

Análisis de la Variabilidad Espacial de las Condiciones de Vida para el Departamento de  
Santander

Dayanna Lucía Gámez González

Trabajo de Grado para Optar al Título de Matemática

Director

Tulia Esther Rivera Flórez

Magister en Estadística

Codirector

Fabio Humberto Sepúlveda Murillo

Doctor en Ciencias Económicas

Universidad Industrial de Santander

Facultad de Ciencias

Escuela de Matemáticas

Matemáticas

Bucaramanga

2024

### **Dedicatoria**

Me enorgullece dedicar este trabajo y logro a las mujeres que han trascendido en mi vida, Martha Lucía, Natalia Fernanda, María Camila y aunque ya no esté conmigo Lady Johanna. A ustedes debo las enseñanzas que han quedado en lo más profundo de mi alma, donde han echado raíces y han moldeado la persona que soy. Gracias por acompañarme en este camino que llamamos *vida*.

### **Agradecimientos**

Agradezco a mis maestros de estos años, especialmente a Tulia y Fabio por transmitirme sus experiencias, orientación y apoyo. Gracias por sembrar en mí enseñanzas que van más allá de lo profesional.

A mis padres, hermanas y hermano, quienes con su amor y esfuerzo me han permitido alcanzar un sueño profesional, gracias por inculcar en mí un ejemplo de resiliencia y apoyarme aún en la distancia.

A Daimon, por contagiarme cada día de su alegría y su infinita paciencia, por apapachar mi corazón y por ser el mejor compañero de vida que se puede tener.

A mis amigos con quienes mutuamente nos acompañamos e impulsamos a llegar siempre un poco más lejos. Gracias por compartir risas y llantos conmigo.

## Tabla de Contenido

	Pág.
Introducción .....	9
1. Antecedentes .....	14
2. Planteamiento del problema.....	16
3. Justificación .....	18
4. Objetivos.....	19
4.1 Objetivo General.....	19
4.2 Objetivos Específicos.....	19
5. Marco Teórico.....	20
5.1 Conceptos Algebraicos .....	20
5.2 Análisis de Componentes Principales.....	21
5.2.1 Matriz de covarianzas y correlaciones.....	22
5.2.2 Cálculo de componentes principales.....	25
5.2.3 Criterios para la selección del número de componentes.....	30
5.3 Medidas de Distancia.....	31
5.3.1 Distancia P2 .....	32
5.4 Condiciones para un indicador sintético.....	34
5.5 Estadística Espacial.....	37
5.5.1 Autocorrelación Espacial.....	38
5.5.2 Índice Global de Moran .....	40
5.5.3 Diagrama de dispersión de Moran .....	42
5.5.4 Indicadores Locales de Autocorrelación Espacial.....	43

6. Metodología .....	47
6.1 Descripción de la muestra.....	47
6.2 Métodos.....	48
6.3 Software .....	50
7. Resultados .....	54
7.1 Análisis descriptivo.....	54
7.2 Análisis de Componentes Principales.....	56
7.2.1 Estandarización .....	56
7.2.2 Correlaciones .....	58
7.2.3 Resultados del Análisis de Componentes Principales (PCA) .....	60
7.2.3.1 Distribución espacial de las variables relevantes para el PCA .....	64
7.2.3.2 Construcción del Índice de Calidad de Vida por PCA ( <i>ICVPCA</i> ) .....	72
7.3 Índice de Calidad de Vida por Distancia P2 ( <i>ICVP2</i> ) .....	76
7.4 Índice de Moran .....	81
7.5 Indicadores Locales de Autocorrelación Espacial ( <i>LISA</i> ).....	83
8. Discusión.....	86
9. Conclusiones.....	97
Referencias Bibliográficas .....	99
Apéndices.....	105

## Listas de Tablas

	Pág.
Tabla 1. Varibales disponibles para construir los índices.....	48
Tabla 2. Descripción de las variables seleccionadas para construir los índice.....	56
Tabla 3. Resumen de estadísticos del PCA.....	60
Tabla 4. Ranking $ICV_{PCA}$ .....	73
Tabla 5. Ranking $ICV_{P2}$ .....	78

**Listas de Figuras**

	Pág.
Figura 1. Tipos de autocorrelación espacial .....	39
Figura 2. Diagrama de dispersión de Moran.....	43
Figura 3. Mapa de significancia LISA .....	45
Figura 4. Mapa clúster LISA .....	46
Figura 5. Biplot PC1 vs PC2.....	62
Figura 6. Box-plots para las variables relevantes en el análisis por componentes principales.....	64
Figura 7. Distribución espacial Bajaedu .....	66
Figura 8. Distribución espacial Internet.....	67
Figura 9. Distribución espacial Basuras.....	68
Figura 10. Distribución espacial Proporción .....	69
Figura 11. Distribución espacial Jóvenes.....	71
Figura 12. Distribución espacial por cuartiles del ICVPCA .....	74
Figura 13. Distribución espacial por cuartiles del ICVP2 .....	79
Figura 14. Test Índice de Moran I para ICVPCA.....	81
Figura 15. Test Índice de Moran I para ICVP2 .....	81
Figura 16. Diagrama de dispersión de Moran para ICVPCA e ICVP2.....	82
Figura 17. Clústers LISA para ICVPCA e ICVP2 .....	83
Figura 18. Mapa de Calor de LISA para ICVPCA e ICVP2.....	84
Figura 19. Mapa de Significancia de LISA para ICVPCA e ICVP2 .....	85
Figura 20. Interfaz del Visualizador del ICV.....	87

**Listas de Apéndices**

	Pág.
Apéndice A. Variables disponibles y su descripción.....	105
Apéndice B. Matriz de correlaciones para las variables seleccionadas .....	106
Apéndice C. Puntaje de las variables para las primeras dos componentes .....	107
Apéndice D. Código en R-Studio .....	108

## Glosario

**Índice de Calidad de Vida (ICV):** herramienta que mide el nivel de bienestar de una población en términos de factores como salud, educación, acceso a servicios públicos y condiciones de vivienda.

**Análisis de Componentes Principales (PCA):** método estadístico que transforma un conjunto de variables potencialmente correlacionadas en un nuevo conjunto de variables no correlacionadas, llamadas componentes principales. Estas componentes se obtienen mediante combinaciones lineales de las variables originales y no están correlacionadas entre sí. El PCA es útil para identificar factores principales que explican las condiciones de vida en diferentes áreas.

**Distancia Pena Trapero:** medida propuesta por el profesor José M. Peña Trapero para evaluar la "distancia" entre diferentes observaciones o regiones en función de sus características de calidad de vida, tomando un punto de referencia.

## Resumen

**Título:** Análisis de la Variabilidad Espacial de las Condiciones de Vida para el Departamento de Santander\*

**Autor:** Dayanna Lucía Gámez González\*\*

**Palabras Clave:** Calidad de vida, PCA, P2, LISA, Moran

**Descripción:** La calidad de vida, según la Organización Mundial de la Salud (OMS), es la percepción de una persona sobre su posición en el contexto cultural y social en el que vive, en relación con sus metas y expectativas. En Colombia, Santander ocupaba en 2018 el sexto lugar en densidad poblacional y obtuvo una calificación de 8,39 en satisfacción general con la vida, según el DANE, superando el promedio nacional. No obstante, los análisis de calidad de vida en la región se enfocan mayormente en Bucaramanga y su área metropolitana, omitiendo otros municipios. Este estudio cuantitativo, de carácter descriptivo y multivariado, analizó la calidad de vida en Santander utilizando dos técnicas: el Análisis de Componentes Principales (PCA) y la distancia P2 de Pena Trapero, con el objetivo de construir índices sintéticos que permitieran comparar el bienestar entre territorios. Ambos métodos mostraron que los municipios del área metropolitana de Bucaramanga presentan mejores condiciones de vida, seguidos de San Gil y Socorro. Sin embargo, el Índice de Calidad de Vida por P2, ofrece una evaluación más estricta y realista de la calidad de vida en la región al tener a Bucaramanga como referencia. El análisis se complementó con el Índice de Moran y los Indicadores Locales de Autocorrelación Espacial (LISA) para un enfoque espacial detallado.

---

\* Trabajo de Grado

\*\* Facultad de Ciencias. Escuela de Matemáticas. Matemáticas. Director: Tulía Esther Rivera Flórez. Magister en Estadística. Codirector: Fabio Humberto Sepúlveda Murillo. Doctor en Ciencias Económicas.

**Abstract**

**Title:** Spatial Variability Analysis of Living Conditions for the Department of Santander\*

**Author:** Dayanna Lucía Gámez González\*\*

**Key Words:** Quality of Life, PCA, P2, LISA, Moran

**Description:** According to the World Health Organization (WHO), quality of life is a person's perception of their position in the cultural and social context in which they live, relative to their goals and expectations. In Colombia, Santander ranked sixth in population density in 2018 and scored 8.39 in overall life satisfaction, according to DANE, surpassing the national average. However, quality of life analyses in the region are primarily focused on Bucaramanga and its metropolitan area, often overlooking other municipalities. This quantitative, descriptive, and multivariate study analyzed quality of life in Santander using two techniques: Principal Component Analysis (PCA) and the P2 distance by Pena Trapero, aiming to build synthetic indices for comparing well-being across territories. Both methods showed that the municipalities in the metropolitan area of Bucaramanga have better living conditions, followed by San Gil and Socorro. However, the Quality of Life Index based on P2 provides a stricter and more realistic assessment of quality of life in the region by using Bucaramanga as a reference. The analysis was complemented by Moran's Index and Local Indicators of Spatial Association (LISA) for a detailed spatial approach.

---

\*Degree Work

\*\* Faculty of Sciences. School of Mathematics. Mathematics. Director: Tulia Esther Rivera Flórez. Master in Statistics. Co-director: Fabio Humberto Sepúlveda Murillo. Doctor in Economic Sciences.

## Introducción

Según la Organización Mundial de la Salud (OMS), la calidad de vida se entiende como “... la percepción que tiene una persona sobre su posición dentro del contexto cultural y el sistema de valores en el que vive, con respecto a sus metas, expectativas, estándares y preocupaciones.” Además, la Organización de las Naciones Unidas (ONU) establece que la calidad de vida está directamente relacionada con la sostenibilidad, la disposición de bienes y servicios, la conectividad a través de vías, el acceso a tecnologías de información y la equidad de género. De estos planteamientos se entiende que la calidad de vida va más allá de la salud física, como era habitual asociarlo. Por lo tanto, garantizar y mejorar la calidad de vida resulta fundamental para alcanzar el desarrollo integral y satisfactorio de cada individuo. Desde 1948 cuando la OMS amplió el concepto, los instrumentos, las dimensiones consideradas, las escalas e índices vienen cambiando, constituyéndose en un tema de relevancia actual.

En Colombia, desde 1986 y con el acompañamiento de Naciones Unidas (UN), el Departamento Nacional de Planeación y UNICEF, el Departamento Administrativo Nacional de Estadística (DANE) inició el trabajo en procura de definir los métodos para la medición de la pobreza, posteriormente vio la necesidad de involucrar otras dimensiones del bienestar como la salud, educación, actividades laborales, tenencia de bienes en el hogar, entre otras que constituyeron la base de lo que hoy se conoce como *Encuesta de Calidad de Vida*, un instrumento que sigue aplicándose con algunas variaciones en cuanto a representatividad y dimensiones que incluye.

De acuerdo con el DANE, en el año 2018 Santander ocupaba la sexta posición en términos de densidad poblacional entre los departamentos de Colombia y en la Encuesta Nacional de Calidad de Vida de ese mismo año, obtuvo una puntuación de 8,39 frente al promedio nacional de 8,26 en cuanto a la calificación de la satisfacción con la vida en general en personas de 15 años y más. Hasta ahora, los análisis sobre la calidad de vida a nivel municipal en Santander se han enfocado principalmente en su capital, Bucaramanga y en ocasiones en su zona metropolitana, Floridablanca, Piedecuesta y Girón, esto significa que solo se ha considerado un pequeño porcentaje, aproximadamente el 4,6%, de los municipios que conforman el departamento.

Con base en lo anterior, este proyecto propone construir un índice sintético compuesto que permita analizar la variabilidad espacial de las discrepancias en el bienestar social para los municipios del departamento de Santander a través de una metodología que combina métodos multivariados y elementos de estadística espacial e involucrando 32 variables agrupadas en nueve dimensiones: servicios de las viviendas, calidad de vida, socio economía, vulnerabilidad, salud, educación, demografía, economía y seguridad y protección, esto demanda mapear las variables relevantes que ayudan a la caracterización del bienestar social de los municipios y finalmente, desarrollar una representación geográfica interactiva que permita visualizar la zona de estudio y la distribución espacial del índice.

## 1. Antecedentes

Como ya se mencionó en este documento, el DANE aplica e informa de manera periódica la Encuesta de Calidad de Vida, en ella se hacen reportes a nivel nacional y a nivel regional (Bogotá, Antioquia, Valle del Cauca, Caribe, Pacífica, Central y Oriental, San Andrés y Providencia, Buenaventura, Orinoquía y Amazonía) \*\*. En el informe técnico se detalla información sobre 40 índices agrupados en las siguientes categorías: Conformación de hogares, Servicios del hogar, Tenencia de bienes y servicios, Percepción sobre ingreso y seguridad, Salud, Atención integral de niños menores de 5 años, Educación, Tecnologías de información y Comunicación.

En la búsqueda de antecedentes relacionados con estudios sobre la calidad de vida urbana en Colombia, nos encontramos con el Instituto de Estudios Urbanos de la Universidad Nacional de Colombia, creado en el 2005, en el que participan Facultades de Artes, Ciencias Económicas, Ciencias Humanas, Derecho, Ciencias Políticas y Sociales y entre sus objetivos se destacan “Contribuir en la comprensión de las dinámicas urbanas colombianas; analizar, evaluar, hacer seguimiento, sugerir y conceptuar sobre las políticas urbana y regional y respecto a las políticas en municipios específicos, así como en general sobre las políticas urbanas y de descentralización en Colombia”.

---

\*\* Departamento Administrativo Nacional de Estadística. (2024). Encuesta de Calidad de Vida. Recuperado de <https://www.dane.gov.co/index.php/comunicados-y-boletines/estadisticas-sociales/calidad-de-vida>

Además, nos proporciona un proyecto de extensión iniciado en Septiembre del 2010 y a la fecha finalizado, titulado Indicadores de Calidad de Vida Urbana en el cual se tuvo como objeto “Desarrollar un documento conceptual y metodológico que contenga los criterios y una batería mínima de indicadores que puedan ser aplicados en el mayor número posible de ciudades que permita realizar comparaciones de las condiciones de calidad de vida urbana en cada una, dentro del marco de Ciudades Justas y Sostenibles de la Red.”

Adicionalmente, encontramos la Red Colombiana de Ciudades Cómo Vamos (RCCV), creada oficialmente en el 2013, que actualmente cuenta con 20 ciudades de Colombia activas en el programa, y cuyo propósito es generar información confiable, imparcial y comparable en torno a temas de calidad de vida urbana y participación ciudadana, enfocándose en la evaluación de la calidad de vida de las ciudades a partir de indicadores técnicos, de percepción y del seguimiento a los resultados de la gestión pública. Además, se incluye la participación de expertos, administradores, académicos, organizaciones sociales y comunitarias, y ciudadanos, entre otros actores del desarrollo.

En estudios de casos particulares a nivel nacional, realizados por otras entidades, encontramos la investigación de los profesores, Sepúlveda, Chica y Soto (2019), cuyo objetivo principal fue estudiar la calidad de vida de los hogares de Medellín a partir de un indicador construido por medio de variables tanto objetivas como subjetivas y teniendo en cuenta la ubicación geográfica de los hogares y técnicas de análisis multivariado como componentes principales convencional y análisis de componentes principales ponderados geográficamente.

## 2. Planteamiento del problema

Santander es considerado una entidad territorial importante en Colombia debido a una serie de factores clave. Su ubicación geográfica le permite ser un punto de conexión entre varias regiones del país. Su economía diversificada, que incluye sectores como la agricultura, la ganadería, la minería, la industria manufacturera, entre otros, desempeña un papel fundamental en el desarrollo económico a nivel nacional. No se puede pasar por alto el valioso legado histórico de Santander en la formación de Colombia como nación. Hoy día, goza de gran popularidad por sus atractivos turísticos, por los servicios médicos de alta calidad y por su destacada contribución en el ámbito académico, especialmente en la investigación científica y tecnológica, lo que impulsa el avance y la innovación a nivel nacional. Estos factores, en conjunto, resaltan la importancia y el impacto significativo que Santander tiene en el contexto colombiano y subrayan su relevancia en múltiples aspectos que intervienen en el progreso del país.

No obstante, al abordar la relevancia de la calidad de vida en el desarrollo regional en primera instancia se encuentran que los análisis se centran en Bucaramanga y municipios de la zona metropolitana, por ejemplo el Informe de Calidad de Vida elaborado por el proyecto Bucaramanga Cómo Vamos<sup>††</sup>, también existe una carencia de estudios que logre una integración completa de componentes fundamentales como servicios de las viviendas, calidad de vida,

---

<sup>††</sup> Bucaramanga Cómo Vamos. (2024). Nuestra historia. Recuperado de <https://www.bucaramangacomovamos.org/our-story>

socioeconomía, vulnerabilidad, salud, educación, demografía, economía y seguridad y protección a nivel municipal en Santander.

Finalmente, la falta de análisis espaciales puede ocultar las necesidades y desigualdades existentes entre los municipios que requieren de atención, lo que resalta la necesidad de un análisis localizado y detallado que proporcione información relevante sobre aspectos a mejorar en cada municipio y promovería el crecimiento integral del departamento.

### **3. Justificación**

La importancia de considerar la calidad de vida en un municipio radica en diversas razones que impactan el bienestar individual y el desarrollo colectivo de la comunidad puesto que evaluar y mejorar la calidad de vida es esencial para asegurar que las personas tengan acceso a condiciones de vida dignas, proporcionar una base sólida para diseñar políticas públicas y plantear programas de desarrollo que incluyan las necesidades y prioridades reales de la población, reducir la desigualdad focalizando los recursos y esfuerzos en las áreas más necesitadas, impactar positivamente en la salud mental y emocional de los individuos, diseñar espacios habitables seguros y propicios para la convivencia, estimular el crecimiento económico de una región al ser atractivo para inversionistas, entre otros aspectos.

El desarrollo de este proyecto se basó en analizar la información disponible sobre calidad de vida en los municipios de Santander considerando dimensiones esenciales, con el fin de proporcionar información clara de los puntos fuertes y las debilidades en cada zona, los resultados podrán sustentar decisiones de política pública orientadas a resolver desigualdades territoriales y fomentar el progreso uniforme en el departamento. Además, se espera que esta investigación sirva como insumo para promover estudios futuros en estos temas, con lo cual sería un aporte valioso de la Universidad Industrial de Santander para nuestra región.

## 4. Objetivos

### 4.1 Objetivo General

Construir un índice sintético compuesto que permita analizar la variabilidad espacial de las discrepancias en bienestar social para los municipios del departamento de Santander.

### 4.2 Objetivos Específicos

- i. Describir y analizar las variables que serán usadas para la construcción del índice.
- ii. Construir dos índices, uno a partir de análisis de componentes principales y otro por medio de la distancia  $DP_2$ .
- iii. Mapear las variables relevantes que ayudan a la caracterización del bienestar social de los municipios.
- iv. Escribir un código computacional para la construcción de un geo portal que permita la visualización de la zona de estudio y de la distribución espacial del índice.

## 5. Marco Teórico

### 5.1 Conceptos Algebraicos

A continuación, se presenta un compendio de las definiciones y propiedades que constituyen la base teórica de algebra matricial que requieren las herramientas de análisis a implementar en este proyecto. Las siguientes definiciones y resultados han sido extraídos de *Applied Multivariate Statistical Analysis* (Johnson & Wichern, 1988, pp. 80-100).

**Definición 5.1.1** Sea  $A = \{a_{ij}\}$  una matriz  $k \times k$ . La traza de la matriz A, denotada  $\text{tr}(A)$  es la suma de los elementos en su diagonal, esto es  $\text{tr}(A) = \sum_{i=1}^k a_{ii}$ .

**Resultado 5.1.2** Sea A y B dos matrices  $k \times k$  y c un escalar, por las propiedades de sumatoria se satisface que:

- i.  $\text{tr}(cA) = c \text{tr}(A)$
- ii.  $\text{tr}(A \pm B) = \text{tr}(A) \pm \text{tr}(B)$
- iii.  $\text{tr}(AB) = \text{tr}(BA)$
- iv.  $\text{tr}(B^{-1}AB) = \text{tr}(A)$
- v.  $\text{tr}(AA') = \sum_{i=1}^k \sum_{j=1}^k a_{ij}^2$

**Definición 5.1.3** La descomposición espectral permite expresar una matriz cuadrada en términos de sus valores y vectores propios. Sea A una matriz  $k \times k$  definida positiva con la descomposición espectral. Entonces  $A = \sum_{i=1}^k \lambda_i e_i e_i'$ . Sea P una matriz  $k \times k$  que tiene a los vectores propios de A en sus columnas,  $P = [e_1, e_2, \dots, e_k]$ .

Entonces

$$A = \sum_{i=1}^k \lambda_i e_i e_i' = P \Lambda P'$$

Donde  $PP' = P'P = I$  y  $\Lambda$  es matriz diagonal. Luego  $\Lambda$  tiene la siguiente forma.

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{bmatrix} \quad \text{con } \lambda_i > 0$$

**Lema 5.1.4** Sea  $B$  una matriz  $p \times p$  definida positiva con valores propios  $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$  y los vectores propios asociados normalizados  $e_1, e_2, \dots, e_p$ . Entonces,

1.  $\max_{x \neq 0} \frac{x' B x}{x' x} = \lambda_1$  (alcanzado cuando  $x = e_1$ )
2.  $\min_{x \neq 0} \frac{x' B x}{x' x} = \lambda_p$  (alcanzado cuando  $x = e_p$ )

Además,

$$3. \max_{x \perp e_1, \dots, e_k} \frac{x' B x}{x' x} = \lambda_{k+1} \quad k = 1, 2, \dots, p-1$$

## 5.2 Análisis de Componentes Principales

En el proceso de comprender el comportamiento de un fenómeno en una población siguiendo un enfoque estadístico implica descomponerlo y estudiarlo a través de variables, analizar su relación conduce a determinar cuáles están correlacionadas entre sí y cuáles no; en este sentido, hoy día es cada vez más probable encontrarse con una gran cantidad de información la cual no necesariamente es sinónimo de calidad porque el número de variables puede ser innecesariamente

alto o contener información redundante, con lo cual se hace necesario optimizar la cantidad de variables con las que se va a adelantar un análisis estadístico.

Dentro de las técnicas estadística disponibles para reducir la dimensionalidad de un análisis multivariado está el análisis de componentes principales, esta técnica permite describir un conjunto de variables a través de nuevas variables construidas a partir de las variables originales que tienen la particularidad de ser incorrelacionadas entre sí y que podrían explicar en un alto porcentaje la variabilidad exhibida en la muestra de datos original. En la siguiente sección se presentarán los fundamentos teóricos que sustentan un análisis de componentes principales.

### 5.2.1 Matriz de covarianzas y correlaciones

Consideremos el vector aleatorio  $\mathbf{X}' = [X_1, X_2, \dots, X_p]$ , donde cada elemento de  $X$  es una variable aleatoria con su propia distribución de probabilidad marginal. Las siguientes definiciones han sido extraídas del libro *Probabilidad y estadística para ingenieros* (Johnson, 2011, pp. 119-125).

**Definición 5.2.1.1** La media y varianza marginal de cada variable aleatoria de  $\mathbf{X}$  están definidas como  $\mu_i = E(X_i)$  y  $\sigma_i^2 = E(X_i - \mu_i)^2$ ,  $i = 1, 2, \dots, p$ , respectivamente.

**Definición 5.2.1.2** La covarianza es el valor que indica el grado de variación conjunta que tienen dos variables aleatorias  $X_i$  y  $X_k$ .

Los posibles valores que puede tomar la covarianza entre dos variables aleatorias son:

$Cov(X_i, X_k) > 0$  existe una relación positiva entre las variables.

$Cov(X_i, X_k) < 0$  existe una relación negativa entre las variables.

$Cov(X_i, X_k) = 0$  no hay relación entre las variables.

