone unturned in light of the COVID-19 faecal-oral hypothesis? A water, sanitation and hygiene (WASH) perspective targeting low-income countries. *The Science of the Total Environment*, 753, 141751. <https://doi.org/10.1016/J.SCITOTENV.2020.141751>
- Herrera, M., Mur, J. & Ruíz, M. (2010). *¿Cuál matriz de pesos espaciales?. Un enfoque sobre selección de modelos.* [https://www.researchgate.net/publication/277211219\\_Cual\\_matriz\\_de\\_pesos\\_espaciales\\_Un\\_enfoque\\_sobre\\_seleccion\\_de\\_modelos](https://www.researchgate.net/publication/277211219_Cual_matriz_de_pesos_espaciales_Un_enfoque_sobre_seleccion_de_modelos)
- Jabłońska, K., Aballéa, S. & Toumi, M. (2021). Factors influencing the COVID-19 daily deaths peak across European countries. *Public Health*, 194, 135–142. <https://doi.org/10.1016/j.puhe.2021.02.037>
- Jin, J.M., Bai, P., He, W., Wu, F., Liu, X.F., Han, D.M., Liu, S. & Yang, J.K. (2020). Gender differences in patients with COVID-19: Focus on severity and mortality. *Frontiers in Public Health*, 8, 152. <https://doi.org/10.3389/FPUBH.2020.00152/PDF>
- Kampf, G., Todt, D., Pfaender, S. & Steinmann, E. (2020). Persistence of coronaviruses on inanimate surfaces and their inactivation with biocidal agents. In *Journal of Hospital Infection* (Vol. 104, Issue 3, pp. 246–251). W.B. Saunders Ltd. <https://doi.org/10.1016/j.jhin.2020.01.022>
- Klemeš, J.J., Fan, Y., Van, Tan, R.R. & Jiang, P. (2020). Minimising the present and future plastic waste, energy and environmental footprints related to COVID-19. *Renewable and Sustainable Energy Reviews*, 127, 109883. <https://doi.org/10.1016/J.RSER.2020.109883>
- León, G. (2020). *Tecnologías estratégicas relacionadas con COVID-19: Impacto y dependencia para España.*
- López, J.F. (2017). *Coeficiente de determinación (R cuadrado) - Qué es, definición y concepto | Economipedia.* <https://economipedia.com/definiciones/r-cuadrado-coeficiente-determinacion.html>
- Martínez, D.R., Albín, J.L., Cabaleiro, J.C., Pena, T.F., Rivera, F.F. & Blanco, V. (2009). *El criterio de información de Akaike en la obtención de modelos estadísticos de rendimiento.* <https://www.researchgate.net/publication/236279245>

- Martínez, E. (2005). Errores frecuentes en la interpretación del coeficiente de determinación lineal. *Anuario Jurídico y Económico Escurialense*, 38, 315–331.
- Mendoza, M.Á. (2016). *Econometría aplicada utilizando R*. 446.
- Ministerio de Sanidad. (2020). *Plan de respuesta temprana en un escenario de control de la pandemia por covid-19*.  
<https://www.google.com/search?q=frecuente+sinonimo&oq=frecuente+&aqs=chrome.69i57j0i512l8j0i10i512.2624j0j7&sourceid=chrome&ie=UTF-8>
- Mollalo, A., Vahedi, B. & Rivera, K.M. (2020). GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. *Science of the Total Environment*, 728, 138884. <https://doi.org/10.1016/j.scitotenv.2020.138884>
- Mondragon, M. (2000). Uso de la correlación de Spearman en un estudio de intervención en fisioterapia. *Movimiento Científico, ISSN-e 2011-7191, Vol. 8, Nº. 1, 2014, Págs. 98-104*, 8(1), 98–104. <https://dialnet.unirioja.es/servlet/articulo?codigo=5156978>
- Mukherjee, S. & Pahan, K. (2021). Is COVID-19 gender-sensitive? *Journal of Neuroimmune Pharmacology*, 16(1), 1. <https://doi.org/10.1007/S11481-020-09974-Z>
- Neal, M.J. (2020). COVID-19 and water resources management: reframing our priorities as a water sector. *Water International*, 45(5), 435–440. <https://doi.org/10.1080/02508060.2020.1773648>
- Nzediegwu, C. & Chang, S.X. (2020). Improper solid waste management increases potential for COVID-19 spread in developing countries. *Resources, Conservation and Recycling*, 161, 104947. <https://doi.org/10.1016/j.resconrec.2020.104947>
- Observatorio del Agua Chillón Rímac Lurín. (2019). *Diagnóstico Inicial para el Plan de Gestión de Recursos Hídricos de las cuencas Chillón, Rímac, Lurín y Chilca*.
- OMS. (2020). *Preguntas y respuestas sobre la enfermedad por coronavirus (COVID-19)*. <https://www.who.int/es/emergencias/diseases/novel-coronavirus-2019/advice-for-public/q-a-coronaviruses>
- Organización Mundial de la salud. (2020). *Alocución de apertura del Director General de la OMS en la rueda de prensa sobre la COVID-19 celebrada el 11 de marzo de 2020*. <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>
- Pan, A., Liu, L., Wang, C., Guo, H., Hao, X., Wang, Q., Huang, J., He, N., Yu, H., Lin, X., Wei, S. & Wu, T. (2020). Association of public health interventions with the epidemiology of the COVID-19 outbreak in Wuhan, China. *JAMA*, 323(19), 1. <https://doi.org/10.1001/JAMA.2020.6130>

- Paraskevis, D., Kostaki, E.G., Alygizakis, Nikiforos, Thomaidis, N.S., Cartalis, C., Tsiodras, S. & Dimopoulos, M.A. (2021). A review of the impact of weather and climate variables to COVID-19: In the absence of public health measures high temperatures cannot probably mitigate outbreaks. In *Science of the Total Environment* (Vol. 768, p. 144578). Elsevier B.V. <https://doi.org/10.1016/j.scitotenv.2020.144578>
- Prüss-Ustün, A., Wolf, J., Bartram, J., Clasen, T., Cumming, O., Freeman, M.C., Gordon, B., Hunter, P.R., Medlicott, K. & Johnston, R. (2019). Burden of disease from inadequate water, sanitation and hygiene for selected adverse health outcomes: An updated analysis with a focus on low- and middle-income countries. *International Journal of Hygiene and Environmental Health*, 222(5), 765–777. <https://doi.org/10.1016/j.ijheh.2019.05.004>
- Rahimi, N.R., Fouladi-Fard, R., Aali, R., Shahryari, A., Rezaali, M., Ghafouri, Y., Ghalhari, M.R., Asadi-Ghalhari, M., Farzinnia, B., Gea, O.C. & Fiore, M. (2021). Bidirectional association between COVID-19 and the environment: A systematic review. In *Environmental Research* (Vol. 194). Elsevier. <https://doi.org/10.1016/J.ENVRES.2020.110692>
- Rivera-Córdova, A. (2021). Relación densidad y pobreza con casos confirmados por coronavirus en la región metropolitana. *Revista Médica Clínica Las Condes*, 32(1), 81–89. <https://doi.org/10.1016/j.rmclc.2020.07.006>
- Rocha da Silva, R., Bezerra dos Santos, M., Dantas dos Santos, A., dos Santos Tavares, D. & Lima dos Santos, P. (2020). Coronavirus disease and basic sanitation: too early to be worried? *Revista Da Sociedade Brasileira de Medicina Tropical*, 53, 1–5. <https://doi.org/10.1590/0037-8682-0345-2020>
- Sánchez, P.A. (2012). *Una nueva metodología de entrenamiento de redes neuronales y sus implicaciones en la selección de modelos*. <http://bdigital.unal.edu.co/6843/1/43632277.2012.pdf%0Ahttp://www.bdigital.unal.edu.co/6843/>
- Sannigrahi, S., Pilla, F., Basu, B., Basu, A.S., & Molter, A. (2020). Examining the association between socio-demographic composition and COVID-19 fatalities in the European region using spatial regression approach. *Sustainable Cities and Society*, 62, 102418. <https://doi.org/10.1016/j.scs.2020.102418>
- Santabárbara, J. (2019). Cálculo del intervalo de confianza para los coeficientes de correlación mediante sintaxis en SPSS. *REIRE Revista d Innovació i Recerca En*

