

**ESCUELA POLITÉCNICA NACIONAL**

**FACULTAD DE CIENCIAS**

**CONSTRUCCIÓN DE UN INDICADOR DE DESIGUALDAD  
SOCIAL CON DEPENDENCIA ESPACIAL MEDIANTE EL  
ANÁLISIS DE COMPONENTES PRINCIPALES NO LINEAL,  
UTILIZANDO LA ENCUESTA DE CONDICIONES DE VIDA  
(2014)**

**TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE  
INGENIERA MATEMÁTICA**

**PROYECTO DE INVESTIGACIÓN**

**JARED ABIGAIL VALENCIA SALVADOR**

jared.valencia@epn.edu.ec

**DIRECTOR: DR. MIGUEL ALFONSO FLORES SÁNCHEZ**

miguel.flores@epn.edu.ec

QUITO, ABRIL 2021

## DECLARACIÓN

Yo **JARED ABIGAIL VALENCIA SALVADOR**, declaro bajo juramento que el trabajo aquí escrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y que he consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración cedo mis derechos de propiedad intelectual, correspondientes a este trabajo, a la Escuela Politécnica Nacional, según lo establecido por la Ley de Propiedad Intelectual, por su reglamento y por la normatividad institucional vigente.

---

**Jared Abigail Valencia Salvador**

## CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por **JARED ABIGAIL VALENCIA SALVADOR**, bajo mi supervisión.

---

**Dr. Miguel Alfonso Flores Sánchez**  
Director del Proyecto

## **AGRADECIMIENTOS**

A Dios, por sus bondades y su amor inagotable,  
a José y Rocío, mis padres, por enseñarme el camino que debía seguir,  
a Diego y Arianna, mis hermanos, por su apoyo incondicional,  
a Nicolás, mi mejor amigo y cómplice, por su amor y dedicación,  
a Alex y Lucía, mi familia, por su corazón bondadoso, a Miguel, mi tutor, por su  
paciencia y dirección para la elaboración de este trabajo.

## **DEDICATORIA**

*A José Valencia, mi padre.  
Mientras más me acerco a la meta más te extraño.*

# Índice general

<b>Resumen</b>	<b>x</b>
<b>Abstract</b>	<b>xI</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Antecedentes . . . . .	1
1.2. Justificación . . . . .	2
1.3. Objetivos . . . . .	3
1.3.1. Objetivo General . . . . .	3
1.3.2. Objetivos Específicos . . . . .	3
<b>2. Marco Teórico</b>	<b>4</b>
2.1. Preliminares . . . . .	4
2.1.1. Análisis de Componentes Principales (PCA) . . . . .	4
2.1.2. Aprendizaje Estadístico . . . . .	5
2.1.3. Decaimiento de distancias . . . . .	5
2.1.4. Espacios de Hilbert . . . . .	7
2.1.5. Métodos kernel . . . . .	8
2.2. Análisis de Componentes Principales No Lineal . . . . .	10
2.3. Análisis de Componentes Principales No Lineal para datos espaciales	12
2.3.1. Análisis de Componentes Principales No Lineal con depen-	
dencia geoespacial basado en un kernel (KNPCA) . . . . .	13
2.4. Sistemas de Indicadores Sociales (SIS) . . . . .	15
2.4.1. Construcción de un Sistema de Indicadores Sociales . . . . .	17

2.5.	Desigualdad Social . . . . .	19
2.5.1.	Indicadores de desigualdad . . . . .	21
2.5.2.	Desigualdad territorial . . . . .	23
2.6.	Construcción del indicador . . . . .	25
<b>3.</b>	<b>Aplicación a la Encuesta de Condiciones de Vida (2014)</b>	<b>26</b>
3.1.	Análisis de datos . . . . .	26
3.1.1.	Descripción de la base de datos . . . . .	26
3.2.	Categorización de datos . . . . .	29
3.3.	Aplicación del NPCA basado en un kernel . . . . .	33
3.3.1.	Matriz de pesos espaciales . . . . .	33
3.3.2.	Caso sin ponderación geográfica . . . . .	34
3.3.3.	Caso con ponderación geográfica . . . . .	37
3.4.	Análisis de resultados . . . . .	41
<b>4.</b>	<b>Conclusiones y Recomendaciones</b>	<b>50</b>
4.1.	Conclusiones . . . . .	50
4.2.	Recomendaciones . . . . .	52
<b>Anexos</b>		<b>53</b>
A.	Pasos en la construcción de la <i>curva de Lorenz</i> . . . . .	53
B.	Centrado de un espacio de gran dimensión . . . . .	55
C.	El problema de valores propios en el espacio de coeficientes de expansión . . . . .	57
D.	Variables categorizadas . . . . .	59
D.1.	Variables originales . . . . .	59
D.2.	Variables creadas . . . . .	60
<b>E.</b>	<b>Implementación de los algoritmos en Rstudio</b>	<b>62</b>
E.1.	Algoritmo árbol de mínima expansión . . . . .	62
E.2.	KPCA - Caso sin ponderación geográfica . . . . .	63
E.3.	KPCA - Caso con ponderación geográfica . . . . .	64



# Índice de cuadros

3.1. Variables seleccionadas . . . . .	30
3.2. Variables utilizadas para la construcción del indicador . . . . .	33
3.3. Resumen del indicador . . . . .	41
3.4. Quintiles del indicador de desigualdad social . . . . .	42
3.5. Grupos del indicador de desigualdad social . . . . .	43
3.6. Sectores por grupo . . . . .	44
3.7. Indicador por parroquia, DMQ . . . . .	46
3.8. Sectores censales en la Parroquia . . . . .	49
A.1. Cuadro de datos . . . . .	54
A.2. Frecuencias . . . . .	54
D.1. La variable original se reemplazó por la unión de <i>De donde se obtiene el agua y Tratamiento que se da al agua.</i> . . . . .	61
D.2. La variable original se reemplazó por la condición de si en ese hogar, se cubre la canasta básica o canasta vital per cápita. . . . .	61

# Resumen

La desigualdad social se produce cuando una persona recibe un trato diferente como consecuencia de su posición social, su situación económica, su género, la religión que profesa, entre otros aspectos. Todos estos factores se entrecruzan en los territorios, por lo cual varios estudios de carácter social sugieren que el lugar donde una persona nace o vive influye en su desarrollo, en el acceso y distribución de oportunidades de bienestar, dando lugar a las desigualdades territoriales.

Por esta razón, el presente trabajo tiene como objetivo analizar las variables de la Encuesta de Condiciones de Vida 2014, Sexta Ronda, para la construcción de un indicador de desigualdad social que considere la influencia espacial y geográfica de los datos para analizar la desigualdad que existe en las parroquias del Distrito Metropolitano de Quito.

**Palabras clave:** Desigualdad social, desigualdad territorial, distancias de decaimiento, métodos kernel, Análisis de Componentes Principales No Lineal basado en un kernel, Distrito Metropolitano de Quito.

# Abstract

Social inequality occurs when a group of people receive different treatment as a consequence of their social position, economic situation, gender, the religion they profess and other aspects. All of the factors mentioned before intersect in territories, that is why a lot of social studies suggested that the place where a person is born or lives influences on their development, on their access and distribution of well-being opportunities, resulting in territorial inequalities.

For that reason, the present investigation aims to analyze the variables of the Survey of Living Conditions 2014, Sixth Round, for the construction of an indicator of social inequality that considers the spatial and geographical influence of the data in order to analyze inequality in Quito Metropolitan District.

**Keywords:** Social inequality, territorial inequality, distance decay, kernel methods, Nonlinear Principal Component Analysis based on a kernel, Quito Metropolitan District.

# Capítulo 1

## Introducción

### 1.1. Antecedentes

En América Latina, la matriz de desigualdad social está fuertemente condicionada por su matriz productiva, por lo cual el primer y más básico determinante de la desigualdad es el estrato socio económico.[11] Sin embargo, las desigualdades de género, desigualdades étnicas y raciales, desigualdades relacionadas con el ciclo de vida y desigualdades territoriales constituyen ejes estructurantes de la matriz de desigualdad social, ya que son determinantes en la disposición de los patrones de desigualdad social y, la magnitud y reproducción de las brechas identificadas en algunos de los principales ámbitos del desarrollo social y del ejercicio de los derechos. [11]

Según la Comisión Económica para América Latina y el Caribe (CEPAL), una de las dimensiones más significativas de los altos niveles de desigualdad en América Latina y el Caribe es la brecha en los niveles de desarrollo de distintas localidades dentro de cada país; por lo cual se ha planteado que en esta región del mundo, “el lugar importa”, es decir no da lo mismo nacer o vivir en cualquier territorio, puesto que este influye en el acceso y distribución de oportunidades de bienestar, dando lugar a las desigualdades territoriales. [10]

Los distintos ejes de la desigualdad social se manifiestan y materializan en las regiones donde viven las personas y se entrecruzan con la heterogeneidad territorial, sumando nuevos factores de rezago, discriminación y exclusión. Por ello, es importante considerar el análisis del territorio como uno de los ejes de las desigualdades sociales que permite identificar nuevas dimensiones del entrecruzamiento de desigualdades y plantea una mirada a lo que ocurre en las localidades y entre ellas.

Por lo tanto, un indicador de desigualdad social que considere el espacio y la geografía, sería un instrumento útil para aproximarnos al análisis de los factores, causantes de las distintas desigualdades, que no pueden ser observados de forma directa, lo que nos permitiría identificar cómo influyen en el territorio analizado y proponer soluciones para las problemáticas encontradas.

Se haría uso de determinados datos locales que sirvan de base para la selección de las zonas donde se deberían plantear propuestas de intervención con base en la orientación del análisis y los problemas detectados en las localidades seleccionadas.

La propuesta metodológica que se desarrollará a continuación, tiene la capacidad de adaptarse a las características propias de cada territorio, ya que viene determinada por el lugar de intervención y por información local que se pueden adaptar al mismo.

## **1.2. Justificación**

El análisis de componentes principales (PCA) es una técnica poderosa para extraer una estructura latente de conjuntos de datos posiblemente de alta dimensión. [18] PCA es una transformación ortogonal del sistema de coordenadas en el que describimos nuestros datos. Los nuevos valores de coordenadas mediante los cuales representamos los datos se denominan componentes principales. A menudo, ocurre que un pequeño número de componentes principales es suficiente para explicar la mayor parte de la estructura de los datos. Sin embargo el PCA tiene dos limitaciones importantes, la primera es que asume que la relación entre variables es lineal, y segundo, asume que todas las variables son de escala numérica. [54] En las ciencias sociales y de comportamiento estas suposiciones, en general, no son justificables, por lo cual este método no es el más apropiado.

Para resolver esos limitantes, se desarrolló el Análisis de Componentes No Lineales (NPCA); una primera versión de este método fue descrita por Guttman [23]. Este método tiene los mismos objetivos del método original pero es adecuado para variables de niveles de medición mixtos (nominal, ordinal y numérico), que pueden estar o no relacionadas linealmente entre ellas. En NPCA, a los niveles categóricos de las variables se asignan cuantificaciones categóricas de tal manera que se tenga en cuenta la mayor parte posible de la varianza en las variables cuantificadas; estas cuantificaciones se obtienen de un proceso llamado cuantificación o escalamiento óptimo.

El Análisis de Componentes Principales considerando un kernel es una extensión de los métodos de PCA no lineales usando núcleos definidos positivos. El objetivo es encontrar variables proyectadas en un kernel inducido, con máxima varianza, del espacio de características. Los métodos kernel se basan en el teorema de Mercer del análisis funcional, que implica que si  $k$  es un kernel continuo de un operador integral positivo, entonces existe una función en un espacio donde  $k$  es un producto escalar. El hecho de necesitar solo productos escalares permite reproducir cualquier algoritmo lineal en un espacio de Hilbert [1].

Por otro lado, las ciencias sociales son ciencias observacionales, no experimentales, por lo cual es importante registrar el lugar y tiempo de los eventos. [3] Estas ciencias estudian procesos en diferentes tipos de lugares y espacios, ya que la estructura de lugares y espacios puede influir en el desarrollo de procesos económicos y sociales, generando así datos con dependencia espacial.

Si se utilizan datos espaciales en modelos clásicos se provocará un gran impacto en la calidad de las estimaciones, que resultarán sesgadas e ineficientes. [14] Por lo cual se debe utilizar otras técnicas del análisis multivariado que serán desarrolladas más adelante.

## **1.3. Objetivos**

### **1.3.1. Objetivo General**

El objetivo general que se plantea es "Diseñar y construir un indicador de desigualdad social, considerando las variables de la Encuesta de Condiciones de Vida realizada en el Ecuador, en el año 2014, que permita identificar grupos en condición de vulnerabilidad".

### **1.3.2. Objetivos Específicos**

1. Construir pesos que representen la dependencia espacial que existe entre cada localidad
2. Construir pesos que representen la dependencia que existe entre cada localidad en función de su ubicación geográfica.
3. Identificar los grupos vulnerables y su ubicación geográfica.

# Capítulo 2

## Marco Teórico

En este capítulo se presentan conceptos importantes de los sistemas de indicadores sociales, aprendizaje estadístico, análisis de componentes principales y econometría. Además, se detalla una breve reseña de la desigualdad y los indicadores comúnmente usados para medir la misma.

Posteriormente, presentaremos la formula teórica del método de Análisis de Componentes Principales No Lineal desarrollada a partir del PCA estandarizado, formulado exclusivamente con productos escalares. También, se tomará en cuenta los métodos kernel y conceptos de econometría espacial, mismos que permiten el desarrollo teórico del método de Análisis de Componentes Principales No Lineal basado en un kernel.

### 2.1. Preliminares

#### 2.1.1. Análisis de Componentes Principales (PCA)

La idea central del Análisis de Componentes Principales es reducir la dimensión de un conjunto de datos de un gran número de variables interrelacionadas, mientras se mantiene la variabilidad presente en la base de datos. Para esto se realiza una transformación ortogonal del sistema de coordenadas en el que describimos los datos a un nuevo conjunto de variables, las componentes principales (PCs), mismas que son no correlacionadas y están ordenadas tal que las primeras conservan la mayor parte de la información presente en todas las variables originales.

## 2.1.2. Aprendizaje Estadístico

- **Espacio de características.** El término espacio de características es usado con frecuencia en la literatura del Aprendizaje Estadístico porque una de sus tareas es la extracción de características. Este espacio es el conjunto de variables relacionadas con algunas propiedades del fenómeno observado. Supongamos que se tiene un ejemplo del fenómeno como un vector de características,  $y_1, \dots, y_n$ . Cada característica o variable puede ser considerada como una "dimensión" del problema y la observación sería un "punto" en el espacio de características de  $n$ -dimensiones.

Es posible representar un espacio de características utilizando solo números, y además, hacerlo de tal manera que el espacio de características se pueda interpretar como un espacio real.

- **Entorno general del aprendizaje estadístico.** El entorno del aprendizaje estadístico a partir de ejemplos tiene tres componentes [61]:
  1. *Un generador de datos de entrada.* Asumiremos que los datos pueden ser representados como vectores de  $\mathbb{R}^N$ , mismos que son independientes e idénticamente distribuidos según una distribución de probabilidad  $p(x)$  desconocida.
  2. *Un supervisor que, dados los datos de entrada, retorne un valor de salida, acorde a una distribución condicional  $p(y|x)$  también fija y desconocida.* Es importante considerar que el supervisor puede estar presente o no.
  3. *Un algoritmo de aprendizaje, capaz de escoger una hipótesis*

$$f(x; \theta)$$

donde la hipótesis  $f$  es una función de  $x$  y depende de un vector de parámetros  $\theta$  que pertenece a un conjunto  $\Theta$ . Así, la hipótesis que corresponde al espacio es,

$$S = \{f(x; \Theta) : \theta \in \Theta\}$$

que es uno a uno con el espacio de parámetros  $\Theta$ .

## 2.1.3. Decaimiento de distancias

Para comprender mejor la definición de *decaimiento de distancias* primero debemos definir los términos: distancia y decaimiento.

La **distancia** significa el mapa físico, y de manera más realista, la distancia de la superficie topográfica. Pero este término como tal, debe definirse con una nueva perspectiva, donde se debe considerar el objeto de estudio y los resultados que se desean obtener.

El término **decaimiento** se refiere a la influencia decreciente de un fenómeno, atributo o actividad cuando dos lugares están lejos uno del otro.

Así, el término decaimiento de distancia ha sido usado para describir el efecto de la *distancia* en las interacciones entre dos localidades separadas, y está relacionado con la primera ley de geografía de Waldo R. Tobler:

Todas las cosas están relacionadas, pero las cosas cercanas están más relacionadas que las lejanas.

Esta distancia es una regla importante en el análisis espacial, especialmente para modelos de interacción espacial y las nociones de difusión cultural; además, hace referencia a la influencia decreciente de un fenómeno, atributo o actividad cuando dos lugares están lejos uno del otro. Por ejemplo, habría menos interacción entre lugares distantes porque la accesibilidad y el valor de la tierra tienden a disminuir desde el centro.

Siempre ha sido controversial si este concepto es de aplicación universal y si la relación se cuantifica correctamente, ya que con la llegada de las tecnologías de transporte y comunicación, el efecto de fricción de la distancia se ha reducido en gran medida. Por lo que la distancia física o topográfica ya no es un factor determinante principal del grado de interactividad entre lugares y personas.

Las interacciones entre lugares y personas siguen dependiendo en gran medida de la conectividad disponible y la eficiencia en el transporte de personas y mercancías o en la transmisión de información. Como tal, el término "distancia" debe definirse con una nueva perspectiva, misma que depende del objetivo del análisis.

Empíricamente, se han identificado algunas variables clave que se pueden usar en lugar de la distancia física:

- Distancia de tiempo
- Distancia de costos
- Conveniencia en términos de alternativas
- Preferencia de comportamiento.

Muchos estudios han trabajado en identificar y analizar patrones de movilidad para ver el alcance de la distancia u otros factores que gobiernan el comportamiento humano, con el fin de calcular o predecir el número de interacciones entre dos lugares separados espacialmente. El impacto de la distancia en una actividad particular puede ser lineal o no lineal y se puede modelar con una función de decaimiento de la distancia.

### 2.1.4. Espacios de Hilbert

Un espacio de Hilbert es un espacio vectorial real completo con producto interno  $\langle \cdot, \cdot \rangle$ , con las siguientes propiedades

1. Simetría:  $\langle f, g \rangle = \langle g, f \rangle$
2. Linealidad:  $\langle \alpha f_1 + \beta f_2, g \rangle = \alpha \langle f_1, g \rangle + \beta \langle f_2, g \rangle$ , para todo número real  $\alpha$  y  $\beta$
3. No negatividad:  $\langle f, g \rangle \geq 0$
4. Cero:  $\langle f, g \rangle = 0 \Rightarrow f = 0$

Dado esto, definamos el concepto de operador.