Las medias y covarianzas de las variables de  $\mathbf{X}$  se pueden expresar como matrices, donde el valor esperado de cada variable está contenido en el vector de medias  $\boldsymbol{\mu} = E(\mathbf{X})$  y las  $p$  varianzas  $\sigma_{ii}$  y  $\frac{p(p-1)}{2}$  distintas covarianzas  $\sigma_{ik}$  con  $i < k$ , estarán contenidas en la matriz de varianzas y covarianzas  $\boldsymbol{\Sigma} = E(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'$ , específicamente

$$E(\mathbf{X}) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \boldsymbol{\mu}$$

y

$$\begin{aligned} \boldsymbol{\Sigma} &= E(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})' = E \left( \begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_p - \mu_p \end{bmatrix} [X_1 - \mu_1, X_2 - \mu_2, \dots, X_p - \mu_p] \right) \\ &= E \begin{bmatrix} (X_1 - \mu_1)^2 & (X_1 - \mu_1)(X_2 - \mu_2) & \cdots & (X_1 - \mu_1)(X_p - \mu_p) \\ (X_2 - \mu_2)(X_1 - \mu_1) & (X_2 - \mu_2)^2 & \cdots & (X_2 - \mu_2)(X_p - \mu_p) \\ \vdots & \vdots & \ddots & \vdots \\ (X_p - \mu_p)(X_1 - \mu_1)(X_p - \mu_p)(X_2 - \mu_2) & \cdots & & (X_p - \mu_p)^2 \end{bmatrix} \\ &= \begin{bmatrix} E(X_1 - \mu_1)^2 & E(X_1 - \mu_1)(X_2 - \mu_2) & \cdots & E(X_1 - \mu_1)(X_p - \mu_p) \\ E(X_2 - \mu_2)(X_1 - \mu_1) & E(X_2 - \mu_2)^2 & \cdots & E(X_2 - \mu_2)(X_p - \mu_p) \\ \vdots & \vdots & \ddots & \vdots \\ E(X_p - \mu_p)(X_1 - \mu_1)E(X_p - \mu_p)(X_2 - \mu_2) & \cdots & & E(X_p - \mu_p)^2 \end{bmatrix} \end{aligned}$$

Que con la notación usual se tendría expresada como

$$\mathbf{\Sigma} = Cov(\mathbf{X}) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix} \quad (5.1)$$

Note que  $\sigma_{ik} = E(X_i - \mu_i)(X_k - \mu_k) = \sigma_{ki}$ . Se debe notar aquí que la matriz de varianzas y covarianzas es simétrica y por tanto, se puede escribir como

$$\mathbf{\Sigma} = Cov(\mathbf{X}) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp} \end{bmatrix} \quad (5.2)$$

**Definición 5.2.1.3** El coeficiente de correlación  $\rho_{ik}$  mide la asociación lineal entre dos variables aleatorias  $X_i$  y  $X_k$ , está definido en términos de la covarianza  $\sigma_{ik}$  y varianzas  $\sigma_{ii}$  y  $\sigma_{kk}$  así

$$\rho_{ik} = \frac{\sigma_{ik}}{\sqrt{\sigma_{ii}\sigma_{kk}}} \quad (5.3)$$

También se tiene que la matriz  $p \times p$  de correlación está dada por una matriz simétrica que en forma general se expresa como:

$$\boldsymbol{\rho} = \begin{bmatrix} \frac{\sigma_{11}}{\sqrt{\sigma_{11}\sqrt{\sigma_{11}}} & \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sqrt{\sigma_{22}}} & \cdots & \frac{\sigma_{1p}}{\sqrt{\sigma_{11}\sqrt{\sigma_{pp}}} \\ \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sqrt{\sigma_{22}}} & \frac{\sigma_{22}}{\sqrt{\sigma_{22}\sqrt{\sigma_{22}}} & \cdots & \frac{\sigma_{2p}}{\sqrt{\sigma_{22}\sqrt{\sigma_{pp}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{1p}}{\sqrt{\sigma_{11}\sqrt{\sigma_{pp}}} & \frac{\sigma_{2p}}{\sqrt{\sigma_{22}\sqrt{\sigma_{pp}}} & \cdots & \frac{\sigma_{pp}}{\sqrt{\sigma_{pp}\sqrt{\sigma_{pp}}} \end{bmatrix} \quad (5.3)$$

$$\boldsymbol{\rho} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{12} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1p} & \rho_{2p} & \cdots & 1 \end{bmatrix} \quad (5.4)$$

Así la matriz  $p \times p$  de desviaciones estándar está dada por:

$$\mathbf{v}^{1/2} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{\sigma_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\sigma_{pp}} \end{bmatrix} \quad (5.5)$$

Además,  $\boldsymbol{\Sigma}$  se puede calcular a partir de  $\mathbf{v}^{1/2}$  y  $\boldsymbol{\rho}$ , mientras que  $\boldsymbol{\rho}$  se puede obtener de  $\boldsymbol{\Sigma}$  a través de las siguientes igualdades:

$$\mathbf{v}^{1/2} \boldsymbol{\rho} \mathbf{v}^{1/2} = \boldsymbol{\Sigma} \quad (5.6)$$

$$\boldsymbol{\rho} = (\mathbf{v}^{1/2})^{-1} \boldsymbol{\Sigma} (\mathbf{v}^{1/2})^{-1} \quad (5.7)$$

### 5.2.2 Cálculo de componentes principales

El Análisis de Componentes Principales (PCA) es una técnica estadística multivariada que se utiliza para reducir el número de variables en un conjunto de datos en un número más pequeño número de “dimensiones”. En términos matemáticos, de un conjunto inicial de  $n$  variables correlacionadas, PCA crea índices o componentes no correlacionados, donde cada componente es una combinación lineal ponderada de variables iniciales (Vyas & Kumaranayake, 2006).

En términos geométricos, el análisis de componentes principales se reduce a encontrar el espacio de menor dimensión que represente adecuadamente los datos, es decir, deseamos encontrar un subespacio de dimensión menor que  $p$  de forma que al hacer la proyección de los datos sobre él se conserve la estructura original con la menor distorsión posible. Las siguientes definiciones,

teoremas y demostraciones han sido extraídas de *Applied Multivariate Statistical Analysis* (Johnson & Wichern, 1988, pp. 430-434).

**Definición 5.2.2.1** Sea el vector aleatorio  $\mathbf{X}' = [X_1, X_2, \dots, X_p]$  con matriz de covarianza  $\mathbf{\Sigma}$ , media  $\boldsymbol{\mu}$  y valores propios  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ . Considere las combinaciones lineales

$$Y_1 = \mathbf{a}'_1 \mathbf{X} = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p$$

$$Y_2 = \mathbf{a}'_2 \mathbf{X} = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p$$

⋮

$$Y_p = \mathbf{a}'_p \mathbf{X} = a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p$$

Las cuales se pueden expresar como

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_p \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pp} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} = \mathbf{C}\mathbf{X}$$

Con

$$\boldsymbol{\mu}_Y = E(\mathbf{Y}) = E(\mathbf{C}\mathbf{X}) = \mathbf{C}\boldsymbol{\mu}_X \tag{5.8}$$

$$\mathbf{\Sigma}_Y = Cov(\mathbf{Y}) = Cov(\mathbf{C}\mathbf{X}) = \mathbf{C}\mathbf{\Sigma}_X\mathbf{C}' \tag{5.9}$$

Donde  $\boldsymbol{\mu}_X$  y  $\mathbf{\Sigma}_X$  son el vector de la media y la matriz de covarianza de  $\mathbf{X}$ , respectivamente. Luego,

$$Var(Y_i) = \mathbf{a}'_i \mathbf{\Sigma} \mathbf{a}_i \tag{5.10}$$

$$Cov(Y_i, Y_k) = \mathbf{a}'_i \mathbf{\Sigma} \mathbf{a}_k \tag{5.11}$$

Donde  $i, k = 1, 2, \dots, p$

De esta forma, las componentes principales son combinaciones lineales no correlacionadas de  $X_1, X_2, \dots, X_p$  ordenadas de tal forma que el valor de sus varianzas, dadas por (5.11), vayan en forma descendente.

La primera componente principal es la combinación lineal con máxima varianza, es decir, maximizar  $Var(Y_i) = \mathbf{a}'_i \boldsymbol{\Sigma} \mathbf{a}_i$ , pero esta se puede incrementar multiplicando por cualquier constante  $a_i$ . Para quitar esta indeterminación, se restringe el vector de coeficientes, de tal forma que tenga norma unitaria. Así, la primera componente principal de  $\mathbf{X}$  será una combinación lineal de  $\mathbf{a}'_1 \mathbf{X}$  que maximiza  $Var(\mathbf{a}'_1 \mathbf{X})$  sujeto a  $\mathbf{a}'_1 \mathbf{a}_1 = 1$ .

La segunda componente principal será una combinación lineal  $\mathbf{a}'_2 \mathbf{X}$  que maximiza  $Var(\mathbf{a}'_2 \mathbf{X})$  sujeto a  $\mathbf{a}'_2 \mathbf{a}_2 = 1$  y  $Cov(\mathbf{a}'_1 \mathbf{X}, \mathbf{a}'_2 \mathbf{X})=0$  y la  $i$ -ésima componente principal estará definida como una combinación lineal  $\mathbf{a}'_i \mathbf{X}$  que maximiza  $Var(\mathbf{a}'_i \mathbf{X})$  sujeto a  $\mathbf{a}'_i \mathbf{a}_i = 1$  y  $Cov(\mathbf{a}'_i \mathbf{X}, \mathbf{a}'_k \mathbf{X})=0$  para  $k < i$ . Por consiguiente, tenemos el siguiente resultado.

**Teorema 5.2.2.2** Sea  $\boldsymbol{\Sigma}$  la matriz de covarianza asociada al vector aleatorio  $\mathbf{X}' = [X_1, X_2, \dots, X_p]$  y cuyos pares de valores y vectores propios están dados por  $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \dots, (\lambda_p, \mathbf{e}_p)$ , donde  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ . Entonces, el  $i$ -ésimo componente principal está dado por:

$$Y_i = \mathbf{e}'_i \mathbf{X} = e_{i1}X_1 + e_{i2}X_2 + \dots + e_{ip}X_p, \quad i = 1, 2, \dots, p$$

Y su varianza y covarianza están dadas por:

$$Var(Y_i) = \mathbf{e}'_i \boldsymbol{\Sigma} \mathbf{e}_i = \lambda_i, \quad i = 1, 2, \dots, p$$

$$Cov(Y_i, Y_k) = \mathbf{e}'_i \boldsymbol{\Sigma} \mathbf{e}_k = 0, \quad i \neq k$$

**Prueba**

En el ítem 1 del Lema 5.1.4 tome  $B = \Sigma$ , luego:

$$\max_{x \neq 0} \frac{x' \Sigma x}{x' x} = \lambda_1 \quad (\text{cuando } x = \mathbf{e}_1)$$

Pero  $\mathbf{e}'_1 \mathbf{e}_1 = 1$  ya que los vectores propios están normalizados. Por lo tanto,

$$\max_{x \neq 0} \frac{x' \Sigma x}{x' x} = \lambda_1 = \frac{(\mathbf{e}'_1 \Sigma \mathbf{e}_1)}{(\mathbf{e}'_1 \mathbf{e}_1)} = \mathbf{e}'_1 \Sigma \mathbf{e}_1 = \text{Var}(Y_1)$$

De manera similar, usando el ítem 2 del Lema 5.1.4, obtenemos:

$$\max_{x \perp \mathbf{e}_1, \dots, \mathbf{e}_k} (x' \Sigma x) (x' x) = \lambda_{\{k+1\}}, \quad k = 1, 2, \dots, p - 1$$

Para  $a = \mathbf{e}_{k+1}$ , con  $\mathbf{e}'_{k+1} \mathbf{e}_i = 0$ , para  $i = 1, 2, \dots, k$  y  $k = 1, 2, \dots, p - 1$ ,

$$\frac{\mathbf{e}'_{k+1} \Sigma \mathbf{e}_{k+1}}{\mathbf{e}'_{k+1} \mathbf{e}_{k+1}} = \mathbf{e}'_{k+1} \Sigma \mathbf{e}_{k+1} = \text{Var}(Y_{k+1})$$

Pero  $\mathbf{e}'_{k+1} (\Sigma \mathbf{e}_{k+1}) = \lambda_{k+1} \mathbf{e}'_{k+1} \mathbf{e}_{k+1} = \lambda_{k+1}$ , así  $\text{Var}(Y_{k+1}) = \lambda_{k+1}$ .

Veamos que  $\text{Cov}(Y_i, Y_k) = 0$ , basta ver que  $\mathbf{e}_i$  es perpendicular a  $\mathbf{e}_k$ . Los vectores propios de  $\Sigma$  son ortogonales si todos los autovalores  $\lambda_1, \lambda_2, \dots, \lambda_p$  son distintos, pero si algunos son iguales, entonces, los vectores propios correspondientes a los valores propios se pueden elegir para que sean ortogonales. Por lo tanto, para cualquiera dos valores propios  $\mathbf{e}_i$  y  $\mathbf{e}_k$  se tiene que  $\mathbf{e}'_i \mathbf{e}_k = 0$ . Como  $\Sigma \mathbf{e}_k = \lambda_k \mathbf{e}_k$ , entonces

$$\text{Cov}(Y_i, Y_k) = \mathbf{e}'_i \Sigma \mathbf{e}_k = \mathbf{e}'_i \lambda_k \mathbf{e}_k = \lambda_k \mathbf{e}'_i \mathbf{e}_k = 0 \quad \text{si } i \neq k.$$

**Teorema 5.2.2.3** Sea  $\mathbf{X}' = [X_1, X_2, \dots, X_p]$  un vector de variables aleatorias, con matriz de covarianzas  $\Sigma$  y pares de valores y vectores propios  $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \dots, (\lambda_p, \mathbf{e}_p)$  donde  $\lambda_1 \geq \lambda_2 \geq$

$\dots \geq \lambda_p \geq 0$ . Sean  $Y_1 = \mathbf{e}'_1 \mathbf{X}, Y_2 = \mathbf{e}'_2 \mathbf{X}, \dots, Y_p = \mathbf{e}'_p \mathbf{X}$  las componentes principales de  $\mathbf{X}$ .

Entonces

$$\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \sum_{i=1}^p \text{Var}(X_i) = \lambda_1 + \lambda_2 + \dots + \lambda_p = \sum_{i=1}^p \text{Var}(Y_i) \quad (5.12)$$

**Prueba**

Por la Definición 5.1,  $\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \text{tr}(\mathbf{\Sigma})$  y por la Definición 5.3, tomando  $\mathbf{A} = \mathbf{\Sigma}$ , podemos expresar  $\mathbf{\Sigma} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}'$  donde  $\mathbf{\Lambda}$  es la matriz diagonal de valores propios y  $\mathbf{P} = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_p]$  con  $\mathbf{P}\mathbf{P}' = \mathbf{P}'\mathbf{P} = \mathbf{I}$  y por el Resultado 5.2 (iii) tenemos

$$\text{tr}(\mathbf{\Sigma}) = \text{tr}(\mathbf{P}\mathbf{\Lambda}\mathbf{P}') = \text{tr}(\mathbf{\Lambda}\mathbf{P}'\mathbf{P}) = \text{tr}(\mathbf{\Lambda}) = \lambda_1 + \lambda_2 + \dots + \lambda_p$$

Así,

$$\sum_{i=1}^p \text{Var}(X_i) = \text{tr}(\mathbf{\Sigma}) = \text{tr}(\mathbf{\Lambda}) = \sum_{i=1}^p \text{Var}(Y_i)$$

**Teorema 5.2.2.4** Sean  $Y_1 = \mathbf{e}'_1 \mathbf{X}, Y_2 = \mathbf{e}'_2 \mathbf{X}, \dots, Y_p = \mathbf{e}'_p \mathbf{X}$  las componentes principales de  $\mathbf{X}$  calculados a partir de la matriz de covarianza  $\mathbf{\Sigma}$  y  $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \dots, (\lambda_p, \mathbf{e}_p)$  los pares de valores y vectores propios de  $\mathbf{\Sigma}$ , entonces

$$\rho_{Z_i, X_k} = \frac{e_{ik}\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}} \quad i, k = 1, 2, \dots, p \quad (5.13)$$

Son los coeficientes de correlación entre la componente  $Y_i$  y la variable  $X_k$ .

**Prueba**

Sea  $\mathbf{a}'_k = [0, \dots, 0, 1, 0, \dots, 0]$  vector de dimensión  $p$  nulo, excepto en la  $k$  –ésima posición.

Entonces,  $X_k = \mathbf{a}'_k \mathbf{X}$  y por (5.10)

$$Cov(X_k, Y_i) = Cov(\mathbf{a}'_k \mathbf{X}, \mathbf{e}'_i \mathbf{X}) = \mathbf{a}'_k \boldsymbol{\Sigma} \mathbf{e}_i$$

Donde  $\boldsymbol{\Sigma} \mathbf{e}_i = \lambda_i \mathbf{e}_i$ , por consiguiente

$$Cov(X_k, Y_i) = \mathbf{a}'_k \lambda_i \mathbf{e}_i = \lambda_i \mathbf{e}_{ik}$$

Y por (5.14)  $Var(Y_i) = \lambda_i$  y  $Var(X_k) = \sigma_{kk}$ , así

$$\rho_{Y_i, X_k} = \frac{Cov(Y_i, X_k)}{\sqrt{Var(Y_i)} \sqrt{Var(X_k)}} = \frac{\lambda_i \mathbf{e}_{ik}}{\sqrt{\lambda_i} \sqrt{\sigma_{kk}}} = \frac{\mathbf{e}_{ik} \sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}} \quad i, k = 1, 2, \dots, p.$$

### 5.2.3 Criterios para la selección del número de componentes

La pregunta sobre cuántos componentes conservar en un análisis de componentes principales no tiene una respuesta definitiva. Hay varios aspectos a considerar, como la cantidad total de varianza explicada, los tamaños relativos de los valores propios (que representan las varianzas de los componentes), y las interpretaciones temáticas de los componentes. Estos factores ayudan a decidir cuántos componentes son necesarios para representar adecuadamente los datos. A continuación, se presentan algunos criterios comunes para la selección de componentes principales.

- i. *Criterio de Kaiser*: se seleccionan las componentes principales cuyo autovalor  $\lambda_i$  sea mayor o igual a 1. (Kaiser, 1960, pp. 141).
- ii. *Contraste de caída*: por medio de un gráfico de sedimentación de los autovalores para cada componente en orden decreciente, se escogen las componentes hasta el punto en el que se evidencie un punto de inflexión, lo que indica que a partir de ese punto la proporción de varianza explicada compartida no es significativa (Jolliffe, 2002, pp. 115).

iii. *Porcentaje de varianza explicada*: radica en determinar el mayor porcentaje de varianza explicada por las componentes hasta llegar a un nivel mínimo (**60 – 70%**) en relación con el número de variables (Jolliffe, 2002, pp. 114).

Una vez definidos los componentes principales se estima el índice empleando un promedio ponderado de las puntuaciones de cada componente principal, ponderados por la raíz cuadrada de la varianza de cada componente (Peters, 1970). Así, el ICV por PCA se calcula como:

$$ICV_{PCA} = \frac{\sum_{i=1}^r Y_{rj} \cdot \sqrt{\lambda_r}}{\sum_{i=1}^r \sqrt{\lambda_r}} \quad (5.18)$$

Donde,  $Y_{rj}$  es la puntuación del componente  $r$ -ésimo para la observación  $j$ -ésima,  $\lambda_r$  es el valor propio para dicho componente y  $r$  es la cantidad de componentes seleccionados. De esta forma se garantiza que los componentes con una mayor varianza explicada tengan una mayor ponderación en la calificación del índice (Jaramillo, 2010, p. 119).

### 5.3 Medidas de Distancia

Las medidas de distancia constituyen un enfoque muy habitual para el diseño de indicadores sintéticos y económicos de coyuntura o medidas del desarrollo, la pobreza y la desigualdad. Como Pena (1977) y Zarzosa (1996) plantean, los indicadores sintéticos elaborados sobre medidas de distancia deben satisfacer las siguientes condiciones:

- i. No negatividad. La distancia es un real no negativo y toma el valor nulo cuando los dos vectores son iguales.

$$d(x, z) > 0 \quad \text{si } x \neq z$$

$$d(x, z) = 0 \Leftrightarrow x = z$$

- ii. Conmutatividad. La distancia se preserva sin importar la dirección.

$$d(x, z) = d(z, x)$$

- iii. Desigualdad triangular. Sean  $x, y, z$  tres vectores definidos en el mismo espacio vectorial.

$$d(x, z) \leq d(x, y) + d(y, z)$$

Estos indicadores pueden usarse para comparaciones entre distintos ámbitos y a lo largo del tiempo, dado que consiguen mantener el orden, es decir, la introducción de nuevos casos no supone la alteración del orden en los casos existentes, hecho que no ocurre en el análisis de componentes principales.

### 5.3.1 Distancia P2

El indicador sintético de distancia P2 es una medida diseñada por Pena (Pena 1977) específicamente para medir el bienestar social de un conjunto de unidades territoriales en un momento dado de tiempo, o de una unidad territorial en distintos momentos de tiempo. Este índice sintético se basa en el concepto de distancia elaborado por Pena Trapero a partir de una modificación de la distancia de Ivanovic (1974). Las siguientes definiciones fueron extraídas del libro *“El valor económico de la calidad ambiental urbana: Un análisis espacial para las comunas y barrios de Cali”* (Jaramillo, 2010, pp. 123-125).

**Definición 5.3.1.1** La distancia I ( $d_I$ ) de Ivanovic es usada para la elaboración de indicadores sintéticos de pobreza y aplicada como medida de disimilitud entre casos o a un caso estándar con relación a un conjunto de p variables incluidas en un vector X.

$$d_I(i, i^*) = \sum_{j=1}^p \frac{|x_{ij} - x_{i^*j}|}{\sigma_j} \prod_{i=1}^{j-1} (1 - r_{ij,1,2,\dots,i-1}), \quad i < j$$

Donde  $r_{ij,1,2,\dots,i-1}$  es el coeficiente de correlación parcial entre el componente i-ésimo y el j-ésimo,  $d_I$  es la medida que representa el valor absoluto de la diferencia entre el conjunto de indicadores ideales ( $x_{i^*j}$ ) con relación a un conjunto de p indicadores simples ( $x_{ij}$ ), tipificados por la inversa de la desviación estándar del indicador simple y corrigiendo la información redundante mediante la inclusión del coeficiente de correlación parcial.

**Definición 5.3.1.2** La distancia P2, desarrollada por Pena Trapero y basada en la distancia de Frechet, utiliza como factor de ponderaciones de las distancias estimadas el coeficiente de determinación  $R^2$  de la siguiente forma:

$$P2 = \sum_{i=1}^p \frac{|x_{ij} - x_{i^*j}|}{\sigma_i} (1 - R_{i,i-1,i-2,\dots,1}^2)$$

Donde  $d_i = |x_{ij} - x_{i^*j}|$  es la distancia del indicador simple de un país, región, ciudad, comuna, etc. Con respecto a un parámetro o base de referencia del indicador simple  $X_*$  y  $\sigma_i$  es la desviación estándar de los valores que toma el indicador simple i-ésimo.

El factor  $\frac{d_i}{\sigma_i}$  soluciona el problema de la heterogeneidad de las unidades de medida de cada indicador simple y actúa como ponderador que da mayor importancia a las distancias con valores de mayor dispersión con respecto a la media.

$R_{i,i-1,i-2,\dots,1}^2$  es el coeficiente de determinación de la regresión de  $x_i$  sobre  $x_{i-1}, x_{i-2}, \dots, x_1$ , este factor nos permite eliminar la información ya contenida en los indicadores simples precedentes (Zarzosa, 1996). Si  $R_1^2 = 0$  es porque la primera componente aporta toda la información al no existir un componente previo. Por esto, su ponderación es la unidad.

Para estimar la distancia P2 es necesario realizar una jerarquización previa de los componentes, aproximándose mediante un método interactivo que parte de la solución inicial hasta encontrar la solución de convergencia. La distancia P2 final para la situación o caso  $i$  respecto a la referencia  $i^*$  muestra el orden de inclusión correcto para los indicadores.

$$P2 = \frac{|x_{i1} - x_{i^*1}|}{\sigma_1} + \frac{|x_{i2} - x_{i^*2}|}{\sigma_2} (1 - R_{2,1}^2) + \dots \tag{5.19}$$

$$+ \frac{|x_{ip} - x_{i^*p}|}{\sigma_p} (1 - R_{p,p-1,p-2,\dots,1}^2)$$

Este indicador sintético de distancia ha sido utilizado para medir variables latentes como el bienestar, desarrollo sostenible, etc. entre distintas regiones de un país (Jaramillo, 2010, p. 123).

#### 5.4 Condiciones para un indicador sintético

Pena Trapero (1977), Zarzosa (1996) y Ebert (2004) presentan una serie de condiciones que a priori debe cumplir el índice sintético para que sea consistente.

