

**NATIONAL UNIVERSITY OF ENGINEERING**

**COLLEGE OF ECONOMICS AND STATISTICAL  
ENGINEERING**

**STATISTICAL ENGINEERING PROGRAM**



**INFORME DE CURSO CONSULTORÍA**

**TITLE**

"Analysis of the spatial distribution of % of people affiliated with the SIS in the departments of Ayacucho, Pasco and Huancavelica"

**TÍTULO**

"Análisis de la distribución espacial del % de personas afiliadas al SIS en los departamentos de Ayacucho, Pasco y Huancavelica"

**Integrantes:**

Oscar Diego Fernando Cutipa Luque  
Rodrigo Pujada Pujada Pacora

**Docente:**

Anna Sikov

**LIMA - PERÚ**

**2023**

## Índice de Contenidos

<b>LISTA DE FIGURAS</b>	<b>4</b>
<b>LISTA DE TABLAS</b>	<b>6</b>
<b>1 Introducción</b>	<b>8</b>
1.1 Definición del problema . . . . .	8
1.2 Descripción de los datos . . . . .	8
1.3 Objetivos de la investigación . . . . .	9
<b>2 Análisis preliminar</b>	<b>11</b>
2.1 Variable de Interés : Porcentaje de afiliados al SIS . . . . .	11
2.1.1 Nivel Provincial . . . . .	11
2.1.2 Nivel Distrital . . . . .	12
2.2 Variables Explicativas . . . . .	13
2.2.1 Departamento de Ayacucho . . . . .	13
2.2.2 Departamento de Huancavelica . . . . .	15
2.2.3 Departamento de Pasco . . . . .	16
2.3 Autocorrelación Espacial . . . . .	17
2.4 Índice de Moran por Departamento a Nivel Distrital . . . . .	18
<b>3 Metodología</b>	<b>20</b>
3.1 Modelo . . . . .	20
3.1.1 Definición . . . . .	20
3.1.2 Supuestos . . . . .	21
3.1.3 Hipótesis . . . . .	22
3.1.4 Pesos espaciales . . . . .	23
<b>4 Resultados</b>	<b>25</b>
4.1 Primeros resultados . . . . .	25
4.2 Departamento Ayacucho . . . . .	26
4.3 Departamento Huancavelica . . . . .	27
4.4 Departamento Pasco . . . . .	28
4.5 Supuestos . . . . .	28
4.6 Hipótesis . . . . .	29
4.6.1 Departamento de Ayacucho . . . . .	29
4.6.2 Departamento de Huancavelica . . . . .	30
4.6.3 Departamento de Pasco . . . . .	30
<b>5 Discusión</b>	<b>31</b>



## LISTA DE FIGURAS

2.1	Distribución espacial del departamento de estudio a nivel provincial del % de afiliados al SIS . . . . .	11
2.2	Distribución espacial del departamento de estudio a nivel distrital del % de afiliados al SIS . . . . .	12
2.3	Distribución espacial del Departamento de Ayacucho al nivel distrital de la Altitud y IDH . . . . .	13
2.4	Distribución espacial del Departamento de Ayacucho al nivel distrital del % de pobreza extrema y tasa de desempleo . . . . .	14
2.5	Distribución espacial del Departamento de Ayacucho al nivel distrital del % sin conexión a internet y % que no sabe leer ni escribir . . . . .	14
2.6	Distribución espacial del Departamento de Huancavelica al nivel distrital de la Altitud y IDH . . . . .	15
2.7	Distribución espacial del Departamento de Huancavelica al nivel distrital del % de pobreza extrema y tasa de desempleo . . . . .	15
2.8	Distribución espacial del Departamento de Huancavelica al nivel distrital del % sin conexión a internet y % que no sabe leer ni escribir . . . . .	16
2.9	Distribución espacial del Departamento de Pasco al nivel distrital de la Altitud y IDH . . . . .	16
2.10	Distribución espacial del Departamento de Pasco al nivel distrital del % de pobreza extrema y tasa de desempleo . . . . .	17
2.11	Distribución espacial del Departamento de Pasco al nivel distrital del % sin conexión a internet y % que no sabe leer ni escribir . . . . .	17

3.1	Grafo no direccionado del departamento de Ayacucho . . . . .	24
-----	--	----

## LISTA DE TABLAS

1.1	Descripción de las variables . . . . .	9
2.1	Indice Moran . . . . .	18
2.2	Indice Moran con covariables Ayacucho . . . . .	18
2.3	Indice Moran con covariables Huancavelica . . . . .	18
2.4	Indice Moran con covariables Pasco . . . . .	19
4.1	P-valor del índice de moran utilizando modelos de regresion lineal para cada departamento . . . . .	25
4.2	Comparación de los modelos SAR, SEM , SARMA considerando vecinos . .	26
4.3	Estimaciones del modelo SARMA en el departamento de Ayacucho . . . . .	26
4.4	Comparación de los modelos SAR, SEM , SARMA considerando vecinos . .	27
4.5	Estimaciones del modelo SEM en el departamento de Huancavelica . . . . .	27
4.6	Comparación de los modelos SAR, SEM , SARMA considerando vecinos . .	28
4.7	Estimaciones del modelo lineal en el departamento de Pasco . . . . .	28
4.8	P-valor de los mejores modelos utilizando la prueba de normalidad . . . . .	29
4.9	P-valor de la prueba de hipótesis para las 3 variables en el departamento de Ayacucho . . . . .	29
4.10	P-valor de la prueba de hipótesis para las 3 variables en el departamento de Huancavelica . . . . .	30

4.11 P-valor de la prueba de hipótesis para las 3 variables en el departamento de Pasco . . . . .	30
---	----

# **Chapter 1**

## **Introducción**

### **1.1 Definición del problema**

La afiliación al SIS es un tema decisivo para evaluar la accesibilidad a la atención médica en el Perú que es un problema actualmente en el país y con mayor concentración en la región central, donde se encontramos a los departamentos de Ayacucho, Huancavelica y Pasco, que se enfrentan desafíos únicos en este aspecto, a pesar de los esfuerzos gubernamentales.

Comprender estos desafíos y factores determinantes es esencial para abordar las disparidades en el acceso a la atención médica en esa región central del país, esto con el fin de que una entidad pública del estado pueda plantear una ruta de acción que permita acortar la brecha de disparidad respecto a otras regiones del país.

Es necesario, tener estimaciones confiables en el modelo para así comprender la distribución espacial del % de afiliados al SIS, por ello es necesario utilizar modelos que capturen la estructura espacial, ya que los distritos más cercanos poseen características similares y los distritos más lejanos características diferentes.

