

UNIVERSIDAD DE EL SALVADOR
FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA
ESCUELA DE MATEMÁTICA
DEPARTAMENTO DE ESTADÍSTICA

TRABAJO DE INVESTIGACIÓN TITULADO:

**“APLICACIÓN DE MODELOS MULTIVARIANTES EN LA CONSTRUCCIÓN
DE INDICADORES SINTÉTICOS DE CALIDAD AMBIENTAL PARA ZONAS
URBANAS DE EL SALVADOR”**



Universidad de El Salvador

Hacia la libertad por la cultura

ASESOR:

DR. JOSÉ NERYS FUNES TORRES.

ESTUDIANTES:

CLAVEL QUIJADA, DONY HENRY.

NAVES SÁNCHEZ, FRANCISCO.

CARNET

CQ02005

NS99003

UNIVERSIDAD DE EL SALVADOR.

RECTOR: ING. RUFINO ANTONIO QUEZADA SÁNCHEZ.

SECRETARIO GENERAL: LIC. DOUGLAS VLADIMIR ALFARO CHÁVEZ.

FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA.

DECANO: DR. RAFAEL ANTONIO GÓMEZ ESCOTO.

SECRETARIA: LIC. MARÍA TRINIDAD TRIGUERROS DE CASTRO.

ESCUELA DE MATEMÁTICA.

DIRECTOR: ING. CARLOS MAURICIO CANJURA LINARES.

TRABAJO DE GRADUACIÓN APROBADO POR:

ASESOR: DR. JOSÉ NERYS FUNES TORRES.

AGRADECIMIENTOS.

Dony Henry Clavel Quijada:

Agradezco muy humildemente al dueño del conocimiento: Dios todopoderoso, que me permitió culminar mis estudios universitarios, a Jesucristo que me da la fortaleza en todo momento, sobretodo en aquellos más difíciles, pues es él quien realiza por mí lo que yo no puedo; a mis padres por el apoyo económico, moral y emocional que durante toda mi vida he tenido de ellos, así como de mis hermanos y amigos, a todos ellos por su confianza y comprensión hacia mí. De manera muy especial quiero agradecer a un amigo que ya no está entre nosotros, Pablo Eusabio Márquez, pero que significó para mí un ejemplo de vida a seguir, pues de él aprendí importantes lecciones de vida; por último agradezco a nuestro asesor por su importante apoyo y colaboración durante todo el proceso hasta poder culminarlo.

Dedico este trabajo a Sabas Clavel Fuentes y Ana Margarita Quijada, mis padres de quienes estoy muy orgulloso y a quienes quiero mucho, además porque ellos me dieron los ánimos para poder continuar cuando ya no tenía motivación y sobretodo porque me han demostrado que creen en mí, además por su paciencia, amor y dedicación para cada uno de sus hijos, por habernos enseñado con su ejemplo el temor a Dios y la calidad de una vida cristiana y mariana (amor a la Virgen Maria).

Francisco Naves Sánchez:

Agradezco de gran manera a nuestra “Alma Mater” (Universidad de El Salvador) la más grande escuela política y académica, que me ha permitido desenvolverme a lo largo de toda la carrera y poder conocer a profesores, amigos y a un descubrimiento trascendental que es la “Memoria Histórica” hilo conductor del desarrollo histórico de la sociedad, que me ha convertido en un sujeto político con identidad.

También agradezco a Carmen Sánchez y Jesús Naves Domínguez, mis padres, que han sido guías fantásticos en mi formación. Todos mis agradecimientos se engloban a un poder superior llamado “Dios”.

ÍNDICE

Contenido	Página
INTRODUCCIÓN.....	i
PLANTEAMIENTO DEL PROBLEMA.....	iii
JUSTIFICACIÓN.....	iv
OBJETIVOS.....	vi
Objetivo General.....	vi
Objetivos Específicos.....	vi

CAPÍTULO 1: MODELOS MULTIVARIANTES COMPONENTES PRINCIPALES

1.1 Prólogo.....	1
1.1.1 Reseña histórica.....	2
1.1.2 Métodos de análisis multivariante.....	3
1.1.3 Descripción general de las técnicas de análisis multivariante.....	4
1.1.4 Análisis de componentes principales.....	5
1.1.5 Utilidades del análisis de componentes principales.....	6
1.1.6 Objetivos del análisis de componentes principales.....	7
1.2 Definición de componentes principales.....	8
1.2.1 Enfoque descriptivo.....	9
1.2.2 Enfoque estadístico.....	11
1.2.3 Enfoque geométrico.....	12
1.2.4 Cálculo de los componentes.....	16
1.2.4.1 Cálculo del primer componente.....	16
1.2.4.2 Cálculo del segundo componente.....	19
1.2.5 Generalización.....	23
1.2.6 Varianza de las componentes.....	25
1.2.7 Propiedades de los componentes.....	27
1.2.8 Análisis normado o con correlaciones.....	30
1.3 Aplicación del método ACP a partir de la matriz de correlaciones.....	32

1.4 Estructura factorial de las componentes principales.....	34
1.5 Puntuaciones o medición de las componentes.....	35
1.6 Calificación de componentes principales.....	36
1.7 Interpretación de los componentes.....	37
1.8 Selección del número de componentes.....	38
1.8.1 Criterio de la media aritmética.....	38
1.8.2 El gráfico de sedimentación.....	39
1.8.3 Variabilidad explicada.....	40
1.9 Representación gráfica.....	40
1.10 Datos atípicos.....	41
1.11 Matriz de cargas.....	42
1.12 Rotación de los componentes.....	44
1.13 Ejemplo usando SPSS.....	46

CAPÍTULO 2: ÍNDICES SINTÉTICOS

2.1 Índices sintéticos de calidad ambiental (ISCA).....	58
2.1.1 Antecedentes.....	59
2.2 Aspectos técnicos para la construcción de los indicadores sintéticos.....	61
2.2.1 Criterios generales para la selección de indicadores ambientales simple y sintéticos.....	62
2.2.2 Supuestos relevantes para la formulación del índice.....	65
2.3 Sistema de indicadores ambientales urbanos.....	67
2.3.1 Marco de referencia de indicadores de calidad ambiental.....	67
2.3.2 Modelo de indicador sintético.....	68
2.4 Estimación del índice resultado del ACP.....	73
2.5 Metodología para construir índices sintéticos a partir del análisis multivariante....	74

CAPÍTULO 3: APLICACIÓN DEL MODELO, CONSTRUCCIÓN DEL ÍNDICE

3.1 Aplicación del modelo de indicador sintético a las zonas urbanas de el salvador...76	
3.1.1 Descripción general del área geográfica del estudio.....	76
3.1.2 Estructura de la población.....	77
3.2 Descripción de la base de datos.....	78
3.3 Representación espacial de los indicadores operativos.....	81

3.4 Procedimientos para la construcción del indicador sintético de calidad ambiental..	86
3.5 Construcción del índice de flujo urbano (IFLU).....	87
3.6 Construcción del índice de medio ambiente urbano (IMAU).....	96
3.7 Construcción del índice sintético de calidad ambiental (ISCA).....	101
ANEXOS.....	114
BIBLIOGRAFÍA.....	144

INTRODUCCIÓN

Uno de los actuales retos en nuestro país es la generación de información estadísticas relevante y oportuna para la toma de decisiones de las políticas públicas. Los indicadores estadísticos, constituyen una de las herramientas indispensables para el logro de estas acciones, y una de las tareas para la Dirección General de Estadísticas y Censos (DIGESTYC) y el Ministerio de Medio Ambiente y Recursos Naturales (MARN) es fomentar la cultura estadística para su correcto uso y aplicación. Es importante conocer lo que son los indicadores, sus características, utilidad, tipos y limitaciones para, enmarcar la importancia que tienen y los retos que imponen a los generadores de estadísticas del país.

No existe una definición específica por parte de algún organismo nacional o internacional para los indicadores, sólo algunas referencias que los describen, como: “Herramientas para clarificar y definir, de forma más precisa, objetivos e impactos”, “son medidas verificables de cambio o resultado, diseñadas para contar con un estándar contra el cual evaluar, estimar o demostrar el progreso”, “con respecto a metas establecidas, facilitan el reparto de insumos, produciendo productos y alcanzando objetivos” y muchas más. Una de las definiciones más utilizadas por diferentes organismos y autores es: *“Los indicadores sociales, económicos, ambientales,... son estadísticas, serie estadística o cualquier forma de indicación que nos facilita estudiar dónde estamos y hacia dónde nos dirigimos con respecto a determinados objetivos y metas, así como evaluar programas específicos y determinar su impacto”*. Si bien los indicadores pueden ser cualitativos o cuantitativos, en este trabajo nos referimos únicamente a los segundos. Por tanto, los indicadores son elementales para evaluar, dar seguimiento y predecir tendencias de la situación de un país, un estado o una región en lo referente a su economía, sociedad, desarrollo humano, etc., así como para valorar el desempeño institucional encaminado a lograr las metas y objetivos fijados en cada uno de los ámbitos de acción de los programas de gobierno.

Hemos presentado hasta aquí una breve introducción general sobre lo que son los indicadores, centraremos ahora nuestra atención a un tipo de ellos, los cuales son la causa de nuestra investigación, los –indicadores ambientales–, desde un punto de vista urbano.

La construcción de un sistema de indicadores simples o índices sintéticos, demanda la síntesis de información que en muchas ocasiones no está disponible.

Al respecto, algunos autores sostienen que generalmente el índice construido no da explicación de todos los factores que pueden describir una *variable latente*¹; sin embargo, son en todo momento una aproximación a ella. Según otros autores, los *índices* son una herramienta cuantitativa que simplifica a través de modelos matemáticos los atributos y pesos de múltiples variables, con la intención de proporcionar una explicación más amplia de un recurso o el atributo a evaluar y gestionar. Lo importante es que el índice garantice ser un *índice consistente*.

En este anteproyecto se presenta la metodología para la construcción de un indicador sintético ambiental, utilizando diferentes técnicas de Análisis Multivariante, como Análisis de Componentes Principales, Análisis de Factores, entre otros. Además se describen los indicadores ambientales simples que servirán de base para la construcción de un indicador sintético ambiental para las zonas urbanas de El Salvador.

¹ **Variable latente:** En [las estadísticas](#) , las **variables latentes** (a diferencia de [las variables observables](#)), son [variables](#) que no están bajo observación directa, sino que son inferidos (a través de un [modelo matemático](#)) a partir de otras variables que se observan (medidas directamente) Los modelos matemáticos que tratan de explicar las variables observadas en términos de variables latentes son llamados *modelos de variables latentes*.

PLANTEAMIENTO DEL PROBLEMA

La bibliografía sobre indicadores sintéticos enfatiza que la selección del conjunto de indicadores apropiados no es una tarea fácil, dado que ello demanda el entendimiento de cómo funciona el sistema o fenómeno que se quiere explicar, y esto no siempre es posible cuando se trabaja con el medio ambiente. Por ello, la selección de indicadores obedece no sólo a la interpretación que el estadista o el investigador y la sociedad hagan de una realidad, sino también a la disponibilidad de la información en un marco analítico que la interprete.

Desde las ciencias sociales, los científicos buscan definir modelos analíticos que delimiten la explicación de *variables latentes* como la *calidad ambiental*, el *desarrollo sostenible*, el *bienestar social* o cualquier otra. Por ello, un modelo puede o no ser apropiado, en función de su poder de explicación de esa realidad y no tanto por la mayor representación de la complejidad de la misma.

Si bien los problemas ambientales urbanos han sido tratados como componentes temáticos en mucho de los sistemas de indicadores de desarrollo sostenible propuestos, la generación de indicadores e índices para medir los problemas ambientales y la sostenibilidad ambiental urbana son más recientes, debido a las restricciones de datos para desarrollar sistemas de indicadores a los niveles más desagregados (para nuestro caso, desagregado por municipio). Abordar metodologías a niveles más desagregados consideramos, es una línea de investigación relevante, dado que las grandes concentraciones urbanas son sistemas complejos en donde las situaciones ambientales (positivas y negativas) pueden presentar diferencias significativas entre zonas de una misma ciudad. Ello demanda la construcción de información adecuada que permita, a los tomadores de decisiones, diseñar y priorizar la ejecución de estrategias ambientales que impacten de forma diferencial las distintas zonas en una ciudad.

El Ministerio de Medio Ambiente y Recursos Naturales de El Salvador, como responsable de la gestión ambiental y tomando en consideración que el medio ambiente está compuesto por diferentes elementos interrelacionados en constante cambio, que inciden directamente en su comportamiento y desarrollo sostenible, presenta a la población "El Medio Ambiente en Cifras" que contiene un conjunto de datos estadísticos e información ambiental de los últimos años. Esta información es evidentemente descriptiva que permite posicionarnos de cómo estamos en el tema de medio ambiente, sin embargo, no se construyen modelos que identifiquen la trayectoria hacia dónde vamos. Es por ello, que en este trabajo se presenta una estructura metodológica para la construcción de un sistema de indicadores ambientales a nivel urbano y se propone un modelo para construir el **Índice Sintético de Calidad Ambiental** (ISCA) de acuerdo a la división político-administrativa de El Salvador por municipio. Este índice es considerado como una **variable latente** explicada por un conjunto de indicadores simples que lo determinan.

JUSTIFICACIÓN

Antecedentes

Desde 1987 se viene desarrollando planes de acción ambiental a nivel mundial, pues para ese año se conoce del informe dirigido por las Naciones Unidas (UNCED/CNUMA) denominado “Nuestro Futuro Común”, en el cual se introduce el concepto de desarrollo sostenible como “el desarrollo que satisface las necesidades presentes sin comprometer la habilidad de las generaciones futuras para satisfacer sus propias necesidades”; para 1992 Naciones Unidas desarrolla la conferencia sobre Ambiente y Desarrollo donde presenta el problema ambiental a nivel mundial, lo cual contribuye significativamente a un nuevo enfoque del concepto de desarrollo sostenible y a sus políticas, uno de los resultados de la conferencia es el documento oficial Agenda Local 21, y su artículo 40 se refiere a la necesidad de disponer de información adecuada para el monitoreo, evaluación y seguimiento de la sostenibilidad del desarrollo. De aquí se estimula la necesidad de indicadores ambientales. También para 1987 en Europa se desarrollo el 4to Programa de Acción Ambiental dirigido por la Unión Europea (EU), en el cual se define el Plan de Acción Ambiental (1987 - 1992); luego para 1990 la Comisión Europea presenta el Libro Verde Sobre Ambiente Urbano, el cual fue un hito sobre ambiente urbano, porque constituyó el primer camino hacia el debate e investigación sobre ambiente urbano y calidad de vida en la Unión Europea, se describen los principales problemas ambientales urbanos, demandando la necesidad de integrar la planificación y gestión urbana de los problemas ambientales. De aquí surge la necesidad de monitorear y evaluar las condiciones ambientales de las ciudades.

Es importante destacar la evolución europea en materia ambiental que se ha venido desarrollando a través de los años, de la cual se han venido elaborando importantes documentos como los mencionados, además de otros como el Tratado sobre la Unión Europea, el 5to (1992) y 6to (2001) Programa de Acción Ambiental, Carta Alborg (1994), Plan de Acción de Lisboa (1996), Foro Urbano de Viena (1998), Campaña de Ciudades Sostenibles “Hannover” (2000); en algunos de los cuales se destaca la importancia de construir indicadores para la toma de decisiones.

La *Organización para la Cooperación y el Desarrollo Económico (OCDE)*, ha definido trece áreas de indicadores, entre las cuales una de ellas se refiere al fenómeno ambiental urbano como área de interés especial, otro programa relevante a nivel mundial es el de la Organización Mundial para la Salud (OMS), que establece un conjunto de indicadores de ciudades saludables, dentro del Programa Salud para Todos en 2000. En España, además del sistema de indicadores ambientales urbanos diseñados por el Ministerio del Medio Ambiente (MMA, 2000), se destacan los sistemas de indicadores desarrollados para el nivel de regiones en Cataluña, Andalucía y el País Vasco, entre otras. Los indicadores desarrollados responden a las condiciones específicas de las regiones, con fines de comparar y diseñar políticas a este nivel.

Estados Unidos presentó una de las iniciativas más destacables a nivel mundial, la “Seattle Ciudad Sostenible”, 1992, la cual fue caracterizada por un proceso de participación con amplia convocatoria comunitaria, ONGs, universidades, académicos, etc. Que llevó a la definición de un conjunto de **indicadores simples** dirigidos a tres áreas temáticas: economía, sociedad y medio ambiente a nivel de la ciudad.

En Latinoamérica se destaca la experiencia de Chile, México y Costa Rica como procesos avanzados en la definición de sistemas de indicadores de desarrollo sostenible. Para el caso de Colombia, son dos las experiencias más relevantes, más que por su trascendencia internacional, porque están relacionadas con el modelo de indicadores que se presenta en nuestra investigación. Una de las experiencias es el sistema de indicadores ambientales para Colombia, realizado por el SISA en 2002, en el que se define un conjunto de 34 indicadores ambientales para su desarrollo sostenible. En la construcción del ISCA en Colombia, se cuenta con 38 indicadores simples agrupados en 10 áreas temáticas. De los indicadores simples derivados se obtienen los indicadores operativos del sistema. Se obtuvo resultados muy relevantes e importantes para dicha ciudad, como concluir que las condiciones de calidad ambiental en una ciudad son heterogéneas. Su importancia operativa indica que se construyó un importante instrumento para la política pública en cuestión de medio ambiente; trabajos similares, han sido realizados en España y otros países desarrollados.

Bajo estas premisas, nos damos cuenta lo importante que es para la toma de decisiones en pro del desarrollo de un país, la construcción de indicadores e índices sintéticos en general y para nuestro caso índices medioambientales. Por lo que nuestra propuesta de desarrollar un modelo general para zonas urbanas es necesaria aunque un tanto limitada por la recopilación de una escasa información estadística existente en los distintos entes rectores, como el MARN, DIGESTYC, Vice Ministerio de Vivienda y Desarrollo Urbano (VMVDU), Servicio Nacional de Estudios Territoriales (SNET), Universidad de El Salvador (UES) y otros. En nuestro país se conoce, sobre todo de investigaciones en torno a indicadores socio-económicos, no así, sobre indicadores sintéticos medioambientales, no se ha desarrollado teoría al respecto, ni aplicaciones prácticas, por lo cual sería éste trabajo el que dé origen a una serie de investigaciones relacionadas a este tópico, para generar una línea base; la cual se tenga como apoyo, para la toma de decisiones futuras de manera adecuada. En la Licenciatura en Estadística, carrera que se imparte en la Facultad de Ciencias Naturales y Matemáticas, no se tiene ninguna materia en el plan de estudio que profundice el estudio de indicadores sintéticos, por lo que éste trabajo puede ser iniciativa para incorporar, ya sea una materia al pensum o al menos una temática en algún curso, para culturizar a los futuros estadistas en el área de los Índices. Además existen muchas herramientas y técnicas estadísticas que se utilizan para la construcción de índices sintéticos, las cuales podemos desarrollar; por ejemplo, técnicas de análisis multivariado como el Análisis de Componentes Principales (ACP), Análisis de Conjuntos Difusos, Análisis de Distancia (DP_2), Análisis por Factores (AF), Análisis de Correlación Canónica (ACC), Análisis de Variables Canónicas (AVC), Análisis Discriminantes Canónico (ADC).

OBJETIVOS

Objetivo general

- Desarrollar una metodología para la construcción de un modelo general de índices sintéticos de calidad ambiental.

Objetivos específicos

- Clasificar adecuadamente la información medioambiental que ha sido proporcionada por MARN, DIGESTYC, VMVDU, SNET y otros.
- Identificar las variables ambientales que inciden en la construcción de indicadores e índices sintéticos.
- Aplicar técnicas de análisis multivariado, en la construcción de un índice sintético.
- Elaborar un modelo general de índices sintéticos, para El Salvador, desagregado por municipio.
- Sistematizar las técnicas de Análisis Multivariante: Componentes Principales, Análisis de Factores, utilizadas en la construcción de un modelo.
- Verificar la eficiencia del modelo respecto a la realidad salvadoreña, a través de la metodología propuesta.
- Establecer una metodología para la construcción de indicadores ambientales y que sirvan de base para futuras investigaciones similares a ésta.
- Aplicar herramientas estadísticas tales como Análisis Descriptivo, Análisis Multivariante, para tener la mayor confiabilidad en los resultados.

CAPÍTULO I

MODELOS MULTIVARIANTES:

-COMPONENTES PRINCIPALES-

1.1. PRÓLOGO.

La teoría estadística y los métodos aplicados a la elaboración de indicadores sintéticos por entidades del orden local, regional e internacional, recomiendan el uso de análisis multivariante para el tratamiento de datos que describen fenómenos que pueden ser explicados en común. Se argumenta que pasar de la recopilación de indicadores simples a la generación de indicadores sintéticos es un imperativo para simplificar los datos y estructurarlos en información más elaborada, de acuerdo a las necesidades de los tomadores de decisión y el público en general.

En este trabajo se realiza una descripción general de la utilidad del análisis multivariante de datos, empleando la técnica de componentes principales con el fin de definir la dimensionalidad real de los datos disponibles, para la estimación de un indicador sintético o índice de calidad ambiental a nivel urbano (en los municipios de El Salvador). El empleo del análisis multivariante para mejorar la interpretación de problemas no es nuevo; los autores Escobar y Bermúdez (2004) presentan un índice de calidad ambiental para las localidades de la ciudad de Bogotá, Colombia, empleando Análisis de Componentes Principales (ACP), indicando con este estudio que la ciudad no es ambientalmente homogénea según el índice construido y que es posible priorizar las inversiones ambientales de acuerdo al valor resultante del índice. En este mismo sentido presentan los detalles de la aplicación del índice de calidad ambiental para la ciudad de Cali, Colombia. A nivel internacional, se recomienda la revisión del estudio de autores como Castro (2004) quien hace una excelente revisión bibliográfica sobre el uso de técnicas de análisis multivariante para la construcción de índices de desarrollo sostenible, y emplea las técnicas de ACP, entre otras para estimar índices de desarrollo sostenible para los municipios de la comunidad de Andalucía (España).

1.1.1. Reseña histórica.

Los primeros estudios descriptivos para encontrar la relación entre variables son debidos a Adolfo Quetelet (1796-1874), astrónomo belga que inició la aplicación de los métodos de probabilidad desarrollados por matemáticos y físicos a las ciencias sociales. Quetelet desarrolló los métodos estadísticos para estudiar problemas demográficos, sociológicos y políticos. Su trabajo influyó mucho en Florence Nightingale (1820-1910), la primera mujer estadística, que consiguió, contra todas las dificultades de su tiempo, estudiar matemáticas en Gran Bretaña. Nightingale utilizó los métodos estadísticos para relacionar las condiciones sanitarias con la mortalidad en el ejército británico, demostrando que la falta de higiene producía más bajas que las acciones enemigas. Fue responsable de la reorganización de la sanidad militar durante la guerra de Crimea y la primera mujer condecorada con la orden del mérito británico. Su trabajo contribuyó a luchar contra los prejuicios imperantes mediante los estudios empíricos. El primer método para medir la relación estadística entre dos variables es debido a Francis Galton (1822-1911), que introduce el concepto de recta de regresión y la idea de correlación entre variables en su libro *Natural Inheritance*, publicado en 1889 cuando Galton tenía 67 años. Estos descubrimientos surgen en sus investigaciones sobre la transmisión de los rasgos hereditarios, motivados por su interés en contrastar empíricamente la teoría de la evolución de las especies, propuesta por su primo Charles Darwin en 1859. El concepto de correlación es aplicado en las ciencias sociales por Francis Edgeworth (1845-1936), que estudia la normal multivariante y la matriz de correlación.

Karl Pearson (1857-1936), distinguido estadístico británico creador del famoso contraste ji-cuadrado que lleva su nombre, obtuvo el estimador del coeficiente de correlación en muestras, y se enfrentó al problema de determinar si dos grupos de personas, de los que se conocen sus medidas físicas, pertenecen a la misma raza. Este problema intrigó a Harold Hotelling (1885-1973), que pasó del periodismo a las matemáticas y la economía y que, atraído por la estadística, entonces una joven disciplina emergente, viaja en 1929 a la estación de investigación agrícola de Rothamsted en el Reino Unido para trabajar con el ya célebre científico y figura destacada de la estadística, R. A. Fisher (1890-1962). Hotelling se interesó por el problema de comparar tratamientos agrícolas en función de varias variables, y descubrió las semejanzas entre este problema y el planteado por Pearson. Debemos a Hotelling

(1931) el contraste que lleva su nombre, que permite comparar si dos muestras multivariantes vienen de la misma población. A su regreso a la Universidad de Columbia en Nueva York, Truman Kelley, profesor de pedagogía en Harvard, planteó a Hotelling el problema de encontrar los factores capaces de explicar los resultados obtenidos por un grupo de personas en test de inteligencia.

El [Análisis](#) de Componentes Principales, es una técnica [estadística](#) del Análisis de Datos Multivariante que fue propuesta a [principios](#) del siglo pasado por Karl Pearson (1901) como parte del análisis de factores. Sin embargo la complejidad de los cálculos retrasó su [desarrollo](#) hasta la aparición de los computadores y su utilización en la segunda mitad del siglo XX. El relativamente reciente progreso de los [métodos](#) basados en componentes principales hace que ellos sean poco utilizados por una gran cantidad de investigadores no especialistas en estadística, por eso nuestro propósito es dar a conocer la [naturaleza](#) del ACP y mostrar algunas de sus posibles aplicaciones, para nuestro caso, la construcción de indicadores sintéticos.

1.1.2. Métodos de análisis multivariante.

Es posible que cualquier investigador se enfrente a teorías científicas o criterios consensuados sobre algún aspecto a investigar, por ejemplo, sociales y ambientales. En cada uno de estos componentes se pueden describir un conjunto grande, complicado y complejo de datos que representan las variables que explican el desarrollo sostenible en las diferentes unidades de observación (por ejemplo países, ciudades, comunidades, barrios, etc.). Es aquí donde tiene sentido el uso de métodos de análisis multivariante dado que ayudan al investigador a resumir grandes cantidades de variables, que pueden estar correlacionadas, por medio de relativamente pocos componentes que los simplifican.

Los métodos de análisis multivariante son un conjunto de técnicas de análisis de datos que permiten al investigador testear la utilidad conjunto de los datos que se emplean en la explicación de los fenómenos de interés analítico, como el bienestar social, la calidad ambiental urbana y el desarrollo sostenible, etc.

Para un empleo acertado de los métodos de análisis multivariante es pertinente primero definir la unidad de observación (cantones, municipios, países, etc.) sobre la cual recae

dicho análisis. Las unidades de observación son la base para el análisis del fenómeno que se quiere estudiar, dado que ello exige una medición y evaluación de los datos al nivel de agregación de la unidad seleccionada.

La aplicación de métodos de análisis multivariante en la medición de la calidad ambiental, el desarrollo sostenible, el bienestar social, la calidad de vida, etc., como una variable latente, conduce a la elaboración de un indicador sintético que resume la información contenida en múltiples datos, dado que este es determinado por una variedad de indicadores simples que inciden positiva y negativamente en su valoración.

Es importante aclarar que las técnicas de análisis multivariante se caracterizan por su tendencia exploratoria en lugar de confirmatoria. De esta forma, se debe entender que la estadística convencional exige al investigador comprobar hipótesis con el uso de los datos, mientras que con el análisis multivariante el investigador intenta derivar de los datos una explicación relevante y consistente, dando respuesta a si existe alguna información valiosa en la estructura de los datos.

1.1.3. Descripción general de las técnicas de análisis multivariante.

Se puede determinar dos grandes categorías de técnicas de análisis multivariante:

1. Las técnicas dirigidas por variables que determinan las correlaciones que podrían existir entre las variables respuestas, a través del análisis de la matriz de correlación. Algunos ejemplos de estas técnicas son el Análisis de Componentes Principales, el Análisis por Factores, el Análisis de Regresión, el Análisis de Correlación Canónica, Análisis de Variables Canónicas, Análisis Discriminante Canónico y el Análisis de Distancia.
2. Las técnicas dirigidas a las unidades de observación se dirigen a determinar la relación que podría existir entre éstas. Ejemplo de estas técnicas son el Análisis Discriminante, Análisis por Agrupación y el Análisis Multivariado de la Varianza.

Presentaremos el marco conceptual para estimar indicadores sintéticos, aplicando la técnica ACP. La selección de esta técnica obedece a que el análisis que se realiza para estimar indicadores sintéticos consiste, a grandes rasgos, en la exploración de la

correlación entre los indicadores simples, la reducción de la dimensionalidad de los datos, la agrupación de los indicadores a través de componentes que incorporen la mayor parte de la varianza contenida en los datos y por último la calificación y ordenación de las unidades de observación, que en el caso de este estudio se refieren a municipios de El Salvador.

1.1.4. Análisis de componentes principales.

En muchas situaciones, los científicos de todas las disciplinas se enfrentan al manejo de un conjunto amplio de datos que representan y explican el comportamiento de un vasto número de unidades de observación. En estos casos, el uso de técnicas multivariante para el análisis de los datos debe partir inicialmente de un examen de los mismos, asegurando el cumplimiento de las propiedades estadísticas básicas para realizar análisis entre una variable latente (el índice) y las variables respuestas (indicadores simples). Es probable que las variables respuestas puedan presentar propiedades no deseables para el análisis multivariados. Por ello se requiere muchas veces tratamientos previos del conjunto de datos, que prueben si la distribución de este conjunto de variables es una distribución normal multivariada. Si esto no es así, es altamente deseable una transformación del conjunto de variables.

El ACP es un procedimiento matemático que transforma un conjunto de variables respuestas correlacionadas en un conjunto menor de variables ortogonales (no relacionables) llamadas componentes principales que tienen como fundamento explicar la mayor parte de la varianza contenida en los datos originales. En este sentido, los componentes principales son combinaciones lineales de las variables originales².

Se advierte que si las variables ya están casi no correlacionadas, el analista no gana nada con aplicar ACP porque la dimensionalidad real del conjunto de datos es igual al número de variables respuestas. Ello demanda la necesidad de aplicar una prueba estadística previa que defina si los datos siguen una distribución normal multivariada.

² Esta propiedad de relaciones lineales entre las variables respuestas es altamente aplicado al marco conceptual del índice, en la medida en que el modelo utilizado parte de relaciones lineales causa-efecto para explicar fenómenos sociales y ambientales.

Aunque el ACP es una técnica descriptiva, no niega la posibilidad de que también pueda ser utilizado con fines de inferencia. Por otra parte, las aplicaciones del ACP son numerosas y entre ellas podemos citar la clasificación de individuos, la comparación de poblaciones, la estratificación multivariada, entre algunos.

Se ha dicho que el problema central en el análisis de datos multivariantes es la reducción de la dimensionalidad, si es posible describir con precisión los valores de p variables por un pequeño subconjunto $k < p$ de ellas, se habrá reducido la dimensión del problema a costa de una pequeña pérdida de información.

El análisis de componentes principales tiene este objetivo: dada n observaciones de p variables, se analiza si es posible representar adecuadamente esta información con un número menor de variables construidas como combinaciones lineales de las originales. Por ejemplo, con variables con alta dependencia es frecuente que un pequeño número de nuevas variables (menos del 20 por 100 de las originales) expliquen la mayor parte (más del 80 por 100 de la variabilidad original).

1.1.5. Utilidades del análisis de componentes principales.

Nos referimos a la utilidad del ACP por las siguientes funciones:

1. Selección (depuración) de datos multivariados. El ACP es útil para el análisis previo de cualquier tipo de análisis multivariado, y puede ayudar a revelar anomalías en los datos y a descubrir datos atípicos (outliers).
2. Agrupación de unidades de observación. El ACP ayuda al investigador a detectar subgrupos de acuerdo al comportamiento común de la varianza de las variables respuesta.
3. Análisis discriminante. En este tipo de análisis, la matriz de varianza-covarianza requiere ser invertida para realizar una regla de discriminación. Cuando el número de variables respuestas es superior al número de unidades de observación, la matriz no se puede invertir. Allí juega un papel importante el ACP para reducir todas las variables respuestas en un número menor de componentes principales que expliquen el mayor porcentaje de varianza de los datos originales.

4. Permite representar óptimamente en un espacio de dimensión pequeña observaciones de un espacio general p -dimensional. En este sentido, componentes principales es el primer paso para identificar las posibles *variables latentes*, o no observadas que generan los datos.
5. Regresión. El ACP ayuda a determinar y corregir los problemas derivados de regresiones que presenten problemas de multicolinealidad.
6. Permite transformar las variables originales, en general correladas, en nuevas variables incorreladas, facilitando la interpretación de los datos.

1.1.6. Objetivos del análisis de componentes principales.

Son dos los objetivos más relevantes asociados al análisis de componentes principales:

1. Reducir la dimensionalidad de un conjunto de datos, descubriendo la verdadera dimensión contenida en ellos. Cuando la dimensionalidad real de los datos es inferior al número original de los datos, entonces el análisis sugiere reemplazar estos últimos por los componentes principales estimados, ayudando a mejorar la interpretación de los datos³.
2. Identificar nuevas variables significativas contenidas en la estructura de los datos. La creación de nuevas variables (componentes principales) es una combiación lineal de las variables originales que debe seguir un orden de importancia, teniendo en cuenta que éstas no están correlacionadas. La primera componente explica la mayor varianza posible en los datos y cada componente adicional adquiere la mayor variabilidad posible restante.

Usualmente, estudios aplicados de análisis multivariado presentan dos procedimientos generales para definir los componentes principales. Mediante el análisis de la matriz de

³ En concreto, el procedimiento realizado define geoméricamente un subespacio creado con las “ m ” primeras componentes, mejorando el ajuste de toda la información de las variables utilizadas mediante la estimación de la suma de los cuadrados de las distancias perpendiculares desde cada punto al subespacio determinado. Es importante reiterar que el ACP agrega valor a los datos, cuando la dimensionalidad de p variables es mayor que los m componentes.

varianza-covarianza y a través de la matriz de correlación (R). Por ejemplo, para la definición formal de la estructura analítica que define los componentes principales, Luis Escobar (2004) desarrolla el análisis de la matriz de correlaciones, aclarando que el uso de la matriz de varianza-covarianza es apropiado cuando el investigador está seguro de que todas las variables del análisis guardan las siguientes condiciones:

- Todas las variables deben estar medidas en las mismas unidades o comparables.
- Las unidades deben tener varianzas similares.

El argumento de estas exigencias para el uso de la matriz de varianza-covarianza es que si alguna de las variables respuesta tiene una unidad de medida o escala distinta, ésta puede tener un efecto significativo sobre los componentes principales, presentándola como importante, cuando en realidad no lo sea. Esto es lo que se llama sesgo de componentes iniciales. Si una de las variables tiene una varianza mucho mayor que las demás, dominará la componente principal, sin importar la estructura de las covarianzas de las variables y en este caso tiene poco objeto la realización de un ACP.

Cuando las variables no presentan fundamentos de escala similares, es necesario la aplicación del ACP a la matriz de correlación de las P variables respuestas tipificadas y no a los datos originales. Por lo tanto, exige al investigador una transformación previa de los datos de tal forma que garantice una escala y unidad de medida común.

Se recomiendan tener cuidado cuando se elige algún criterio de tipificación para el conjunto de variables, dado que con ello se está diciendo a priori que todas las variables respuestas tienen igual importancia. Es conveniente por ello hacer un examen de los datos, de tal forma que permita al investigador observar la relevancia de utilizar la matriz de varianza-covarianza o la matriz de correlación.