- i. Existencia y determinación.* La función matemática que define el indicador sintético debe ser tal que exista y tenga solución para todos los valores del índice en cada unidad experimental.
- ii. Monotonía.* El índice debe responder positivamente a una modificación positiva de los componentes y negativamente a una modificación negativa. En la práctica, esto exige que en el análisis multivariante se introduzca la información con el signo con el que se espera estén relacionados los indicadores simples y el objetivo que se va a medir.
- iii. Unicidad.* El índice debe representar un único valor para una situación dada.
- iv. Invariancia.* Como consecuencia de la propiedad “iii”, el índice debe ser invariante respecto a un cambio de origen y/o de escala de medición de los indicadores simples.
- v. Homogeneidad.* La función matemática que defina el índice debe ser homogénea de grado uno ( $f(tI_1, tI_2, \dots, tI_n) = tf(I_1, I_2, \dots, I_n)$ ). Esto asegura que si los indicadores simples aumentan o disminuyen en igual proporción lo hace el indicador sintético.
- vi. Transitividad.* Suponga que a, b y c son tres estados distintos medidos por el índice, e  $I(a)$ ,  $I(b)$  e  $I(c)$  son valores del índice correspondiente a esos tres estados. Debe verificarse que:

$$\left. \begin{array}{l} I(a) > I(b) \\ I(b) > I(c) \end{array} \right\} \Rightarrow I(a) > I(c)$$

- vii. Exhaustividad.* El índice debe ser tal que aproveche al máximo y de forma útil la información suministrada por cada uno de los indicadores.

Johnson (2000) se refiere a la utilidad del PCA por las siguientes razones:

- a) Cribado de datos multivariantes. El PCA es útil para el análisis previo de cualquier tipo de análisis multivariantes, dado que puede ayudar a revelar anomalías en los datos y a descubrir datos atípicos (outliers).
- b) Agrupación de unidades experimentales. El PCA ayuda al investigador a detectar subgrupos de acuerdo con el comportamiento común de la varianza de las variables respuesta.
- c) Análisis discriminantes. En este tipo de análisis la matriz de varianzas-covarianza requiere ser invertida para realizar una regla de discriminación. Cuando el número de variables respuestas es superior al número de unidades experimentales, la matriz no se puede invertir. Allí juega un papel importante el PCA para reducir todas las variables respuestas a un número menor de componentes principales que expliquen el mayor porcentaje de varianza de los datos originales.
- d) Regresión. El PCA ayuda a determinar y corregir los problemas derivados de regresiones que presenten problemas de multicolinealidad.

Sin embargo, los indicadores basados en PCA no satisfacen totalmente las condiciones de

- a. *Homogeneidad.* Las medidas derivadas del PCA son ordinales, por lo tanto las estimaciones que se realicen cumplen esta propiedad, indicando sólo una comparación ordinal y no cardinal con respecto al valor ideal de los indicadores simples.
- b. *Transitividad.* El valor del índice puede variar si se incluye una unidad experimental adicional al conjunto de datos. Lo cual hace que este índice no sea universal, sino referido a la muestra de las unidades experimentales que se evalúan en el momento.

Por otra parte, los indicadores sintéticos elaborados con la distancia P2 cumplen con las siete condiciones presentadas anteriormente y una propiedad adicional, invarianza respecto de la base de referencia, es decir, los indicadores simples son tipificados.

Lo anterior indica que con base en las propiedades matemáticas que deben cumplir los indicadores sintéticos, la distancia P2 es más consistente que el PCA. Sin embargo, Pena Trapero argumenta la complementariedad de fases en los dos métodos (Jaramillo, 2010, p. 125). Dado que el objetivo de esta investigación es construir un índice sintético compuesto para analizar la variabilidad espacial de las discrepancias en bienestar social en los municipios de Santander, se utilizarán ambos métodos. Esto permitirá evaluar las discrepancias individuales de cada uno y la complementariedad entre ambos.

## **5.5 Estadística Espacial**

En su forma más simple, el espacio juega un importante papel en definir cuantos eventos discretos, personas u objetos existen en un área (Alegret, Herrera & Grau 2008, p. 3). En el cálculo de índices sintéticos para medir la calidad de vida en un área específica, está implícita la consideración del espacio. Una ventaja de utilizar estadísticas espaciales es que ellas son capaces de resumir el complejo patrón espacial, haciendo esta complejidad digerible para los ojos y la mente humana, lo que hace más factible la toma de decisiones y la intervención (Alegret, Herrera & Grau 2008, p. 3).

El término estadística espacial se refiere a un conjunto de técnicas estadísticas que cuantifican aspectos relacionados con la estructura de las distribuciones espaciales. La característica distintiva del análisis estadístico de datos espaciales es que el patrón espacial de las

localizaciones (objetos espaciales), la asociación espacial entre los valores observados en diferentes localizaciones (dependencia espacial) y la variación sistemática del fenómeno en las distintas localizaciones (heterogeneidad espacial) se convierte en el mayor foco de investigación (Alegret, Herrera & Grau 2008, p. 3).

### **5.5.1 Autocorrelación Espacial**

Según Celemín (2009) las características socioeconómicas y ambientales propias de la Geografía tienden a mostrar cierto grado de similitud, ya que a menos que existan factores de ruptura o de discontinuidad muy marcados, la situación normalmente esperable sería la de cierta homogeneidad espacial. En general se acepta que existe autocorrelación espacial (AE) siempre que haya una variación espacial sistemática en los valores de una variable a través de un mapa, es decir un patrón en el comportamiento de la variable según la ubicación geográfica del dato.

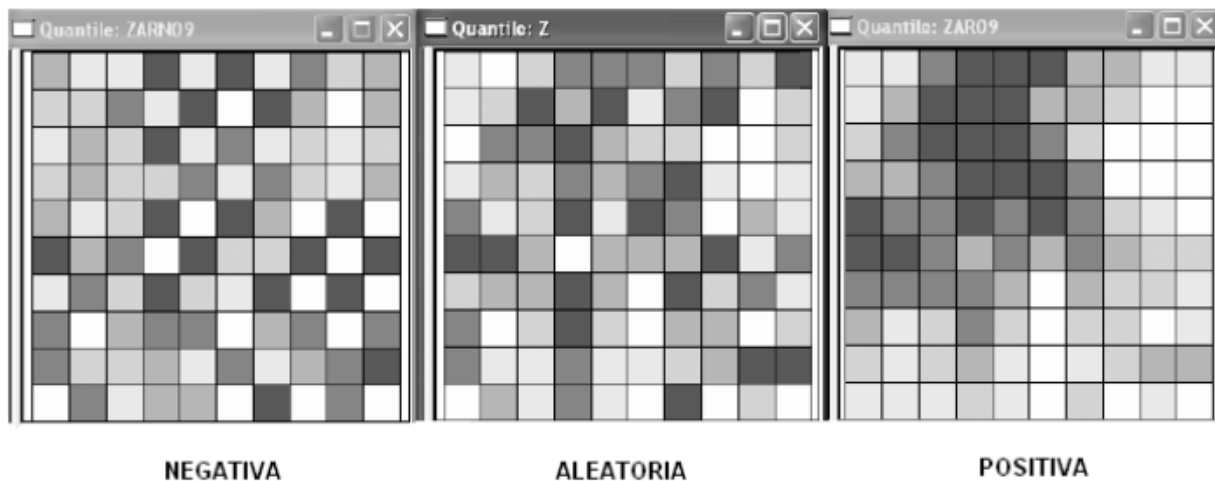
La autocorrelación o dependencia espacial se define como una relación funcional de lo que sucede entre dos puntos diferentes de un mismo espacio (Moreno y Vayá 2000). La AE es "especialmente indicada en fenómenos de propagación, es decir, que se adapten al modelo de difusión epidémica y en situaciones con un fuerte componente social, puesto que la población suele residir de un modo segregado" (Gamir Orueta, 1995: 216). En otras palabras, "la utilidad de la AE se encuentra en su capacidad para estudiar la forma en que un fenómeno se irradia a través de las unidades espaciales, y si tal conducta corresponde a algún modelo de difusión conocido o bien registra la segregación espacial de alguna característica. En definitiva, refleja el grado en que objetos o actividades en una unidad geográfica son similares a los objetos o actividades en unidades geográficas próximas" (Vilalta y Perdomo, 2005: 326)

La correlación de una misma variable en unidades espaciales diferentes y contiguas da lugar a tres posibilidades:

- i. Autocorrelación espacial positiva que indica una tendencia al agrupamiento de las unidades espaciales y se da cuando estas unidades vecinas presentan valores próximos.
- ii. Autocorrelación espacial negativa que indica una tendencia de dispersión en las unidades espaciales, producto de valores lejanos entre estas.
- iii. Sin autocorrelación, producida cuando las unidades espaciales presentan valores aleatorios y por tanto, no se puede detectar alguno de los casos anteriores.

**Figura 1.**

*Tipos de autocorrelación espacial*



*Nota.* La imagen muestra los 3 tipos de autocorrelación espacial que existen. Negativa, aleatoria y positiva. Tomada de: *Autocorrelación espacial e indicadores locales de asociación espacial. Importancia, estructura y aplicación* (Figura 1: Tipos de autocorrelación espacial), Celemín, 2009.

Existen diferentes tipos de índices que permiten medir la autocorrelación espacial y cuya estructura general es la siguiente

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

Donde  $n$  es el total de zonas en el espacio,  $W_{ij}$  son los elementos de la matriz de pesos espaciales, cuyos valores son una función de alguna medida de contigüidad en la matriz de los datos originales y el valor  $C_{ij}$  es una medida de la distancia (euclidiana, esférica, definida por el investigador, etc.) entre los valores  $i$  y  $j$  en alguna dimensión (Celemín, 2009, p. 3).

Los índices de autocorrelación espacial global evalúan asociaciones en el conjunto de valores del fenómeno de estudio, comparando de forma directa los valores de cada unidad de análisis con la media global del fenómeno ( $\bar{x}$ ). Sin embargo, no son sensibles a espacios que presentan heterogeneidad en el comportamiento del fenómeno geográfico (Siabato & Guzmán, 2019, p. 10).

### 5.5.2 Índice Global de Moran

El Índice Global de Moran es una medida estadística desarrollada por Alfred Pierce Moran (1950) que analiza de forma integral las variaciones de autocorrelación espacial entre valores vecinos más cercanos, los mismos que pueden clasificarse como positivo, negativo y sin autocorrelación espacial. Cuando los valores tienden a agruparse, se habla de una autocorrelación espacial positiva, pero si estos valores se dispersan, entonces se convierte en una autocorrelación negativa, y si los valores se encuentran dispersos o distribuidos de forma aleatoria, entonces no hay autocorrelación espacial entre los valores analizados (Hidalgo, 2018, p. 97)

El Índice Global de Moran está basado en la siguiente expresión analítica:

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x}) \bar{x}}{\sum_{i=2}^n (x_i - \bar{x})^2} \quad \text{para } i \neq j \quad (5.20)$$

Donde,  $x_i$  es la  $i$ -ésima observación,  $\bar{x}$  es el promedio de la variable de interés y  $w_{ij}$  es la ponderación espacial entre las observaciones  $i$  y  $j$ . El Índice Global de Moran consiste en la medición de la presencia o ausencia de autocorrelación espacial de una variable. La autocorrelación espacial está indicada por medio de valores que oscilan entre +1 y -1, en donde +1 indica autocorrelación positiva perfecta, -1 expresa autocorrelación negativa perfecta y un valor 0 muestra la presencia de patrones completamente aleatorios en su distribución espacial (Chasco, 2003).

El resultado de la fórmula 5.20 arroja una regla de decisión que valida la hipótesis de investigación de la siguiente forma (Moran, 1948):

Ho: Índice Moran = 0 → Ho: Índice Global de Moran es igual a 0, entonces no existe autocorrelación espacial, más bien se observa una distribución aleatoria de los valores de los ICV propuestos y por consiguiente se acepta la hipótesis nula.

H1: Índice Moran ≠ 0 → H1: Índice Global de Moran no es igual a 0, por tanto si existe autocorrelación espacial, entonces es posible afirmar que la calidad de vida en los municipios de Santander no se distribuye aleatoriamente, sino pueden encontrarse dispersos o agrupados en el espacio geográfico del departamento. Por tanto, mientras más pequeño sea el valor de  $p$ , mayor probabilidad hay que exista autocorrelación espacial.

### 5.5.3 Diagrama de dispersión de Moran

En el año 1993, Anselin presenta el Moran Scatterplot, una herramienta de análisis que permite observar en un gráfico de dispersión el comportamiento de cada unidad espacial. Esta novedad fue uno de los primeros pasos hacia el análisis local, es decir, la desagregación del valor global de AE. El scatterplot se puede dividir en cuatro cuadrantes (Figura 2) comenzando por el primero en la parte superior derecha y siguiendo en sentido de las agujas del reloj con los siguientes.

En el eje de las x aparecen los valores estandarizados de una variable para cada unidad espacial del área estudiada, y en el eje y los valores estandarizados del promedio de los valores en unidades vecinas de la misma variable (análisis univariado) u otra variable (análisis bivariado). En el cuadrante I podemos identificar las unidades espaciales con valores superiores a la media que, a su vez, también cuentan con vecinos con valores altos (situación alto-alto, también conocidos como hot spots en el scatterplot de Moran).

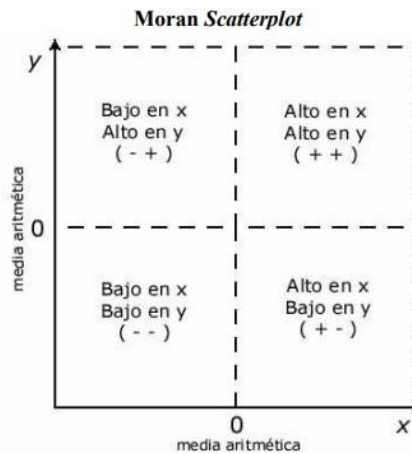
La situación inversa se registra en el cuadrante III (situación bajo-bajo, también denominados cold spots). Ambos cuadrantes permiten detectar los clusters o agrupamientos de unidades espaciales con valores similares a los de sus vecinos. En contraparte, los outliers espaciales responden a contextos mixtos, en otras palabras, unidades espaciales con valores bajos (inferiores a la media) con vecinos que registran valores altos (situación bajo-alto) en el cuadrante IV. El escenario opuesto (situación alto-bajo) se encuentra en el cuadrante II.

La importancia de los outliers radica en que no responden al principio de AE, ya que el valor de las unidades espaciales que poseen estas características se diferencia de los valores de sus

vecinos. Además, por su carácter atípico sirven, inicialmente, para detectar errores en el ingreso de la información, ya sea de forma accidental o deliberada (Celemín, 2009, pp. 5-6).

**Figura 2.**

*Diagrama de dispersión de Moran*



*Nota.* La figura 2 muestra las posibles ubicaciones de puntos en un diagrama de dispersión de Moran. Tomada de: *Autocorrelación espacial e indicadores locales de asociación espacial. Importancia, estructura y aplicación* (Figura 3: Moran Scatterplot), Celemín, 2009.

**5.5.4 Indicadores Locales de Autocorrelación Espacial**

El Índice Local de Asociación Espacial (LISA), permite la identificación de patrones locales de asociación espacial, descomponiendo el Índice Moran para evaluar la influencia de ubicaciones individuales en la estadística global que amplía las capacidades de visualización de los valores analizados a través del uso de Sistemas de Información Geográfica (SIG) (Anselin, 1995).

Este índice se encarga de representar aquellas localizaciones con valores significativos en indicadores estadísticos de asociación espacial local, alertando así de la presencia de puntos

calientes hot spots o atípicos espaciales, cuya intensidad depende de la significativa asociada de los datos estadísticos analizados. Este análisis se basa en la autocorrelación espacial local que representa las localizaciones con valores significativos en determinados indicadores estadísticos de asociación espacial local (Lizama, 2012).

Anselin (1995) define los indicadores locales de autocorrelación espacial (LISA) como estadísticos que satisfacen dos condiciones:

- a.** Para cada observación, el LISA da una medida del alcance o nivel de agrupación espacial significativa con relación a valores que se encuentran localizados a su alrededor.
- b.** La suma de las medidas LISA de todas las observaciones es proporcional a un indicador global de asociación espacial.

Como se ha indicado previamente, los índices de autocorrelación espacial global tienen restricciones que no posibilitan la identificación de las particularidades del fenómeno de estudio. Por ejemplo, no logran mostrar la existencia de agrupaciones regionales o locales dentro del contexto geográfico. Es precisamente en esta situación donde los indicadores locales adquieren relevancia para el análisis espacial. Dado que los fenómenos sociales no son espacialmente estacionarios, dado que sus características cambian en función a la ubicación geográfica, puesto que existen variaciones espaciales significativas en sus valores, Anselín (1995) propuso un indicador local de autocorrelación espacial (Índice local de Moran) que permite determinar la presencia de clústers o conglomerados espaciales en ubicaciones específicas dentro del área de estudio. El índice local de Moran está determinado por

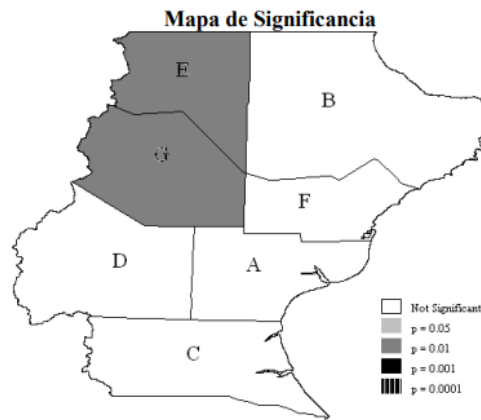
$$I_i = \frac{n(x_i - \bar{x}) \sum_{i=1}^n w_{ij}(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{para } i \neq j \quad (5.21)$$

Donde,  $x_i$  es la  $i$ -ésima observación,  $\bar{x}$  es el promedio de la variable de interés y  $w_{ij}$  es la ponderación espacial entre las observaciones  $i$  y  $j$ .

El resultado tangible de este proceso radica en la visualización del comportamiento mediante la generación de dos tipos de mapas. El primero se denomina mapa de significancia (Figura 3), parte de un procedimiento de aleatorización y se muestra para cada unidad espacial la probabilidad de que sus relaciones de contigüidad se produzcan de manera aleatoria. El segundo mapa se denomina de agrupamiento o clúster (Figura 4). En este se observa como cada unidad espacial se diferencia de sus unidades espaciales vecinas (Buzai y Baxendale, 2006).

**Figura 3.**

*Mapa de significancia LISA*

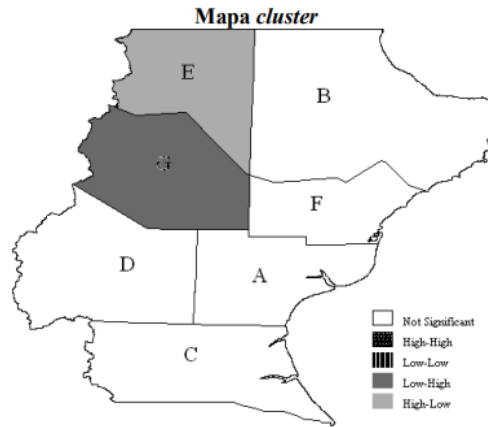


*Nota.* Tomado de: *Autocorrelación espacial e indicadores locales de asociación espacial.*

*Importancia, estructura y aplicación* (Figura 7: Mapa de significancia), Celemín, 2009.

**Figura 4.**

*Mapa clúster LISA*



*Nota.* Tomado de: *Autocorrelación espacial e indicadores locales de asociación espacial.*

*Importancia, estructura y aplicación* (Figura 8: Mapa Cluster), Celemín, 2009.

## 6. Metodología

### 6.1 Descripción de la muestra

Para esta investigación se tuvo en cuenta el departamento de Santander, ubicado en la zona norte central de Colombia y caracterizado por ser el cuarto departamento más influyente del país, a nivel económico y poblacional. Según cifras ofrecidas por el DANE en 2018, Santander cuenta con 2 millones 8 mil habitantes y se organiza, territorialmente, en 87 municipios agrupados en 8 provincias.

Para el análisis de la calidad de vida urbana en cada municipio, se construyó una base de datos de 116 variables, a partir de diferentes fuentes con información recolectada del año 2018. Primordialmente del DANE con el censo nacional de población y vivienda, complementado con el Observatorio de Medicina Legal y la Cámara de Comercio. En la Tabla 1 se presentan una selección de variables a disposición para construir el índice, agrupadas en nueve dimensiones, calidad de vivienda, servicios públicos, vulnerabilidad, socioeconomía, educación, salud, demografía, economía y seguridad y protección. En el Apéndice A se encuentra la descripción de cada una de las variables presentadas a continuación.

A excepción de las variables que representan tasas y números enteros (en el caso de las variables de Economía y Seguridad y Protección), las demás variables miden el porcentaje de viviendas o personas del municipio que poseen la característica explicada por la variable. En general, estas variables toman valores entre 0 y 1, donde 0 indica que ninguna vivienda o persona posee la característica y 1 indica que todas las viviendas o personas la poseen.

**Tabla 1.**

*Variables disponibles para construir los índices*

VARIABLES DISPONIBLES		
<b>Calidad de vivienda</b>	<b>Socioeconomía</b>	<b>Demografía</b>
Tipo de vivienda	Estrato 1 Estrato 2 Estrato 3 Estrato > 4	Índice de feminidad Índice de Masculinidad Relación de dependencia Jóvenes 15 a 29 años Relación personas +65 años/ 14 años
Principal componente de las paredes	Trabajo informal Desempleo de larga duración	<b>Economía</b> Peso relativo municipal en el PIB departamental % de inversión financiada con recursos propios Ingresos Tributarios Per - Cápita Municipal (promedio 4 años)
Principal componente del piso	<b>Educación</b> Personas que asisten a alguna institución educativa Alfabetismo 15 y Más Componente Inasistencia Analfabetismo Bajo logro educativo Barreras a servicios para cuidado de la primera infancia Inasistencia escolar Rezago escolar Sabe leer y escribir Puntaje saber 11 Matemáticas Puntaje saber 11 Lenguaje Básica Primaria Completa Básica secundaria completa Médica académica o técnica Técnica profesional o Tecnológica Universitario Especialización, Maestría, Doctorado	<b>Seguridad y Protección</b> Homicidios mujer Homicidios hombres Suicidios hombres Suicidios mujeres Muertes por accidente de tránsito Hombres Muertes por accidente de tránsito Mujeres Muertes accidentales Hombres Muertes accidentales Mujeres Violencia interpersonal Hombres Violencia Interpersonal Mujeres Presunto delito sexual en hombres Presunto delito sexual en mujeres Violencia contra niños, niñas y adolescentes por hombres Violencia contra niños, niñas y adolescentes por mujeres Violencia de Pareja Hombres Violencia de Pareja Mujer Violencia entre otros familiares Hombres Violencia entre otros familiares Mujeres Violencia adulto mayor Hombres Violencia adulto mayor Mujeres Lesiones por accidentes de transporte Hombres Lesiones por accidentes de transporte Mujeres Tasa de hurtos por 10.000 habitantes Tasa de homicidios por 10.000 habitantes Tasa de casos de violencia intrafamiliar por 10.000 habitantes
<b>Servicios públicos</b>		
Servicio de Acueducto		
Serv Alcantarillado		
Gas natural		
Recolección Basuras		
Internet		
Inodoro alcantarillado		
Inodoro pozo séptico		
Sin acceso a fuente de agua mejorada		
Penetración de la banda ancha		
<b>Vulnerabilidad</b>	<b>Salud</b>	
Número de cuartos por vivienda	Personas con dificultad de oír Personas con dificultad para hablar o conversar Discapacidad de origen visual Dificultad para: Ver de cerca, de lejos o alrededor Dificultad para: Mover el cuerpo, caminar o subir y bajar escaleras Dificultad para: Agarrar o mover objetos con las manos Dificultad para: Entender, aprender, recordar o tomar decisiones por sí mismo Dificultad para: Comer, vestirse o bañarse por sí mismo Dificultad para: Relacionarse o interactuar con las demás personas Personas Con Dificultades	
Número de hogares por vivienda		
Número de personas por vivienda		
Trabajo infantil		
Prop de Personas en NBI (%)	Muertes	
Hacinamiento crítico	Hijos Nacidos vivos	
Sin aseguramiento en salud	Promedio de hijos por mujer	
Prop de Personas en miseria	Razón de niños <5 años/ mujeres	
Componente vivienda	Adultos mayores con algún problema de salud en los últimos 30 días	
Componente Servicios	Cobertura en salud	
Componente Hacinamiento	Tasa de mortalidad infantil	
Barreras de acceso a servicios de salud	Cobertura en vacunación pentavalente	
Componente dependencia económica		

## 6.2 Métodos

El tipo de estudio que se realizó fue de carácter cuantitativo, enmarcado en un enfoque descriptivo a nivel multivariado. En particular para analizar la calidad de vida urbana de los

departamentos de Santander se hizo uso de dos técnicas estadísticas, con el fin de comparar y determinar la que mejor se adapte a los datos proporcionados. La primer técnica que se aplicó fue el análisis de componentes principales, método multivariante, cuyo objetivo principal es representar la información suministrada por las variables de estudio en un menor número de variables, llamadas componentes principales, resignando una baja cantidad de información, estas nuevas variables son combinaciones lineales de las originales y no están correlacionados entre sí, además se encuentran ordenados de manera que el primer componente principal explica el mayor porcentaje de variabilidad de los datos.