Con esta definición nos planteamos la siguiente pregunta: ¿Cuál es el mejor modelo por departamento para estimar las covariables que permitan tener resultados confiables y capturar los patrones espaciales?

### **1.2 Descripción de los datos**

En el presente estudio, donde nuestra variable de interés es el % de afiliados al Sistema Integrado de Salud (SIS) en cada distrito de los departamentos de Ayacucho, Huancavelica y Pasco, esta información se obtuvo del Censo Nacional 2017 realizado por el Instituto Nacional de Estadística e Informática (INEI), asimismo incorporamos variables socioeconómicas basándonos en (Mezones-Holguín et al., ) donde se realiza un estudio de la cobertura de



seguros de salud basandose en factores similares a los que estamos considerando.

Del Censo se obtuvo la información de las covariables del % de la población que no tiene conexión a internet y % de la población que no sabe leer ni escribir en cada distrito de los departamentos en estudio, por otro lado las covariables de Altitud, Índice de Desarrollo Humano (IDH), Tasa de Desempleo y el % de población con pobreza extrema se obtuvo de la información para el planeamiento a nivel departamental, provincial y distrital que proporciona el Centro Nacional de Planeamiento Estratégico (CEPLAN).

Variable	Descripción
AFSIS	Porcentaje de la población afiliada al SIS en cada distrito de los departamentos de Ayacucho, Huánuco y Pasco
ALT	Altitud en (msnm) promedio en cada distrito
IDH	Índice de Desarrollo Humano por distrito
PE	% de la población que presenta pobreza extrema
TD	Tasa de Desempleo por distrito
SCI	% de la población que no tiene acceso a internet por distrito
NLE	% de la población que no sabe leer ni escribir por distrito

**Tabla 1.1:** Descripción de las variables

### 1.3 Objetivos de la investigación

Evaluar la influencia de las variables de pobreza y demográficas en el % de afiliados al SIS de manera distrital en los departamentos de Ayacucho, Huancavelica y Pasco. De manera específica:

- Comparar los modelos de regresión espaciales con los modelos lineales identificando cuáles presentan mejores estimaciones en cada departamento en estudio.
- Contrastar si la altitud promedio en cada departamento en estudio presenta un efecto negativo en el % afiliados al SIS.
- Contrastar si el % de la población sin acceso a internet en cada departamento en estudio presenta un efecto negativo en el % afiliados al SIS.
- Contrastar si IDH (índice de Desarrollo Humano) en cada departamento en estudio presenta un efecto negativo en el % afiliados al SIS.

En el presente trabajo emplearemos modelos que capturen la estructura espacial (SAR, SEM y SARMA) para medir el comportamiento del % de afiliados al SIS a nivel distrital en cada departamento en estudio (Ayacucho , Huancavelica y Pasco) . la ventaja de emplear estos modelos espaciales es que nos brindan mejores estimaciones, ya que la variable de interés no se trabaja de manera independiente , sino que se espera que distritos cercanos presenten valores similares , mientras que para distritos lejanos los valores presentarían disparidades.

Además los objetivos de uso de los modelos espaciales son para estimar el efecto de las covariables y para contrastar nuestras hipótesis de investigación.

Implementaremos funciones en el software R que permitan estimar modelos espaciales (SAR, SEM y SARMA), además de que podremos comparar estos modelos mediante una métrica (AIC), con el fin de poder seleccionar el mejor modelo para cada departamento en estudio.

Los resultados que pensamos obtener son :

- El mejor modelo para los departamentos que posean las 6 covariables con correlación espacial significativa pueda ser un modelo de regresión lineal, debido que la estructura espacial de la variable de interés sea capturado mediante las 6 covariables.
- Un resultado interesante sería que a mayor IDH el % de afiliados al SIS disminuya, ya que tener un mayor IDH es sinónimo de un mejor crecimiento de un distrito y por lo tanto una buena proporción de la población tendría acceso a otros tipos de seguros de salud.

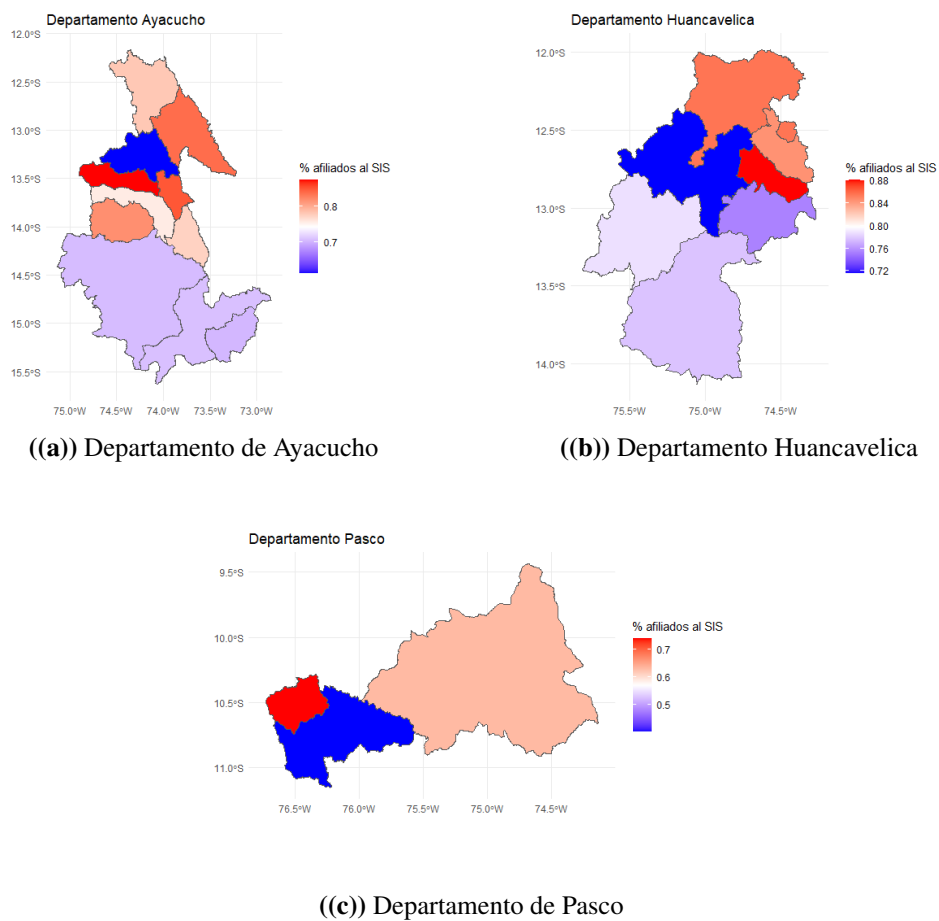
Finalmente los modelos seleccionados deberán cumplir con el supuesto de independencia y normalidad de los residuales, esto para validar y obtener estimaciones mas confiables.