1.2. DEFINICIÓN DE COMPONENTES PRINCIPALES

Supongamos que se dispone de los valores de p-variables en n elementos de una población dispuestos en una matriz X de dimensión $n \times p$, donde las columnas contienen las variables y las filas los elementos. Supondremos que previamente hemos restado a cada variable su media, de manera que las variables de la matriz X tienen media cero y su matriz de covarianzas vendrá dada por $\frac{1}{n} X'X$.

El problema que se desea resolver es encontrar un espacio de dimensión más reducida que represente adecuadamente los datos. Puede abordarse desde tres perspectivas equivalentes.

1.2.1. Enfoque descriptivo.

Se desea encontrar un subespacio de dimensión menor que p tal que al proyectar sobre él los puntos conserven su estructura con la menor distorsión posible. Veamos cómo convertir esta noción intuitiva en un criterio matemático operativo. Consideremos primero un subespacio de dimensión uno, una recta. Se desea que las proyecciones de los puntos sobre esta recta mantengan, lo más posible, sus posiciones relativas. Para concretar, consideremos el caso de dos dimensiones ($p = 2$). La figura 1.1 indica el diagrama de dispersión y una recta que, intuitivamente, proporciona un buen resumen de los datos, ya que la recta pasa cerca de todos los puntos y las distancias entre ellos se mantienen aproximadamente en su proyección sobre la recta. La condición de que la recta pase cerca de la mayoría de los puntos puede concretarse exigiendo que las distancias entre los puntos originales y sus proyecciones sobre la recta sean lo más pequeñas posibles. En consecuencia, si consideramos un punto x_i y una dirección $a_1 = (a_{11}, \dots, a_{1p})'$, definida por un vector a_1 de norma unidad, la proyección del punto x_i sobre esta dirección es el escalar:

$$Z_i = a_{11}x_{i1} + \dots + a_{1p}x_{ip} = a_1'X_i \quad (1)$$

y el vector que representa esta proyección será $Z_i a_1$. Llamando r_i a la distancia entre el punto X_i , y su proyección sobre la dirección a_1 , este criterio implica:

$$\text{minimizar } \sum_{i=1}^n r_i^2 = \sum_{i=1}^n |X_i - Z_i a_1|^2 \quad (2)$$

donde $|X_i - Z_i a_1|$ es la norma euclídea o módulo del vector $X_i - Z_i a_1$.

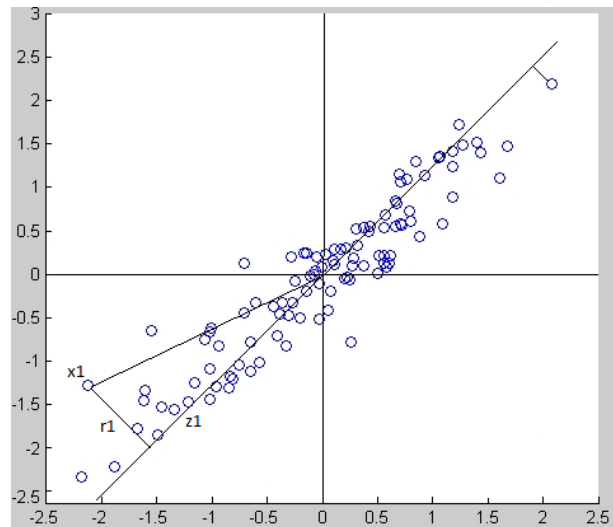
La figura 1.1 muestra que al proyectar cada punto sobre la recta se forma un triángulo rectángulo donde la hipotenusa es la distancia del punto al origen, $(x_i'x_i)^{\frac{1}{2}}$, y los catetos la proyección del punto sobre la recta (Z_i) y la distancia entre el punto y su proyección (r_i). Por el teorema de Pitágoras, podemos escribir:

$$X'_i X_i = Z_i^2 + r_i^2 \quad (3)$$

Y sumando esta expresión para todos los puntos, se obtiene:

$$\sum_{i=1}^n X'_i X_i = \sum_{i=1}^n Z_i^2 + \sum_{i=1}^n r_i^2 \quad (4)$$

Figura 1.1. Ejemplo de la recta que minimiza las distancias ortogonales de los puntos a ella.



Como el primer miembro es constante, minimizar $\sum_{i=1}^n r_i^2$, la suma de las distancias a la recta de todos los puntos, es equivalente a maximizar $\sum_{i=1}^n Z_i^2$, la suma al cuadrado de los valores de las proyecciones. Como las proyecciones Z_i son, por (1) variables de media cero, maximizar la suma de sus cuadrados equivale a maximizar su varianza, y obtenemos el criterio de encontrar la dirección de proyección que maximice la varianza de los datos proyectados. Este resultado es intuitivo: la recta de la figura 1.1 parece adecuada porque conserva lo más posible la original de los puntos. El lector puede convencerse considerando una dirección de proyección perpendicular a la de la recta en esta figura: los puntos tendrían muy poca variabilidad y perderíamos la información sobre sus distancias en el espacio. Si en lugar de buscar la dirección que pasa cerca de los puntos buscamos la dirección tal que los puntos proyectados sobre ella conserven lo más posible sus distancias originales entre los puntos y $D = \sum_i \sum_j (d_{ij}^2 - \hat{d}_{ij}^2)$ sea

mínima. Como la suma de las distancias originales es fija, minimizar D requiere maximizar $\sum_i \sum_j \hat{d}_{ij}^2$, las distancias entre los puntos proyectados.

1.2.2. Enfoque estadístico.

Representar puntos p dimensionales con la mínima pérdida de información en un espacio de dimensión uno es equivalente a sustituir las p variables originales por una nueva variable, Z_1 , que resuma óptimamente la información. Esto supone que la nueva variable debe tener globalmente máxima correlación con las originales o, en otros términos, debe permitir prever las variables originales con la máxima precisión. Esto no será posible si la nueva variable toma un valor semejante en todos los elementos, la condición para que podamos prever con la mínima pérdida de información los datos observados, es utilizar la variable de máxima variabilidad.

Volviendo a la figura 1.1 se observa que la variable escalar obtenida al proyectar los puntos sobre la recta sirve para prever bien el conjunto de los datos. La recta indicada en la figura no es la línea de regresión de ninguna de las variables con respecto a la otra, que se obtienen minimizando las distancias verticales u horizontales, sino la que minimiza las distancias ortogonales o entre los puntos y la recta y se encuentra entre ambas rectas de regresión.

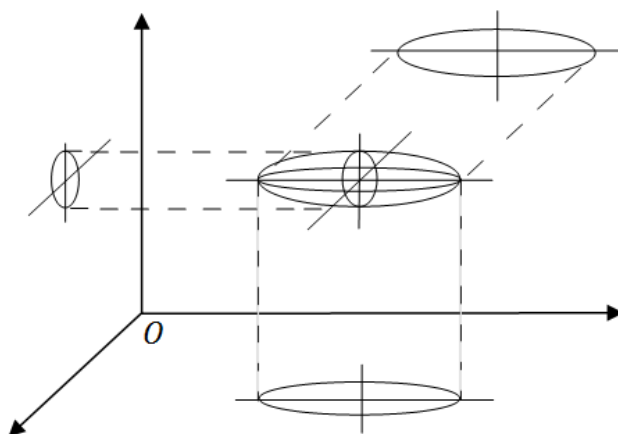
Este enfoque puede extenderse para obtener el mejor subespacio resumen de los datos de dimensión 2. Para ello, calcularemos el plano que mejor aproxima a los puntos. El problema se reduce a encontrar una nueva dirección definida por un vector unitario, a_2 , que, sin pérdida de generalidad, puede tomarse ortogonal a a_1 , y que verifique la condición de que la proyección de un punto sobre este eje maximice las distancias entre los puntos proyectados. Estadísticamente esto equivale a encontrar una segunda variable z_2 , incorrelada con la anterior, y que tenga varianza máxima. En general, la componente Z_k ($k < p$) tendrá varianza máxima entre todas las combinaciones lineales de las p variables originales, con la condición de estar incorreladas con las Z_1, \dots, Z_{k-1} previamente obtenidas.

1.2.3. Enfoque geométrico.

El problema puede abordarse desde un punto de vista geométrico con el mismo resultado final. Si consideramos la nube de puntos de la figura 1.1 vemos que los puntos se sitúan siguiendo una elipse figura 1.2 y podemos describirlos por su proyección en la dirección del eje mayor de la elipse. Puede demostrarse que este eje es la recta que minimiza las distancias ortogonales, con lo que volvemos al problema que ya hemos resuelto. En varias dimensiones tendremos elipsoides, y la mejor aproximación a los datos es la proporcionada por su proyección sobre el eje mayor del elipsoide. Intuitivamente la mejor aproximación en dos dimensiones es la proyección sobre el plano de los dos ejes mayores del elipsoide y así sucesivamente. Considerar los ejes del elipsoide como nuevas variables originales supone pasar de variables correladas a variables ortogonales o incorreladas como veremos a continuación.

Se puede realizar una representación gráfica de la nube de puntos X_1, X_2, \dots, X_p , columnas de la matriz $X_{(n,p)}$ de casos-variables resultante de medir las variables X_1, X_2, \dots, X_p sobre una muestra de tamaño n . Si únicamente se consideran dos variables, una nube de puntos en un plano puede ser asimilada, de forma simple, a una elipse como envolvente de todos ellos. Si se manejan tres variables, la figura en relieve resultante sería un elipsoide. Y si fueran más de tres ($p > 3$), sería necesario un esfuerzo de generalización para imaginar un hiperelipsoide p -dimensional. En general, los ejes principales de estas figuras o cuerpos geométricos presentaran una inclinación espacial cualquiera respecto a los ejes que representan las variables originales, es decir, no tienen por qué ser paralelos a estos. Si lo fueran, en un caso hipotético, las proyecciones del elipsoide sobre los planos que definen dos a dos las variables (figura 1.2) representarían la máxima dispersión de los puntos de la nube (máximas “sombras” del elipsoide) y las variables originales, cada una por sí sola, contendrían la máxima información de la nube en una de las p dimensiones.

Figura 1.2. Interpretación geométrica de la nube de puntos. ACP.



Como, además, los ejes son perpendiculares, la dispersión que condensa una de las variables no tiene nada que ver con la que condensan las demás, de modo que cada una recoge la parte de información que no pueden recoger las otras, como ocurre fotografiando un objeto desde tres direcciones perpendiculares.

Pero como la nube, en general es espacialmente oblicua, las variables originales no recogen solas la información optimizada (máxima “sombras” o dispersión); y habrá otros p ejes (también entre sí perpendiculares y transformados de los ejes iniciales), que serán precisamente los ejes reales del elipsoide, que sí optimizarán la información. Las ecuaciones de estos nuevos ejes serán, como siempre en geometría analítica, combinaciones lineales de las p variables iniciales, que los posicionaran en el espacio. Y se calcularán bajo las condiciones matemáticas de que pasen por el centro de gravedad de la nube, que sean perpendiculares entre sí, que el primero de ellos (el más largo) haga máxima (derivada=0) la dispersión de la nube sobre él, que el segundo (el siguiente en importancia) haga máxima la dispersión en un plano perpendicular al primero y así los demás, hasta obtener los p nuevos ejes. Para dos (y hasta para tres) variables iniciales, este cálculo es abordable por métodos habituales de geometría analítica. Pero cuando son más de tres, se llega a tal magnitud de cálculos que es preciso sistematizar la técnica matemática convencional recurriendo a la única herramienta que lo hace posible: el cálculo matricial.

Si los nuevos ejes obtenidos se cortan, como debe ser, en el centro de gravedad de la nube, sus ecuaciones, al no pasar por el origen O , tendrán un término constante

(ordenada en el origen). Pero, como a efecto de información recogida en las proyecciones del elipsoide, no importa más que la dirección de ellos, se puede obligar a que pasen por O con lo que todos sus términos independientes desaparecerán.

Sus ecuaciones representan nuevas variables ficticias, combinación lineal de las reales, que deberán ser interpretadas y si tienen sentido, etiquetadas o nombradas. Se trata de variables sintéticas (componentes principales) que, aunque no han sido medidas en los individuos, si pueden ser calculadas a través de los valores que estos presenten en todas las variables originales. En cada una de estas p componentes principales habrá algunas variables que contribuyan más a su configuración, o sea, “pesarán” más en su diseño. Este peso viene dado por los valores de los coeficientes de las variables originales en la ecuación de la componente. Y otras variables pesaran menos, revelando una menor influencia de ellas en la componente. Algunas de estas componentes, precisamente las últimas que surjan en el cálculo de la progresivamente menor dispersión perpendicular, pueden presentar escasa dispersión de la nube desde esas direcciones. (El elipsoide, si fueran tres dimensiones, puede ser tan “aplanado”, visto desde la última componente, que considerarlo como una elipse al ser tan estrecho hará perder escasa información). Desechando estas últimas componentes, previa cuantificación, que permite el cálculo, de la poca información perdida, se habrá simplificado la dimensión del problema que pasa de p variables iniciales a k , si bien cada una de las componentes conservadas mantienen en su ecuación a todas las variables originales.

Se puede ver así al elipsoide p -dimensional que contiene toda la información del problema estudiado, desde unas cuantas (k) direcciones principales que presentan k puntos de vista simplificados que el investigador deberá, si puede, interpretar. La adecuada interpretación de estas nuevas variables sintéticas va a depender de que cada una de ellas agrupe con más peso algunas de las variables originales de significado parecido y con menos peso, las demás. (Por ejemplo, un componente que aparecería con coeficientes altos para la talla y el peso, y bajos para la presión arterial y el colesterol, podría definirse o etiquetarse como una nueva variable llamada tamaño, que no ha sido medida en el estudio).

Pero no siempre será tan destacada la selección útil de las variables en las ecuaciones de los componentes principales y en consecuencia, no resultará fácil la identificación o interpretación de los k “puntos de vista”. En tal caso, se debe sacrificar la ubicación

óptima de los componentes (ejes ideales del elipsoide), haciéndolos rotar algo (poco, o incluso mucho) para que, perdiendo en la dispersión que cada uno explica, pero no en la dispersión total explicada, ganen en agrupación verosímil de variables y por lo tanto, en interpretabilidad práctica. Esta rotación última del proceso puede mejorar mucho la utilidad del análisis de componentes principales en la consecución perseguida de un equilibrio entre la reducción de las dimensiones del problema estudiado y la más fácil interpretabilidad de lo que se conserva.

Esta idea intuitiva geométrica es equivalente a la representación matemática que hemos utilizado para la obtención de las componentes principales Z_1, \dots, Z_p , como combinación lineal de las p variables iniciales X_1, \dots, X_p de la forma siguiente:

$$\begin{aligned} Z_1 &= a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \\ Z_2 &= a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p \\ &\dots \\ Z_p &= a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p \end{aligned}$$

Donde, los a_{ij} representan los pesos o cargas factoriales de cada variable en cada componente. Existirán tantos componentes Z_1, \dots, Z_p como número de variables, definidas por p series de coeficientes $a_1 = (a_{11}, \dots, a_{1p}), \dots, a_p = (a_{p1}, \dots, a_{pp})$. Cada componente explica una parte de la varianza total, considera esta como una manera de valorar la información total de la tabla de datos. Si se consigue encontrar pocos componentes (k), capaces de explicar casi toda la varianza total, podrán sustituir a las variables primitivas con mínima pérdida de información. De esta forma se dispondrá de unas variables ficticias que, siendo pocas, contiene a todas las originales. Este es el objetivo del análisis en componentes principales: simplificar o reducir la tabla inicial, de $n \times p$ a $n \times k$. Naturalmente, si en vez de seleccionar k componentes principales, se tomaran los p posibles, no existiría pérdida alguna de información, pero no se habría conseguido simplificar el problema.

1.2.4. Cálculo de los componentes.

En el análisis en componentes principales se dispone de una muestra de tamaño n acerca de p variables X_1, X_2, \dots, X_p (tipificadas o expresadas en desviaciones respecto de su media) inicialmente correlacionadas, para posteriormente obtener a partir de ellas un número $k < p$ de variables incorreladas Z_1, Z_2, \dots, Z_p que sean combinación lineal de las variables iniciales y que expliquen la mayor parte de su variabilidad.

1.2.4.1. Cálculo del primer componente.

El primer componente principal se define como la combinación lineal de las variables originales que tiene varianza máxima. Los valores en este primer componente de los n individuos se representarán por un vector Z_1 , dado por

$$Z_1 = Xa_1 \quad (5)$$

$$Z_{1i} = a_{11}X_{1i} + a_{12}X_{2i} + \dots + a_{1p}X_{pi} \quad (6)$$

En forma de matrices:

$$\begin{bmatrix} Z_{11} \\ Z_{12} \\ \vdots \\ Z_{1n} \end{bmatrix} = \begin{bmatrix} X_{11} & X_{21} & \dots & X_{p1} \\ X_{12} & X_{22} & \dots & X_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1n} & X_{2n} & \dots & X_{pn} \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \\ \vdots \\ a_{1p} \end{bmatrix} \quad (7)$$

Tanto si las X_j están tipificadas, como si están expresadas en desviaciones respecto de su media muestral, la media de Z_1 es cero: $E(Z_1) = E(Xa_1) = E(X)a_1 = 0$.

Como las variables originales tienen media cero también Z_1 tendrá media nula.

La varianza de Z_1 será:

$$V(Z_1) = \frac{\sum_{i=1}^n Z_{1i}^2}{n} = \frac{1}{n} Z_1' Z_1 = \frac{1}{n} (a_1' X)' (X a_1) = \frac{1}{n} a_1' X' X a_1 = a_1' \left[\frac{1}{n} X' X \right] a_1 = a_1' S a_1$$

$$V(Z_1) = a_1' S a_1 \quad (8)$$

Si las variables están expresadas en desviaciones respecto de la media, la expresión $\frac{1}{n} X' X$ (*matriz de inercia*) es la matriz de covarianza muestral a la que denominaremos S ,

y para variables tipificadas $\frac{1}{n}X'X$ es la matriz de correlaciones R . Es obvio que podemos maximizar la varianza sin límite aumentando el módulo del vector a_1 . Para que la maximización de (8) tenga solución debemos imponer una restricción al módulo del vector a_1 , y, sin pérdida de generalidad, impondremos que $\sum_{i=1}^p a_{1i}^2 = a_1' a_1 = 1$. Introduciremos esta restricción mediante el multiplicador de Lagrange:

$$L = a_1' S a_1 - \lambda(a_1' a_1 - 1) \quad (9)$$

y maximizaremos esta expresión de la forma habitual derivando respecto a los componentes de a_1 e igualando a cero. Entonces

$$\frac{\partial L}{\partial a_1} = 2S a_1 - 2\lambda a_1 = 0 \Rightarrow \boxed{S a_1 = \lambda a_1} \quad (10)$$

Que implica que a_1 es un vector propio de la matriz S , y λ su correspondiente valor propio. Además:

$$S a_1 = \lambda a_1 \Rightarrow \boxed{(S - \lambda I) a_1 = 0} \quad (11)$$

Se trata de un sistema homogéneo en a_1 , que sólo tiene solución si el determinante de la matriz de los coeficientes es nulo, es decir, $|S - \lambda I| = 0$. En la ecuación (11) podemos multiplicar por a_1' a la derecha, con lo que se tiene:

$$a_1'(S - \lambda I) a_1 = 0 \Rightarrow a_1' S a_1 = \lambda a_1' a_1 \Rightarrow \boxed{V(Z_1) = \lambda} \quad (12)$$

concluimos, por (12), que λ es la varianza de Z_1 . Como ésta es la cantidad que queremos maximizar, λ será el mayor valor propio de la matriz S . Su vector asociado, a_1 , define los coeficientes de cada variable en el primer componente principal.

Tomando λ_1 como el mayor valor propio de S y tomando a_1 como su vector propio asociado normalizado ($a_1' a_1 = 1$), ya tenemos definido el vector de ponderaciones que se aplica a las variables iniciales para obtener la primera componente principal, componente que vendrá definida como:

$$Z_1 = X a_1.$$

Ejemplo 1.1: Ilustraremos con detalle el cálculo de la primera componente, el siguiente conjunto de datos (esta matriz de datos podemos encontrarla en la tabla A.1, anexos) contiene 34 observaciones que corresponden a distintas acciones que cotizan en el mercado continuo español, y 3 variables que corresponden a medidas de rentabilidad de estas acciones durante un período de tiempo. Las variables son:

X1: rentabilidad efectiva por dividendos.

X2: proporción de beneficios que va a dividendos.

X3: ratio entre precio por acción y beneficios.

Para calcular el primer componente principal necesitamos la matriz de varianzas-covarianzas.

Como las dimensiones de las variables son distintas, los cálculos se realizaron transformando las variables para facilitar su interpretación. Tomando logaritmos en las variables, la matriz de covarianzas es:

$$S = \begin{bmatrix} 0.35 & 0.15 & -0.19 \\ 0.15 & 0.13 & -0.03 \\ -0.19 & -0.03 & 0.16 \end{bmatrix}$$

Los valores propios son las raíces de la ecuación:

$$|S - \lambda I| = \begin{vmatrix} 0.35 - \lambda & 0.15 & -0.19 \\ 0.15 & 0.13 - \lambda & -0.03 \\ -0.19 & -0.03 & 0.16 - \lambda \end{vmatrix} = 0.000382 - 0.0628\lambda + 0.64\lambda^2 - \lambda^3 = 0$$

Al resolver el polinomio $0.000382 - 0.0628\lambda + 0.64\lambda^2 - \lambda^3 = 0$, obtenemos las siguientes raíces:

$$\lambda_1 = 0.521, \quad \lambda_2 = 0.113, \quad \lambda_3 = 6.51 \times 10^{-3}$$

Obsérvese que:

$$\lambda_1 > \lambda_2 > \lambda_3.$$

El vector propio asociado a λ_1 nos da los pesos de la primera componente principal.

Para calcularlo resolvemos el sistema:

$$S a_1 = \lambda_1 a_1$$

$$\begin{bmatrix} 0.35 & 0.15 & -0.19 \\ 0.15 & 0.13 & -0.03 \\ -0.19 & -0.03 & 0.16 \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \end{bmatrix} = 0.521 \times \begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \end{bmatrix}$$

$$\begin{bmatrix} -0.171a_{11} + 0.15a_{12} - 0.19a_{13} \\ 0.150a_{11} - 0.391a_{12} - 0.03a_{13} \\ -0.190a_{11} - 0.030a_{12} - 0.36a_{13} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Este sistema es compatible indeterminado. Para encontrar una de las infinitas soluciones tomemos la primera variable como parámetro, x , y resolvamos el sistema en función de x .

La solución:

$$\{a_{11} = x, \quad a_{12} = 0.427x, \quad a_{13} = -0.562x\}$$

El valor de x se obtiene ahora imponiendo que el vector tenga norma unidad, con lo que resulta:

$$a_1 = \begin{bmatrix} -0.817 \\ -0.349 \\ 0.459 \end{bmatrix}$$

El primer componente es:

$$Z_1 = -0.817X_1 - 0.349X_2 + 0.459X_3$$

Donde X_1, X_2 , y X_3 son las variables en logaritmos.

1.2.4.2. Cálculo del segundo componente

La segunda componente principal, al igual que las restantes, se expresa como combinación lineal de las variables originales:

$$Z_{2i} = a_{21}X_{1i} + a_{22}X_{2i} + \dots + a_{2p}X_{pi} \quad (13)$$

Para el conjunto de las n observaciones muestrales ésta ecuación se puede expresar matricialmente:

$$\begin{bmatrix} Z_{21} \\ Z_{22} \\ \vdots \\ Z_{2n} \end{bmatrix} = \begin{bmatrix} X_{11} & X_{21} & \dots & X_{p1} \\ X_{12} & X_{22} & \dots & X_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1n} & X_{2n} & \dots & X_{pn} \end{bmatrix} \begin{bmatrix} a_{21} \\ a_{22} \\ \vdots \\ a_{2p} \end{bmatrix} \quad (14)$$

En notación abreviada tenemos:

$$Z_2 = Xa_2 \quad (15)$$

Tanto si las X_j están tipificadas, como si están expresadas en desviaciones respecto de su media muestral, la media de Z_2 es cero:

$$E(Z_2) = E(Xa_2) = E(X)a_2 = 0.$$

La varianza de Z_2 será:

$$V(Z_2) = \frac{\sum_{i=1}^n Z_{2i}^2}{n} = \frac{1}{n} Z_2' Z_2 = \frac{1}{n} (a_2 X)' (X a_2) = \frac{1}{n} a_2' X' X a_2 = a_2' \left[\frac{1}{n} X' X \right] a_2 = a_2' S a_2$$

$$V(Z_2) = a_2' S a_2 \quad (16)$$

La segunda componente Z_2 se obtiene de forma que su varianza sea máxima sujeta a la restricción de que la suma de los pesos a_{2j} al cuadrado sea igual a la unidad, es decir, la variable de los pesos o ponderaciones $(a_{21}, a_{22}, \dots, a_{2p})'$ se toma normalizada ($a_2' a_2 = 1$).

Por otra parte como Z_1 y Z_2 han de estar incorrelacionadas se tiene que:

$$0 = E(Z_2' Z_1) = E(a_2' X' X a_1) = a_2' E(X' X) a_1 = a_2' S a_1$$

También sabemos que $S a_1 = \lambda_1 a_1$ (ya que a_1 es el vector propio de S asociado a su mayor valor propio λ_1). Si multiplicamos por a_2' a la derecha tenemos:

$$0 = a_2' S a_1 = \lambda_1 a_2' a_1 \Rightarrow a_2' a_1 = 0$$

Con lo que a_1 y a_2 son ortogonales.

Se trata entonces de hallar Z_2 maximizando (16), sujeta a las restricciones $a_2' a_2 = 1$ y $a_2' S a_1 = 0$.

Para resolver este problema de optimización con dos restricciones se aplica el método de los multiplicadores de Lagrange considerando la función Lagrangiana:

$$L = a_2' S a_2 - 2\mu(a_2' S a_1) - \lambda(a_2' a_2 - 1) \quad (17)$$

Derivando respecto de a_2 e igualando a cero, se tiene:

$$\frac{\partial L}{\partial a_2} = 2Sa_2 - 2\mu Sa_1 - 2\lambda a_2 = 0$$

Dividiendo por 2 y premultiplicando por a'_1 tenemos:

$$a'_1 Sa_2 - \mu a'_1 Sa_1 - \lambda a'_1 a_2 = 0$$

Y como $Sa_1 = \lambda_1 a_1$ (ya que a_1 es el vector propio de S asociado a su mayor valor propio λ_1), entonces $a'_1 S = \lambda_1 a'_1$, y podemos escribir la igualdad anterior como:

$$\lambda_1 a'_1 a_2 - \mu S[Z_1] - \lambda a'_1 a_2 = 0$$

Pero:

$$a'_1 a_2 = 0 \Rightarrow \mu S[Z_1] = 0 \Rightarrow \mu = 0$$

De donde:

$$\frac{\partial L}{\partial a_2} = 2Sa_2 - 2\lambda a_2 = 0 \Rightarrow (S - \lambda I)a_2 = 0$$

$$(S - \lambda I)a_2 = 0 \quad (18)$$

Se trata de un sistema homogéneo en a_2 , que sólo tiene solución si el determinante de la matriz de los coeficientes es nulo, es decir, $|S - \lambda I| = 0$. Pero la expresión $|S - \lambda I| = 0$ es equivalente a decir que λ es un valor propio de la matriz S .

En la ecuación $(S - \lambda I)a_2 = 0$ podemos multiplicar por a'_2 a la izquierda, con lo que se tiene:

$$Sa_2 = \lambda_2 a_2 \Rightarrow a'_2 Sa_2 = \lambda \quad (19)$$

$$a'_2 Sa_2 = \lambda \Rightarrow V(Z_2) = \lambda \quad (20)$$

Por lo tanto, para maximizar $V(Z_2)$ hemos de tomar el segundo mayor valor propio λ de la matriz S .

Tomando λ_2 como el segundo mayor valor propio de S y tomando a_2 como su vector propio asociado normalizado ($a'_2 a_2 = 1$), ya tenemos definido el vector de ponderaciones que se aplica a las variables iniciales para obtener la segunda componente principal, componente que vendrá definida como:

$$Z_2 = Xa_2.$$

De forma similar, la componente principal h -ésima se define como $Z_h = Xa_h$ (21) donde a_h es el vector propio de S asociado a su h -ésimo mayor valor propio. Suele denominarse también a a_h eje factorial h -ésimo.

En general, la ecuación $|S - \lambda I| = 0$ tiene n raíces $\lambda_1, \lambda_2, \dots, \lambda_n$, que podemos ordenarlas de mayor a menor $\lambda_1 > \lambda_2 > \dots > \lambda_n$.

Ejemplo 1.2: Para calcular el segundo componente del ejemplo anterior, como ya aplicamos la ecuación (9) en el ejemplo 1.1, basta con aplicar (20) y tenemos:

$$Sa_2 = \lambda_2 a_2$$

$$\begin{bmatrix} 0.35 & 0.15 & -0.19 \\ 0.15 & 0.13 & -0.03 \\ -0.19 & -0.03 & 0.16 \end{bmatrix} \begin{bmatrix} a_{21} \\ a_{22} \\ a_{23} \end{bmatrix} = 0.113 \times \begin{bmatrix} a_{21} \\ a_{22} \\ a_{23} \end{bmatrix}$$

$$\begin{bmatrix} 0.237a_{21} + 0.15a_{22} - 0.19a_{23} \\ 0.15a_{21} + 0.017a_{22} - 0.03a_{23} \\ -0.19a_{21} - 0.03a_{22} + 0.047a_{23} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

La solución:

$$\{a_{21} = x, \quad a_{22} = 16.467x, \quad a_{23} = 14.546x\}$$

El valor de x se obtiene de nuevo imponiendo que el vector a_2 tenga norma unidad, con lo que resulta:

$$a_2 = \begin{bmatrix} 0.046 \\ 0.749 \\ 0.662 \end{bmatrix}$$

El segundo componente es:

$$Z_2 = 0.046X_1 + 0.749X_2 + 0.662X_3$$

1.2.5. Generalización.

Puede demostrarse análogamente que el espacio de dimensión r que mejor representa a los puntos viene definido por los vectores propios asociados a los r mayores valores propios de S . Estas direcciones se denominan direcciones principales de los datos y a las nuevas variables por ellas definidas componentes principales. En general, la matriz X (y por tanto S) tiene rango p , existiendo entonces tantas componentes principales como variables que se obtendrán calculando los valores propios o raíces características, $\lambda_1, \lambda_2, \dots, \lambda_p$, de la matriz de varianzas y covarianzas de las variables, S , mediante:

$$|S - \lambda I| = 0 \quad (22)$$

Y sus vectores asociados son:

$$(S - \lambda_i I)a_i = 0 \quad (23)$$

Los términos λ_i son reales, al ser la matriz S simétrica, y positivos, ya que S es definida positiva. Por ser S simétrica si λ_j y λ_h son dos raíces distintas sus vectores asociados son ortogonales. Si S fuese semidefinida positiva de rango $k < p$, lo que ocurriría si $p - k$ variables fuesen combinación lineal de las demás, habría solamente k raíces características positivas y el resto serían ceros.

Llamando Z a la matriz cuyas columnas son los valores de los p componentes en los n individuos, estas nuevas variables están relacionadas con las originales mediante:

$$Z = XA \quad (24)$$

Donde $A'A = I$.

Calcular los componentes principales equivale a aplicar una transformación ortogonal A a las variables X (ejes originales) para obtener unas nuevas variables Z incorreladas entre sí. Esta operación puede interpretarse como elegir unos nuevos ejes coordenados, que coincidan con los “ejes naturales” de los datos.

Ejemplo 1.3: Encuesta de Presupuestos Familiares (EPF) en España, presenta los gastos medios de las familias españolas para las 51 provincias, en nueve epígrafes (detallada en el anexo). La matriz de covarianzas (tabla 1.1) resume la variabilidad de estas 9 variables en los 51 elementos observados. Como las distribuciones de los gastos son muy asimétricas, las variables se han expresado en logaritmos. El vector propio asociado al mayor valor propio, 0.348, define la siguiente variable, primer componente principal:

Tabla 1.1: Matriz de covarianzas de EPF.

	lnx1	lnx2	lnx3	lnx4	lnx5	lnx6	lnx7	lnx8	lnx9
lnx1	0.0125	0.0098	0.0116	0.0121	0.0157	0.0129	0.0215	0.0105	0.0136
lnx2	0.0098	0.0267	0.0190	0.0211	0.0272	0.0217	0.0293	0.0156	0.0206
lnx3	0.0116	0.0190	0.0423	0.0304	0.0423	0.0334	0.0538	0.0341	0.0289
lnx4	0.0121	0.0211	0.0304	0.0418	0.0457	0.0357	0.0518	0.0329	0.0336
lnx5	0.0157	0.0272	0.0423	0.0457	0.0946	0.0509	0.0752	0.0456	0.0452
lnx6	0.0129	0.0217	0.0334	0.0357	0.0509	0.0531	0.0549	0.0344	0.0404
lnx7	0.0215	0.0293	0.0538	0.0518	0.0752	0.0549	0.0987	0.0547	0.0458
lnx8	0.0105	0.0156	0.0341	0.0329	0.0456	0.0344	0.0547	0.0422	0.0305
lnx9	0.0136	0.0206	0.0289	0.0336	0.0452	0.0404	0.0458	0.0305	0.0577

Primer componente principal.

$$Z_1 = 0.12X_1 + 0.18X_2 + 0.3X_3 + 0.31X_4 + 0.46X_5 + 0.34X_6 + 0.5X_7 + 0.31X_8 + 0.31X_9$$

Se observa que Z_1 es una suma ponderada de todos los gastos, con mayor peso, de los gastos en enseñanza y cultura (X_7) y sanitarios (X_5). El menor peso lo tiene el gasto en alimentación (X_1). Si calculamos los valores de Z_1 para las provincias españolas y las ordenamos por esta nueva variable las provincias quedan prácticamente ordenadas por su renta. La primera componente principal tiene, pues, en este caso, una explicación inmediata: redescubre la renta de cada provincia.

Segundo componente principal. Asociado al segundo valor propio mayor que es 0.032.

$$Z_2 = 0.05X_1 + 0.16X_2 - 0.17X_3 + 0.07X_4 - 0.21X_5 + 0.29X_6 - 0.40X_7 - 0.17X_8 + 0.78X_9$$

$$Z_2 = (0.05X_1 + 0.16X_2 + 0.07X_4 + 0.29X_6 + 0.78X_9) - (0.17X_3 + 0.21X_5 + 0.40X_7 + 0.17X_8)$$

Esta variable es aproximadamente la diferencia entre dos medias ponderadas de los gastos. La primera, da sobretodo peso a otros gastos (X_9) y transporte (X_6). En otros

gastos están incluidas las transferencias fuera de la provincia a miembros de la familia mayores de 14 años que no residan en ella. Es también significativo que estas provincias tienen altos gastos en transporte. La primera media ponderada puede considerarse un indicador de cómo ésta provincia envía recursos a otras. La segunda media da mayor peso a las variables enseñanza y cultura (X_7) y gastos sanitarios (X_5).

Tercer componente principal.