Por otra parte, la segunda técnica usada fue la distancia  $P_2$  elaborado por el profesor Pena Trapero, la cual proporciona un índice sintético para medir distancias y realizar comparaciones inter espaciales con el fin de medir el bienestar social de un conjunto de unidades territoriales en un tiempo dado. A diferencia del Análisis de Componentes Principales, la Distancia  $P_2$  no se ve afectada ante la adición de nuevos datos. A continuación se presentan las fases desarrolladas para dar cumplimiento al objetivo general y los objetivos específicos.

Fase 1. Construcción de la base de datos.

Fase 2. Edición de la base de datos.

Fase 3. Selección de las variables, por criterios específicos (presentados en la sección 7.1), para la construcción de los índices.

Fase 4. Programación de funciones en R-Studio para construir los índices por los métodos presentados anteriormente.

Fase 5. Programación del FlexDashboard para la visualización de los datos en forma de Geoportal.

Fase 6. Análisis de los resultados y escritura del informe final.

### 6.3 Software

Para el manejo de los datos recolectados, su exploración, procesamiento, mapeo y aplicación de las técnicas estadísticas mencionadas anteriormente, se manejó el software libre R Studio, un entorno de desarrollo integrado para el lenguaje de programación R. Para la construcción de los índices por ambos métodos se crearon dos funciones, una para cada índice, para realizar el geo portal que permite la visualización de los resultados a nivel espacial se construyó un dashboard interactivo con gráficos y mapas. A continuación se muestran las librerías utilizadas y una breve descripción de su propósito en tanto en el script como en el dashboard.

- iv. `sf`: Utilizada para manejar datos espaciales, en particular para leer y manipular archivos shapefile (`read_sf` y operaciones con geometrías).
- v. `readr`: Para leer datos (en este caso, shapefiles).
- vi. `ggplot2`: Utilizada para crear gráficos, en este caso, un mapa con la función `geom_sf` y la personalización de gráficos con escalas y flechas norte (`annotation_scale` y `annotation_north_arrow`).
- vii. `ggspatial`: Extiende `ggplot2` para facilitar la visualización de datos espaciales y la inclusión de elementos de mapas, como escalas y leyendas de coordenadas.
- viii. `dplyr`: Utilizada para manipulación de datos, en particular para filtrar y unir data frames (`filter`, `merge`, `subset`).
- ix. `factoextra`: Utilizada para análisis de componentes principales, en particular para obtener los valores propios (`get_eigenvalue`).

- x. `kableExtra`: Utilizada para presentar tablas de manera más atractiva (`kable`, `kable_styling`, `add_header_above`).
- xi. `knitr`: Utilizada para crear tablas (por la función `kable`).
- xii. `magrittr`: Utilizada para realizar múltiples operaciones secuenciales sobre los datos. (`%>%`).
- xiii. `spdep`: Proporciona herramientas para el análisis de dependencia espacial en datos. Incluye funciones para calcular y analizar matrices de pesos espaciales y realizar análisis espaciales.
- xiv. `stats`: Utilizada para funciones estadísticas y la regresión en el índice por  $DP_2$  (`lm`, `summary`)
- xv. `flexdashboard`: Utilizada para crear dashboards interactivos, organizando los gráficos y visualizaciones en un formato de panel.
- xvi. `plotly`: Utilizada para crear gráficos interactivos que se pueden explorar en el dashboard.
- xvii. `maps`: Proporciona datos geográficos y funciones para trabajar con mapas.
- xviii. `shiny`: Añade interactividad al dashboard, permitiendo a los usuarios interactuar con los gráficos y mapas.
- xix. `viridis`: Proporciona paletas de colores perceptualmente uniformes para los gráficos.
- xx. `tidyr`: Facilita la manipulación y transformación de datos, asegurando que los datos estén en el formato adecuado para su visualización.
- xxi. `tidyverse`: Un conjunto de paquetes que incluyen herramientas como `dplyr`, `ggplot2`, y `tidyr` para la manipulación y visualización de datos.

xxii. `ggiraph`: Añade interactividad a los gráficos creados con `ggplot2`, permitiendo a los usuarios interactuar con los gráficos en el dashboard.

xxiii. `ggrepel`: Extiende `ggplot2` para mejorar la visualización de etiquetas en gráficos. Evita la superposición de etiquetas y facilita la creación de gráficos más legibles.

xxiv. `cowplot`: Proporciona herramientas para combinar y organizar gráficos creados con `ggplot2`.

xxv. `svDialogs`: Permite la creación de diálogos gráficos interactivos en aplicaciones R, lo que es útil para obtener entradas del usuario de manera más visual.



## 7. Resultados

### 7.1 Análisis descriptivo

Se consolidó una base de datos compuesta por 117 observaciones (ver Tabla 1) y 87 unidades muestrales (municipios de Santander). Aproximadamente el 2,6% de las variables fueron recopilados del Observatorio de Competitividad de la Cámara de Comercio de Bucaramanga, 21,4% del Observatorio de Violencia de Medicina Legal y el 76% del DANE. Posteriormente, se removieron 85 variables por los siguientes criterios de selección.

Primero, más del 60% de los municipios presentaban valores nulos en la variable (vivienda en área protegida, lesiones por accidentes de transporte hombres y lesiones por accidentes de transporte mujeres). Segundo, todos los municipios presentaban valores muy cercanos a 0 o 1 (vaciado concreto, servicio de energía, apartamentos, casas, material de piso mármol).

Tercero, tenían correlaciones cercanas a -1 con otras variables que tenían información similar (paredes de tapia pisada, adobe, bahareque con correlación de -0.83 con paredes de ladrillo; pisos de cemento con correlación de -0.77 con pisos de baldosa; inodoro pozo séptico correlación con inodoro alcantarillado de -0.87; número de cuartos, número de personas y número de hogares en la vivienda correlacionadas con hacinamiento crítico; servicio de acueducto correlación de -0.92 con sin acceso a fuente de agua mejorada; sabe leer y escribir correlación de -0.85 con alfabetismo).

Cuarto, tenían correlaciones cercanas a 1 con otras variables que tenían información similar (personas con dificultad para ver, oír, tomar objetos, razonar, hablar, ver alrededor, mover el cuerpo, vestirse o comer por sí mismo e interactuar con los demás son resumidas por la variable

personas con dificultades; inodoro alcantarillado correlación con servicio de alcantarillado de 0.99; penetración banda ancha correlación con internet de 0.91).

Quinto, tenían correlaciones débiles, no mayores a 0.3, con las demás variables (pisos en tierra o arena, cobertura en vacunación pentavalente, tasa de mortalidad infantil, cobertura en salud, barreras de acceso a servicio de salud, muertes, hijos nacidos vivos, promedio hijos por mujer, razón de niños menores de 5 años y mujeres, adultos mayores con dificultades de salud en el último mes, personas que asisten a alguna institución educativa, barreras a servicios para cuidado de la primera infancia, inasistencia escolar, rezago escolar, índice de masculinidad, relación de dependencia).

Por último, estas variables se agrupan en otra variable que muestra una mejor correlación con las demás. Por ejemplo, los niveles educativos como básica primaria completa, básica secundaria completa, media académica o técnica, técnica profesional o tecnológica, y especialización o maestría se explican mediante la variable universitario. Del mismo modo, se incluyen variables de seguridad como homicidios de mujeres y hombres, suicidios de hombres y mujeres, muertes de hombres y mujeres en accidentes de tránsito, muertes accidentales de hombres y mujeres, violencia intrafamiliar de hombres y mujeres, presunto delito sexual de hombres y mujeres, violencia contra niños o adolescentes por hombres y mujeres, violencia de pareja de hombres y mujeres, violencia entre otros familiares con hombres y mujeres, y violencia contra adultos mayores por hombres y mujeres.

Estas variables se explican a través de tasas como la tasa de hurtos por 10,000 habitantes, la tasa de homicidios por 10,000 habitantes, y la tasa de casos de violencia intrafamiliar por 10,000

habitantes. A continuación, se presenta una descripción breve de las variables consideradas en el estudio.

**Tabla 2.**

*Descripción de las variables seleccionadas para construir los índices*

VARIABLES SELECCIONADAS			
Variable	Descripción	Variable	Descripción
pladrillo	Proporción de viviendas cuyo principal componente de sus paredes sea bloque, ladrillo, piedra o madera pulida	hacinamiento	Porcentaje de viviendas con más de tres personas por cuarto
pmadera	Proporción de viviendas cuyo principal componente de sus paredes sea madera burda, tabla o tablón	dependencia económica	Porcentaje de personas naturales que dependen de otra persona para suplir sus necesidades económicas
baldosa	Proporción de viviendas cuyo principal componente de sus pisos sea baldosa, vinilo, tableta, ladrillo, laminado	salud	Porcentaje de personas sin aseguramiento a un régimen de salud
madera	Proporción de viviendas cuyo principal componente de sus pisos sea madera burda, tabla o tablón	difsalud	Personas con dificultades de salud
alcantarillado	Proporción de viviendas con servicio de alcantarillado	analfabetismo	Porcentaje de personas de 15 años o más que no saben leer ni escribir
gas	Proporción de viviendas con servicio de gas	bajaedu	Porcentaje de viviendas con promedio menor a 9 de los años de escolaridad de las personas de 15 años y más en el hogar
basuras	Proporción de viviendas con servicio de recolección de basuras	icfesm	Promedio del puntaje de la sección de matemáticas en el saber 11 del 2017
internet	Proporción de la población con servicio de internet móvil o fijo	icfesl	Promedio del puntaje de la sección de lenguaje en el saber 11 del 2017
sfam	Proporción de la población con acceso a fuentes mejoradas de abastecimiento de agua potable	universitario	Proporción de personas con título universitario para el 2017
estrato1	Proporción de viviendas en estrato 1	indicefem	Número de mujeres por cada 100 hombres en la población
estrato3	Proporción de viviendas en estrato 3	jovenes	Proporción de personas con años entre 15 a 29
desempleo	Proporción de la población desempleada	viv/per	Proporción de viviendas del municipio con respecto a la densidad poblacional
tinformal	Proporción de la población con trabajo informal	proporción	Relación de personas mayores a 65 años con respecto a menores de 14 años
pib	Peso relativo municipal en el PIB departamental	hurtos	Tasa de hurtos por 10.000 habitantes
tinfantil	Proporción de la población menor a 18 años que realice alguna actividad económica o trabajo	homicidios	Tasa de homicidios por 10.000 habitantes
miseria	Proporción de la población que carecen de dos o más necesidades básicas	vif	Tasa de casos de violencia intrafamiliar por 10.000 habitantes

## 7.2 Análisis de Componentes Principales

### 7.2.1 Estandarización

Con el fin de satisfacer la condición de *invarianza* en el índice sintético que mide la calidad de vida por PCA se procede primero a resolver el problema de escala entre las variables originales. Jackson (1991) señaló que la normalización en el Análisis de Componentes Principales (PCA) es fundamental para garantizar que todas las variables tengan la misma escala, con el fin de facilitar la comparación entre ellas y evitar que las variables con escalas diferentes dominen el análisis. De

igual manera, Hastie, Tibshirani, & Friedman, 2009, mencionaron cómo la normalización facilita la interpretación de los resultados en el PCA, al permitir que los pesos de las componentes principales se basen en la estructura de covarianza o correlación entre las variables originales, en lugar de sus escalas (p. 547).

Por otra parte, la importancia de normalizar los datos en valores entre 0 y 1 para el PCA, con el fin de asegurar que todas las variables tengan la misma escala relativa y facilitar la comparación entre ellas. Para este estudio, se usó la siguiente ecuación para normalizar la variable  $i$  para el  $j$ -ésimo municipio (Hair et al. , 2019).

$$x'_{ji} = \frac{|\mathbf{Max}(x_i) \text{ o } \mathbf{Min}(x_i) - x_{ij}|}{\mathbf{Max}(x_i) - \mathbf{Min}(x_i)} \quad (7.1)$$

Dónde  $\mathbf{Max}(x_i)$  y  $\mathbf{Min}(x_i)$  son el máximo y mínimo valor de la variable  $i$ , respectivamente,  $x_{ij}$  es el valor de la observación de la variable  $i$  para el municipio  $j$  y los valores máximo y mínimo en el numerador dependerán del efecto que tenga la variable en el Índice de Calidad de Vida (ICV). Es decir, si se considera que un valor alto en la variable tiene un impacto positivo en el ICV, entonces se usa el mínimo en el numerador, en cambio, si se considera que un valor alto en la variable tiene un impacto negativo en el ICV, se usa el máximo en el numerador.

Por ejemplo, para la variable Sin acceso a fuente de agua mejorada se usa el máximo, así el indicador es 1 para el municipio que tiene menos viviendas sin acceso a fuente de agua potable y 0 para el municipio que más viviendas tiene sin acceso a fuente de agua potable. Para la variable Internet se usa el mínimo y de esta forma el municipio para el cual el indicador es 1 es aquel que tiene más viviendas con servicio de internet y 0 para el que tiene poca o nula conexión a internet.

### 7.2.2 Correlaciones

La correlación es un concepto importante en el PCA por su capacidad de identificar relaciones lineales entre variables de un conjunto de datos, desarrollar una interpretación más precisa de los componentes principales y seleccionar las variables relevantes para reducir la dimensionalidad de los datos. Para este proceso se usó la función *cor* que permite calcular el coeficiente de Pearson para variables cuantitativas.

En el Apéndice B se observa que la variable Pladrillo, que representa el porcentaje de viviendas del municipio cuyas paredes están construidas en materiales como bloque, ladrillo, piedra o madera pulida, presenta correlaciones moderadamente negativas (de -0,6 a -0,4) con Hurtos y moderadamente positivas (de 0,4 a 0,6) con 11 de las 32 variables analizadas. Entre estas se destacan Baldosa, que contiene las viviendas con pisos en baldosa, vinilo o tableta, lo que hace suponer una relación positiva en la calidad de los materiales de construcción de las paredes y los pisos en las viviendas de los municipios. De manera similar, la variable Pmadera, que indica el porcentaje de viviendas con paredes de madera burda, tabla o tablón, tiene una correlación positiva alta (de 0,6 a 0,8) con Madera, asociada a los pisos en las viviendas construidos con dicho material.

Por otra parte, la variable Alcantarillado, que refleja el acceso a servicio de alcantarillado en las viviendas, muestra una correlación alta positiva con 9 variables, incluyendo Gas, Internet y Estrato 3, y correlaciones positivas fuertes (mayores a 0,8) con Basuras, lo que sugiere que las viviendas con servicio de alcantarillado cuentan con servicios de gas, internet y basura. De manera similar, la variable Internet presenta correlaciones fuertes con Baldosa, Basura y Universitario, y correlaciones altas con 8 variables, tales como Estrato 3 y Trabajo Informal. Esto sugiere que las viviendas con acceso a internet suelen ser de estrato 3, tener pisos en baldosa, contar con servicio

de recolección de basuras y sus habitantes tienen niveles de educación universitaria o se desempeñan en trabajos informales.

Además, la variable Hurto, que mide la tasa de hurtos por cada 10.000 habitantes, muestra correlaciones altas negativas con Alcantarillado, Gas, Basura, Internet, Universitario, Tinformal y PIB; y correlación alta positiva con Salud. Esto indica que, a medida que la tasa de hurtos por cada 10.000 habitantes aumenta, la presencia o acceso a servicios como alcantarillado, gas, recolección de basura, acceso a internet, niveles de educación universitaria, empleo informal y el peso del producto interno bruto (PIB) del departamento tienden a disminuir y aumentan las personas sin aseguramiento de salud.

Sin embargo, se observaron variables que presentaron correlaciones débiles (de 0.2 a 0.4) o muy bajas ( $> 0.0$  a 0.2) con las demás variables, pero se tienen en cuenta debido a su importancia en la medición de Calidad de Vida según la literatura. Por ejemplo, Homicidios, que mide la cantidad de homicidios por cada 10.000 habitantes, y Vif, que mide la tasa de casos de violencia intrafamiliar (IF) por cada 10.000 habitantes, muestran este tipo de correlaciones, pero se consideran en el estudio dado que la incidencia de victimización y las percepciones de inseguridad afectan el bienestar y la salud en América Latina (Graham y Chaparro, 2021).

Asimismo, la variable Índicefem muestra algunas correlaciones moderadas, aunque en su mayoría son bajas. No obstante, se conserva con el propósito de proporcionar una perspectiva de equidad de género en el área de estudio y evaluar las condiciones y oportunidades que tienen las mujeres para desarrollarse en los ámbitos familiar, social, laboral y político en el municipio donde residen.

**7.2.3 Resultados del Análisis de Componentes Principales (PCA)**

Para el cálculo de los Componentes Principales (PC), se hizo uso de la función *prcomp* del software *R-studio*, que toma la siguiente forma, *prcomp (objeto, centro, escala)*, donde *objeto* es la matriz de datos y *centro* y *escala* son valores lógicos que indican si se quiere centrar y escalar las variables antes de realizar el análisis. La función devuelve variedad de información útil para examinar los resultados, incluye la desviación estándar de cada componente principal, *sdev*, la matriz de pesos, *rotation*, la matriz, *x*, con las puntuaciones del componente r-ésimo en la región j-ésima y los valores usados para centrar y escalar.

El comando *summary (prcomp ())* devuelve un resumen del PCA, que incluye la desviación estándar de cada componente, la proporción de la varianza que explica cada uno y la proporción de varianza acumulada hasta dicho componente. En la tabla 3, se presentan los resultados de la función *summary* para los primeros 8 componentes principales y adicionalmente una fila con sus respectivos valores propios.

**Tabla 3.**

*Resumen de estadísticos del PCA*

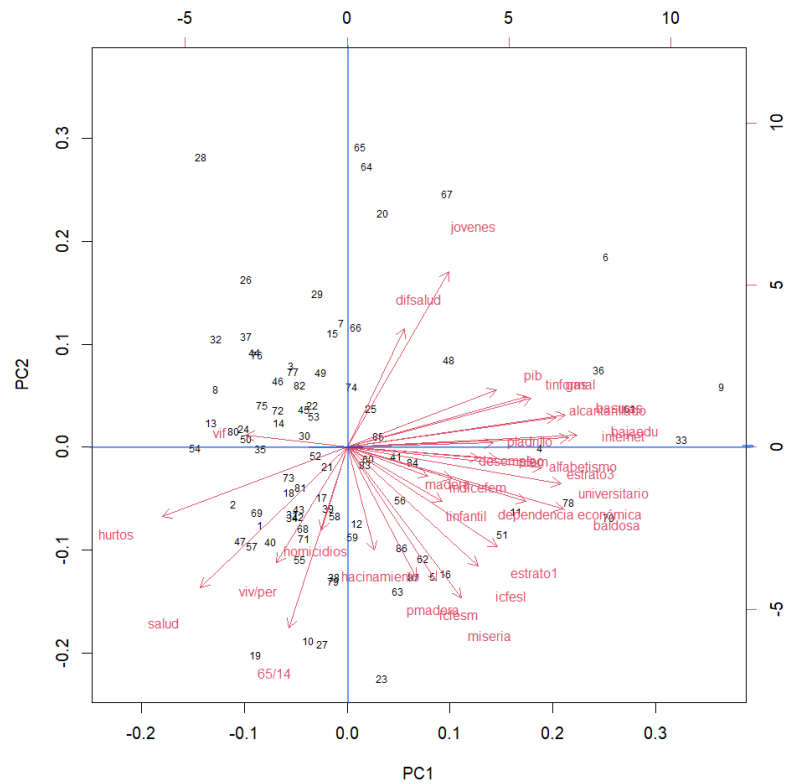
	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Desviación estándar	3.514	2.019	1.547	1.398	1.183	1.046	1.000	0.971
Proporción varianza	0.386	0.127	0.075	0.061	0.044	0.034	0.031	0.029
Proporción acumulada	0.386	0.513	0.588	0.649	0.693	0.727	0.758	0.788
Valores propios	12.34	4.077	2.394	1.953	1.400	1.093	1.000	0.942

Se aplicó el criterio de Kaiser, que sugiere retener únicamente los componentes principales cuyos valores propios sean mayores que 1 (Kaiser, 1960). De esta forma se seleccionan los 7 primeros componentes principales, los cuales explican un 75.8% de la variabilidad de los datos.

Para analizar el peso que tiene cada variable en las componentes se usó la matriz *rotation* que proporciona la función *prcomp*, una vez realizado el PCA, los puntajes de cada variable para las primeras dos componentes se encuentran en el Apéndice C. En la Figura 3 se visualiza el biplot de las dos primeras componentes, donde los vectores representan las variables y los números los municipios de Santander ordenados alfabéticamente. Las variables que se asocian al primer componente principal son aquellas que están más cercanas al eje horizontal o que el ángulo de su respectivo vector sea cercano a  $0^\circ$ , mientras que las que se asocian al segundo componente principal son las más cercanas al eje vertical o que el ángulo de su respectivo vector sea cercano a  $90^\circ$ .

**Figura 5.**

*Biplot PC1 vs PC2*



Las variables asociadas a las componentes principales cuya longitud del vector es mayor son más importantes en la explicación de la variabilidad en los datos y los vectores con la misma dirección indican que sus respectivas variables tienen correlación positiva y viceversa, los vectores con direcciones opuestas indican correlación negativa en sus respectivas variables. Además, la proximidad la posición de las observaciones a los vectores indican que tan alto es el valor de la variable para este municipio.

De este modo, la variable con mayor peso en la primera componente principal es Bajaedu, que corresponde a la proporción de población con un bajo logro educativo. Esta variable se mide

a través de la escolaridad promedio de las personas de 15 años y más en el hogar. Se considera que un hogar está privado cuando el promedio del logro educativo de las personas de 15 años y más que lo componen es menor a 9 años escolares (Departamento Nacional de Planeación, 2011, p. 15).

La segunda variable que tiene mayor peso para la primera componente principal es Internet, que corresponde a la proporción de hogares que cuentan con servicio de internet móvil o fijo. La tercera variable con mayor peso es Basuras, que indica la proporción de hogares que cuentan con servicio de recolección de basuras. En Colombia, se considera que una persona cuenta con servicio de recolección de basuras cuando a su vivienda una entidad pública, privada o comunal legalmente establecida le presta este servicio en forma regular o esporádica, en carretas, zorras, etc., destinados para este fin (DANE, 2017, p. 1).

Por otra parte, para la segunda componente principal las variables con mayor peso son Proporción y Jóvenes. Donde la primera se refiere a la proporción de personas mayores a 65 años con respecto a los menores de 14 años en el municipio y la segunda indica la proporción de habitantes que tienen una edad comprendida entre los 15 y 29 años.

Se observa que las variables de la primera componente principal tienen correlación positiva entre ellas y por el contrario las variables de la segunda componente principal están correlacionadas negativamente. Además, Bucaramanga (9), Floridablanca (33), Piedecuesta (61) y Girón (36) tiene valores altos en las variables con más peso, que se asocian a la primera componente principal y Sabana de Torres (67), Cimitarra (20), Cabrera (10) y Chipatá (19) reflejan valores altos en las variables con más peso, asociadas a la segunda componente principal.

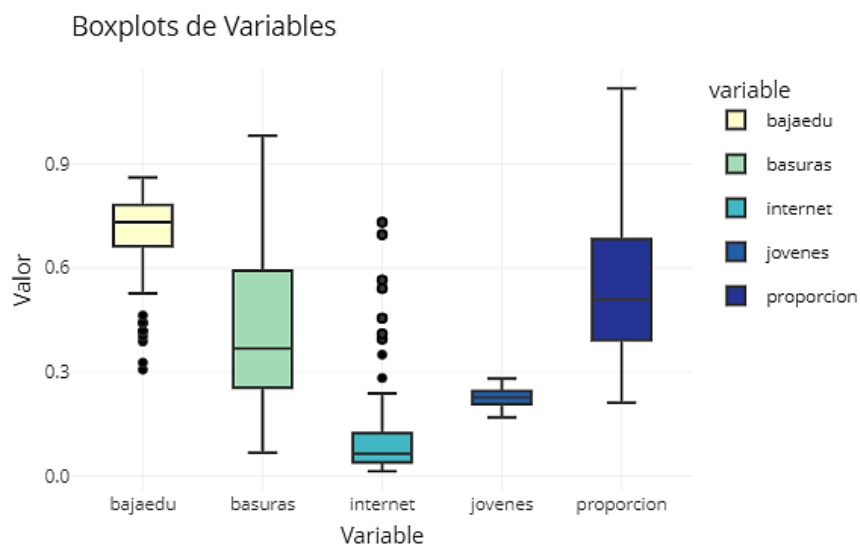
Asimismo, se observa una agrupación de observaciones que están cercanas de una o dos variables, pero lejanas de las variables con más peso para las dos primeras componentes principales. Estas observaciones corresponden a municipios como Carcasí (13), Molagavita (54), Macaravita (50) y Coromoro (24) ubicados en dirección inversa a las variables asociadas al primer componente.