## Chapter 2

### Análisis preliminar

## 2.1 Variable de Interés : Porcentaje de afiliados al SIS

### 2.1.1 Nivel Provincial



**Figura 2.1:** Distribución espacial del departamento de estudio a nivel provincial del % de afiliados al SIS

Según el gráfico 2.1 observamos que para el departamento de Ayacucho y Huancavelica el % de afiliados al SIS aparentemente presenta una estructura espacial, por ejemplo se observa que en el Departamento de ayacucho el % de afiliados al SIS que presentan una proporción de 0.7 aprox se encuentran en el sur.

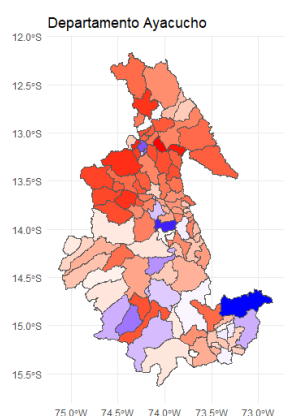
Por otro lado en el Departamento de Huancavelica, se evidencia que las provincias ubicadas en el norte presentan un alto % afiliados al SIS, además conforme nos acercamos al sur el % afiliados al SIS disminuye.

Por último en el departamento de Pasco, debido a que solo se tiene 3 provincias no se evidencia una estructura espacial clara.

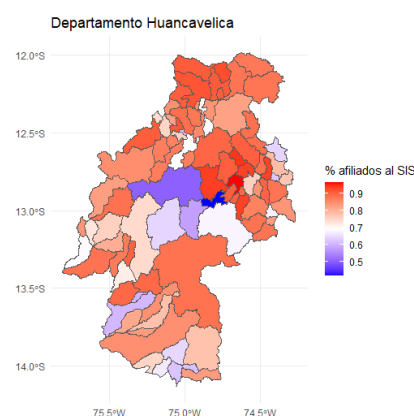
Ahora nuestro foco será en poder ver la estructura espacial de manera distrital, ya que aquí se podrá evidenciar de mejor manera la estructura espacial.

## 2.1.2 Nivel Distrital

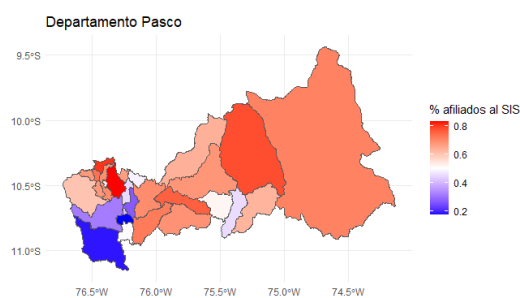
### Departamento de Ayacucho



((a)) Departamento de Ayacucho



((b)) Departamento de Huancavelica



((c)) Departamento de Pasco

**Figura 2.2:** Distribución espacial del departamento de estudio a nivel distrital del % de afiliados al SIS

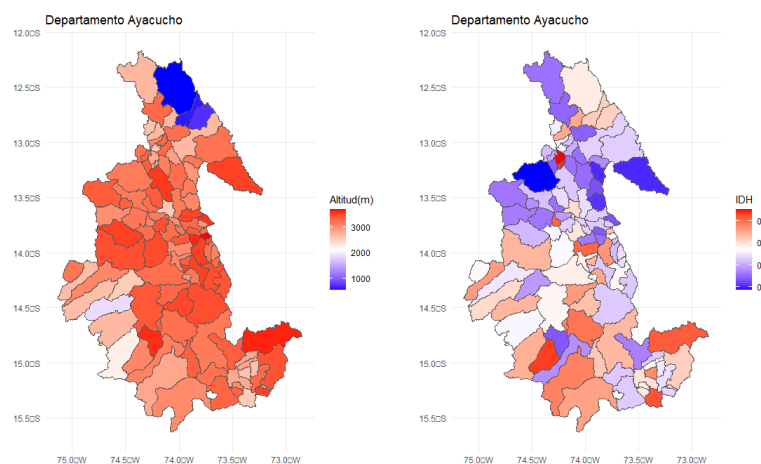
En la figura 2.2 observamos que en el departamento de Ayacucho, el mayor % de afiliados al sis se encuentra en el norte del departamento, por otro lado observamos que en el sur esta proporción disminuye.

En el departamento de Huancavelica se observa que en el centro de este departamento la % afiliados al SIS presenta una baja proporción, en comparación con el norte y el sur.

En el departamento de Pasco, se muestra una mayor % de afiliados al SIS en la parte derecha, mientras que en el lado izquierdo presenta una proporción menor.

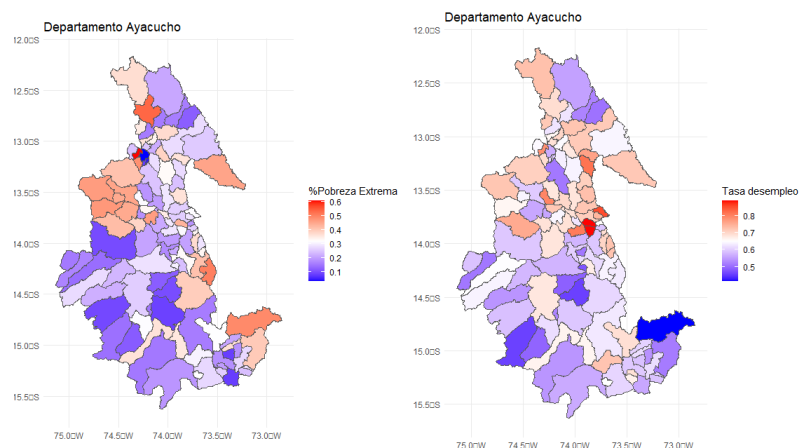
## 2.2 Variables Explicativas

### 2.2.1 Departamento de Ayacucho



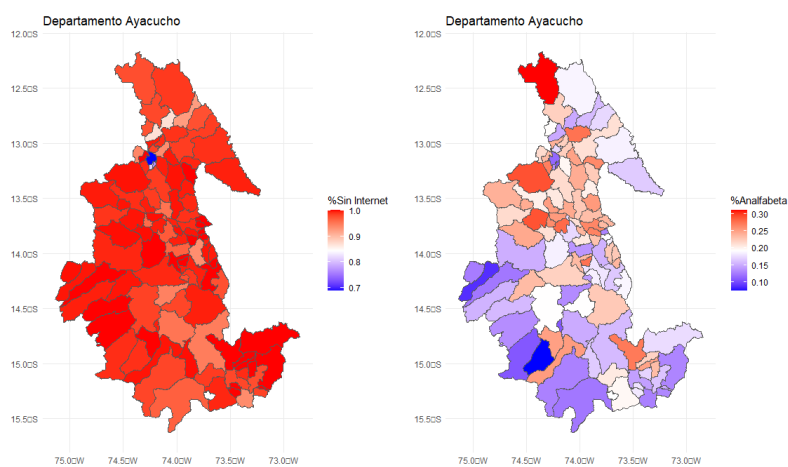
**Figura 2.3:** Distribución espacial del Departamento de Ayacucho al nivel distrital de la Altitud y IDH