Un operador lineal  $\varphi$  asigna una función  $f$  en un espacio de Hilbert a otra función  $g$  en el mismo espacio de Hilbert o en otro, es decir,

$$\varphi f = g$$

Este operador tiene la siguiente propiedad

$$\varphi(\alpha f + \beta g) = \alpha \varphi f + \beta \varphi g.$$

Intuitivamente, se puede pensar en las funciones como vectores y los operadores como matrices; ya que, en términos algebraicos lineales, una matriz normalmente proyecta un conjunto de vectores a otro conjunto de vectores.

Por tanto, el efecto del operador es transformar una función en un espacio de Hilbert a otra función en el mismo u otro espacio de Hilbert.

Ahora, una función tal que

$$k : \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R} \tag{2.1}$$

es llamada un kernel en  $\mathbb{R}^N$ , si existe un espacio del Hilbert  $F$  con producto interno  $\langle \cdot, \cdot \rangle$  y una función

$$\Phi : \mathbb{R}^N \rightarrow F, \quad x \mapsto X. \quad (2.2)$$

tal que para todo  $x, y \in \mathbb{R}^N$  se tiene

$$k(x, y) = \langle \Phi(x), \Phi(y) \rangle \quad (2.3)$$

Una función kernel puede ser interpretada como un tipo de medida entre las matrices  $x$  y  $y$ .

### Kernels definidos positivos

Sea  $\chi$  un conjunto no vacío. Una función simétrica  $k : \chi \times \chi \rightarrow \mathbb{R}$  es un *kernel definido positivo* si para todo  $N \in \mathbb{N}$  y para cualquier tupla  $(x_1, x_2, \dots, x_N) \in \chi$ , la matriz Gram  $K$  definida por las entradas  $K_{nm} = k(x_n, x_m)$ , satisface<sup>1</sup>

$$\alpha^T K \alpha = \sum_{n=1}^N \sum_{m=1}^N K_{nm} \alpha_n \alpha_m \geq 0, \quad \text{para todo } \alpha \in \mathbb{R}^N. \quad (2.4)$$

En particular, supongamos que  $F$  es algún espacio de Hilbert con producto interno  $\langle \cdot, \cdot \rangle$ . Entonces para cualquier función  $\Phi : \chi \rightarrow F$  tenemos

$$\sum_{n=1}^N \sum_{m=1}^N \langle \Phi(x_n), \Phi(x_m) \rangle \alpha_n \alpha_m = \sum_{n=1}^N \sum_{m=1}^N \langle \alpha_n \Phi(x_n), \alpha_m \Phi(x_m) \rangle = \left\| \sum_{n=1}^N \alpha_n \Phi(x_n) \right\|^2 \geq 0 \quad (2.5)$$

De la primera expresión podemos ver que, en el sentido de lo explicado anteriormente,

$$k : (x, y) \mapsto \langle \Phi(x), \Phi(y) \rangle$$

es un kernel definido positivo. Un kernel  $k$  continuo y definido positivo se conoce como **kernel de Mercer** [32].

### 2.1.5. Métodos kernel

Los métodos kernel son una clase especial de los algoritmos de aprendizaje. Su idea principal consiste en mapear los puntos de entrada, generalmente representados por elementos de  $\mathbb{R}^N$ , dentro de un espacio  $F$  de alta dimensión con producto interno, llamado **espacio de características**. El mapeo se realiza mediante la función

<sup>1</sup>Note que, por definición, un kernel definido positivo satisface  $k(x, x) \geq 0$  para cualquier  $x \in \chi$

de características  $\Phi$ , definida como en 2.2,

$$\Phi : \mathbb{R}^N \rightarrow F, \quad x \mapsto X.$$

Luego, se aborda la tarea de aprendizaje encontrando un modelo lineal en el espacio de características de acuerdo con los puntos de entrenamiento  $\Phi(x_1), \dots, \Phi(x_N) \in F$ . Ahora, dado que la función de características normalmente se elige para que sea no lineal, un modelo lineal en el espacio de características corresponde a una regla no lineal en  $\mathbb{R}^N$ .

Si el algoritmo se puede expresar únicamente en términos de productos internos, se puede plantear el problema en función de las imágenes de una *función del kernel* (2.1),

$$k : \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R}, \quad (x, y) \mapsto k(x, y). \quad (2.6)$$

con,

$$k(x, y) = \Phi(x)^T \Phi(y) \quad (2.7)$$

Esto, denominado como el **truco del kernel**, es de gran interés para los casos en los que el espacio de características es de dimensión infinita.

En la práctica, es común comenzar con el diseño de un kernel definido positivo, que garantiza la existencia de una función de características que satisfaga 2.6.

Notemos que, en la ecuación 2.7 notamos  $\langle \Phi(x), \Phi(y) \rangle$  como  $\Phi(x)^T \Phi(y)$ , implícitamente asumiendo que el espacio de características  $F$  es un espacio Ecludideo de dimensión finita. Sin embargo es posible demostrar que el espacio de características asociado a un kernel definido positivo es un espacio de Hilbert de dimensión infinita.

## 2.2. Análisis de Componentes Principales No Lineal

Dado un conjunto de datos centrados  $x_k, k = 1, \dots, M, x_k \in \mathbb{R}^N, \sum_{k=1}^M x_k = 0$ , el PCA diagonaliza la matriz de covarianza,

$$C = \frac{1}{M} \sum_{j=1}^M \langle x_j, x_j \rangle \quad (2.8)$$

para lo cual se debe resolver la ecuación de valores propios,

$$Cv = \lambda v, \quad (2.9)$$

con los valores propios  $\lambda \geq 0$  y  $v \in \mathbb{R}^N \setminus \{0\}$ . Como,

$$\frac{1}{M} \sum_{j=1}^M x_j \langle x_j, v \rangle = \lambda v \quad (2.10)$$

todas las soluciones  $v$  con  $\lambda \neq 0$  deben estar en el espacio generado por  $x_1, \dots, x_M$ ; por lo tanto la ecuación 2.9 es equivalente a

$$\langle x_k, Cv \rangle = \lambda \langle x_k, v \rangle, \quad \text{para todo } k = 1, \dots, M \quad (2.11)$$

Ahora, describiremos el mismo cálculo considerando otro espacio con producto interno  $F$ , mismo que está relacionado con el espacio de entrada mediante una función posiblemente no lineal,

$$\Phi : \mathbb{R}^N \rightarrow F, \quad x \mapsto X. \quad (2.12)$$

Notemos que  $F$  puede tener una dimensionalidad arbitrariamente grande, posiblemente infinita. Aquí y en lo siguiente, los caracteres en mayúscula serán usados para elementos de  $F$ , y los caracteres en minúscula denotarán elementos de  $\mathbb{R}^N$ .

De manera análoga a lo anterior, asumimos que se trabaja con datos centrados, es decir  $\sum_{k=1}^M \Psi(x_k) = 0^2$  Usando la matriz de covarianza en  $F$ ,

$$\bar{C} = \frac{1}{M} \sum_{j=1}^M \langle \Psi(x_j), \Psi(x_j) \rangle \quad (2.13)$$

tenemos que encontrar los valores propios  $\lambda \geq 0$  y los vectores propios  $V \in F \setminus \{0\}$  que satisfacen,

$$\bar{C}V = \lambda V. \quad (2.14)$$

---

<sup>2</sup>Por simplicidad, asumimos que las observaciones están centradas. Conseguir esto en el espacio de entrada es fácil, pero es complicado en  $F$  porque no se puede calcular la media de  $\Psi(x_j)$  en  $F$  de forma explícita. Sin embargo hay una manera de hacerlo, misma que consta en el anexo B.

donde:

$$\langle V, V \rangle = 1, \Psi(x_i) = \Phi(x_i) - \frac{1}{M} \sum_{i=1}^M \Phi(x_i) \quad i = 1, \dots, M \quad (2.15)$$

Por lo cual, todas las soluciones  $V$  con  $\lambda \neq 0$  están en el espacio generado por  $\Psi(x_1), \dots, \Psi(x_M)$ . De esto se tiene dos consecuencias útiles.

Primero, podemos considerar el conjunto de ecuaciones,

$$\langle \Psi(x_k), \bar{C}V \rangle = \lambda \langle \Psi(x_k), V \rangle, \quad \text{para todo } k = 1, \dots, M, \quad (2.16)$$

y, segundo, existen coeficientes  $\alpha_i$ , con  $i = 1, \dots, M$  tal que,

$$V = \sum_{i=1}^M \alpha_i \Psi(x_i). \quad (2.17)$$

Combinando las ecuaciones 2.16 y 2.17, tenemos

$$\frac{1}{M} \sum_{i=1}^M \alpha_i \langle \Psi(x_k), \sum_{j=1}^M \Psi(x_j) \rangle \langle \Psi(x_j), \Psi(x_i) \rangle = \lambda \sum_{i=1}^M \alpha_i \langle \Psi(x_k), \Psi(x_i) \rangle, \quad \text{para todo } k = 1, \dots, M. \quad (2.18)$$

Ahora, definamos una matriz  $K_{M \times M}$ ,

$$K_{ij} := k(x_i, x_j) = \langle \Psi(x_i), \Psi(x_j) \rangle, \quad i, j = 1, \dots, M \quad (2.19)$$

y consideremos la matriz de centrado

$$H = I_M - \frac{1}{M} \mathbf{1}_M \mathbf{1}_M^T, \quad \mathbf{1}_M \in \mathbb{R}^{M \times 1} \quad (2.20)$$

tal que  $\tilde{K} = HKH$ . Así, reemplazando lo anterior en la ecuación 2.18, se tiene

$$\tilde{K}^2 \alpha = M \lambda \tilde{K} \alpha, \quad (2.21)$$

donde  $\alpha$  es un vector con entradas  $\alpha_1, \dots, \alpha_M$ .

Luego, para encontrar la solución de la ecuación 2.21, resolvemos el siguiente problema de valores propios,

$$\tilde{K} \alpha = M \lambda \alpha, \quad (2.22)$$

para valores propios distintos de cero <sup>3</sup>.

Sean  $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_M$  valores propios de  $\tilde{K}$ , es decir las soluciones  $M\lambda$  de la ecuación 2.22, y  $\alpha^1, \dots, \alpha^M$  el conjunto completo correspondiente de vectores propios, con  $\lambda_p$  como el primer valor propio distinto de cero (asumiendo  $\Psi \neq 0$ ).

<sup>3</sup>La justificación de este procedimiento se encuentra en el anexo C

Normalizamos  $\alpha^1, \dots, \alpha^M$  de tal manera que los vectores correspondientes en el espacio  $F$  estén normalizados, esto es,

$$\langle V^k, V^k \rangle = 1 \quad \text{para todo } k = p, \dots, M \quad (2.23)$$

por lo tanto, considerando las ecuaciones 2.17 y 2.22, lo anterior se tiene como una condición de normalización para  $\alpha^1, \dots, \alpha^M$ :

$$1 = \sum_{i,j=1}^M \alpha_i^k \alpha_j^k \langle \Psi(x_i), \Psi(x_j) \rangle = \sum_{i,j=1}^M \alpha_i^k \alpha_j^k \tilde{K}_{ij} = \langle \alpha^k, \tilde{K} \alpha^k \rangle = M \lambda_k \langle \alpha^k, \alpha^k \rangle. \quad (2.24)$$

entonces, los vectores  $\alpha^1, \dots, \alpha^M$  deben ser normalizados para que tengan tamaño  $\frac{1}{\sqrt{M\lambda}}$ . Notemos que, de la expresión 2.24 se sigue que  $\tilde{K}$  es un kernel definido positivo, lo que garantiza la existencia de  $\Psi$ .

Luego, para extraer los componentes principales necesitamos calcular las proyecciones en los vectores propios  $V^k$  en  $F$  con  $k = p, \dots, M$ .

Sea  $x$  un punto de prueba, con imagen  $\Psi(x)$  en  $F$ ; entonces

$$\langle V^k, \Psi(x) \rangle = \sum_{i=1}^M \alpha_i^k \langle \Psi(x_i), \Psi(x) \rangle \quad (2.25)$$

pueden llamarse las componentes principales no lineales correspondientes a  $\Psi$ .

### 2.3. Análisis de Componentes Principales No Lineal para datos espaciales

Dos propiedades que caracterizan a los datos espaciales son la heterogeneidad y la autocorrelación espacial.

La heterogeneidad espacial se refiere a la no estacionariedad de los procesos geográficos, es decir, los procesos pueden variar localmente y no necesariamente son los mismos en cada ubicación espacial. La autocorrelación espacial es la tendencia de los atributos a estar relacionados en algún lugar del espacio.[13]

La presencia de estas propiedades invalida dos supuestos de algunos análisis estadísticos estándar, *los datos son independientes e idénticamente distribuidos*. Por lo cual, el uso de metodologías estándar, incluido el PCA, plantea desafíos particulares, ya que la calidad de cualquier estimación resultará sesgada e ineficiente.

A continuación, describiremos dos técnicas de Análisis de Componentes Princi-

pales (PCA) que consideran el efecto espacial en datos espaciales.

- **PCA ponderado localmente (LWPCA).** Esta técnica es aplicada cuando el conjunto universal de componentes principales no representa bien el conjunto de datos originales. A continuación, describiremos dos técnicas de LWPCA.

1. Consideremos que hay regiones localizadas en el espacio de atributos donde un conjunto de componentes principales proporciona una mejor descripción. Esta técnica utiliza una ponderación de ventana móvil en el espacio de datos donde se encuentran las componentes principales en la localidad de algún punto  $x$  que pertenece al espacio de datos. Para cada LWPCA alrededor de  $x$ , los puntos de datos vecinos son ponderados primero según alguna función kernel de decaimiento de distancia<sup>4</sup>. Luego, cada observación es multiplicada por su respectivo peso y se aplica el algoritmo del PCA estándar a los datos ponderados.
2. Supongamos que se tiene un conjunto de datos  $X$ , para los que se tiene la matriz de covarianzas  $C = X^T X$ . Ahora, sea  $W$  una matriz de pesos calculada mediante una función de ponderación para cada una de las observaciones de  $X$ ; así, la matriz ponderada de covarianzas sería  $C_w = X^T W X$ . Entonces, para tener una versión del Análisis de Componentes Principales ponderada localmente basta extraer los valores y vectores propios de la matriz  $C_w$ .

- **PCA ponderado geográficamente (GWPCA).** Esta técnica es similar a los modelos LWPCA, pero en este caso se supone que hay regiones del espacio geográfico en los que se debe considerar distintos modelos de PCA, ya que los datos de estudio pueden ser espacialmente heterogéneos y deben ser modelados como tal. Además, las distancias se basan en distancias geográficas, usualmente la distancia euclidiana.

### 2.3.1. Análisis de Componentes Principales No Lineal con dependencia geoespacial basado en un kernel (KNPCA)

En esta sección, presentaremos dos casos [33]:

1. Caso sin pesos geográficos

---

<sup>4</sup>Por ejemplo: bi-cuadrado, gaussiano, entre otros.

## 2. Caso con pesos geográficos

Para continuar, es importante considerar lo desarrollado en la sección 2.2 y la segunda técnica descrita para el PCA ponderado localmente.

### Caso sin pesos geográficos

Sea  $x_1, \dots, x_n$  un conjunto de datos tal que  $x_i \in \mathbb{R}^{t \times p}$  con  $i = 1, \dots, n$ , donde  $p$  es el número de variables,  $t$  el número de observaciones y  $n$  el número de regiones. Se define la función,

$$\Phi : \mathbb{R}^{t \times p} \rightarrow H$$

donde  $H$  es un espacio de Hilbert con producto interno  $\langle \cdot, \cdot \rangle$ . Supongamos que primero mapeamos los datos de forma no lineal en  $H$ ; es decir, se tiene  $\Phi(x_1), \dots, \Phi(x_n)$ . Notemos que  $H$  puede ser de gran dimensión, posiblemente infinita.

Para aplicar el análisis de componentes principales en el espacio de características  $H$ , procedemos de manera análoga a lo detallado en la sección 2.2. A continuación, con los supuestos considerados en esta sección, presentamos esquemáticamente el algoritmo de PCA basado en una función kernel de decaimiento de distancia.

Algoritmo sistemático del PCA
<ol style="list-style-type: none"> <li>1. Dado un conjunto de datos <math>x_1, \dots, x_n</math>, con <math>x_i \in \mathbb{R}^{t \times p}</math>, <math>i = 1, \dots, n</math>.</li> <li>2. Calcular la matriz kernel <math>K = (K_{ij})</math>, <math>K_{ij} = k(x_i, x_j)</math></li> <li>3. Calcular la matriz del kernel centrado <math>\tilde{K} = HKH</math>, <math>H = I_n - \frac{1}{n}1_n1_n^T</math></li> <li>4. Calcular los vectores propios <math>\alpha_i</math> y los valores propios <math>\lambda_i</math> de <math>\tilde{K}</math>, <math>i = 1, \dots, n</math></li> <li>5. Escoger el <math>i</math>-ésimo vector <math>\alpha_i</math> sobre el cual se va a proyectar</li> <li>6. Normalizar <math>\alpha_i</math> para que tenga longitud <math>\frac{1}{\sqrt{n\lambda_i}}</math></li> <li>7. Luego, para la muestra <math>x_i</math>, <math>i = 1, \dots, n</math>, el valor de la proyección de <math>\Phi_i \in H</math> sobre el <math>j</math>-ésimo vector <math>u_j</math> del operador de covarianza de la muestra en <math>H</math> es <math>\langle \Psi(x_i), u_j \rangle = \sum_{k=1}^n \alpha_k^j \langle \Psi(x_i), \Psi(x_k) \rangle = \sum_{k=1}^n \tilde{K}_{ik} \alpha_k^j</math></li> </ol>



En el algoritmo no se ha definido algún kernel en específico, por lo que utilizaremos el kernel Gaussiano<sup>5</sup> que es una función kernel de decaimiento de distancia,

$$k(x, x') = \exp(-\gamma \|x - x'\|_F^2), \quad \gamma > 0$$

<sup>5</sup>Es claro que el kernel Gaussiano es definido positivo, por lo cual se cumple la condición que garantiza la existencia de  $\Psi$ .

donde:

$$\|A\|_F = \sqrt{\text{tr}(A^T A)}$$

es la norma de Frobenius. La constante  $\gamma > 0$  se selecciona adecuadamente para los datos  $x_1, \dots, x_n$ , con  $x_i \in \mathbb{R}^{t \times p}$  para  $i = 1, \dots, n$ .