Los restantes valores propios de la matriz de covarianza de los datos de la EPF son 0.027, 0.0175, 0.0126, 0.0107, 0.010, 0.0059, y 0.00526. A partir del tercero son muy pequeños. El tercer componente principal es

$$Z_3 = 0.12X_1 + 0.05X_2 + 0.34X_3 + 0.11X_4 - 0.85X_5 + 0.04X_6 - 0.30X_7 + 0.20X_8 + 0.003X_9$$

$$Z_3 = (0.12X_1 + 0.05X_2 + 0.34X_3 + 0.11X_4 + 0.04X_6 + 0.20X_8 + 0.003X_9) - (0.85X_5 + 0.30X_7)$$

Y puede interpretarse como la diferencia entre dos medias ponderadas. La primera da sobretodo peso a las variables: (3, vivienda), (8, turismo y ocio), (1, alimentación) y (4, mobiliario doméstico).

1.2.6. Varianza de las componentes.

En el proceso de obtención de las componentes principales presentado anteriormente hemos visto que la varianza de la componente h-ésima es:

$$V(Z_h) = a'_h S a_h = \lambda_h \quad (25)$$

Es decir, la varianza de cada componente es igual al valor propio de la matriz S al que va asociada. Si, como es lógico, la medida de la variabilidad de las variables originales es suma de sus varianzas, dicha variabilidad será:

$$\sum_{h=1}^p V(X_h) = \text{traza}(S) \quad (26)$$

Ya que las varianzas de las variables son los términos que aparecen en la diagonal de la matriz de varianzas covarianzas S .

Ahora bien, como, S es una matriz real simétrica, por la teoría de diagonalización de matrices, existe una matriz ortogonal P ($P^{-1} = P'$) tal que $P'SP = D$, siendo D

diagonal con los valores propios de S ordenados de mayor a menor en la diagonal principal. Por lo tanto:

$$\text{traza}(P'SP) = \text{traza}(D) = \sum_{h=1}^p \lambda_h \quad (27)$$

Pero: $\text{traza}(P'SP) = \text{traza}(SPP') = \text{traza}(S.I) = \text{traza}(S)$

Con lo que ya podemos escribir:

$$\begin{aligned} \sum_{h=1}^p V(X_h) &= \text{traza}(S) = \text{traza}(P'SP) \\ &= \text{traza}(D) \\ &= \sum_{h=1}^p \lambda_h = \sum_{h=1}^p V(Z_h) \\ \sum_{h=1}^p V(Z_h) &= \sum_{h=1}^p \lambda_h \quad (28) \end{aligned}$$

Hemos comprobado, además, que la suma de las varianzas de las variables (inercia total de la nube de puntos) es igual a la suma de las varianzas de las componentes principales e igual a la suma de los valores propios de la matriz de varianzas covarianzas muestral S .

La proporción de la variabilidad total recogida por la componente principal h -ésima (porcentaje de inercia explicada por la componente principal h -ésima) vendrá dada por:

$$\frac{\lambda_h}{\sum_{h=1}^p \lambda_h} = \frac{\lambda_h}{\text{traza}(S)} \quad (29)$$

Si las variables están tipificadas, $S = R$ y $\text{traza}(S) = \text{traza}(R) = p$, con lo que la proporción de la componente h -ésima en la variabilidad total será λ_h/p .

También se define el porcentaje de inercia explicada por las k primeras componentes (o ejes factoriales) como:

$$\frac{\sum_{h=1}^k \lambda_h}{\sum_{h=1}^p \lambda_h} = \frac{\sum_{h=1}^k \lambda_h}{\text{traza}(S)} \quad (30)$$

1.2.7. Propiedades de los componentes.

Los componentes principales son nuevas variables con las propiedades siguientes:

1. Conservan la variabilidad inicial: la suma de las varianzas de los componentes es igual a la suma de las varianzas de las variables originales, y la varianza generalizada de los componentes es igual a la original. Comprobemos el primer punto. Como $Var(Z_h) = \lambda_h$ y la suma de los valores propios es la traza de la matriz:

$$tr(S) = Var(X_1) + \dots + Var(X_p) = \lambda_1 + \dots + \lambda_p \quad (31)$$

por tanto, $\sum_{i=1}^p Var(X_i) = \sum \lambda_i = \sum_{i=1}^p Var(Z_i)$. Las nuevas variables Z_i tienen conjuntamente la misma variabilidad que las variables originales.

Los componentes principales también conservan la *Varianza generalizada* (determinante de la matriz de covarianzas de las variables). Como el determinante es el producto de los valores propios, llamando S_z a la matriz de covarianzas de los componentes, que es diagonal con términos λ_i :

$$|S_x| = \lambda_1 \dots \lambda_p = \prod_{i=1}^p Var(Z_i) = |S_z| \quad (32)$$

2. La proporción de variabilidad explicada por un componente es el cociente entre su varianza, el valor propio asociado al vector propio que lo define, y la suma de los valores propios de la matriz. En efecto, la varianza del componente h es λ_h , y la suma de las varianzas de las variables originales es $\sum_{i=1}^p \lambda_i$, igual, como acabamos de ver, a la suma de las varianzas de los componentes. La proporción de variabilidad total explicada por la componente h es $\lambda_h / \sum \lambda_i$.
3. Las covarianzas entre cada componente principal y las variables X vienen dadas por el producto de las coordenadas del vector propio que define el componente por su valor propio:

$$Cov(Z_i; X_1, \dots, X_p) = \lambda_i a_i = (\lambda_i a_{i1}, \dots, \lambda_i a_{ip}) \quad (33)$$

Donde a_i , es el vector de coeficientes de la componente Z_i .

Para justificar este resultado, vamos a calcular la matriz $p \times p$ de covarianzas entre los componentes y las variables originales. Esta matriz es:

$$Cov(Z, X) = \frac{1}{n} Z'X \quad (34)$$

Y su primera fila proporciona las covarianzas entre la primera componente y las p variables originales. Como $Z = XA$, sustituyendo

$$Cov(Z, X) = \frac{1}{n} A'X'X = A' \left(\frac{1}{n} X'X \right) = A'S = DA' \quad (35)$$

Donde A contiene en columnas los vectores propios de S y D es la matriz diagonal de los valores propios. En consecuencia, la covarianza entre, por ejemplo, el primer componente principal y las p variables vendrá dada por la primera fila de $A'S$, es decir a'_1S o también $\lambda_1 a'_1$, donde a'_1 es el vector de coeficientes de la primera componente principal.

4. La correlación entre un componente principal y una variable X es proporcional al coeficiente de esa variable en la definición del componente, y el coeficiente de proporcionalidad es el cociente entre la desviación típica del componente y la desviación típica de la variable.

Para comprobarlo:

$$Corr(Z_i; X_j) = \frac{Cov(Z_i, X_j)}{\sqrt{Var(Z_i)Var(X_j)}} = \frac{\lambda_i a_{ij}}{\sqrt{\lambda_i s_j^2}} = a_{ij} \frac{\sqrt{\lambda_i}}{s_j} \quad (36)$$

5. Las k componentes principales ($k < p$) proporcionan la predicción lineal óptima con k variables del conjunto de variables X .

Esta afirmación puede demostrarse de dos formas. La primera demostrando que la mejor predicción lineal con k variables de las variables originales se obtiene utilizando las k primeras componentes principales. La segunda demostrando que la mejor aproximación de la matriz de datos que puede construirse con una matriz de rango k se obtiene construyendo esta matriz con los valores de los k primeros componentes principales.

6. Si estandarizamos los componentes principales, dividiendo cada uno por su desviación típica, se obtiene la estandarización multivariante de los datos originales.

Estandarizando los componentes Z por sus desviaciones típicas, se obtienen las nuevas variables

$$Y_c = ZD^{-1/2} = XAD^{-1/2} \quad (37)$$

Donde $D^{-1/2}$ es la matriz que contiene las inversas de las desviaciones típicas de las componentes. La estandarización multivariante de una matriz de variables X de media cero se define como:

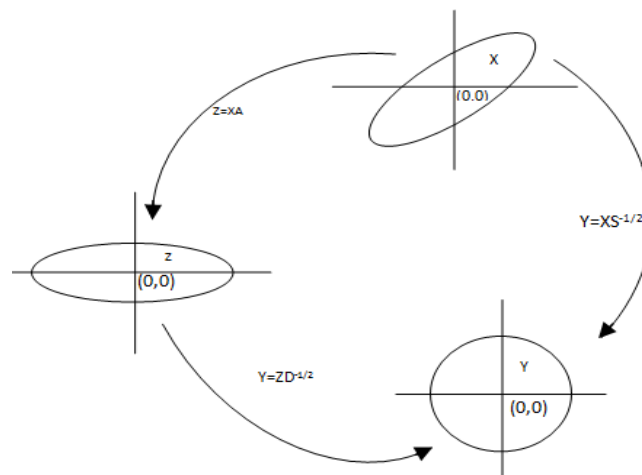
$$Y_s = XAD^{-1/2}A' \quad (38)$$

Tanto las variables Y_c , como las Y_s tienen matriz de covarianzas identidad, pero unas pueden ser una rotación de las otras. Esto no altera sus propiedades, y la estandarización puede interpretarse como:

- (1) Obtener los componentes principales;
- (2) Estandarizarlos para que tengan todos la misma varianza.

Esta relación se presenta gráficamente en la figura 1.3. La transformación mediante componentes principales conduce a variables incorreladas pero con distinta varianza. Puede interpretarse como rotar los ejes de la elipse que definen los puntos para que coincidan con sus ejes naturales. La estandarización multivariante produce variables incorreladas con varianza unidad, lo que supone buscar los ejes naturales y luego estandarizarlos. En consecuencia, si estandarizamos los componentes se obtiene las variables estandarizadas de forma multivariante.

Figura 1.3. Representación gráfica de la relación entre componentes principales y estandarización multivariante.



1.2.8. Análisis normado o con correlaciones.

Los componentes principales se obtienen maximizando la varianza de la proyección. En términos de las variables originales esto supone maximizar:

$$M = \sum_{i=1}^p a_i^2 s_i^2 + 2 \sum_{i=1}^p \sum_{j=i+1}^p a_i a_j s_{ij} \quad (39)$$

Con la restricción $a'a = 1$. Si alguna de las variables, por ejemplo la primera, tiene una varianza s_1^2 , mayor que las demás, la manera de aumentar M es hacer tan grande como podamos la coordenada a_1 asociada a esta variable. En el límite, si una variable tiene una varianza mucho mayor que las demás, el primer componente principal coincidirá muy aproximadamente con esta variable.

Cuando las variables tienen unidades distintas esta propiedad no es conveniente: si disminuimos la escala de medida de una variable cualquiera, de manera que aumente en magnitud sus valores numéricos (pasamos por ejemplo de medir en km a medir en metros), el peso de esa variable en el análisis aumentará, ya que en (39):

- (1) Su varianza será mayor y aumentará su coeficiente en el componente, a_i^2 , pues contribuye más a aumentar M ;
- (2) Sus covarianzas con todas las variables aumentarán, con el consiguiente efecto de incrementar a_i .

En resumen, cuando las escalas de medida de las variables son muy distintas, la maximización de (39) dependerá decisivamente de estas escalas de medida y las variables con valores más grandes tendrán más peso en el análisis. Si queremos evitar este problema, conviene estandarizar las variables antes de calcular los componentes, de manera que las magnitudes de los valores numéricos de las variables X sean similares.

La estandarización resuelve otro posible problema. Si las variabilidades de las X son muy distintas, las variables con mayor varianza van a influir más en la determinación de la primera componente. Este problema se evita al estandarizar las variables, ya que entonces las varianzas son la unidad, y las covarianzas son los coeficientes de correlación. La ecuación a maximizar se transforma en:

$$M' = 1 + 2 \sum_{i=1}^p \sum_{j=i+1}^p a_i a_j r_{ij} \quad (40)$$

Siendo r_{ij} el coeficiente de correlación lineal entre las variables i y j . en consecuencia, la solución depende de las correlaciones y no de las varianzas.

Los componentes principales normados se obtienen calculando los vectores y valores propios de la matriz R , de coeficientes de correlación. Llamando λ_p^R a las raíces características de esa matriz, que suponemos no singular, se verifica que:

$$\sum_{i=1}^p \lambda_i^R = \text{traza}(R) = p \quad (41)$$

Las propiedades de los componentes extraídos de R son:

1. La proporción de variación explicada por λ_p^R será:

$$\frac{\lambda_p^R}{p} \quad (42)$$

2. Las correlaciones entre cada componente Z_j y las variables X originales vienen dados directamente por $a'_j \sqrt{\lambda_j}$ (43) siendo $Z_j = Xa_j$.

Cuando las variables X originales están en distintas unidades conviene aplicar el análisis de la matriz de correlaciones o análisis normado. Cuando las variables tienen las mismas unidades, ambas alternativas son posibles. Si las diferencias entre las varianzas de las variables son informativas y queremos tenerlas en cuenta en el análisis, no debemos estandarizar las variables: por ejemplo, supongamos dos índices con la misma base pero no fluctúa mucho y el otro es casi constante. Este hecho es informativo, y para tenerlo en cuenta no se deben estandarizar las variables, de manera que el índice de mayor variabilidad tenga más peso. Por el contrario, si las diferencias de variabilidad no son relevantes se eliminan con el análisis normado. En caso de duda, conviene realizar ambos análisis, y seleccionar aquél que conduzca a conclusiones más informativas.

1.3. APLICACIÓN DEL MÉTODO ACP A PARTIR DE LA MATRIZ DE CORRELACIONES.

Estimar los componentes principales a partir de la matriz de correlaciones implica un análisis matricial de los datos tipificados, reduciendo la estructura de ellos en una misma unidad de escala. A continuación se presenta formalmente el procedimiento matemático y los algoritmos asociados a la estimación de las m componentes principales.

Sea $X = (X_{n1}, X_{n2}, \dots, X_{np})$ un conjunto de p variables respuestas que representan n observaciones seleccionadas de una población con media, \bar{X} y matriz de correlación R . La matriz de correlaciones R , derivada de la correlación lineal de cada par de variables originales, está dada por:

$$R = \begin{bmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{21} & 1 & \dots & r_{2p} \\ \dots & \dots & \dots & \dots \\ r_{p1} & r_{p2} & \dots & 1 \end{bmatrix} \quad (44)$$

En su forma matricial ampliada las componentes principales están definidas por y son una combinación lineal de las p variables originales, de tal forma que se puede representar como:

$$\begin{bmatrix} Z_1 \\ Z_2 \\ \dots \\ Z_p \end{bmatrix} = \begin{bmatrix} a_{11}X_1 & a_{12}X_2 & \dots & a_{1p}X_p \\ a_{21}X_1 & a_{22}X_2 & \dots & a_{2p}X_p \\ \dots & \dots & \dots & \dots \\ a_{p1}X_1 & a_{p2}X_2 & \dots & a_{pp}X_p \end{bmatrix} \Rightarrow Z = AX \quad (45)$$

En forma resumida, el primer componente estaría determinado por $Z_1 = a_1'(X - \mu)$, en donde a_1 se elige teniendo en cuenta que la varianza de $a_1'(X - \mu)$ se maximice sobre todos los vectores a_1 que satisfagan $a_1'a_1 = 1$.

Como la condición más importante que se le ha impuesto al primer componente es que maximice la varianza contenida en todos los datos, ésta ocurriría cuando a_1 es un autovector de R correspondiente al autovalor λ_1 y que satisface $a_1'a_1 = 1$. Esto implica maximizar la función objetivo, que es la varianza de:

$$V(Z_1) = \frac{\sum_{i=1}^n Z_{1i}^2}{n} = \frac{1}{n} Z_1' Z_1 = \frac{1}{n} (a_1 X)' (X a_1) = \frac{1}{n} a_1' X' X a_1 = a_1' \left[\frac{1}{n} X' X \right] a_1$$

donde $\left[\frac{1}{n} X' X \right]$ es equivalente a la matriz R de las variables estandarizadas. En este caso la expresión de la varianza del primer componente se describe como $V(Z_1) = a_1' R a_1$. Esta función se maximiza, sujeta a la restricción $a_1' a_1 = 1$ y resolviendo el siguiente lagrangiano:

$$L = a_1' R a_1 - \lambda (a_1' a_1 - 1).$$

Derivando con respecto a a_1 e igualando a cero se obtiene: $\frac{\partial L}{\partial a_1} = 2R a_1 - 2\lambda a_1 = 0$; $(R - \lambda I) a_1 = 0$. La solución de esta expresión permite derivar el autovector y el autovalor del primer componente a_1 y λ_1 respectivamente.

La segunda componente principal se define por $Z_2 = a_2' (X - \mu)$, donde a_2 se elige teniendo en cuenta que la varianza de $a_2' (X - \mu)$ se maximice entre todas las combinaciones lineales de X que no estén correlacionadas con la primera componente principal ($a_2' a_1 = 0$) y que satisfaga $a_2' a_2 = 1$. Dando como resultado el autovector a_2 y el autovalor λ_2 . Esto garantiza, como se ha indicado anteriormente, que el segundo componente capturará la mayor parte de la varianza restante en los datos, luego de obtener el primer componente.

En forma general, cada uno de los j -ésimos ($j=3,4,\dots,p$) componentes principales restantes se puede representar por $Z_j = a_j' (X - \mu)$ en donde a_j se elige teniendo en cuenta que la varianza de $a_j' (X - \mu)$, se maximice entre todas las combinaciones lineales de X que no estén correlacionadas con la $j-1$ componentes principales y que satisfagan $a_j' a_j = 1$.

De lo anterior se generaliza que los autovalores y autovectores de R se expresan así: $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ y $a_1 \geq a_2 \geq \dots \geq a_p$, respectivamente.

Para analizar la importancia de los componentes estimados, se parte de la varianza de la j -ésima componente (Z_j) que es λ_j . Si definimos la sumatoria de las λ_j , ello es equivalente a la Traza de la matriz de correlación $[Tr(R)]$, en el sentido que mide la variación total explicada por todas las variables componentes principales. Así, la

importancia de la j -ésima componente principal está definida por $R_j = \frac{\lambda_j}{Tr(P)}$ para $j = 1, 2, \dots, p$.

Por tanto se definen tres condiciones básicas:

1. Los vectores son estandarizados de tal forma que: $|a_1| = |a_2| = \dots = |a_p| = 1$

$$\begin{cases} a_i' a_i = 1 \\ a_i' a_j = 0; i \neq j \end{cases}$$

2. Los componentes Z_1, Z_2, \dots, Z_p no están correlacionados entre sí.
3. $V(ZZ_1) \geq V(Z_2) \geq \dots \geq V(Z_p)$, siendo $\sum_{i=1}^p V(X_i) = \sum_{h=1}^p V(Z_h)$.

1.4. ESTRUCTURA FACTORIAL DE LAS COMPONENTES PRINCIPALES

Se denomina estructura factorial de las componentes principales a la matriz de correlaciones entre las componentes Z_h y las variables originales X_j .

Consideremos los vectores muestrales relativos a X_j y Z_h respectivamente:

$$X_j = \begin{bmatrix} X_{j1} \\ X_{j2} \\ \dots \\ X_{jn} \end{bmatrix} \quad Z_h = \begin{bmatrix} Z_{h1} \\ Z_{h2} \\ \dots \\ Z_{hn} \end{bmatrix} \quad (46)$$

La covarianza muestral entre Z_h y X_j viene dada por:

$$Cov(X_j, Z_h) = \frac{1}{n} X_j' Z_h \quad (47)$$

El vector X_j se puede expresar en función de la matriz X utilizando el vector de orden p , al que denominamos por δ , que tiene un 1 en la posición j -ésima y 0 en las posiciones restantes. La forma de expresar X_j en función de la matriz X a través del vector p es la siguiente:

$$X'_j = \delta' X' = [0 \dots 1 \dots 0] \begin{bmatrix} X_{11} & \dots & X_{1i} & \dots & X_{1n} \\ \vdots & & \vdots & & \vdots \\ X_{j1} & \dots & X_{ji} & \dots & X_{jn} \\ \vdots & & \vdots & & \vdots \\ X_{p1} & \dots & X_{pi} & \dots & X_{pn} \end{bmatrix} \quad (48)$$

Teniendo en cuenta que $Z_h = Xa_h$ podemos escribir:

$$\begin{aligned} Cov(X_j, Z_h) &= \frac{1}{n} X'_j Z_h = \frac{1}{n} \delta' X' X a_h = \delta' \frac{1}{n} X' X a_h = \delta' S a_h = \delta' \lambda_h a_h = \lambda_h \delta' a_h \\ Cov(X_j, Z_h) &= \lambda_h a_{hj} \quad (49) \end{aligned}$$

Por lo tanto, podemos escribir la correlación existente entre la variable X_j y la componente Z_h de la siguiente forma:

$$r_{jh} = \frac{Cov(X_j, Z_h)}{\sqrt{V(X_j)}\sqrt{V(Z_h)}} = \frac{\lambda_h a_{hj}}{\sqrt{V(X_j)}\sqrt{\lambda_h}} \quad (50)$$

Si las variables originales están tipificadas, la correlación entre las variables X_j y la componente Z_h es la siguiente:

$$r_{jh} = \frac{\lambda_h a_{hj}}{\sqrt{V(X_j)}\sqrt{\lambda_h}} = \frac{\lambda_h a_{hj}}{\sqrt{\lambda_h}} = a_{hj} \sqrt{\lambda_h} \quad (51)$$

1.5. PUNTUACIONES O MEDICIÓN DE LAS COMPONENTES

El análisis en componentes principales es en muchas ocasiones un paso previo a otros análisis, en los que se sustituye el conjunto de variables originales por las componentes obtenidas. Por ejemplo, en el caso de estimación de modelos afectados de multicolinealidad o correlación serial (autocorrelación). Por ello, es necesario conocer los valores que toman las componentes en cada observación.

Una vez calculados los coeficientes a_{hj} (componentes del vector propio normalizado asociado al valor propio h-ésimo de la matriz $S = X'X/n$ relativo a la componente principal Z_{hj}), se pueden obtener las puntuaciones Z_{hj} , es decir, los valores de las componentes correspondientes a cada observación, a partir de la siguiente relación:

$$Z_{hj} = a_{h1}X_{1i} + a_{h2}X_{2i} + \dots + a_{hp}X_{pi} \quad h = 1 \dots p \quad i = 1 \dots n \quad (52)$$

Si las componentes se dividen por su desviación típica se obtienen las componentes tipificadas. Por lo tanto, si llamamos Y_h a la componente Z_h tipificada tenemos:

$$Y_h = \frac{Z_h - E(Z_h)}{\sqrt{V(Z_h)}} = \frac{Z_h}{\sqrt{\lambda_h}} \quad h = 1 \dots p \quad (53)$$

Por lo tanto, las puntuaciones tipificadas serán:

$$\frac{Z_{hi}}{\sqrt{\lambda_h}} = \frac{a_{h1}}{\sqrt{\lambda_h}} X_{1i} + \frac{a_{h2}}{\sqrt{\lambda_h}} X_{2i} + \dots + \frac{a_{hp}}{\sqrt{\lambda_h}} X_{pi} \quad h = 1 \dots p \quad i = 1 \dots n \quad (54)$$

Expresión que puede escribirse como:

$$Y_{hi} = c_{h1} X_{1i} + c_{h2} X_{2i} + \dots + c_{hp} X_{pi} \quad c_{hi} = \frac{a_{hi}}{\sqrt{\lambda_h}} \quad h = 1 \dots p \quad i = 1 \dots n \quad (55)$$

La matriz formada por los coeficientes c_{hi} suele denominarse matriz de coeficientes de puntuaciones de los factores (factor score coefficient matrix).

1.6. CALIFICACIÓN DE COMPONENTES PRINCIPALES.

Una vez estimados los componentes principales, su utilidad en el análisis estadístico está asociada a que su calificación pueda dar contenido a la variable latente que se intenta explicar. Entonces el procedimiento para calificar cada una de las unidades de observación podemos obtenerlo siguiendo el análisis anterior de componentes principales, de donde, si X_r es el vector de variables observadas para la r -ésima unidad de observación, entonces, la calificación de la j -ésima unidad de observación viene dada por: $Z_{rj} = a'_j(X_r - \mu)$, para $j = 1, 2, \dots, p$ y $r = 1, 2, \dots, n$.

Esta calificación proporciona un valor ordinal de las unidades de observación, en la medida en que indican su ubicación relativa en un conjunto de datos o variables con respecto a los ejes componentes principales que se han definido.

1.7. INTERPRETACIÓN DE LOS COMPONENTES.

Componentes de tamaño y forma.

Cuando existe una alta correlación positiva entre todas las variables, el primer componente principal tiene todas sus coordenadas del mismo signo y puede interpretarse como un promedio ponderado de todas las variables, o un factor global de “tamaño”. Los restantes componentes se interpretan como factores “de forma” y típicamente tienen coordenadas positivas y negativas, que implica que contraponen unos grupos de variables frente a otros. Estos factores de forma pueden frecuentemente escribirse como medias ponderadas de dos grupos de variables con distinto signo y contraponen las variables de un signo a las del otro.

Ejemplo 1.4: *El segundo componente principal de los datos del ejemplo 1.2, puede escribirse aproximadamente, despreciando los coeficientes pequeños (menores que 0.1):*

$$Z_2 = (0.05X_1 + 0.16X_2 + 0.07X_4 + 0.29X_6 + 0.78X_9) \\ - (0.17X_3 + 0.21X_5 + 0.40X_7 + 0.17X_8) \cong I_O - I_S$$

Donde:

$$I_O = 0.16X_2 + 0.29X_6 + 0.78X_9$$

Es un indicador de los gastos de transporte y transferencias a otras provincias y

$$I_S = 0.17X_3 + 0.21X_5 + 0.40X_7 + 0.17X_8$$

Es un indicador de gastos en servicios (educación y sanidad). Además, cuando las variables van en logaritmos, los componentes suelen poder escribirse como ratios de promedios geométricos de las variables.

La interpretación de los componentes se simplifica suponiendo que los coeficientes pequeños son cero y redondeando los coeficientes grandes para expresar el componente como cocientes, diferencias o sumas entre variables. Estas aproximaciones son razonables si modifican poco la estructura del componente y mejoran su interpretación. Una medida del cambio introducido al modificar un vector propio de a_i , a a_iM es el cambio en la proporción de variabilidad explicada por el componente. Si el valor propio

asociado a a_i es λ_i , el componente explica el $\lambda_i/\sum \lambda_j$ de la variabilidad. Si ahora modificamos el vector a a_{iM} , la varianza de la proyección de los datos sobre este componente es $\lambda_{iM} = a'_{iM} S a_{iM} = (\tilde{X} a_{iM})' (\tilde{X} a_{iM})/n$, la varianza del componente, y la proporción de variabilidad explicada será $\lambda_{iM}/\sum \lambda_j$. El cambio relativo será $(\lambda_i - \lambda_{iM})/\lambda_i$, ya que siempre $\lambda_i \geq \lambda_{iM}$, y si este cambio es pequeño, está justificada la modificación si favorece la interpretación.

1.8. SELECCIÓN DEL NÚMERO DE COMPONENTES.

En general, el objetivo de la aplicación de los componentes principales es reducir las dimensiones de las variables originales, pasando de p variables originales a $m < p$ componentes principales. El problema que se plantea es como fijar m , o dicho de otra forma, ¿Qué número de componentes se deben retener?. Aunque para la extracción de componentes principales no hace falta plantear un modelo estadístico previo, alguno de los criterios para determinar cuál debe ser el número óptimo de componentes a retener requiere la formulación previa de hipótesis estadísticas.

Se sugieren distintos criterios para seleccionar el número de componentes:

1.8.1. Criterio de la media aritmética.

Según este criterio se selecciona aquellas componentes cuya raíz característica λ_j exceda de la media de las raíces características. Recordemos que la raíz característica asociada a una componente es precisamente su varianza.

Analíticamente este criterio implica retener todas aquellas componentes en que se verifique que:

$$\lambda_h > \bar{\lambda} = \frac{\sum_{j=1}^p \lambda_h}{p} \quad (56)$$

Si se utilizan variables tipificadas, entonces, como ya se ha visto, se verifica que $\sum_{j=1}^p \lambda_h = p$, con lo que para variables tipificadas se retiene aquellas componentes tales que $\lambda_h > 1$.

Dicho de otra manera, se trata de desechar aquellos componentes asociados a valores propios inferiores a una cota, que suele fijarse como la varianza media, $\sum \lambda_i/p$. En particular, cuando se trabaja con la matriz de correlación, el valor medio de los componentes es 1, y esta regla lleva a seleccionar los valores propios mayores que la unidad. De nuevo esta regla es arbitraria: una variable que sea independiente del resto suele llevarse un componente principal y puede tener un valor propio mayor que la unidad. Sin embargo, si está incorrelada con el resto puede ser una variable poco relevante para el análisis, y no aportar mucho a la comprensión del fenómeno global.

1.8.2. El gráfico de sedimentación.

Realizar un gráfico de sedimentación, de λ_i frente a i . Comenzar seleccionando componentes hasta que los restantes tengan aproximadamente el mismo valor de λ_i . La idea es buscar un “codo” en el gráfico, es decir, un punto a partir del cual los valores propios son aproximadamente iguales. El criterio es quedarse con un número de componentes que excluya los asociados a valores pequeños y aproximadamente del mismo tamaño. Lo que se representa son las combinaciones de $(1, \lambda_1), (2, \lambda_2), \dots, (p, \lambda_p)$. Se supone que cuando la gráfica tiende a nivelarse, sus autovalores tienden a cero y su explicación como un componente adicional probablemente pueda estar midiendo ruido aleatorio. Por ello se debe contar con un buen criterio.

El gráfico de sedimentación se obtiene al representar en ordenadas las raíces características y en abscisas los números de componentes principales correspondientes a cada raíz características en orden decreciente. Uniendo todos los puntos se obtiene una figura que, en general, se parece al perfil de una montaña con una pendiente fuerte hasta llegar a la base, formada por una meseta con una ligera inclinación. Continuando con el símil de la montaña, en esa meseta es donde se acumula los guijarros caídos desde la cumbre, es decir, donde se sedimenta. Su denominación en inglés es scree plot. De acuerdo con el criterio gráfico se retiene todas aquellas componentes previas a la zona de sedimentación.

1.8.3. Variabilidad explicada

Seleccionar componentes hasta cubrir una proporción determinada de varianza, como el 80 o el 90 por 100. Esta regla es arbitraria y debe aplicarse con cierto cuidado. Por ejemplo, es posible que un único componente de “tamaño” recoja el 90 por 100 de la variabilidad y, sin embargo, pueden existir otros componentes que sean muy adecuados para explicar la “forma” de las variables. Para el caso de estudios que toman muestras de laboratorio, se espera reunir del 90% al 95% de la variabilidad total con pocos componentes. Sin embargo para datos de investigación social, es posible que se requieran más de tres componentes principales para explicar entre el 70% y 75% de la variación total.

Cuando los componentes principales son estimados a través de la matriz de correlación, el criterio a elegir, además de los anotados antes, es el que la dimensión real de los datos es igual al número de autovalores mayor o igual a 1. La lógica de este criterio es que cuando se usan variables tipificadas para estimar los componentes principales, la varianza explicada de ellos no puede ser inferior a la que resulta de las variables tipificadas, que en este caso es igual a la unidad.

1.9. REPRESENTACIÓN GRÁFICA.

La interpretación de los componentes principales se favorece representando las proyecciones de las observaciones sobre un espacio de dimensión dos, definido por parejas de los componentes principales más importantes. Este punto se ha ilustrado en los ejemplos anteriores, donde se ha indicado que la proyección de cualquier observación sobre un componente es directamente el valor del componente para esa observación. La representación habitual es tomar dos ejes ortogonales que representen los dos componentes considerados, y situar cada punto sobre ese plano por sus coordenadas con relación a estos ejes, que son los dos primeros componentes, las coordenadas del punto \mathbf{x}_i son $z_{1i} = \mathbf{a}'_1 \mathbf{x}_i$ y $z_{2i} = \mathbf{a}'_2 \mathbf{x}_i$.

La interpretación se favorece representando en el mismo plano, además de las observaciones, las variables originales. Esto puede hacerse utilizando como

coordenadas su coeficiente de correlación con cada uno de los ejes. El vector de correlaciones entre el primer componente y las variables originales viene dado por $\lambda_1^{1/2} a'_1 D$, donde D es una matriz diagonal cuyos términos son las inversas de las desviaciones típicas de cada variable. La matriz de correlaciones R_{cv} entre los p componentes y las p variables tendrán como filas los términos $\lambda_j^{1/2} a'_j D$ y puede escribirse

$$R_{cv} = {}^{1/2}AD \quad (57)$$

Donde A es la matriz de vectores propios, ${}^{1/2}$ es la matriz diagonal con términos $\sqrt{\lambda_i}$ y en el análisis normado como las variables se estandarizan a varianza unidad las correlaciones serán simplemente ${}^{1/2}A$.

Conviene investigar si transformando las variables se obtiene una interpretación más simple. Como regla general, cuando al tomar logaritmos las variables X tienen una distribución aproximadamente simétrica, conviene realizar el análisis de componentes principales sobre los logaritmos de las variables.

Es importante recordar que las covarianzas (o correlaciones) miden únicamente las relaciones lineales entre las variables. Cuando entre ellas existan relaciones fuertes no lineales el análisis de componentes principales puede dar una información muy parcial de las variables.

1.10. DATOS ATÍPICOS

Antes de obtener los componentes principales conviene asegurarse de que no existen datos atípicos, ya que, como es sabido, los atípicos pueden distorsionar totalmente la matriz de covarianzas. Para ilustrar su efecto sobre los componentes, supongamos el caso más simple en que un error de medida introduce un valor atípico grande en la primera variable. Su efecto será aumentar mucho la varianza de esta variable y disminuir las covarianzas con las restantes, con lo que, si hacemos el atípico muy grande, la matriz S será, aproximadamente:

$$\begin{bmatrix} \sigma_1^2 & 0' \\ 0 & S_{22} \end{bmatrix} \quad (58)$$

Donde $0' = (0, 0, \dots, 0)$. Esta matriz tiene un vector propio $(1, 0, \dots, 0)$ unido al valor propio σ_1^2 y, si σ_1^2 es muy grande, éste será el primer componente principal. Por tanto, un valor atípico suficientemente grande distorsiona todos los componentes que podemos obtener de la matriz afectada.

El resultado anterior sugiere que las componentes principales podrían utilizarse para detectar datos atípicos multivariantes, ya que un valor muy extremo se llevará un componente principal y aparecerá como extremo sobre esta componente. Desgraciadamente, aunque los componentes pueden identificar atípicos aislados, no hay garantía de que funcionen cuando existen grupos de atípicos, debido al problema de enmascaramiento. Por esta razón, conviene utilizar para detectarlos el método basado en proyecciones sobre las direcciones extremas de kurtosis, que al ser capaz de identificar todos los posibles atípicos permite calcular la matriz de covarianzas libre de distorsiones graves.