Estas dos variables también presentan una clara estructura espacial, en el caso de la altitud, se evidencia que casi todo el Departamento de Ayacucho presenta una elevada altitud y solo un pequeño distrito presenta una baja altitud. Además la variable IDH también presenta una estructura espacial.



**Figura 2.4:** Distribución espacial del Departamento de Ayacucho al nivel distrital del % de pobreza extrema y tasa de desempleo

En el gráfico 2.4 observamos que tanto el % de pobreza extrema como la tasa de desempleo presentan una cierta estructura espacial.

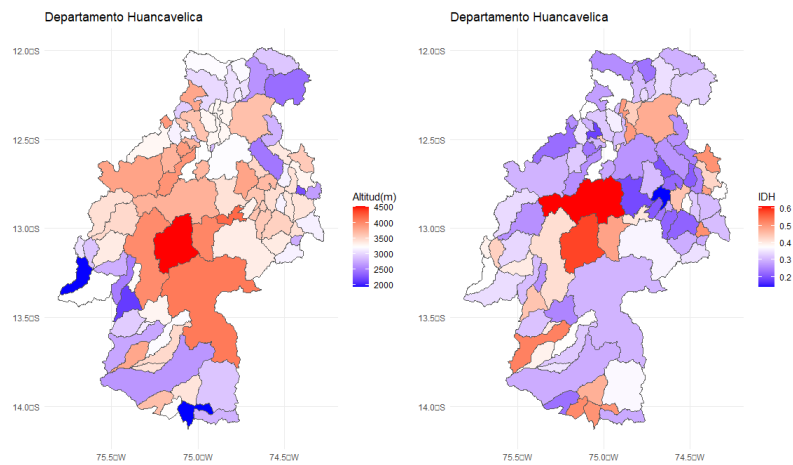


**Figura 2.5:** Distribución espacial del Departamento de Ayacucho al nivel distrital del % sin conexión a internet y % que no sabe leer ni escribir

En el gráfico 2.4 observamos que en casi todo el departamento presenta una baja % de proporción sin conexión a internet.

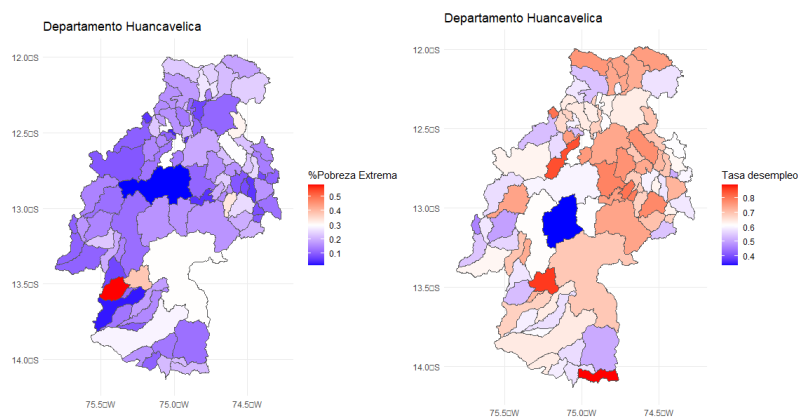
Por otro lado en la variable % que no sabe leer ni escribir si se evidencia que los distritos más cercanos presentan similitudes, mientras que distritos más lejanos presentan proporciones mas diferentes.

### 2.2.2 Departamento de Huancavelica



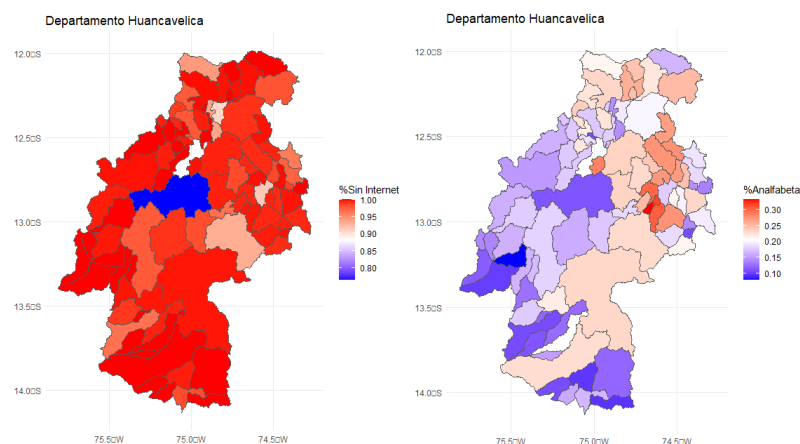
**Figura 2.6:** Distribución espacial del Departamento de Huancavelica al nivel distrital de la Altitud y IDH

Se evidencia que para ambas variables Altitud e IDH existe cierto patrón espacial, por ejemplo en el centro del departamento se observa una elevada altitud y un alto IDH.



**Figura 2.7:** Distribución espacial del Departamento de Huancavelica al nivel distrital del % de pobreza extrema y tasa de desempleo

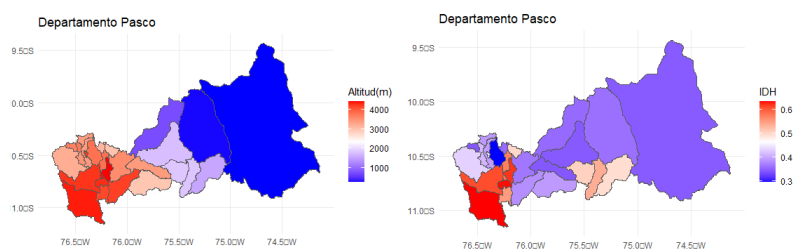
En la figura 2.7 tanto % de pobreza de extrema y la tasa de desempleo tiene ciertas peculiaridades, ya que los distritos cercanos presentan valores en las variables similares, por lo tanto presentan cierto patrón espacial.