**7.2.3.1 Distribución espacial de las variables relevantes para el PCA**

Como se mencionó, las variables con mayor aporte en el primer componente son Bajaedu, Internet y Basuras, y en el segundo componente, Proporción y Jóvenes. Para saber cómo es el comportamiento de las variables, se mostrarán estadísticos descriptivos y la distribución espacial de cada una de ellas dentro de los municipios de Santander

**Figura 6.**

*Box-plots para las variables relevantes en el análisis por componentes principales*



*Nota.* Box-plots para las variables con mayor peso en la primera y segunda componente principal.

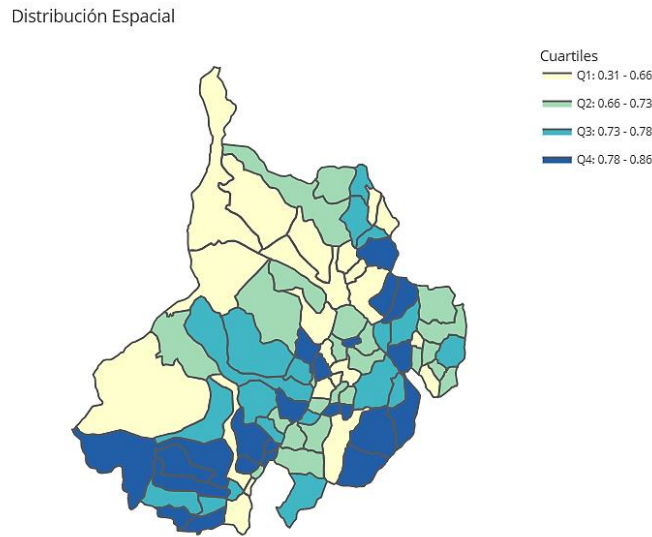
En la Figura 6 se observa que los indicadores Bajaedu e Internet presentan municipios con valores atípicos. Para Bajaedu, su distribución es sesgada hacia la izquierda, tiene variabilidad moderada y sus valores atípicos están entre 0.3 y 0.5. Además, el 75% de los municipios presentan puntajes mayores a 0.66, lo que indica que la mayoría de los municipios tienen un bajo logro educativo en los hogares.

Por otra parte, para Internet su distribución es sesgada hacia la derecha, tiene variabilidad moderada y sus valores atípicos son mayores a 0.25. Esto sugiere que el acceso a internet es limitado en la mayoría de los municipios, siendo excepcionales aquellos con una cobertura de internet superior al 25% de su población.

Las variables Basuras y Proporción también presentan una distribución sesgada hacia la derecha, con alta variabilidad pero sin valores atípicos. La alta variabilidad en la cobertura de recolección de basuras indica disparidades significativas en el acceso a este servicio esencial. Por otro lado, la variable Jóvenes presenta la distribución más simétrica entre todas, lo que indica una mayor uniformidad en la proporción de jóvenes a lo largo de los municipios.

**Figura 7.**

*Distribución espacial Bajaedu*



*Nota.* Mapa de los municipios de Santander con la distribución espacial por cuartiles de la variable Bajaedu.

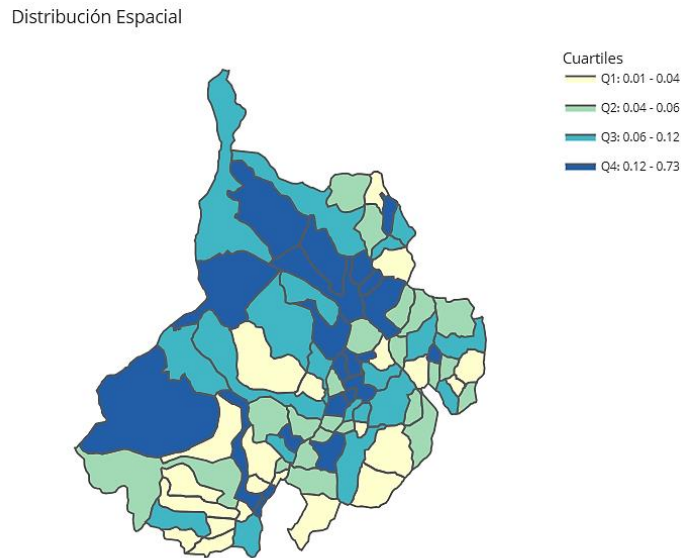
Debido a la forma en que se construye la variable Bajaedu, un puntaje bajo indica que en la mayoría de los hogares, uno o más miembros han alcanzado el título de Bachiller. Se observa, en la Figura 5, que los municipios cercanos a Bucaramanga y ubicados en las zonas suroeste obtuvieron puntajes dentro del Cuartil 1 (Q1: 0.31 – 0.66), que corresponden a valores entre 0.31 y 0.66. Esto sugiere que en estos municipios hay una probabilidad entre 0.34 y 0.69 de encontrar hogares con al menos un Bachiller y, por tanto, presencia de instituciones educativas de mayor calidad.

Adicionalmente, se observa que la mitad de los municipios del departamento tienen un puntaje mayor a 0.73 y se encuentran ubicados, en su mayoría, en el suroeste y sureste de Santander. Esta distribución muestra que en estas zonas el acceso a la educación básica es limitado

y posiblemente nulo en educación superior, puesto que la probabilidad de encontrar hogares con mínimo un bachiller es menor al 0.27.

**Figura 8.**

*Distribución espacial Internet*



*Nota.* Mapa de los municipios de Santander con la distribución espacial por cuartiles de la variable Internet.

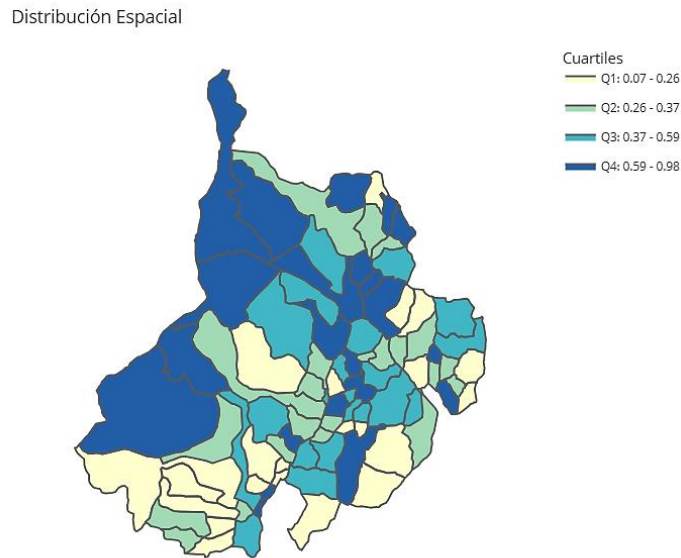
Se observa que la disponibilidad de acceso a internet en los hogares de Santander presenta una notable disparidad. En un amplio porcentaje, específicamente en más del 88% de los hogares del 75% de los municipios, no se cuenta con acceso a internet móvil ni fijo, como lo demuestra el puntaje inferior a 0.12 registrado en esta variable.

Además, el máximo valor registrado para la variable de Internet fue 0.73, esto indica que en al menos un municipio, el 73% de los hogares disfrutaban de acceso a internet. Sin embargo, este dato también sugiere que más de una cuarta parte de los hogares en este municipio específico aún

carecen de este servicio, lo que subraya la necesidad de ampliar la infraestructura de conectividad para garantizar la equidad en el acceso a la información y los recursos en línea.

**Figura 9.**

*Distribución espacial Basuras*



*Nota.* Mapa de los municipios de Santander con la distribución espacial por cuartiles de la variable Basuras.

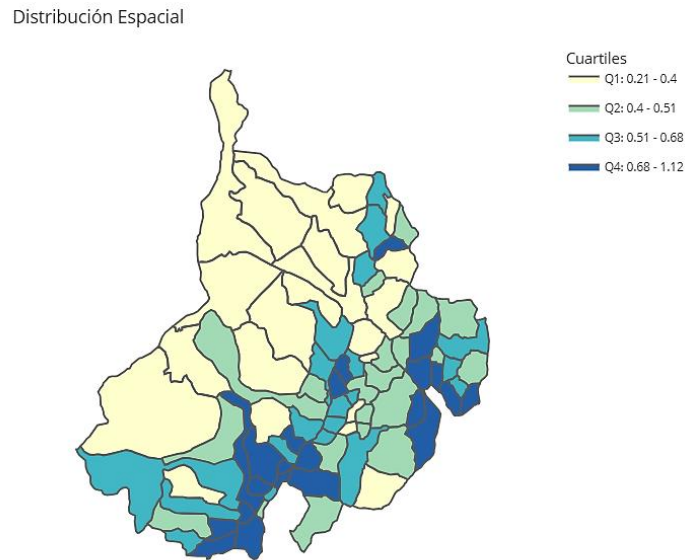
En la mitad de los municipios de Santander, el servicio de recolección de basura solo llega a un máximo del 37% de la población residente. Esta cifra pone de manifiesto una deficiencia significativa en la gestión de residuos en estas áreas, lo que puede conllevar problemas de salud pública, contaminación ambiental y deterioro del entorno.

La falta de acceso adecuado a la recolección de basura puede resultar en la acumulación de desechos sólidos en espacios públicos, aumentando el riesgo de enfermedades y contribuyendo a la degradación del paisaje urbano. Además, es preocupante que solo el 25% de los municipios en

el departamento cuenten con un servicio de recolección que llegue al menos a la mitad de sus habitantes.

**Figura 10.**

*Distribución espacial Proporción*



*Nota.* Mapa de los municipios de Santander con la distribución espacial por cuartiles de la variable Proporción.

Según proyecciones para Colombia, se estima que para el año 2050, el 18.7% de la población será mayor de 60 años, convirtiéndose esto en un motivo de preocupación acuciante porque este envejecimiento de la población conducirá a reducir el crecimiento económico del país, debido a que los costos derivados de la satisfacción de las necesidades de este grupo poblacional crecerán mucho más rápido que el recaudo de recursos para asegurar su sostenibilidad (Fajardo, Córdoba, & Enciso, 2016, p. 2).

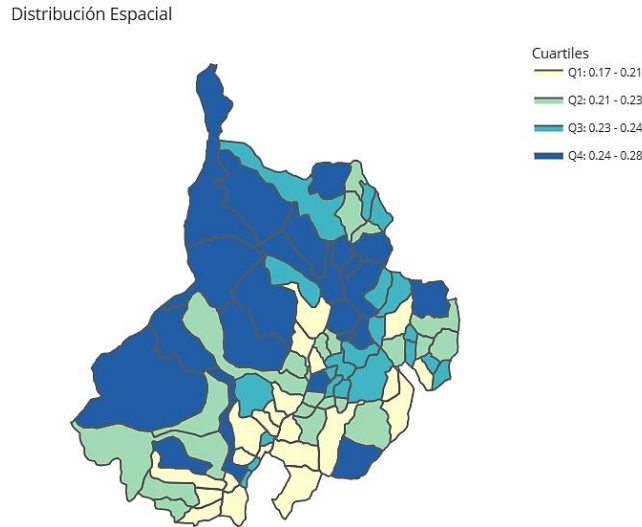
Sin embargo, este problema latente podría ser una realidad en estos momentos en algunos municipios de Santander, donde el valor del indicador Proporción toma valores mayores a 0.68, pertenecientes al cuarto cuartil (Q4: 0.68 - 1.12) y coloreadas en color azul oscuro. Estas áreas enfrentan ya los desafíos significativos del envejecimiento poblacional, con implicaciones directas en la demanda de servicios de salud, pensiones y otros recursos sociales.

En cambio, para los municipios del primer cuartil (Q1: 0.21 – 0.4), la proporción de personas mayores de 65 años es relativamente baja en comparación con la de menores de 14 años. Estas áreas, indican una población más joven y, por lo tanto, podrían enfrentarse menos a los desafíos económicos y sociales asociados con un alto porcentaje de población envejecida en el corto plazo. Sin embargo, es crucial que estos municipios implementen políticas preventivas para manejar el envejecimiento futuro.

Por último, los municipios en el segundo y tercer cuartil (Q2: 0.4 - 0.51 y Q3: 0.51 - 0.68), coloreados en tonos de verde, presentan una proporción moderada de personas mayores en relación con los menores. Estos municipios están en una etapa de transición demográfica y necesitan prepararse para el incremento esperado en la población mayor en las próximas décadas. En resumen, aunque el envejecimiento de la población es una tendencia futura proyectada para Colombia en general, ya es una realidad presente en varios municipios de Santander.

**Figura 11.**

*Distribución espacial Jóvenes*



*Nota.* Mapa de los municipios de Santander con la distribución espacial por cuartiles de la variable Jóvenes.

La distribución del indicador Jóvenes presenta una baja variabilidad en los municipios de Santander. Como se muestra en la Figura 11, el valor mínimo de esta variable es 0.17, mientras que el valor máximo es 0.28. Esta estrecha diferencia sugiere que, en general, los municipios de Santander tienen una proporción de población joven relativamente homogénea.

Los municipios con valores más bajos en esta variable podrían estar comenzando a enfrentar los desafíos asociados con una población que envejece. Por otro lado, aquellos con valores más altos tienen una ventaja demográfica, ya que una mayor proporción de jóvenes puede actuar como un motor para el crecimiento económico y la innovación.

### 7.2.3.2 Construcción del Índice de Calidad de Vida por PCA ( $ICV_{PCA}$ )

Como se mencionó en la sección 5.2.3, se adaptó una fórmula matemática usando los indicadores compuestos obtenidos por medio del análisis de componentes principales, con el fin de medir las condiciones de vida puesto que la calidad de vida es un fenómeno multidimensional.

$$ICV_{PCA} = \frac{\sum_{i=1}^r Y_{rj} \cdot \sqrt{\lambda_r}}{\sum_{i=1}^r \sqrt{\lambda_r}} \quad (7.2)$$

Donde,  $Y_{rj}$  es la puntuación del componente  $r$ -ésimo para la observación  $j$ -ésima,  $\lambda_r$  es el valor propio para dicho componente y  $r$  es la cantidad de componentes seleccionados, en este caso, siete. Luego se estandarizó por medio de la formula

$$\frac{ICV_{PCA} - \min(ICV_{PCA})}{\max(ICV_{PCA}) - \min(ICV_{PCA})} \quad (7.3)$$

Obteniendo el índice entre 0 y 1, que refleja la calidad de vida en un municipio con un nivel desfavorable asumirá valores cercanos a 0, mientras que un valor cercano a 1 indica una situación favorable.

En la Tabla 4 se muestra un ranking de los 7 mejores y 7 peores puntajes en el  $ICV_{PCA}$ . Se observa que los municipios con puntajes altos en las variables asociadas al primer componente lideran el ranking, mientras que municipios con valores bajos en estas se ubican en las últimas posiciones, como es el caso de Carcasí y Molagavita.

Además, se evidencia un puntaje alto para Barrancabermeja, lo que supone que el papel de capital petrolera del país, el puerto fluvial y el patrimonio histórico que tiene esta impacta positivamente en el  $ICV_{PCA}$ , permitiendo que Barrancabermeja supere en calificación a municipios del Área Metropolitana de Bucaramanga, como Girón y Piedecuesta.

Por otro lado, Carcasí, ubicado en la región Comunera, es el municipio con puntaje en el  $ICV_{PCA}$  más bajo, debido al ser un municipio rural y carecer de infraestructuras básicas como viviendas adecuadas, acceso confiable a servicios de salud y educación, así como sistemas de suministro de agua potable y saneamiento eficientes y oportunidades de empleo, lo que puede resultar en altos niveles de desempleo o empleo informal.

**Tabla 4.**

*Ranking  $ICV_{PCA}$*

Top 5 Valores Más Altos			Top 5 Valores Más Bajos		
	NOMBRE_MPI	indacp		NOMBRE_MPI	indacp
9	BUCARAMANGA	1.0000000	13	CARCASI	0.0000000
33	FLORIDABLANCA	0.8588792	54	MOLAGAVITA	0.0229606
6	BARRANCABERMEJA	0.8418819	19	CHIPATA	0.0438145
61	PIEDECUESTA	0.7678408	37	GUACA	0.0634904
36	GIRON	0.7217098	57	ONZAGA	0.0768886
70	SAN GIL	0.6544909	50	MACARAVITA	0.1004102
78	SOCORRO	0.6113128	31	ENCISO	0.1025736

En la Figura 12 se presenta el mapa de Santander con sus respectivos municipios y sus resultados obtenidos del  $ICV_{PCA}$ . Se evidencia que los municipios del Área Metropolitana de Bucaramanga presentan mejores condiciones en la calidad de vida, según el  $ICV_{PCA}$ , seguido de San Gil y Socorro, ubicados en la ruta principal que comunica a Bucaramanga con Bogotá, DC.

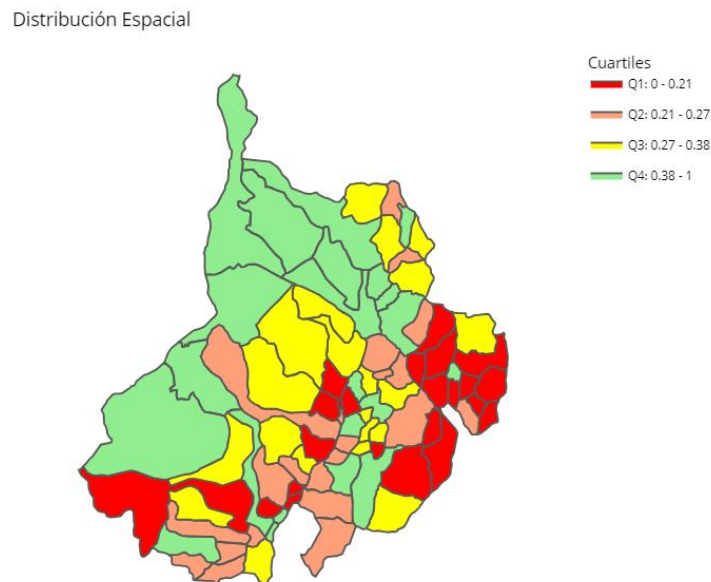
Los municipios dentro del Área Metropolitana de Bucaramanga, al encontrarse en una región urbanizada y desarrollada, suelen contar con una infraestructura y servicios públicos

superiores en comparación con otros municipios. Esto abarca desde servicios de salud y educación hasta transporte, actividades recreativas y culturales, vitales para el bienestar de sus residentes.

Por otra parte, San Gil y Socorro son destinos turísticos populares debido a su patrimonio cultural, ofertas de educación y actividades recreativas y de turismo, lo que contribuye positivamente a la calidad de vida de sus habitantes al generar empleo, impulsar la economía local y proporcionar oportunidades de recreación y esparcimiento.

**Figura 12.**

*Distribución espacial por cuartiles del  $ICV_{PCA}$*



*Nota.* Mapa de los municipios de Santander con la distribución espacial por cuartiles del  $ICV_{PCA}$ .

Los municipios situados en la zona sureste y suroeste del departamento presentan los puntajes más bajos, correspondientes al Cuartil 1 (Q1: 0 – 0.21) del  $ICV_{PCA}$ , representados en color

rojo y donde el máximo valor del índice es de 0.21. En estas mismas zonas se presentaban los puntajes más bajos para las variables con mayor peso en el primer componente principal.

Estos resultados revelan que estas regiones podrían estar enfrentando desafíos significativos en términos de acceso a servicios básicos, infraestructura, y oportunidades educativas y económicas lo que se traduce en una menor calidad de vida para sus habitantes, como se había mencionado en la sección 7.3.2.1.

Además, los municipios representados en color naranja, correspondientes al Cuartil 2 (Q2: 0.21 – 0.27), no superan un puntaje de 0.27 en el  $ICV_{PCA}$ . Estos municipios se ubican mayoritariamente en la zona sur del departamento, con algunos situados en las zonas periféricas. Lo cual sugiere que, aunque estos municipios muestran una ligera mejora en la calidad de vida en comparación con aquellos en el grupo anterior, todavía enfrentan desafíos significativos.

Por otro lado, los municipios en color amarillo, pertenecientes al Cuartil 3 (Q3: 0.27 – 0.38) presentan puntajes del  $ICV_{PCA}$  entre 0.27 y 0.38 y se encuentran distribuidos principalmente en la zona central del departamento; estos, podrían estar beneficiándose de un acceso más equilibrado a servicios y oportunidades en comparación con los municipios agrupados en los cuartiles Q1 y Q2. La ubicación central de estos municipios podría facilitar una mejor conectividad y acceso a infraestructuras clave, como centros educativos, de salud y redes de transporte, contribuyendo así a una calidad de vida superior.

No obstante, aunque muestran una situación relativamente favorable, sus puntajes no superan el 0.38. Esto indica que, aunque tienen mejores condiciones que los municipios en los cuartiles Q1 y Q2, las condiciones de calidad de vida de los habitantes de estos municipios aún están lejos de alcanzar los niveles observados en Bucaramanga.

Por último, los municipios denotados en color verde en la Figura 12 pertenecen al Cuartil 4 (Q4: 0.38 - 1) y tienen los puntajes más altos en el  $ICV_{PCA}$ , indicando mejores condiciones de calidad de vida. Estos municipios se concentran en el norte, noreste y algunas zonas centrales del departamento. Además, los municipios pertenecientes a este cuartil, también presentaron los mejores valores en al menos una variable de las analizadas en la sección 7.3.2.1.

De los 22 municipios agrupados en el cuartil 4 (Q4), la mitad, incluyendo Bucaramanga, tienen puntajes superiores a 0.5 en el  $ICV_{PCA}$  lo cual se traduce en que únicamente el 12,5% de los municipios de Santander alcanzan un  $ICV_{PCA}$  mayor a 0.5. Esta situación refleja que más del 80% estos municipios enfrentan desafíos considerables en términos de infraestructura, acceso a servicios esenciales, oportunidades de empleo y desarrollo socioeconómico.

La baja calificación en el  $ICV_{PCA}$  de esos municipios señala una calidad de vida más precaria para sus residentes, lo que posiblemente esté relacionado con la falta de inversión en estas áreas y la ausencia de políticas dirigidas a mejorar las condiciones de vida de la población local. Esto sugiere que la situación de precariedad no se limita a algunos municipios específicos, sino que es un problema generalizado en todo el departamento.

### **7.3 Índice de Calidad de Vida por Distancia P2 ( $ICV_{P2}$ )**

Como se mostró en la sección 5.3.1, la distancia P2 usa la distancia de Frechet, el coeficiente de correlación de Pearson asociado a la distancia de Frechet, la  $j$ -ésima variable y un objeto hipotético o real que servirá como punto de comparación para medir, en este caso, la calidad de vida de los municipios de Santander. Para este propósito, se seleccionó como referencia la capital del departamento, Bucaramanga. A continuación, se describe el proceso utilizado:

Paso 1. Crear un vector  $w_j = 1 \forall j = 1, 2, \dots, m$  y defina  $\varepsilon = 0.00001$ .

Paso 2. Definir  $\delta_{ij} = \frac{d_{ij}}{\sigma_j} \forall j = 1, 2, \dots, m, i = 1, 2, \dots, n$  donde  $d_{ij} = |x_{ij} - x_{\rho j}|$  y  $\rho$  corresponde al objeto de comparación.

Paso 3. Calcular la distancia de Frechet,  $DF_i = \sum_{j=1}^m \left[ \left( \frac{d_{ij}}{\sigma_j} \right) w_j \right]; i = 1, 2, \dots, n$

Paso 4. Computar el coeficiente de correlación de Pearson  $r(DF, \delta_j) \forall j = 1, 2, \dots, m$ .

Paso 5. Ordenar decrecientemente  $|r(DF, \delta_j)|$  y reindexar las variables asociadas,  $\delta_j$ , correspondientemente.

Paso 6. Computar  $Z_i = \sum_{j=1}^m \left[ \left( \frac{d_{ij}}{\sigma_j} \right) w_j \right]; i = 1, 2, \dots, n; w_j = (1 - R_{j,j-1,j-2,\dots,1}^2)$  para  $j = 2, 3, \dots, m$  y  $w_1 = 1$ .

Paso 7. Si  $\sum_{i=1}^n (DF_i - Z_i)^2 \geq \varepsilon$  entonces, reemplazar  $DF$  por  $Z$  y volver al paso 4, de lo contrario el proceso ha terminado y  $Z_i$  tendrá la distancia  $P_2$  de cada observación al objeto de comparación. De esta forma, el segundo índice para medir la calidad de vida en el departamento de Santander está dado por:

$$ICV_{P_2} = \sum_{j=1}^m \left[ \left( \frac{d_{ij}}{\sigma_j} \right) w_j \right] i = 1, 2, \dots, n, j = 2, 3, \dots, m; w_j = (1 - R_{j,j-1,j-2,\dots,1}^2) \text{ y } w_1 = 1$$

A continuación se presentan los resultados obtenidos para el  $ICV_{P_2}$ .