- Educació*, 12 (2), 1–14. <https://doi.org/10.1344/reire2019.12.228245>
- Saravia Matus, S., Gil, M., Blanco, E., Llavona, A. & Naranjo, L. (2020). Desafíos hídricos en Chile y recomendaciones para el cumplimiento del ODS 6 en América Latina y el Caribe. *198*, 77.
- Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI). (2014). *Estudio de indicadores de alteración hidrológica del río Rímac*.
- Silva, R.R., Ribeiro, C.J.N., Moura, T.R., Santos, M.B., Santos, A.D., Tavares, D.S. & Santos, P.L. (2021). Basic sanitation: a new indicator for the spread of COVID-19? *Transactions of The Royal Society of Tropical Medicine and Hygiene*, 0, 1–9. <https://doi.org/10.1093/trstmh/traa187>
- Tumialán De la Cruz, P. (2016). Rasgos geológicos del río Rimác, Abastecimiento de agua en su cono de deyección. *Revista de Ciencias*, 11, 19–26. [https://doi.org/10.31381/revista\\_ciencias.v11i0.575](https://doi.org/10.31381/revista_ciencias.v11i0.575)
- UNICEF. (2020). *UNICEF WASH Programme contribution to COVID- 19 prevention and response*. <https://www.who.int/docs/default-source/coronaviruse/covid-19-sprp-unct->
- Velasco, J.M., Tseng, W.C. & Chang, C.L. (2021). Factors Affecting the Cases and Deaths of COVID-19 Victims. *International Journal of Environmental Research and Public Health*, 18(2), 1–10. <https://doi.org/10.3390/IJERPH18020674>
- Yojcon, D. & Ruiz, J.C. (2020). Los factores sociales asociados a los modelos matemáticos en un contexto de pandemia. *Revista Latinoamericana de Etnomatemática Perspectivas Socioculturales de La Educación Matemática*, 13(1), 355–374. <https://doi.org/10.22267/relatem.20131.47>
- Yrigoyen, C.C. (2003). *Econometría espacial aplicada a la Predicción-Extrapolación De Datos*. <http://www.madrid.org/bvirtual/BVCM005618.pdf>



## **VIII. ANEXOS**

## ANEXO 1: ACCESO A LOS SERVICIOS DE SANEAMIENTO POR DISTRITO

**Tabla 15: Población y nivel de acceso a agua potable y alcantarillado por distrito en la cuenca Rímac**

Provincia / Distrito	Población (habitantes)	Hombres	Mujeres	Agua potable				Alcantarillado			
				Conexión a red pública (%)				Sin conexión (%)	Conexión a red pública (%)		Sin conexión (%)
				Todos los días			No todos los días		Dentro de la vivienda	Fuera de la vivienda	
				24 h	18 h	9 h					
<i>PROVINCIA DE HUAROCHIRÍ</i>											
Callahuanca	798	421	377	90.2	0.0	0.4	2.8	6.5	61.0	14.6	24.4
Carampoma	331	168	163	82.9	3.2	3.2	2.5	8.2	46.8	38.0	15.2
Chicla	3,826	2,565	1,261	72.5	3.0	0.9	1.7	21.9	19.2	59.3	21.5
Huachupampa	609	314	295	91.1	0.0	1.8	1.8	5.3	30.8	38.5	30.8
Huanza	875	512	363	39.1	41.9	15.1	1.1	2.8	37.0	19.0	44.0
Laraos	666	338	328	90.9	0.0	0.0	3.5	5.6	35.4	23.2	41.4
Matucana	4,058	2,053	2,005	22.7	29.9	17.1	2.3	28.0	56.3	10.8	32.9
Ricardo Palma	6,542	3,266	3,276	20.2	1.5	19.3	48.8	10.3	70.3	11.0	18.7
San Andrés de Tupicocha	1,320	683	637	57.0	3.4	11.3	4.8	23.5	45.9	4.3	49.8
San Antonio	912	606	306	20.0	1.6	1.2	1.8	75.4	0.2	0.8	99.0
San Bartolomé	1,139	581	558	40.4	0.7	40.4	3.5	14.9	49.4	16.4	34.2
San Juan de Iris	699	339	360	64.0	2.9	0.7	2.2	30.2	33.8	20.9	45.3
San Mateo	4,245	2,145	2,100	52.4	15.1	13.3	6.7	12.5	47.8	34.5	17.7
San Mateo de Otao	1,335	696	639	9.8	0.7	4.8	2.1	82.7	7.1	5.0	87.9
San Pedro de Casta	928	462	466	50.5	10.7	3.2	21.5	14.2	41.3	29.0	29.7
Santa Eulalia	12,636	6,175	6,461	10.8	8.0	62.9	6.0	12.3	63.3	12.0	24.8
Santiago de tuna	411	222	189	8.9	12.3	34.6	33.0	11.2	40.2	16.8	43.0
Sta. Cruz de Cocachacra	2,486	1,303	1,183	67.1	18.9	5.8	5.3	2.9	80.8	5.8	13.4
Surco	1,407	734	673	55.6	9.3	0.7	3.3	31.1	52.2	7.3	40.4
<i>PROVINCIA DEL CALLAO</i>											
Bellavista	74,851	35,673	39,178	96.4	2.3	0.5	0.7	0.1	88.0	12.0	0.1
Callao	451,260	221,758	229,502	61.2	31.3	3.1	2.6	1.8	87.4	10.8	1.8
Carmen de La Legua	42,240	20,586	21,654	65.4	11.0	20.9	1.8	0.9	82.5	16.4	1.0
La Perla	61,417	29,085	32,332	96.6	1.0	0.9	1.4	0.1	88.6	11.3	0.0
La Punta	3,829	1,756	2,073	92.7	6.7	0.4	0.2	0.0	79.1	20.8	0.1

«Continuación»

*PROVINCIA DE LIMA*

Ate	599,196	293,019	306,177	75.5	6.5	3.1	5.9	9.0	76.7	9.9	13.4
Barranco	34,378	15,949	18,429	98.2	0.4	0.4	0.9	0.1	82.9	17.1	0.0
Breña	85,309	40,081	45,228	95.2	3.1	0.8	0.9	0.0	80.8	19.2	0.0
Chaclacayo	42,912	20,655	22,257	32.5	16.6	39.8	7.8	3.3	84.7	12.4	2.9
Chorrillos	314,241	152,926	161,315	87.2	2.4	2.4	3.5	4.4	85.6	11.5	2.9
El Agustino	198,862	97,512	101,350	81.7	3.0	11.9	2.6	0.8	89.2	10.2	0.6
Independencia	211,360	103,694	107,666	41.4	14.1	31.5	8.7	4.4	85.2	9.0	5.8
Jesús María	75,359	33,692	41,667	97.9	1.2	0.7	0.3	0.0	78.6	21.4	0.0
La Molina	140,679	65,682	74,997	95.0	1.1	1.4	1.9	0.7	88.7	10.5	0.8
La Victoria	173,630	85,108	88,522	80.8	6.2	10.7	2.1	0.2	81.8	18.2	0.1
Lima	268,352	129,630	138,722	91.7	4.6	1.9	1.6	0.3	83.2	16.6	0.2
Lince	54,711	25,085	29,626	98.0	1.0	0.7	0.3	0.0	80.9	19.1	0.0
Los Olivos	325,884	156,847	169,037	95.9	1.2	1.2	1.6	0.1	90.7	9.2	0.1
Lurigancho	240,814	118,926	121,888	23.4	2.4	14.8	15.7	43.7	41.3	8.6	50.1
Magdalena del mar											
Miraflores	99,337	45,152	54,185	99.0	0.3	0.4	0.3	0.0	79.3	20.7	0.0
Pueblo Libre	83,323	37,628	45,695	98.3	0.6	0.6	0.5	0.0	82.3	17.7	0.0
Rímac	174,785	85,138	89,647	69.6	7.0	16.4	5.0	2.1	83.8	13.8	2.4
San Borja	113,247	51,704	61,543	98.4	0.5	0.6	0.5	0.0	83.6	16.4	0.0
San Isidro	60,735	26,906	33,829	98.9	0.3	0.4	0.4	0.0	78.6	21.4	0.0
San Juan de Lurigancho	1,038,495	518,537	519,958	52.3	4.3	6.4	26.3	10.6	75.2	8.2	16.6
San Juan de Miraflores	355,219	173,683	181,536	87.4	1.0	1.4	2.8	7.3	84.6	6.3	9.1
San Luis	52,082	25,081	27,001	96.1	2.4	0.7	0.8	0.1	83.3	16.7	0.1
San Martín de Porres	654,083	317,786	336,297	91.7	2.1	1.3	2.1	2.7	88.0	9.1	2.9
San Miguel	155,384	72,499	82,885	98.5	0.3	0.5	0.6	0.1	80.9	19.1	0.0
Santa Anita	196,214	96,080	100,134	88.3	7.0	2.2	2.2	0.3	85.6	14.0	0.4
Santiago de Surco	329,152	152,312	176,840	98.0	0.5	0.6	0.6	0.3	82.4	17.3	0.3
Surquillo	91,023	42,439	48,584	97.9	0.4	1.0	0.6	0.0	80.8	19.2	0.0
Villa El Salvador	393,254	193,833	199,421	87.3	1.6	1.6	5.4	4.1	88.4	5.7	5.9
Villa María del Triunfo	398,433	195,222	203,211	73.5	2.2	2.2	6.9	15.2	74.7	6.0	19.2