**Figura 2.8:** Distribución espacial del Departamento de Huancavelica al nivel distrital del % sin conexión a internet y % que no sabe leer ni escribir

Nuevamente en la figura 2.8 nuevamente las variables % personas sin conexión a internet y % personas que no saben presentan cierto patrón espacial, lo cual será necesario para comprender como esto afecta al % afiliados al SIS.

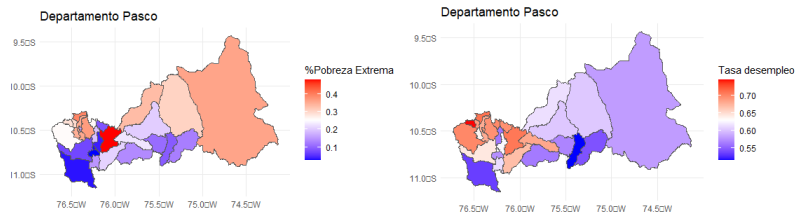
### 2.2.3 Departamento de Pasco



**Figura 2.9:** Distribución espacial del Departamento de Pasco al nivel distrital de la Altitud y IDH

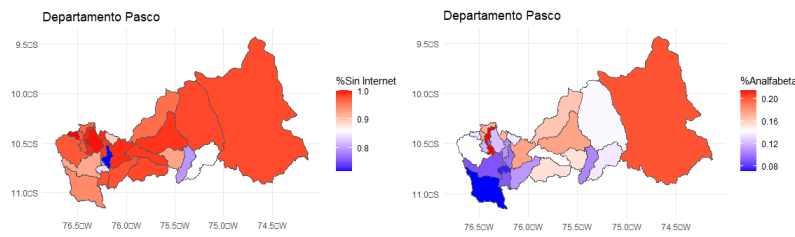
Según la figura 2.9 se existe un compartamiento similar en los distritos de la derecha y una muy diferente en los distritos de la izquierda para las variables altitud e IDH.





**Figura 2.10:** Distribución espacial del Departamento de Pasco al nivel distrital del % de pobreza extrema y tasa de desempleo

En el gráfico 2.10 se observa algo similar a las 2 variables anteriores, ese comportamiento disjunto entre los distritos de la derecha e izquierda se mantiene para las variables % de pobreza extrema y tasa de desempleo.



**Figura 2.11:** Distribución espacial del Departamento de Pasco al nivel distrital del % sin conexión a internet y % que no sabe leer ni escribir

Por último en el gráfico 2.11 se evidencia que % de personas sin conexión a internet y la % que no sabe leer y escribir, posee una distribución espacial diferente a las 4 variables anteriores, podemos ver que para la primera variable mencionada solo hay distrito que presenta una muy baja % de conexión a internet.

## 2.3 Autocorrelación Espacial

Para ello utilizaremos el índice de Moran para determinar si presenta o autocorrelación espacial. Esto bajo la siguiente hipótesis estadística:

$H_0$  : Existe correlación negativa o no hay correlación espacial.

$H_1$  : Existe correlación espacial positiva.

$$I = \frac{n \sum_i \sum_j w_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{(\sum_{i \neq j} \sum_i (Y_i - \bar{Y})^2)}$$

## 2.4 Índice de Moran por Departamento a Nivel Distrital

A continuación se presentan los resultados de los Índices de Moran por Departamento a Nivel Distrital, esto considerando un nivel de significancia del 5%.

Departamento	Índice de Moran	p-valor
Ayacucho	0.258	0.000
Huancavelica	0.202	0.001
Pasco	0.426	0,000

**Tabla 2.1:** Índice Moran

En la tabla 2.1 observamos que la variable % de afiliados al SIS presenta correlación espacial en los 3 departamentos, esto nos da indicios que debemos trabajar con modelos que capturen la estructura espacial.

Covariables	MoranT	MoranMC	Moran Statitics
ALT	0.0000	0.0005	0.3601
IDH	0.0022	0.0015	0.17
PE	0.0004	0.0009	0.2017
TD	0.0000	0.0004	0.2782
SCI	0.0000	0.0009	0.3099
NLE	0.0000	0,0005	0.2509

**Tabla 2.2:** Índice Moran con covariables Ayacucho

En la tabla 2.2 observamos que todas las covariables en estudio para el departamento de Ayacucho , presentan estructura espacial debido a los valores obtenidos con el índice de moran , lo cual verifica lo descrito por nuestro análisis preliminar de los gráficos.

Covariables	MoranT	MoranMC	Moran Statitics
ALT	0.0000	0.0005	0.2825
IDH	0.0156	0.0209	0.1344
PE	0.4311	0.4117	0.0002
TD	0.043	0.0489	0.1043
SCI	0.7637	0.7596	-0.0511
NLE	0.0000	0,0005	0.4574

**Tabla 2.3:** Índice Moran con covariables Huancavelica

En la tabla 2.3 observamos que 4 covariables en estudio para el departamento de Huancavelica , presentan estructura espacial según el índice de moran, las covariables de % de la población en pobreza extrema y % de la población que no tiene acceso a conexión al internet , no presentan una estructura espacial.

Covariables	MoranT	MoranMC	Moran Statitics
ALT	0.0000	0.0005	0.7718
IDH	0.0001	0.0005	0.4313
PE	0.0012	0.0019	0.3388
TD	0.0001	0.0005	0.4191
SCI	0.0053	0.0135	0.2607
NLE	0.0024	0,0035	0.3097

**Tabla 2.4:** Indice Moran con covariables Pasco

En la tabla 2.4 observamos en el departamento de Pasco que al igual que en Ayacucho todas las covariables presentan una estructura espacial , según el indice de moran.

## Chapter 3

### Metodología

### 3.1 Modelo

#### 3.1.1 Definición

Para el presente trabajo de investigación emplearemos el modelo SARMA(*Spatially Autoregressive Moving Average*), que lo representamos matemáticamente como:

$$Y = \rho WY + X\beta + u$$

$$u = \lambda Wu + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

Este modelo presenta dos componentes: **los rezagos en la variable de interés y los rezagos en los residuos del modelo.**

Teniendo en cuenta ello, presentamos:

- **Y:** % de afiliados al SIS en los distritos de un departamento.