1.11. MATRIZ DE CARGAS FACTORIALES

La dificultad en la interpretación de los componentes radica en la necesidad de que tengan sentido y midan algo útil en el contexto del fenómeno estudiado. Por tanto, es indispensable considerar el peso que cada variable original tiene dentro del componente elegido, así como las correlaciones existentes entre variables y factores. Un componente es una función de todas las variables, pero puede estar muy bien correlacionado con algunas de ellas, y menos con otras. Ya hemos visto que el coeficiente de correlación entre una componente y una variable se calcula multiplicando el peso de la variable en esa componente por la raíz cuadrada de su valor propio:

$$r_{jh} = u_{hj}\sqrt{\lambda_h}$$

Se demuestra también que estos coeficientes r representan la parte de varianza de cada variable que explica cada factor. De este modo, cada variable puede ser representada como una función lineal de los k componentes retenidos, donde los pesos o cargas de cada componente o factor (cargas factoriales) en la variable coinciden con los coeficientes de correlación. El cálculo matricial permite obtener de forma inmediata la tabla de coeficientes de correlación variables-componentes ($p \times k$) que se denominan

matriz de cargas factoriales. Las ecuaciones de las variables en función de las componentes (factores), traspuestas las inicialmente planteadas, son de mayor utilidad en la interpretación de los componentes y se expresan como sigue:

$$\begin{array}{ccc}
 Z_1 = r_{11}X_1 + \dots + r_{1p}X_p & & X_1 = r_{11}Z_1 + \dots + r_{k1}Z_k \\
 Z_2 = r_{21}X_1 + \dots + r_{2p}X_p & \Rightarrow & X_2 = r_{12}Z_1 + \dots + r_{k2}Z_k \\
 \dots & & \dots \\
 Z_k = r_{k1}X_1 + \dots + r_{kp}X_p & & X_p = r_{1p}Z_1 + \dots + r_{kp}Z_k
 \end{array}$$

Por las propiedades del coeficiente de correlación se deduce que la suma en horizontal de los cuadrados de las cargas factoriales de una variable en todos los factores (componentes) retenidos es la parte de dispersión total de la variable explicada por el conjunto de k componentes. Esta suma de cuadrados se denomina **comunalidad**. Por ejemplo, para la primera variable, la comunalidad será $r_{11}^2 + \dots + r_{k1}^2 = V(X_1) = h_1^2$. Por consiguiente, la suma de las comunalidades de todas las variables representa la parte de inercia global de la nube original explicada por los k factores retenidos y coincide con la suma de los valores propios de estas componentes. La comunalidad proporciona un criterio de calidad de la representación de cada variable, de modo que, variables totalmente representadas tienen de comunalidad la unidad. También se demuestra que la suma en vertical de los cuadrados de las cargas factoriales de todas las variables en un componente es su valor propio. Por ejemplo, el valor propio del primer componente será $r_{11}^2 + \dots + r_{1p}^2 = \lambda_1$.

Es evidente que, al ser las cargas factoriales los coeficientes de correlación entre variables y componentes, su empleo hace comparables los pesos de cada variable en la componente y facilita su interpretación. En este mismo sentido, su representación gráfica puede orientar al investigador en una primera aproximación a la interpretación de los componentes. Como es lógico, esta representación sobre un plano sólo puede contener los factores de dos en dos, por lo que se pueden realizar tantos gráficos como parejas de factores retenidos. Estos gráficos se denominan **círculos de correlación** y están formados por puntos que representan cada variable por medio de dos coordenadas que miden los coeficientes de correlación de dicha variable con los dos factores o

componentes considerados. Todas las variables estarán contenidas dentro de un círculo de radio unidad.

1.12. ROTACIÓN DE LOS COMPONENTES

Es frecuente no encontrar interpretaciones verosímiles a los factores (componentes) obtenidos, ya que se ha organizado el estudio partiendo de un primer componente principal que condensaba la máxima inercia de la nube.

Sin embargo, no tiene por qué coincidir esta máxima inercia del primer factor, que condiciona el cálculo de los restantes, con la óptima interpretación de cada uno de los componentes. Sería deseable, para una más fácil interpretación, que cada componente estuviera muy bien relacionada con pocas variables (coeficientes de correlación r próximos a 1 ó -1) y mal con las demás (r próximos a 0). Esta optimización se obtiene por una adecuada **rotación de los ejes** que definen los componentes principales. Sería tolerable sacrificar la no coincidencia del primero de ellos con el eje principal del elipsoide, si así mejorara la interpretabilidad del conjunto. Rotar un conjunto de componentes no cambia la proporción de inercia total explicada, como tampoco cambia las comunalidades de cada variable, que no son sino la proporción de varianza explicada por todos ellos. Sin embargo, los coeficientes, que dependen directamente de la posición de los componentes respecto a las variables originales (cargas factoriales y valores propios), se ven alterados por la rotación.

Se puede comprobar que la rotación así descrita equivale a rotar los componentes en el círculo de correlación. Por ejemplo, en el Gráfico anterior puede observarse cómo una rotación que acercara la primera componente a las variables x_1 y x_2 conseguiría una proyección (carga factorial) máxima sobre ella de dichas variables y mínima en la componente 2 que se mantendría perpendicular al ser arrastrada en el giro. Los nuevos ejes rotados tendrán una correlación con los correspondientes primitivos, que será menor cuanto mayor sea el ángulo de giro.

Existen varios tipos de rotaciones, las más utilizadas son la rotación VARIMAX y la QUARTIMAX. La rotación VARIMAX se utiliza para conseguir que cada componente rotado (en vertical, en la matriz de cargas factoriales) presente altas correlaciones sólo con unas cuantas variables, rotación a la que suele aplicarse la llamada normalización de

Kaiser para evitar que componentes con mayor capacidad explicativa, que no tienen por qué coincidir con la mejor interpretabilidad, pesen más en el cálculo y condicionen la rotación. Esta rotación, la más frecuentemente utilizada, es adecuada cuando el número de componentes es reducido. La rotación QUARTIMAX se utiliza para conseguir que cada variable (en horizontal, en la matriz de cargas factoriales) tenga una correlación alta con muy pocos componentes cuando es elevado el número de éstos. Tanto Varimax como Quartimax son **rotaciones ortogonales**, es decir, que se mantiene la condición de perpendicularidad entre cada uno de los ejes rotados. Sin embargo, cuando las componentes, aún rotadas ortogonalmente, no presentan una clara interpretación, cabe todavía la posibilidad de intentar mejorarla a través de **rotaciones oblicuas**, que no respetan la perpendicularidad entre ellos. Piénsese en espacios multidimensionales para comprender la complejidad de los cálculos necesarios. De entre las diversas rotaciones oblicuas desarrolladas, la PROMAX, aplicada normalmente sobre una VARIMAX previa, es la más utilizada dada su relativa simplicidad.

En las soluciones oblicuas varían, lógicamente, no sólo los valores propios sino también las comunalidades de las variables y se mantiene, por supuesto, la varianza explicada por el modelo. Además, es importante tener en cuenta que la no perpendicularidad entre los ejes surgida tras una rotación oblicua produce una correlación entre ellos antes existente, por lo que parte de varianza de una variable explicada por una componente no es ya independiente de los demás factores. Deberá valorarse esta relación en la interpretación de los componentes. No se puede decir que una rotación sea mejor que otra, ya que desde un punto de vista estadístico todas son igualmente buenas. La elección entre diferentes rotaciones se basa en criterios no estadísticos, ya que la rotación preferida es aquella más fácilmente interpretable. Si dos rotaciones proponen diferentes interpretaciones no deben ser consideradas discordantes sino como dos enfoques diferentes de un mismo fenómeno que el investigador deberá realizar. La interpretación de una componente es un proceso subjetivo al que la rotación puede restar parte de subjetividad.

1.13. EJEMPLOS USANDO SPSS:

Es importante familiarizarnos con los paquetes computarizados para desarrollar cálculos y efectuar análisis, ya que en la práctica es por medio de ellos que resolvemos los problemas. Es así que a continuación presentamos un par de ejemplos usando SPSS.

Ejemplo 1.5: Usaremos los datos tabla A.4 Empresas que corresponde a empresas de 9 sectores de actividad para 26 distintos países. Se trata de realizar un análisis en componentes principales de todas las variables con la finalidad de reducirlas a un conjunto menor de variables con la menor pérdida de información posible.

Comenzamos eligiendo en los menús *Analizar//Reducción de datos//Análisis factorial*, y seleccionando las variables y las especificaciones para el análisis (figura 1.4), los botones *Descriptivos*, *Extracción*, *Rotación*, *Puntuaciones* y *Opciones* se rellenan como se indica en las figuras 1.5 a la 1.19.

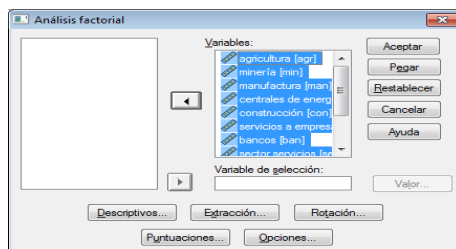


Figura 1.4

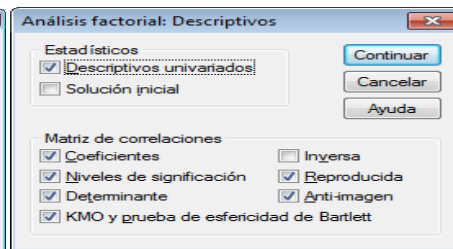


Figura 1.5

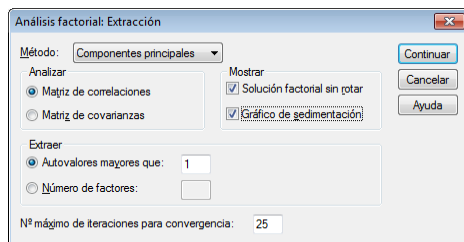


Figura 1.6

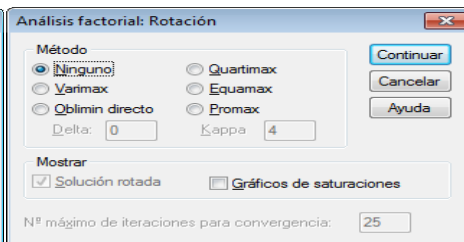


Figura 1.7

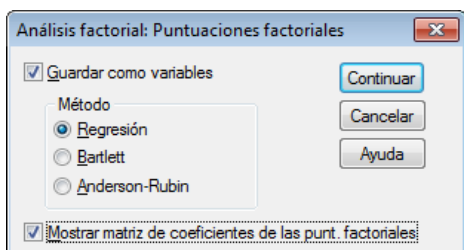


Figura 1.8

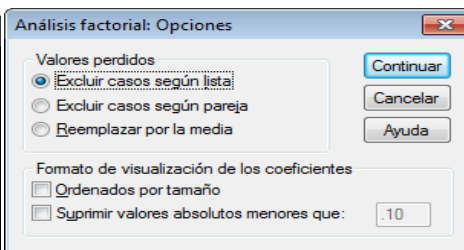


Figura 1.9

Obtenemos los siguientes resultados:

	Media	Desviación típica	N del análisis
agricultura	19,131	15,5466	26
minería	1,254	,9700	26
manufactura	27,008	7,0078	26
centrales de energía	,908	,3762	26
construcción	8,165	1,6456	26
servicios a empresas	12,958	4,5753	26
bancos	4,000	2,8066	26
sector servicios	20,023	6,8295	26
transporte y comunicaciones	6,546	1,3915	26

Figura 1.10

	Agri	Min	Man	Cen	Con	Ban	Secser	Tc
Correlación								
agricultura	1.000	.036	-.671	-.400	-.538	-.220	-.747	-.565
minería	.036	1.000	.445	.405	-.026	-.443	-.281	.157
manuf actura	-.671	.445	1.000	.385	.494	-.156	.154	.351
centrales de energía	-.400	.405	.385	1.000	.060	.110	.132	.375
construcción	-.538	-.026	.494	.060	1.000	.016	.158	.388
servicios a empresas	-.737	-.397	.204	.202	.356	.366	.572	.188
bancos	-.220	-.443	-.156	.110	.016	1.000	.108	-.246
sector serv icios	-.747	-.281	.154	.132	.158	.108	1.000	.568
transporte y comunicaciones	-.565	.157	.351	.375	.388	-.246	.568	1.000
Sig. (Unilateral)								
agricultura		.431	.000	.021	.002	.140	.000	.001
minería	.431		.011	.020	.451	.012	.082	.222
manuf actura	.000	.011		.026	.005	.224	.226	.040
centrales de energía	.021	.020	.026		.386	.297	.260	.029
construcción	.002	.451	.005	.386		.469	.220	.025
servicios a empresas	.000	.022	.159	.161	.037	.033	.001	.179
bancos	.140	.012	.224	.297	.469		.300	.113
sector serv icios	.000	.082	.226	.260	.220	.300		.001
transporte y comunicaciones	.001	.222	.040	.029	.025	.113	.001	

a. Determinante = 2.38E-006

Figura 1.11

La tabla de la figura 1.10 nos muestra los estadísticos univariados correspondientes, en la figura 1.11 tenemos la matriz de correlaciones cuyo determinante es $2,38 \cdot 10^{-6}$ que al ser muy pequeño indica que el grado de intercorrelación entre las variables es muy alto, condición inicial que debía cumplir.

KMO y prueba de Bartlett

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		.134
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	274.053
	gl	36
	Sig.	.000

Figura 1.12

Otro resultado importante es el de la figura 1.12, test de esfericidad de Barlett que permite contrastar formalmente la existencia de correlación entre las variables. Como en este caso su p-valor es 0,000, se puede concluir que existe correlación significativa entre las variables. Esta tabla incluye el estadístico KMO, cuyo valor tan pequeño (alejado de la unidad) indica una mala adecuación de la muestra a este análisis.

Luego tenemos la matriz de correlaciones anti-imagen formada por los coeficientes de correlación parcial entre cada par de variables cambiada de signo. Estos coeficientes deben ser bajos para que las variables compartan factores comunes. Los elementos de la diagonal de esta matriz son similares al estadístico KMO para cada par de variables e interesa que estén cercanos a la unidad. Observando la matriz (figura 1.13) vemos que no tenemos buenos resultados:

Matrices anti-imagen

		Agri	Min	Man	Cen	Con	Ser	Ban	Secser	Tc
Covarianza anti-imagen	agricultura	7E-005	.001	.000	.002	.001	.000	.000	.000	.001
	minería	.001	.018	.002	.035	.011	.004	.006	.003	.012
	manufactura	.000	.002	.000	.005	.001	.001	.001	.000	.002
	centrales de energía	.002	.035	.005	.096	.023	.008	.013	.005	.024
	construcción	.001	.011	.001	.023	.007	.002	.004	.002	.007
	servicios a empresas	.000	.004	.001	.008	.002	.001	.001	.001	.003
	bancos	.000	.006	.001	.013	.004	.001	.002	.001	.004
	sector servicios	.000	.003	.000	.005	.002	.001	.001	.000	.002
	transporte y comunicaciones	.001	.012	.002	.024	.007	.003	.004	.002	.009
Correlación anti-imagen	agricultura	.235 ^a	.975	1.000	.892	.993	.999	.998	.999	.987
	minería	.975	.101 ^a	.972	.826	.971	.977	.978	.975	.963
	manufactura	1.000	.972	.140 ^a	.890	.991	.998	.998	.999	.987
	centrales de energía	.892	.826	.890	.100 ^a	.903	.884	.879	.895	.847
	construcción	.993	.971	.991	.903	.099 ^a	.990	.990	.994	.971
	servicios a empresas	.999	.977	.998	.884	.990	.155 ^a	.997	.998	.989
	bancos	.998	.978	.998	.879	.990	.997	.060 ^a	.997	.989
	sector servicios	.999	.975	.999	.895	.994	.998	.997	.151 ^a	.983
	transporte y comunicaciones	.987	.963	.987	.847	.971	.989	.989	.983	.136 ^a

a. Medida de adecuación muestral

Figura 1.13

Ahora analizaremos el número de componentes con el que nos quedaremos, que normalmente son las relativas a valores propios mayores que la unidad. Observando la tabla de la varianza total explicada (figura 1.14) vemos que la primera componente explica un 38.7% de la varianza total y las dos siguientes un 23.6% y un 12.2% (un 74.6% entre las tres). El gráfico de sedimentación (figura 1.15) muestra que sólo hay tres componentes con autovalor mayor que 1.

Componente	Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado
1	3.487	38.746	38.746
2	2.130	23.669	62.415
3	1.099	12.211	74.625

Método de extracción: Análisis de Componentes principales.

Figura 1.14

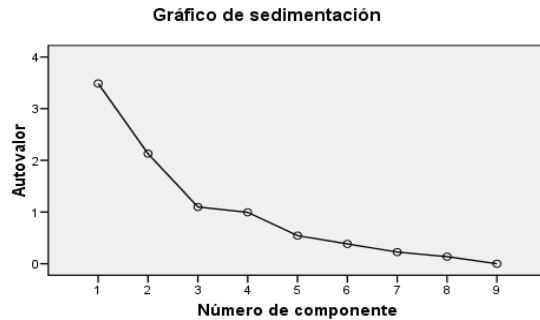


Figura 1.15

Ahora veamos cómo utilizar las variables que se agrupan en torno a cada componente y cuya combinación lineal define la componente. Para ello representamos cada una de las nueve variables por medio de los tres factores extraídos utilizando la matriz de componentes (Figura 1.16). Podemos escribir lo siguiente:

$$\text{Agricultura} = -0.978F_1 + 0.078F_2 - 0.510F_3$$

$$\text{Minería} = -0.020F_1 + 0.902F_2 + 0.211F_3$$

...

$$\text{T. y comu} = 0.0685F_1 + 0.296F_2 - 0.393F_3$$

	Componente		
	1	2	3
agricultura	-.978	.078	-.051
minería	-.002	.902	.211
manuf actura	.649	.518	.158
centrales de energía	.478	.381	.588
construcción	.607	.075	-.161
servicios a empresas	.708	-.511	.121
bancos	.139	-.662	.616
sector servicios	.723	-.323	-.327
transporte y comunicaciones	.685	.296	-.393

Método de extracción: Análisis de componentes principales.

a. 3 componentes extraídos

Figura 1.16

	Extracción
agricultura	.965
minería	.858
manufactura	.714
centrales de energía	.719
construcción	.400
servicios a empresas	.776
bancos	.837
sector servicios	.735
transporte y comunicaciones	.711

Método de extracción: Análisis de Componentes principales.

Figura 1.17

Para ver qué variables se agrupan en cada componente (factor) hay que observar las variables cuyas cargas sean altas en un factor y bajas en otros (valores menores que 0.25 suelen considerarse bajos).

En la primera componente está representada claramente Agricultura y en la segunda Minería (sus valores en la matriz de componentes son muy altos). Sector Servicios y Transporte y Comunicaciones están representadas en las tres componentes, lo mismo que Centrales de energía. Servicios a empresas y Manufacturas están representadas en la primera y segunda componentes y Bancos en la segunda y la tercera. Se observa entonces que es difícil agrupar las variables en componentes con lo que procedería realizar una rotación, que se realizará posteriormente. Es importante observar que las sumas de los cuadrados de los elementos de las columnas de la matriz de componentes es igual a los valores propios significativos.

A continuación se analiza la comunalidad (figura 1.17) de cada variable (suma de cuadrados de sus cargas factoriales definidas en la matriz de componentes) después de la extracción de los factores (componentes). La comunalidad es la parte de variabilidad de cada variable explicada por los factores. Antes de la extracción de los factores la comunalidad de cada variable es la unidad, e interesa que después de la extracción siga siendo alta.

Correlaciones reproducidas

	Agri	Min	Man	Cen	Con	Ser	Ban	Serser	Tc	
Correlación reproducida	agricultura	.965 ^b	.062	-.602	-.467	-.580	-.738	-.219	-.716	-.627
	minería	.062	.858 ^b	.499	.466	.032	-.437	-.468	-.362	.182
	manuf actura	-.602	.499	.714 ^b	.600	.407	.213	-.156	.250	.536
	centrales de energía	-.467	.466	.600	.719 ^b	.224	.214	.176	.030	.208
	construcción	-.580	.032	.407	.224	.400 ^b	.372	-.064	.468	.501
	servicios a empresas	-.738	-.437	.213	.214	.372	.776 ^b	.511	.638	.286
	bancos	-.219	-.468	-.156	.176	-.064	.511	.837 ^b	.113	-.343
	sector servicios	-.716	-.362	.250	.030	.468	.638	.113	.735 ^b	.529
	transporte y comunicaciones	-.627	.182	.536	.208	.501	.286	-.343	.529	.711 ^b
	Residual ^a		-.026	-.069	.067	.042	.001	-.001	-.031	.062
	agricultura	-.026	-.054	-.061	-.058	.041	.025	.081	-.025	
	minería	-.069	-.054	-.215	.087	-.010	E-.005	-.096	-.185	
	manuf actura	.067	-.061	-.215	-.164	-.012	-.066	.102	.167	
	centrales de energía	.042	-.058	.087	-.164	-.016	.080	-.309	-.114	
	construcción	.001	.041	-.010	-.012	-.016	-.145	-.066	-.099	
	servicios a empresas	-.001	.025	7E-005	-.066	.080	-.145	-.006	.097	
	bancos	-.031	.081	-.096	.102	-.309	-.066	-.006	.039	
	sector servicios	.062	-.025	-.185	.167	-.114	-.099	.097	.039	
	transporte y comunicaciones									

Método de extracción: Análisis de Componentes principales.

a. Los residuos se calculan entre las correlaciones observadas y reproducidas. Hay 22 (61.0%) residuales no redundantes con valores absolutos mayores que 0,05.

b. Comunalidades reproducidas

Figura 1.18

Es posible calcular los coeficientes de correlación entre cada dos variables después de que estén en función de las componentes, denominados *coeficientes de correlación reproducidas*. Estos coeficientes de correlación reproducidos no tienen por qué coincidir con los de la matriz de correlaciones inicial, pero no deben diferenciarse en más de 0.05 (residuos menores que 0.05), porque entonces la bondad del modelo factorial será discutible. En la figura 1.18 se presenta la *matriz de correlaciones reproducidas* en la que se observa que un 61% de los errores son mayores que 0.05, lo que indica que la bondad del modelo es discutible.

Ahora calcularemos las puntuaciones factoriales, que no son más que los valores que toman cada uno de los individuos en las tres componentes seleccionadas. Serán entonces tres variables sustitutas de las iniciales que representan su reducción y que recogen el 74.6% de la variabilidad total. Estas variables son las que se utilizarán como sustitutas de las iniciales para análisis posteriores tales como el análisis de la regresión con problemas de multicolinealidad y el análisis cluster. SPSS incorpora estas variables al conjunto de datos si así se le pide en la figura 1.8.

Matriz de coeficientes para el cálculo de las puntuaciones en las componentes

	Componente		
	1	2	3
agricultura	-.280	.037	-.046
minería	-.001	.423	.192
manuf actura	.186	.243	.144
centrales de energía	.137	.179	.535
construcción	.174	.035	-.146
servicios a empresas	.203	-.240	.110
bancos	.040	-.311	.560
sector serv icios	.207	-.152	-.298
transporte y comunicaciones	.196	.139	-.358

Método de extracción: Análisis de componentes principales.
Puntuaciones de componentes.

Figura 1.19

De la matriz de coeficientes de las puntuaciones en las componentes anteriores podemos deducir la siguiente relación entre componentes y variables:

$$\mathbf{Z1} = -2.800\mathbf{Agri} - 0.010\mathbf{Min} + \dots + 01.96\mathbf{Tc}$$

$$\mathbf{Z2} = 0.037\mathbf{Agri} + 0.423\mathbf{Min} + \dots + 0.139\mathbf{Tc}$$

$$\mathbf{Z3} = -0.046\mathbf{Agri} + 0.192\mathbf{Min} + \dots - 0.358\mathbf{Tc}$$

Cuando se realizó el análisis de la matriz de componentes se observó que es difícil agrupar las variables en componentes, con lo que procedería realizar una rotación. Realizamos una rotación *Varimax* que tiene la propiedad de que los factores siguen siendo incorrelados. Para ello marcamos las opciones de la figura 1.7, como aparece a continuación en la figura 1.20:

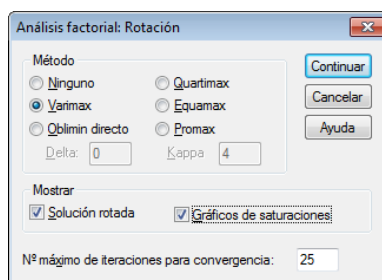


Figura 1.20

Obtenemos la *matriz de componentes rotados*, que muestra cómo la variable **Servicios a empresas** se sitúa en la primera componente, la variable **Centrales de energía** se sitúa en la segunda componente y la variable **Construcción** se sitúa en la primera componente. A pesar de la rotación, no se ven claro los grupos de variables.

Matriz de componentes rotados^a

	Componente		
	1	2	3
agricultura	-.871	-.343	-.299
minería	-.186	.743	-.520
manufactura	.465	.692	-.136
centrales de energía	.146	.809	.207
construcción	.607	.174	-.030
servicios a empresas	.643	-.006	.602
bancos	-.060	-.005	.913
sector servicios	.824	-.157	.177
transporte y comunicaciones	.751	.205	-.325

Método de extracción: Análisis de componentes principales.

Método de rotación: Normalización Varimax con Kaiser.

a. La rotación ha convergido en 5 iteraciones.

Figura 1.21

El gráfico de componentes en el espacio rotado (figura 1.22) tampoco ayuda mucho a la detección de los grupos de variables. En este gráfico, dos variables correladas positivamente forman un ángulo desde el origen de 0 grados, de 180 si lo están negativamente y de 90 si están incorreladas.

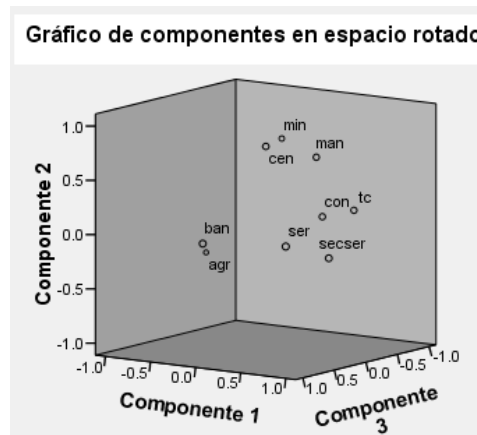


Figura 1.22

***Ejemplo 1.6:** Para el siguiente ejemplo con SPSS, utilizaremos la Encuesta de Hogares y Propósitos Múltiples de El Salvador del año 2009, del cual se tomarán 6 variables detalladas a continuación: INGFA (Ingreso Familiar), INGPE (Ingreso Personal), GASTOHOG (Gasto Hogar), GHALI, GASHE, GHUL6, el tamaño de la población analizada es censal y es de 20361.*

Realizando los mismos pasos del las figuras 1.5 a la 1.19 obtenemos los siguientes resultados:

Estadísticos descriptivos

	Media	Desviación típica	N del análisis
INGFA	428.9582	503.64031	20361
INGPE	122.8751	155.49808	20361
Gasto del Hogar	313.8120	268.20303	20361
GHALI	131.2558	77.86865	20361
GHASE	32.1878	34.91760	20361
GHUL6	19.6703	64.35499	20361

Figura 1.23

La tabla de la figura 1.23 nos muestra los estadísticos univariados correspondientes, en la figura 1.24 tenemos la matriz de correlaciones cuyo determinante es 0.25 que al ser muy pequeño indica que el grado de intercorrelación entre las variables es muy alto, condición inicial que debía cumplir.

Matriz de correlaciones

		INGFA	INGPE	Gasto del Hogar	GHALI	GHASE	GHUL6
Correlación	INGFA	1.000	.787	.722	.542	.538	.353
	INGPE	.787	1.000	.515	.321	.406	.315
	Gasto del Hogar	.722	.515	1.000	.763	.696	.526
	GHALI	.542	.321	.763	1.000	.520	.270
	GHASE	.538	.406	.696	.520	1.000	.282
	GHUL6	.353	.315	.526	.270	.282	1.000
Sig. (Unilateral)	INGFA		.000	.000	.000	.000	.000
	INGPE	.000		.000	.000	.000	.000
	Gasto del Hogar	.000	.000		.000	.000	.000
	GHALI	.000	.000	.000		.000	.000
	GHASE	.000	.000	.000	.000		.000
	GHUL6	.000	.000	.000	.000	.000	

a. Determinante = .025

Figura 1.24

KMO y prueba de Bartlett

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		.734
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	75444.047
	gl	15
	Sig.	.000

Figura 1.25

La figura 1.25, nos muestra el test de esfericidad de Barlett como su p-valor es 0,000, se puede concluir que existe correlación significativa entre las variables. Esta tabla incluye el estadístico KMO, cuyo valor 0.734 (> 0.5) cercano a la unidad es un indicativo de suficiente correlación y, por lo tanto, el análisis factorial es una técnica útil para el estudio. Luego tenemos en la figura 1.26 la matriz de correlaciones anti-imagen formada por los coeficientes de correlación parcial entre cada par de variables cambiada de signo. Estos coeficientes deben ser bajos para que las variables compartan factores comunes. Los elementos de la diagonal de esta matriz son similares al estadístico KMO para cada par de variables e interesa que estén cercanos a la unidad. Observando la matriz vemos que no tenemos buenos resultados:

Matrices anti-imagen

		INGFA	INGPE	Gasto del Hogar	GHALI	GHASE	GHUL6
Covarianza anti-imagen	INGFA	.238	-.207	-.072	-.022	-.003	.041
	INGPE	-.207	.360	.015	.049	-.022	-.050
	Gasto del Hogar	-.072	.015	.176	-.156	-.135	-.163
	GHALI	-.022	.049	-.156	.386	.021	.115
	GHASE	-.003	-.022	-.135	.021	.501	.085
	GHUL6	.041	-.050	-.163	.115	.085	.659
Correlación anti-imagen	INGFA	.743 ^a	-.706	-.354	-.074	-.009	.104
	INGPE	-.706	.701 ^a	.060	.132	-.052	-.104
	Gasto del Hogar	-.354	.060	.698 ^a	-.598	-.453	-.478
	GHALI	-.074	.132	-.598	.752 ^a	.047	.227
	GHASE	-.009	-.052	-.453	.047	.847 ^a	.148
	GHUL6	.104	-.104	-.478	.227	.148	.669 ^a

a. Medida de adecuación muestral

Figura 1.26

Veamos ahora el número de componentes con el que nos quedaremos. Observando la tabla de la varianza total explicada (figura 1.27) vemos que la primera componente explica un 59.83% de la varianza total y como es la única que aparece, esto nos sugiere que solo se utilizara una componente principal, esto lo podemos corroborar con el grafico de la figura 1.28.

Varianza total explicada

Componente	Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado
1	3.590	59.834	59.834

Método de extracción: Análisis de Componentes principales.

Figura 1.27

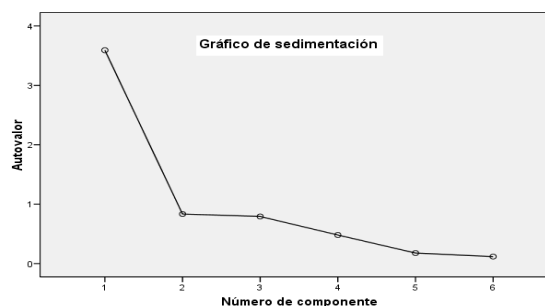


Figura 1.28

Ahora veamos cómo utilizar las variables que se agrupan en torno a cada componente y cuya combinación lineal define la componente. Para ello representamos cada una de las seis variables por medio del factor extraído utilizando la matriz de componentes (Figura 1.29). Podemos escribir lo siguiente:

$$\begin{aligned}
 \text{INGFA} &= 0.870F1 \\
 \text{INGPE} &= 0.728F1 \\
 \text{GASTOHOG} &= 0.925F1 \\
 \text{GHALI} &= 0.754F1 \\
 \text{GHASE} &= 0.755F1 \\
 \text{GHUL6} &= 0.556F1
 \end{aligned}$$

Matriz de componentes^a

	Componen te
	1
INGFA	.870
INGPE	.728
Gasto del Hogar	.925
GHALI	.754
GHASE	.755
GHUL6	.556

Método de extracción: Análisis de componentes principales.

a. 1 componentes extraídos

Figura 1.29

Comunalidades

	Extracción
INGFA	.756
INGPE	.530
Gasto del Hogar	.856
GHALI	.569
GHASE	.571
GHUL6	.309

Método de extracción: Análisis de Componentes principales.

Figura 1.30

En la primera componente está representada claramente GASTOHO (0.925) y podemos ver el resto en la tabla. A continuación se analiza la comunalidad (figura 1.30) de cada variable (suma de cuadrados de sus cargas factoriales definidas en la matriz de componentes) después de la extracción de los factores (componentes). La comunalidad es la parte de variabilidad de cada variable explicada por los factores. Antes de la extracción de los factores la comunalidad de cada variable es la unidad, e interesa que después de la extracción siga siendo alta.

Correlaciones reproducidas

		INGFA	INGPE	Gasto del Hogar	GHALI	GHASE	GHUL6
Correlación reproducida	INGFA	.756 ^b	.633	.804	.656	.657	.483
	INGPE	.633	.530 ^b	.673	.549	.550	.405
	Gasto del Hogar	.804	.673	.856 ^b	.698	.699	.514
	GHALI	.656	.549	.698	.569 ^b	.570	.419
	GHASE	.657	.550	.699	.570	.571 ^b	.420
	GHUL6	.483	.405	.514	.419	.420	.309 ^b
Residual ^a	INGFA		.155	-.082	-.114	-.119	-.131
	INGPE		.155	-.159	-.228	-.144	-.089
	Gasto del Hogar		-.082	-.159	.065	-.003	.012
	GHALI		-.114	-.228	.065	-.049	-.149
	GHASE		-.119	-.144	-.003	-.049	-.138
	GHUL6		-.131	-.089	.012	-.149	-.138

Método de extracción: Análisis de Componentes principales.

a. Los residuos se calculan entre las correlaciones observadas y reproducidas. Hay 12 (80.0%) residuales no redundantes con valores absolutos mayores que 0,05.

b. Comunalidades reproducidas

Figura 1.31

En la figura 1.31 se presenta la *matriz de correlaciones reproducidas* en la que se observa que un 80% de los errores son mayores que 0.05, lo que indica que la bondad del modelo es discutible.

Ahora calcularemos las puntuaciones factoriales. Será entonces una variable sustituta de las iniciales que representa su reducción y que recoge el 59% de la variabilidad total. De la matriz de coeficientes de las puntuaciones en las componentes anteriores podemos deducir la siguiente relación entre componentes y variables:

Matriz de coeficientes para el cálculo de las puntuaciones en las componentes

	Componen te
	1
INGFA	.242
INGPE	.203
Gasto del Hogar	.258
GHALI	.210
GHASE	.210
GHUL6	.155

Método de extracción: Análisis de componentes principales.
Puntuaciones de componentes.

Figura 1.32

$$Z_1 = 0.242INGFA + 0.203INGPE + 0.258GASHOG + 0.210GHALI + 0.210GHASE + 0.155GHUL6$$

CAPÍTULO II
ÍNDICES SINTÉTICOS
-ÍNDICES MEDIOAMBIENTALES-

2.1. ÍNDICES SINTÉTICOS DE CALIDAD AMBIENTAL (ISCA).

Las estadísticas ambientales constituyen la fuente primaria de información ambiental que necesitan los países para ser incorporados en el ciclo de políticas públicas, y para la información ciudadana. Sin una producción sistemática, oficial y de calidad, los indicadores ambientales no podrán ser construidos ni utilizados para aquellos ámbitos de decisión que son más importantes según la realidad del territorio en cuestión. De la experiencia mundial y regional, se constata que en un contexto de escasez de recursos, las estadísticas e indicadores son herramientas adecuadas para mejorar la gestión ambiental y de sostenibilidad en múltiples espacios, incluyendo las políticas públicas, la gestión empresarial y el accionar de las organizaciones ciudadanas. Y es precisamente por la escasez de recursos financieros asignados a medio ambiente, que se requiere de estas herramientas para focalizar y hacer más efectivo y transparente el ciclo de política pública. No solamente los expertos regionales, sino también las agencias internacionales, como la División de Estadística de las Naciones Unidas (DENU) (United Nations Statistics Division (UNSD)), recomiendan que debe privilegiarse la producción de estadística básica ambiental para el poblamiento y sostenimiento de los indicadores.