**Tabla 16: Cobertura de agua potable por provincia en la cuenca Rímac**

Detalle	Provincia		
	<i>Huachirí</i>	<i>Callao</i>	<i>Lima</i>
Red pública (%)	79.0	99.4	96.3
<i>Dentro de la vivienda</i>	56.5	85.1	80.5
<i>Fuera de la vivienda</i>	22.5	14.3	15.8
No conectado a la red pública (%)	21.0	0.6	3.7

FUENTE: Elaborado a partir de la información del Censo Nacional 2017: XII de Población, VII Vivienda y III de Comunidades Indígenas, del Instituto Nacional de Estadística e Informática (INEI)

**Tabla 17: Cobertura por red pública de alcantarillado por provincia en la cuenca Rímac**

Cobertura	Provincia		
	<i>Huachirí</i>	<i>Lima</i>	<i>Callao</i>
Red pública (%)	62.4	95.5	99.4
<i>Dentro de la vivienda</i>	43.1	81.4	85.1
<i>Fuera de la vivienda</i>	19.3	14.1	14.3
No conectado a la red pública (%)	37.6	4.5	0.6

FUENTE: Elaborado a partir de la información del Censo Nacional 2017: XII de Población, VII Vivienda y III de Comunidades Indígenas, del Instituto Nacional de Estadística e Informática (INEI)

**Tabla 18: Porcentaje de composición de residuos sólidos por provincia en la cuenca Rímac**

Composición de los residuos (%)	Provincia		
	<i>Huachirí</i>	<i>Callao</i>	<i>Lima</i>
Orgánica	53.8	59.4	55.7
Inorgánica	17.8	18.5	21.2
No aprovechable	19.7	12.8	13.4
Peligroso	8.7	9.3	9.7

FUENTE: Elaborado a partir de la información del portal “Seguimiento e información sobre gestión de residuos sólidos” del Ministerio del Ambiente (MINAM)

## ANEXO 2: CARACTERIZACIÓN DE LAS VARIABLES SOCIODEMOGRÁFICAS

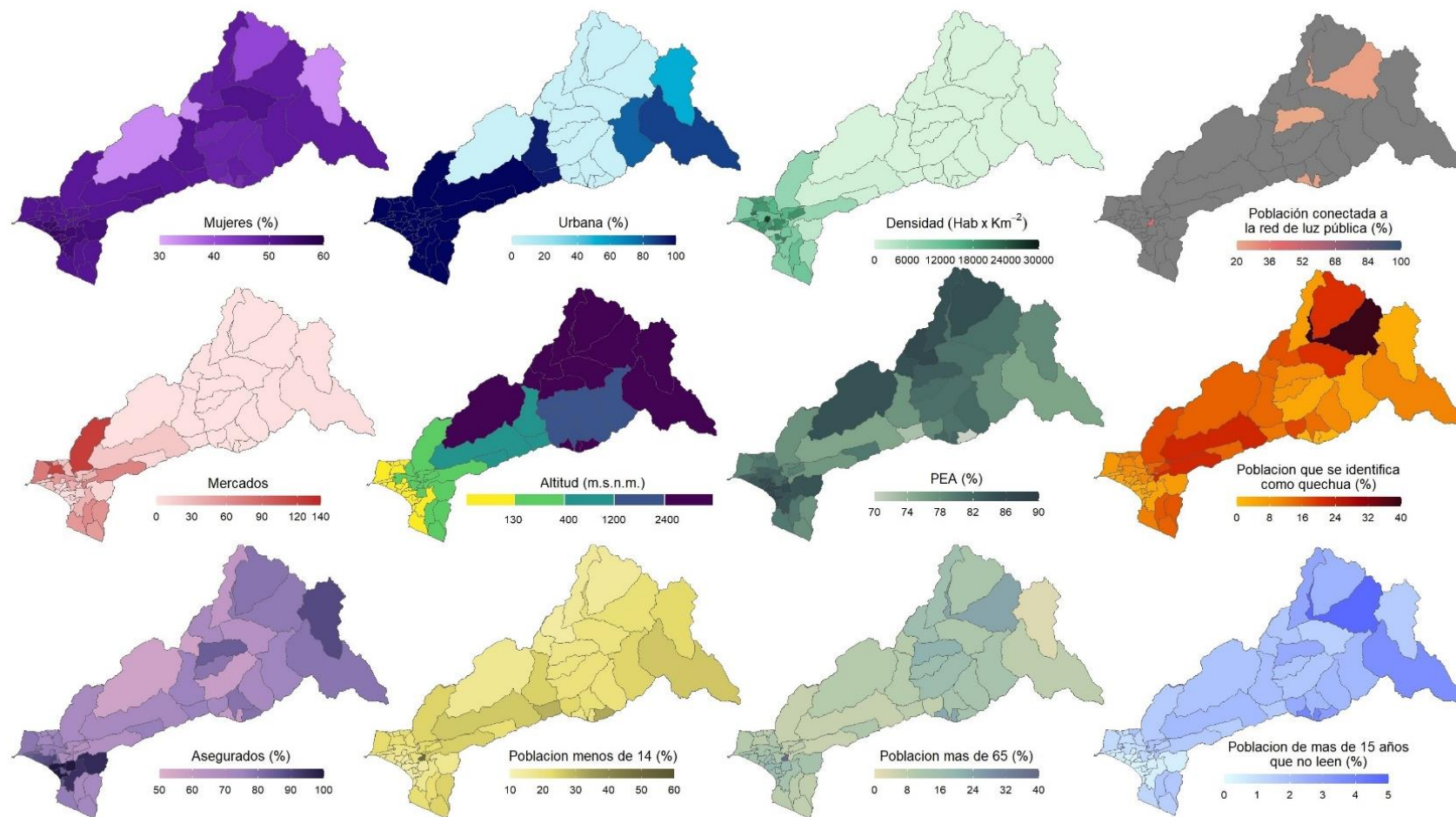
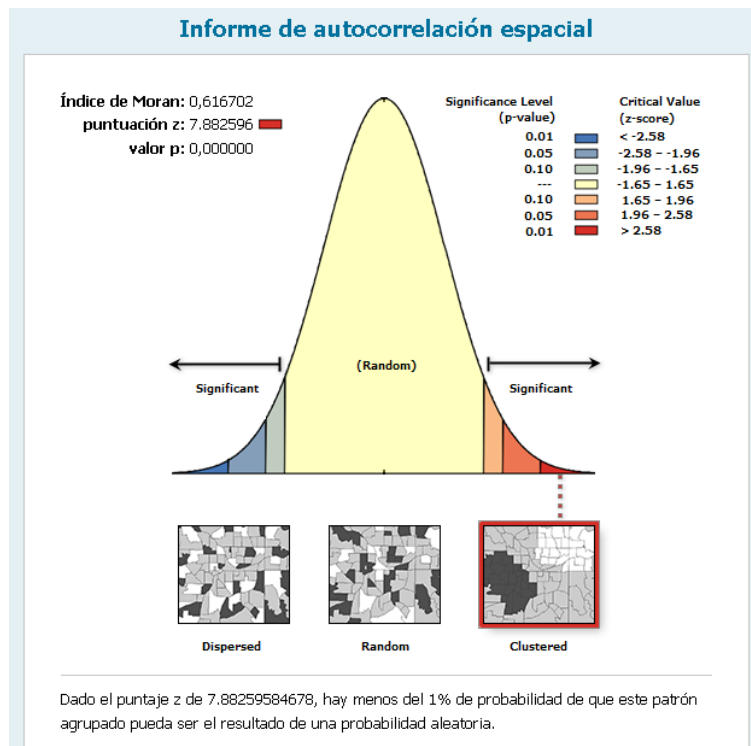
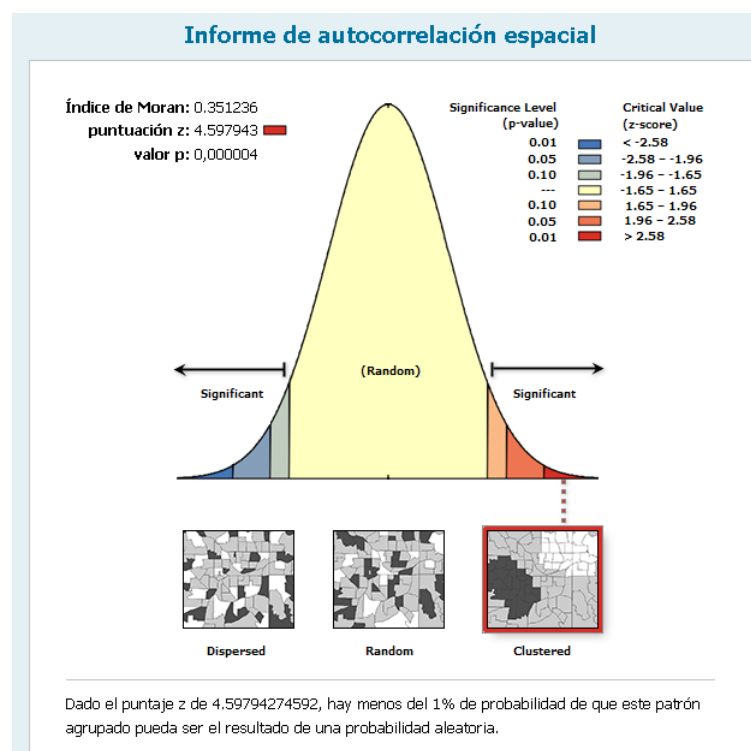