Esta variable va a ser explicada por variables sociodemográficas y por un efecto espacial. Las variables sociodemográficas van a ser representadas por  $X$  que representa una matriz donde cada fila representa una observación de las  $p$  variables explicativas. A continuación mencionaremos dichas variables explicativas.

- **$X_1$ :** Altitud Promedio en msnm.
- **$X_2$ :** Índice de Desarrollo Humano.
- **$X_3$ :** % de la población con pobreza extrema.
- **$X_4$ :** % de desempleo.

- $X_5$ : % de la población que no tiene conexión a internet.
- $X_6$ : % de la población analfabeta.

Por último se definen los parámetros y la matriz  $W$ :

- $\rho$ : Efecto de los rezagos espaciales del % de afiliados al SIS.
- $\beta$ : Vector efectos de las variables regresoras
- $u$ : Vector rezagos
- $\lambda$ : Efecto de los rezagos espaciales del residuo.
- $\sigma^2$ : Varianza del error aleatorio del modelo.
- $W$ : Matriz de pesos o ponderación espacial del departamento de interés

### 3.1.2 Supuestos

Todo modelo estadístico se construye en base a supuestos. En este caso el modelo SARMA presenta los siguientes supuestos en base a los residuales. Para que el modelo nos brinde buenas estimaciones tenemos que comprobar que cumpla con los siguientes supuestos los residuales:

#### Supuesto de Normalidad:

Para poder comprobar el supuesto emplearemos la **Prueba de Shapiro-Wilk**.

En esta prueba validaremos el supuesto de normalidad a través de la prueba de Shapiro Wilks, primeramente enunciamos las hipótesis a contrastar:

$H_0$ : Los errores siguen una distribución normal.

$H_1$ : Los errores no siguen una distribución normal.

Para esta prueba tenemos el estadístico W de Shapiro-Wilk

$$W = \frac{(\sum_{i=1}^n a_i e_{(i)})^2}{\sum_{i=1}^n (e_i - \bar{e})^2}$$

Si W es significativamente menor que 1, esto sugiere que los residuos no siguen una distribución normal

#### Supuesto de Independencia:

Para este supuesto emplearemos la **Prueba de Moran**. En esta prueba validaremos el supuesto de homocasticidad, primeramente enunciamos las hipótesis a contrastar:

$H_0$ : La varianza de los errores no está correlacionada espacialmente.

$H_1$ : La varianza de los errores está correlacionada espacialmente.

**Supuesto de Homocedasticidad:** Para este supuesto emplearemos la **Prueba de White**.

En esta prueba validaremos el supuesto de homocesticidad, primeramente enunciamos las

hipotesis a contrastar:

$H_0$ : La varianza de los errores no está correlacionada espacialmente.

$H_1$ : La varianza de los errores está correlacionada espacialmente

### 3.1.3 Hipótesis

Realizaremos el modelo SARMA para cada Departamento(Ayacucho, Huancavelica, Pasco). Todo modelo SARMA se puede reducir a un modelo SAR cuando  $\lambda = 0$  y puede reducirse a un modelo SEM cuando  $\rho = 0$ . Por último cuando se tiene que tanto  $\lambda = 0$  y  $\rho = 0$  entonces se reduce a un modelo de regresión lineal.

Por ello será importante probar primero de manera conjunta si podríamos utilizar el modelo SARMA en cada uno de los departamentos.

$$H_0: \begin{pmatrix} \rho \\ \lambda \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{vs} \quad H_1: \begin{pmatrix} \rho \\ \lambda \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (3.1)$$

Estas hipótesis la probaremos utilizando la prueba de Wald y la distribución asintótica cuando  $n \rightarrow \infty$ , de esta manera:

$H_0: C\theta = c$

$H_1: C\theta \neq c$

Se define el siguiente estadístico:

$$W = (C\hat{\theta} - c)'(-CH(\hat{\theta})C')^{-1}(C\hat{\theta} - c) \sim \chi^2_{(k)}$$

donde  $k = \dim(\theta) - \dim(\theta_0)$  y  $\theta = (\beta', \sigma, \rho, \lambda)'$ . Como ultimo detalle  $H(\theta)$  es el hesiano del logaritmo de la función de verosimilitud y se define de la siguiente manera:

$$H(\theta) = \frac{\partial^2 \log L(\text{Data}|\theta)}{\partial \theta^2}$$

teniendo en cuenta que  $L(\text{Data}|\theta)$  : Función de verosimilitud

Una vez definido el mejor modelo para cada departamento, obtaremos por medir el grado de significancia de las variables. Por ello para cada modelo utilizaremos las siguientes hipótesis:

$H_0: \beta_i = 0$ .

$H_1: \beta_i \neq 0$ .

para  $i \geq 2$ , dado que no nos importa por que el intercepto sea significativo.

Para esta hipótesis se utilizará el siguiente estadístico:

$$W = \frac{\hat{\beta}_i}{\sqrt{-H_{ii}^{-1}(\hat{\theta})}} \sim \mathcal{N}(0, 1)$$

donde  $i \in \{2, 3, 4, 5, 6, 7\}$

Por último cuando queramos saber la significancia individual de  $\rho$  y  $\lambda$  para determinar si un modelo pueda ajustarse mejor con el SAR o SEM. Empezaremos por  $\rho$ , las hipótesis serían:

$$H_0: \rho = 0.$$

$$H_1: \rho \neq 0.$$

Utilizaremos un estadístico similar:

$$W = \frac{\hat{\rho}_i}{\sqrt{-H_{99}^{-1}(\hat{\theta})}} \sim \mathcal{N}(0, 1)$$

Para el otro parámetro  $\lambda$ , tenemos las hipótesis:

$$H_0: \lambda = 0.$$

$$H_1: \lambda \neq 0.$$

El estadístico es:

$$W = \frac{\hat{\lambda}_i}{\sqrt{-H_{(10)(10)}^{-1}(\hat{\theta})}} \sim \mathcal{N}(0, 1)$$

### 3.1.4 Pesos espaciales

En esta subseccion trataremos de explicar como hemos trabajado con los pesos espaciales, ya que este sería el último paso para realizar tanto las estimaciones de los modelos espaciales como las hipótesis planteadas.

Por ello primero definamos a que vamos a considerar vecinos en nuestro modelo. En este ejemplo, la matriz de Pesos espaciales  $W$  será representado de la siguiente manera considerando (Cerdeña, 2010):

$$w_{ij} = \begin{cases} 1 & \text{si el área } j \text{ comparte una frontera con } i \\ 0 & \text{caso contrario} \end{cases}$$

Ahora para que la matriz  $W$  esté bien especificada es necesario que se les de un peso similar para cada vecino, es decir si un distrito tiene 3 vecinos, se le estaría asignando 0.3 a cada vecino de peso.