En este trabajo se tomará el valor de  $\gamma$  como el recíproco de la media aritmética de los elementos de la matriz triangular inferior que tiene su  $(i, j)$ -ésimo elemento dado por  $\|x_i - x_j\|_F^2$  con  $i, j = 1, \dots, n$ .

Es importante notar que los datos están dispuestos para aplicar las técnicas descritas al inicio de esta sección, además están agrupados por localidades. Como se menciona en el LWPCA, los valores y vectores propios se extraen de la matriz de covarianzas  $C$ , misma que está definida por una función de decaimiento de distancia.

### **Caso con pesos geográficos**

En este caso, la matriz de observaciones  $x_i$  es remplazada por la matriz  $w_i X_i$ , donde  $w_i$  con  $i = 1, \dots, n$  es el peso geográfico positivo asociado con el  $i$ -ésimo sitio de observación.

Dado que los datos pueden depender del lugar geográfico de observación, para eliminar, al menos parcialmente, esta relación, se asignan pesos geográficos a estos lugares. El resto de la teoría desarrollada en las secciones 2.2 y 2.3.1 permanece sin cambios.

## **2.4. Sistemas de Indicadores Sociales (SIS)**

El crecimiento económico se ha considerado, históricamente, como sinónimo de bienestar. Se asumía que entre ambos existía una relación causal directa, y que promoviendo un desarrollo económico estábamos promoviendo mejoras en las condiciones de vida, sin necesidad de consideraciones adicionales. Sin embargo, a lo largo del tiempo esta condición se ha revelado como necesaria pero no suficiente.[3]

Por lo cual se empieza a dar plantear herramientas que permitan medir la incidencia de ciertos problemas específicos a cada área de las condiciones de vida, como la educación, el mercado de trabajo y las oportunidades de empleo, la composición de los hogares y sus ingresos, la distribución del bienestar, la salud y la desigualdad de género, dando lugar a la construcción de indicadores sociales.

Los Sistemas de Indicadores Sociales son todo intento sistemático e integrado para conceptualizar, operacionalizar y medir, por medio de un conjunto de indicadores sociales, la diversidad de aspectos que conforman un concepto (pluridimensional) de bienestar.[4]

De forma consensual, se considera como el origen de los indicadores sociales, dentro del campo de las ciencias sociales, al intento de la Agencia Espacial Americana de detectar los efectos colaterales del programa espacial en la sociedad americana, mismo que fue realizado a mediados de los años setenta. Los participantes en el estudio confirmaron que el programa tendría consecuencias sociales importantes, pero al mismo tiempo les era imposible obtener datos que permitieran realizar un análisis cuantitativo detallado. Además, se encontraron con una ausencia total de marco teórico y metodología para tratar estos temas.

Raymond A. Bauer en su obra, *Social Indicators*, aborda esta experiencia y trata sobre las condiciones para construir un sistema de información adecuado, más allá de los que muestran los indicadores económicos, además de la necesidad de acordar indicadores sociales que reflejen el mayor número de aspectos de la sociedad. Por lo cual esto constituye un pilar básico y ha dado nombre al movimiento de los indicadores sociales.

El perfeccionamiento alcanzado a lo largo del tiempo en las estadísticas administrativas y herramientas que proporcionan los datos básicos para la elaboración de indicadores, la identificación de campos de preocupación social generalmente estudiados, el acuerdo de sistemas de indicadores sectoriales y el desarrollo de programas de armonización internacional en la estadística social, son algunas de las consecuencias positivas que se atribuyen a este movimiento.[4]

Más allá del objetivo general de mejorar la base de información de las sociedades, pueden distinguirse dos enfoques fundamentales en los Sistemas de Indicadores Sociales. De un lado, encontramos el enfoque no orientado de políticas, sino dirigido a la información social del usuario en general sin importar si es investigador o no.

Por otro lado, tenemos enfoques que plantean que los indicadores e informes sociales deberían permitir el análisis de las políticas sociales, detectando objetivos para la acción, proporcionando información de apoyo al análisis y al seguimiento, y así, evaluar el impacto de esas políticas.

Los elementos centrales de la definición de Sistema de Indicadores Sociales, que permite caracterizarlos como tales son [3]:

1. El intento de conceptualización y medida tiene como finalidad la medida del

bienestar, individual o social.

2. El bienestar se considera como un término multidimensional, compuesto por partes o áreas diversas

Ahora, es importante considerar que la construcción de cualquier sistema requiere de la existencia de un marco teórico referencial coherente, que hace posible la estructuración de los distintos campos o dominios sociales.[3]

### **2.4.1. Construcción de un Sistema de Indicadores Sociales**

Para construir un Sistema de Indicadores Sociales es necesario establecer el entorno del sistema y el enfoque operativo que va a orientado. Por ejemplo, no es lo mismo un SIS para evaluar una política concreta en un barrio de una ciudad, que un SIS diseñado para dar una visión global de un país o cuya finalidad es realizar comparaciones internacionales, o atender disparidades regionales dentro de ese país.[9]

Luego, se debe tener en cuenta el objetivo del sistema, que puede ser proporcionar un conocimiento cuantitativo de las condiciones de vida y procesos sociales, medir los rasgos principales de la sociedad en cuestión, su interrelación y cambio. Por tanto el SIS se concentraría en la descripción de las condiciones de vida de la población objetivo y también puede prestar atención a la forma como estas condiciones son percibidas por las personas.

El concepto de bienestar o calidad de vida, caracterizado por los campos, dimensiones o dominios contemplados o el grado de atención al análisis de distribución, es decir, las discrepancias entre grupos sociales y entre distintas unidades territoriales son también aspectos principales que determinan la orientación del sistema. A todo ello hay que añadir la disponibilidad de datos que permitan calcular los indicadores juzgados como idóneos por el sistema según su enfoque específico.

Así, de lo descrito sobre los sistemas de indicadores se concluye que no existe un modelo de referencia único que pueda atender a los distintos tipos de necesidades.

### **Método deductivo y Método inductivo**

Entre las alternativas metodológicas para la construcción de un SIS, haremos referencia al método deductivo y al método inductivo. [44]

El enfoque deductivo parte de la construcción teórica y, tras el diseño de los indicadores necesarios, se plantea la forma de poder obtener los indicadores seleccionados. Un enfoque inductivo, por el contrario, presta una atención a la disponibilidad de datos para adaptar los indicadores a la información disponible.[3][44]

En un estado puro, un enfoque inductivo implica un aprovechamiento exclusivo de las fuentes de información existentes, sin contemplar dimensiones teóricas que no puedan ser cubiertas por el sistema de información de referencia. Este método, en el peor de los casos, por su propia dinámica de elaboración, implica que el sistema de indicadores renuncia a una de sus funciones principales como es la detección de campos o subcampos de preocupación social que presentan lagunas de información.[3]

El enfoque deductivo puro se basa en un modelo teórico previo que no toma en cuenta en ningún momento la disponibilidad de datos existentes. Este enfoque basado en construcciones ideales, deriva fácilmente en la implantación de una encuesta social general que cubra completamente los requerimientos del sistema.[3]

Este método, además de cubrir las lagunas de información detectadas entre las fuentes de información existentes, prefiere obtener la información disponible en el sistema de forma que sea posible realizar análisis en el nivel del microdato.[3][44]

Además, la opción deductiva depende fuertemente de la base teórica de partida, por lo que resulta rígido a la hora de incorporar desarrollos metodológicos ulteriores, que impliquen un cambio sustancial de la orientación teórica de partida. Así pues, una condición indispensable es tener una aportación teórica importante junto con la posibilidad real de implantar las encuestas necesarias. [4]

Por otro lado, los enfoques deductivos, a través de sus construcciones teóricas proporcionan líneas de desarrollo para las estadísticas y su implantación de manera coherente y comparable.[3][44]

En cualquier caso, el razonamiento entre ambos planteamientos implica, de una parte, que el modelo teórico no puede construirse de forma totalmente ajena a la disponibilidad de las fuentes de información y, por otro lado, éstas deben prestar atención a los indicadores que paulatinamente se van determinando como importantes en la tarea de caracterización del bienestar y calidad de vida.

## 2.5. Desigualdad Social

La desigualdad es una medida relativa ya que no proporciona información sobre la situación de los individuos en términos absolutos, sino que, es un *concepto racional*. Es decir, es importante analizar la distribución completa de la desigualdad dentro de una sociedad, no solo en cómo la desigualdad afecta los resultados individuales, sino también en cómo afecta las relaciones de las personas entre sí y con las instituciones sociales. [38]

La desigualdad social aborda diferentes áreas que pretenden mostrar el nivel de diferenciación entre individuos, generalmente para ver su grado de pertenencia a una clase social determinada. Estas áreas abarcan aspectos relacionados con el acceso a educación, salud, trabajo, libertad de expresión, entre otros.[59]

Por tanto las características centrales de las desigualdades sociales que debemos considerar son [59]:

1. Las desigualdades son parte de un proceso de cambio que se propaga a través del tiempo (dimensión temporal).
2. Involucran un grado de pertenencia local y nacional que se fortalece en el ámbito regional e internacional como resultado de la creciente interdependencia entre las diversas regiones del mundo. Esto hace que, si bien, las desigualdades tienen un sentido de pertenencia en lo nacional, existe un grado de jerarquización de las desigualdades, que va de lo internacional a lo local.
3. Las desigualdades contienen una dimensión transnacional que afecta al grado de percepción sobre la igualdad de derechos y oportunidades entre los individuos del planeta, principalmente, en lo referente a ciudadanía y justicia social, que puede alimentar el conflicto social.
4. Tienen implicaciones sobre los cimientos del desarrollo humano y la sostenibilidad del planeta.

La desigualdad también puede definirse sobre la base de sus tres extensiones: vertical, horizontal y estructural.

Según Sergio Tezanos, la **desigualdad individual o vertical** es la desigualdad entre personas independientemente del lugar donde vivan o del grupo social al que pertenezcan y se da entre personas en función del acceso a recursos tangibles e intangibles. En cambio, la **desigualdad horizontal** es el producto de la desigualdad

entre grupos de personas o familias por motivos de raza, sexo, orientación sexual, lugar de nacimiento u otros factores. Estas dos extensiones suelen reforzarse mutuamente y pueden crear y perpetuar ciclos de pobreza a través de las generaciones.

Por otro lado, la **desigualdad estructural** resulta de la gestión y distribución del poder y de los legados históricos. Estos resaltan aspectos alojados en la creación de instituciones, organizaciones, grupos sociales, etc., los cuales alimentan la exclusión y la marginalización de grupos e individuos, así como dinámicas de dualismo económico, que no favorecen el cierre de la brecha de oportunidades y la justicia social a lo largo del tiempo. [59]

La discusión sobre la desigualdad por lo regular se focaliza en la desigualdad de resultados desde el punto de vista de los ingresos porque generalmente es más fácil de observar. La distribución de los ingresos puede medirse por la forma en que estos se reparten entre los factores de producción básicos y el rendimiento que generan. Sin embargo, el ingreso como medida de la desigualdad, es imperfecta. [59]

Los expertos han determinado que [38]:

"los parámetros de medición de los ingresos se refieren a lo que la gente tiene, en vez de lo que puede hacer o ser, es decir, lo que Amartya Sen denomina sus capacidades o aptitudes para alcanzar metas razonadas [...] El enfoque de estas capacidades se basa en que las capacidades son ideas de libertad en el sentido positivo: qué oportunidades reales tiene uno con respecto a la vida que pueda llevar"

En el Informe sobre Desarrollo Humano 2011, se concluye que "la desigualdad en los resultados [...] es en gran medida consecuencia del acceso dispar a capacidades" [50]. Por lo tanto, lo ideal sería tener sociedades que proporcionen igualdad de capacidades y oportunidades, pero no necesariamente igualdad de resultados, ya que la gente tiene preferencias y valores diferentes.

Desde los inicios de los años 2000, América Latina ha vivido un período de transformaciones políticas, económicas, sociales y, culturales profundas que permitieron por ejemplo, que en la actualidad la democracia sea el sistema de gobierno que predomina en los países de la región; además, se reflejaron en aspectos como la música, en las aspiraciones de las personas y su búsqueda de nuevas perspectivas.[52] Sin embargo, la discriminación y la pobreza constituyen el desafío principal de las democracias en el siglo XXI, por lo que entender que este es el siglo de la inclusión social y de la erradicación de todo tipo de discriminación es un gran reto que tiene como fin que todas las personas del continente ejerzan sus derechos en igualdad.[63]

La desigualdad deteriora las bases de sustentación de la democracia y pone en peligro las importantes conquistas sociales de un conjunto de políticas de inclusión cuyos beneficios disminuyen y desestabilizan mientras no se identifican y transforman las causas que originan y mantienen la injusticia y la exclusión social.[47]

Así pues, descubrir las dinámicas históricas, económicas, políticas, sociales, culturales y territoriales de la desigualdad es, además de un gran desafío analítico, una urgente necesidad política.

### 2.5.1. Indicadores de desigualdad

Existen diferentes medidas de desigualdad. Tres de las más utilizadas dentro del análisis de la desigualdad son [4]:

- Curva de Lorenz
- Coeficiente de Gini
- Índice de Theil

asimismo, otros dos indicadores han cobrado importancia:

- Índice de Desarrollo Humano ajustado por desigualdad (IDH-D)
- Índice de Oportunidad Humana (IOH)

#### Curva de Lorenz, coeficiente de Gini e índice de Theil

La **curva de Lorenz**<sup>6</sup> muestra en términos gráficos la relación existente entre los porcentajes acumulados de población y los porcentajes acumulados de ingreso, a partir de dos condiciones [4]:

1. Ordenar la relación entre ambas variables de forma creciente respecto del valor de su ingreso hasta cubrir el total de la población.
2. Considerar una misma amplitud o tamaño de cada intervalo.<sup>7</sup>

La representación gráfica que se tiene de esta desigualdad relativa muestra la inclinación de la curva (curva de Lorenz) respecto de una línea diagonal de equidad

---

<sup>6</sup>En el anexo A se presenta el método para construir la curva de Lorenz y un ejemplo de aplicación.

<sup>7</sup>Normalmente de 10 %: denominados deciles.

perfecta. Entre más se aproxime la curva a esta línea se observará más igualdad en el país o grupo de sociedades analizadas, mientras que a mayor distancia habrá más desigualdad. Dos aspectos adicionales contiene la curva de Lorenz [41]:

1. El área de concentración, que se ubica entre la línea de equidad perfecta y la curva de Lorenz. Su valor es finalmente lo que mide el coeficiente de Gini.
2. Información sobre el grado de progresividad o regresividad de la distribución del ingreso en una sociedad.

El **coeficiente de Gini** se computa a partir de la curva de Lorenz. Este coeficiente mide el grado de desigualdad en la distribución del ingreso entre individuos en un determinado periodo. La perfecta igualdad está representada por el 0 y la inequidad perfecta por el 100. Tradicionalmente se utilizan dos variables para medir la desigualdad: el ingreso per capita y el ingreso total del hogar. [37]

El **índice de Theil** se basa en medidas de entropía, lo que significa que es una medida de desorden de un sistema, y en nuestro caso de la desviación respecto de la perfecta igualdad del ingreso. Este índice expresa la comparación del ingreso relativo de todos los individuos ponderados por la población o por el ingreso. [41]

### **Índice de Desarrollo Humano ajustado por desigualdad (IDH-D)**

El **Índice de Desarrollo Humano ajustado por Desigualdad (IDH-D)**, elaborado por el PNUD<sup>8</sup>, es un indicador del nivel de desarrollo humano de las personas de una sociedad que tiene en cuenta el grado de desigualdad. Este indicador consta de tres componentes [50]:

1. El índice de esperanza de vida al nacer ajustado por la desigualdad en la distribución de la esperanza de vida basada en las tablas de mortalidad.
2. El índice de educación ajustado por la desigualdad en la distribución de los años de escolaridad basado en datos de encuestas de hogares.
3. El índice de ingresos ajustado por la desigualdad en la distribución de ingresos basado en datos de encuestas de hogares.

Se considera que el IDH-D es un indicador de desarrollo humano “potencial”, ya que intenta mostrar el nivel máximo de desarrollo humano que se podría alcanzar si no hubiese desigualdades.[59]

---

<sup>8</sup>Programa de las Naciones Unidas para el Desarrollo

## Índice de Oportunidad Humana

El **Índice de Oportunidad Humana (IOH)**, elaborado por el Banco Mundial, mide el acceso a servicios básicos y su distribución bajo el principio de igualdad. Tiene como referencia un contexto de cobertura universal, por lo que el índice toma valores entre 0 y 100. Si una sociedad tiene una cobertura de servicios básicos muy baja se acercará al 0, mientras que si tiene cobertura universal tendrá un valor de 100. El IOH tiene tres características [40]:

1. *Efecto escala*: el valor del IOH se modifica por cambios en la tasa de cobertura de los servicios.
2. *Efecto composición*: el IOH cambia cuando se modifican las circunstancias de vida de las personas.
3. *Efecto distribución*: el IOH crece o decrece porque los servicios se distribuyen de forma menos desigual.

### 2.5.2. Desigualdad territorial

La desigualdad territorial es un concepto que vincula a las sociedades con los espacios que ocupan, revelando la complejidad de relaciones sociales abusivas y las asimetrías de poder, orientándose especialmente en los escenarios de aquellos pobladores cuyos derechos de acceso al suelo y a viviendas dignas han sido suprimidos y que circunscriben otras manifestaciones de desigualdades espaciales.[5]

Las mencionadas asimetrías de poder se manifiestan, entre otras formas, en fenómenos como la segregación residencial socioeconómica. Al referirnos a la Segregación Residencial Socioeconómica (SRS) describimos aquella que vincula a actores sociales definidos por su poder económico o su nivel socio económico en la que el acceso al suelo urbano se determina por el mercado inmobiliario y los mecanismos de renta del suelo. Por tanto, es la renta la que determina los precios de suelo y establece quienes, en función de su capacidad de pago, tienen las posibilidades de acceder a las diferentes localizaciones en la ciudad.[5]

Es preciso considerar que el precio del suelo no está determinando mayormente por las características del suelo mismo, sino por su localización en la ciudad; lo que significa que una determinada ubicación le ofrece más o menos ventajas. Entre otros elementos que influyen en el precio son: la accesibilidad que está influida por el costo del transporte, las expectativas existentes en cuanto al uso y precio que tendrá

el suelo en el futuro, entre otros. Además del aspecto socio económico, es en los territorios donde se hacen presentes otras desigualdades.