Figura 22: Distribución espacial de las variables sociodemográficas en la cuenca Rímac

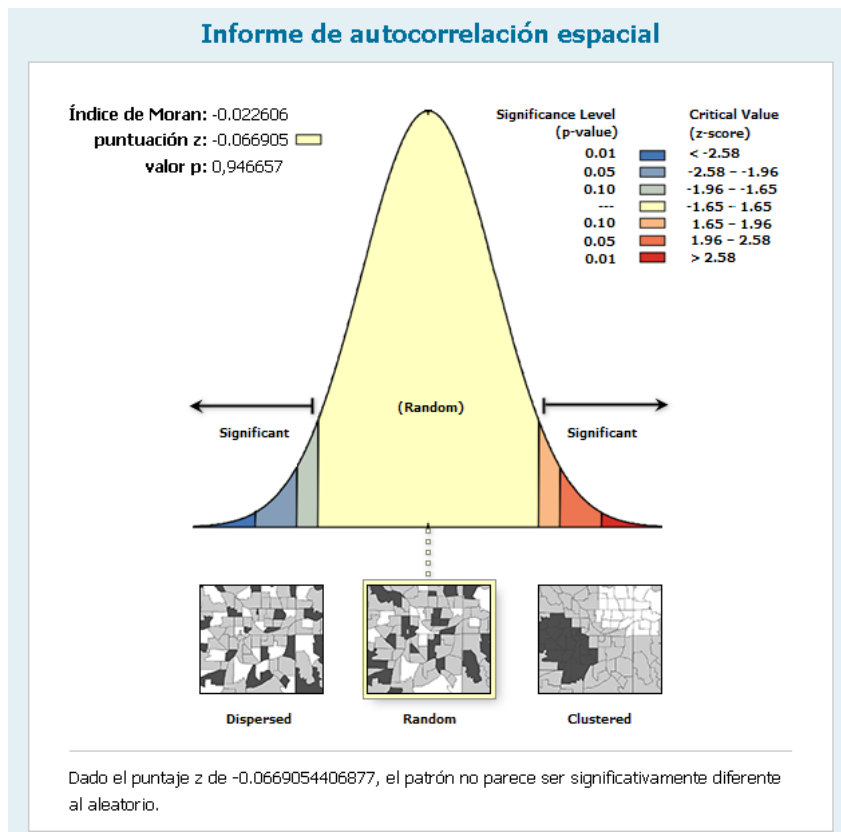
### ANEXO 3: TEST DE AUTOCORRELACIÓN DE LA VARIABLE TASA DE INCIDENCIA DE COVID-19



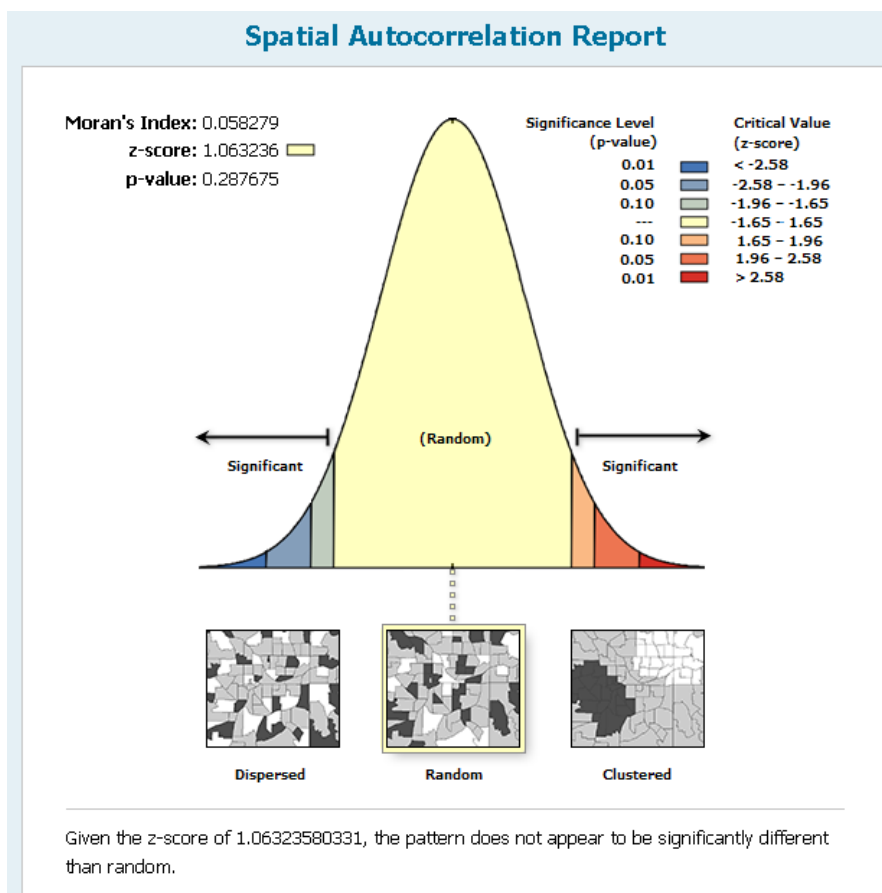
**Figura 23: Índice de Morán para la tasa de incidencia del mes de marzo**



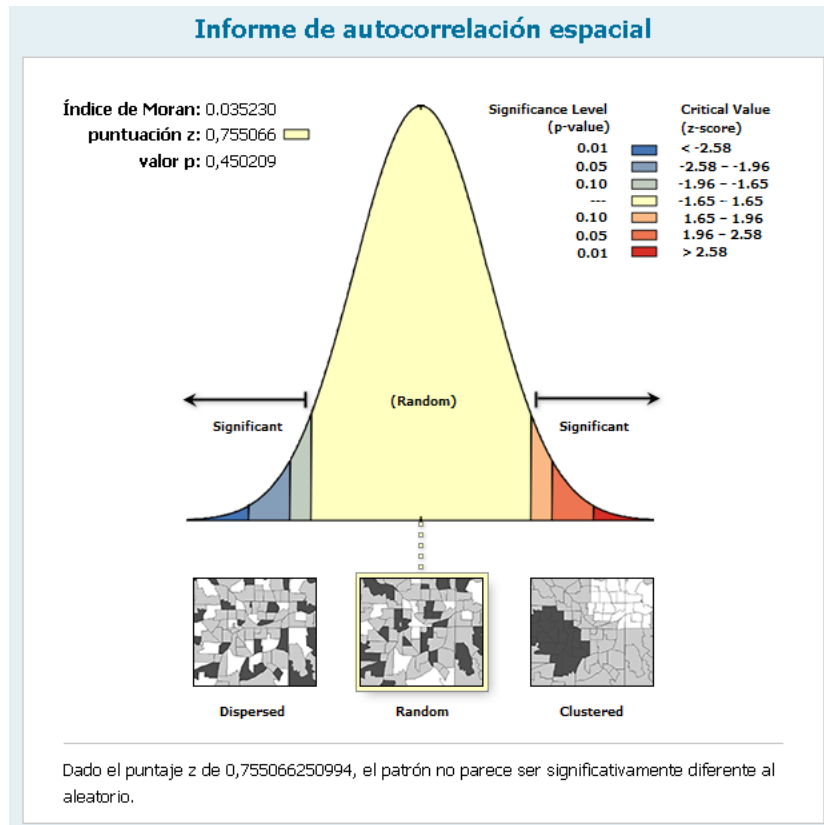
**Figura 24: Índice de Morán para la tasa de incidencia del mes de abril**