**Tabla 5.**

*Ranking  $ICV_{P2}$*

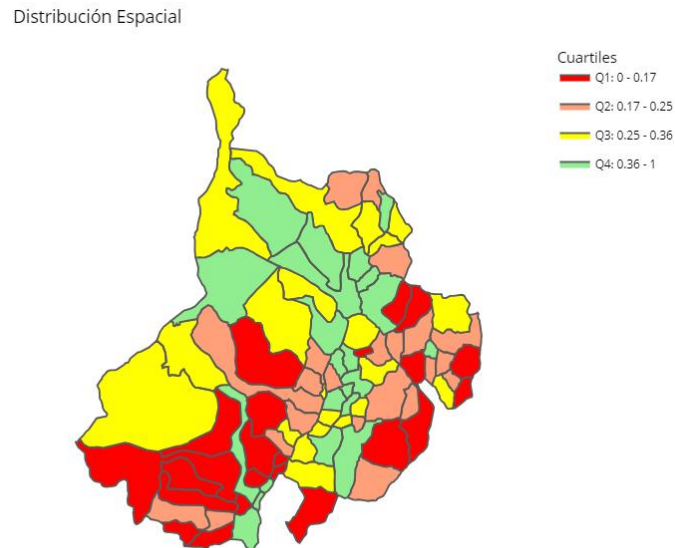
Top 7 Valores Más Altos			Top 7 Valores Más Bajos		
	NOMBRE_MPI	dp2		NOMBRE_MPI	dp2
9	BUCARAMANGA	1.0000000	28	EL PEÑON	0.0000000
33	FLORIDABLANCA	0.8362188	54	MOLAGAVITA	0.0184335
36	GIRON	0.7349444	13	CARCASI	0.0423320
61	PIEDECUESTA	0.7327667	32	FLORIAN	0.0493328
6	BARRANCABERMEJA	0.7189584	2	ALBANIA	0.0543760
70	SAN GIL	0.6946071	8	BOLIVAR	0.0708302
78	SOCORRO	0.6703410	80	SUCRE	0.0797435

En la Tabla 5 se presentan los 7 puntajes más altos y bajos del  $ICV_{P2}$ . Se observa que Floridablanca es el municipio que muestra mayor similitud en términos de calidad de vida respecto a Bucaramanga. Girón y Piedecuesta tienen puntajes muy similares, cercanos a 0.735, mientras que Barrancabermeja ocupa la quinta posición, en comparación con la tercera que tenía en el  $ICV_{PCA}$ . San Gil y Socorro mantienen sus posiciones, destacándose consistentemente en ambos índices.

Por otro lado, El Peñón es el municipio con el puntaje más bajo en el  $ICV_{P2}$ . Molagavita y Carcasí continúan ocupando puestos bajos, al igual que en el  $ICV_{PCA}$ . Además, aparecen municipios como Florián, Albania, Bolívar y Sucre, entrando en el grupo de los 7 municipios con los puntajes más bajos en el  $ICV_{P2}$ . Estos municipios ocupaban posiciones más altas en el  $ICV_{PCA}$ , encontrándose en puestos como el 62, 59, 65 y 55, respectivamente.

**Figura 13.**

*Distribución espacial por cuartiles del  $ICV_{P2}$*



En la Figura 13 se observan la distribución espacial por cuartiles del  $ICV_{P2}$  para los municipios de Santander. Los colores indican los diferentes cuartiles de puntajes, proporcionando una visión clara de las variaciones en la calidad de vida a lo largo del departamento.

Los municipios ubicados en la zona sureste y suroeste del departamento, representados en color rojo, pertenecen al Cuartil 1 (Q1: 0 – 0.17) del  $ICV_{P2}$  con un valor máximo de 0.17 en el índice. Esto indica que el 25% de los municipios de Santander presentan puntajes menores a 0.17, en comparación con el puntaje máximo registrado en el  $ICV_{PCA}$  para el Q1 cuyo valor era de 0.21.

Por otra parte, los municipios representados en color naranja corresponden al Cuartil 2 (Q2: 0.17 – 0.25) del  $ICV_{P2}$  con puntajes inferiores a 0.25. Estos municipios se encuentran en las zonas sur y periféricas del departamento. Aunque muestran una ligera mejora en los valores de los puntajes, siguen siendo valores bajos para el  $ICV_{P2}$ .

Los municipios en color amarillo pertenecen al Cuartil 3 (Q3: 0.25 – 0.36), presentan puntajes del  $ICV_{P2}$  entre 0.25 y 0.36 y se encuentran distribuidos en la zona central, el norte y las periferias del departamento. Además, se encuentran rodeados o próximos a municipios con mejores puntajes en el  $ICV_{P2}$ , pertenecientes al Cuartil 4 (Q4: 0.36 - 1).

Finalmente, los municipios denotados en color verde pertenecen al Q4 y alcanzan puntajes entre 0.36 y 1. Estos municipios se concentran en el norte, noreste y algunas zonas centrales del departamento. De los municipios del Q4, solo 9 (excluyendo Bucaramanga) tienen puntajes superiores a 0.5, lo que indica que solo el 10% de los municipios tienen al menos la mitad de las condiciones de vida registradas en Bucaramanga.

Sin embargo, se debe mencionar que los puntajes para los cuartiles del  $ICV_{P2}$  son notablemente menores que los registrados para el  $ICV_{PCA}$ . Lo que sugiere que el método de Distancia P2 produce valores menores para los puntajes del  $ICV_{P2}$  en comparación con los del  $ICV_{PCA}$ . Esta diferencia puede ser indicativa de que el  $ICV_{P2}$  proporciona una evaluación más rigurosa y posiblemente más precisa de la calidad de vida en los municipios, reflejando de manera más efectiva las carencias y desigualdades presentes en la región.

Además, esta observación podría implicar que el  $ICV_{PCA}$  podría estar sobreestimando las condiciones de vida en ciertos municipios debido a su metodología, mientras que el  $ICV_{P2}$ , al ser más estricto en su análisis, revela un panorama más realista y menos optimista. Esta discrepancia subraya la importancia de utilizar múltiples métodos de evaluación para obtener una comprensión más completa y equilibrada de la calidad de vida en cualquier región.

### 7.4 Índice de Moran

Con la finalidad de evaluar si los valores de los Índices de Calidad de Vida ( $ICV_{PCA}$ ,  $ICV_{P2}$ ), construidos en las secciones 7.3 y 7.2, presentan un patrón espacial y determinar hasta qué punto están correlacionados espacialmente, se realizó el Test de Índice de Moran en R-Studio utilizando la función *moran.test()*. A continuación, se presentan los resultados obtenidos.

#### Figura 14.

*Test Índice de Moran I para  $ICV_{PCA}$*

```

Moran I test under randomisation

data: indpca
weights: pesos

Moran I statistic standard deviate = 5.137, p-value = 1.396e-07
alternative hypothesis: greater
sample estimates:
Moran I statistic      Expectation      Variance
    0.317796373      -0.011627907      0.004112318
    
```

#### Figura 15.

*Test Índice de Moran I para  $ICV_{P2}$*

```

Moran I test under randomisation

data: indp2
weights: pesos

Moran I statistic standard deviate = 4.3854, p-value = 5.789e-06
alternative hypothesis: greater
sample estimates:
Moran I statistic      Expectation      Variance
    0.269821435      -0.011627907      0.004118934
    
```

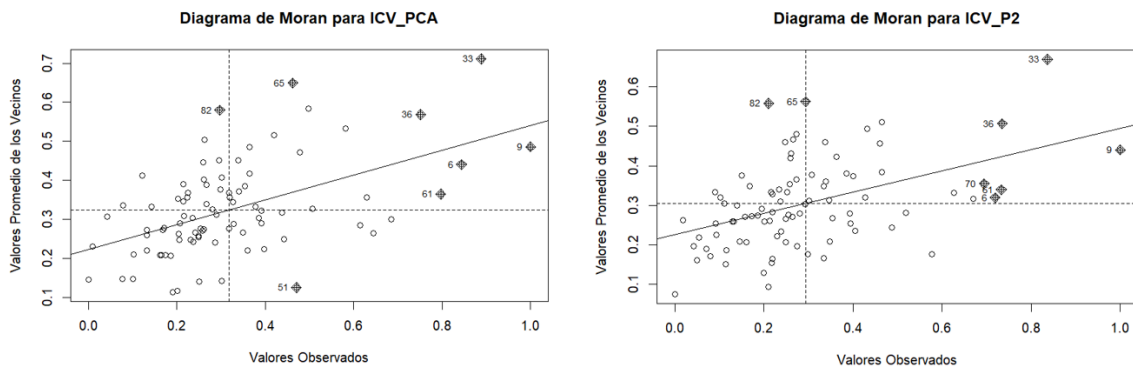
En las Figuras 14 y 15 se muestra que para el  $ICV_{PCA}$  el Índice de Moran tiene un valor de 0.317 y para el  $ICV_{P2}$  el Índice de Moran es de 0.269. Además, su *p*-valor es menor a 0.05. Esto indica que, independientemente del índice de calidad de vida considerado, existe una

autocorrelación espacial positiva significativa. En otras palabras, estos resultados ponen en evidencia que los municipios con valores altos (o bajos) de los índices tienden a estar agrupados.

Esta afirmación se verifica en los diagramas de dispersión presentados en la Figura 16, donde se observa que tanto para el  $ICV_{P2}$  y el  $ICV_{PCA}$  la ubicación de los municipios tiende a estar entre el III y I cuadrante. En el caso del  $ICV_{PCA}$  se observa que los valores de los municipios presentan mayor dispersión en comparación con los valores para  $ICV_{P2}$ . Esto se puede interpretar como una mayor diversidad en los valores del índice de calidad de vida ajustado por componentes principales, lo cual podría estar influenciado por una mayor cantidad de factores que afectan la calidad de vida en los municipios.

**Figura 16.**

*Diagrama de dispersión de Moran para  $ICV_{PCA}$  e  $ICV_{P2}$*



Además, se evidencia que los municipios que se encuentran en el top de los valores más altos en ambos índices (Tabla 4 y 5) son los que se encuentran ubicados con valores atípicos en el diagrama de Moran. Por otro lado, se distingue un agrupamiento de puntos en el cuadrante I del Diagrama de Moran para  $ICV_{P2}$ , los cuales representan a Piedecuesta (61), Barrancabermeja (6),

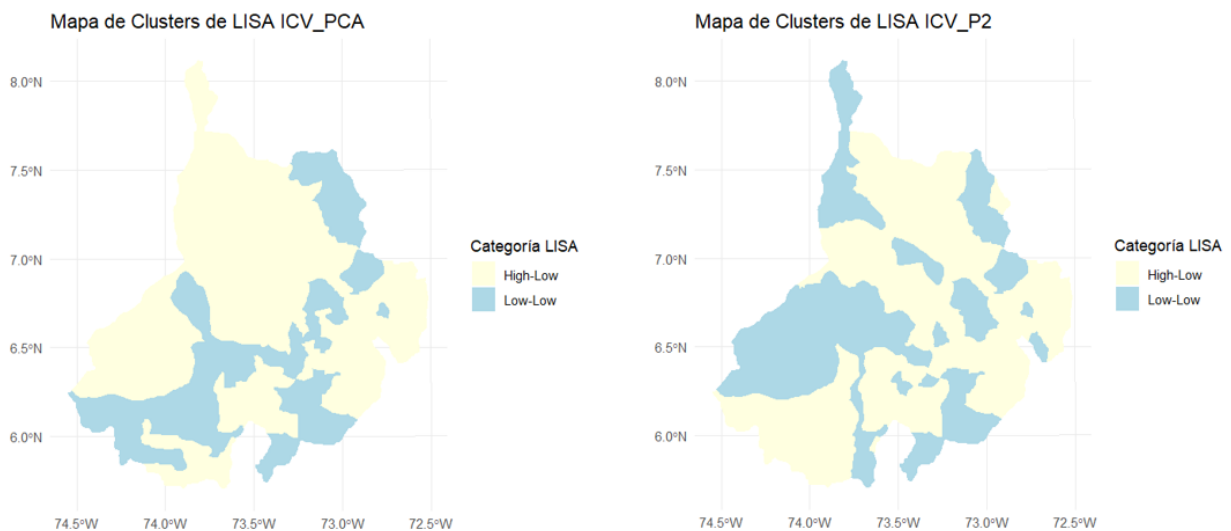
San Gil (70) y otros dos municipios. Los puntos en este cuadrante muestran que los municipios con valores altos en el  $ICV_{P2}$  tienen vecinos que también tienen valores altos.

### 7.5 Indicadores Locales de Autocorrelación Espacial (LISA)

En la sección anterior se demostró que los valores del  $ICV_{P2}$  y el  $ICV_{PCA}$  presentan un patrón espacial y se identifican grupos o clústers de municipios con valores similares. Para complementar este análisis, se elaboraron dos gráficos LISA. Estos gráficos permiten identificar áreas específicas donde los municipios tienen valores similares para los  $ICV$ , visualizar el patrón espacial, detectar outliers espaciales (aquellos municipios con valores muy diferentes a sus vecinos) y analizar la influencia local de cada municipio.

**Figura 17.**

*Clústers LISA para  $ICV_{PCA}$  e  $ICV_{P2}$*



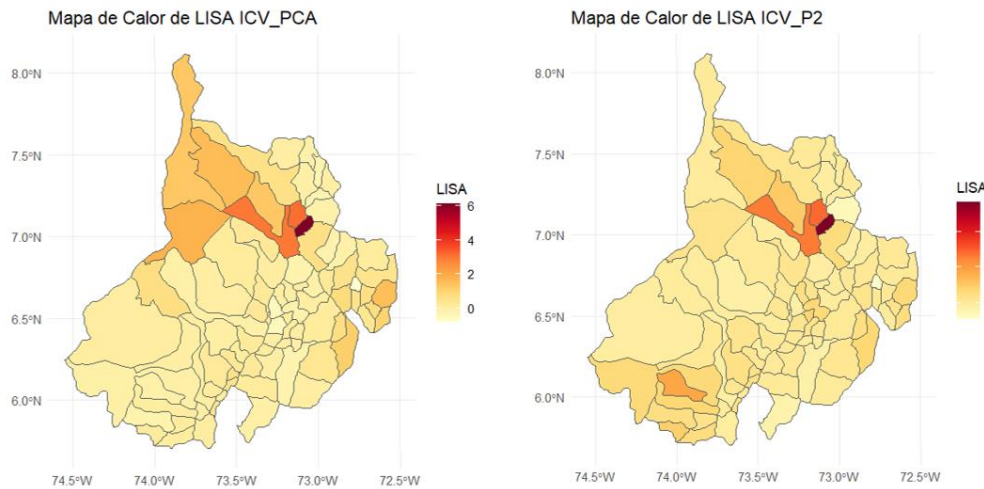
En la Figura 17 se presenta el primer resultado del LISA para  $ICV_{P2}$  e  $ICV_{PCA}$ . En estos mapas, los municipios en amarillo presentan asociación Alta-Baja. Es decir, municipios que tienen

valores altos para este índice y están rodeados de municipios con valores bajos. Por otro lado, los municipios en celeste muestran asociación Baja-Baja. Esto indica que estos municipios tienen valores bajos están rodeados de municipios con valores igualmente bajos en este índice.

Se observa en ambos mapas que en el noreste del departamento se encuentran municipios con asociaciones positivas fuertes contiguos a municipios con asociaciones negativas fuertes. Esta proximidad de asociaciones contrastantes sugiere la presencia de factores locales que podrían estar influyendo en la variabilidad de los índices construidos en esta zona del departamento. Además, es notable que el  $ICV_{P2}$  presenta mayor variabilidad en la distribución de las relaciones.

**Figura 18.**

*Mapa de Calor de LISA para  $ICV_{PCA}$  e  $ICV_{P2}$*



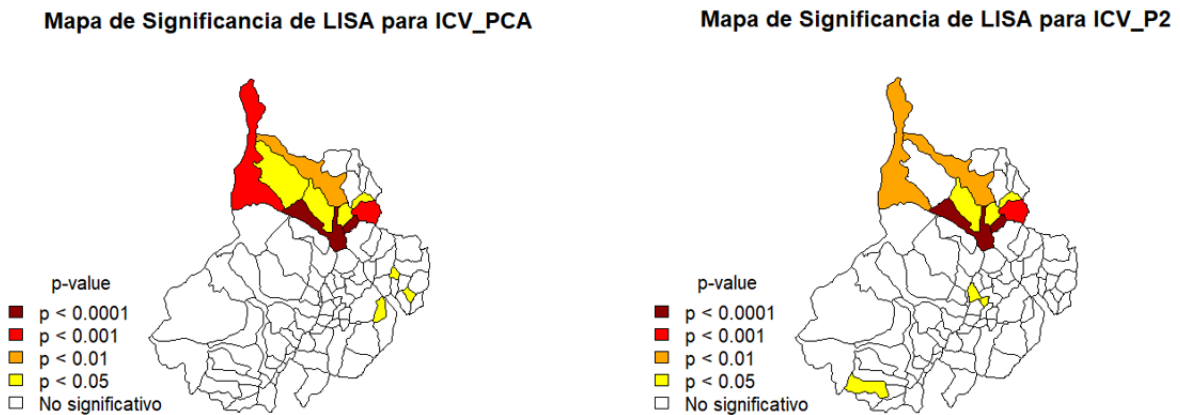
Se observa que en los mapas de calor de LISA para ambos índices los municipios del noreste muestran una fuerte concentración de valores altos de LISA, indicando una asociación significativa. Esto indica que estos municipios tienen una relación espacial muy intensa con sus vecinos. En el caso del  $ICV_{P2}$  la intensidad máxima es ligeramente menor que para el  $ICV_{PCA}$ . Por

otro lado, al oeste y sur del departamento la intensidad del LISA disminuye, sugiriendo asociaciones positivas débiles o neutrales en los municipios ubicados en estas zonas.

Además, en el caso del  $ICV_{PCA}$ , la homogeneidad es más evidente, especialmente en el noreste. Para  $ICV_{P2}$ , la variabilidad es mayor, lo que sugiere la existencia de diversos factores que influyen en la estructura espacial de los municipios.

**Figura 19.**

*Mapa de Significancia de LISA para  $ICV_{PCA}$  e  $ICV_{P2}$*



Para los mapas de significancia de los  $ICV$  se observa que los municipios del norte, especialmente los que limitan con la capital del departamento, Bucaramanga, presentan autocorrelación espacial significativa. Los municipios con autocorrelación significativa forman clústers espaciales, lo que sugiere que los valores de  $ICV_{PCA}$  e  $ICV_{P2}$  no son aleatorios y tienden a estar agrupados. La comparación entre ambos mapas muestra que los patrones espaciales son consistentes, con varios municipios mostrando significancia en ambos índices.

Sin embargo, el valor  $p$  varía dependiendo del índice evaluado. Por ejemplo, el municipio de Sabana de Torres tiene una autocorrelación espacial significativa con valor  $p < 0.05$  para el  $ICV_{PCA}$ , mientras que para el  $ICV_{P2}$  no muestra autocorrelación espacial significativa.

## 7.6 Visualizador Índice de Calidad de Vida en los municipios de Santander

Para la visualización de los resultados obtenidos y un acercamiento más efectivo a los datos, se desarrolló un Geoportal utilizando el software R, empleando las bibliotecas FlexDashboard y Shiny. Este Geoportal tiene como objetivo proporcionar información accesible al público, permitiendo a los usuarios explorar la distribución espacial de los Índices de Calidad de Vida desarrollados ( $ICV_{PCA}$  e  $ICV_{P2}$ ), así como las variables más relevantes que los explican.

A través de una interfaz intuitiva y amigable, los usuarios pueden interactuar de manera efectiva con los datos, realizando comparaciones detalladas de los valores en los municipios de Santander para los indicadores evaluados. Esta funcionalidad permite no solo observar la distribución de los indicadores de calidad de vida, sino también explorar patrones y tendencias que pueden no ser evidentes a simple vista. Además, el Geoportal ofrece resultados sobre indicadores espaciales, como el análisis LISA, que ayuda a identificar áreas de concentración y relación entre variables. Esta herramienta permite a los usuarios detectar clústers de municipios con altos y bajos índices de calidad de vida, facilitando la comprensión de las dinámicas regionales.

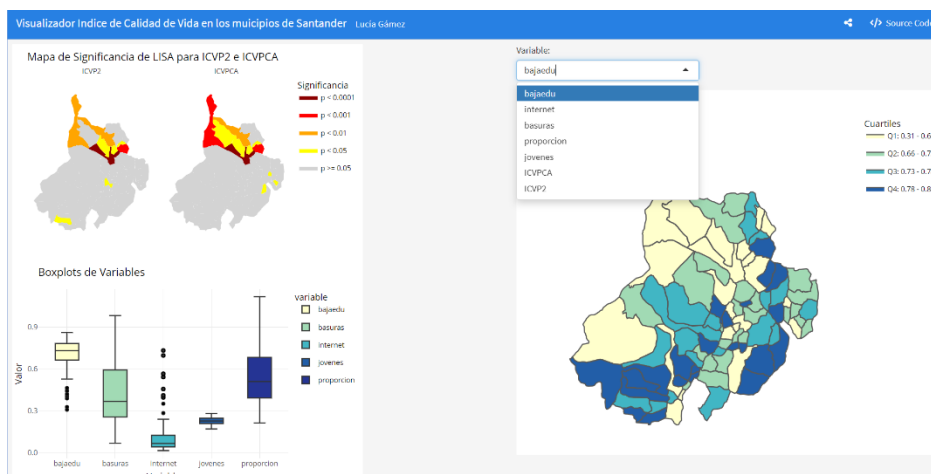
En la Figura 20 se presenta la interfaz del Geoportal, que se organiza en tres secciones claramente definidas. En la parte superior izquierda, se pueden visualizar los mapas de significancia del análisis LISA, los cuales permiten identificar grupos de municipios que presentan patrones de asociación significativos. Estos mapas destacan las áreas donde se concentran

municipios con puntajes altos en los índices de calidad de vida, así como aquellos con puntajes bajos. Esta visualización facilita el análisis de cómo las condiciones socioeconómicas se distribuyen en la región, ayudando a identificar clusters de similitudes y disparidades entre los municipios de Santander.

En la parte inferior izquierda se observa la distribución de las variables más relevantes para el  $ICV_{PCA}$ , esto con el fin de observar el comportamiento de estas a nivel departamental. Por último, en la parte derecha de la interfaz se presenta la distribución espacial de estas variables y de los índices desarrollados en el departamento. Esta visualización se clasifica en cuartiles, lo que facilita la identificación de áreas con diferentes niveles de calidad de vida. Además, el Geoportal ofrece la opción de filtrar los datos según la variable o índice que se desee explorar, permitiendo a los usuarios personalizar su análisis y profundizar en la información específica que les interesa. En el Apéndice D se incluye el repositorio donde los usuarios pueden acceder y utilizar el Geoportal desarrollado.

**Figura 20.**

*Interfaz del Visualizador del ICV*



## 8. Discusión

Como se evidenció en la sección de resultados, los municipios de Santander presentaron condiciones de vida poco favorables, con puntuaciones bajas en diversas variables relacionadas con la calidad de vida. Entre las principales dificultades se destacan los altos índices de pobreza, la limitada infraestructura de servicios públicos y la falta de acceso a oportunidades educativas y laborales, especialmente en las zonas rurales. Los municipios más pequeños y alejados del área metropolitana de Bucaramanga enfrentan mayores desafíos, tanto en términos de desarrollo económico como de bienestar social.

Aunque Bucaramanga logró mantener niveles de empleo más estables en 2018 en comparación con otras ciudades principales del país, la realidad en los municipios rurales es bastante diferente. El desempleo rural, impulsado por una economía dependiente de la agricultura y la escasez de industrias que diversifiquen las oportunidades laborales, sigue siendo un problema significativo. Según datos del Departamento Nacional de Planeación (DNP) en 2018, algunos municipios rurales de Santander alcanzaron tasas de desempleo superiores al 10%, frente a una tasa de desempleo en Bucaramanga del 8.8%, lo que refleja las dificultades estructurales para generar empleo fuera de las grandes urbes.

Además, la migración de jóvenes hacia las ciudades en busca de mejores oportunidades económicas deja a las zonas rurales con una fuerza laboral envejecida, lo que dificulta aún más la creación de empleo sostenible y el crecimiento económico en estas áreas. Este fenómeno, acompañado por la falta de inversión en infraestructura y servicios públicos, perpetúa las brechas entre el desarrollo urbano y rural en Santander.

Por otra parte, El Informe Nacional de Calidad de Vida 2018 elaborado por el DANE subrayaba que en muchos municipios rurales de Santander, una proporción significativa de la población no alcanzaba el nivel educativo esperado para su edad. En algunas localidades, los bajos niveles de cobertura y la limitada infraestructura escolar afectaban de manera directa los logros educativos de niños y adolescentes. Según los datos del DANE, el bajo logro educativo era más pronunciado en las zonas rurales del departamento, donde hasta un 20% de la población mayor de 15 años no había culminado la educación básica secundaria.