No existe un procedimiento único para diseñar e implementar estadísticas e indicadores ambientales o de sostenibilidad que sean adecuados. Cada país o institución deberá realizar un diseño propio, de acuerdo a sus recursos y necesidades, que garantice que las estadísticas e indicadores producidos sean realmente útiles para la toma de decisiones; justificando, por tanto, la no despreciable inversión de recursos en su producción y mantención. El mejor criterio para desarrollar indicadores, es que éstos deben responder a las necesidades de los usuarios en forma oportuna y costo efectiva. Respecto de la utilización del marco conceptual y metodológico, se cuenta con varios enfoques y marcos referenciales que explican la relación sociedad-naturaleza, pero a la hora de aplicarlos en el desarrollo de indicadores, se evidencia su alta complejidad, y relación de contenidos de distintas especialidades. También, es necesario avanzar en la transformación de los indicadores como producto al servicio de la información, mediante la comunicación efectiva, para que la herramienta sea realmente utilizada.

Es importante que las estadísticas ambientales, los sistemas de información ambiental y los indicadores ambientales o sobre sostenibilidad, no se conviertan en una parcela tecnocrática (de ingenieros, informáticos, y/o estadísticos), ya que se trata de un ámbito transversal, tanto científico como instrumental y político. Un primer paso imprescindible para avanzar en el desarrollo de indicadores y sistemas de información es preparar la base estadística ambiental que la mayoría de los países requiere sistematizar.

Al avanzar en la producción de estadística, información e indicadores, es importante garantizar su uso en forma sistemática. Para esto pueden ser dispuestos diversos mecanismos, pero la participación temprana de actores desde el diseño de los sistemas es primordial. Luego, es necesario construir capacidades para su uso, tanto en la ciudadanía, como en el sector privado y en el sector público. La plataforma de comunicación de las estadísticas, así como los sistemas de información e indicadores resulta crucial; siendo este un desafío que poco se ha abordado en nuestro país. Hay que trabajar en mejorar el acopio de información ambiental y sobre la sostenibilidad del desarrollo.

2.1.1. Antecedentes.

Desde 1987 se viene desarrollando planes de acción ambiental a nivel mundial, pues para ese año se conoce del informe dirigido por las Naciones Unidas (UNCED/CNUMA) denominado “Nuestro Futuro Común”, en el cual se introduce el concepto de desarrollo sostenible como “el desarrollo que satisface las necesidades presentes sin comprometer la habilidad de las generaciones futuras para satisfacer sus propias necesidades”; para 1992 Naciones Unidas desarrolla la conferencia sobre Ambiente y Desarrollo donde presenta el problema ambiental a nivel mundial, lo cual contribuye significativamente a un nuevo enfoque del concepto de desarrollo sostenible y a sus políticas, uno de los resultados de la conferencia es el documento oficial Agenda Local 21, y su artículo 40 se refiere a la necesidad de disponer de información adecuada para el monitoreo, evaluación y seguimiento de la sostenibilidad del desarrollo. De aquí se estimula la necesidad de indicadores ambientales. También para 1987 en Europa se desarrollo el 4to Programa de Acción Ambiental dirigido por la

Unión Europea (EU), en el cual se define el Plan de Acción Ambiental (1987 - 1992); luego para 1990 la Comisión Europea presenta el Libro Verde Sobre Ambiente Urbano, el cual fue un hito sobre ambiente urbano, porque constituyó el primer camino hacia el debate e investigación sobre ambiente urbano y calidad de vida en la Unión Europea, se describen los principales problemas ambientales urbanos, demandando la necesidad de integrar la planificación y gestión urbana de los problemas ambientales. De aquí surge la necesidad de monitorear y evaluar las condiciones ambientales de las ciudades.

Es importante destacar la evolución europea en materia ambiental que se ha venido desarrollando a través de los años, de la cual se han venido elaborando importantes documentos como los mencionados, además de otros como el Tratado sobre la Unión Europea, el 5to (1992) y 6to (2001) Programa de Acción Ambiental, Carta Alborg (1994), Plan de Acción de Lisboa (1996), Foro Urbano de Viena (1998), Campaña de Ciudades Sostenibles “Hannover” (2000); en algunos de los cuales se destaca la importancia de construir indicadores para la toma de decisiones.

La *Organización para la Cooperación y el Desarrollo Económico (OCDE)*, ha definido trece áreas de indicadores, entre las cuales una de ellas se refiere al fenómeno ambiental urbano como área de interés especial, otro programa relevante a nivel mundial es el de la Organización Mundial para la Salud (OMS), que establece un conjunto de indicadores de ciudades saludables, dentro del Programa Salud para Todos en 2000. En España, además del sistema de indicadores ambientales urbanos diseñados por el Ministerio del Medio Ambiente (MMA, 2000), se destacan los sistemas de indicadores desarrollados para el nivel de regiones en Cataluña, Andalucía y el País Vasco, entre otras. Los indicadores desarrollados responden a las condiciones específicas de las regiones, con fines de comparar y diseñar políticas a este nivel.

Estados Unidos presentó una de las iniciativas más destacables a nivel mundial, la “Seattle Ciudad Sostenible”, 1992, la cual fue caracterizada por un proceso de participación con amplia convocatoria comunitaria, ONGs, universidades, académicos, etc. Que llevó a la definición de un conjunto de **indicadores simples** dirigidos a tres áreas temáticas: economía, sociedad y medio ambiente a nivel de la ciudad.

En Latinoamérica se destaca la experiencia de Chile, México y Costa Rica como procesos avanzados en la definición de sistemas de indicadores de desarrollo sostenible. Para el caso de Colombia, son dos las experiencias más relevantes, más que por su

trascendencia internacional, porque están relacionadas con el modelo de indicadores que se presenta en nuestra investigación. Una de las experiencias es el sistema de indicadores ambientales para Colombia, realizado por el SISA en 2002, en el que se define un conjunto de 34 indicadores ambientales para su desarrollo sostenible.

2.2. ASPECTOS TÉCNICOS PARA LA CONSTRUCCIÓN DE LOS INDICADORES SINTÉTICOS

El indicador sintético o índice está formado por una serie de componentes que aportan información valiosa acerca del objetivo a medir, ya sea la calidad ambiental, el desarrollo sostenible, etc., en las distintas unidades de observación.

Es deseable que la información recopilada para la medición de los componentes sea objetivamente medida, dado que es una expresión matemática (un dato) acerca del estado del mismo. Una medición subjetiva puede conducir erróneamente a múltiples estados y por lo tanto puede afectar de distintas formas y magnitudes el índice⁴.

Esta medición de los componentes indica la magnitud del estado del mismo. Por ello técnicamente se denomina indicador simple o parcial. Este indicador puede ser un dato o la combinación de más de uno para conformar o calificar el estado de un componente de interés. Por ejemplo en el modelo presentado en la tabla A.1 (ver anexo) presentamos un conjunto de 16 indicadores simples que influyen en la calidad ambiental urbana de nuestro país, del cual se deduce que los datos originales con que se valorarían estos indicadores simples son ampliamente mayores. Imaginemos el indicador *densidad de área verde por habitante* (para nuestro caso no fue posible obtener esta variable), que demanda la combinación matemática de dos tipos de datos: áreas verdes (en m²) dividido por la población total en una unidad de observación. El indicador sintético o índice es una combinación matemática de los indicadores parciales.

⁴ La objetividad en la medición de los componentes asegura la no violación de supuestos como el de invarianza del índice.

2.2.1. Criterios generales para la selección de indicadores ambientales simple y sintéticos.

Los criterios de selección de indicadores juegan un papel determinante cuando se aborda el diseño de un sistema de indicadores simples o la construcción de un indicador sintético, dado que éstos demandan información cuantitativa de múltiples componentes e indicadores que lo explican. Pero, la tarea es saber seleccionar estos componentes e indicadores, además, el número de componentes e indicadores que se deberá incluir en el índice sintético.

En todo caso es bueno saber si son los indicadores una buena aproximación a la medición del fenómeno a evaluar. Los criterios de selección tienen que ser, por un lado, el filtro para resumir una gran cantidad de datos en un número reducido de indicadores por temas, áreas, componentes, preocupaciones sociales, etc., y por otro, también deben permitir dotar el sistema de indicadores de mayor calidad estadística en la información y de un método científico que delimite la frontera del análisis.

En su mayoría, los criterios empleados para la selección de indicadores coinciden en que el primer filtro lo constituye la definición de componentes, temas o áreas de interés de acuerdo al modelo de análisis seleccionado, ya sea basado en el enfoque de desarrollo sostenible o la ciudad como ecosistema urbano. El siguiente paso es definir una serie de indicadores que se pueden medir de acuerdo a su **estado** u otra categoría. En nuestra investigación nos interesa resaltar, por ejemplo que si el objetivo es medir la **calidad ambiental urbana**, donde el componente principal es el **medio ambiente urbano**; un tema sería el **agua**, y una posible medida o indicador podría ser **cobertura de agua (% de viviendas con sistema de agua potable)**, etc.

En cuanto a la selección podríamos tener en cuenta la lista de filtros para la definición de un conjunto de indicadores:

- a. Deben describir un resultado final que concierne a la preocupación social.
- b. Deben describir situaciones sociales de elección pública, es decir, que sean susceptibles de mejorar mediante la gestión social.

- c. Deben estar referidos a un campo de aplicación temporal y espacial definido, de forma que permita la comparación intertemporal y entre regiones.
- d. Deben tener la posibilidad de agregación y desagregación, lo cual plantea un desarrollo que sea independiente de las instituciones que proporcionan la información.
- e. Deben estar integrados en sus definiciones, especificaciones, directrices estadísticas y categorías clasificatorias, con otros sistemas de estadísticas sociales, demográficas y económicas, con las cuales se puedan relacionar.
- f. Deben basarse en la validez científica. Es decir, el conocimiento científico de las relaciones de causalidad, sus atributos y su significado deben estar bien fundamentados.
- g. Debe ser sensible a cambios, en la medida que deben señalar cambios de tendencia en las situaciones que representan, preferiblemente en el corto plazo.
- h. Deben ser predictivos, de forma tal que brinden señales de posibles tendencias futuras de lo que miden.
- i. Existencia de los datos como un criterio de selección posterior a la definición de la lista amplia de indicadores.
- j. Representatividad de las zonas descritas, de tal forma que permita comparar las posiciones relativas.
- k. Se prefieren los indicadores directos del fenómeno a explicar, pero en caso de no existir, se recomienda el uso de indicadores indirectos que no generen ambigüedad en la explicación del factor que se quiere medir (justo nuestro caso).
- l. Los indicadores deben tener una gran capacidad de discriminación del conjunto de unidades de observación para el cual se están diseñando (municipalidades, ciudades, países, etc.).

En general, este amplio resumen de criterios genéricos para la selección de indicadores muestra que no existe un consenso generalizado al respecto; sin embargo, son una buena colección de criterios que los investigadores tenemos al alcance para definir esa primera lista de indicadores que ayuden a sustentar las hipótesis de trabajo que se relacionan con esta primera fase de definición y recopilación de datos: un buen conjunto de indicadores debe ser “completo” para medir el índice, y cada uno de los indicadores debe ser una “buena” medida de los estados (u otra categoría) de la situación ambiental.

De acuerdo al análisis realizado hasta aquí, se pueden resaltar básicamente tres aspectos determinantes en la definición de indicadores ambientales urbanos:

- Los sistemas de indicadores están referidos a un marco analítico.
- La elaboración de indicadores ambientales urbanos tienen la función de evaluar la política ambiental local y hacer un seguimiento y comparación a nivel de agregación demandado.
- La referencia espacial para los indicadores es básicamente de ciudad, y en algunos casos de municipios. Por lo tanto, la evaluación de la gestión desde lo local es bastante agregada y no da cuenta de los detalles de unidades espaciales de una mayor escala, como colonias, barrios, cantones, etc.

Desarrollar un sistema de indicadores ambientales urbanos a una escala espacial amplia (a nivel de colonias, por ejemplo) proporcionaría a los tomadores de decisiones una importante información para ejecutar planes de acción con proyectos y actividades diferenciadas. Además son una fuente de información importante para la medición y seguimiento de la gestión institucional y el control de la opinión pública por parte de los ciudadanos. Sin embargo en El Salvador no se dispone de la información a esos niveles de desagregación.

A continuación se presenta la manera de construir un modelo de un sistema de indicadores ambientales urbanos, que permita ordenar los municipios de El Salvador de acuerdo al Índice Sintético de Calidad Ambiental (ISCA).

2.2.2. Supuestos relevantes para la formulación del índice

Medir la calidad ambiental, el desarrollo sostenible o cualquier otra variable latente, como la agregación de un conjunto de datos e indicadores simples en un índice,

demanda la formulación de una serie de supuestos que aseguran la consistencia de los resultados, partiendo de la adecuada selección de los datos.

Estos supuestos consisten en:

- ✓ **Completitud:** Los indicadores seleccionados para construir el índice de calidad ambiental son todos los que explican el objetivo a medir.
- ✓ **Bondad de los indicadores que determinan el índice:** Los indicadores seleccionados deben medir adecuadamente los atributos que se describan en relación con el índice a estimar, tanto en el espacio como en el tiempo.
- ✓ **Objetividad en la valoración de los indicadores:** Los datos que se utilicen deben ser una representación objetiva de la valoración de su estado. Es decir, están ausentes de sus propiedades intrínsecas, los juicios de valor con respecto a su magnitud en el espacio y el tiempo.

Con los supuestos planteados se han asegurado las condiciones ideales de los datos utilizados. El siguiente paso es determinar las propiedades matemáticas que garantizan la consistencia de un indicador sintético.

Condiciones a priori para que un indicador sintético sea consistente:

- i. **Existencia y Determinación:** la función matemática que define el indicador sintético debe ser tal que exista y tenga solución para todos los valores del índice en cada unidad de observación.
- ii. **Monotonía:** El índice debe responder positivamente a una modificación positiva de los componentes y negativamente a una modificación negativa. En el modelo presentado en la tabla 2.1 (Ecuaciones 59 – 65, además Ec. 66 – 68) se define el sentido que tienen los componentes de primer nivel cuando entran al indicador de segundo, tercer y cuarto nivel. En la práctica, esto exige que en el análisis multivariado se introduzca la información con el signo con que se espera estén relacionados los indicadores simples y el objetivo a medir.
- iii. **Unicidad:** El índice debe representar un único valor, para una situación dada.
- iv. **Invariancia:** Como consecuencia de la propiedad anterior, el índice debe ser invariante respecto a un cambio de origen y/o de escala de medición de los indicadores simples.

- v. **Homogeneidad:** La función matemática que define el índice, debe ser homogénea de grado uno. Esto asegura que si los indicadores simples aumentan o disminuyen, en igual proporción lo hace el indicador sintético.

$$f(c.I_1, c.I_2, \dots, c.I_n) = c.f(I_1, I_2, \dots, I_n)$$

- vi. **Transitividad:** Suponga que a , b , y c son tres estados distintos medidos por el índice e $I(a)$, $I(b)$ e $I(c)$ son valores del índice correspondiente a esos tres estados. Debe verificarse que:

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

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

Una vez definidas las propiedades matemáticas que deben cumplir los indicadores simples y el índice, se debe resolver el problema de agregación dado que lo más probable es que cada uno de los componentes del índice tenga distintas unidades de medida.

Se propone que técnicamente cualquier proceso de tipificación pueda transformar todos los componentes en las mismas unidades y escala, asegurando que esta transformación no modifique el orden del estado del indicador en las distintas unidades de observación.

Otro problema que quedaría por resolver es cómo ponderar los indicadores simples en un indicador agregado o sintético. Ello demandaría un sistema de ponderación que otorgue peso o importancia a cada indicador simple. Algunos criterios podrían ser:

- Desde el punto de vista de los objetivos sociales, se puede asignar ponderaciones de acuerdo a la importancia de cada uno de los indicadores simples.
- El peso de cada indicador simple vendría dado por la información útil que tuviera cada uno de acuerdo a la varianza explicada del índice.

Este último criterio es el que emplea la mayoría de estudios que estiman índices, a través de los métodos de análisis multivariante como el que se emplea en este trabajo.

2.3. SISTEMA DE INDICADORES AMBIENTALES URBANOS.

2.3.1. Marco de referencia de indicadores de calidad ambiental

El concepto de calidad ambiental como el estado de un conjunto multidimensional de indicadores ambientales dada una función de utilidad, donde las preferencias de los agentes son débilmente separables entre las viviendas y las características ambientales del entorno, es posible definir una función de elección de localización i , $b_{ij} = b_{ij}(q_i, Q_{ij}^*, S_i, N_i, u^*)$, en la que los Q_{ij}^* es todo el conjunto de atractivos ambientales asociadas a la vivienda i para cada localidad j , excepto q_i que es una característica ambiental específica asociada a la vivienda i , S_i y N_i representan las características socioeconómicas y estructurales de la vivienda, respectivamente, u^* es un nivel de utilidad dado. Esto indicaría que los agentes eligen su localización en un entorno urbano teniendo en cuenta el valor que toman las variables que determinan su función de elección. En este trabajo presentamos la propuesta de un modelo de indicadores ambientales y su síntesis en un índice de tal forma que recoja Q_{ij}^* como una agregación de indicadores multidimensionales que determinan la calidad ambiental del entorno.

2.3.2. Modelo de indicador sintético

La Figura 2.1 muestra el esquema de cómo se simplifica la información derivada del sistema de indicadores ambientales presentado en la tabla 2.1. Este proceso muestra las

acciones a seguir para construir un indicador sintético (ISCA) que se utiliza para evaluar cada una de las unidades experimentales. En este caso nos referimos a las municipalidades de zona urbana. Partimos de la definición de calidad ambiental urbana y de sus dos componentes. De aquí se derivan un conjunto de indicadores simples, a los cuales les aplicamos los criterios definidos anteriormente, resultando los indicadores operativos del sistema.

Estos indicadores operativos como ideal se deben presentar en la escala espacial definida, mediante el empleo de un **Sistema de Información Geográfico** que homogenice espacialmente toda la base de datos.

Figura 2.1. *Proceso de cálculo y representación espacial del ISCA.*



Obviamente, este esquema ilustra el proceso de cálculo ideal de un indicador sintético, sin embargo, en nuestro estudio se modificará al momento de construir el indicador, pero en general realizaremos los siguientes pasos:

1. Definir los indicadores operativos y especializarlos.
2. Emplear la técnica de análisis multivariado Componentes Principales para sintetizar un conjunto de indicadores simples en un índice.
3. Pasar de indicadores de primer nivel a indicadores sintéticos de nivel II, III y IV, tal como se expresa en la tabla 2.1.

Simplificando el desarrollo conceptual expresado hasta aquí, a continuación describimos un conjunto de ecuaciones que explican el proceso a seguir para derivar matemáticamente el ISCA. Este proceso consiste en tres fases, en las que de manera consecutiva se simplifica la información de indicadores de primer nivel, para traducirlos en índices temáticos (nivel II), luego a índices de componentes (nivel III) y finalmente al indicador sintético de calidad ambiental (nivel IV).

En la primera fase (ver Tabla 2.1) se puede observar la forma en que se pasa de indicadores de primer a segundo nivel. Lo importante aquí son las hipótesis de relación lineal entre el conjunto de *indicadores simples* y los índices parciales (temáticos y de componentes) e indicador sintético de calidad ambiental. El signo que toma cada una de las variables es determinante para que en el proceso matemático de síntesis de la información sea correcto, dado que el modelo (ACP) debe incluir el signo que lleva esta variable como una explicación de la variación del índice ante un cambio en el *indicador simple*.

Tabla 2.1. Sistema de indicadores ambientales, paso de indicadores de primer nivel a indicadores de segundo nivel. Ver en anexos la definición y construcción de cada uno de estos indicadores.

Paso de indicadores de primer nivel a indicadores de segundo nivel					ECUACIÓN
$Irs_j = \sum_{j=1}^n W_i RS_{ij}$	$\frac{\partial Irs_j}{\partial RS_{1j}} < 0$	$\frac{\partial Irs_j}{\partial RS_{2j}} > 0$	$\frac{\partial Irs_j}{\partial RS_{3j}} < 0$	$\frac{\partial Irs_j}{\partial RS_{4j}} > 0$	(59)
$Ice_j = \sum_{j=1}^n W_i CE_{ij}$	$\frac{\partial Ice_j}{\partial CE_{1j}} < 0$	$\frac{\partial Ice_j}{\partial CE_{1j}} < 0$	$\frac{\partial Ice_j}{\partial CE_{1j}} < 0$		(60)
$Imi_j = \sum_{j=1}^n W_i MI_{ij}$	$\frac{\partial Imi_j}{\partial MI_{1j}} < 0$				(61)
$Itr_j = \sum_{j=1}^n W_i TR_{ij}$	$\frac{\partial Itr_j}{\partial TR_{1j}} < 0$				(62)
$Icv_j = \sum_{j=1}^n W_i CV_{ij}$	$\frac{\partial Icv_j}{\partial CV_{1j}} < 0$	$\frac{\partial Icv_j}{\partial CV_{2j}} < 0$			(63)
$Iag_j = \sum_{j=1}^n W_i AG_{ij}$	$\frac{\partial Iag_j}{\partial AG_{1j}} > 0$	$\frac{\partial Iag_j}{\partial AG_{2j}} > 0$	$\frac{\partial Iag_j}{\partial AG_{3j}} > 0$	$\frac{\partial Iag_j}{\partial AG_{4j}} > 0$	(64)
$Imo_j = \sum_{j=1}^n W_i MO_{ij}$	$\frac{\partial Imo_j}{\partial MO_{1j}} < 0$				(65)

Los $j=1,2,3,\dots,n$ representan las unidades experimentales. Las i representa el subíndice de cada una de las variables que componen la ecuación de cada indicador de segundo nivel y que se detallan en la tabla A.5. W_i es el conjunto de pesos relativos de cada indicador de primer nivel. Como las variables que componen cada indicador de segundo nivel tienen diferentes unidades de medida y escala, se emplean técnicas de análisis multivariantes como el Análisis de Componentes Principales o Análisis de Distancia. En la Ecuación (59) de la tabla 2.1, $\frac{\partial Irs_j}{\partial RS_{1j}} < 0$, indica la contribución negativa al índice, por ende $\frac{\partial Irs_j}{\partial RS_{2j}} > 0$, indica la contribución positiva al mismo. De la misma manera se debe interpretar el resto de ecuaciones.

Una vez estimados los componentes principales y las tasas de contribución de los mismos para las ecuaciones de la (59) a la (65), se procede a calificar cada una de las unidades experimentales de acuerdo al índice que se esté derivando, teniendo en cuenta una ecuación general que consiste en un promedio ponderado de las puntuaciones de cada componente principal, ponderados por la raíz cuadrada de la varianza de cada componente. En este sentido, el índice para cada unidad experimental se calcula como se describe en la formulación (66).

$$I_{mj} = \frac{\sum_{i=1}^r Z_{rj} \sqrt{\lambda_r}}{\sum_{i=1}^r \sqrt{\lambda_r}} \quad (66)$$

En la fórmula (66), I_{mj} representa I_{rs} , I_{ce} , I_{tr} , I_{cv} , I_{mo} , I_{ag} , e I_{mi} separadamente para cada unidad experimental j -ésima, Z_{rj} la puntuación del componente r -ésimo para la unidad experimental j -ésima, y $\sqrt{\lambda_r}$ la raíz cuadrada del autovalor para dicho componente, garantizando así que los componentes con una mayor varianza explicada tengan una mayor ponderación en la calificación del índice. Los $j = 1, 2, \dots, n$ representan las unidades experimentales. Las i representan el subíndice de cada una de las variables que componen la ecuación de cada indicador de segundo nivel y que se detallan en la tabla 2.1. Como las variables que componen cada indicador de segundo nivel tienen diferentes unidades de medida y escala, se emplean técnicas de análisis multivariantes como el Análisis de Componentes Principales (ACP)

Una vez estimados los indicadores de segundo nivel (índices temáticos), se avanza a la fase II para construir los indicadores de tercer nivel (índices de componentes). El procedimiento general consiste en seleccionar los indicadores de primer nivel que tengan *comunalidades*⁵ extraídas superiores al 40% para cada índice temático estimado en la fase I. A los indicadores seleccionados se les aplica el ACP y se estima la fórmula (67), resultando la ordenación de las unidades experimentales de acuerdo a los índices de componentes *Índice de Flujo Urbano (IFLU)* e *Índice de Medio Ambiente Urbano (IMAU)*. Teniendo en cuenta que por IFLU entenderemos como aquel que incluye procesos urbanos que modifican la calidad ambiental del entorno, y por IMAU nos

⁵ Comunalidad: la varianza observada se descompone en *varianza común (comunalidad)* y *varianza específica*.

referimos al estado de indicadores que describen una característica ambiental determinada, por ejemplo, calidad del agua (para nuestro caso, cobertura de agua potable, etc.).

$$IU_{mj} = \frac{\sum_{i=1}^r Z_{rj} \sqrt{\lambda_r}}{\sum_{i=1}^r \sqrt{\lambda_r}} \quad (67)$$

En la fórmula (67), IU_{mj} representa IFLU e IMAU separadamente para cada unidad experimental j -ésima. La parte derecha de la ecuación es igual a la descrita en (66).

En la fase III pasamos de indicadores de tercer nivel al índice de calidad ambiental urbana. Al igual que en la fase II, seleccionamos los indicadores que tienen correlación superior al 40% con el índice de componente estimado. A los indicadores seleccionados se les aplica el ACP y se estima la fórmula (68), resultando la ordenación de las unidades experimentales de acuerdo a los índices de calidad ambiental (ICA).

$$ISCA_j = \frac{\sum_{i=1}^r Z_{rj} \sqrt{\lambda_r}}{\sum_{i=1}^r \sqrt{\lambda_r}} \quad (68)$$

La fórmula (68) representa el *índice sintético* de calidad ambiental, que en el período $t=0$ indica la calidad ambiental relativa entre las distintas unidades experimentales. Sin embargo, de forma dinámica, el índice puede indicar cambio del período $t-1$ al período t , tal como lo expresan las ecuaciones (69a, b, c).

$$(69a). \Delta ISCA_j = \frac{ISCA_{j,t} - ISCA_{j,t-1}}{ISCA_{j,t}} > 0$$

$$(69b). \Delta ISCA_j = \frac{ISCA_{j,t} - ISCA_{j,t-1}}{ISCA_{j,t}} = 0$$

$$(69c). \Delta ISCA_j = \frac{ISCA_{j,t} - ISCA_{j,t-1}}{ISCA_{j,t}} < 0$$

El resultado final del índice de calidad ambiental que se formula en este documento es una variable que recoge las características ambientales de los distintos sectores en que se puede subdividir una ciudad; una representación del nivel de calidad ambiental “agregada” del entorno Q_{ij}^* .

2.4. ESTIMACIÓN DEL ÍNDICE RESULTADO DEL ACP

Cuando se tiene definidos los componentes debe interesarse en incorporarlos al análisis del indicador sintético que se quiere construir. Si solo se llega a seleccionar un componente principal, éste actuaría como indicador sintético del conjunto de variables respuestas, pero cuando los componentes son más de uno, ello exige un tratamiento para utilizarlos en el análisis de los resultados. Aquí se propone, que el índice sea calculado empleando un promedio de las puntuaciones de cada componente principal, ponderados por la raíz cuadrada de la varianza de cada componente. En este sentido, el índice para cada unidad de observación se debe calcular como:

$$Indice_j = \frac{\sum_{i=1}^r Z_{rj} \sqrt{\lambda_r}}{\sum_{i=1}^r \sqrt{\lambda_r}} \quad (70)$$

$$j = 1, 2, \dots, 262$$

$$i = 1, 2, \dots, r \text{ componentes.}$$

Donde Z_{rj} corresponde a la puntuación del componente r -ésimo para la unidad de observación j -ésima, y $\sqrt{\lambda_r}$ la raíz cuadrada del autovalor para dicho componente, garantizando así que los componentes con una mayor varianza explicada tenga una mayor ponderación en la calificación del índice.

2.5. METODOLOGÍA PARA CONSTRUIR ÍNDICES SINTÉTICOS A PARTIR DEL ANÁLISIS MULTIVARIANTE

Comenzamos primero en conocer los pasos con los cuales procederemos en la utilización del análisis multivariante para el estudio de indicadores sintéticos.

- Se parte de un conjunto de variables que conceptualmente responden a un modelo de componentes principales para el estudio de indicadores sintéticos.
- Se analizan los datos para determinar si se emplea el análisis de varianzas-covarianzas o el de correlaciones. Ello depende del grado de homogeneidad de la escala en que se encuentran los datos de las variables respuestas.
- Se prueba la independencia de las variables respuestas. Es decir, se debe probar si estas variables son independientes o no correlacionadas. En cuyo caso el ACP no operaría.
- Determinar si existen datos ausentes, atípicos, etc., y definir un procedimiento a emplear para su tratamiento en el conjunto del ACP.
- Modelar la base de datos (puede ser en SPSS o cualquier otro paquete estadístico), obteniendo los componentes principales y el conjunto de estimaciones que permitirán probar la consistencia de los resultados. Estos paquetes estadísticos estiman automáticamente los valores de los componentes seleccionados y a juicio del investigador se deben definir los componentes relevantes para la construcción del índice. Un criterio útil es que se deben seleccionar los componentes que tengan autovalores mayor o igual a 1.
- Estimar el orden de las unidades de observación, de acuerdo al procedimiento de agregación de componentes.
- Análisis espacial de los resultados, de tal forma que se puedan identificar patrones espaciales del índice por cada unidad de observación.
- Presentación de los resultados, mediante informe escrito, cuadros, gráficos o mapas según a quien se dirija la investigación. Aquí juega un papel importante la vinculación de las distintas bases de datos a un Sistema de Información Geográfica, que permita representar los resultados.

Es probable que las variables respuestas puedan presentar propiedades no deseables para los análisis multivariados. Por ello se requiere muchas veces tratamientos previos

del conjunto de datos, que prueben si la distribución de este conjunto de variables es una distribución normal multivariada. Si esto no es así, es altamente deseable una transformación del conjunto de variables, luego se procede con los pasos anteriores.

CAPÍTULO III

APLICACIÓN DEL MODELO -CONSTRUCCIÓN DEL ÍNDICE-

3.1. APLICACIÓN DEL MODELO DE INDICADOR SINTÉTICO A LAS ZONAS URBANAS DE EL SALVADOR

3.1.1. Descripción general del área geográfica del estudio

Es importante conocer las condiciones geográficas, espaciales y temporales en el cual se está aplicando el índice sintético de calidad ambiental, por lo cual a continuación presentamos una descripción geográfica de El Salvador en el contexto de nuestro estudio.

Desde el punto de vista de su división político administrativa, El Salvador está dividido en 14 departamentos y 262 municipios (véase figura 3.1.1).

Figura 3.1.1 – Mapa de El Salvador, división político-administrativa.



El territorio nacional de la República de El Salvador tiene una superficie de 21,040.79 km² y sus límites están fijados por las fronteras con Honduras, al norte y al este, y con Guatemala, al oeste. Las aguas del océano Pacífico bañan un extenso litoral que ocupa el sector meridional del país; al sudeste el golfo de Fonseca lo separa de Nicaragua. El Salvador es el único país centroamericano que no tiene salida al océano Atlántico.

3.1.2. Estructura de la población

Según las edades, la población salvadoreña es predominantemente joven. Casi un 30% de los habitantes tiene menos de 15 años, mientras que apenas algo más del 5% pertenecen al grupo de mayores de 65 años. Los individuos adultos comprendidos entre los 15 y 64 años de edad, por su parte, representan el 65% de la población total. La distribución por sexo revela un porcentaje superior de mujeres (52.7%) que de hombres (47.3%). Las tasas de fecundidad han experimentado un cierto descenso en los últimos años debido a la reducción registrada en San Salvador y en las zonas rurales, aunque se han mantenido en el resto de las zonas urbanas. Así, el número de hijos por mujer que en la década de 1980 era de casi seis había pasado a cuatro en la de 1990, y apenas superaba los tres hijos en los primeros años del siglo XXI. De acuerdo con estas orientaciones y sobre la base de una población de 5,744,113 habitantes⁶, la proyección de población calculada para 2010 era de 7,440,662 (ver tablas demográficas en anexos). Esto indica que la población de El Salvador tuvo un índice de crecimiento del 1.9% hasta el 2000, año en que comenzó a disminuir, para situarse en el 1.4% en el 2010. La población activa representa algo más de un 33% del total de habitantes. En cuanto a la distribución por actividades económicas, una quinta parte de la población ocupada trabaja en el campo; algo más de la mitad de la población ocupada está vinculado al sector servicios y el resto, una cuarta parte, se dedica a la industria.

El flujo migratorio desde las zonas rurales hacia las urbes comenzó a producirse de forma masiva en las décadas de 1950 y 1970. Así fue consolidándose la población del área metropolitana de San Salvador, que continúa siendo el mayor polo de atracción del país, junto con las cabeceras de los municipios de Santa Ana, Chalatenango, La Libertad, San Salvador, Cuscatlán, San Miguel, Cabañas, Morazán y La Unión. La mayor concentración se da por tanto en el centro del país, en especial en los municipios circundantes a San Salvador (Mejicanos, Soyapango, Ciudad delgado, Ayutuxtepeque, Apopa, San Marcos e Ilopango), y también se extiende hacia los de otros departamentos cercanos como La Libertad (Santa Tecla y Zaragoza). En la zona oriental y occidental se están constituyendo otros núcleos que atraen a importantes contingentes de población.

⁶ Según VI Censo Nacional de Población 2007.

3.2. DESCRIPCIÓN DE LA BASE DE DATOS.

La preparación de un modelo de indicadores sintéticos que responda a la calidad ambiental, nos exige que se reúnan datos empíricos existentes al nivel de desagregación por municipio. Un primer inconveniente que se da para el desarrollo de investigaciones en El Salvador es la deficiente creación de registros, no almacenamiento de información, la burocracia y la privacidad de la información que debería ser pública. No está de más mencionar que muchas variables importantísimas no se miden en nuestro país, y para el caso de algunas variables que se miden, lamentablemente no se hace a un nivel de desagregación más amplio, por ejemplo las variables: *ruido, riesgo de inundación, microgramos de partículas en suspensión, NO_x, SO_x, CO₂*, se mide solo en algunas zonas, sectores o puntos específicos, por lo cual no es posible realizar estudios como el nuestro a niveles desagregados por municipios, considerando esto, podemos darnos cuenta que nos harán falta unas áreas temáticas muy importantes a la hora de construir nuestro ISCA, como por ejemplo, *calidad del aire*.

Pese a las limitaciones en cuanto a la información disponible, contamos con 16 indicadores simples con los cuales trabajaremos. A continuación describimos el conjunto de datos con el cual realizaremos nuestra investigación.

Inicialmente contamos con 23 variables o indicadores simples, pero luego de aplicar los criterios de selección tenemos 16 indicadores simples, de los cuales 12 fueron obtenidos a partir del VI Censo de Población y V de Vivienda 2007; tres de la publicación del Segundo Censo de Desechos Sólidos, y una de El Vice Ministerio de Vivienda y Desarrollo Urbano (VMVDU). Gestiones realizadas con los Ministerios de Medio Ambiente y Recursos Naturales, Vice Ministerio de Tránsito Terrestre, y el Servicio Nacional de Estudios Territoriales (SNET), a pesar de contar con muy buena información, ésta no cumplía los requisitos de validez para nuestra investigación, ya que se encontraba incompleta, en cuanto al nivel de desagregación exigido, por lo cual no fueron consideradas por razones prácticas.