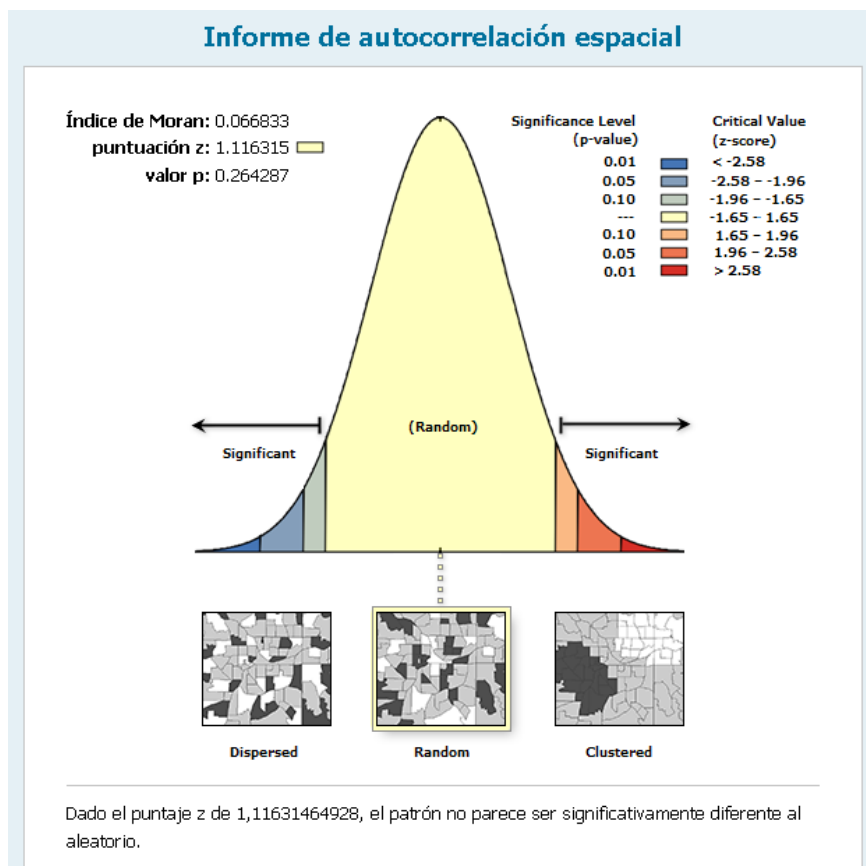
**Figura 25: Índice de Morán para la tasa de incidencia del mes de mayo**