Esta disparidad en los logros educativos entre Bucaramanga y los municipios rurales se veía agravada por la falta de recursos en las instituciones educativas rurales y las barreras geográficas que dificultaban el acceso a la educación. Mientras Bucaramanga se beneficiaba de una mayor oferta educativa y acceso a programas de formación técnica y universitaria, en las zonas más alejadas del área metropolitana las oportunidades eran considerablemente más limitadas, perpetuando las brechas de desarrollo humano.

En 2018, los valores asociados con el servicio de recolección de basura presentaron una notable disparidad entre Bucaramanga y los demás municipios de Santander. Según el Departamento Administrativo Nacional de Estadística (DANE), Bucaramanga se destacó por ofrecer un servicio de recolección de basura más eficiente y accesible en comparación con muchos de los municipios rurales del departamento. En su informe, el DANE indicó que la cobertura del servicio de recolección en Bucaramanga alcanzaba más del 95%, lo que garantizaba un manejo adecuado de los residuos sólidos y contribuía a mantener condiciones sanitarias favorables en la ciudad.

Por otro lado, muchos municipios rurales de Santander enfrentaban serias dificultades en la gestión de residuos. Según el Informe de Calidad de Vida 2018 del DANE, en varias localidades de la región, la cobertura del servicio de recolección de basura era significativamente más baja, llegando a menos del 70% en algunos casos. Esta falta de acceso a servicios de recolección eficientes no solo afectaba la salubridad pública, sino que también generaba problemas ambientales, al permitir que los residuos se acumularan en las calles y espacios públicos (DANE, 2018).

La diferencia en la calidad de los servicios de recolección de basura refleja una de las muchas brechas que existen entre Bucaramanga y los municipios rurales de Santander, exacerbando las desigualdades en términos de calidad de vida. Mientras Bucaramanga puede invertir en infraestructura y tecnología para la gestión de residuos, muchos municipios rurales continúan lidiando con una infraestructura insuficiente, lo que limita su capacidad para ofrecer servicios básicos adecuados a sus ciudadanos.

Por otro lado, la proximidad a Bucaramanga generó una serie de beneficios significativos para los municipios circundantes en Santander, así como dificultades para aquellos más alejados de la capital. La ubicación estratégica de los municipios cercanos a Bucaramanga les permitió acceder a una infraestructura de servicios públicos más desarrollada, oportunidades de empleo en sectores diversos y una mayor oferta educativa y de salud. Según un estudio del Observatorio de Desarrollo Regional de la Universidad Industrial de Santander (UIS, 2018), los municipios que limitan con Bucaramanga experimentaron un incremento en su desarrollo económico debido a la facilidad de acceso a mercados, recursos y servicios.

Además, la cercanía a Bucaramanga facilitó la movilidad de los residentes, permitiéndoles beneficiarse de la infraestructura de transporte y la disponibilidad de servicios especializados que la capital ofrece. Por ejemplo, la Secretaría de Planeación de Santander (2018) destacó que los municipios aledaños lograron mantener índices de bienestar superiores, en gran medida gracias a su acceso a oportunidades laborales y educativas en Bucaramanga.

Se destaca que para el  $ICV_{P2}$  Barrancabermeja obtuvo el puntaje más alto entre los municipios de Santander, solo superado por aquellos de la zona metropolitana de Bucaramanga. Según un informe de la Alcaldía de Barrancabermeja (2018), la ciudad cuenta con una infraestructura robusta y servicios públicos relativamente bien desarrollados, lo que ha contribuido a mejorar las condiciones de vida de sus habitantes. Además, la oferta de empleo en el sector energético ha generado una base económica sólida que ha permitido un crecimiento constante en el ingreso per cápita de sus ciudadanos.

Un estudio realizado por el Observatorio de Desarrollo Económico de la Universidad Industrial de Santander (UIS) en 2018 resalta que Barrancabermeja no solo ha atraído inversión en infraestructura, sino que también ha desarrollado programas sociales que han fortalecido la educación y la salud en la región. Esto ha resultado en una reducción significativa de la pobreza y una mejora en los indicadores de bienestar social (UIS, 2018). Esta combinación de factores económicos y educativos permite que Barrancabermeja, a pesar de su ubicación geográfica, logre ofrecer una calidad de vida superior a la de muchos municipios rurales en Santander.

Adicionalmente, San Gil, que ocupa la 6ª posición, es conocido por su atractivo turístico y su desarrollo económico basado en el ecoturismo. La Alcaldía de San Gil (2018) informa que el municipio ha promovido una serie de iniciativas para mejorar la infraestructura turística, lo que ha

incrementado el flujo de visitantes y, a su vez, ha generado empleos y oportunidades económicas para los residentes. El crecimiento del sector turístico ha permitido que más ciudadanos accedan a servicios básicos de calidad, lo que se traduce en una mejora en la calidad de vida.

Un estudio realizado por la Secretaría de Planeación de Santander (2018) menciona que la inversión en educación y la creación de espacios recreativos han contribuido significativamente a la satisfacción de la población local. La combinación de estos factores ha llevado a San Gil a posicionarse favorablemente en el ranking de calidad de vida, destacándose por sus altos índices de felicidad y bienestar social.

Por su parte, Socorro se ubicó en la 7ª posición, gracias a su desarrollo integral y sus políticas locales enfocadas en el bienestar social. Según el DANE (2018), Socorro ha implementado programas de desarrollo rural y urbano que han mejorado la infraestructura de servicios básicos, como agua potable y educación. Esta mejora en la infraestructura ha permitido que más habitantes tengan acceso a condiciones de vida dignas.

Un informe del Centro de Estudios Económicos de la Universidad de Los Andes (2018) destaca que la inversión en proyectos sociales y la promoción de la cultura local han jugado un papel crucial en la mejora de la calidad de vida en Socorro. Los programas orientados a la inclusión social han permitido que la población participe activamente en la construcción de su entorno, lo que ha generado un sentido de pertenencia y cohesión social que se refleja en los índices de calidad de vida.

En contraste, los municipios más alejados enfrentaron serias dificultades que limitaron su desarrollo. La distancia geográfica complicaba el acceso a servicios básicos como salud, educación y transporte, lo que se traducía en condiciones de vida menos favorables. Según el DANE (2018),

los municipios remotos presentaban índices más altos de pobreza y menores niveles de satisfacción en términos de calidad de vida. La falta de infraestructura adecuada en estas áreas dificultaba la implementación de programas sociales y de desarrollo económico, perpetuando ciclos de pobreza y exclusión social.

Málaga se encuentra a aproximadamente 130 kilómetros al norte de Bucaramanga. Esta distancia impacta negativamente en el acceso a servicios básicos y oportunidades de empleo. La infraestructura vial es deficiente en varias áreas, lo que complica el transporte de productos agrícolas y limita el acceso a mercados más grandes. Según el DANE (2018), la cobertura de servicios de salud en Málaga es limitada, con muchas comunidades rurales que no tienen acceso a atención médica adecuada. Esta situación contribuye a que el municipio presente altos índices de pobreza y baja calidad de vida en comparación con las áreas más urbanizadas.

Vélez está situado a cerca de 120 kilómetros de Bucaramanga y también enfrenta serias dificultades debido a su lejanía. El acceso a servicios educativos es un desafío, especialmente para los jóvenes que desean continuar su formación en niveles superiores. La distancia a Bucaramanga limita las opciones de educación técnica y profesional, lo que afecta las oportunidades laborales de los residentes. Según un informe de la Secretaría de Planeación de Santander (2018), Vélez presenta índices de desempleo más altos que la media departamental, exacerbados por la falta de industrias y la dependencia de la agricultura, que no proporciona suficientes empleos estables.

Guapotá, ubicado a aproximadamente 135 kilómetros al sureste de Bucaramanga, también experimenta dificultades relacionadas con su lejanía. La escasa cobertura de servicios públicos, como agua potable y saneamiento básico, es una preocupación constante en el municipio. Según el DANE (2018), muchos hogares en Guapotá no cuentan con acceso adecuado a estos servicios,

lo que repercute en la salud y calidad de vida de sus habitantes. Además, la falta de infraestructura adecuada limita el desarrollo económico y social, perpetuando condiciones de pobreza y exclusión.

La situación de los municipios de Santander en relación con su proximidad a Bucaramanga y los desafíos que enfrentan debido a su lejanía se refleja en patrones similares en otras regiones de Colombia. Por ejemplo, en el departamento de Antioquia, se observa una disparidad significativa entre Medellín y sus municipios vecinos. Según el Informe de Calidad de Vida de la Alcaldía de Medellín (2018), mientras que Medellín goza de una infraestructura sólida y acceso a servicios públicos eficientes, los municipios aledaños como San Pedro de los Milagros y Caldas presentan índices de pobreza más altos y menor acceso a servicios básicos, lo que agrava las desigualdades regionales.

Además, en la región del Caribe, se pueden observar problemas similares. En departamentos como Cesar y La Guajira, la lejanía de los centros urbanos, como Valledupar y Riohacha, contribuye a condiciones de vida precarias. Según un estudio del Banco Mundial (2018), la falta de infraestructura vial y la escasa cobertura de servicios de salud en estas regiones han perpetuado ciclos de pobreza y han limitado el desarrollo económico. La falta de acceso a mercados y oportunidades educativas también ha generado brechas significativas en la calidad de vida de los habitantes de municipios más alejados.

En resumen, la situación de los municipios de Santander, marcada por disparidades significativas en acceso a servicios y oportunidades debido a su ubicación, refleja patrones que se pueden observar en diversas regiones de Colombia. Las diferencias en calidad de vida entre áreas urbanas y rurales, así como entre municipios cercanos y lejanos a las capitales, son un desafío común que requiere atención y políticas públicas efectivas a nivel nacional.

La brecha entre los municipios cercanos y lejanos a Bucaramanga resalta la importancia de la localización geográfica en el desarrollo regional, indicando que la proximidad a la capital no solo influye en el acceso a recursos, sino también en la calidad de vida de los habitantes. Para abordar estos problemas, se pueden implementar varias soluciones políticas que busquen mejorar la conectividad, el acceso a servicios básicos y la calidad de vida en los municipios más distantes.

En primer lugar, es crucial mejorar la infraestructura vial, ya que la construcción y mantenimiento de carreteras y vías de acceso facilitarán el transporte de personas y mercancías entre los municipios lejanos y Bucaramanga, fomentando el comercio local y la integración regional. Además, se podrían establecer programas de subsidios y apoyo a la educación que ofrezcan becas a estudiantes de municipios lejanos para acceder a educación superior en Bucaramanga, incluyendo incentivos para instituciones educativas que amplíen sus servicios en estas áreas.

También es fundamental fortalecer los servicios de salud mediante la creación de unidades móviles que visiten regularmente los municipios más alejados y la formación de profesionales de la salud locales. Asimismo, el fomento del desarrollo económico local a través de incentivos fiscales y financieros para atraer inversiones en sectores como la agricultura sostenible y el ecoturismo contribuirá a diversificar la economía y crear empleos. Por último, aumentar el acceso a Internet y las tecnologías de la información en los municipios lejanos es esencial para mejorar la educación y el desarrollo económico, lo que se podría lograr mediante asociaciones público-privadas que expandan la infraestructura de telecomunicaciones.

En conjunto, estas políticas públicas, bien diseñadas y ejecutadas en colaboración con diferentes niveles de gobierno y comunidades locales, contribuirán a cerrar las brechas y promover

un desarrollo más equitativo en Santander. Además, la participación activa de las comunidades en la planificación e implementación de estas iniciativas es fundamental para asegurar que las soluciones sean adecuadas a sus necesidades específicas. La promoción de un diálogo constante entre los ciudadanos y las autoridades permitirá identificar y priorizar los problemas más urgentes, fomentando un sentido de pertenencia y responsabilidad compartida en el desarrollo local.

## 9. Conclusiones

Los resultados obtenidos a través del  $ICV_{PCA}$  y el  $ICV_{P2}$  para el año 2018 revelan una notable desigualdad en la calidad de vida entre los municipios de Santander. Los municipios del Área Metropolitana de Bucaramanga, como Floridablanca, Girón y Piedecuesta, muestran condiciones de vida comparables a la capital del departamento, Bucaramanga (Ver Tablas 4 y 5). Esta ventaja se debe principalmente a su ubicación en áreas urbanizadas, que les proporcionan mejor infraestructura, servicios públicos eficientes y acceso a educación de alta calidad.

De igual forma, destinos turísticos como San Gil y Socorro también muestran condiciones relativamente favorables en comparación con otros municipios, especialmente aquellos situados en las zonas sureste y suroeste del departamento (Ver Figura 12 y 13). En estos municipios se encuentran instituciones de educación superior que, junto con la actividad turística, generan empleo, impulsan la economía local y mejoran el acceso a servicios públicos.

La alta variabilidad observada en la cobertura de servicios como la recolección de basuras e internet sugiere disparidades significativas entre los municipios (Ver Figura 8 y 9). Los valores atípicos en estas variables indican que solo unos pocos municipios cuentan con una cobertura adecuada, mientras que la mayoría enfrenta limitaciones. Invertir en mejorar la infraestructura y el acceso a estos servicios es crucial para reducir estas disparidades.

Ambos índices muestran que el noreste del departamento, particularmente los municipios que limitan con Bucaramanga, es una región con asociaciones positivas fuertes (Ver Figura 19). Esto sugiere la presencia de áreas con características económicas o sociales similares y robustas, lo que podría indicar una alta homogeneidad en estos aspectos.

El  $ICV_{P2}$  parece proporcionar una evaluación más estricta y realista de la calidad de vida, muestra valores más bajos en comparación con el  $ICV_{PCA}$  (Ver Tabla 5). Al compararse con Bucaramanga, el  $ICV_{P2}$  utiliza un punto de referencia que ayuda a entender mejor las disparidades entre los municipios. Esta comparación permite identificar no solo las deficiencias, sino también las áreas donde existen oportunidades para el desarrollo.

Además, la menor variabilidad en los puntajes del  $ICV_{P2}$  proporciona una visión clara y concisa de las necesidades de cada municipio (Ver Figura 17). Al destacar las áreas que requieren mayor atención, permite una focalización más eficiente de los recursos, esto es fundamental para la planificación estratégica permitiendo a los responsables de la toma de decisiones diseñar intervenciones específicas que aborden las desigualdades de manera más efectiva. También, puede servir como una herramienta de monitoreo continuo para evaluar el impacto de las políticas implementadas.

En resumen, los análisis del  $ICV_{PCA}$  y el  $ICV_{P2}$  proporcionan una visión detallada de las precarias condiciones de vida en los municipios de Santander. El fortalecimiento de la infraestructura, la mejora en el acceso a servicios básicos y la promoción del desarrollo económico y educativo son vitales para elevar la calidad de vida en los municipios más rezagados. Es esencial desarrollar estrategias de intervención que aborden las disparidades regionales para promover la distribución equitativa de los recursos y oportunidades.

**Referencias Bibliográficas**

- Alcaldía de Barrancabermeja. (2018). Informe de Desarrollo Económico y Social. Barrancabermeja, Colombia. Recuperado de <http://www.barrancabermeja.gov.co>
- Alcaldía de Medellín. (2018). Informe de Calidad de Vida 2018. Medellín, Colombia. Recuperado de <https://www.medellin.gov.co>
- Alcaldía de San Gil. (2018). Plan de Desarrollo Municipal 2016-2019. San Gil, Colombia. Recuperado de <http://www.sangil.gov.co>
- Alegret Rodríguez, M., Herrera, M., & Grau Abalo, R. (2008). Las técnicas de estadística espacial en la investigación salubrista. Caso síndrome de Down. *Revista Cubana de Salud Pública*, 34(4), 1-11.
- Anselin, L. (1995). *Geographical analysis* (Vol. 27). Ohio State University Press.
- Anselin, L. (2003). Spatial autocorrelation refresher. CSISS, Dept. Agricultural and Consumer Economics, University of Illinois, Urbana-Champaign.
- Banco Mundial. (2018). Informe sobre Desarrollo en Colombia. Washington, D.C. Recuperado de <https://www.bancomundial.org>
- Chasco, C. (2003). Métodos gráficos del análisis exploratorio de datos espaciales [Resumen]. Universidad Autónoma de Madrid.
- Celemín, Juan Pablo. (2009). Autocorrelación espacial e indicadores locales de asociación espacial: Importancia, estructura y aplicación. *Revista Universitaria de Geografía*, 18(1), 11-31. Recuperado en 26 de julio de 2024, de

[http://www.scielo.org.ar/scielo.php?script=sci\\_arttext&pid=S1852-42652009000100002&lng=es&tlng=es](http://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S1852-42652009000100002&lng=es&tlng=es)

Centro de Estudios Económicos de la Universidad de Los Andes. (2018). Análisis de la calidad de vida en los municipios de Santander: el caso de Socorro. Bogotá, Colombia. Recuperado de <http://www.uniandes.edu.co>

Departamento Administrativo Nacional de Estadística (DANE). (2018). Boletín Técnico: Mercado Laboral en Bucaramanga y otras ciudades principales, 2018. Bogotá, Colombia. Recuperado de <https://www.dane.gov.co>

Departamento Administrativo Nacional de Estadística (DANE). (2018). Informe sobre la Calidad de Vida en Socorro. Bogotá, Colombia. Recuperado de <https://www.dane.gov.co>

DANE. (2017). Hogares con acceso a recolección de residuos. <https://www.dane.gov.co/files/investigaciones/pib/ambientales/IIac/18-Recoleccion-de-desechos-2017/Hogares-con-acceso-a-recoleccion-de-residuos.pdf>

Departamento Nacional de Planeación. (2011, 8 de noviembre). Índice de Pobreza Multidimensional para Colombia (IPM-Colombia) 1997-2010. Recuperado de <https://colaboracion.dnp.gov.co/cdt/estudios%20econmicos/382.pdf>

Departamento Nacional de Planeación (DNP). (2018). Informe de Desempeño del Mercado Laboral Rural en Colombia 2018. Bogotá, Colombia. Recuperado de <https://www.dnp.gov.co>

Departamento Administrativo Nacional de Estadística (DANE). (2018). Informe Nacional de Calidad de Vida 2018. Bogotá, Colombia. Recuperado de <https://www.dane.gov.co>

- Departamento Administrativo Nacional de Estadística (DANE). (2018). Informe sobre la Calidad de Vida en los Municipios de Santander. Bogotá, Colombia. Recuperado de <https://www.dane.gov.co>
- Departamento Nacional de Planeación (DNP). (2019). Indicadores de Logro Educativo en Municipios de Colombia. Bogotá, Colombia. Recuperado de <https://www.dnp.gov.co>
- Departamento Nacional de Planeación (DNP). (2018). Informe sobre la Gestión de Residuos Sólidos en Colombia. Bogotá, Colombia. Recuperado de <https://www.dnp.gov.co>
- DeGroot, M. H., & Schervish, M. J. (2012). Probability and statistics (4th ed.). Pearson
- Fajardo Ramos, E., Córdoba Andrade, L., & Enciso Luna, J. E. (2016). El envejecimiento poblacional en América Latina y el Caribe: desafíos de las políticas sociales ante el nuevo siglo. Recuperado de [https://ve.scielo.org/scielo.php?script=sci\\_arttext&pid=S1690-32932016000200005](https://ve.scielo.org/scielo.php?script=sci_arttext&pid=S1690-32932016000200005)
- Gamir Orueta, Agustín, Seguí Pons, Joana María, et al. (1995) Prácticas de análisis espacial. Barcelona: Oikos-Tau.
- Graham, C., & Chaparro, J. C. (2021). Inseguridad, salud y bienestar: Una exploración inicial basada en encuestas sobre la felicidad en América Latina y el Caribe. Revista Latinoamericana de Opinión Pública.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). Multivariate Data Analysis (8th ed.). Cengage Learning.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer. <https://doi.org/10.1007/978-0-387-84858-7>

- Hidalgo Bucheli, G. E. (2019). Uso del índice de Moran y LISA para explicar el ausentismo electoral rural en Ecuador. *Revista Geográfica*, 160, 1-15.
- Kaiser, H.F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141-151.
- Jackson, J. E. (1991). *A User's Guide to Principal Components*. John Wiley & Sons.
- Jaramillo, L. A. E. (2010). El valor económico de la calidad ambiental urbana: Un análisis espacial para las comunas y barrios de Cali. Universidad del Valle.
- Jaramillo, L. A. E. (2008). Indicadores ambientales sintéticos: una aproximación conceptual desde la estadística multivariante. *Gestión y Ambiente*, 11(1), 121-140.  
<http://www.redalyc.org/pdf/1694/169414452008.pdf>
- Johnson, R. A., & Wichern, D. W. (1988). *Applied Multivariate Statistical Analysis*. Biometrics
- Johnson, R. A. (2011). *Probabilidad y estadística para ingenieros (8a ed.)*. Pearson Educación.
- Jolliffe, I. T. (2002). *Principal Component Analysis*. Springer.
- Montgomery, D. C., & Runger, G. C. (2018). *Estadística aplicada y probabilidad para ingenieros (6a ed.)*. Wiley.
- Moran, P. A. P. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society: Series B (Methodological)*, 10(2), 243-251. <http://www.jstor.org/pss/2983777>
- Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO). (2018). *Informe sobre la Migración Rural y Juventud en Colombia*. Bogotá, Colombia.  
Recuperado de <https://www.fao.org>
- Pena Traperero, J. B. (1977). *Problemas de la Medición del Bienestar y Conceptos Afines: una aplicación al caso español*. INE, Madrid.

Secretaría de Planeación de Santander. (2018). Estudio sobre el Impacto del Turismo en la Calidad de Vida en San Gil. Bucaramanga, Colombia. Recuperado de

<http://www.santander.gov.co>

Secretaría de Planeación de Santander. (2018). Informe de Desarrollo Económico y Social de los Municipios de Santander. Bucaramanga, Colombia. Recuperado de

<http://www.santander.gov.co>

Sepúlveda Murillo, F. H., Chica-Olmo, J., & Builes, N. M. S. (2019). Spatial variability analysis of quality of life and its determinants: a case study of Medellín, Colombia. *Quality of Life Research*.

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

Universidad Industrial de Santander (UIS). (2018). Observatorio de Desarrollo Económico: Análisis de la calidad de vida en Barrancabermeja. Bucaramanga, Colombia. Recuperado de <http://www.uis.edu.co>

Universidad Industrial de Santander (UIS). (2018). Observatorio de Desarrollo Regional: Impacto de la Proximidad a Bucaramanga en el Desarrollo de Santander. Bucaramanga, Colombia. Recuperado de <http://www.uis.edu.co>

Vilalta y Perdomo, Carlos J. (2005). Como enseñar autocorrelación espacial. *\*Economía, Sociedad y Territorio\**, 5(18). El Colegio Mexiquense, México.

- Vilalta y Perdomo, C. J. (2005). Sobre la espacialidad de los procesos electorales urbanos y una comparación entre las técnicas de regresión OLS y SAM. *Estudios Demográficos y Urbanos*, 21(1), 83-122. El Colegio de México.
- Vyas, S., & Kumaranayake, L. (2006). Constructing socio-economic status indices: How to use principal components analysis. *Health Policy and Planning*, 21(6), 459-468.
- Zarzosa, P. (1996). Aproximación de la medición del bienestar social. Universidad de Valladolid.

Apéndices

Apéndice A.