La mayor parte de indicadores (14 de los 16) están expresados en porcentajes, excepto para el caso de RS1 y CV2, las cuales están expresadas en valores que resultan directos de la medición de la variable. En este caso RS1 representa la cantidad de toneladas de desechos sólidos que se generan por municipio. Al consultar con la base de datos por

ejemplo tenemos que para el municipio de Ahuachapán, se generan 17.29 toneladas de desechos sólidos que equivale aproximadamente a un 34% de la producción de todo el departamento, de manera análoga se analiza para el resto de municipios. Por otra parte la variable RS4 nos dice que de esa cantidad (17.29 ton.) se evacua el 83.06%, por lo que podemos deducir que el 16.94% son problemas de contaminación en el municipio. La variable CV2 nos representa la densidad poblacional por vivienda por municipio, en otras palabras, expresa el promedio del número de personas que habitan por vivienda; tomando nuevamente de ejemplo Ahuachapán se tiene que la densidad poblacional es de 2.66 hab/viv., esto nos dice que en Ahuachapán en promedio habitan aproximadamente 3 personas por vivienda.

El resto de índices fueron creados a partir del Censo de Población de El Salvador 2007, con la ayuda del software REDATAM y de la función de cruce de variables, se pudo obtener dichas variables o índices. Para el caso de CE3 este indicador simple se construyó de la siguiente manera: en la pregunta del censo *¿Cuál es el principal combustible que utiliza para cocinar?*, se tienen las diferentes opciones: *electricidad (1), gas propano (2), kerosene (3), leña (4), carbón de leña (5), paja-palma (6), desechos (7), otro (8)*; sumamos los resultados de las opciones desde la 4 hasta la 8 y esa suma lo dividimos entre el total, así obtenemos el índice CE3; nótese que al cocinar con una de las opciones mencionadas se provoca contaminación, de ahí la importancia de incorporar éste índice al conjunto de la base de datos, podemos mencionar que esta variable contribuye de manera negativa al índice sintético que queremos construir. De manera similar podemos analizar el resto de índices que hemos considerado para la investigación. Es importante tener en cuenta que el indicador sintético a modelar se puede crear con un nivel de exactitud mayor, con un conjunto de indicadores simples significativos, que con los cuales por falta de información en El Salvador no se cuenta.

Los criterios empleados para seleccionar las variables que finalmente se utilizan para construir el índice de calidad ambiental son: *validez científica; capacidad de discriminación; existencia de datos; definición de situación de elección pública; representación de la zona descrita; sensibilidad a cambios; costo-eficiencia*. Estos criterios son el filtro para la selección de los indicadores que finalmente harán parte del índice, dado un conjunto de indicadores teóricos que se presentan en la Tabla 3.2.1.

Tabla 3.2.1 – Indicadores de primer nivel

$Irs_j = \sum_{j=1}^n W_i RS_{ij}$	$\frac{\partial Irs_j}{\partial RS_{1j}} < 0$	$\frac{\partial Irs_j}{\partial RS_{2j}} > 0$	$\frac{\partial Irs_j}{\partial RS_{3j}} < 0$	$\frac{\partial Irs_j}{\partial RS_{4j}} > 0$
$Ice_j = \sum_{j=1}^n W_i CE_{ij}$	$\frac{\partial Ice_j}{\partial CE_{1j}} < 0$	$\frac{\partial Ice_j}{\partial CE_{1j}} < 0$	$\frac{\partial Ice_j}{\partial CE_{1j}} < 0$	
$Imi_j = \sum_{j=1}^n W_i MI_{ij}$	$\frac{\partial Imi_j}{\partial MI_{1j}} < 0$			
$Itr_j = \sum_{j=1}^n W_i TR_{ij}$	$\frac{\partial Itr_j}{\partial TR_{1j}} < 0$			
$Icv_j = \sum_{j=1}^n W_i CV_{ij}$	$\frac{\partial Icv_j}{\partial CV_{1j}} < 0$	$\frac{\partial Icv_j}{\partial CV_{2j}} < 0$		
$Iag_j = \sum_{j=1}^n W_i AG_{ij}$	$\frac{\partial Iag_j}{\partial AG_{1j}} > 0$	$\frac{\partial Iag_j}{\partial AG_{2j}} > 0$	$\frac{\partial Iag_j}{\partial AG_{3j}} > 0$	$\frac{\partial Iag_j}{\partial AG_{4j}} > 0$
$Imo_j = \sum_{j=1}^n W_i MO_{ij}$	$\frac{\partial Imo_j}{\partial MO_{1j}} < 0$			

De la tabla 3.2.1 anterior podemos pensar distintas situaciones posibles dentro de éste marco de análisis propuesto, en primer lugar supongamos unas circunstancias en la cual al aplicar ecuaciones de la tabla 3.2.1 a los índices simples (en la fase 1, para pasar del nivel 1 al nivel 2) obtenemos una lista de indicadores de segundo nivel a los cuales les aplicamos ACP y para éstas componentes principales resultantes, el índice para cada unidad experimental se calcula por medio de la ecuación (11). Una vez estimados los indicadores de segundo nivel, aplicamos nuevamente ACP y estimamos la ecuación (12), resultando la clasificación de las unidades experimentales de acuerdo a los índices de componentes (IFLU e IMAU); estamos en la fase 2, para pasar a la fase 3 (estimación del ISCA) realizamos el mismo procedimiento que en la fase 2 y se estima el ISCA por medio de la ecuación (13). En la figura 2.1 se detalla cómo obtener el ISCA

pasando por cada una de las fases. Pero esta situación corresponde a una idealización, pero puede darse el siguiente caso: suponemos ahora que las componentes principales obtenidas del primer nivel solo pueden clasificarse a áreas temáticas que a su vez corresponden al subcomponente IFLU, en este caso el ISCA se obtendrá siguiendo el proceso conocido, pero sin tomar en cuenta el IMAU en el momento de aplicar la fórmula (12). Ahora pensemos en un caso extremo en el cual al aplicar ACP a los indicadores de primer nivel (aplicando ecuaciones de tabla 1 a índices simples), se obtiene un solo componente principal, entonces, éste es el único componente que al aplicar la ecuación (12) nos proporcionaría nuestro ISCA.

Como observamos se pueden dar muchas situaciones, para lo cual hemos sintetizado el procedimiento para obtener el ISCA en algunas de ellas, sucediendo algún otro caso que no hemos considerado se utilizaran criterios similares pero muy objetivos.

3.3. REPRESENTACIÓN ESPACIAL DE LOS INDICADORES OPERATIVOS.

Es importante tener una visualización espacial de los indicadores simples que estamos midiendo a nivel nacional por municipio, esto para poder tener una idea del comportamiento del mismo y para poder comparar regiones, sectores o municipios. Es así que a continuación mostramos algunos indicadores representados en el mapa político-administrativo de El Salvador.

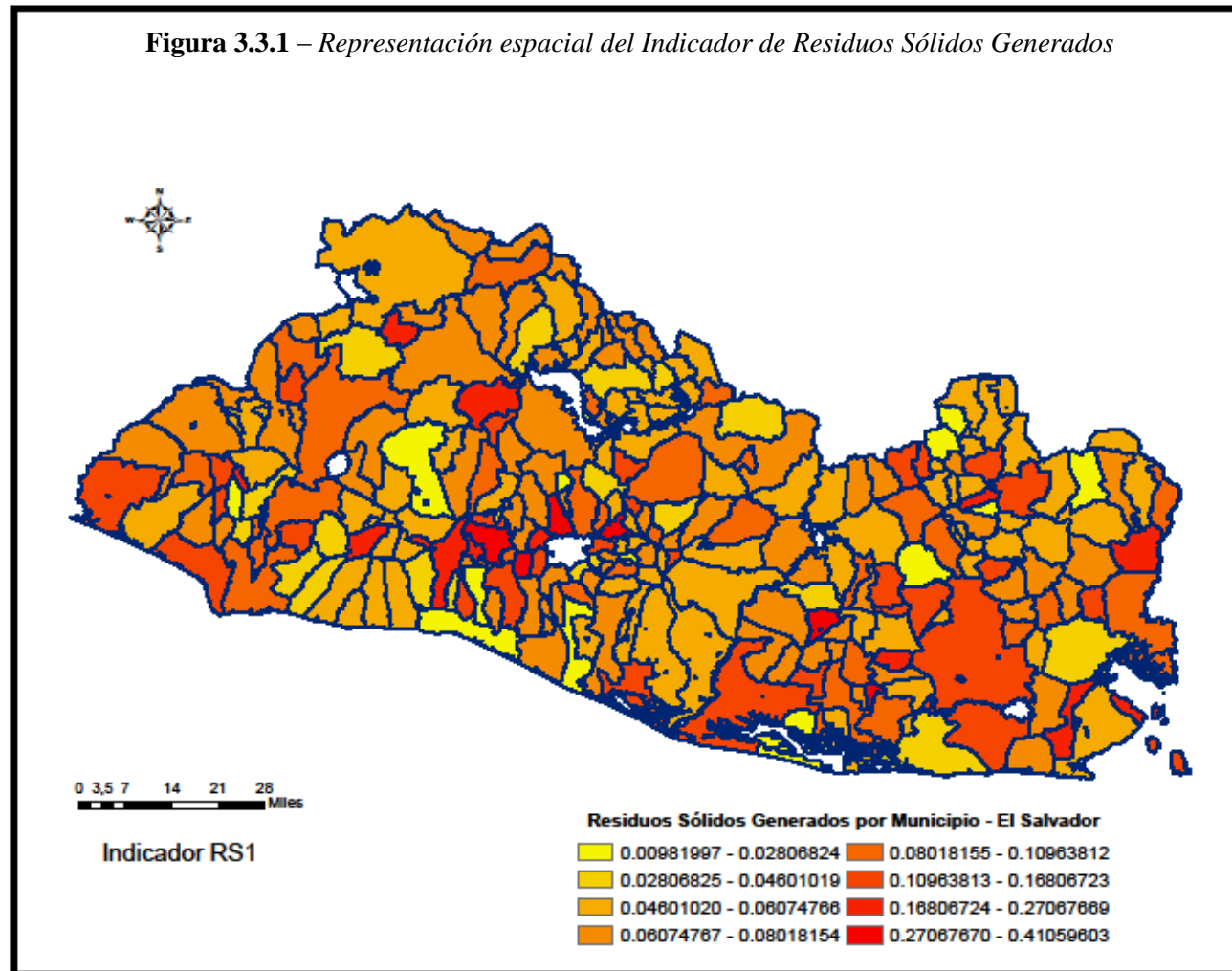
El indicador RS1 es la medida de la generación de residuos sólidos generados por cada uno de los municipios y al referirnos a la figura 3.3.1 podemos observar la siguiente tabla 3.3.1:

Figura 3.3.1 – Representación espacial del Indicador de Residuos Sólidos Generados

ESTADO	TABLA 3.3.1 – RESUMEN DE CONDICIÓN DEL INDICADOR RS1		
	REGIÓN OCCIDENTAL	REGIÓN CENTRAL	REGIÓN ORIENTAL
LEVE	METAPÁN, TEXISTEPEQUE	TEJÚTLA, LA LIBERTAD	CHAPELTIQUE, PERQUÍN, ARAMBALA, TOROLA, SAN FERNANDO
MODERADO	SANTA ANA, AHUACHAPAN	SAN JOSÉ GUAYABAL, SAN LUÍS	CORINTO, SOCIEDAD, CONCHAGUA
CRÍTICO	MASAHUAT, SAN FRANCISCO MENÉNDEZ	MEJICANOS, SAN MARTÍN, SOYAPANGO, EL PAISNAL	MEANGUERA, SAN MIGUEL, EL SAUCE, LA UNIÓN, CHIRILAGUA

Podemos interpretar la tabla anterior de la siguiente manera, los municipios que se encuentran en estado ‘leve’ quiere decir que el porcentaje de desechos sólidos generado en esos municipios es bajo, aquellos que se encuentran en estado ‘moderado’ producen un porcentaje muy considerable de basura y los que se encuentran en estado ‘critico’ son municipios que llegan a altos grados de generación de basura o desechos sólidos medido por toneladas. En el mapa anterior podemos apreciar el estado del resto de municipios según este indicador.

Figura 3.3.1 – Representación espacial del Indicador de Residuos Sólidos Generados



Veamos un segundo caso interesante de analizar, ya que sería como una apreciación de cómo se contrarresta el efecto de RS1, el indicador de Evacuación de Desechos Sólidos, para ello tenemos la figura 3.3.2 y la tabla 3.3.2 que nos proporcionan la siguiente información:

Figura 3.3.2 – Representación espacial del Indicador de Evacuación de Residuos Sólidos

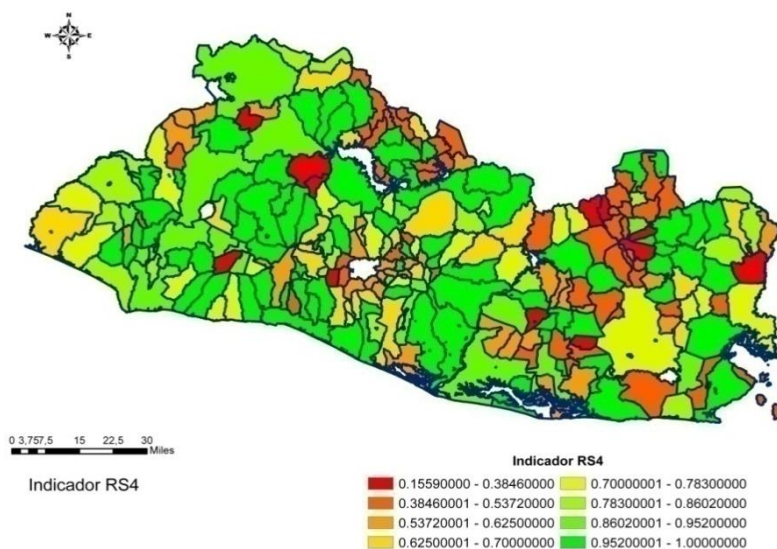


TABLA 3.3.2 – RESUMEN DE CONDICIÓN DEL INDICADOR RS4			
ESTADO	REGIÓN OCCIDENTAL	REGIÓN CENTRAL	REGIÓN ORIENTAL
	BUENO	ACAJUTLA, IZALCO	AGUA CALIENTE, SUCHITOTO
REGULAR	TACUBA, JUJUTLA	GUAZAPA, APOPA	YUCUAIQUÍN, BOLIVAR
MALO	MASAHUAT, EL PORVENIR, TEPECOYO	EL PAISNAL, AGUILARES, SANTO TOMAS	TOROLA, ALEGRÍA SAN ANTONIO, SAN JORGE CAROLINA, EL SAUCE

La manera de interpretar la figura 3.3.2 y tabla 3.3.2 anteriores, es que todos los municipios en estado ‘bueno’ representa una evacuación de desechos con calidad, para el caso de Acajútla se espera que un alto porcentaje de los desechos sólidos generados sean removidos del municipio por algún medio de tratamiento de la basura por lo general con el tren de aseo, tenemos a Guazapa en un estado de ‘regular’ lo que significa que un porcentaje de basura no es evacuado lo cual puede generar problemas

de contaminación, pero que no llega a niveles críticos, en cambio para el caso de Alegría que se encuentra en un nivel ‘malo’ nos indica que en ese municipio los métodos de tratamiento de basura son muy malos o no existen y por tanto esto puede repercutir en problemas de contaminación perjudicando la salud de los ciudadanos.

Otro indicador muy interesante es el de uso de carbón, leña o de otro material para cocinar, realizando el mismo análisis que los casos anteriores tenemos:

Figura 3.3.3 – Representación espacial del Indicador de Uso de Carbón, Leña y otros materiales para cocinar.

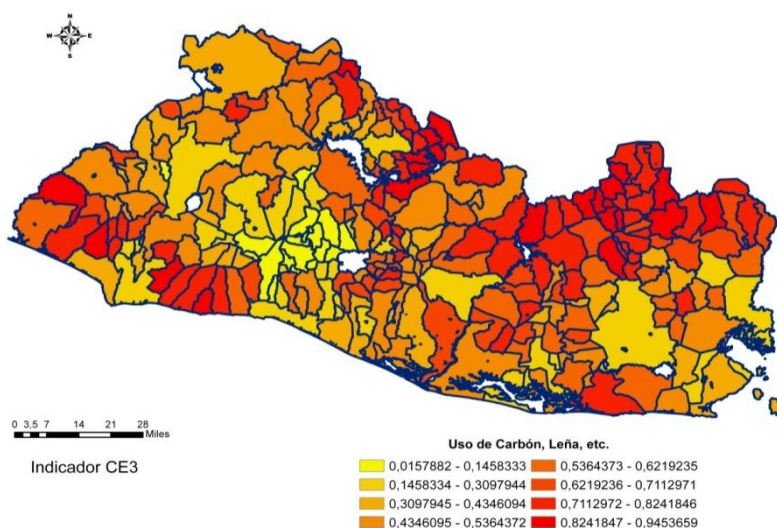


TABLA 3.3.3 – RESUMEN DE CONDICIÓN DEL INDICADOR CE3			
ESTADO	REGIÓN OCCIDENTAL	REGIÓN CENTRAL	REGIÓN ORIENTAL
POCO	SONSONATE, ACAJUTLA	SANTA TECLA, CIUDAD DELGADO	SAN MIGUEL, PASAQUINA, SANTA ROSA DE LIMA
MODERADO	ATQUIZAYA, NAHUIZALCO	ZACATECOLUCA, ILOBASCO	SAN JOSÉ, SAN ALEJO, ANAMOROS
MUCHO	TACUBA, CONCEPCIÓN DE ATACO, JUJUTLA	JUTIAPA, CINQUERA, ARCATAO	JUCUARAN, GUATAJIAGUA, JOATECA

En este caso nos referimos a la figura 3.3.3 y a la tabla 3.3.3, que nos dicen que los municipios en estado ‘poco’ son los que menos contaminan el ambiente con el uso reducido de carbón, leña, u otros materiales similares o de uso rústico para cocinar incluyendo plástico, etc. Los otros estados ‘moderado’ y ‘mucho’ se analizan de la misma manera como lo hemos venido haciendo, tomando en consideración que donde

más se utilizan estos materiales para cocinar se debe a falta de recurso o acceso a gas, o cocina eléctrica.

En los anexos presentamos el resto de indicadores espacializados en el mapa a similitud de los expuestos. La manera de interpretar cada uno de ellos es análoga, por lo cual no consideramos necesario incorporarlos todos en esta sección.

3.4. PROCEDIMIENTOS PARA LA CONSTRUCCIÓN DEL INDICADOR SINTÉTICO DE CALIDAD AMBIENTAL

Siguiendo los criterios empleados para agregar la información de indicadores de primer nivel hasta la construcción del ISCA, se encontró que, aplicando ACP en la primera fase del modelo, en principio todos los indicadores de primer nivel tenían *comunalidades extraídas* superior al 40%, por lo tanto el ISCA se estima con los 15 indicadores disponibles. Se eliminó la variable RS3 ya que era totalmente dependiente con la variable RS2, lo cual introducirá información redundante.

Tabla 3.4.1 - Comunalidades

	Inicial	Extracción
RS1	1,000	,732
RS2	1,000	,696
RS4	1,000	,763
AG1	1,000	,569
AG2	1,000	,667
AG3	1,000	,526
AG4	1,000	,666
CE1	1,000	,709
CE2	1,000	,890
CE3	1,000	,899
MO1	1,000	,419
MI1	1,000	,831
TR1	1,000	,783
CV1	1,000	,489
CV2	1,000	,520

Ahora nos interesa conocer el indicador de flujo urbano (IFLU), el cual construiremos siguiendo los pasos definidos con anterioridad aplicando componentes y siguiendo el

diagrama de flujo de la figura 2.1. Del capítulo 2. Para la construcción del ISCA utilizaremos las 15 variables ya que las comunales resultantes son mayores al 40%.

Antes de continuar es importante mencionar que las variables fueron sometidas a un tratamiento y análisis previo para estudiar la normalidad en ellas, además se realizó una estandarización para todas ellas debido que para la construcción del indicador partiremos de variables estandarizadas.

3.5. CONSTRUCCIÓN DEL ÍNDICE DE FLUJO URBANO (IFLU)

Primero obsérvese que para la construcción del IFLU tenemos definidas 7 áreas temáticas que incluyen 16 indicadores operativos (ver anexos tabla A.5), a los cuales aplicamos ACP y obtuvimos los siguientes resultados.

Partimos de la matriz de correlaciones siguiente:

TABLA 3.5.1 - Matriz de correlaciones(a)										
	RS1	RS2	RS4	CE1	CE2	CE3	MI1	TR1	CV1	CV2
RS1	1	0,1819	-0,416	0,1713	0,1953	-0,203	-0,03	0,2725	-0,014	0,0602
RS2	0,1819	1	0,2936	0,4541	0,645	-0,68	-0,248	0,4922	-0,139	0,1962
RS4	-0,416	0,2936	1	0,0696	0,1863	-0,183	-0,131	0,0733	-0,134	0,0382
CE1	0,1713	0,4541	0,0696	1	0,7186	-0,709	0,1498	0,6281	-0,118	0,0379
CE2	0,1953	0,645	0,1863	0,7186	1	-0,984	-0,112	0,7457	-0,215	0,1383
CE3	-0,203	-0,68	-0,183	-0,709	-0,984	1	0,1274	-0,74	0,2013	-0,128
MI1	-0,03	-0,248	-0,131	0,1498	-0,112	0,1274	1	0,248	0,0944	-0,065
TR1	0,2725	0,4922	0,0733	0,6281	0,7457	-0,74	0,248	1	-0,171	0,1103
CV1	-0,014	-0,139	-0,134	-0,118	-0,215	0,2013	0,0944	-0,171	1	-0,108
CV2	0,0602	0,1962	0,0382	0,0379	0,1383	-0,128	-0,065	0,1103	-0,108	1

Determinante = 0,001

En la tabla 3.5.1 tenemos la matriz de correlaciones cuyo determinante es 0,001 que al ser muy pequeño indica que el grado de intercorrelación entre las variables es muy alto, condición inicial que debía cumplir.

Tabla 3.5.2 – KMO y prueba de Bartlett

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		,744
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	1748,461
	gl	45
	Sig.	,000

El resultado de la tabla 3.5.2 es importante, es el test de esfericidad de Barlett que permite contrastar formalmente la existencia de correlación entre las variables. Como en este caso su p-valor es 0,000, se puede concluir que existe correlación significativa entre las variables. Esta tabla incluye el estadístico KMO, cuyo valor 0,744 (cercano a la unidad) indica una buena adecuación de la muestra a este análisis.

A continuación tenemos la matriz de correlaciones anti-imagen formada por los coeficientes de correlación parcial entre cada par de variables cambiada de signo. Estos coeficientes deben ser bajos para que las variables compartan factores comunes. Los elementos de la diagonal de esta matriz son similares al estadístico KMO para cada par de variables e interesa que estén cercanos a la unidad. Observando la matriz de la tabla 3.5.3 vemos que tenemos buenos resultados:

Tabla 3.5.3 – Matrices anti-imagen

		RS1	RS2	RS4	CE1	CE2	CE3	MI1	TR1	CV1	CV2
Covarianza	RS1	,686	-,088	,334	-,004	6,16E-005	-,003	,081	-,093	,005	-,008
	RS2	-,088	,434	-,157	-,032	,024	,035	,113	-,039	-,024	-,102
	RS4	,334	-,157	,668	,032	-,012	-,010	,030	-,012	,070	,025
	CE1	-,004	-,032	,032	,422	-,018	,001	-,140	-,008	-,011	,054
	CE2	6,16E-005	,024	-,012	-,018	,028	,026	,010	-,012	,012	-,017
	CE3	-,003	,035	-,010	,001	,026	,027	-,009	,006	,008	-,015
	MI1	,081	,113	,030	-,140	,010	-,009	,611	-,219	-,066	-,001
	TR1	-,093	-,039	-,012	-,008	-,012	,006	-,219	,305	,039	-,010
	CV1	,005	-,024	,070	-,011	,012	,008	-,066	,039	,926	,069
	CV2	-,008	-,102	,025	,054	-,017	-,015	-,001	-,010	,069	,937
Correlación	RS1	,546 ^a	-,162	,493	-,007	,000	-,018	,125	-,203	,006	-,010
	RS2	-,162	,815 ^a	-,291	-,074	,220	,321	,220	-,106	-,038	-,160
	RS4	,493	-,291	,512 ^a	,060	-,088	-,077	,047	-,026	,089	,032
	CE1	-,007	-,074	,060	,933 ^a	-,167	,011	-,275	-,023	-,017	,086
	CE2	,000	,220	-,088	-,167	,723 ^a	,937	,074	-,133	,075	-,105
	CE3	-,018	,321	-,077	,011	,937	,723 ^a	-,068	,070	,052	-,098
	MI1	,125	,220	,047	-,275	,074	-,068	,330 ^a	-,506	-,088	-,002
	TR1	-,203	-,106	-,026	-,023	-,133	,070	-,506	,850 ^a	,073	-,018
	CV1	,006	-,038	,089	-,017	,075	,052	-,088	,073	,836 ^a	,074
	CV2	-,010	-,160	,032	,086	-,105	-,098	-,002	-,018	,074	,642 ^a

a. Medida de adecuación muestral

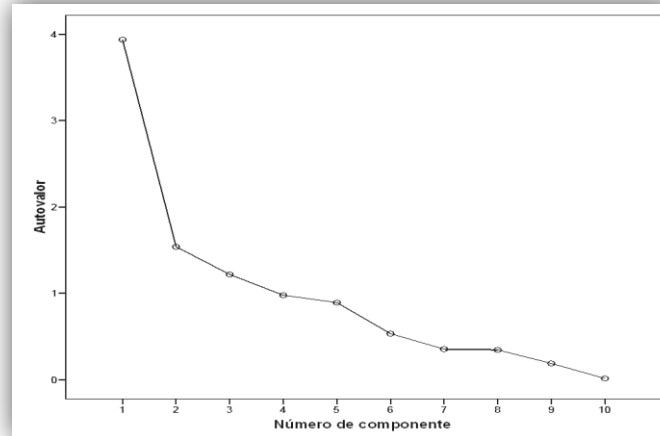
Ahora veamos el número de componentes con el que nos quedaremos, que normalmente son las relativas a valores propios mayores que la unidad. Observando la tabla 3.5.4 de la varianza total explicada vemos que la primera componente explica un 39.39% de la varianza total y las dos siguientes un 15.40% y un 12.18% haciendo un total de 66.97% entre las tres. El gráfico de sedimentación figura 3.5.1 muestra que sólo hay tres componentes con autovalor mayor que 1.

Tabla 3.5.4 – Varianza total explicada

Compo nente	Autovaleores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	% acumulad o	Total	% de la varianza	% acumulad o	Total	% de la varianza	% acumulad o
1	3,939	39,389	39,389	3,939	39,389	39,389	3,765	37,646	37,646
2	1,540	15,396	54,785	1,540	15,396	54,785	1,485	14,851	52,498
3	1,218	12,184	66,969	1,218	12,184	66,969	1,447	14,472	66,969
4	,978	9,777	76,747						
5	,892	8,921	85,667						
6	,533	5,325	90,992						
7	,353	3,535	94,527						
8	,345	3,453	97,980						
9	,188	1,880	99,860						
10	,014	,140	100,000						

Método de extracción: Análisis de Componentes principales.

Figura 3.5.1 – Sedimentación



Ahora construiremos las variables que se agrupan en torno a cada componente y cuya combinación lineal define la componente. Para ello representamos cada una de las diez variables por medio de los tres factores extraídos utilizando la matriz de componentes tabla 3.5.5. Podemos realizar lo siguiente:

Tabla 3.5.5 – Matriz de componentes rotados

	Componente		
	1	2	3
RS1	,233	-,175	,829
RS2	,656	-,488	-,093
RS4	,174	-,193	-,840
CE1	,843	,101	,029
CE2	,919	-,250	-,025
CE3	-,917	,267	,020
MI1	,168	,854	,055
TR1	,882	,100	,103
CV1	-,196	,319	,142
CV2	,087	-,440	,097

Tabla 3.5.6 - Comunalidades

	Inicial	Extracción
RS1	1,000	,773
RS2	1,000	,678
RS4	1,000	,773
CE1	1,000	,722
CE2	1,000	,908
CE3	1,000	,913
MI1	1,000	,760
TR1	1,000	,798
CV1	1,000	,161
CV2	1,000	,210

Para ver qué variables se agrupan en cada componente (factor) hay que observar las variables cuyas cargas sean altas en un factor y bajas en otros (valores menores que 0.25 suelen considerarse bajos).

En la primera componente están representadas claramente RS2, CE1, CE2, CE3 y TR1; en la segunda componente tenemos a RS2, y MI1, CV2 se excluye debido a que su

comunalidad es menor del 40%; finalmente en la tercera componente tenemos RS1 y RS4.

Analizamos de la tabla 3.5.6 la comunalidad de cada variable (suma de cuadrados de sus cargas factoriales definidas en la matriz de componentes) después de la extracción de los factores (componentes). La comunalidad es la parte de variabilidad de cada variable explicada por los factores. Antes de la extracción de los factores la comunalidad de cada variable es la unidad, e interesa que después de la extracción siga siendo alta. Obsérvese que las comunalidades de CV1 y CV2 son menores del 40% por lo que no serán incluidas para estimar el IFLU.

Es posible calcular los coeficientes de correlación entre cada dos variables después de que estén en función de las componentes, denominados *coeficientes de correlación reproducidas*. Estos coeficientes de correlación reproducidos no tienen porqué coincidir con los de la matriz de correlaciones inicial, pero no deben diferenciarse en más de 0.05 (residuos menores que 0.05), porque entonces la bondad del modelo factorial será discutible. En la tabla 3.5.7 se presenta la *matriz de correlaciones reproducidas* en la que se observa que un 42% de los errores son mayores que 0.05, lo que indica que la bondad del modelo no es discutible.

Tabla 3.5.7 – Correlaciones reproducidas

		RS1	RS2	RS4	CE1	CE2	CE3	MI1	TR1	CV1	CV2
Correlación reproducida	RS1	,773 ^b	,161	-,622	,203	,237	-,244	-,065	,273	,016	,178
	RS2	,161	,678 ^b	,287	,501	,728	-,734	-,312	,521	-,298	,263
	RS4	-,622	,287	,773 ^b	,103	,229	-,228	-,182	,048	-,215	,019
	CE1	,203	,501	,103	,722 ^b	,749	-,746	,230	,757	-,129	,032
	CE2	,237	,728	,229	,749	,908 ^b	-,910	-,060	,783	-,263	,188
	CE3	-,244	-,734	-,228	-,746	-,910	,913 ^b	,075	-,780	,268	-,196
	MI1	-,065	-,312	-,182	,230	-,060	,075	,760 ^b	,239	,248	-,355
	TR1	,273	,521	,048	,757	,783	-,780	,239	,798 ^b	-,126	,043
	CV1	,016	-,298	-,215	-,129	-,263	,268	,248	-,126	,161 ^b	-,144
	CV2	,178	,263	,019	,032	,188	-,196	-,355	,043	-,144	,210 ^b
	Residual ^a	RS1		,021	,207	-,032	-,042	,041	,035	-,001	-,030
RS2		,021		,007	-,047	-,083	,055	,064	-,028	,159	-,067
RS4		,207	,007		-,033	-,043	,045	,051	,025	,081	,019
CE1		-,032	-,047	-,033		-,030	,037	-,080	-,129	,011	,006
CE2		-,042	-,083	-,043	-,030		-,074	-,051	-,037	,049	-,049
CE3		,041	,055	,045	,037	-,074		,052	,040	-,067	,068
MI1		,035	,064	,051	-,080	-,051	,052		,009	-,153	,291
TR1		-,001	-,028	,025	-,129	-,037	,040	,009		-,044	,067
CV1		-,030	,159	,081	,011	,049	-,067	-,153	-,044		,036
CV2		-,117	-,067	,019	,006	-,049	,068	,291	,067	,036	

Método de extracción: Análisis de Componentes principales.

a. Los residuos se calculan entre las correlaciones observadas y reproducidas. Hay 19 (42,0%) residuales no redundantes con valores absolutos mayores que 0,05.

Ahora calcularemos las puntuaciones factoriales, que no son más que los valores que toman cada uno de los individuos en las tres componentes seleccionadas. Serán entonces tres variables sustitutas de las iniciales que representan su reducción y que recogen el 66.97% de la variabilidad total. Estas variables son las que se utilizarán como sustitutas de las iniciales para análisis posteriores tales como el análisis de la regresión con problemas de multicolinealidad y el análisis cluster. SPSS incorpora estas variables al conjunto de datos si así se le pide, como se muestra en la figura 1.8 del capítulo 1.

Tabla 3.5.8 – Matriz de coeficientes para el cálculo de las puntuaciones en las componentes

	Componente		
	1	2	3
RS1	,026	-,161	,587
RS2	,128	-,268	-,043
RS4	,044	-,057	-,576
CE1	,256	,181	-,006
CE2	,233	-,064	-,019
CE3	-,230	,077	,014
MI1	,157	,647	-,028
TR1	,265	,180	,045
CV1	-,018	,200	,080
CV2	-,034	-,320	,098

De la matriz de coeficientes de las puntuaciones en las componentes anteriores podemos deducir la siguiente relación entre componentes y variables:

CUADRO 3.5.1 – Componentes principales para construir el IFLU.

$$Z_1 = 0.128RS2 + 0.256CE1 + 0.233CE2 - 0.230CE3 + 0.265TR1$$

$$Z_2 = -0.268RS2 + 0.647MI1$$

$$Z_3 = 0.587RS1 - 0.576RS4$$

Con estas tres componentes construiremos el IFLU a continuación.

TABLA 3.5.9 – Componentes principales IFLU.

INDICADOR	Componente			Comunalidad
	1	2	3	
RS1		-0,1609075	0,58749526	0,772984236
RS2	0,1283749			0,677978987
RS4		-0,0573539	-0,5764550	0,773084597
CE1	0,2555567			0,722192764
CE2	0,2333574			0,907888308
CE3	-0,230419			0,91317863
MI1		0,64686634		0,760458618
TR1	0,2648513			0,79828301
CV1				0,160653345
CV2				0,210240378
AUTOVALORES	3,939	1,54	1,218	6,697

Para permitir una lectura más clara de la matriz de correlación se eliminan las cargas factoriales estimadas con valores absolutos inferiores a 0,40 como para CV1 y CV2. Sin embargo, la estimación de los factores por cada unidad experimental se realiza en SPSS utilizando todas las cargas factoriales.

Las cargas factoriales más significativas en la matriz de componentes principales, indican que el primer componente explica claramente la variabilidad de indicadores de *respuesta* como disposición de residuos sólidos así como cobertura o al uso “común” de energía e infraestructura pública urbana (cobertura de energía, gas y aseo), y el segundo componente parece explicar los factores de **presión** sobre el entorno urbano, como migración, los componentes con signo negativo reflejan un aporte adverso al IFLU, muchos autores afirman que no siempre los componentes pueden tener una explicación directa; en este caso, los tres componentes seleccionados son consistentes con la explicación de tres grupos de variables claramente expresadas por su grado de correlación.

TABLA 3.5.10 – Factores e IFLU.