**Figura 26: Índice de Morán para la tasa de incidencia del mes de junio**

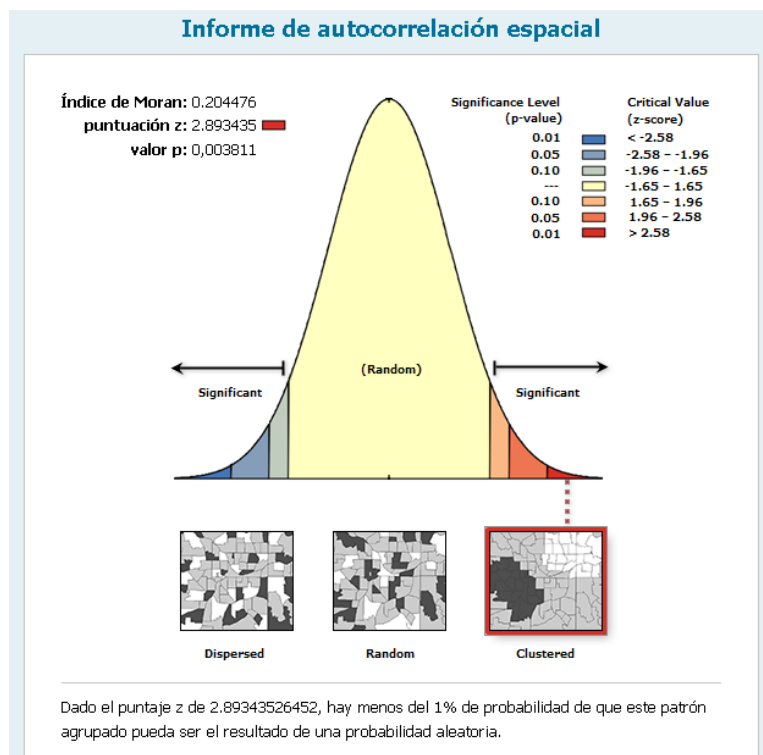


**Figura 27: Índice de Morán para la tasa de incidencia del mes de julio**

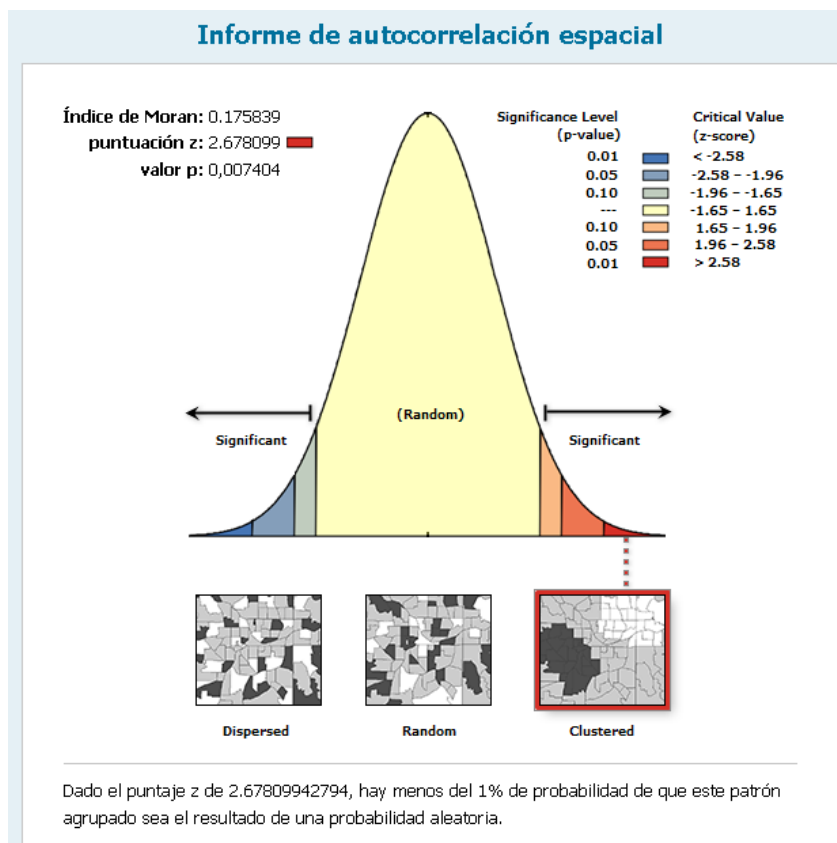


**Figura 28: Índice de Morán para la tasa de incidencia del mes de agosto**

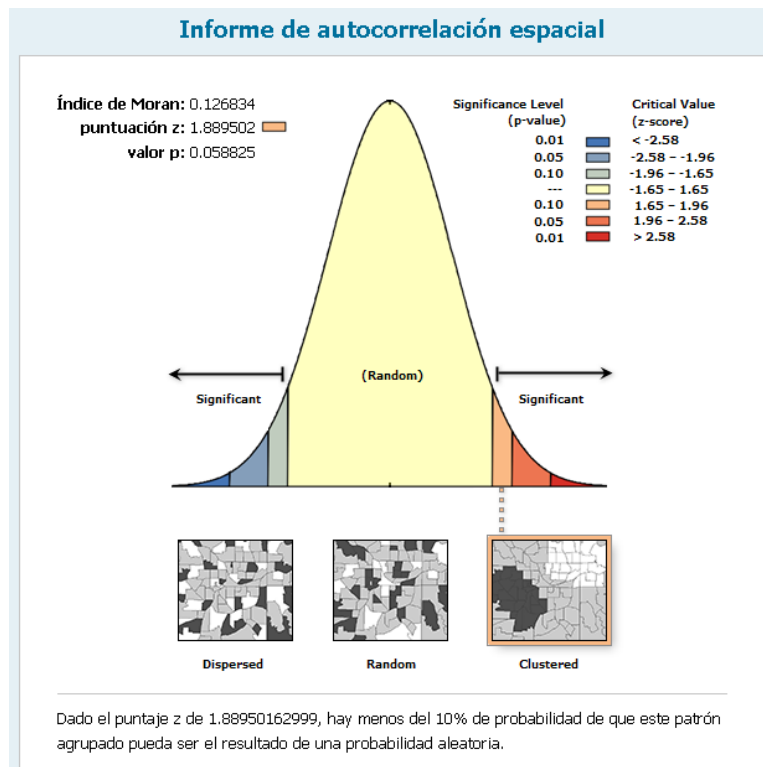




**Figura 29: Índice de Morán para la tasa de incidencia del mes de septiembre**

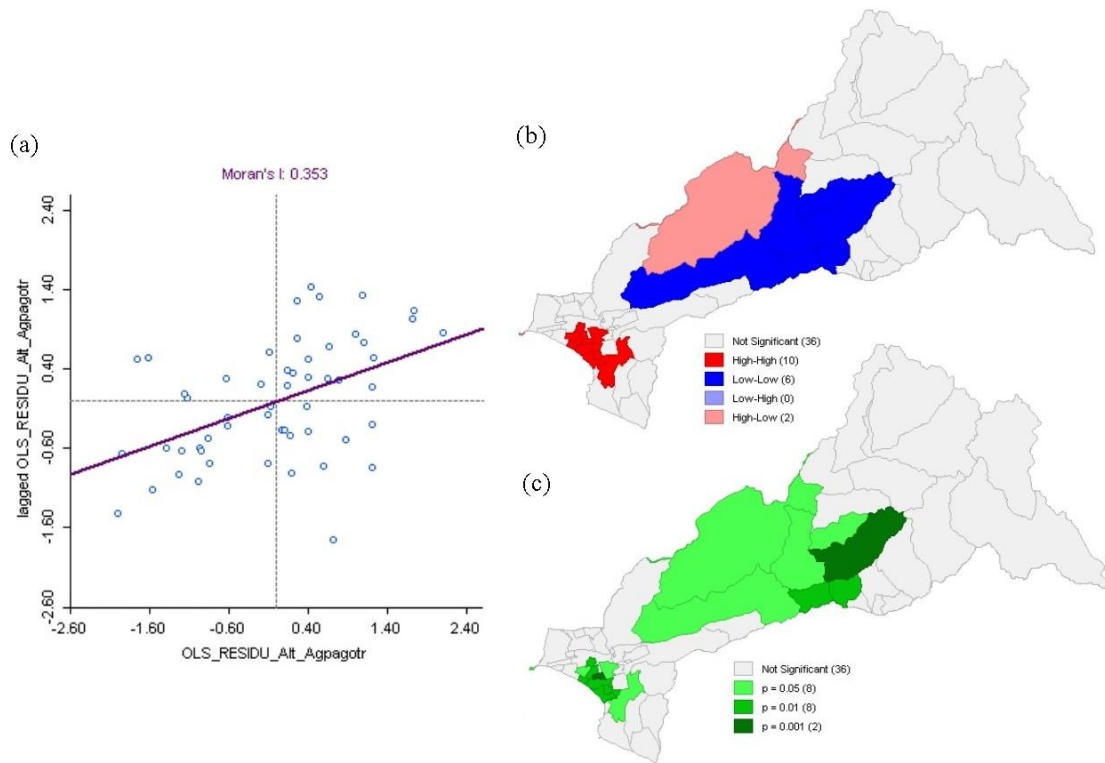


**Figura 30: Índice de Morán para la tasa de incidencia del mes de octubre**

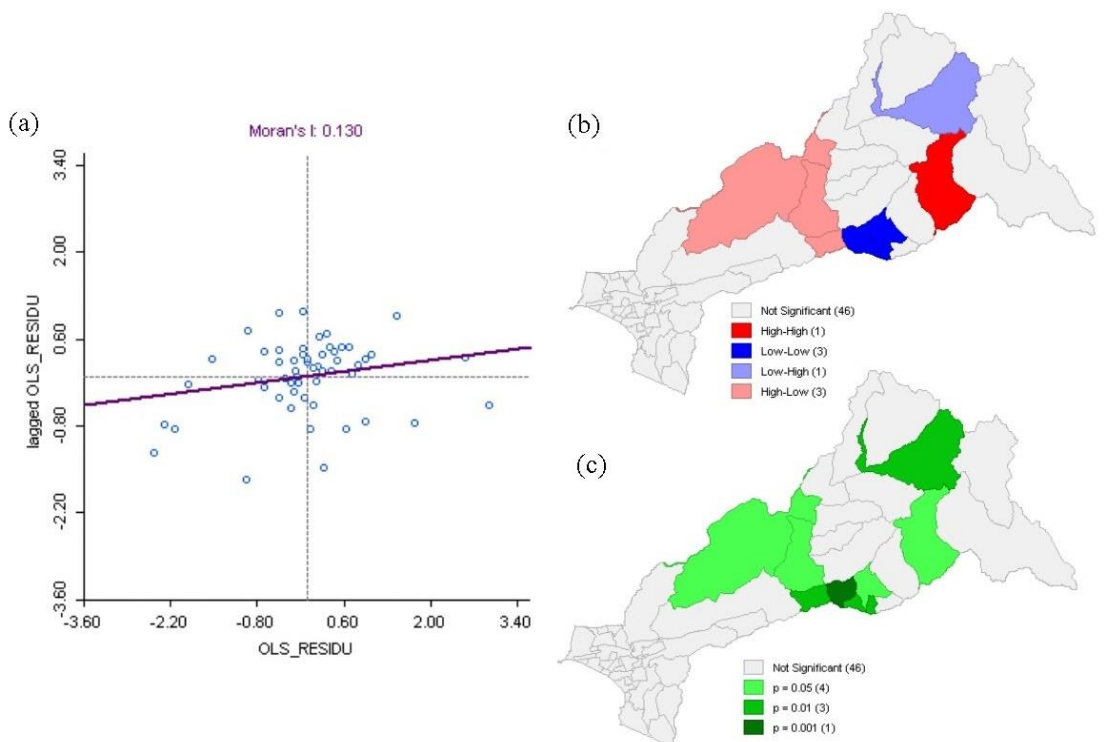


**Figura 31: Índice de Morán para la tasa de incidencia del mes de noviembre**

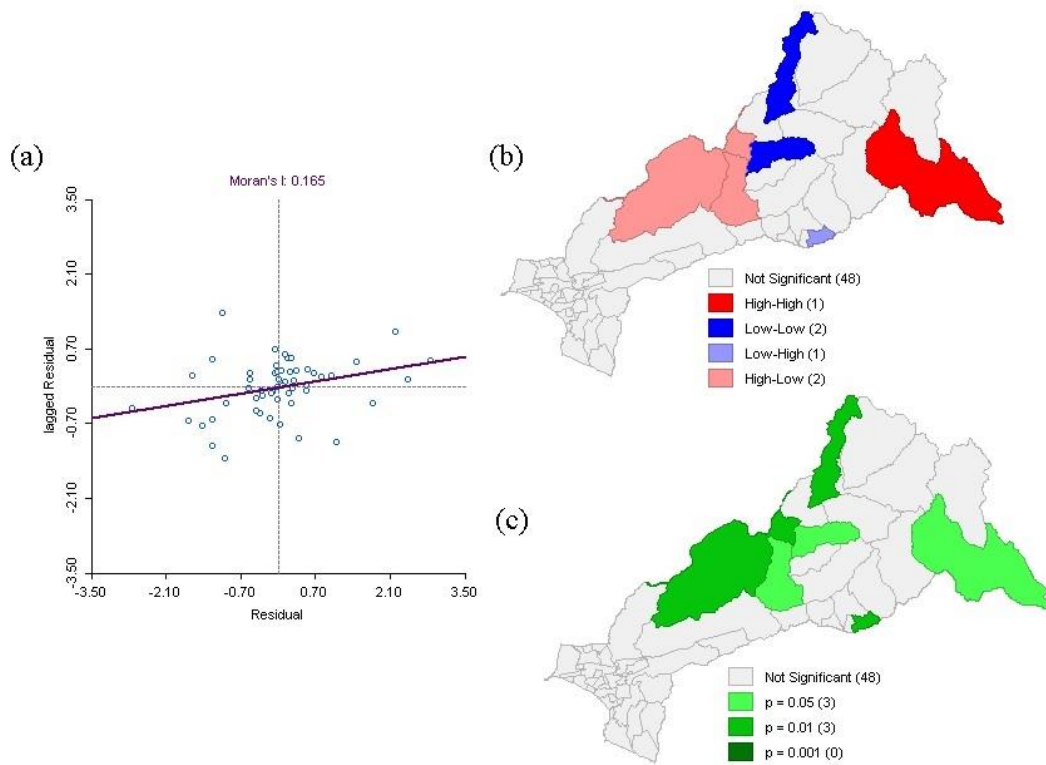
**ANEXO 4: TEST DE ÍNDICE DE MORÁN PARA LOS RESIDUOS DEL MODELO DE MÍNIMOS CUADRADOS ORDINARIOS**