La presencia de desigualdades territoriales es una realidad que se manifiesta en países desarrollados y en desarrollo. Es el resultado de un proceso de organización territorial de distribución de las actividades, cuya dinámica no es fácil de controlar, y siempre acarrea importantes problemas de desigualdad social que pueden tener repercusiones sociales relevantes.[45]

## 2.6. Construcción del indicador

La construcción del indicador se realizará en las siguientes etapas [33]:

1. Seleccionar y crear las variables de acuerdo al área de análisis para la definición del índice.
2. Aplicar el criterio de escalado óptimo a las variables.[22]
3. Construir la matriz de pesos con influencia espacial.
4. Identificar los pesos geográficos para cada sector.
5. Aplicar el método de Análisis de Componentes Principales no Lineal basado en un kernel.
6. Obtener, por medio de lo realizado en la etapa anterior, las componentes principales necesarias para recoger al menos el 70 % de variabilidad total de los datos.
7. Definir el indicador, para cada sector, mediante la suma de sus componentes principales. Es decir, si  $U_{n \times m}$  es la matriz de componentes principales del paso anterior, con  $n$  igual al número de localidades a analizar y  $m$  el número de componentes principales que recogen al menos el 70 % de variabilidad total de los datos. El indicador de desigualdad  $d_i$  con  $i = 1, \dots, n$  estará definido como:

$$d_i = \sum_{j=1}^m U_{i,j}$$

8. Expresar el indicador en una nueva escala para interpretar mejor los resultados.[39]  
El indicador escalado  $d_i^T$  con  $i = 1, \dots, n$  estará dado por:

$$d_i^T = 1 - \frac{d_i - \min(d_i)}{\max(d_i) - \min(d_i)}$$

donde, si el valor de  $d_i^T$  es cercano a 0 significará que existe poca desigualdad y si es cercano a 1 que existe mucha desigualdad.

## Capítulo 3

# Aplicación a la Encuesta de Condiciones de Vida (2014)

En este capítulo se presenta un análisis descriptivo de las variables de la Encuesta de Condiciones de Vida realizada en el período noviembre 2013–octubre 2014, cuyos resultados definitivos se presentaron en abril de 2015 [30], con el objetivo de analizar, crear, determinar y categorizar las variables que serán incluidas en el análisis y la construcción del indicador de desigualdad. El objetivo es obtener ese indicador dentro del Distrito Metropolitano de Quito (DMQ) para cada sector censal considerado en la encuesta y deducir los resultados para las parroquias del DMQ.

Con las variables seleccionadas se construirá la matriz de pesos basada en una función kernel considerando las que aportan información relevante a nuestra investigación. Dada esta matriz podemos aplicar el Análisis de Componentes Principales No Lineal y extraer los componentes principales con las que obtendremos el indicador.

### 3.1. Análisis de datos

#### 3.1.1. Descripción de la base de datos

En la década de los noventa, el Gobierno del Ecuador inicia el estudio y desarrollo de investigaciones y análisis interrelacionados sobre la situación de los niveles de vida de la sociedad ecuatoriana, con el fin de identificar, comprender y evaluar el efecto de las decisiones económicas familiares y el efecto de las políticas sociales y económicas en los hogares. Así, el país decide implementar en 1994 la Primera

Ronda de la Encuesta de Condiciones de Vida.[30]

Los períodos de recolección de información referidos a las distintas Rondas de esta encuesta, ejecutadas en el Ecuador, han sido los siguientes:

Primera Ronda	junio - octubre 1994
Segunda Ronda	agosto - noviembre 1995
Tercera Ronda	febrero - mayo 1998
Cuarta Ronda	octubre 1998 - septiembre 1999
Quinta Ronda	noviembre 2005 - octubre 2006
Sexta Ronda	noviembre 2013 - octubre 2014

Las Encuestas de Condiciones de Vida (ECV) proporcionan un instrumento estadístico de primer orden para el estudio de la pobreza y desigualdad, el seguimiento de la cohesión social en el territorio de su ámbito, el estudio de las necesidades de la población y el impacto de las políticas sociales y económicas sobre los hogares y las personas. Además, permite obtener indicadores sobre los niveles de vida y bienestar de la población relacionando varios factores como educación, salud, pobreza, inequidad, hábitos, prácticas y uso del tiempo de los hogares, bienestar psicosocial, percepción del nivel de vida, capital social, seguridad ciudadana y retorno migratorio. Dado que la ECV es una encuesta multipropósito de cobertura nacional, busca determinar, no solo cuántos pobres hay y dónde viven sino por qué lo son, para este fin se busca recoger información específica sobre las principales variables mencionadas anteriormente.[30]

Los objetivos de esta encuesta, en términos generales, son los siguientes [31]:

- Medir el impacto que las medidas de ajuste macroeconómico y las de compensación social que han tenido sobre el nivel de vida de los hogares y de la población ecuatoriana.
- Entregar medidas monetarias y no monetarias completas de la distribución del bienestar y del nivel de pobreza de los hogares.
- Facilitar la formulación de políticas y el diseño de acciones destinadas a reducir los niveles de pobreza de la población en general.

En este trabajo utilizaremos los datos de la Encuesta de Condiciones de Vida Sexta Ronda, realizada en los años 2014 y 2015.

El universo bajo estudio de la ECV VI Ronda constituyen los hogares del área urbana y rural de la República del Ecuador. Esta encuesta tiene una cobertura a ni-

vel nacional, regional, provincial y cuatro ciudades auto representadas. La división política del Ecuador abarca las 24 provincias en que está dividido el país. La ECV, recogió información en las 24 provincias del país continental e insular, y las ciudades de Quito, Guayaquil, Cuenca y Machala.

El diseño muestral de la ECV se caracteriza por ser probabilístico, en consecuencia, los resultados obtenidos de la encuesta pueden generalizarse a toda la población. El diseño, a su vez, es estratificado y proporcional al tamaño de la población, lo que permite mejorar la eficiencia del diseño muestral; y también es bietápico, donde la unidad última de selección es la vivienda y la unidad de observación es el hogar.[31]

Para la selección de la muestra se definieron las unidades primarias de muestreo (UPM) a partir de la agrupación de sectores censales, en tanto que las unidades de segunda etapa están representadas por las viviendas.

Las UPM fueron seleccionadas con probabilidad proporcional al tamaño utilizando como medida de tamaño el total de viviendas reportadas en la cartografía censal. Por su parte, las viviendas se seleccionaron asignando la misma probabilidad de selección a todas las viviendas que conforman una unidad primaria, es decir, con muestreo aleatorio simple.[31] El tamaño de muestra se definió teniendo en cuenta: dominios de estudio, nivel de confianza y la precisión de las estimaciones, errores de muestreo para variables importantes y la no respuesta o no entrevista.

Bajo estos parámetros estadísticos, se determinó un tamaño de muestra para los dominios de estimación: Nacional, provincial y cuatro ciudades principales (Quito, Guayaquil, Cuenca, Machala).

El Distrito Metropolitano de Quito está dividido en 8551[29] sectores censales , y bajo el diseño muestral descrito anteriormente, basta considerar 192 sectores censales para que la muestra sea representativa en el DMQ. [31] Sin embargo, después de la depuración de la base generada por la ECV y su validación con otras fuentes de información acerca de los sectores censales, se tienen 191 sectores para los cuales se construirá el indicador. A partir de estos resultados, se deducirá el indicador para las 65 parroquias del DMQ.

Ahora, la muestra seleccionada es representativa para todo el DMQ, pero la distribución de los sectores censales no necesariamente es representativa para cada una de las parroquias. Por esta razón, es posible tener desde un solo sector hasta varios sectores en una parroquia. Entonces, si existe un único sector en la parroquia entonces se le asignará el valor de indicador del sector y en caso de que exista más de un

sector se le asignará la media de los indicadores de los sectores.

Inicialmente se consideraron 33 de 80 variables que constan en la encuesta, presentadas en el cuadro 3.1, seleccionadas bajo el criterio de contar menos del 30 % de datos faltantes. De hecho, al momento de aplicar dicho filtro a la base, las variables no tenían ningún dato faltante; además, estarían apoyadas en la propuesta de las necesidades a escala humana [36], de las capacidades humanas [57][20][15], y de la justicia social [15][46][60]. En las cuales se tienen en cuenta otros factores diferentes al ingreso económico, para considerar las condiciones de vida. Estas variables se pueden agrupar en tres dimensiones, *hogar y medio ambiente*, *aspecto económico* (distinto a la distribución del ingreso), y *constitución familiar*.

La definición de estas dimensiones se ha realizado con base en el conjunto de preguntas realizadas en la ECV; ahora, esto no quiere decir que esas sean las únicas dimensiones a considerar, de hecho se puede considerar *seguridad, salud, movilidad*, entre otros, como dimensiones del análisis de calidad de vida.[43] Sin embargo, la estructura de la ECV no cuenta con preguntas que tengan relación con las dimensiones antes mencionadas, razón por la cual el análisis se plantea con base en las variables presentes en la encuesta.<sup>1</sup>

## 3.2. Categorización de datos

El método del PCA No Lineal permite analizar una base de datos con variables cualitativas (nominales u ordinales) y cuantitativas (numéricas). Dado que el objetivo es construir un indicador que cuantifique el nivel de desigualdad entre los sectores del DMQ con el fin de estratificarlos, es necesario categorizar las variables a un nivel ordinal ya que esta categorización está restringida por el orden. Esto nos permitirá crear el indicador entre los valores 0 y 1, relacionados con menor desigualdad y mayor desigualdad respectivamente.

En este proceso creamos nuevas variables a partir de la unión de las ya existentes en la encuesta, con el fin de discriminar mejor características de accesibilidad a servicios básicos y percepción de bienestar.

Se creó una variable para identificar si un hogar tiene acceso a la canasta vital y a la canasta básica a partir de la variable *Monto mínimo para vivir bien* de la ECV.

También se creó una variable a partir de la información de dónde obtienen agua

---

<sup>1</sup>Las variables definitivas a utilizar en el modelo serán seleccionadas después del tratamiento de los datos; además, de ser necesario se crearán nuevas variables a partir de otras.

Variable	Abreviatura	Escala
Tipo de servicio higiénico	HVI001	Nominal
Dispone el hogar de ducha	HVI004	Nominal
Tiene servicio telefónico convencional	HVI006	Nominal
Tiene servicio de internet	HVI007	Nominal
Tiene servicio TV por cable	HVI008	Nominal
Cómo eliminan la basura	HVI014	Nominal
Clasifica los desechos en su hogar	HVI016	Nominal
De dónde se obtiene el agua	HVI017	Nominal
Tratamiento que se le da al agua	HVI018	Nominal
Tienen acceso - a veredas	HVI019	Nominal
Tienen acceso - a parques	HVI020	Nominal
Tipo de vivienda	HVI024	Nominal
Tenencia de la vivienda	HVI025	Nominal
Tiene focos ahorradores en el hogar	HVI028	Nominal
Acceso a espacios verdes	HVI029	Nominal
Vías de acceso al hogar	HVI030	Nominal
Las últimas dos semanas tuvo suficiente comida	HVI033	Nominal
Las últimas dos semanas tuvo problemas para pagar gastos en alimentos	HVI034	Nominal
En últimos 12 meses realizaron préstamos	HVI036	Nominal
Realizaron compras a crédito en últimos 12 meses	HVI037	Nominal
Cómo viven con los ingresos que tienen	HVI039	Nominal
Monto mínimo mensual para vivir bien	HVI040	Numérica
En la actual situación económica,..principalmente	HVI041	Nominal
Usted considera que su hogar es	HVI042	Nominal
El nivel de vida de su hogar en los últimos 12 meses	HVI043	Nominal
El hogar tiene negocios	HVI044	Nominal
Tienen tierras propias	HVI045	Nominal
Perceptores por hogar	HVI060	Numérica
Sexo del jefe del hogar	HJH001	Nominal
Grupos <sup>2</sup> de edad del Jefe de Hogar	HJH003	Nominal
Nivel de instrucción del Jefe de Hogar	HJH006	Nominal
Afiliación a Seguro de Salud del Jefe de Hogar	HJH009	Nominal
Relación de dependencia del jefe de hogar	HJH011	Nominal

**Cuadro 3.1:** Variables seleccionadas

los hogares (HVI017) y el tratamiento que se le da a la misma después de obtenerla (HVI018). Ya que filtrar agua que se ha obtenido de una red pública de agua potable no es igual a filtrar agua obtenida de un río.

Luego, cambiamos la escala original de las variables en el cuadro 3.1 y las variables creadas, excepto la variable *Perceptores por hogar* que es de tipo numérico, y las transformamos a ordinales. De este modo se asignó la puntuación más baja (1) a las categorías o atributos de menor valor o peor condición, y de manera análoga, la puntuación más alta a los atributos de mayor valor o mejor condición. En su mayoría, las categorizaciones son triviales, pero existen algunas que requieren un sustento teórico.

Respecto a la variable *HJH001*, el género del jefe del hogar, ésta no se plantea

con la intención de señalar que un hombre es mejor que una mujer, más bien se la realizó considerando que las desigualdades de género en el Ecuador, vinculadas con las desigualdades sociales, no dejan de ser problemas no resueltos. [62][7]

De hecho, las desigualdades sociales entre los hombres y mujeres se trata de un proceso histórico y complejo de relaciones sociales. Según Vizcarra Bordi (2008), estas desigualdades se basan en la creencia de que las diferenciaciones sexuales, donde lo femenino es inferior a lo masculino, justifican relaciones de dominación y privilegios de unos sobre otras en todos los referentes sociales.

Además, en un estudio realizado acerca de la viejas y nuevas desigualdades en Quito, al momento de analizar los resultados de las elecciones generales de 2013 se pudo ver que existía discrepancia de género en espacios locales de decisión política, desigualdad que articula a otro tipo de desigualdades, como la brecha salarial entre hombres y mujeres, dificultad para acceder al mercado laboral, entre otros.[8]

Por otro lado, es importante resaltar que la variable *HJH006* es el nivel de estudios del jefe de hogar, razón por la cual no se puede combinar con la variable *HVI060* que es el número de perceptores por hogar.

En el caso de tener el nivel de estudios de cada uno de los perceptores, sería interesante realizar un estudio considerando la unión de esas variables ya que las interacciones entre mercado de trabajo y el hogar modifican la capacidad redistributiva del sistema familiar y pueden hacer aparecer nuevas vulnerabilidades. [17]

La categorización de la variable *HJH009* se realizó tomando en cuenta que la seguridad social, además de reconocerse como un derecho contemplado y regulado en la actual Constitución del Ecuador, se considera un mecanismo jurídico que condiciona la consecución de otros derechos humanos y tiene como principal objetivo cubrir las necesidades básicas de las personas ante posibles situaciones sociales de desventaja. [25]

Conforme a informes emitidos, el Instituto Ecuatoriano de Seguridad Social (IESS) cuenta con un excedente significativo, con el que se podría garantizar algunos años de prestaciones a sus afiliados; pero, dada la inclusión de personas que no aportan al seguro social (cónyuges e hijos menores de edad), algunos de los trabajadores afiliados al IESS ya no utilizan sus servicios por considerarlos deficientes y optan por la contratación de seguros privados.[21] Así, vemos que las personas no afiliadas a ningún tipo de seguro, se encuentran en situación de desigualdad respecto a los afiliados, ya que no se garantiza su derecho a la seguridad social; ahora, los afiliados al IESS que pueden contratar un seguro privado, además de tener garantizado ese

derecho, también tienen acceso a una mayor eficiencia en la gestión del sistema, a una mejor cobertura y la pronta atención médica.[21][48]

Según la Encuesta de percepción de usuarios y beneficiarios del IESS sobre el desempeño del Instituto en términos de gobernanza, cobertura, calidad y sostenibilidad realizada por la Organización Internacional del Trabajo, los afiliados voluntarios, los afiliados al Seguro Social Campesino, y las trabajadoras no remuneradas del hogar expresaban inconformidad con respecto a la eficacia al momento de acceder a los servicios del IESS. En este aspecto, las necesidades de estos grupos no han sido satisfechas en su totalidad vulnerando su derecho a salud, seguridad social e igualdad.[2]

La variable *HJH011* tiene que ver con la Relación de dependencia del jefe de hogar, para la cual se ha considerado tres opciones dentro de la encuesta, mismas que, en este trabajo, han sido categorizadas de acuerdo con la situación de desigualdad en la que se encuentran.

Los trabajadores no remunerados se dividen en dos grupos: hogar y no hogar. El primer grupo, aunque realice alguna labor no recibe ninguna retribución por ello, por lo cual estaría en una condición de vulnerabilidad. Luego, aunque en la Ley Orgánica de Justicia Laboral y Reconocimiento del trabajo no remunerado del hogar existan normativas que promuevan al igualdad para esta minoría, no tiene mecanismos normativos que realmente garanticen el ejercicio de los derechos fundamentales para este grupo, ubicándolos así, en una posición de desigualdad.[2]

Los trabajadores formales e informales pueden ser dependientes o independientes y se estima según la ECV más del 52 % de personas trabajaban en la informalidad, dentro de este grupo el 26,20 % era informal dependiente y el 26,60 % era informal independiente; el 47 % de la población tenía un empleo formal, donde el 40,64 % era formal dependiente y sólo el 6,56 % formal independiente.

Con estos resultados vemos que la mayoría de trabajadores independientes pertenecen al empleo informal, y una de las afectaciones más importantes para estos trabajadores es la pobreza y la desigualdad social, esto como consecuencia de los bajos salarios y la poca, o incluso nula, protección social; esto en comparación con un trabajador formal dependiente que cuenta con mejores garantías laborales y de seguridad social. [6]

En el anexo D, presentamos las cuantificaciones obtenidas de las variables seleccionadas y creadas a partir de la Encuesta de Condiciones de Vida.

### 3.3. Aplicación del NPCA basado en un kernel

Previo a la aplicación de los algoritmos<sup>3</sup> descritos en la sección 2.3.1 es necesario realizar un escalado óptimo a los datos<sup>4</sup> ya que las variables están categorizadas. Además, tomaremos en cuenta dos casos de aplicación como en la teoría desarrollada en el capítulo anterior.

#### 3.3.1. Matriz de pesos espaciales

Con la base de datos escalada procedemos a construir la matriz de pesos espaciales que, con base los métodos LWPCA (Sección 2.3) y el algoritmo descrito en la sección 2.3.1, es el kernel calculado mediante la función de decaimiento de distancia de Gauss. Para lo cual utilizaremos las variables detalladas en el cuadro 3.2, combinación con la cual se obtuvo el mayor porcentaje de representatividad en las dos primeras componentes al aplicar el algoritmo del PCA No Lineal.