**Figura 3.1:** Grafo no direccionado del departamento de Ayacucho

A continuación veremos un ejemplo de un grafo para el departamento de Ayacucho en 3.1, en esta se presenta un grafo definido por un vértice que hace alusión al centroide de cada distrito y una arista que asocia a un vecino de otro.

En este gráfico no solo se observa las relaciones sino que también se evidencia distritos que poseen más vecinos que otros. Podríamos decir que esta estructura captura el patrón espacial con el que trabajaremos modelos espaciales.



## Chapter 4

### Resultados

#### 4.1 Primeros resultados

Primero definamos un modelo de regresión lineal para cada departamento, recordemos que este modelo cumple el supuesto de errores independientes.

Ahora procederemos a aplicar la prueba de Moran a los residuales en el espacio distrital, con el fin de ver si existe correlación espacial o no.

Departamento	P-valor
Ayacucho	0.0001
Huancavelica	0.004
Pasco	0.917

**Tabla 4.1:** P-valor del índice de moran utilizando modelos de regresion lineal para cada departamento

De la tabla 4.1 podemos podemos indicar que para los departamentos **Ayacucho** y **Huancavelica**, presentan residuales que poseen alguna estructura espacial. Sin embargo los residuales del departamento de **Pasco** son independientes, por lo que el modelo de regresion lineal es adecuado.

Para los modelos que necesiten una estructura espacial, trabajaremos con estos 3: Modelos **SAR**, **SEM** y **SARMA**.

A continuación presentaremos una comparación de los diferentes modelos utilizados, tomando en cuenta el número de parámetros a estimar, el Criterio de información de Akaike, el p-valor de la prueba de Moran para los residuales de cada modelo, por último el p-valor de la significancia de  $\rho$  y  $\lambda$ .

## 4.2 Departamento Ayacucho

A continuación presentamos una comparación de los modelos utilizados usando los criterios mencionados anteriormente.

Modelo	N°Param	AIC	p-valor	p-valor $\rho$	p-valor $\lambda$
Lineal	7	-279.95	0.0001	—	—
SAR	8	-284.69	0.024	0.009	—
SEM	8	-294.27	0.550	—	0.000
SARMA	9	-294.63	0.622	0.029	0.000

**Tabla 4.2:** Comparación de los modelos SAR, SEM , SARMA considerando vecinos

En la tabla 4.2 considerando el AIC más bajo y el número de parámetros más bajo podemos ver que los mejores modelos son el SARMA y el SEM. Además de que ambos modelos presentan residuales independientes, esto significa que ambos modelos capturan bien la estructura espacial.

### Modelo SARMA Ayacucho

Coefficientes	Estimación	Std.Error	p-valor
Intercepto	0.361	0.232	0.120
ALT	-0.000	0.000	0.361
IDH	-0.584	0.113	0.000
PE	-0.091	0.065	0.160
TD	0.204	0.084	0.015
SCI	0.707	0.171	0.000
NLE	0.536	0.155	0.001

**Tabla 4.3:** Estimaciones del modelo SARMA en el departamento de Ayacucho

En la tabla 4.3, se observa las estimaciones del modelo utilizando el mejor modelo (SARMA), con ello se obtiene:

- La altitud, y el porcentaje de la población en pobreza extrema en los diferentes distritos de Ayacucho no influye en el porcentaje de personas afiliadas al SIS.
- Por otro lado, las demás variables como el Índice de desarrollo humano, tasa de desempleo, población que no tiene acceso a internet y población que no sabe leer ni escribir tienen efecto en el promedio de las personas afiliadas al SIS.
- Además un incremento del 10 % en el porcentaje de la población en pobreza extrema disminuye en 0.918 % en promedio en el porcentaje de afiliados al SIS.

### 4.3 Departamento Huancavelica

Modelo	NºParam	AIC	p-valor	p-valor $\rho$	p-valor $\lambda$
Lineal	7	-230.93	0.004	—	—
SAR	8	-236.26	0.092	0.006	—
SEM	8	-237.67	0.444	—	0.003
SARMA	9	-236.65	0.445	0.307	0.092

**Tabla 4.4:** Comparación de los modelos SAR, SEM , SARMA considerando vecinos

En la tabla 4.6 observamos que según AIC, el mejor modelo es el SEM. Tanto el SEM como el SARMA presentan residuales sin estructura espacial, es decir, residuales independientes. Además en el modelo SARMA  $\rho$  no es significativo, por lo que el mejor modelo en parsimonia sería el modelo SEM.

A continuación veremos las estimaciones del modelo SEM en el departamento de Huancavelica.

#### Modelo SEM Huancavelica

Coefficientes	Estimación	Std.Error	p-valor
Intercepto	0.843	0.244	0.001
ALT	0.000	0.000	0.713
IDH	-0.753	0.099	0.000
PE	0.182	0.088	0.038
TD	-0.146	0.077	0.057
SCI	0.235	0.211	0.265
NLE	0.227	0.166	0.171

**Tabla 4.5:** Estimaciones del modelo SEM en el departamento de Huancavelica

En la tabla 4.5 se observan las estimaciones del modelo SEM, a continuación se presentan los siguientes resultados:

- Las variables Altitud, población que no tiene acceso a internet y la población que no sabe leer ni escribir no presentan efecto en promedio en el porcentaje de afiliados al SIS.
- Por otro lado, las variables de Índice de desarrollo humano, población en pobreza extrema y tasa de desempleo tienen un efecto en el promedio del porcentaje de afiliados al SIS.
- Un incremento en 0.1 en el Índice de Desarrollo Humano disminuye en promedio en 7.5 % en el porcentaje de afiliados al SIS.

## 4.4 Departamento Pasco

Modelo	N°Param	AIC	p-valor	p-valor $\rho$	p-valor $\lambda$
Lineal	7	-60.57	0.92	—	—
SAR	8	-58.59	0.93	0.905	—
SEM	8	-64.09	0.39	—	0.018
SARMA	9	-63.59	0.44	0.225	0.000

**Tabla 4.6:** Comparación de los modelos SAR, SEM , SARMA considerando vecinos

En la tabla 4.6 observamos que tanto los modelos espaciales como el modelo de regresión lineal, presentan residuales independientes.

Además según el primer resultado, observamos que ya no es necesario utilizar modelos espaciales, ya que el lineal a pesar de que no presenta el mejor AIC, es el modelo que tiene menor cantidad de parámetros.