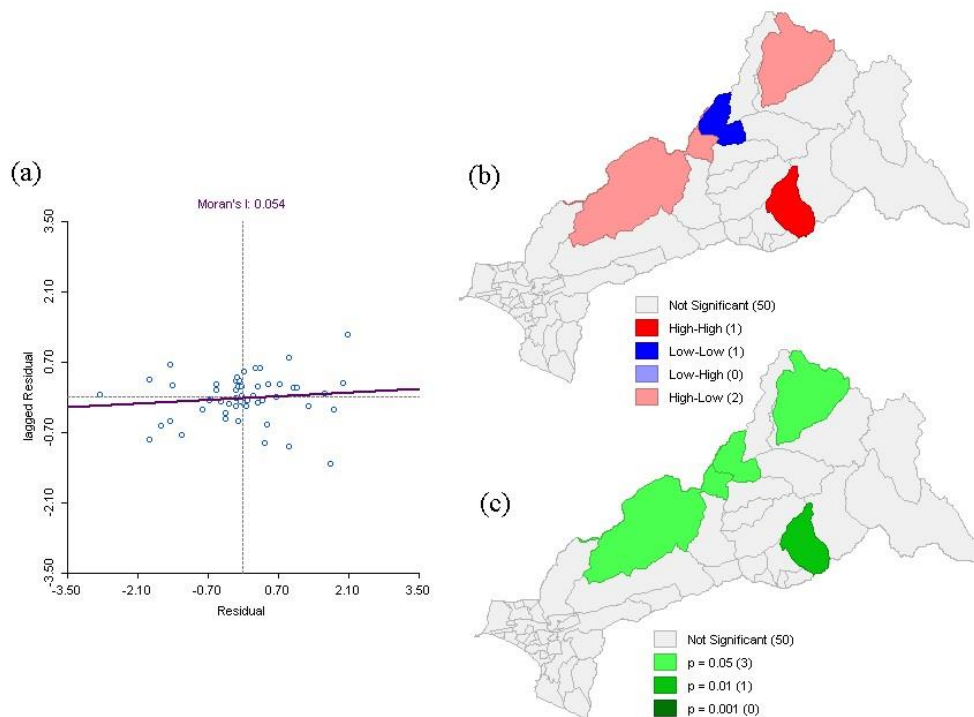
**Figura 32: Índice de Morán de los residuos del modelo para el mes de marzo: diagrama de dispersión (a), mapa de conglomerados (b) y mapa de significancia (c)**



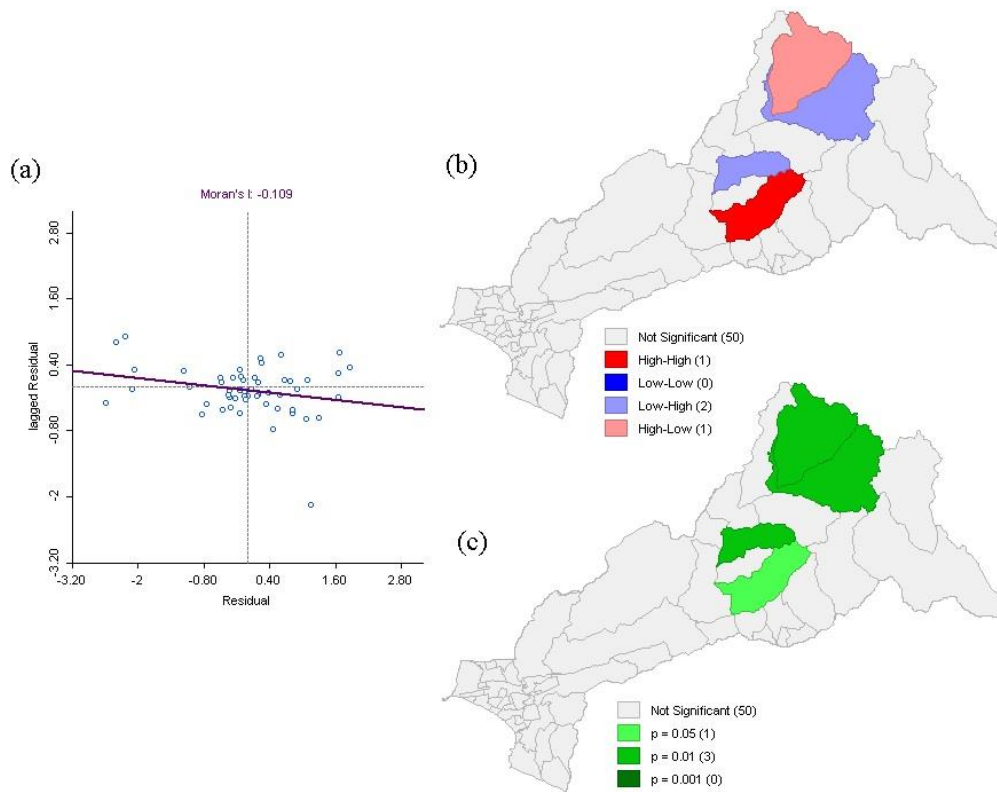
**Figura 33: Índice de Morán de los residuos del modelo para el mes de abril: diagrama de dispersión (a), mapa de conglomerados (b) y mapa de significancia (c)**



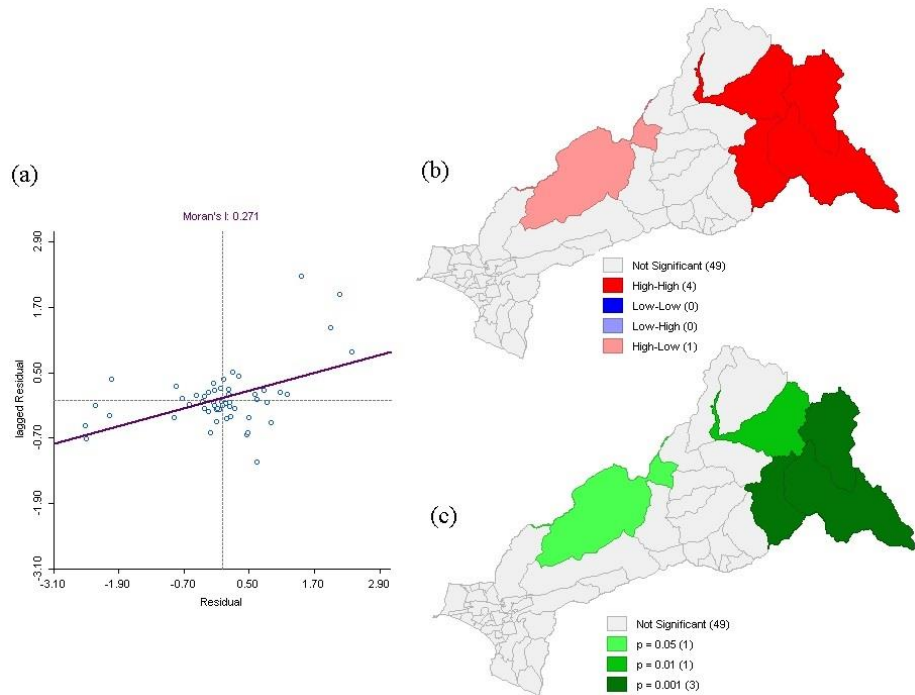
**Figura 34: Índice de Morán de los residuos del modelo para el mes de mayo: diagrama de dispersión (a), mapa de conglomerados (b) y mapa de significancia (c)**