Variable	Abreviatura
Tipo de servicio higiénico	HVI001
Tienen acceso a parques	HVI020
Acceso a espacios verdes	HVI029
En los dos últimos meses realizaron préstamos	HVI036
Realizaron compras a crédito en los últimos 12 meses	HVI037
El nivel de vida de su hogar en los últimos 12 meses	HVI043
Perceptores por hogar	HVI060
Sexo del jefe del hogar	HJH001
Grupos de edad del jefe de hogar	HJH003
Relación de dependencia del jefe de hogar	HJH011

**Cuadro 3.2:** Variables utilizadas para la construcción del indicador

---

<sup>3</sup>Los algoritmos se implementaron en el programa Rstudio. El código se presenta en el Anexo E.

<sup>4</sup>El escalado óptimo asigna valores a las categorías o atributos de una variable categórica mediante algún criterio de optimización que separe, o discrimine, los grupos de casos que se han formado al cruzar los casos con dicha variable

### 3.3.2. Caso sin ponderación geográfica

Dada la matriz de pesos espaciales o matriz kernel  $K$ , aplicamos el método del PCA No Lineal descrito en la sección 2.3.1 a partir del paso 3. La primera componente principal  $U_1$  explica 46.21 % de la variabilidad total y la segunda componente  $U_2$  explica 10.93 %. Entonces, las dos componentes explican el 57.14 % de variabilidad total, lo que sugiere que es necesario considerar más componentes para recoger la mayor parte de información de los datos. Los 191 sectores están representados por puntos en el sistema, en base a las dos primeras componentes principales; además, construimos un árbol de mínima expansión con estos puntos, lo que indica la división de los sectores en grupos homogéneos. (Ver figura 3.1)

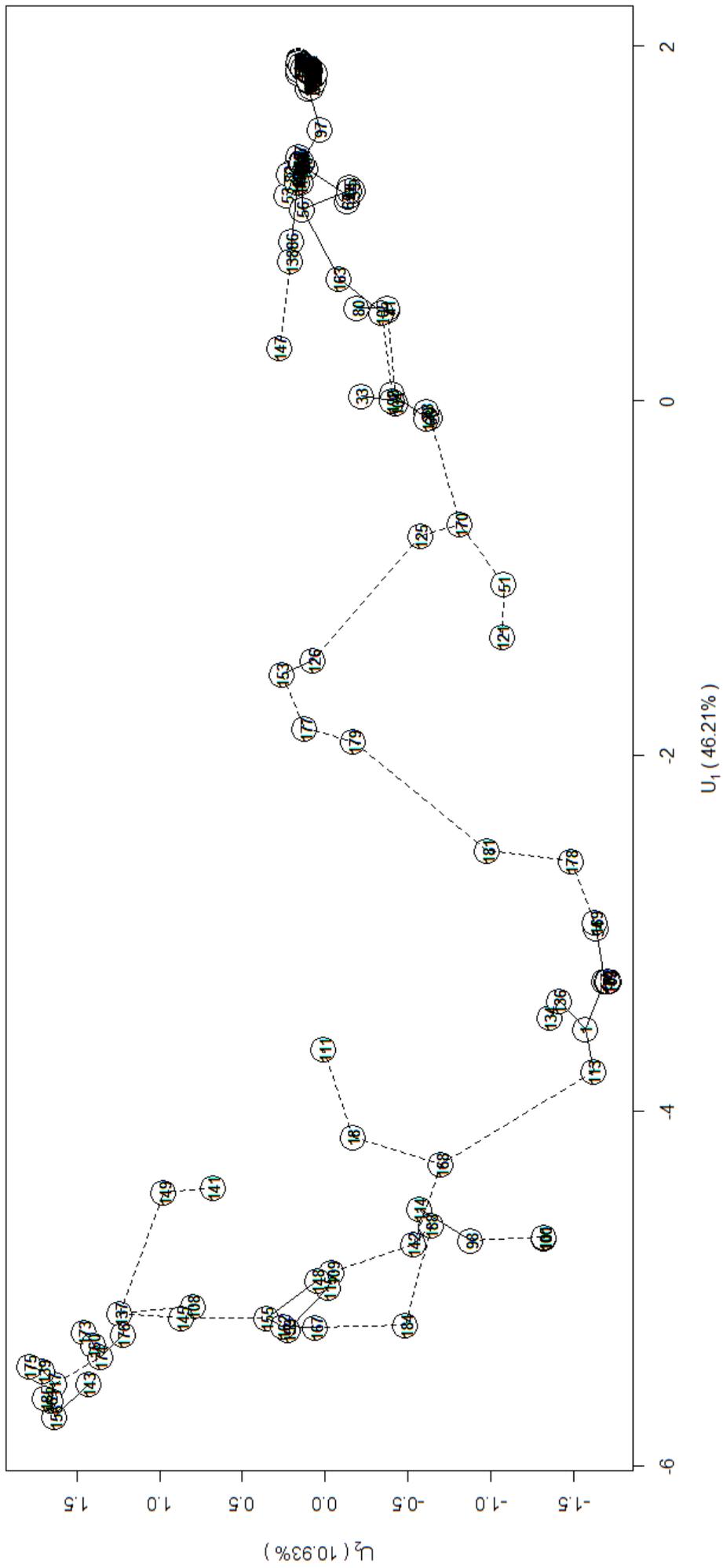
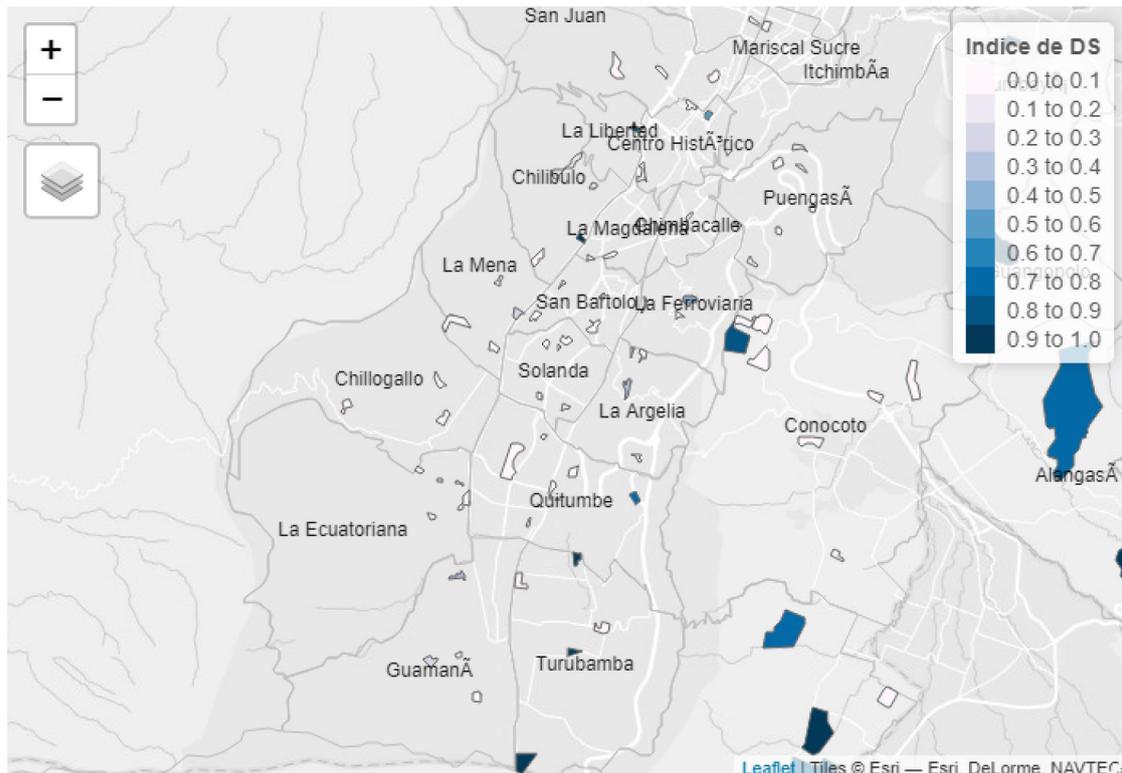


Figura 3.1: Resultados KNPFA sin ponderación geográfica





**Figura 3.3:** Indicador DS para los sectores censales sur del DMQ, sin ponderación geográfica

### 3.3.3. Caso con ponderación geográfica

Según Schölkopf [54], la accesibilidad de transporte de una región a otra es una característica que considera la influencia de la ubicación geográfica de una región y toman como pesos geográficos un índice de accesibilidad de transporte. Por otro lado, la tierra es un factor importante para la planificación urbana y es un instrumento básico para políticas de vivienda y construcción del espacio público. Por lo que analizar cuánto vale un determinado terreno urbano y rural es importante, ya que este valor incrementa a través del tiempo debido al mejoramiento de diferentes factores como su ubicación dentro del entorno urbano, servicios e infraestructura, valor urbano e arquitectónico, y accesibilidad. [49]

Así, valor del suelo es una característica que, además de accesibilidad de transporte, toma en cuenta atributos que describen mejor la ubicación geográfica de un lugar.

Con base en la metodología de Schölkopf [54], para este trabajo consideraremos como pesos geográficos  $w_i$  el costo del metro cuadrado de cada sector  $i, i = 1, \dots, 191$ , o valor del suelo, cuyo cálculo se detalla en la Ordenanza No.152 emitida por el Consejo del Distrito Metropolitano de Quito en el 2011[16], multiplicado por la pro-

porción del número de viviendas y el número de hogares en cada sector, éstos dos obtenidos de la Cartografía del INEC 2010 [29].

De este modo, multiplicamos los  $w_i, i = 1, \dots, 191$  por la matriz de datos escalada y aplicamos el método de manera análoga a la sección anterior. En este caso, la primera componente principal  $U_1$  explica 38.4% de la variabilidad total y la segunda componente  $U_2$  explica 20.7%. Entonces, las dos componentes explican el 59.1% de variabilidad total. Así, es necesario considerar más componentes para recoger la mayor parte de información de los datos. En la figura 3.4 presentamos los 191 sectores tomando en cuenta las dos primeras componentes principales y el árbol de mínima expansión.

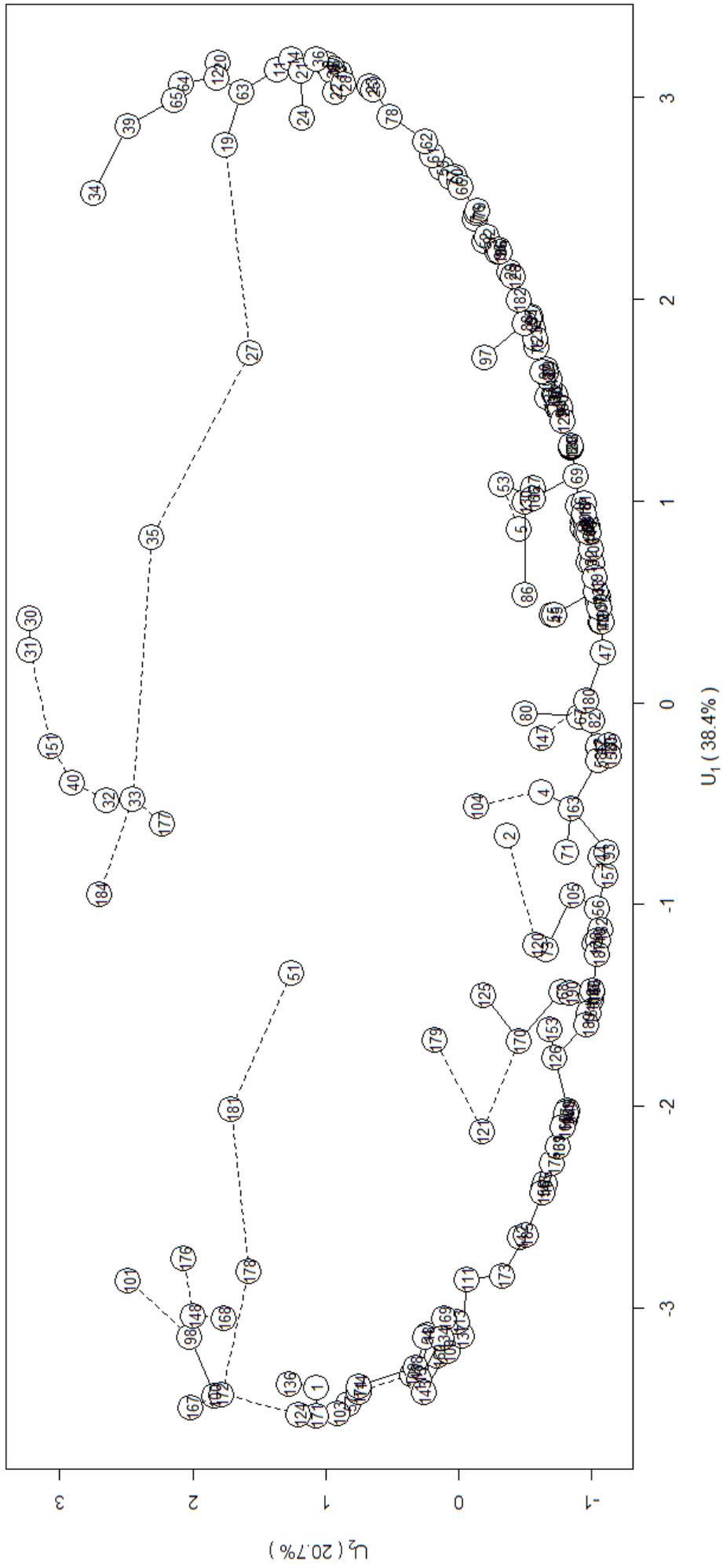
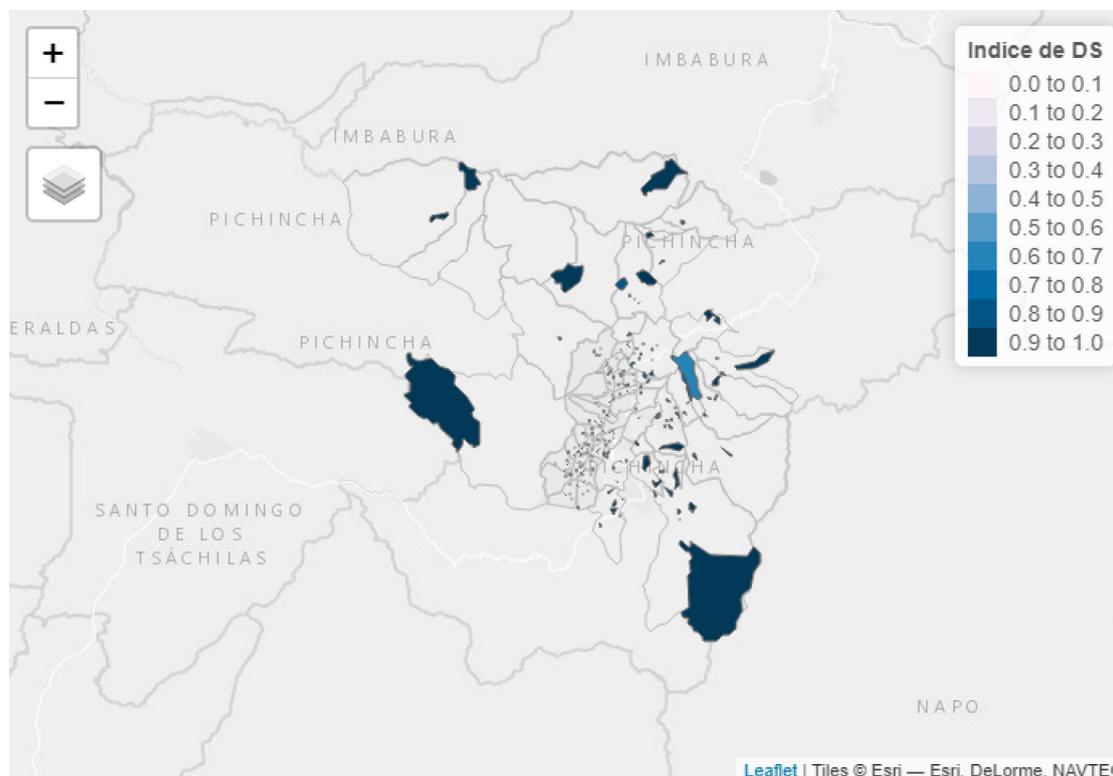


Figura 3.4: Resultados KNPCA sin ponderación geográfica

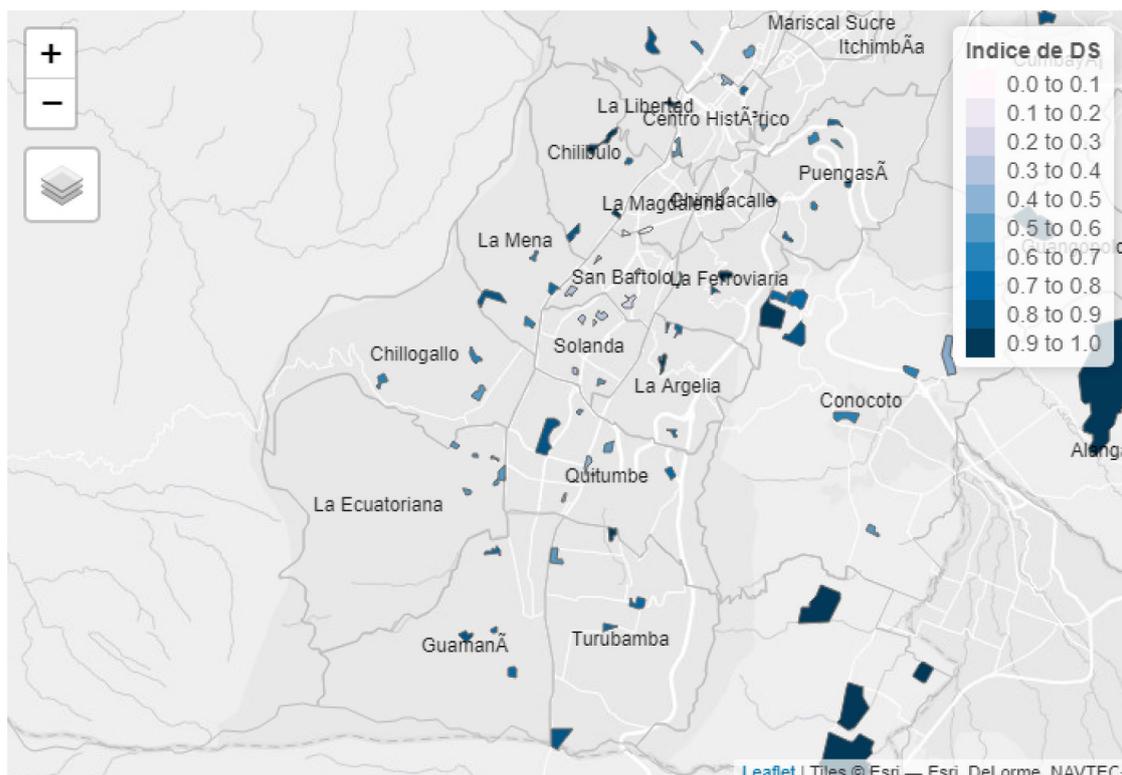
De manera análoga a la sección anterior, presentamos el indicador DS para los sectores censales. (Figura 3.5)



**Figura 3.5:** Indicador DS para los sectores censales, con ponderación geográfica

Para continuar con lo planteado en el caso anterior, veamos el mapa con el indicador de los sectores censales del sur de la ciudad. (Figura 3.6) Es claro que este resultado va acorde con la situación del DMQ descrita en el estudio de evolución de la pobreza [53] realizado por el Municipio.