Variables disponibles y su descripción

Nombre de la variable	Notación final	Descripción	Medida	Nombre de la variable	Notación final	Descripción	Medida	Nombre de la variable	Notación final	Descripción	Medida
Vivienda en un área protegida		Proporción de viviendas en un área definida geográficamente que haya sido designada, regulada y administrada a fin de alcanzar objetivos específicos de conservación.	0 a 1	Si acceso a fuente de agua mejorada	efan	Vivienda sin acceso a fuente de agua mejorada	0 a 1	Barreras de acceso a servicios de salud		Proporción de personas que tienen una discapacidad vinculada a los servicios de salud	0 a 1
Casa		Porcentaje de casas	0 a 1	Componente Servicios		Puntuación en el componente de servicios según el DANE	0 a 1	Componente dependencia económica	dependencia económica	Puntuación en el componente de dependencia económica según el DANE	0 a 1
Apartamento		Porcentaje de apartamento	0 a 1	Puntuación de la banda ancha		Porcentaje de hogares o usuarios en una región que tienen acceso a servicios de internet de alta velocidad (banda ancha).	0 a 1	Trabajo informal	informal	Porcentaje de personas en trabajo informal	0 a 1
Bloque, ladrillo, piedra, madera pulida	pladrillo	Viviendas cuyos paredes están hechas de Bloque, ladrillo, piedra, madera pulida	0 a 1	Numero de cuartos	0 a 2	Porcentaje de viviendas cuyo numero de cuartos es de 0 a 2	0 a 1	Estrato 1	estrato1	Porcentaje de viviendas en estrato 1	0 a 1
Concreto vaciado		Viviendas cuyos paredes están hechas de Concreto vaciado	0 a 1	3 a 5		Porcentaje de viviendas cuyo numero de cuartos es de 3 a 5	0 a 1	Estrato 2		Porcentaje de viviendas en estrato 2	0 a 1
Tapa plada, bahareque, adobe		Viviendas cuyos paredes están hechas de Tapa plada, bahareque, adobe	0 a 1	> 6		Porcentaje de viviendas cuyo numero de cuartos es > 6	0 a 1	Estrato 3	estrato3	Porcentaje de viviendas en estrato 3	0 a 1
Madera burda, tablón, tablón	pmadera	Viviendas cuyos paredes están hechas de Madera burda, tablón, tablón	0 a 1	0		Porcentaje de viviendas cuyo numero de hogares es 0	0 a 1	Estrato > 4		Porcentaje de viviendas en estrato > 4	0 a 1
Mierro, parquet, madera pulida y lacada		Viviendas cuyos pisos están hechos de Mierro, parquet, madera pulida y lacada	0 a 1	1		Porcentaje de viviendas cuyo numero de hogares es 1	0 a 1	Desempeño de lenguaje	desempeño	Porcentaje de personas en desempeño durante el mes 12 meses	0 a 1
Baldosa, vinilo, tabeta, linóleo, laminado	balbosa	Viviendas cuyos pisos están hechos de Baldosa, vinilo, tabeta, linóleo, laminado	0 a 1	2		Porcentaje de viviendas cuyo numero de hogares es 2	0 a 1	Personas que asisten a alguna institución educativa		Porcentaje de personas que asisten a alguna institución educativa	0 a 1
Cerámico, grava		Viviendas cuyos pisos están hechos de Cerámico, grava	0 a 1	>3		Porcentaje de viviendas cuyo numero de hogares es >3	0 a 1	Aflabafismo 15 y Nis		Porcentaje de personas con aflabafismo	0 a 1
Madera burda, tabla, tablón, otro vegetal	medera	Viviendas cuyos pisos están hechos de Madera burda, tabla, tablón, otro vegetal	0 a 1	0 a 2		Porcentaje de viviendas cuyo numero de personas es 0 a 2	0 a 1	Componente Inasistencia		Puntuación del componente inasistencia según el DANE	0 a 1
Terra, arena, barro		Viviendas cuyos pisos están hechos de Tierra, arena, barro	0 a 1	3 a 5		Porcentaje de viviendas cuyo numero de personas es 3 a 5	0 a 1	Analfabetismo	analfabetismo	Porcentaje de personas en analfabetismo	0 a 1
Vivienda con acceso a Servicio de energía		Vivienda con acceso a Servicio de energía	0 a 1	>6		Porcentaje de viviendas cuyo numero de personas es >6	0 a 1	Raño libro educativos	hajaedu	Porcentaje de viviendas cuyo promedio de escolaridad de habitantes mayores a 15 años es menor a 9	0 a 1
Servicio de Acueducto		Vivienda con acceso a Servicio de Acueducto	0 a 1	Trabajo infantil	infantil	Porcentaje de menores de edad que realiza algún trabajo	0 a 1	Barreras a servicios para cuidado de la primera infancia		Porcentaje de personas con barreras a servicios para cuidado de la primera infancia	0 a 1
Servicio Alcantarillado	alcantarillado	Vivienda con acceso a Servicio Alcantarillado	0 a 1	Prop de Personas en MIE (%)		Proporción de personas con Necesidades Básicas Insatisfechas (NBI)	0 a 1	Inasistencia escolar		Porcentaje de personas en inasistencia escolar	0 a 1
Gas natural	gas	Vivienda con acceso a Gas natural	0 a 1	Hacinamiento cúbico	hacinamiento o	Proporción de viviendas que presentan más de tres personas por cuarto	0 a 1	Riesgo escolar		Porcentaje de personas en riesgo escolar	0 a 1
Recolección Basuras	basuras	Vivienda con acceso a Recolección Basuras	0 a 1	Si aseguramiento en salud	salud	Proporción de personas sin aseguramiento de salud	0 a 1	Saber leer y escribir		Porcentaje de personas que no saben leer y escribir	0 a 1
Internet	internet	Vivienda con acceso a internet	0 a 1	Prop de Personas en MIE	mieira	Proporción de personas que carecen de dos o más NBI	0 a 1	Puntaje saber 11 Matemáticas	icfem	Puntaje promedio de Matemáticas en el ICES de 11 grado	Mayor a 0
Indoora Alcantarillado		Vivienda con acceso a Indoora Alcantarillado	0 a 1	Componente vivienda		Puntuación en el componente de vivienda según el DANE	0 a 1	Puntaje saber 11 Lenguaje	icel	Puntaje promedio de Lenguaje en el ICES de 11 grado	Mayor a 0
Indoora piso láctico		Vivienda con acceso a Indoora piso láctico	0 a 1	Componente Hacinamiento		Puntuación en el componente de hacinamiento según el DANE	0 a 1	Bases Primaria Completa		Porcentaje de personas cuyo máximo grado de escolaridad es Básica Primaria Completa	0 a 1
Básica secundaria completa		Porcentaje de personas cuyo máximo grado de escolaridad es Básica secundaria completa	0 a 1	Dificultad para Relaciones o interacción con las demás personas		Porcentaje de dificultad para relacionarse o interactuar con las demás personas	0 a 1	Muertes por accidentes de tránsito Niños		Muertes por accidentes de tránsito Niños	Mayor a 0
Medica académica o técnica		Porcentaje de personas cuyo máximo grado de escolaridad es Medica académica o técnica	0 a 1	Personas Con Dificultades	difultad	Puntuación si el personas con dependencia económica según el DANE	0 a 1	Muertes accidentales Hombres		Muertes accidentales Hombres	Mayor a 0
Técnica profesional o Tecnológica		Porcentaje de personas cuyo máximo grado de escolaridad es Técnica profesional o Tecnológica	0 a 1	Cobertura de salud		Porcentaje de cobertura de salud	0 a 1	Muertes accidentales Mujeres		Muertes accidentales Mujeres	Mayor a 0
Universitario	universitario	Porcentaje de personas cuyo máximo grado de escolaridad es Universitario	0 a 1	Tasa de mortalidad infantil		Tasa de mortalidad infantil para el 2018	Mayor a 0	Violencia interpersonal Hombres		Violencia interpersonal Hombres	Mayor a 0
Especialización, Maestría, Doctorado		Porcentaje de personas cuyo máximo grado de escolaridad es Especialización, Maestría, Doctorado	0 a 1	Cobertura en vacunación Pentavalente		Proporción de niños con vacuna para difteria, tétanos, Sarampión, poliomielitis e infecciones producidas por Haemophilus influenzae tipo b	0 a 1	Violencia Interpersonal Mujeres		Violencia Interpersonal Mujeres	Mayor a 0
Tasa de mortalidad		Tasa de mortalidad para el 2018	Mayor a 0	Índice de Femenidad	indicefem	Cantidad de mujeres por cada 100 hombres	Mayor a 0	Presunto delito sexual en hombres		Presunto delito sexual en hombres	Mayor a 0
Hijos nacidos vivos		Proporción de hijos nacidos vivos respecto a total de nacimientos	0 a 1	Razón de dependencia		Cociente de las personas en edades «dependientes» (menores de 15 y mayores de 60 años) respecto a las personas en edades «productivas» (población entre 15 y 59 años de edad).	Mayor a 0	Presunto delito sexual en mujeres		Presunto delito sexual en mujeres	Mayor a 0
Promedio de hijos por mujer		Promedio de hijos por mujer	Mayor a 0	Jovenes 15 a 29 años	jovenes	Porcentaje de jovenes 15 a 29 años	0 a 1	Violencia contra niños, niñas y adolescentes por hombres		Violencia contra niños, niñas y adolescentes por hombres	Mayor a 0
Razón de niños < 5 años / mujeres		Razón de niños < 5 años / mujeres	Mayor a 0	Relación personas > 65 años / 14 años	proporcion	Cociente de personas > a 65 años con < a 14 años	0 a 1	Violencia contra niños, niñas y adolescentes por mujeres		Violencia contra niños, niñas y adolescentes por mujeres	Mayor a 0
Adultos mayores con algún problema de salud en los últimos 30 días		Porcentaje de adultos mayores con algún problema de salud en los últimos 30 días	0 a 1	Peso relativo municipal en el PIB departamental	pi	Peso relativo municipal en el PIB departamental	Mayor a 0	Violencia de Pareja Hombres		Violencia de Pareja Hombres	Mayor a 0
Personas con dificultad de oír		Porcentaje de personas con dificultad de oír	0 a 1	% de inversión financiada con recursos propios		Porcentaje de inversión financiada con recursos propios	0 a 1	Violencia de Pareja Mujeres		Violencia de Pareja Mujeres	Mayor a 0
Personas con dificultad para hablar o conversar		Porcentaje de personas con dificultad para hablar o conversar	0 a 1	Ingresos 7 Rubros Por Cápita Municipal (promedio 4 años)		Promedio de los últimos 4 años de los ingresos Por Cápita Municipal	Mayor a 0	Violencia entre otros familiares Hombres		Violencia entre otros familiares Hombres	Mayor a 0
Discapacidad de origen visual		Porcentaje de personas en discapacidad durante el mes 12 meses	0 a 1	Viviendas/personas	viv/per	Proporción de viviendas en el municipio respecto a la densidad poblacional	0 a 1	Violencia entre otros familiares Mujeres		Violencia entre otros familiares Mujeres	Mayor a 0
Dificultad para: Ver de cerca, de lejos o alrededor		Porcentaje de dificultad para: ver de cerca, de lejos o alrededor	0 a 1	Hombres mujer		Hombres mujer	Mayor a 0	Violencia adulto mayor Hombres		Violencia adulto mayor Hombres	Mayor a 0
Dificultad para: Mover el cuerpo, caminar o subir y bajar escaleras		Porcentaje de dificultad para: mover el cuerpo, caminar o subir y bajar escaleras	0 a 1	Hombres hombres		Hombres hombres	Mayor a 0	Violencia adulto mayor Mujeres		Violencia adulto mayor Mujeres	Mayor a 0
Dificultad para: Agarrar o mover objetos con las manos		Porcentaje de dificultad para: agarrar o mover objetos con las manos	0 a 1	Sukides hombres		Sukides hombres	Mayor a 0	Lesiones por accidentes de transporte Hombres		Lesiones por accidentes de transporte Hombres	Mayor a 0
Dificultad para: Escuchar, aprender, recordar o tomar decisiones por sí mismo		Porcentaje de dificultad para: entender, aprender, recordar o tomar decisiones por sí mismo	0 a 1	Sukides mujeres		Sukides mujeres	Mayor a 0	Lesiones por accidentes de transporte Mujeres		Lesiones por accidentes de transporte Mujeres	Mayor a 0
Dificultad para: Comer, vestirse o bañarse por sí mismo		Porcentaje de dificultad para: comer, vestirse o bañarse por sí mismo	0 a 1	Muertes por accidentes de tránsito Hombres		Muertes por accidentes de tránsito Hombres	Mayor a 0	Tasa de hurtos por 10.000 habitantes	hurtos	Tasa de hurtos por 10.000 habitantes	Mayor a 0
				Tasa de casos de violencia masculista por 10.000 habitantes	vif	Tasa de casos de violencia masculista por 10.000 habitantes	Mayor a 0	Tasa de homicidios por 10.000 habitantes	homicidios	Tasa de homicidios por 10.000 habitantes	Mayor a 0

Apéndice B.

Matriz de correlaciones para las variables seleccionadas

	Pladri	Pladerj	Baldos	madera	alcant	gas	basura	intern	sfam	estrat	estrat	desemp	tinfor	pib	tinfan	miseri	hacina	depend	salud	dif sa	analfa	bajaed	icfetm	icfesl	univer	indice	juvene	vivien	propor	hurto	homicid	vif
Pladrillo	1.000	0.269	0.535	0.381	0.435	0.504	0.493	0.415	0.394	0.420	0.473	0.346	0.361	0.298	0.275	0.387	0.282	0.453	-0.387	0.332	0.408	0.517	0.067	0.220	0.498	0.134	0.337	-0.106	-0.260	-0.426	-0.047	-0.208
Pladera	0.269	1.000	0.392	0.692	0.111	0.072	0.167	0.221	0.028	0.339	0.229	0.111	0.065	0.044	0.163	0.554	-0.110	0.290	0.193	-0.254	0.310	0.211	0.388	0.385	0.324	-0.043	-0.127	-0.210	0.303	-0.056	0.292	-0.195
Baldosa	0.535	0.392	1.000	0.327	0.763	0.609	0.791	0.814	0.542	0.688	0.794	0.471	0.556	0.404	0.387	0.580	0.228	0.663	-0.347	0.099	0.734	0.855	0.454	0.578	0.860	0.360	0.230	-0.194	-0.051	-0.570	0.020	-0.322
madera	0.381	0.692	0.327	1.000	0.240	0.171	0.286	0.206	0.117	0.196	0.204	0.075	0.276	0.107	0.198	0.151	-0.273	0.223	-0.113	-0.107	0.332	0.261	0.163	0.245	0.241	-0.015	0.163	-0.299	-0.063	-0.194	0.140	-0.193
alcantarillado	0.435	0.111	0.763	0.240	1.000	0.717	0.950	0.778	0.615	0.465	0.659	0.433	0.669	0.477	0.380	0.251	-0.021	0.668	-0.341	0.118	0.654	0.850	0.177	0.295	0.707	0.447	0.387	-0.221	-0.251	-0.657	-0.159	-0.373
gas	0.504	0.072	0.609	0.171	0.717	1.000	0.727	0.736	0.483	0.343	0.605	0.354	0.521	0.546	0.351	0.200	0.042	0.580	-0.573	0.255	0.451	0.722	0.039	0.162	0.641	0.396	0.421	-0.154	-0.159	-0.635	-0.221	-0.304
basuras	0.493	0.167	0.791	0.286	0.950	0.727	1.000	0.810	0.606	0.466	0.652	0.470	0.747	0.498	0.372	0.289	-0.034	0.649	-0.554	0.171	0.671	0.881	0.212	0.378	0.718	0.450	0.441	-0.290	-0.309	-0.695	-0.120	-0.386
Internet	0.415	0.221	0.814	0.206	0.778	0.736	0.810	1.000	0.500	0.555	0.757	0.477	0.651	0.661	0.309	0.368	0.064	0.623	-0.578	0.201	0.676	0.897	0.327	0.495	0.874	0.468	0.406	-0.313	-0.185	-0.704	-0.097	-0.349
sfam	0.394	0.028	0.542	0.117	0.615	0.606	0.606	0.500	1.000	0.458	0.484	0.343	0.497	0.313	0.282	0.367	0.164	0.410	-0.292	0.114	0.336	0.532	0.187	0.329	0.492	0.342	0.186	-0.042	-0.141	-0.400	-0.178	-0.231
estrat01	0.420	0.339	0.688	0.196	0.465	0.343	0.466	0.555	0.458	1.000	0.656	0.337	0.291	0.266	0.205	0.477	0.307	0.479	-0.077	-0.114	0.515	0.535	0.423	0.502	0.608	0.258	0.021	0.001	0.002	-0.270	0.029	-0.314
estrat03	0.473	0.229	0.794	0.204	0.659	0.605	0.652	0.757	0.484	0.656	1.000	0.450	0.510	0.391	0.294	0.387	0.194	0.547	-0.442	0.205	0.592	0.753	0.311	0.426	0.758	0.365	0.274	-0.174	-0.138	-0.553	-0.105	-0.261
desempleo	0.346	0.111	0.471	0.075	0.433	0.354	0.470	0.473	0.343	0.237	0.450	1.000	0.386	0.206	-0.185	0.411	0.216	0.666	-0.346	0.297	0.366	0.448	0.258	0.329	0.433	0.132	0.214	-0.173	-0.139	-0.292	0.026	-0.082
tinfor	0.361	0.065	0.556	0.276	0.669	0.521	0.747	0.651	0.497	0.291	0.510	0.396	1.000	0.420	0.411	0.211	0.022	0.462	-0.581	0.231	0.574	0.748	0.178	0.353	0.522	0.217	0.408	-0.288	-0.315	-0.652	-0.183	-0.311
pib	0.298	0.044	0.404	0.107	0.477	0.546	0.498	0.661	0.313	0.266	0.391	0.206	0.420	1.000	0.211	0.136	-0.092	0.329	-0.583	0.236	0.404	0.583	0.120	0.240	0.572	0.304	0.319	-0.185	-0.180	-0.724	-0.166	-0.274
tinfantil	0.275	0.163	0.387	0.198	0.380	0.351	0.372	0.309	0.282	0.205	0.294	-0.185	0.411	0.211	1.000	0.223	0.077	0.213	-0.079	-0.153	0.298	0.461	0.041	0.195	0.376	0.265	-0.131	0.117	0.231	-0.343	-0.011	-0.278
miseria	0.387	0.554	0.580	0.151	0.253	0.200	0.289	0.368	0.367	0.477	0.387	0.411	0.211	0.136	0.223	1.000	0.490	0.602	0.050	0.017	0.438	0.363	0.445	0.517	0.474	0.051	-0.201	0.071	0.299	-0.151	0.071	-0.143
hacinamiento	0.282	-0.110	0.228	-0.273	-0.021	0.042	-0.034	0.064	0.164	0.307	0.194	0.256	-0.022	-0.092	0.077	0.490	1.000	0.205	0.005	0.171	0.071	0.037	0.140	0.174	0.115	0.042	-0.275	0.541	0.266	0.069	0.054	-0.041
dependencia económica	0.453	0.290	0.665	0.223	0.668	0.380	0.649	0.623	0.410	0.470	0.547	0.666	0.463	0.329	0.213	0.602	0.205	1.000	-0.300	0.110	0.666	0.687	0.263	0.392	0.630	0.397	0.112	-0.073	0.013	0.453	-0.044	-0.360
salud	-0.387	0.193	-0.347	-0.113	-0.543	-0.577	-0.554	-0.576	-0.292	-0.077	-0.442	-0.346	-0.581	-0.583	-0.079	0.050	0.005	-0.300	1.000	-0.525	-0.391	-0.586	0.075	-0.108	-0.426	-0.261	-0.645	0.316	0.507	0.676	0.173	0.238
dif salud	0.332	-0.254	0.099	-0.107	0.118	0.253	0.171	0.201	0.114	-0.114	0.205	0.297	0.231	0.236	-0.153	0.017	0.171	0.120	-0.525	1.000	0.152	0.169	-0.109	-0.044	0.103	0.004	0.400	-0.150	-0.440	-0.281	-0.292	0.012
analfabetismo	0.409	0.310	0.734	0.332	0.654	0.450	0.671	0.676	0.336	0.515	0.592	0.366	0.574	0.404	0.298	0.438	0.071	0.666	-0.391	0.152	1.000	0.768	0.399	0.456	0.682	0.155	0.301	-0.345	-0.271	-0.553	-0.029	-0.347
bajaedu	0.517	0.211	0.855	0.163	0.850	0.722	0.881	0.897	0.532	0.535	0.753	0.448	0.748	0.583	0.461	0.363	0.037	0.687	-0.586	0.169	0.768	1.000	0.309	0.467	0.895	0.444	0.455	-0.299	-0.219	-0.727	-0.084	-0.351
icfetm	0.067	0.388	0.454	0.163	0.177	0.039	0.212	0.327	0.187	0.423	0.311	0.258	0.178	0.120	0.041	0.445	0.140	0.263	0.075	-0.109	0.399	0.309	1.000	0.872	0.435	0.198	-0.172	-0.032	0.159	-0.123	0.123	-0.111
icfesl	0.220	0.385	0.576	0.245	0.295	0.162	0.378	0.495	0.329	0.502	0.426	0.329	0.353	0.240	0.195	0.517	0.174	0.392	-0.108	-0.044	0.458	0.467	0.872	1.000	0.546	0.298	-0.040	-0.083	0.055	-0.259	0.185	-0.205
universitario	0.498	0.324	0.860	0.241	0.707	0.641	0.718	0.874	0.492	0.608	0.758	0.435	0.542	0.572	0.376	0.474	0.115	0.630	-0.479	0.103	0.682	0.895	0.435	0.546	1.000	0.386	0.306	-0.234	-0.041	0.640	-0.024	-0.169
Indicfem	0.134	-0.043	0.360	-0.015	0.457	0.396	0.450	0.468	0.342	0.258	0.365	0.132	0.217	0.304	0.265	0.051	0.042	0.397	-0.261	0.004	0.155	0.444	0.198	0.298	0.396	1.000	-0.020	0.254	0.184	-0.313	-0.089	-0.186
juvenes 15 a 29 años	0.337	-0.127	0.230	0.163	0.387	0.421	0.441	0.406	0.186	0.021	0.274	0.214	0.408	0.319	-0.201	0.275	0.112	-0.645	0.400	0.301	0.455	-0.172	-0.040	0.306	0.020	1.000	-0.633	-0.758	-0.418	-0.129	-0.032	
vivienda/personas	-0.106	-0.210	-0.194	-0.299	-0.221	-0.154	-0.290	-0.313	-0.042	0.001	-0.174	-0.173	-0.288	-0.185	0.117	0.071	0.541	-0.073	0.316	-0.150	-0.345	-0.299	-0.032	-0.083	-0.234	0.254	-0.633	1.000	0.610	0.264	0.039	-0.001
proporcion	-0.260	0.303	-0.051	-0.063	-0.251	-0.159	-0.309	-0.185	-0.141	0.002	-0.139	-0.139	-0.315	-0.180	0.231	0.299	0.266	0.013	0.503	-0.440	-0.271	-0.219	0.159	0.055	-0.041	0.184	-0.758	0.610	1.000	0.270	0.148	0.103
hurto	-0.426	-0.056	-0.370	-0.194	-0.657	-0.635	-0.695	-0.704	-0.400	-0.270	-0.553	-0.292	-0.652	-0.724	-0.343	-0.151	0.069	-0.453	0.676	-0.281	-0.553	0.727	-0.123	-0.259	-0.640	-0.313	-0.418	0.264	0.270	1.000	0.168	0.433
homicidios	-0.047	0.292	0.020	0.140	-0.159	-0.221	-0.120	-0.097	-0.178	0.029	-0.105	0.026	-0.183	-0.166	-0.011	0.071	0.054	-0.044	0.173	-0.292	-0.029	-0.084	0.123	0.185	-0.034	-0.089	-0.129	0.039	0.148	0.168	1.000	-0.036
vif	-0.208	-0.185	-0.322	-0.193	-0.373	-0.394	-0.386	-0.349	-0.231	-0.314	-0.261	-0.082	-0.311	-0.374	-0.278	-0.143	-0.041	-0.269	0.239	0.012	-0.347	-0.351	-0.111	-0.205	-0.269	-0.186	-0.092	-0.001	0.103	0.453	-0.026	1.000

Nota: Las casillas en rojo representan las correlaciones bajas (entre -0.4 y 0.4), las amarillas representan las correlaciones moderadas (entre (-0.6, -0.4) y (0.4 y 0.6)) y las verdes representan las correlaciones fuertes mayores/menores a ±0.8.

**Apéndice C.**

*Puntaje de las variables para las primeras dos componentes*

	<b>PC1</b>	<b>PC2</b>
pladrillo	0.17227901	0.008653245
pmadera	0.08122893	-0.267311465
baldosa	0.25433099	-0.126975985
madera	0.09515798	-0.059243317
alcantarillado	0.24576661	0.061047095
gas	0.21640123	0.099840333
basuras	0.25661700	0.065977025
internet	0.26062530	0.019466002
sfam	0.17650905	-0.022452049
estrato1	0.17668729	-0.204172557
estrato3	0.22942373	-0.042581836
desempleo	0.15478961	-0.022182213
tinformal	0.21090446	0.103872001
pib	0.17559134	0.117440904
tinfantil	0.11205348	-0.111509522
miseria	0.13432283	-0.307907918
hacinamiento	0.03181006	-0.210585847
dependencia económica	0.21030046	-0.110830132
salud	-0.17263766	-0.287297024
difsalud	0.06762628	0.242502596
analfabetismo	0.22245515	-0.030269029
bajaedu	0.27094274	0.024775706
icfesm	0.10529623	-0.273225776
icfsl	0.15402917	-0.242422539
universitario	0.25130796	-0.074254942
indicefem	0.12331078	-0.064508335
jovenes	0.11954584	0.359886110
viv/per	-0.08295727	-0.236691419
65/14	-0.06848948	-0.369580977
hurtos	-0.21742257	-0.141599610
homicidios	-0.02996307	-0.169388385
vif	-0.11989217	0.024784779

**Apéndice D.***Código en R-Studio*

*El código utilizado y el Geoportal desarrollado en este proyecto está disponible en GitHub. A continuación se proporciona el enlace al repositorio que contiene el código fuente y los datos relevantes.*

*Lucia-Gamez. (2024). ICV [Repositorio en GitHub]. Disponible en <https://github.com/Lucia-Gamez/ICV>*