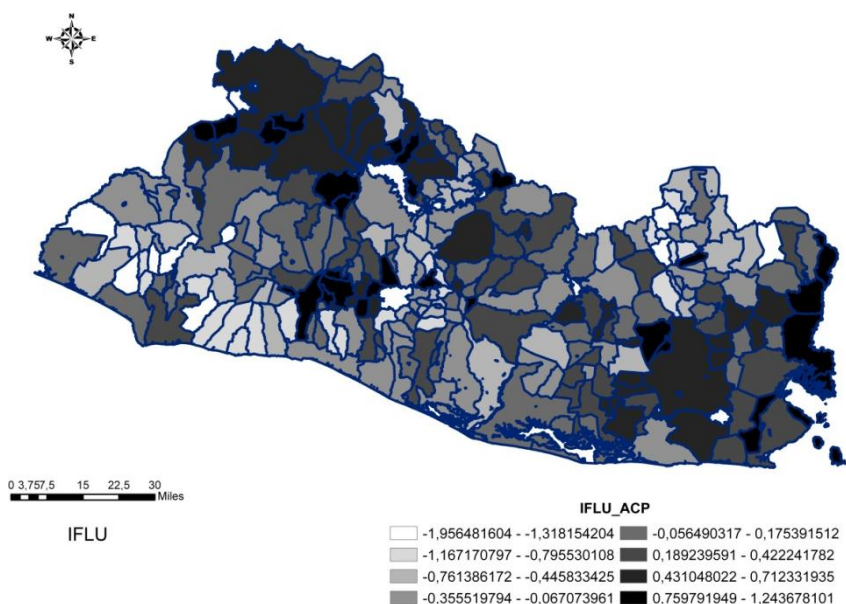
MUNICIPIO	Z1_IFLU	Z2_IFLU	Z3_IFLU	IFLU
AHUACHAPAN	0,35189	-0,48904	-0,64684	-0,14375614
APANECA	0,01573	-1,7858	-0,28609	-0,57760936
ATIQUIZAYA	0,23545	-0,17793	-0,65895	-0,11104524
CONCEPCION DE ATACO	-0,81974	-2,15922	-0,04863	-1,00712193
EL REFUGIO	1,06622	0,23822	-1,23227	0,24294296
GUAYMANGO	-2,30583	-1,28854	-0,25385	-1,49113479
JUJUTLA	-1,18539	0,12672	-0,67684	-0,67964025
SAN FRANCISCO MENENDEZ	-0,15447	-0,09761	0,92265	0,13641003
SAN LORENZO	-0,39445	0,15551	-0,47372	-0,25701459
SAN PEDRO PUXTLA	-1,45348	-2,09445	1,11284	-0,98299956

En la tabla 3.5.10 se presentan solo 10 casos del total de 262 por razones prácticas la tabla completa la podemos consultar en los anexos, pero podemos apreciar los valores del IFLU para estos municipios, el IFLU fue calculado a partir de los factores obtenidos y la ecuación:

$$IU_{mj} = \frac{\sum_{i=1}^r Z_{rj} \sqrt{\lambda_r}}{\sum_{i=1}^r \sqrt{\lambda_r}} \quad (3.5.1)$$

El resultado final es el arreglo de los municipios de acuerdo al valor que resulta del índice parcial estimado. Éste se distribuye espacialmente como se observa en la figura 3.5.2.

Figura 3.5.2 – Representación espacial del IFLU



ESTADO DEL INDICADOR	TABLA 3.5.11 – DISTRIBUCIÓN DEL IFLU		
	REGIÓN OCCIDENTAL	REGIÓN CENTRAL	REGIÓN ORIENTAL
BUENA	TACUBA, GUAYMANGO	SAN JOSÉ GUAYABAL, SAN PEDRO PERULAPAN, VICTORIA	TOROLA, SAN FERNANDO, PERQUIN, CORINTO, CACAOPERA, JOATECA, GUATAJIAGUA
REGULAR	JUAYUA, APANECA, COATEPEQUE	SUCHITOTO, GUAZAPA, TONACATEPEQUE	SOCIEDAD, SAN ALEJO, JUCUARAN, YAYANTIQUE
DEFICIENTE	TEXISTEPEQUE, SAN ANTONIO PAJONAL, SANTIAGO DE LA FRONTERA, MASAHUAT	ILOBASCO, EL PAISNAL, SAN SALVADOR, SAN MARTIN, SANTA RITA, SANTO TOMAS	SAN MIGUEL, MONCAGUA, JOCORO, SANTA ROSA DE LIMA, EL SAUCE, CONCEPCIÓN ORIENTE, PASAQUINA, LA UNIÓN

Para realizar una interpretación un tanto resumida de la calidad ambiental según el Índice de Flujo Urbano (IFLU), éste mide variables de procesos urbanos que modifican la calidad del entorno y nos proporciona la siguiente información: para el occidente del país tenemos a Tacuba con una buena calidad ambiental, a Juayua en estado regular de la misma, y a Texistepeque con una deficiente calidad ambiental. En la tabla 3.5.11 podemos apreciar el estado de algunos de los municipios según este indicador en las regiones Central y Oriental del país y en la figura 3.5.2 podemos tener una apreciación general del estado del resto de municipios en este contexto.

3.6. CONSTRUCCIÓN DEL ÍNDICE DE MEDIO AMBIENTE URBANO (IMAU)

En este caso tenemos dos áreas temáticas definidas, Agua y Mortalidad (véase tabla A.5), las cuales definen únicamente 5 indicadores simples, una de las limitantes mencionadas es la falta de otros indicadores que no se encuentran disponibles, los cuales son muy importantes a la hora de estimar un índice sintético, por consecuencia realizamos nuestro análisis con la información disponible.

Tabla 3.6.1 – Matriz de correlaciones

		AG1	AG2	AG3	AG4	MO1
Correlación	AG1	1,000	,219	,140	,082	,097
	AG2	,219	1,000	,399	,219	,205
	AG3	,140	,399	1,000	,094	,089
	AG4	,082	,219	,094	1,000	,028
	MO1	,097	,205	,089	,028	1,000
Sig. (Unilateral)	AG1		,000	,012	,092	,059
	AG2	,000		,000	,000	,000
	AG3	,012	,000		,065	,075
	AG4	,092	,000	,065		,325
	MO1	,059	,000	,075	,325	

a. Determinante = ,724

Por el determinante de la matriz de correlaciones (tabla 3.6.1) que es alto 0.724, podemos notar que el grado de intercorrelación entre las variables es bajo. Y con la prueba de Bartlett identificamos que existe una cierta correlación significativa entre las variables, siendo su p-valor de 0.000 como se observa en la tabla 3.6.2.

Tabla 3.6.2 – KMO y prueba de Bartlett

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		,600
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	83,463
	gl	10
	Sig.	,000

A la vez nótese que el valor 0.600 nos dice que la adecuación de la muestra al análisis es buena.

Tabla 3.6.3 – Matrices anti-imagen

		AG1	AG2	AG3	AG4	MO1
Covarianza anti-imagen	AG1	,945	-,134	-,052	-,035	-,052
	AG2	-,134	,760	-,293	-,162	-,148
	AG3	-,052	-,293	,838	-,005	-,005
	AG4	-,035	-,162	-,005	,950	,019
	MO1	-,052	-,148	-,005	,019	,955
Correlación anti-imagen	AG1	,720 ^a	-,158	-,058	-,037	-,055
	AG2	-,158	,568 ^a	-,366	-,191	-,174
	AG3	-,058	-,366	,586 ^a	-,005	-,005
	AG4	-,037	-,191	-,005	,629 ^a	,020
	MO1	-,055	-,174	-,005	,020	,641 ^a

a. Medida de adecuación muestral

La matriz de correlaciones anti-imagen presenta coeficientes altos, por lo que las variables no comparten factores comunes. Los elementos de la diagonal de esta matriz son similares al estadístico KMO para cada par de variables e interesa que estén cercanos a la unidad. Observando la matriz de la tabla 3.6.3 vemos que no tenemos resultados ideales.

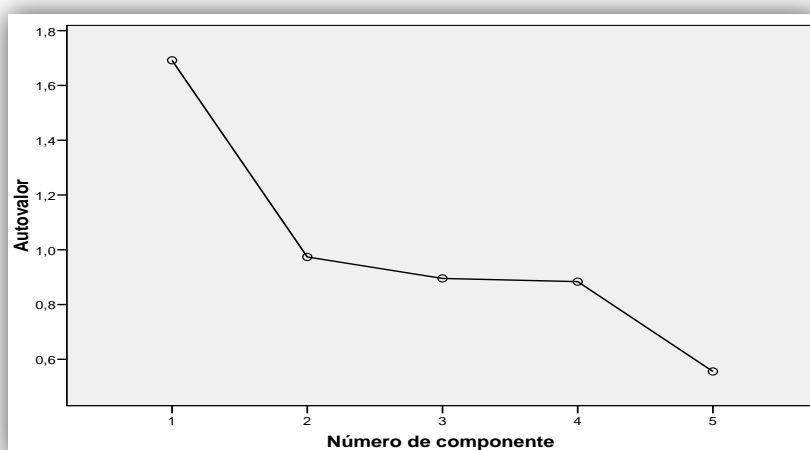
Tabla 3.6.4 – Varianza total explicada

Componente	Autov alores iniciales		
	Total	% de la varianza	% acumulado
1	1,692	33,834	33,834
2	,974	19,474	53,308
3	,895	17,907	71,216
4	,884	17,673	88,888
5	,556	11,112	100,000

Método de extracción: Análisis de Componentes principales

Según la tabla 3.6.4 nos quedaremos con una sola componente principal la cual explica el 33.83% de la varianza total. Esto se verifica también con el gráfico de sedimentación de la figura 3.6.1.

FIGURA 3.6.1 – Sedimentación



Construiremos las variables que se agrupan en torno a la componente y cuya combinación lineal la define. Representamos cada una de las cinco variables por medio del único factor extraído utilizando la matriz de componentes. Podemos realizar lo siguiente:

Tabla 3.6.4 - Comunalidades

	Inicial
AG1	1,000
AG2	1,000
AG3	1,000
AG4	1,000
MO1	1,000

Según la matriz de componente (la cual no mostramos) se extrajo una sola componente, entonces las variables se agruparan en esa única componente.

Los coeficientes de correlación entre cada dos variables indican según la tabla 3.6.5 que la bondad del modelo factorial será discutible, ya que el 100% de los errores son mayores que 0.05. La explicación a este fenómeno es que para el área temática de medio ambiente no se han incorporado muchas variables que influyen en el índice y que de seguro estarían mejor relacionadas con las existentes.

Tabla 3.6.5 – Correlaciones reproducidas

		AG1	AG2	AG3	AG4	MO1
Correlación reproducida	AG1	,248 ^b	,400	,336	,210	,205
	AG2	,400	,643 ^b	,540	,338	,330
	AG3	,336	,540	,453 ^b	,284	,277
	AG4	,210	,338	,284	,178 ^b	,173
	MO1	,205	,330	,277	,173	,169 ^b
Residual ^a	AG1		-,180	-,196	-,128	-,108
	AG2	-,180		-,141	-,119	-,125
	AG3	-,196	-,141		-,190	-,188
	AG4	-,128	-,119	-,190		-,145
	MO1	-,108	-,125	-,188	-,145	

Método de extracción: Análisis de Componentes principales.

a. Los residuos se calculan entre las correlaciones observadas y reproducidas. Hay 10 (100,0%) residuales no redundantes con valores absolutos mayores que 0,05.

b. Comunalidades reproducidas

A continuación calcularemos las puntuaciones factoriales a partir de la tabla 3.6.6, como solo tenemos una componente, será este el único que participará para la construcción del IMAU y nótese que éste será la misma componente, ya que al aplicar la ecuación (67) obtenemos el mismo valor.

Tabla 3.6.6 – Matriz de coeficientes

	Componente
	1
AG1	,295
AG2	,474
AG3	,398
AG4	,249
MO1	,243

Tenemos la componente:

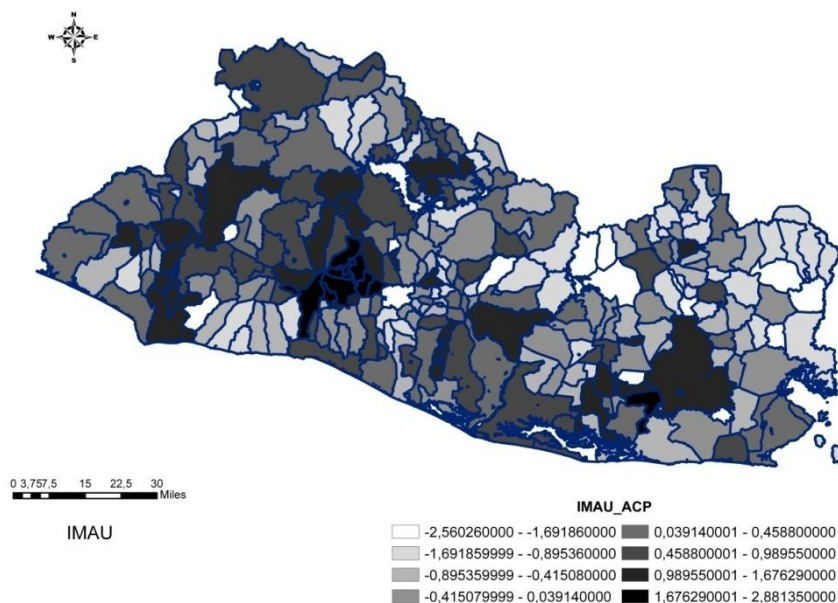
CUADRO 3.6.1 – Componentes principales para construir el IMAU.

$$Z_1 = 0.295AG1 + 0.474AG2 + 0.398AG3 + 0.249AG4 + 0.243MO1$$

TABLA 3.6.7 - IMAU	
MUNICIPIO	IMAU
AHUACHAPAN	0,43686
APANECA	0,70677
ATQUIZAYA	0,58534
CONCEPCIÓN DE ATACO	1,13412
EL REFUGIO	-1,12714
GUAYMANGO	-0,98946
JUJUTLA	-0,62245
SAN FRANCISCO MENENDEZ	0,12796
SAN LORENZO	0,36891
SAN PEDRO PUXTLA	-0,80605

La tabla del IMAU completa la encontramos en anexos. Procedemos a espacializar este indicador en la figura 3.6.1.

Figura 3.6.1 – Representación espacial del IMAU



ESTADO DEL INDICADOR	TABLA 3.6.8 – DISTRIBUCIÓN DEL IMAU		
	REGIÓN OCCIDENTAL	REGIÓN CENTRAL	REGIÓN ORIENTAL
BUENA	COATEPEQUE, JUJUTLA, GUAYMANGO	JUTIAPA, VICTORIA, TAPALHUACA	ANAMOROS, EL SAUCE, PASAQUINA, LISLIQUE, NUEVA ESPARTA, SESORI, SAN GERARDO
REGULAR	TEXISTEPEQUE, MASAHUAT	SAN ISIDRO, SAN SEBASTIAN, ZACATECOLUA	CACAOPELA, CONCEPCIÓN BATRES, CAROLINA, PERQUÍN
DEFICIENTE	METAPAN, SANTA ANA, SONSONATE	EL PAISNAL, AGUILARES, GUAZAPA, SAN SALVADOR, SOYAPANGO	SAN MIGUEL, EL TRANSITO, SANTA ELENA, CIUDAD BARRIOS, SAN FRANCISCO GOTERA

Esta tabla la interpretaremos de la misma manera que la tabla 3.5.11, pero con la diferencia de que estamos frente al Índice de Medio Ambiente Urbano que mide el estado de variables que describen una característica ambiental determinada.

3.7. CONSTRUCCIÓN DEL ÍNDICE SINTÉTICO DE CALIDAD AMBIENTAL (ISCA)

A continuación procedemos a realizar la última fase para la construcción del ISCA, previamente con la información organizada en el SIG se elaboró la línea base de todos los indicadores ambientales simples distribuidos en el espacio. Dado que el interés básico de este estudio es calificar y organizar jerárquicamente los municipios de acuerdo al ISCA, se requiere la utilización de técnicas de análisis multivariante para sintetizar un conjunto de variables relacionadas que explican una **variable latente** como la calidad ambiental.

Aplicados los criterios para la selección de los indicadores idóneos para la construcción del índice sintético de calidad ambiental, se tienen 15 indicadores simples. De ellos, diez participaron en la construcción del IFLU, y los restantes del IMAU con los cuales estimaremos el índice sintético que andamos buscando. A continuación aplicamos el

análisis de componentes principales para calcular el ISCA, procedimiento que hemos utilizado para el IFLU e IMAU.

De la matriz de correlaciones podemos rescatar que el grado de intercorrelación entre las variables es muy alto debido a que su determinante es bastante bajo 0.000.

Tabla 3.6.9 – KMO y prueba de Bartlett

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		,779
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	2201,742
	gl	105
	Sig.	,000

El p-valor es de 0.000 lo cual indica una intercorrelación significativa entre las variable, y una buena adecuación de la muestra al análisis es la conclusión a la que llegamos debido a que la medida de Kaiser-Meyer-Olkin es de 0.779, muy cercana a la unidad (véase tabla 3.6.9).

Al explorar la matriz anti-imagen observamos por lo general valores bajos, lo que indica que las variables comparten factores comunes. Además como se espera los valores de la diagonal de esta matriz son similares al KMO y por lo general también cercanos a la unidad.

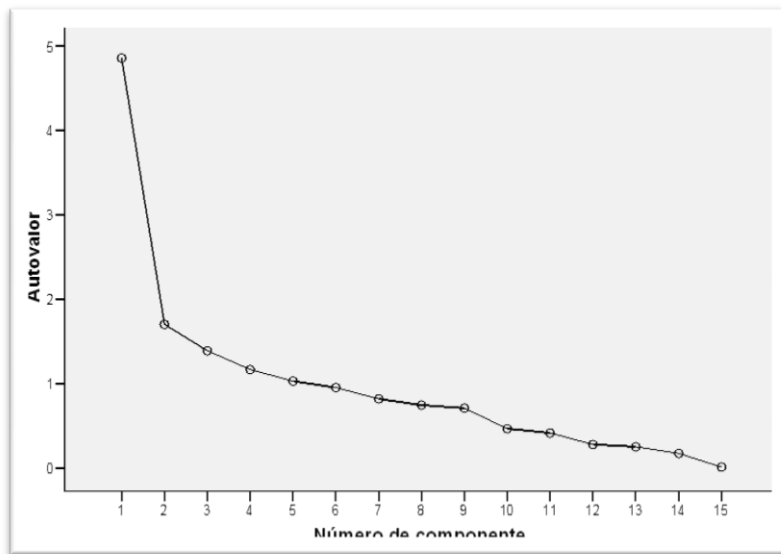
Tabla 3.6.10 – Varianza total explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	4,860	32,403	32,403	4,860	32,403	32,403	4,595	30,636	30,636
2	1,705	11,366	43,769	1,705	11,366	43,769	1,660	11,069	41,704
3	1,392	9,278	53,047	1,392	9,278	53,047	1,478	9,856	51,560
4	1,169	7,796	60,843	1,169	7,796	60,843	1,232	8,216	59,776
5	1,032	6,879	67,722	1,032	6,879	67,722	1,192	7,946	67,722
6	,955	6,370	74,092						
7	,821	5,477	79,568						
8	,747	4,982	84,550						
9	,711	4,737	89,287						
10	,468	3,120	92,407						
11	,417	2,778	95,185						
12	,280	1,869	97,054						
13	,254	1,693	98,747						
14	,175	1,169	99,915						
15	,013	,085	100,0						

Método de extracción: Análisis de Componentes principales.

Según la tabla 3.6.10 y la figura 3.6.2 tomaremos 5 componentes para el análisis debido que es ése el número de autovalores con valor mayor a la unidad. Notemos que el porcentaje de la varianza total que explica la primera componente es de 32.40% y entre las cinco componentes explican el 67.72% de la variabilidad total.

Figura 3.6.2 – Sedimentación



Para construir las variables que se agrupan en torno a cada componente y cuya combinación lineal define la componente representamos las 15 variables por medio de los cinco factores extraídos utilizando la matriz de componentes tabla 3.6.11. Podemos realizar lo siguiente:

Tabla 3.6.11 – Matriz de componentes rotados

	Componente				
	1	2	3	4	5
RS1	,241	-,058	-,813	,079	,053
RS2	,722	-,301	,092	,062	,270
RS4	,187	-,122	,831	,124	,081
AG1	,273	-,694	,039	,082	-,068
AG2	,651	-,180	,080	,012	,453
AG3	,512	-,082	,138	-,371	,317
AG4	,000	,029	,000	,041	,814
CE1	,822	,065	-,058	-,030	-,160
CE2	,912	-,160	-,016	,180	-,030
CE3	-,919	,178	,018	-,144	-,025
MO1	,400	,284	,204	,368	-,035
MI1	,074	,903	-,027	-,057	-,079
TR1	,827	,235	-,136	,161	,002
CV1	-,142	,119	-,119	-,649	,139
CV2	,033	-,061	-,106	,649	,287

Tabla 3.6.12 - Comunalidades

	Inicial	Extracción
RS1	1,000	,732
RS2	1,000	,696
RS4	1,000	,763
AG1	1,000	,569
AG2	1,000	,667
AG3	1,000	,526
AG4	1,000	,666
CE1	1,000	,709
CE2	1,000	,890
CE3	1,000	,899
MO1	1,000	,419
MI1	1,000	,831
TR1	1,000	,783
CV1	1,000	,489
CV2	1,000	,520

De las tablas anteriores tomamos los siguientes resultados: en la primera componente se representan RS2, AG2, AG3, CE1, CE2, CE3, MO1 y TR1; en la segunda, AG1 y MI1; para la tercera componente tenemos RS1 y RS4; para la cuarta componente CV1 y CV2; y para la quinta AG2 y AG4. Debemos mencionar que según la matriz de correlaciones reproducidas, es indiscutible la bondad del modelo.

Para representar las componentes tomamos los coeficientes de la tabla 3.28 y obtendremos las combinaciones lineales que definen las componentes principales para la construcción de nuestro indicador sintético.

Tabla 3.6.12 – Coeficientes para el cálculo de las puntuaciones

	Componente				
	1	2	3	4	5
RS1	,040	-,057	-,563	,062	,054
RS2	,132	-,121	,038	-,033	,148
RS4	,022	-,012	,557	,062	,016
AG1	,028	-,429	-,007	,007	-,148
AG2	,109	-,034	,030	-,058	,328
AG3	,135	-,006	,089	-,377	,214
AG4	-,076	,095	-,032	,060	,733
CE1	,224	,066	-,026	-,116	-,210
CE2	,200	-,049	-,019	,049	-,118
CE3	-,200	,056	,022	-,018	,072
MO1	,078	,227	,147	,283	-,039
MI1	,080	,570	,032	-,017	,005
TR1	,199	,195	-,082	,064	-,041
CV1	,029	,045	-,064	-,536	,135
CV2	-,092	,014	-,104	,569	,271

Cuadro 3.6.2 – Componentes principales para construir el ISCA.

$$Z_1 = 0.132RS2 + 0.109AG2 + 0.135AG3 + 0.224CE1 + 0.200CE2 - 0.200CE3 + 0.078MO1 + 0.199TR1$$

$$Z_2 = -0.429RS2 + 0.57MI1$$

$$Z_3 = -0.563RS1 + 0.554RS4$$

$$Z_4 = -0.536CV1 + 0.569CV2$$

$$Z_5 = 0.328AG2 + 0.733AG4$$

Según las cargas factoriales más significativas en la matriz de componentes principales, el primer componente como para el caso del IFLU explica claramente la variabilidad de indicadores de respuesta como disposición de residuos sólidos así como cobertura o al uso “común” de energía e infraestructura pública urbana (cobertura de energía, gas y aseo), y el segundo componente parece explicar los factores de presión sobre el entorno urbano, migración, el tercero explica factores de disposición de residuos, la cuarta componente explica factores sobre calidad de vivienda, y el quinto componente sobre

calidad del servicio de agua, los componentes con signo negativo reflejan un aporte adverso al ISCA, los cinco componentes seleccionados son consistentes con la explicación de grupos de variables claramente expresadas por su grado de correlación.

Aplicamos la siguiente ecuación, la cual ya conocemos, y de esa manera calculamos el ISCA de lo cual presentamos solo 10 casos en la tabla 3.6.13.

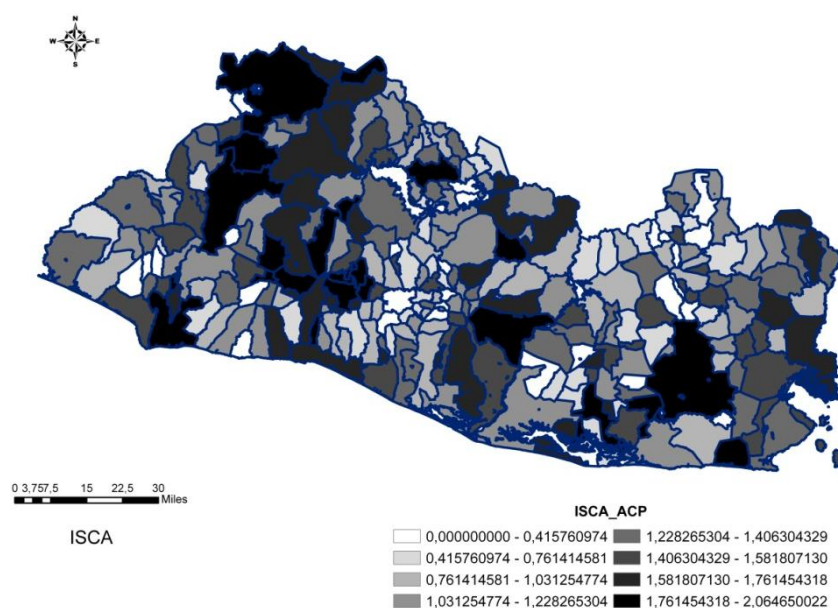
$$ISCA_{mj} = \frac{\sum_{i=1}^r Z_{rj} \sqrt{\lambda_r}}{\sum_{i=1}^r \sqrt{\lambda_r}} \quad (3.6.1)$$

Tabla 3.6.13 – Factores y valores del ISCA.

MUNICIPIO	Z1_ISCA	Z2_ISCA	Z3_ISCA	Z4_ISCA	Z5_ISCA	ISCA
AHUACHAPAN	0,55082	-0,36517	0,67627	-0,39772	-0,12928	0,14350794
APANECA	0,1972	-0,99614	0,31527	0,3208	1,12866	0,14724922
ATIQUIZAYA	0,48335	-0,57095	0,68488	-1,00439	0,10139	0,02138367
CONCEPCION DE ATACO	-0,57658	-1,76664	0,15098	1,33411	1,3626	-0,084437
EL REFUGIO	1,04032	-0,07697	1,12294	-1,13645	-2,34073	-0,01308599
GUAYMANGO	-1,94514	-1,47573	0,40113	-0,46644	-0,13903	-0,94109026
JUJUTLA	-1,31905	0,09213	0,67182	0,38886	0,22554	-0,19822735
SAN FRANCISCO MENENDEZ	-0,39711	0,37436	-0,96067	1,01717	1,2486	0,12496125
SAN LORENZO	-0,29093	-0,83706	0,39818	-1,08992	0,235	-0,32476891
SAN PEDRO PUXTLA	-1,3657	-2,21057	-1,26394	0,84692	-0,37484	-1,00977359

El resultado final es la ordenación de los municipios de acuerdo al valor que resulta del índice sintético estimado. Éste se distribuye espacialmente como se observa en la figura 3.6.3.

Figura 3.6.3 – Representación espacial del ISCA



Esta última fase de la aplicación del modelo es la síntesis de la información contenida en las 15 variables seleccionadas (se incluyeron todas las variables disponibles ya que sus comunalidades eran mayor al 40%) para construir el ISCA. Básicamente se ha pasado de indicadores simples a indicadores sintéticos de componentes (IFLU e IMAU). La razón por la cual los indicadores sintéticos de las 7 áreas temáticas no se han estimado es porque la disponibilidad de información no ha sido suficiente para aplicar ACP a cada una de ellos. Luego de calcular los indicadores sintéticos en el nivel de componentes se construyó el indicador sintético global (ISCA), que resume en cinco componentes el 67,72% de toda la información (de la varianza total) contenida en las variables incorporadas.

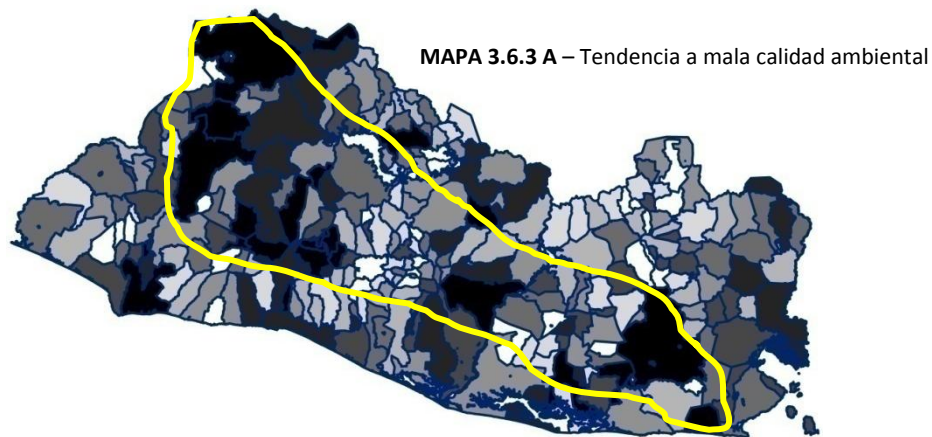
Debido a que se está usando puntajes z en las combinaciones lineales al calcular las puntuaciones de las componentes, los valores de estas pueden ser positivos o negativos. Para evitar que el índice sintético sea negativo se hace una transformación de este índice cambiando el origen 0 de la escala al valor negativo menor que se presenta en el conjunto de los valores del índice. Esto se logra sumando con signo positivo el mínimo valor del conjunto de los valores del índice, a cada uno del conjunto de datos cuyo mínimo es -1,17645, luego se suma 1,17645 a cada uno de los valores del índice, para

tener un nueva escala positiva de 0 a 2,10. El ISCA asume valores entre 0 y 2,10, en donde valores más cercanos a 2,10 son indicativos de calidad ambiental deteriorada.

Los componentes extraídos se expresan en cargas factoriales que tienen un sentido que orienta el análisis de grupos de indicadores simples como explicación de la variable latente ISCA. El primer componente se refiere a indicadores de respuesta a eventuales problemas de recolección de residuos sólidos, abastecimiento de agua potable y disposición de aguas residuales o grises, abastecimiento de energía y gas, explica además, problemas de estado como mortalidad y tráfico; indicando un impacto positivo de algunas de estas variables y negativo de otras en el ISCA. El componente dos no tiene un significado completamente claro, pero se destaca la participación de recolección de residuos y migración. El tercer componente incluye exclusivamente variables de generación y recolección de desechos sólidos. El cuarto componente se enfoca en la situación de vivienda y la quinta explica variables que describen la disponibilidad o cobertura tanto de aguas grises como potable y se relacionan positivamente con el ISCA.

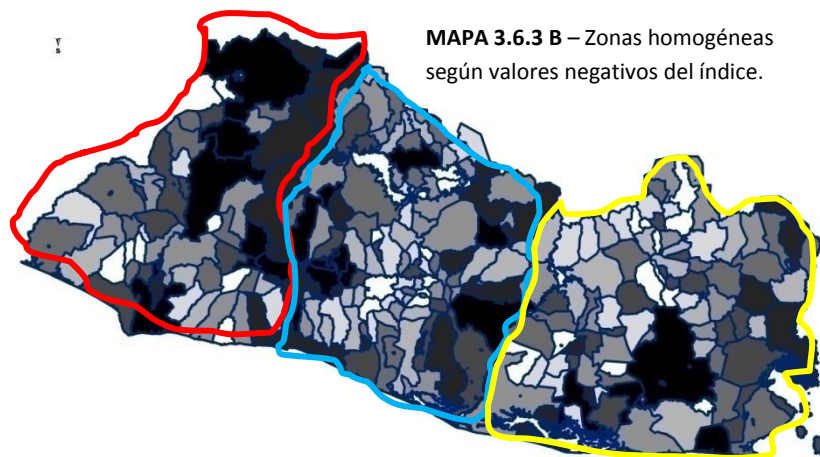
En la tabla 3.6.13 presenta los factores calculados para cada una de las 262 municipalidades de El Salvador y se estima el ISCA aplicando la ecuación 3.6.1. El resultado final es la ordenación de los municipios de acuerdo al indicador sintético de calidad ambiental, que se distribuye espacialmente como se observa en el mapa 3.6.3.

La distribución espacial del índice sintético de calidad ambiental por municipios en El Salvador indica zonas que pueden ser, a este nivel de agregación, definidas como relativamente homogéneas. Este mapa (3.6.3 A) indica cómo en la medida que las ciudades se extienden de occidente a oriente, se forma una ‘diagonal’ de las zonas con mala calidad ambiental. Nos concentraremos más en las zonas de valores negativos ya que es allí donde se presentan deficiencias en la calidad del medio ambiente, para que en materia de política pública se puedan tomar decisiones enfocadas a dar soluciones acertadas.



En general, la distribución espacial de la calidad ambiental de los municipios muestra tres áreas relativamente homogéneas, que se diferencian con respecto al valor promedio del índice (véase mapa 3.6.3 B). Es decir, que en la medida que el valor del índice es superior o inferior al promedio (de acuerdo al sentido de cada indicador simple en el índice), éste tiende a mostrar mejor o peor calidad ambiental.

Coincide la menor calidad ambiental con los municipios de las zonas socioeconómicas más vulnerables del país; por ejemplo, la zona occidental (municipios Metapán, Texistepeque, Santa Ana) y en la zona central los municipios San Salvador, Soyapango, Quezaltepeque, Aguilares, San Vicente, influenciados de la contaminación por diversas causas. En la zona oriental podemos ver claramente a San Miguel como el más afectado en calidad ambiental así como El Tránsito, Intipuca, Usulután. Esto se resume mejor en el cuadro 3.6.14.



ESTADO	CUADRO 3.6.14 – RESUMEN DEL ESTADO DEL ISCA POR REGIÓN		
	REGIÓN OCCIDENTAL	REGIÓN CENTRAL	REGIÓN ORIENTAL
BUENO	Guaymango, Jujutla, Tacuba	San Miguel Tepezontes, San Pedro Perulapan	San Agustín, Alegría, Guatajiagua, Yamabal, Arambala, San Gerardo
REGULAR	Ahuachapán, San Francisco Menéndez	Suchitoto, Ilobasco, Victoria, San José Guayabal	Berlín, Ciudad Barrios, Guatajiagua, San alejo, Bolivar, Sensembra.
MALO	Santa Ana, Metapán, Texistepeque	San Salvador, Soyapango, San Vicente, Zacatecoluca	Usulután, El Tránsito, San Miguel, Intipuca, Santa Rosa de Lima, Pasaquina.

Desde el punto de vista de la gestión ambiental urbana, el modelo y los resultados presentados en esta investigación son una importante herramienta para diferenciar la gestión de los tomadores de decisión en la ciudad como territorio ambientalmente no homogéneo, de tal forma que puedan objetivamente priorizar zonas a intervenir y los recursos a invertir en distintos tipos de acciones que afecten la calidad ambiental en cada municipio, incidiendo esto en el uso eficiente de los recursos públicos.