La comparación de entre los valores del indicador para los dos casos confirma que añadir un peso geográfico proporcionó resultados más adecuados que el KNP-CA sin ponderación geográfica, por lo cual construiremos la aproximación del indicador de desigualdad social parroquial utilizando las componentes principales obtenidas de KNPCA con ponderación geográfica.



**Figura 3.6:** Indicador DS para los sectores censales sur del DMQ, con ponderación geográfica

### 3.4. Análisis de resultados

De acuerdo con los resultados del PCA para tener un 70% de variabilidad total de los datos necesitamos 5 componentes principales y utilizamos las componentes principales resultantes del KNPCA ponderado geográficamente, ya que sus resultados reflejaban mejor la realidad comparados con el caso sin ponderación. La construcción del indicador se realizó de acuerdo con los puntos 7 y 8 del procedimiento descrito en la sección 2.6.

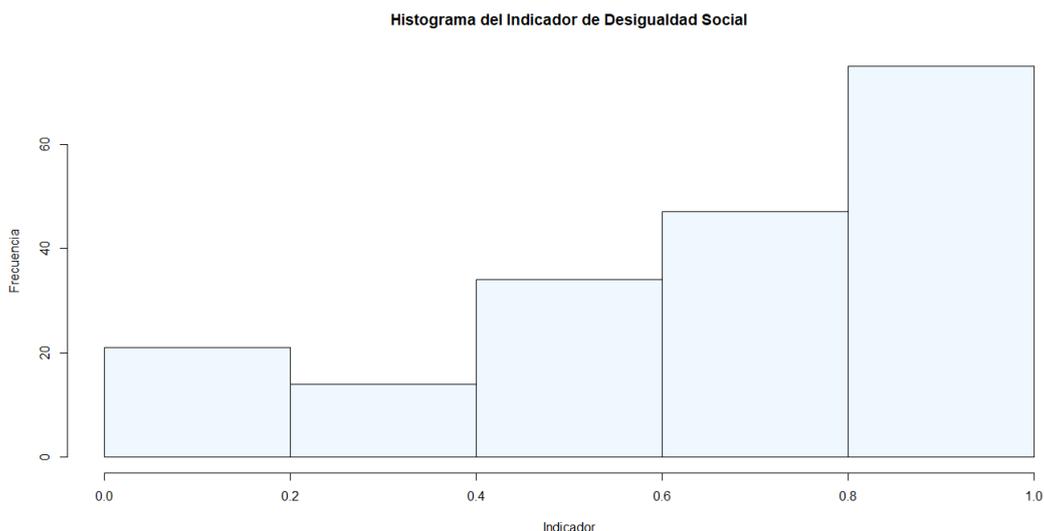
A continuación, en el cuadro 3.3 presentamos las medidas de tendencia central y dispersión del indicador. Además, en el cuadro 3.4 se detallan su quintiles e histograma en la figura 3.7.

Media	0.6645
Mediana	0.718
Desviación estándar	0.276
Varianza	0.076
Coefficiente de variación	0.415

**Cuadro 3.3:** Resumen del indicador

0 %	0.00
20 %	0.436
40 %	0.630
60 %	0.784
80 %	0.933
100 %	100

**Cuadro 3.4:** Quintiles del indicador de desigualdad social



**Figura 3.7:** Histograma del Indicador de Desigualdad social (cinco clases)

Ahora, veamos el histograma considerando 10 clases o grupos. (Ver figura 3.8).

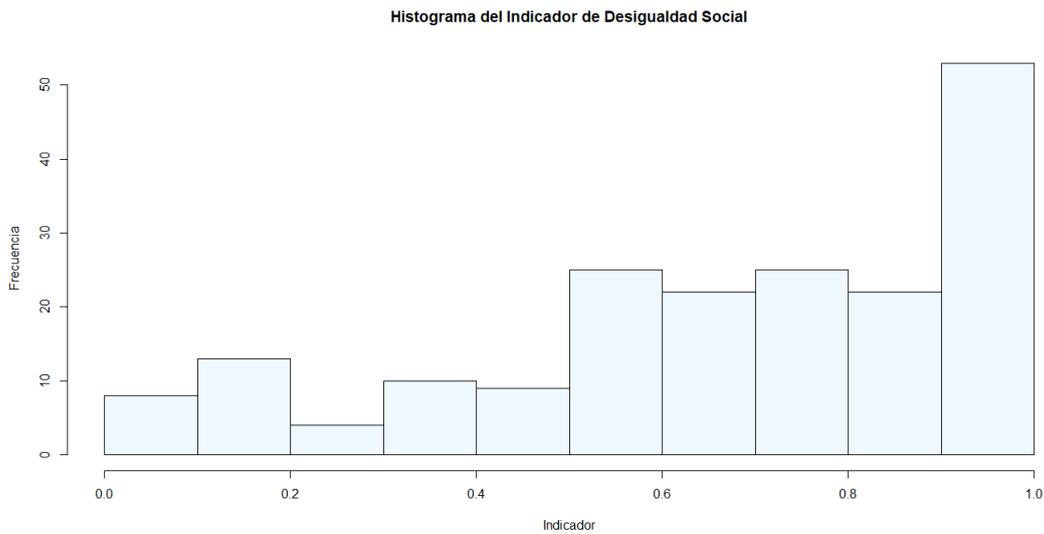
Si comparamos las figuras 3.7 y 3.8 es claro que en el histograma con 10 grupos tenemos más información respecto a la distribución del indicador, comparado con el histograma elaborado con cinco clases.

Así, para nuestro análisis agruparemos los sectores en diez grupos determinados por intervalos. En el grupo 1 estarán los sectores que tengan un indicador de desigualdad menor o igual a 0.1; el grupo dos estará conformado por los sectores cuyo indicador es mayor a 0.1, pero menor o igual a 0.2, y así sucesivamente. Donde el grupo 1 presenta menor desigualdad y el grupo 10 mayor desigualdad.

Esto se realiza con el objetivo de discriminar mejor los resultados obtenidos e identificar los sectores donde realmente existe desigualdad.

Se presenta un resumen de los grupos en el cuadro 3.5.

A partir de esta agrupación y considerando los 191 sectores censales de la ECV, se tiene que 53 sectores, es decir el 27.7%, pertenecen al grupo 10, siendo el grupo con mayor cantidad de sectores censales. Mientras que, solo el 2.09% de sectores



**Figura 3.8:** Histograma del Indicador de Desigualdad social (diez clases)

<b>Grupo</b>	<b>Intervalo</b>
1	(0 , 0.1]
2	(0.1 , 0.2]
3	(0.2 , 0.3]
4	(0.3 , 0.4]
5	(0.4 , 0.5]
6	(0.5 , 0.6]
7	(0.6 , 0.7]
8	(0.7 , 0.8]
9	(0.8 , 0.9]
10	(0.9 , 1.0]

**Cuadro 3.5:** Grupos del indicador de desigualdad social

pertenecen al grupo 3, siendo el grupo con menor cantidad de sectores censales (Ver cuadro 3.6).

Los grupos 1, 2 y 3, dentro de los cuales están los sectores de menor desigualdad, contienen al 13.08 % del total de sectores. Por otro lado, en los grupos 7, 8, 9, 10, donde se presenta mayor desigualdad, se encuentra el 63.84 % de los sectores.

Además, la diferencia de la cantidad de sectores entre el grupo 5 y el grupo 6, que nos da una ligera impresión de que existe una brecha de desigualdad dentro del DMQ.

Partiendo de estos resultados y de que la muestra analizada es representativa para el DMQ, es posible afirmar que el tema de la desigualdad en la ciudad, hasta el momento en que se realizó la ECV en el 2014, era una problemática social que

Grupo	Número se SC <sup>5</sup>	Porcentaje
1	8	4.18 %
2	13	6.80 %
3	4	2.09 %
4	10	5.23 %
5	9	4.71 %
6	25	13.08 %
7	22	11.51 %
8	25	13.08 %
9	22	11.51 %
10	53	27.74 %

**Cuadro 3.6:** Sectores por grupo

no había sido totalmente identificada, ya que los procesos de urbanización y modernización de las ciudades generan injusticia social provocando el surgimiento de nuevas geografías centrales y marginales, y con esto nuevas desigualdades. [8]

De hecho, según el INEC el Coeficiente de Gini en el 2007 se ubicó en 0.49 y en el 2015 en 0.453, lo que quiere decir que en ocho años hubo un descenso poco significativo en los niveles de desigualdad por ingresos en la ciudad.

A partir del indicador obtenido para cada uno de los sectores censales se deduce el indicador para las parroquias del DMQ con base en la información del *Diagnóstico del Territorio del DMQ* [42] realizado el 7 de noviembre del 2014.

Este indicador tiene la intención de reflejar la desigualdad que existe en una parroquia, sin embargo por cómo se deduce el indicador a partir de los sectores censales es necesario considerar que en las parroquias pueden existir barrios con menor desigualdad de la que se le ha asignado. No obstante, debemos enfocarnos en los lugares donde sí se presenta dicha desigualdad para dar solución a esta problemática, y garantizar bienestar y calidad de vida a toda la población.

En el cuadro 3.7 y en la figura 3.9 se presentan los resultados del indicador para las parroquias del distrito.

PARROQUIA	INDICADOR
Gualea	0,98
Chavezpamba	0,98
Perucho	0,98
Puéllaro	0,98
El Quinche	0,98

Sigue en la página siguiente.

<b>PARROQUIA</b>	<b>INDICADOR</b>
Nono	0,97
Nanegal	0,97
San José de Minas	0,97
Atahualpa	0,97
Calacalí	0,97
Lloa	0,97
La Merced	0,97
Nanegalito	0,96
Píntag	0,96
Pífo	0,96
Pomasqui	0,96
Pacto	0,95
Yaruquí	0,95
Cochapamba	0,94
Guangopolo	0,94
Amaguaña	0,92
La Ferroviaria	0,91
Checa	0,91
Puengasí	0,9
La Libertad	0,9
La Argelia	0,89
Chilibulo	0,89
Alangasí	0,89
Guamaní	0,88
La Mena	0,88
San Juan	0,88
Guayllabamba	0,88
Comité del Pueblo	0,87
Calderón	0,87
Zámbiza	0,87
Chillogallo	0,86
Conocoto	0,86
San Antonio de Pichincha	0,85
El Condado	0,83

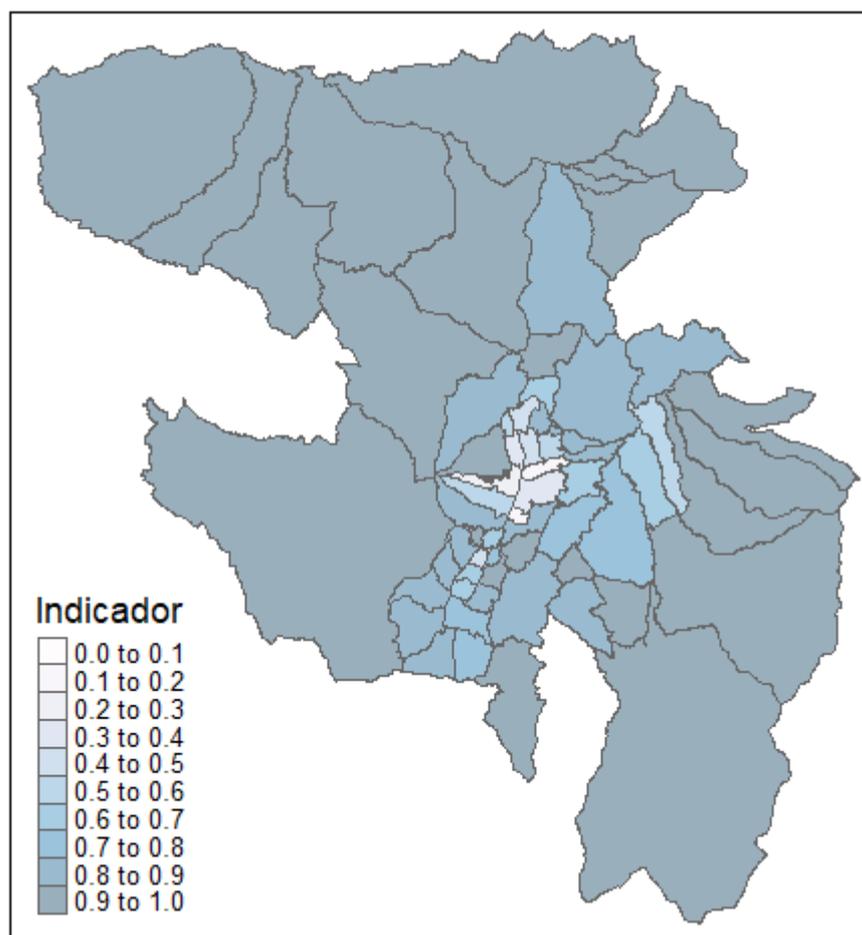
Sigue en la página siguiente.

PARROQUIA	INDICADOR
Llano Chico	0,82
La Ecuatoriana	0,81
Itchimbía	0,8
Turubamba	0,78
Quitumbe	0,77
Cumbayá	0,76
Tumbaco	0,74
Chimbacalle	0,72
Puembo	0,69
Solanda	0,68
Nayón	0,68
Centro Histórico	0,66
San Bartolo	0,64
Carcelén	0,63
Cotocollao	0,59
Tababela	0,59
Belisario Quevedo	0,58
San Isidro del Inca	0,57
Ponceano	0,49
Kennedy	0,48
La Magdalena	0,41
Concepción	0,39
Iñaquito	0,34
Rumipamba	0,26
Mariscal Sucre	0,16
Jipijapa	0,16

**Cuadro 3.7:** Indicador por parroquia, DMQ

Vemos que la tendencia espacial es de adentro hacia afuera, razón por la cual el desarrollo de equipamientos, infraestructura y accesibilidad se ha visto beneficiado en el centro de Quito y se tiene la carencia de estos factores en el norte, sur y periferia de la ciudad, acentuando la brecha de desigualdad.[35]

Este fenómeno tiene que ver con la forma como han crecido las ciudades de América Latina, que han llevado a la *fragmentación socio-espacial* en las ciudades del



**Figura 3.9:** Mapa del Distrito Metropolitano de Quito

continente, ya que es en las periferias donde se distingue el crecimiento de áreas metropolitanas fragmentadas que incrementan la desigualdad en el acceso a la ciudad y a sus bienes, servicios y oportunidades, al sistema educativo y al mercado de trabajo.[56].

Las parroquias Mariscal Sucre y Jipijapa son los territorios que presentan menor desigualdad. En la tesis *Administración zonal norte de Quito "Eugenio Espejo"* [35], realizada en el 2014 se describe la situación de la parroquia Jipijapa,

Sus pobladores no tienen problemas de infraestructura y mucho menos de dotación de servicios básicos. Ellos atribuyen este logro al desarrollo del lugar, al trabajo que los comités directivos han realizado durante todos estos años. Este barrio capitalino destaca por sus parques, que están estratégicamente distribuidos, lo que ha dotado de un buen espacio de área verde al lugar.

comentario que va acorde al valor del indicador para esta localidad.

Por otro lado, en el artículo Evolución de la pobreza y la desigualdad en Quito [53], se presenta un estudio acerca de las necesidades básicas insatisfechas de las parroquias del DMQ mediante la construcción de un índice de Necesidades Básicas Insatisfechas (NBI) que consideró la falta de acceso de la población a servicios básicos, acceso a la educación, hacinamiento en la vivienda y la dependencia demográfica dentro del hogar. Esta investigación registró que las parroquias El Quinche, Perucho, Guayallabamba, Guangopolo, La Merced, Pifo, Yaruquí, Chavezpamba, Lloa, Atahualpa, Nanegalito, Pintag, Checa, Puéllaro, San José de Minas, Nanengal, Pacto, Gualea y Nono, presentan más del 50 % de insatisfacción.

De hecho, Gualea, Chavezpamba, Perucho, Puéllaro y El Quinche presentan 98 % de desigualdad social; Nono, Nanengal, San José de Minas, Atahualpa, Lloa, La Merced presentan 97 % de desigualdad social, resultados que nos permiten ver que en efecto las parroquias que presentan mayor desigualdad según el cuadro 3.9 también presentaron mayor insatisfacción en el estudio mencionado anteriormente.

La unidad de estudio de la ECV son hogares dentro de un sector censal, pero no nos entrega la ubicación geográfica<sup>6</sup> de cada unidad. Pero si un sector tiene un indicador de desigualdad cercano a 1, entonces allí existen hogares en situación de vulnerabilidad.

Ahora, dado que los sectores censales son parte de la información cartográfica, entonces podemos obtener la ubicación geográfica del grupo de hogares encuestados, y para ubicar de forma exacta grupos vulnerables bastaría tener la ubicación geográfica de los hogares pero esa información es de difícil acceso. Razón por la cual utilizaremos los sectores censales para señalar la ubicación de grupos en una posible condición de vulnerabilidad.

Los sectores censales no tienen un nombre definido pero sí un identificador único compuesto, que consta del código de la provincia, seguido del código de la ciudad, de la zona, y del sector. Como se ha mencionado anteriormente, para la ECV se tomó una muestra de 191 sectores representativa para el DMQ, pero es importante tomar en cuenta que estos sectores no necesariamente colindan entre sí, de hecho la mayoría está distribuida de forma dispersa en todo el territorio del distrito.

Así pues, con el fin de ilustrar el hecho de que se puede ubicar geográficamente grupos en condición de vulnerabilidad, consideramos 4 sectores censales de la parroquia San Isidro del Inca, con su respectivo valor del indicador (Ver cuadro 3.8).