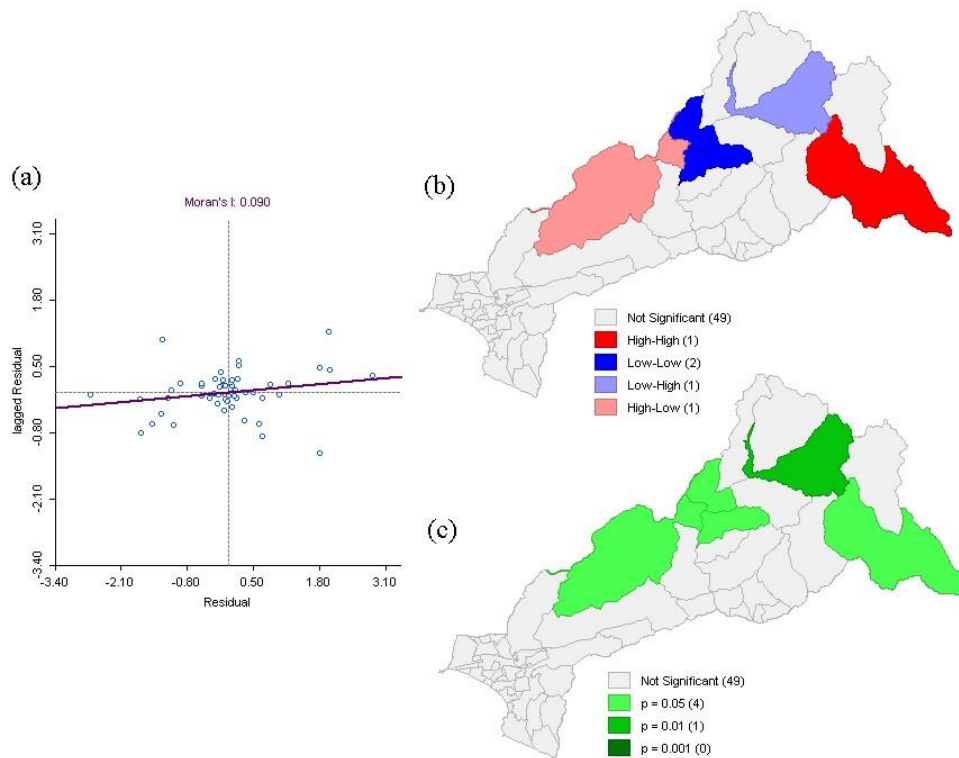
**Figura 35: Índice de Morán de los residuos del modelo para el mes de junio: diagrama de dispersión (a), mapa de conglomerados (b) y mapa de significancia (c)**



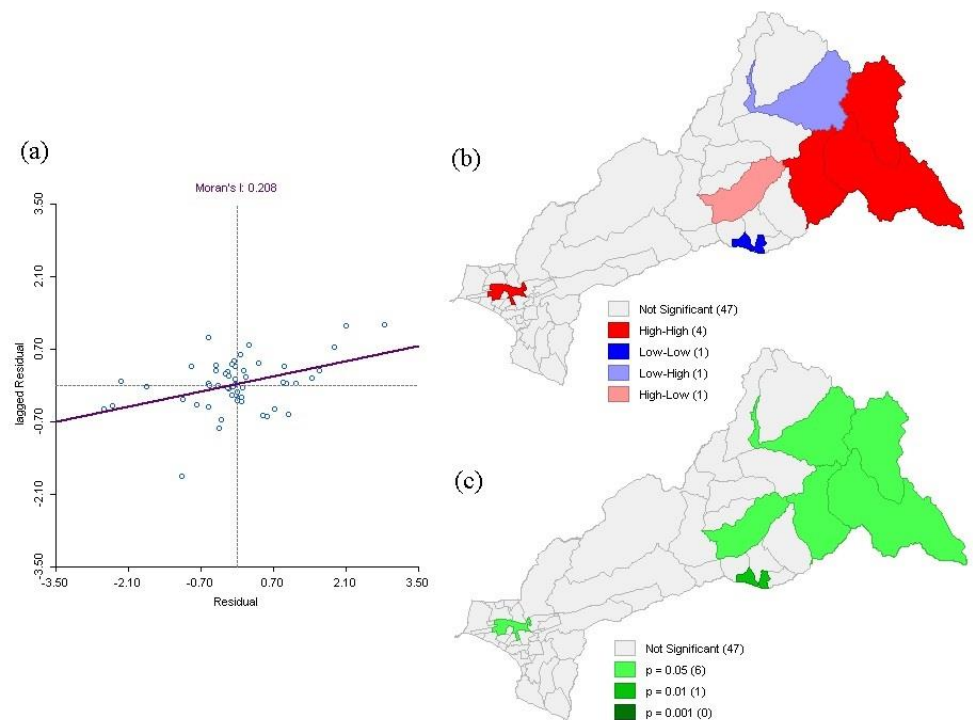
**Figura 36: Índice de Morán de los residuos del modelo para el mes de julio: diagrama de dispersión (a), mapa de conglomerados (b) y mapa de significancia (c)**



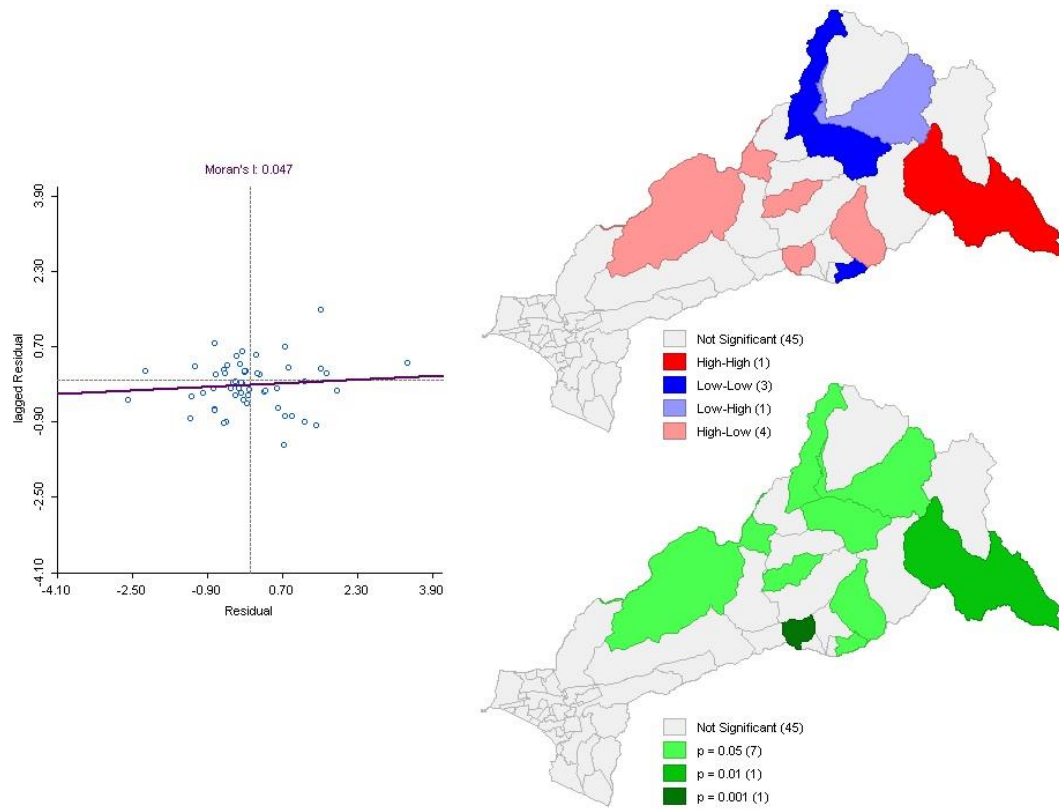
**Figura 37: Índice de Morán de los residuos del modelo para el mes de agosto: diagrama de dispersión (a), mapa de conglomerados (b) y mapa de significancia (c)**



**Figura 38: Índice de Morán de los residuos del modelo para el mes de septiembre: diagrama de dispersión (a), mapa de conglomerados (b) y mapa de significancia (c)**



**Figura 39: Índice de Morán de los residuos del modelo para el mes de octubre: diagrama de dispersión (a), mapa de conglomerados (b) y mapa de significancia (c)**



**Figura 40: Índice de Morán de los residuos del modelo para el mes de noviembre: diagrama de dispersión (a), mapa de conglomerados (b) y mapa de significancia (c)**