### Modelo de Regresión Lineal - Pasco

Coeficientes	Estimación	Std.Error	p-valor
Intercepto	1.294	0.454	0.010
ALT	0.000	0.000	0.756
IDH	-1.431	0.307	0.000
PE	0.229	0.209	0.285
TD	-0.171	0.385	0.661
SCI	-0.074	0.284	0.798
NLE	0.307	0.613	0.622

**Tabla 4.7:** Estimaciones del modelo lineal en el departamento de Pasco

En la tabla 4.7 se observa las estimaciones por regresión lineal en el departamento de Pasco, algunos resultados son:

- Para este caso solo la variable de Índice de desarrollo humano presenta efecto en promedio en el porcentaje de afiliados al SIS.
- Un aumento de 0.1 en el IDH disminuiría en promedio 14.31 % en el porcentaje de afiliados al SIS.

## 4.5 Supuestos

Probaremos el supuesto de normalidad con la prueba de **Anderson Darling** Por ello se tiene la siguiente tabla:

Departamento	p-valor
Ayacucho (SARMA)	0.046
Huancavelica (SEM)	0.003
Pasco (Lineal)	0.777

**Tabla 4.8:** P-valor de los mejores modelos utilizando la prueba de normalidad

En la siguiente tabla 4.8 tenemos lo siguiente:

- Para  $\alpha$  de 1 % el modelo para Ayacucho y el modelo para Pasco cumplen este supuesto.
- A pesar que el modelo Huancavelica no cumple con este supuesto de normalidad, hemos analizado a través del AIC y la independencia de residuos que el mejor modelo es el SEM.

## 4.6 Hipótesis

Ahora responderemos los objetivos de investigación, por ello vamos a plantear las siguientes hipótesis:

$$H_0: \beta_i = 0$$

$$H_1: \beta_i < 0.$$

Esto considerando  $i \in \{1, 2, 5\}$ , estos efectos hacen referencia a las covariables Altitud, % de población sin acceso a internet e índice de desarrollo humano.

Los resultados se consideran utilizando un nivel de significancia de 5%. A continuación presentaremos por departamento el p-valor para esta prueba de hipótesis:

### 4.6.1 Departamento de Ayacucho

Covariables	Estimado	Desviación E.	p-valor
Altitud	-0.00001	0.00001	0.18043
SCI	0.70746	0.17075	0.99998
IDH	-0.5844	0.11266	0.00000

**Tabla 4.9:** P-valor de la prueba de hipótesis para las 3 variables en el departamento de Ayacucho

Segun la tabla 4.9 podemos decir que tanto la altitud como el % de población sin acceso a internet no tienen un efecto negativo sobre el % afiliados al SIS. Además el IDH si tiene un efecto negativo significativo sobre % afiliados al SIS.

#### 4.6.2 Departamento de Huancavelica

Covariables	Estimado	Desviación E.	p-valor
Altitud	-0.00001	0.00001	0.64363
SCI	0.23473	0.21075	0.86731
IDH	-.075294	0.09939	0.00000

**Tabla 4.10:** P-valor de la prueba de hipótesis para las 3 variables en el departamento de Huancavelica

En la tabla 4.10 se evidencia nuevamente que tanto la variable altitud como el % de población sin acceso a internet no tienen un efecto negativo sobre % afiliados al SIS. Ahora el IDH si presenta de manera significativa un efecto negativo, esto quiere decir que si el IDH aumenta, el % de afiliados al SIS disminuye.

#### 4.6.3 Departamento de Pasco

Covariables	Estimado	Desviación E.	p-valor
Altitud	-0.000006	0.00002	0.62366
SCI	-0.0737	0.2843	0.39773
IDH	-1.431	0.3066	0.000002

**Tabla 4.11:** P-valor de la prueba de hipótesis para las 3 variables en el departamento de Pasco

Por último en el departamento de Pasco según la tabla 4.11, la única variable significativa y que tiene un efecto negativo en el % de afiliados al SIS es el índice de desarrollo humano. Las otras variables no presentan dicho efecto negativo, tampoco positivo, simplemente no afectan en promedio al % de afiliados al SIS.

## **Chapter 5**

### **Discusión**

De los resultados anteriores se logró ver que para el departamento de Ayacucho el mejor modelo es el SARMA, para el departamento de Huancavelica es el SEM; mientras que para el departamento de Pasco el modelo de regresión lineal es el mejor.

La variable Altitud no presenta efecto negativo en el % afiliados al SIS esto para los 3 departamentos de manera particular, sino al contrario estos no tienen efecto significativo en la variable de interés mencionada. Esto quiere decir que distritos en un departamento de estudio que presentan altitudes diferentes no afectará en el % afiliados al SIS.

Además no existe evidencia suficiente para que la variable % población sin acceso a internet presente efecto negativo en el % afiliados al SIS para los 3 departamentos en estudio. Por lo que esta variable no es relevante para nuestra variable de interés.

Por último el Índice de desarrollo humano si presenta un efecto negativo significativo con el % afiliados al SIS para los 3 departamentos en estudio. Esto implica que un aumento en el IDH disminuye la % afiliados al SIS, una posible causa es que el IDH mide el nivel de desarrollo de un país, si este aumenta las personas tendrían mejor esperanza de vida y mejor calidad lo que ocasionaría que el % afiliados al SIS disminuya y esta población utilice otros seguros particulares.

## References

- Centro Nacional de Planeamiento Estratégico. (n.d.). *Información sobre zonas y departamentos del Perú*. Retrieved from <https://www.ceplan.gob.pe/informacion-sobre-zonas-y-departamentos-del-peru/> (Accedido el 22 de septiembre de 2023)
- Cerda, J., Anna Sikov. (2021). *Introducción a la estadística espacial para ingeniería, teoría y aplicaciones con r*. Universidad Nacional de Ingeniería.
- Instituto Nacional de Estadística e Informática (INEI). (n.d.). *Censos Nacionales 2017 - Redatam*. Retrieved from <https://censos2017.inei.gob.pe/redatam/> (Accedido el 22 de septiembre de 2023)
- Mezones-Holguín, E., Amaya, E., Bellido-Boza, L., Mougenot, B., Murillo, J. P., Villegas-Ortega, J., Del Carmen Sara, J. C. (2019). Cobertura de aseguramiento en salud: el caso peruano desde la ley de aseguramiento universal. *Revista Peruana de Medicina Experimental y Salud Pública*, 36(2), 196-206. doi: 10.17843/rpmesp.2019.362.3998
- Moraga, P. (2023). *Spatial statistics for data science: Theory and practice with r*. Chapman Hall/CRC.