---

<sup>6</sup>La **ubicación geográfica** es la identificación de un lugar específico del planeta, mediante el uso de diversas herramientas como mapas, brújulas, coordenadas o sistemas de geolocalización.

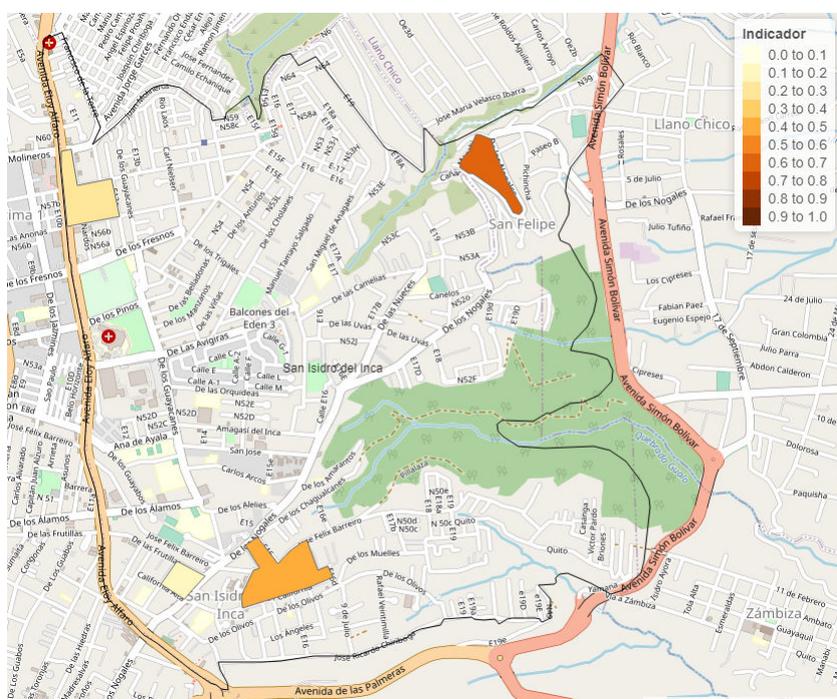
De acuerdo con el indicador de desigualdad calculado para los 4 sectores censales, es posible que dentro del sector censal 170165001012 existan más hogares vulnerables, comparado con el sector 170150120005.

En la figura 3.10 presentamos la ubicación de los sectores en el mapa. Esto nos permite identificar dónde se encuentran los sectores para determinar los factores que influyen en la desigualdad de esa región.

De hecho, es interesante ver que aunque los sectores se encuentran en una misma parroquia existe heterogeneidad entre ellos. Es probable que esto se deba al aumento descontrolado de los precios del suelo entre el 2010 y el 2014 pues se produjo una segregación en la que zonas de altos ingresos llegaron a residir en proximidad espacial con zonas de bajos ingresos, sin que esto signifique una integración ni interacción social.[8]

Id Sector	Indicador
170150112005	0,53
170150122006	0,69
170150120005	0,48
170165001012	0,89

**Cuadro 3.8:** Sectores censales en la Parroquia



**Figura 3.10:** Ubicación de los sectores censales de la parroquia San Isidro del Inca

# Capítulo 4

## Conclusiones y Recomendaciones

### 4.1. Conclusiones

La metodología utilizada para la construcción del indicador de desigualdad social se diferencia de los indicadores descritos en la sección 2.5.1 porque considera variables de condiciones de vida que van más allá del ingreso. Además, toma en cuenta el espacio y el lugar geográfico donde se encuentra la unidad de estudio, características que no son tomadas en cuenta por ninguno de los indicadores antes mencionados.

Por lo cual, con las variables finales se logró construir la matriz de pesos espaciales mediante la función de decaimiento de distancia de Gauss y la norma de Frobenius, ya que el decaimiento de distancia describe el efecto de la distancia en las interacciones espaciales.

Asimismo, al considerar el valor del suelo y la proporción del número de viviendas y el número de hogares como peso geográfico, se pudo describir la influencia geográfica de las regiones de acuerdo con su ubicación.

Dado que, a partir del indicador por sectores censales se dedujo el indicador para las parroquias del Distrito Metropolitano de Quito, se consiguió identificar las parroquias junto con su ubicación geográfica, en las cuales existen grupos en condición de vulnerabilidad respecto a la desigualdad social que presentan esas regiones.

Este indicador *parroquial* no proporciona toda la información que se desea obtener, aunque nos permite medir la condición de desigualdad de la población de una parroquia y dónde está ubicada, no es posible obtener la ubicación exacta de los grupos u hogares vulnerables dentro de esa localidad. Esto, no se debe a la metodología,

sino a la muestra generada para el levantamiento de los datos de la ECV.

Si bien, el desarrollo para cada parroquia es diferente, aunque las etapas básicas siguen patrones similares, es importante comprender los factores de desigualdad que influyen en el territorio, y aunque las parroquias forman parte de un eje importante para el desarrollo del DMQ, es necesario considerar niveles más bajos, como barrios y manzanas.

Hasta el levantamiento de la información de la encuesta, en el 2014, se puede ver que el indicador de desigualdad es bastante alto, por lo que sería necesario hacer una evaluación de la situación actual de las parroquias por parte de las administraciones zonales respectivas y, de ser imprescindible, plantear nuevas políticas de desarrollo e inclusión; ya que el proceso expansivo del Distrito Metropolitano de Quito no garantiza la eliminación de desigualdades socioespaciales, más bien implica la superposición de éstas en el tiempo y el espacio.[8]

Si bien, la desigualdad social presente en el DMQ no depende solo de la alcaldía, sería eficiente si esta entidad aborda esta problemática tomando consideraciones particulares para la realidad de Quito, con el fin de acortar la brecha entre grupos que ejercen una ciudadanía plena y grupos en condición de vulnerabilidad que residen en la ciudad.

En consecuencia, dados los resultados de este trabajo, el indicador de desigualdad social podría ser una herramienta útil para la creación y modificación de políticas públicas implementadas por los Gobiernos Autónomos Descentralizados.

Por otro lado, la metodología desarrollada en este trabajo se puede extender a datos temporales y, asimismo, se puede generalizar para datos funcionales.

## 4.2. Recomendaciones

- Debido a la falta de actualización de datos puede existir una interpretación parcial de los resultados. Para superar este inconveniente será necesario complementarlos con otro tipo de datos secundarios, estudios, informes, entre otros.
- Se recomienda construir el indicador de desigualdad sobre una encuesta que sea actualizada y que tenga una mejor representación de la población, algo que no fue posible realizar en este trabajo ya que la Encuesta de Condiciones de Vida que tuvo lugar en el 2014 es la más reciente y porque el acceso a información desagregada es limitado.
- Se sugiere complementar el indicador de desigualdad social con indicadores de pobreza para tener un mejor panorama del contexto social del Distrito Metropolitano de Quito. Esto se podría realizar mediante la construcción de un indicador multidimensional de desigualdad y pobreza, donde además de atributos correspondientes a condiciones de vida, que fueron incluidos en este trabajo, se puede considerar atributos de accesibilidad de transporte, calidad de red de telefonía móvil por zonas, entre otros.
- Considerando las recomendaciones anteriores, se recomienda a la Secretaría de Planificación del Municipio del Distrito Metropolitano de Quito estudiar las condiciones de las parroquias que presentan mayor desigualdad, con el fin de identificar los factores que contribuyen a este problema estructural y promover políticas sociales que den solución a esta problemática garantizando el bienestar de la población en cuestión.
- Se recomienda realizar este estudio en otras ciudades del Ecuador con el fin de identificar factores propios de las localidades que influyan en la condición de desigualdad de su población. Además, al trabajar con Gobiernos Autónomos Descentralizados se puede tener más información de los hogares y así un indicador más completo que refleje de manera más clara la situación de la región estudiada. En este caso, sería interesante considerar ciudades en las que es posible que los patrones de desigualdad hayan cambiado debido a fenómenos externos.

# Anexo A

## Pasos en la construcción de la *curva de Lorenz*

El proceso<sup>1</sup> para construir la curva de Lorenz está dado de la siguiente manera:

1. Ordenar individuos u hogares de la población en forma ascendente.
2. Definir el tamaño de los intervalos(deciles, quintiles,etc.).
3. Construir frecuencias relativas simple<sup>2</sup> y acumulada<sup>3</sup> de las variables de renta y población.
4. Graficar las columnas de frecuencia relativa acumulada de renta e individuos, tomando en cuenta que la curva de Lorenz nace desde el origen del eje cartesiano.

Ahora, consideremos la renta de 10 individuos, ver cuadro A.1.

Calculamos la frecuencia relativa y frecuencia absoluta, ver cuadro A.2. Graficamos las columnas de frecuencia relativa acumulada de renta e individuos, y se tiene la figura A.1.

---

<sup>1</sup>Este proceso está basado en el Manual sobre cooperación y desarrollo: Desarrollo humano, pobreza y desigualdades.

<sup>2</sup>La **frecuencia relativa simple** equivale al porcentaje que se obtiene del cociente entre el número de observaciones que pertenecen a cada intervalo y el número total de observaciones.

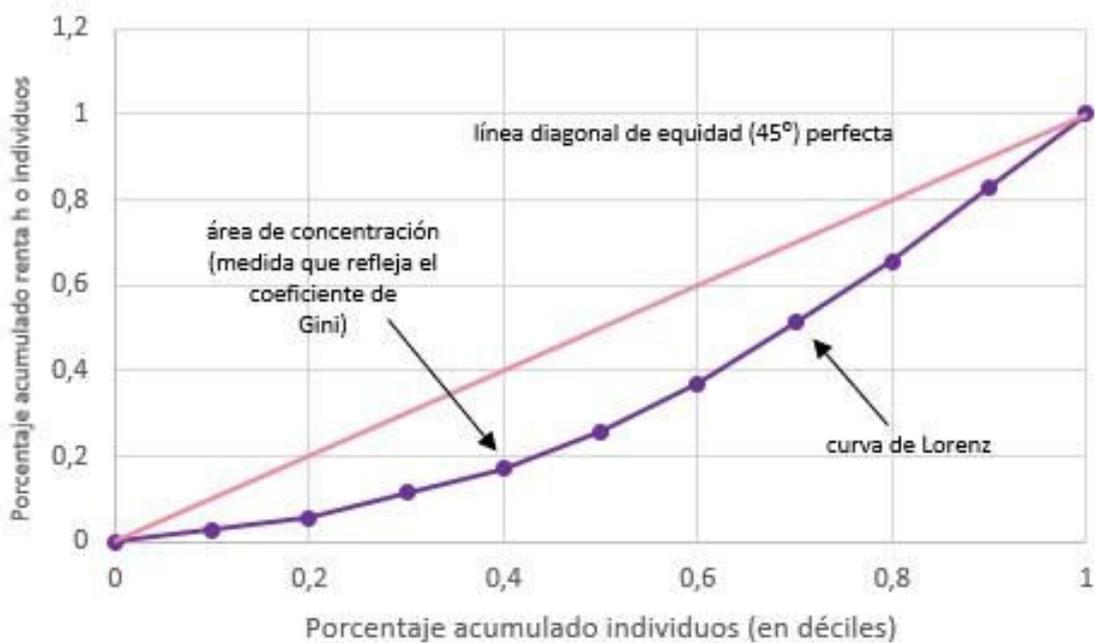
<sup>3</sup>La **frecuencia relativa acumulada** muestra la frecuencia relativa que se va acumulando en orden ascendente hasta llegar al último intervalo.

Deciles	Individuos	Renta
1	1	1
2	1	1
3	1	2
4	1	2
5	1	3
6	1	4
7	1	5
8	1	5
9	1	6
10	1	6

**Cuadro A.1:** Cuadro de datos

Frecuencia relativa		Frecuencia acumulada	
Individuo %	Renta %	Individuos %	Renta %
10.00	3	10	3
10.00	3	20	6
10.00	6	30	11
10.00	6	40	17
10.00	9	50	26
10.00	11	60	37
10.00	14	70	51
10.00	14	80	66
10.00	17	90	83
10.00	17	100	100

**Cuadro A.2:** Frecuencias



**Figura A.1:** Curva de Lorenz. Elaboración: autor

## Anexo B

# Centrado de un espacio de gran dimensión

Dada cualquier función  $\Phi$  y cualquier conjunto de observaciones  $x_1, \dots, x_M$ , los puntos

$$\tilde{\Phi}(x_i) := \Phi(x_i) - \frac{1}{M} \sum_{i=1}^M \Phi(x_i) \quad (\text{B.1})$$

están centrados. Por lo tanto, los supuestos de la sección 2.2 ahora se cumplen.

Definamos la matriz de covarianza y  $\tilde{K}_{ij} = (\tilde{\Phi}(x_i), \tilde{\Phi}(x_j))$  en  $F$ . Así, llegamos al problema de los valores propios,

$$\tilde{\lambda} \tilde{\alpha} = \tilde{K} \tilde{\alpha} \quad (\text{B.2})$$

con  $\tilde{\alpha}$  siendo los coeficientes de expansión de un vector propio, en  $F$ , en términos de los puntos en la ecuación B.1,  $\tilde{V} = \sum_{i=1}^M \tilde{\alpha} \tilde{\Phi}(x_i)$ .

Como no tenemos datos centrados (Ver ecuación B.1), no se puede calcular  $\tilde{K}$  directamente; sin embargo, podemos expresarlo en términos de su contraparte no centrada  $K$ .

Ahora, notemos

$$\begin{aligned} K_{ij} &= (\Phi(x_i), \Phi(x_j)), \\ 1_{ij} &= 1 \quad \text{para todo } i, j, \quad \text{y} \\ (1_M)_{ij} &:= \frac{1}{M} \end{aligned}$$

para calcular  $\tilde{K}_{ij}=(\tilde{\Phi}(x_i), \tilde{\Phi}(x_j))$ :

$$\begin{aligned}\tilde{K}_{ij} &= \left( (\Phi(x_i) - \frac{1}{M} \sum_{m=1}^M \Phi(x_m)), (\Phi(x_j) - \frac{1}{M} \sum_{n=1}^M \Phi(x_n)) \right) \\ &= K_{ij} - \frac{1}{M} \sum_{m=1}^M 1_{im}K_{mj} - \frac{1}{M} \sum_{n=1}^M K_{in}1_{nj} + \frac{1}{M^2} \sum_{m,n=1}^M 1_{im}K_{mn}1_{nj} \\ &= (K - 1_M K - K 1_M + 1_M K 1_M)_{ij}.\end{aligned}\tag{B.3}$$

Por tanto, podemos calcular  $\tilde{K}$  de  $K$  y luego resolver el problema de valores propios de la ecuación B.2. Como en la ecuación 2.24, las soluciones  $\tilde{\alpha}^k$  se normalizan mediante la normalización de los vectores correspondientes  $\tilde{V}^k$  en  $F$ , lo que se traduce en,

$$\tilde{\lambda}_k(\tilde{\alpha}^k, \tilde{\alpha}^k) = 1.$$

Para la extracción de características, calculamos proyecciones de las  $\Phi$ -imagenes centradas de los patrones de prueba  $t$  sobre los vectores propios de la matriz de covarianza de los puntos centrados,

$$(\tilde{V}^k, \tilde{\Phi}(t)) = \sum_{i=1}^M \tilde{\alpha}_i^k(\tilde{\Phi}(x_i), \tilde{\Phi}(t))\tag{B.4}$$

Consideremos un conjunto de punto de prueba  $t_1, \dots, t_L$ , y definamos dos matrices de tamaño  $L \times M$  como:

$$\begin{aligned}K_{ij}^{test} &= (\Phi(t_i), \Phi(x_j)) \\ \tilde{K}_{ij}^{test} &= ((\Phi(t_i) - \frac{1}{M} \sum_{m=1}^M \Phi(x_m)), (\Phi(x_j) - \frac{1}{M} \sum_{n=1}^M \Phi(x_n)))\end{aligned}$$

Como en la ecuación B.3, expresamos  $\tilde{K}^{test}$  en términos de  $K^{test}$ , de donde se tiene que  $\tilde{K}^{test} = K^{test} - 1_M' K - K^{test} 1_M + 1_M' K 1_M$ , donde la matriz  $1_M'$  es de tamaño  $L \times M$

## Anexo C

# El problema de valores propios en el espacio de coeficientes de expansión

Siendo simétrica,  $K$  tiene una base ortonormal de vectores propios  $(\beta^i)_i$  con sus valores propios correspondientes  $u_i$ . Entonces, para todo  $i$ , tenemos  $K\beta^i = u_i\beta^i$  con  $i = 1, \dots, M$ .

Para comprender mejor la relación entre las ecuaciones 2.21 y 2.22, procedemos de la siguiente manera. Primero, supongamos que  $\lambda, \alpha$  satisfacen la ecuación 2.21. Ahora, podemos expresar  $\alpha$  en términos de la base de vectores propios de  $K$  como  $\alpha = \sum_{i=1}^M a_i\beta^i$ , reemplazando esto en la ecuación 2.21 tenemos,

$$M\lambda \sum_i a_i u_i \beta^i = \sum_i a_i u_i^2 \beta^i$$

que es equivalente a, para todo  $i = 1, \dots, M$ ,

$$M\lambda a_i u_i = a_i u_i^2.$$

esto a su vez significa que para todo  $i = 1, \dots, M$ ,

$$M\lambda = u_i \quad \text{o} \quad a_i = 0 \quad \text{o} \quad u_i = 0. \quad (\text{C.1})$$

<sup>1</sup> Ahora, asumimos que  $\lambda$  y  $\alpha$  satisfacen la ecuación 2.22 para realizar una derivación similar. En este caso se tiene que la ecuación 2.22 es equivalente a,

$$M\lambda \sum_i a_i \beta^i = \sum_i a_i u_i \beta^i$$

---

<sup>1</sup>Notemos que los  $\text{o}$  no son exclusivos.

esto es, para todo  $i = 1, \dots, M$ ,

$$M\lambda = u_i \quad \text{o} \quad a_i = 0. \quad (\text{C.2})$$

Comparando las ecuaciones C.1 y C.2, vemos que todas las soluciones de esta última satisfacen a la primera. Sin embargo, no dan su conjunto completo de soluciones: dada una solución de ecuación 2.22, siempre podemos añadir múltiples vectores propios de  $K$  con valor propio 0 y satisfacer la ecuación 2.21 con el mismo valor propio.

Esto significa que existen soluciones de la ecuación 2.21 que pertenecen a diferentes valores propios pero que no son ortogonales en el espacio de  $\alpha^k$ . Sin embargo, eso no implica que los vectores propios de  $\tilde{C}$  en  $F$  no son ortogonales.

En efecto, si  $\alpha$  es un vector propio de  $K$  con valor propio 0, entonces el vector correspondiente  $\sum_i \alpha_i \Psi(x_i)$  es ortogonal a todos los vectores del espacio de  $\Psi(x_j)$  en  $F$ , pues

$$(\Psi(x_j), \sum_i \alpha_i \Psi(x_i)) = (K\alpha)_j = 0, \text{ para todo } j,$$

lo que significa que  $\sum_i \alpha_i \Psi(x_i) = 0$ .

Así, la diferencia anterior entre las soluciones de las ecuaciones 2.21 y 2.22 es irrelevante, pues estamos interesados en los vectores en  $F$  en lugar de los vectores en el espacio de expansión de coeficientes de la ecuación 2.17. Por lo tanto, solo necesitamos diagonalizar  $K$  para encontrar todas las soluciones relevantes de la ecuación 2.21

# Anexo D

## VARIABLES CATEGORIZADAS

En esta sección se presenta un detalle de categorización variables seleccionadas y las variables creadas, antes de aplicar el método KNPCA.

### D.1. Variables originales

<b>HVI001</b>		<b>HVI019</b>	
No tiene	1	No aplica	1
Letrina	2	No	2
Inodoro y pozo ciego	3	Si	3
Inodoro y pozo séptico	4	<b>HVI020</b>	
Inodoro y alcantarillado	5	No aplica	1
<b>HVI004</b>		No	2
No	1	Si	3
Si	2	<b>HVI024</b>	
<b>HVI006</b>		Otro tipo	1
No	1	Cuarto en casa de inquilinato	2
Si	2	Mediagua	3
<b>HVI007</b>		Departamento en casa o edificio	4
No	1	Casa/Villa	5
Si	2	<b>HVI025</b>	
<b>HVI008</b>		Otro	1
No	1	Arrendada	2
Si	2	Propia	3
<b>HVI014</b>		<b>HVI028</b>	
Otros	1	No aplica	1
La entierran	2	No	2
Servicio municipal	3	Si	3
<b>HVI016</b>		<b>HVI033</b>	
No	1	No	1
Si	2	Si	2

<b>HVI029</b>		<b>HVI030</b>	
No tiene acceso/No aplica	1	Otra vía de acceso	1
Veredas	2	Lastrado/calle de tierra	2
Parques	3	Empedrado	3
Plaza o plazoleta	4	Carretera, calle pavimentada o adoquinada	4
Malecón o bulevar	5	<b>HVI036</b>	
<b>HVI034</b>		Si	1
Si	1	No	2
No	2	<b>HVI039</b>	
<b>HVI037</b>		Viven mal	1
Si	1	Viven más o menos bien	2
No	2	Viven bien	3
<b>HVI041</b>		<b>HVI042</b>	
Se ven obligados a gastar ahorros	1	Muy pobre	1
Se ven obligados a endeudarse	2	Pobre	2
Apenas logran equilibrar ingresos y gastos	3	Más o menos pobre	3
Logran ahorrar dinero	4	No pobre	4
<b>HVI043</b>		<b>HVI044</b>	
Empeoró	1	No	1
Está igual	2	Si	2
Mejóro	3	<b>HVI045</b>	
<b>HVI060</b>		No	1
Ninguno	1	Si	2
1 perceptor	2	<b>HJH001</b>	
2 perceptores	3	Mujer	1
3 perceptores	4	Hombre	2
Más de 4 perceptores	5	<b>HJH003</b>	
<b>HJH006</b>		10 a 24 años	5
Ninguno	1	25 a 34 años	5
Primaria	2	35 a 44 años	4
Secundaria	3	45 a 54 años	3
Superior	4	55 a 64 años	2
<b>HJH009</b>		65 años y más	1
Ninguno	1	<b>HJH011</b>	
IESS, Seguro Voluntario/Campesino	2	Trabajadores hogar/no hogar no remunerado	1
IESS, Seguro General	3	Trabajadores Independientes	2
Seguro de Salud Privado	4	Trabajadores Dependientes	3

## D.2. Variables creadas

Las variables que se detallan a continuación se crearon a partir de las variables presentes en la encuesta de condiciones de vida, con el objetivo de tener una mejor categorización de los datos.

Para la variable *HVI017* hicimos una combinación de los atributos de las variables *De donde se obtiene el agua* y *Tratamiento que se da al agua* pues el tratamiento será una necesidad o un lujo dependiendo de donde se obtiene el agua, además las herramientas utilizadas para la purificación del agua no serán las mismas en todos los casos.

También se tomó en cuenta condiciones que no constan en la ECV, como el costo

HVI017	
Otra fuente - Ninguno, tal como la obtiene	1
Otra fuente - La hierven	2
Otra fuente - Otros	3
Red pública - Ninguno, tal como la obtiene	4
Red pública - La hierven	5
Red pública - Otros	6

**Cuadro D.1:** La variable original se reemplazó por la unión de *De donde se obtiene el agua* y *Tratamiento que se da al agua*.

HVI040	
No cubre canasta vital	1
Cubre canasta vital	2
Cubre canasta básica	3

**Cuadro D.2:** La variable original se reemplazó por la condición de si en ese hogar, se cubre la canasta básica o canasta vital per cápita.

de la canasta básica y la canasta vital. Para la variable *HVI040*, utilizamos los montos per cápita de cada una de las canastas y con la variable *Monto considerado para vivir bien* clasificamos los hogares en torno a si cubren o no alguna de las canastas.

Es importante notar que la variable *Monto considerado para vivir bien* no son los ingresos de los hogares, sino que es el monto que consideran necesario para garantizar su bienestar.

# Anexo E

## Implementación de los algoritmos en Rstudio

### E.1. Algoritmo árbol de mínima expansión

```
dendrite <- function(D)
{
  D <- as.matrix(D); M <- mst(D); lw=nrow(M); den <- matrix(0,lw-1,3); k <- 0;
  for (i in 1:(lw-1))
    for (j in (i+1):lw)
      {
        if (M[i,j]==1)
          {
            k <- k+1
            den[k,1] <- i
            den[k,2] <- j
            den[k,3] <- D[i,j]
          }
      }
  par <- matrix(0,5,1)
  par[1,1] <- mean(den[,3])
  par[2,1] <- sd(den[,3])
  par[3,1] <- par[1,1]+par[2,1]
  par[4,1] <- par[1,1]+2*par[2,1]
  par[5,1] <- par[1,1]+3*par[2,1]
  wynik <- list(den=den,par=par)
  return(wynik)
}
```

## E.2. KPCA - Caso sin ponderación geográfica

```
D <- base
p <- ncol(D[[1]])
T<-nrow(D[[1]])
N <- length(D)
F <- matrix(0,nrow=N,ncol=N)
for (i in 1:N)
  for (j in 1:N)
  {
    m <- t(as.matrix(D[[i]]-D[[j]])) %*% (as.matrix(D[[i]]-D[[j]]))
    F[i,j] <- sqrt(sum(diag(m)))
  }
K1 <- matrix(0,nrow=p,ncol=p)
K1 <- exp(-F^2/(sum(F^2)/(N*(N-1))))
I <- diag(N)
J <- matrix(1,nrow=N,ncol=N)
J <- (1/N)*J
P <- I-J
K1 <- P%*%K1%*%P
u1 <- eigen(K1)
uK1 <- u1$values
a<-5
m1 <- matrix(0,nrow=N,ncol=a);
for (k in 1:a)
{
  V<- u1$vectors[,k]
  for (i in 1:N)
  {
    suma <- 0
    for (j in 1:N)
    {
      suma <- suma+V[j]*K1[i,j]
    }
    m1[i,k] <- suma
  }
}
```

### E.3. KPCA - Caso con ponderación geográfica

```
pesos<-plusvalia*NumeroViviendas/NumeroHogares
DK=as.numeric(pesos); DK <- DK/sum(DK);
D <- base
p <- ncol(D[[1]]); T<-nrow(D[[1]]); N <- length(D); X<-D;
for (i in 1:N)
{
  X[[i]] <- X[[i]]*DK[i]
}
F <- matrix(0,nrow=N,ncol=N)
for (i in 1:N)
  for (j in 1:N)
  {
    m <- t(as.matrix(X[[i]]-X[[j]]))%*%(as.matrix(X[[i]]-X[[j]]))
    F[i,j] <- sqrt(sum(diag(m)))
  }
K1 <- matrix(0,nrow=N,ncol=N)
K1 <- exp(-F^2/(sum(F^2)/(N*(N-1))))
I <- diag(N)
J <- matrix(1,nrow=N,ncol=N)
J <- (1/N)*J
P <- I-J
K1 <- P%*%K1%*%P
u1 <- eigen(K1)
uK1 <- u1$values
m <- matrix(0,nrow=N,ncol=a)
for (k in 1:a)
{
  V<- u1$vectors[,k]
  for (i in 1:N)
  {
    suma <- 0
    for (j in 1:N)
    {
      suma <- suma+V[j]*K1[i,j]
    }
    m[i,k] <- suma
  }
}
}
```

# Bibliografía

- [1] Alzate, C. A Weighted Kernel PCA Formulation with Out-of-Sample Extensions for Spectral Clustering Methods. *International Joint Conference on Neural Networks*, págs. 138–144, 2006.
- [2] Bastidas Medina, J. C. La afiliación por trabajo no remunerado del hogar frente al principio de los derechos constitucionales a la igualdad, a la salud ya la seguridad social. B.S. thesis, 2015.
- [3] Bauer, R. A. *Social indicators*. 1966.
- [4] BID, BIRF, y CEPAL. Indicadores sobre el desarrollo social. 2000.
- [5] Cabrera Arias, M. Desigualdad territorial y exclusión social. claves para una lectura crítica de ciudad de panamá. *Revista Vivienda y Ciudad*, 6, 2019.
- [6] Calo, S. y Vinicio, K. Empleo informal y depresión en el ecuador: Resultados de la vi encuesta de condiciones de vida. 2019.
- [7] Camacho, G. La violencia de género contra las mujeres en el ecuador: Análisis de los resultados de la encuesta nacional sobre relaciones familiares y violencia de género contra las mujeres. 2014.
- [8] Carrión, F. y Pinto, J. P. Producción y organización espacial de viejas y ‘nuevas’ desigualdades en quito. *Andamios*, 16(39):101–125, 2019.
- [9] Cecchini, S. *Indicadores sociales en América Latina y el Caribe*. Cepal, 2005.
- [10] CEPAL. La hora de la igualdad: brechas por cerrar, caminos por abrir. *CEPAL*, 2010.
- [11] CEPAL. La matriz de la desigualdad social en América Latina. *CEPAL*, 2016.
- [12] CEPAL. Desarrollo y migración: Desafíos y oportunidades en los países del norte de Centroamérica. *CEPAL*, 2019.

- [13] Charlton, M., Brunsdon, C., Demšar, U., Harris, P., y Fotheringham, S. Principal components analysis: from global to local. En *13 th AGILE International Conference on Geographic Information Science 2010*, 2010.
- [14] Cliff, A. Spatial autocorrelation. 1973.
- [15] Colmenarejo, R. Enfoque de capacidades y sostenibilidad. aportaciones de amartya sen y martha nussbaum. *Ideas y valores*, 65(160):121–149, 2016.
- [16] Consejo Metropolitano de Quito, . Ordenanza 152. 2011.
- [17] Cordon, J. A. F. Desigualdades, una perspectiva demográfica. *Edita*, 2013.
- [18] Demšar, U. Principal Component Analysis on Spatial Data: An overview. *Annals of the Association of American Geographers*, 103:106–128, 2013.
- [19] Domínguez Serrano, M., Blancas Peral, F. J., Guerrero Casas, F. M., y González Lozano, M. Una revisión crítica para la construcción de indicadores sintéticos. *Revista de Métodos Cuantitativos para la Economía y la Empresa*, 11:41–70, 2011.
- [20] Formichella, M. M. Análisis del concepto de equidad educativa a la luz del enfoque de las capacidades de amartya sen. *Revista educación*, 35(1):1–36, 2011.
- [21] Galiano Maritan, G. y Bravo Placeres, I. La seguridad social en ecuador: breves apuntes sobre sus deficiencias y beneficios the social security in ecuador: brief notes about their deficiencias and benefits.
- [22] Greenacre, M. *La práctica del análisis de correspondencias*. 01 2008.
- [23] Guttman, L. The quantification of a class of attributes: A theory and method. 1941.
- [24] Górnjak, J. Identification of transport accessibility of Polish cities based on their transport infrastructures. *Studia Ekonomiczne. Zeszyty Naukowe UE w Katowicach*, 249:145–154, 2015.
- [25] Hidalgo, F. y Rodríguez, I. V. El respeto de los principios laborales y el acceso a la justicia en los procesos de empleo público y seguridad social tramitados ante la jurisdicción contencioso administrativa. *Revista Espiga*, 14(30):65–74, 2015.
- [26] Hoffmann, H., Schaal, S., y Vijayakumar, S. Local Dimensionality Reduction for Non-Parametric Regression. *Neural Processing Letters*, 29:109–131, 04 2009.

- [27] Instituto de fomento, empleo y formación, A. d. C. Diseño de indicadores de desigualdad social en las ciudades.
- [28] Instituto Nacional de Estadística y Censos, . Indicadores de pobreza y desigualdad publicados en 2016. 2016.
- [29] Instituto Nacional de Estadísticas y Censos, . Información cartográfica. [urlhttps://www.ecuadorencifras.gob.ec/category/cartografia-2/](https://www.ecuadorencifras.gob.ec/category/cartografia-2/), 2010.
- [30] Instituto Nacional de Estadísticas y Censos, I. *Compendio de Resultados Encuesta Condiciones de Vida ECV, Sexta Ronda 2015*. 2015.
- [31] Instituto Nacional de Estadísticas y Censos, I. Metodología de la Encuesta de Condiciones de Vida ECV. 2015.
- [32] Johan A.K. Suykens, Marco Signoretto, A. A. *Regularization, Optimization, Kernels, and Support Vector Machines*. Chapman and Hall/CRC, 2015.
- [33] Krzyśko, M. L. Nonlinear Principal Component Analysis for Geographically Weighted Temporal-spatial Data. *Folia Oeconomica*, págs. 169–181, 2018.
- [34] Lopez Menendez, A. *Indicadores económicos de la desigualdad y la pobreza*, págs. 77–105. 07 2012.
- [35] López, L. Administración zonal norte de quito “eugenio espejo”. 2014.
- [36] Max-Neef, M. A., Elizalde, A., y Hopenhayn, M. *Desarrollo a escala humana: conceptos, aplicaciones y algunas reflexiones*, volume 66. Icaria Editorial, 2006.
- [37] Medina, F. y Galván, M. Descomposición del coeficiente de gini por fuentes de ingreso: Evidencia empírica para américa latina 1999-2005. 2008.
- [38] Melamed y Samman. Equity, Inequality and Human Development in a Post-2015 Framework. pág. 4.
- [39] Mideros, A. Ecuador: Definición y medición multidimensional de la pobreza, 2006-2010. *Revista Cepal*, 2012.
- [40] Molinas, J., Paes de Barro, R., Saavedra, J., y Giugale, M. "do our children have a chance?"the 2010 human opportunity report for latin america and the caribbean. *The World Bank*, 2012.
- [41] Montes, B. A., Mancero, X., y Rojas, V. T. Herramientas para el análisis de las desigualdades y del efecto redistributivo de las políticas públicas. 2018.

- [42] Municipio, D. M. Q. Diagnóstico del Territorio del Distrito Metropolitano de Quito. 2014.
- [43] Murillo, F. H. S. Spatial Variability Analysis of Quality of Life and Its Determinants: A Case Study of Medellín, Colombia. *Social Indicators Research: An International and Interdisciplinary Journal for Quality-of-Life Measurement*, Springer, págs. 1233–1256, 2019.
- [44] Newman, G. D. El razonamiento inductivo y deductivo dentro del proceso investigativo en ciencias experimentales y sociales. *Laurus*, 12(Ext):180–205, 2006.
- [45] Nicolini, R. Desigualdades territoriales y estrategias de crecimiento. *Foreign Affairs*, 13, 2013.
- [46] Nussbaum, M. Capacidades como titulaciones fundamentales: Sen y la justicia social. *Books*, 1, 2005.
- [47] Organización de Estados Americanos, O. *Desigualdad e inclusión social en las Américas : 14 ensayos*.
- [48] Organización Internacional del Trabajo, . Encuesta de percepción de usuarios y beneficiarios del iess sobre el desempeño del instituto en términos de gobernanza, cobertura, calidad y sostenibilidad.
- [49] Pérez Torres, F. J. Economía política y métodos de avalúo del suelo. *Equidad y Desarrollo*, 1(24):53–95, 2015.
- [50] Programa de las Naciones Unidas para el Desarrollo, P. Informe sobre desarrollo humano 2011. 2011.
- [51] Pun-Cheng, L. *Distance Decay*. 02 2016.
- [52] Rodríguez, C. *América latina en la segunda mitad del siglo XX*. 2013.
- [53] Samaniego Ponce, P. Evolución de la pobreza y la desigualdad en quito. *Questions Urbano Regionales*.
- [54] Schölkopf, B. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10:1299 – 1319, 1998.
- [55] Secretaría de territorio, hábitat y vivienda DMQ, D. Plan Metropolitano de Ordenamiento Territorial 2012-2022.

- [56] Segura, R. El espacio urbano y la (re) producción de desigualdades sociales.: Desacoples entre distribución del ingreso y patrones de urbanización en ciudades latinoamericanas. 2014.
- [57] Sen, A. *La agenda ética pendiente de América Latina*. Idb, 2005.
- [58] Skočaj, D., Leonardis, A., y Bischof, H. Weighted and robust learning of subspace representations. *Pattern Recognition*, 40:1556–1569, 05 2007.
- [59] Tezanos Vázquez, S., Quiñones Montellano, A., Gutierrez Sobrao, D., y Madrueño Aguilar, R. *Desarrollo humano, pobreza y desigualdades*. 2013.
- [60] Torrecilla, F. J. M. y Castilla, R. H. Hacia un concepto de justicia social. *REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 9(4):7–23, 2011.
- [61] Vapnik, V. *The Nature of Statistical Learning Theory*. Springer, 1995.
- [62] Vizcarra Bordi, I. Entre las desigualdades de género: un lugar para las mujeres pobres en la seguridad alimentaria y el combate al hambre. *Argumentos (México, DF)*, 21(57):141–173, 2008.
- [63] Ziccardi, A. Pobreza y exclusión social en las ciudades del siglo xxi. *Procesos de urbanización de la pobreza y nuevas formas de exclusión social. Los retos de las políticas sociales de las ciudades latinoamericanas del siglo XXI*, págs. 9–33, 2008.