Para precisar aún más lo expresado en el párrafo anterior, en la tabla 3.6.15 se ordenan de manera general los indicadores simples de acuerdo a su estructura de correlación con el ISCA, presentando la importancia relativa de cada una de las variables que componen el índice. De otro lado, también se describen, para cada municipio, cuáles son las variables que más inciden en la modificación del ISCA, identificando el número de desviaciones estándar de cada variable en cada municipio, que se aleja positiva o negativamente con respecto a la media. De esta forma, los tomadores de decisiones tienen información útil no sólo para priorizar la intervención por municipio o distintos sectores del país, sino que pueden priorizar sobre qué variables o factores intervenir en cada una de ellas de manera específica.

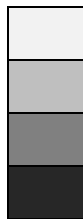
Podemos dar una descripción tanto general a nivel de país, como también por departamento, y dado a la dificultad de presentar una tabla con toda la información agregada tenemos un par de departamentos para los cuales analizamos su situación en torno al ISCA. En la tabla 3.6.15 tenemos el indicador sintético de calidad ambiental detallado para el departamento de Ahuachapán.

La siguiente tabla es importante porque indica la correlación del ISCA con los indicadores simples por municipio, se presentan los valores de menor a mayor en escala de grises para darle un significado visual representativo de la condición ambiental en que se encuentra cada lugar. Se definen además tres zonas o áreas de interés la primera es aquella en que el indicador presenta una *buena* calidad ambiental, la segunda es la que clasificamos como calidad ambiental *regular* y la tercera zona de mucho interés (sin restarle importancia a las otras dos) es la de calidad ambiental *deficiente*, zona en la cual la intervención de política pública debe priorizarse. Aunque para los municipios Atiquizaya, San Francisco Menéndez, Ahuachapan y Apaneca se tienen valores menores o negativos (aquellos que tienen un color moderadamente oscuro), se han clasificado éstos con una calidad ambiental ‘buena’ por su apreciación general. En cambio para El Refugio, San Pedro Puxtla, Tacuba y Guaymango observamos una mayor frecuencia de valores muy negativos. Es así como podríamos interpretar y analizar esta tabla para los otros departamentos y para el país en general.

Tabla 3.6.15 – Importancia relativa de cada variable en relación al ISCA

ÍNDICE SINTÉTICO DE CALIDAD AMBIENTAL		AHUACHAPAN											
MUNICIPIO	CORRELACIÓN ISCA E INDICADORES	ATIQUIZAYA	SAN FRANCISCO MENENDEZ	AHUACHAPAN	APANECA	TURIN	CONCEPCION DE ATACO	JUJUTLA	SAN LORENZO	EL REFUGIO	SAN PEDRO PUXTLA	TACUBA	GUAYMANGO
		Calidad ambiental buena				Calidad ambiental regular				Calidad ambiental deficiente			
CE3 = Uso de carbón, leña, etc	-0,61	-0,37	0,13	-0,41	-0,05	-0,89	0,6	0,86	0,7	-1,1	0,99	1,41	1,54
CE2 = Cobertura de gas	0,61	0,4	-0,09	0,45	0,08	0,95	-0,58	-0,84	-0,69	1,16	-0,95	-1,32	-1,52
MO1 = Mortalidad	0,582	0,23	0,22	0,05	-0,17	-1,32	0,48	-0,12	-0,51	0,02	-2,34	-0,5	-0,71
TR1 = Tráfico	0,55	0,2	-0,26	0	-0,17	0,28	-0,49	-0,77	-0,21	0,41	-0,67	-1,1	-1,06
AG2 = Cobertura de aguas grises	0,517	0,42	-0,42	0,19	0,18	-0,59	1,35	-0,71	-0,42	-0,72	-0,62	-0,22	-0,67
RS2 = Cobertura de recolección	0,503	0,09	-0,5	0,36	1,41	0,26	1,19	-1,05	-0,42	1,4	-0,61	-0,62	-0,87
CE1 = Cobertura de energía eléctrica	0,485	0	-0,43	0,1	-0,34	0,52	-1,45	-1,75	-0,06	0,48	-1,78	-2,77	-3
RS4 = Evacuación de desechos sólidos	0,455	0,95	-0,33	0,91	1,43	1,6	0,71	0,53	0,71	0,95	-0,66	0,41	0,95
AG3 = Cobertura de agua	0,347	0,14	-0,92	0,67	0,81	1,28	-0,57	-0,56	-0,17	0,32	0,11	0,12	-0,45
MI1 = Migración	0,308	-0,85	0,23	-0,77	-1,08	-0,91	-1,15	-0,29	-0,84	-0,45	-1,17	-1,32	-1,16
AG4 = Servicio de agua	0,296	-0,11	1,27	-0,14	0,65	0,05	0,71	0,96	0,97	-1,84	-1,3	0,64	-1,2
CV2 = Densidad poblacional	0,27	-0,49	0,85	-1,04	-0,03	-0,05	0,86	-1,04	-1,95	-2,28	1,4	0,38	0,3
CV1 = Hacinamiento	0,199	0,21	-0,22	-0,63	-0,76	-0,93	-0,55	-0,8	0,98	-0,49	-0,93	0,98	0,21
RS1 = Residuos sólidos generados	-0,14	-0,43	0,8	-0,16	0,78	-0,51	0,22	-0,47	-0,13	-0,57	0,75	-0,31	-0,14
AG1 = Inodoro conectado a alcantarillado	-0,07	0,4	0,22	-0,34	-0,18	-1,17	1,34	-0,08	0,85	-0,57	0,93	0,11	0,29

Muy bueno
Bueno
Regular
Deficiente



$\sigma > 1$
 $0 < \sigma \leq 1$
 $-1 < \sigma \leq 0$
 $\sigma \leq -1$

Los valores indican el número de desviaciones con respecto a la media de cada indicador y los tonos de blanco a grises indican agrupaciones de valores inferiores y superiores a la media tal como se expresa en la matriz de la tabla 3.6.15.

ANÁLISIS DE RESULTADO Y CONCLUSIONES.

En términos generales, podemos decir que los resultados indican que las condiciones de calidad ambiental en las distintas regiones serán heterogéneas en algunas regiones y no así en otras, habiéndose logrado representar en el espacio geográfico el valor que tiene cada unidad experimental, se ha determinado finalmente un conjunto de municipios que comparten valores del índice de calidad ambiental similares y que pueden ser clasificadas como relativamente homogéneas de acuerdo al índice de calidad ambiental estimado.

La importancia operativa de estos resultados indican que se ha construido un importante instrumento para la definición de política pública en materia de medio ambiente urbano, con la cual se puede diferenciar espacialmente el grado de intervención pública en las diversas zonas y establecer criterios de prioridad en la asignación de recursos. Es decir, que el tomador de decisiones tendrá una herramienta cuantitativa importante que le permitirá diferenciar qué factores ambientales son más importantes en cada una de las zonas (municipio, departamentos, etc.) y dirigir su gestión e inversión donde genere el mayor impacto. También será una importante herramienta para definir una línea base de la situación ambiental de cada una de las zonas, con fines de establecer mecanismos de seguimiento, control y evaluación de la política pública.

Técnicamente aquí se ha estimado el ISCA como una variable latente que ha demandado la construcción de un sistema de indicadores ambientales coherente y conceptualmente consistente con lo que se ha intentado representar: una medida cuantitativa del valor relativo de las condiciones ambientales de las unidades experimentales. Consideramos que el ISCA, a este nivel de agregación, será un poderoso instrumento que permitirá resumir una gran cantidad de datos en un solo índice que resume sintéticamente la mayor parte de la información (67%) contenida en las variables que determinan la calidad ambiental en zonas urbanas y a la escala de análisis que se desarrolle esta investigación, es importante recalcar que el 33% de la variabilidad no explicada nos indica que es hay algunos municipios clasificados con una buena calidad ambiental podrían tener una peor o mala calidad ambiental realmente.

Pero de manera general, se tiene la metodología a seguir para la construcción de indicadores de este tipo, y para el caso de las zonas urbanas a nivel de municipios de El Salvador podemos percibir una mala calidad en zonas urbanas masivas como las capitales más importantes de cada departamento, como era de esperarse.

ANEXOS

ANEXOS

Tabla A.1 (Datos del ejemplo 1.1): Este conjunto contiene 34 observaciones y 3 variables. Las observaciones corresponden a distintas acciones que cotizan en el mercado continuo español y las variables a tres medidas de rentabilidad de estas acciones durante un período de tiempo. Las variables son: X1 es la rentabilidad efectiva por dividendos, X2 es la proporción de beneficios que va a dividendos y X3 el ratio entre precio por acción y beneficios.

ACCIONES

	X1	X2	X3		X1	X2	X3
1	3.4	89.7	30.2	18	4.4	58.5	12.1
2	5.1	55.7	9.9	19	7.8	84.3	11
3	4.5	52.3	11.5	20	16	96.5	6
4	3.5	47	11.2	21	16.7	100	6.8
5	5.9	42.7	7	22	15.2	92.3	5.2
6	5.1	30.6	6.9	23	17.5	99.9	6.8
7	4.6	64.4	11.8	24	16.2	93.5	5.2
8	5	51	9.6	25	14.7	100	6.6
9	3.2	54.4	14.7	26	15.3	99.9	5.9
10	3.4	45.7	13.2	27	15.8	100	6.9
11	6.5	39.9	5.2	28	18.3	96.3	5.7
12	4.4	40.3	13.7	29	15.9	100	6.1
13	5.1	52.4	11	30	16.1	92.5	6.1
14	5.8	43.9	8	31	9.7	87.6	7.7
15	4.6	52.8	14.4	32	6.9	53.6	6.6
16	7.2	65.8	7.8	33	14.4	87.8	5.2
17	7.2	58.1	7.7	34	14.9	34.5	4.7

Tabla A.2 (Datos del ejemplo 1.3- EPF): Estos datos corresponden a 51 observaciones y 9 variables. Las observaciones son las provincias españolas más Ceuta y Melilla, que aparecen

unidas como una única provincia, y las variables los nueve epígrafes en los que se desglosa la Encuesta de Presupuestos Familiares en España. Las variables son: X1=alimentación, X2=vestido y calzado, X3=vivienda, X4=mobiliario doméstico, X5=gastos sanitarios, X6=transporte, X7=enseñanza y cultura, X8=turismo y ocio, X9=otros gastos.

ciudad	x1	x2	x3	x4	x5	x6	x7	x8	x9
almeria	618957	295452	522128	167067	58288	280035	129219	307967	107334
cadiz	683940	203619	426690	124162	60657	285888	127792	313238	83523
cordoba	590770	265604	487143	113386	37131	237320	116764	247536	79924
granada	547353	238822	458338	119540	40340	236694	103901	272308	72813
huelva	649225	245722	570631	99250	61953	253286	123244	238880	83070
jaen	556210	183295	332662	86364	37160	136992	57607	189811	57311
malaga	617778	201348	508252	121010	63518	256973	128336	323632	93971
sevilla	621570	208156	549399	137408	45101	298000	118269	308524	84514
huesca	577107	249310	412907	107976	39602	335334	90547	227266	92103
teruel	545238	199788	343919	122154	42281	224286	90291	237747	77938
zaragoza	556737	266468	496989	132517	54106	235188	118931	282369	79718
asturias	624941	280273	530828	132066	57679	340013	149265	315478	120856
baleares	564220	226816	602397	144005	86803	358290	150551	351555	131802
las palmas	632640	201704	522846	153775	84148	327988	173031	305628	114627
tenerife	523476	171072	467424	118857	65247	303598	142620	283563	80959
cantabria	604083	287943	654929	119269	63320	302277	116752	276663	105421
avila	543595	242609	388063	92808	47035	254563	74522	250853	82061
burgos	602307	255567	600121	162166	51308	280023	132161	301813	111224
leon	623047	245240	500414	136030	41667	333066	119657	267506	146434
palencia	589710	206048	449113	113714	34787	248620	115825	294068	109264
salamanca	488784	167814	400049	78217	24476	195065	69846	193056	54442
segovia	528040	184840	455368	103446	46337	217156	91436	259705	116303
soria	679722	232673	503695	129768	55000	272249	117587	300014	120803
valladolid	567361	223201	566433	140573	46111	254216	149041	327774	98430
zamora	544527	178835	402354	99953	32143	227163	70283	231577	125332
albacete	535939	199559	425598	137799	55967	232209	104866	291708	91735
ciudad real	545912	227255	487651	125740	44001	230820	88650	230213	90886
cuenca	506814	194156	420488	109533	50961	220678	78673	270038	103288
guadalajara	546909	179824	477446	115585	40129	299174	94923	287703	87720
toledo	583325	255527	411896	130747	65345	282127	105872	241749	122189
barcelona	702920	257429	702315	168696	97595	365255	239187	379319	99929
gerona	684186	285047	566149	149308	77553	259839	191400	329089	134786
lerida	696542	283134	508906	146773	90828	402073	180652	353124	152924
tarragona	586122	283112	557653	150464	62853	331848	185713	381485	114876
alicante	579355	205685	490235	134254	68141	297939	117710	316675	111756
castellon	496559	201606	411972	107739	42939	212051	84610	241795	77370
valencia	539570	228072	464127	138419	62471	285948	134751	384939	96564
badajoz	430442	204529	332948	91831	41112	187500	77481	203808	61478

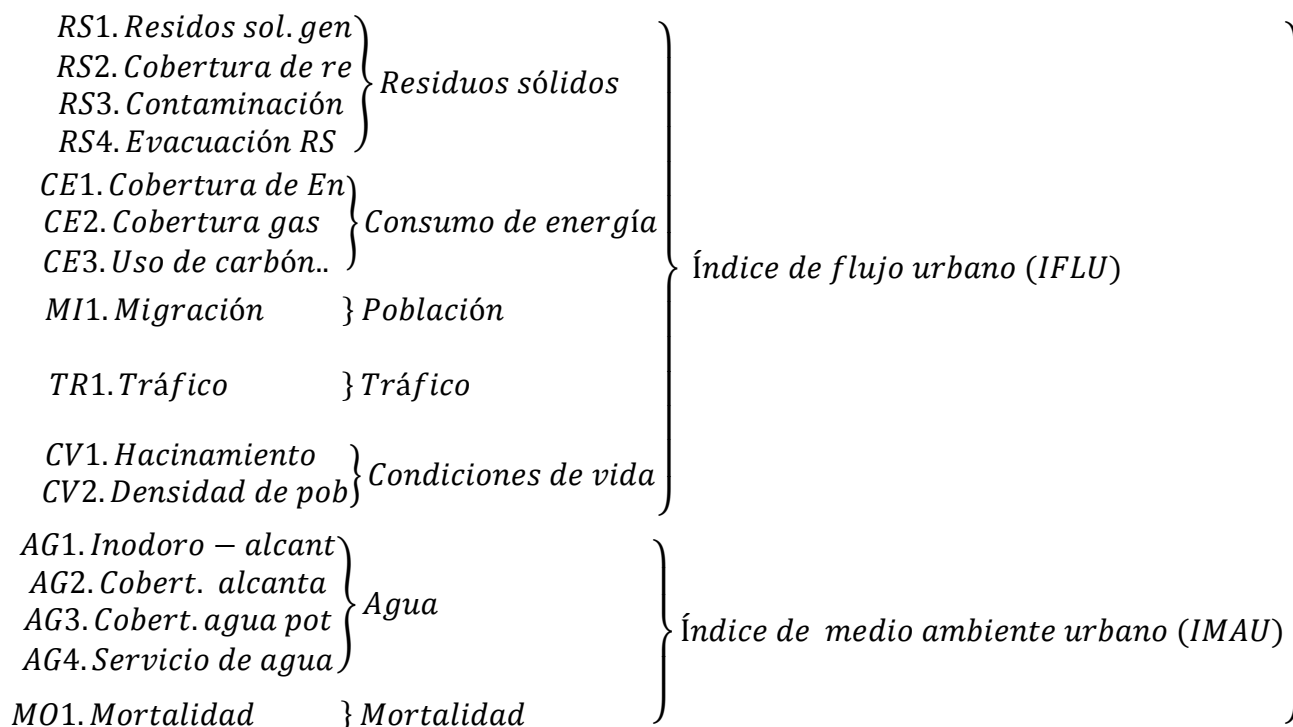
caceres	569808	222756	403547	119078	47904	248571	100282	285880	89736
la coruña	692445	249121	506616	141805	56114	277401	142246	289111	108489
lugo	719078	286277	414893	142629	62779	301867	101889	216693	119398
oreense	598669	182378	370866	106873	31779	221028	114728	205921	90184
pontevedra	736441	263479	468984	136204	50815	344289	129685	309349	100255
madrid	670577	253928	864553	148014	86601	393664	232387	440275	130200
murcia	610718	210169	470859	128627	46866	318508	102978	311262	114457
navarra	669082	324877	704572	221954	81180	415313	185493	411027	156493
alava	664450	234132	631137	189169	58406	313033	164730	355280	98942
guipuzcoa	643141	254653	668435	151454	61985	302491	169527	405259	109995
vizcaya	635929	283160	677817	156612	67899	337253	176222	423122	132572
la rioja	634839	209753	542656	127615	54684	269843	126717	322845	121844
ceuta y mililla	678733	192344	362317	81673	27191	138705	81979	226279	65135

TABLA A.3. EMPRESAS

id	agr	min	man	cen	con	ser	ban	secser	tc
Bélgica	3,3	,9	27,6	,9	8,2	19,1	6,2	26,6	7,2
Dinamarca	9,2	,1	21,8	,6	8,3	14,6	6,5	32,2	7,1
Francia	10,8	,8	27,5	,9	8,9	16,8	6,0	22,6	5,7
Alemania O	6,7	1,3	35,8	,9	7,3	14,4	5,0	22,3	6,1
Irlanda	23,2	1,0	20,7	1,3	7,5	16,8	2,8	20,8	6,1
Italia	15,9	,6	27,6	,5	10,0	18,1	1,6	20,1	5,7
Luxemburgo	7,7	3,1	30,8	,8	9,2	18,5	4,6	19,2	6,2
Holanda	6,3	,1	22,5	1,0	9,9	18,0	6,8	28,5	6,8
Reino Unido	2,7	1,4	30,2	1,4	6,9	16,9	5,7	28,3	6,4
Austria	12,7	1,1	30,2	1,4	9,0	16,8	4,9	16,8	7,0
Finlandia	13,0	,4	25,9	1,3	7,4	14,7	5,5	24,3	7,6
Grecia	41,4	,6	17,6	,6	8,1	11,5	2,4	11,0	6,7
Noruega	9,0	,5	22,4	,8	8,6	16,9	4,7	27,6	9,4
Portugal	27,8	,3	24,5	,6	8,4	13,3	2,7	16,7	5,7
España	22,9	,8	28,5	,7	11,5	9,7	8,5	11,8	5,5
Suecia	6,1	,4	25,9	,8	7,2	14,4	6,0	32,4	6,8
Suiza	7,7	,2	37,8	,8	9,5	17,5	5,3	15,4	5,7
Turquía	66,8	,7	7,9	,1	2,8	5,2	1,1	11,9	3,2
Bulgaria	23,6	1,9	32,3	,6	7,9	8,0	,7	18,2	6,7
Checoslovaqu	16,5	2,9	35,5	1,2	8,7	9,2	,9	17,9	7,0
Alemania E	4,2	2,9	41,2	1,3	7,6	11,2	1,2	22,1	8,4
Hungría	21,7	3,1	29,6	1,9	8,2	9,4	,9	17,2	8,0
Polonia	31,1	2,5	25,7	,9	8,4	7,5	,9	16,1	6,9
Rumanía	34,7	2,1	30,1	,6	8,7	5,9	1,3	11,7	5,0
Rusia	23,7	1,4	25,8	,6	9,2	6,1	,5	23,6	9,3
Yugoslavia	48,7	1,5	16,8	1,1	4,9	6,4	11,3	5,3	4,0

Tabla A.4. Sistema de indicadores de calidad ambiental urbana.

Nivel I - Indicadores simples nivel II - Área temática
 nivel III - Componentes nivel IV - Índice



FICHAS TÉCNICAS DE INDICADORES

Los indicadores que se presentan en la tabla 2.1 están clasificados por grupos temáticos, debido a su naturaleza y forma de cálculo. Es así que, por ejemplo, Irs_j , representa el grupo de indicadores del área temática de Residuos Sólidos Generados. A continuación presentamos con detalle, tanto la definición como la forma de cálculo, para cada indicador de la tabla 2.1.

1. Indicador de Residuos Sólidos Generados

Definición: Entenderemos por Indicador de *Residuos Sólidos Generados*, aquel que mide el promedio en toneladas de la cantidad de desechos sólidos que se generan por la población de un municipio. Por tanto tenemos:

$$Irs_j = \sum_{j=1}^n W_1 RS_{1j}$$

Donde $j = 1, 2, \dots, 262$; W_1 es el conjunto de pesos relativos del indicador.

Entonces $Irs_j = \sum_{j=1}^n W_1 RS_{1j}$, sería el *Indicador de Residuos Sólidos Generados* (en toneladas).

Fuente y forma de cálculo: los datos han sido recopilados del Segundo Censo de Desechos Sólidos. Y es el cociente:

$$\frac{\text{Desechos Sólidos Generados}}{\text{Población Atendida}} \times 1000$$

Interpretación: Por ejemplo, para el municipio de Aguilares el valor que tenemos en éste índice es de 13.74 toneladas, esto nos dice que por cada mil habitantes se genera esa proporción de residuos sólidos.

2. Indicador de Cobertura de Recolección.

Definición: Entenderemos por Indicadores de Cobertura de Recolección, aquel que mide la proporción de la población con acceso a recolección o servicio de aseo, por alguna forma de recolección que por lo general es con el tren de aseo municipal, servicio particular u otra. Entonces:

$$Irs_j = \sum_{j=1}^n W_2 RS_{2j}$$

Donde $j = 1, 2, \dots, 262$; W_2 es el conjunto de pesos relativos del indicador.

Fuente y forma de cálculo: los datos han sido recopilados del VI Censo de Población y V de vivienda 2007. Y es el cociente:

$$\frac{\text{Servicio municipal} + \text{Servicio particular} + \text{La deposita en contenedores}}{\text{Total de población atendida}} \times 100$$

Interpretación: Para el caso en que se tenga, por ejemplo, para el municipio de Ahuachapán un valor de 36.21 esto indica que del total de la población en ese municipio, únicamente el 36.21% tiene servicio de aseo, cualquiera sea la manera de éste, servicio de aseo municipal, particular u otra forma.

3. Indicador de Contaminación por Desechos Sólidos.

Definición: El *Indicador de Contaminación por Desechos Sólidos* se definirá como el procedimiento que se le da a los desechos sólidos generados que provocan contaminación al medio ambiente. Podemos verlo también como el porcentaje de población que al no contar con servicio de aseo, obligatoriamente contamina su medio.

$$Irs_j = \sum_{j=1}^n W_3 RS_{3j}$$

Donde $j = 1, 2, \dots, 262$; W_3 pesos relativos del indicador.

Fuente y forma de cálculo: los datos han sido recopilados del VI Censo de Población y V de vivienda 2007. Y es el cociente:

$$\frac{La\ quema + La\ entierra + La\ tira\ en\ la\ calle, barranca\ o\ baldio + La\ tira\ en\ el\ rio, lago\ o\ mar + Otra}{Total\ de\ poblaci3n}$$

Interpretación: Por ejemplo, si para el municipio de Sonsonate, el valor que se presenta en este indicador, es 41.17, esto nos dice que el porcentaje de habitantes que genera contaminación (se genera contaminación por el hecho de que de alguna manera la población sin servicio de tren de aseo se deshace de la basura, quemándola, tirándola en los ríos, enterrándola, etc.) es del 41.17% en ese municipio.

4. Indicador de Evacuación de Desechos.

Definición: El *Indicador de Evacuación de Desechos Sólidos* le definimos como el porcentaje evacuado por el tren de aseo, de residuos sólidos generados en las zonas urbanas en toneladas, por municipio.

$$Irs_j = \sum_{j=1}^n W_4 RS_{4j}$$

Donde $j = 1, 2, \dots, 262$; W_4 pesos relativos del indicador.

Fuente y forma de cálculo: Los datos han sido recopilados del Segundo Censo de Desechos Sólidos. Y es el cociente:

$$\frac{Total\ de\ residuos\ s3lidos\ retirado}{Total\ de\ residuos\ s3lidos\ generados} \times 100$$

Interpretación: Del total de basura que se genera en las zonas urbanas, se evacua un porcentaje. Por ejemplo, para el municipio de San Miguel si tenemos 0.7094, esto quiere decir que de toda la basura que se genera en la zona urbana, solamente el 70.94% es retirada o evacuada.

5. Indicador Cobertura de Energía Eléctrica.

Definición: Entenderemos por indicadores de *Consumo de Energía* aquellos que por su naturaleza, en cuanto a la recolección de los datos y a su forma de cálculo, se clasifican en ésta área temática. Dentro de esta clase tenemos:

$$Ice_j = \sum_{j=1}^n W_i CE_{ij}$$

Entonces $Ice_j = \sum_{j=1}^n W_1 CE_{1j}$, sería el *indicador Cobertura de Energía Eléctrica*. Que se define como el porcentaje de la población que tiene acceso a energía eléctrica por municipio.

Fuente y forma de cálculo: los datos han sido recopilados del IV Censo de Población y V de Vivienda 2007. Y es el cociente:

$$\frac{\text{Población con acceso a Energía Eléctrica por Municipio}}{\text{Población Total del Municipio}}$$

Interpretación: Por ejemplo, si para el municipio de El Refugio en Ahuachapán observamos el valor 0.8816, eso nos dice que el 88.16% de la población de ese municipio tiene acceso a energía eléctrica.

6. Indicador de Cobertura de Gas.

Definición: El Indicador Cobertura de Gas se define como el porcentaje de la población que tiene acceso a Gas, sea propano o kerosene, por municipio. Y tenemos:

$$Ice_j = \sum_{j=1}^n W_2 CE_{2j}$$

Donde $j = 1, 2, \dots, 262$; W_2 pesos relativos del indicador.

Fuente y forma de cálculo: los datos han sido recopilados del IV Censo de Población y V de Vivienda 2007. Y es el cociente:

$$\frac{\text{Población con acceso a Gas por Municipio}}{\text{Población Total del Municipio}}$$

Interpretación: Por ejemplo, si para el municipio de Coatepeque observamos el valor 0.4567, eso nos dice que el 45.67% de la población de ese municipio tiene acceso a gas propano o kerosene.

7. Indicador Uso de Carbón, Leña u otro material contaminante.

Definición: porcentaje de la población que utiliza carbón, leña, plástico, paja, batirá de carro, desechos u otro material contaminante, sea para cocinar o para alumbrar, por municipio.

$$Ice_j = \sum_{j=1}^n W_3 CE_{3j}$$

Donde $j = 1, 2, \dots, 262$; W_3 pesos relativos del indicador.

Fuente y forma de cálculo: los datos han sido recopilados del IV Censo de Población y V de Vivienda 2007. Y es el cociente:

$$\frac{\text{Población que utiliza Carbón, Leña, plástico, etc., por Municipio}}{\text{Población Total del Municipio}}$$

Interpretación: Por ejemplo, si para el municipio de Guazapa tenemos un valor de 0.2860, eso nos advierte que el 28.60% de la población de ese municipio utiliza o quema material contaminante, como lo es el carbón, leña, plástico u otro.

8. Indicador de Emigración

Entenderemos por *indicadores de Población* aquellos que por su naturaleza, en cuanto a la recolección de los datos y a su forma de cálculo, se clasifican en ésta área temática. Tenemos:

$$Ipo_j = \sum_{j=1}^n W_i PO_{ij}$$

En nuestro caso $i = 1$, entonces $Ipo_j = \sum_{j=1}^n W_1 EM_{1j}$, sería el *indicador Emigración*. Que se define como el porcentaje de la población que ha experimentado la emigración de al menos un familiar.

Fuente y forma de cálculo: los datos han sido recopilados del IV Censo de Población y V de Vivienda 2007. Y es el cociente:

$$\frac{\text{Población con al menos un familiar emigrado por municipio}}{\text{Población Total del Municipio}}$$

Interpretación: Por ejemplo, si para el municipio de San Miguel se presenta el valor 0.1394, esto significa que el 13.94% de la población de ese municipio ha experimentado la emigración de al menos uno de sus miembros. Por el aporte negativo que este índice ofrece, entonces consideramos $\frac{\partial mi_j}{\partial MI_{1j}} < 0$

9. Indicador de Densidad de Coches

Definición: Entenderemos por *indicadores de Tráfico* aquellos que por su naturaleza, en cuanto a la recolección de los datos y a su forma de cálculo, se clasifican en ésta área temática. Tenemos:

$$Itr_j = \sum_{j=1}^n W_i TR_{ij}$$

Para $i = 1$, entonces $Itr_j = \sum_{j=1}^n W_1 TR_{1j}$, sería el *indicador de Densidad de Coches por municipio*. Entendiendo éste como el porcentaje de la población que posee vehículo.

Fuente y forma de cálculo: los datos han sido recopilados del IV Censo de Población y V de Vivienda 2007. Y es el cociente:

$$\frac{\text{Población que posee vehículo por municipio}}{\text{Población Total del Municipio}}$$

Interpretación: Por ejemplo, si para el municipio de San Salvador se presenta el valor 0.4381, esto significa que el 43.81% de la población de ese municipio posee vehículo, obviamente esto representa una contribución negativa al medio por las emisiones de gas, es así que consideramos que el aporte de este índice es negativo: $\frac{\partial Itr_j}{\partial TR_{1j}} < 0$

10. Indicador de Hacinamiento.

Entenderemos por *indicadores de Condiciones de Vida* aquellos que por su naturaleza, en cuanto a la recolección de los datos y a su forma de cálculo, se clasifican en ésta área temática. Veamos:

$$Icv_j = \sum_{j=1}^n W_i CV_{ij}$$

Para $i = 1$, entonces $Icv_j = \sum_{j=1}^n W_1 CV_{1j}$, sería el *indicador de Hacinamiento*. Entendiendo éste como el número de viviendas con más de tres hogares.

Fuente y forma de cálculo: los datos han sido obtenidos del Viceministerio de Vivienda y Desarrollo Humano. Y es el cociente:

$$\frac{\text{Número de viviendas con más de tres hogares}}{\text{Número de hogares por Municipio}}$$

Interpretación: Por ejemplo, si para el municipio de Soyapango tenemos un valor 0.3094, esto significa que el 30.94% de las viviendas de este municipio tienen más de tres hogares. El aporte de este indicador es negativo por lo que: $\frac{\partial Icv_j}{\partial CV_{1j}} < 0$

11. Indicador de Densidad Habitacional.

Definición: Para $i = 2$, entonces $Icv_j = \sum_{j=1}^n W_2 CV_{2j}$, representa el *indicador de Densidad Habitacional*. Entendiendo éste como el número de habitantes por vivienda.

Fuente y forma de cálculo: los datos han sido recopilados del Segundo Censo de Desechos Sólidos. Y es el cociente:

$$\frac{\text{Número de habitantes por municipio}}{\text{Número de viviendas por Municipio}}$$

Interpretación: Si para el municipio de Acajútla tenemos un valor 3.79, esto significa que en ese municipio viven aproximadamente 4 personas por vivienda. El aporte de este indicador es negativo por lo que: $\frac{\partial Icv_j}{\partial CV_{2j}} < 0$

12. Indicadores de Uso de Inodoro.

Definición: Entenderemos por *indicadores de Agua*, aquellos que por su naturaleza, en cuanto a la recolección de los datos y a su forma de cálculo, se clasifican en ésta área temática. Veamos:

$$Iag_j = \sum_{j=1}^n W_i AG_{ij}$$

Para $i = 1$, entonces $Iag_j = \sum_{j=1}^n W_1 AG_{1j}$, representa el *indicador de uso de Inodoro*. Entendiendo éste como el número de viviendas que poseen inodoro.

Fuente y forma de cálculo: los datos han sido recopilados del IV Censo de Población y V de vivienda. Y es el cociente:

$$\frac{\text{Número de viviendas con Inodoro}}{\text{Número de viviendas por Municipio}}$$

Interpretación: Si para el municipio de Arambala tenemos un valor 0.7858, esto nos indica que en ese municipio el 78.58% de las viviendas poseen inodoro. El aporte de este indicador es positivo por lo que: $\frac{\partial Iag_j}{\partial AG_{1j}} > 0$

13. Indicador de Cobertura de Aguas Grises

Definición: Entendiendo éste como el porcentaje de viviendas que poseen acceso a la red de aguas grises. Entonces $I_{ag_j} = \sum_{j=1}^n W_2 AG_{2j}$, representa el *indicador de Cobertura de Aguas Grises*.

Fuente y forma de cálculo: los datos han sido recopilados del IV Censo de Población y V de vivienda. Y es el cociente:

$$\frac{\text{Número de viviendas con acceso a red de alcantarillado para aguas grises}}{\text{Número de viviendas por Municipio}}$$

Interpretación: Si para el municipio de Ayutuxtepeque tenemos un valor 0.6580, esto nos indica que en ese municipio el 65.80% de las viviendas poseen conexión a tuberías para aguas grises. El aporte de este indicador es positivo por lo que: $\frac{\partial I_{ag_j}}{\partial AG_{2j}} > 0$

14. Indicador de Cobertura de Agua Potable.

Definición: Entenderemos éste como el porcentaje de viviendas que poseen acceso a agua potable. Entonces $I_{ag_j} = \sum_{j=1}^n W_3 AG_{3j}$, representa el *indicador de Cobertura de Agua Potable*.

Fuente y forma de cálculo: los datos han sido recopilados del IV Censo de Población y V de vivienda. Y es el cociente:

$$\frac{\text{Número de viviendas con acceso a tubería para agua potable}}{\text{Número de viviendas por Municipio}}$$

Interpretación: Si para el municipio de Apopa tenemos un valor 0.8996, esto nos indica que en ese municipio el 89.96% de las viviendas poseen conexión a tuberías para agua potable. El aporte de este indicador es positivo por lo que: $\frac{\partial I_{ag_j}}{\partial AG_{3j}} > 0$

15. Indicador de Porcentaje de Servicio de Agua.

Definición: Entendiendo éste como la proporción del número de días a la semana que recibe agua potable. Entonces $Iag_j = \sum_{j=1}^n W_4 AG_{4j}$, representa el *indicador de porcentaje de servicio de agua*.

Fuente y forma de cálculo: los datos han sido recopilados del IV Censo de Población y V de vivienda. Y es el cociente:

$$\frac{\text{Número viviendas que reciben agua potable todos los días}}{\text{Total de viviendas que reciben agua potable}}$$

Interpretación: Si para el municipio de Chalatenango tenemos un valor 0.7980, esto nos indica que en ese municipio el 79.80% de las viviendas reciben servicio de agua potable todos los días. El aporte de este indicador es positivo por lo que: $\frac{\partial Iag_j}{\partial AG_{4j}} > 0$

16. Indicadores de Mortalidad

Definición: Entenderemos por *indicadores de Mortalidad* aquellos que por su naturaleza, en cuanto a la recolección de los datos y a su forma de cálculo, se clasifican en ésta área temática. Veamos:

$$Imo_j = \sum_{j=1}^n W_1 MO_{1j}$$

Entonces $Imo_j = \sum_{j=1}^n W_1 MO_{1j}$, representa el *indicador de Mortalidad*. Interpretaremos como el número de viviendas donde se ha tenido más de un fallecido en el último año (fecha en la que se realizó el censo, 2007).

Fuente y forma de cálculo: los datos han sido recopilados del IV Censo de Población y V de vivienda. Siendo:

$$\frac{\text{Número viviendas que perdió al menos un familiar}}{\text{Total de viviendas}}$$

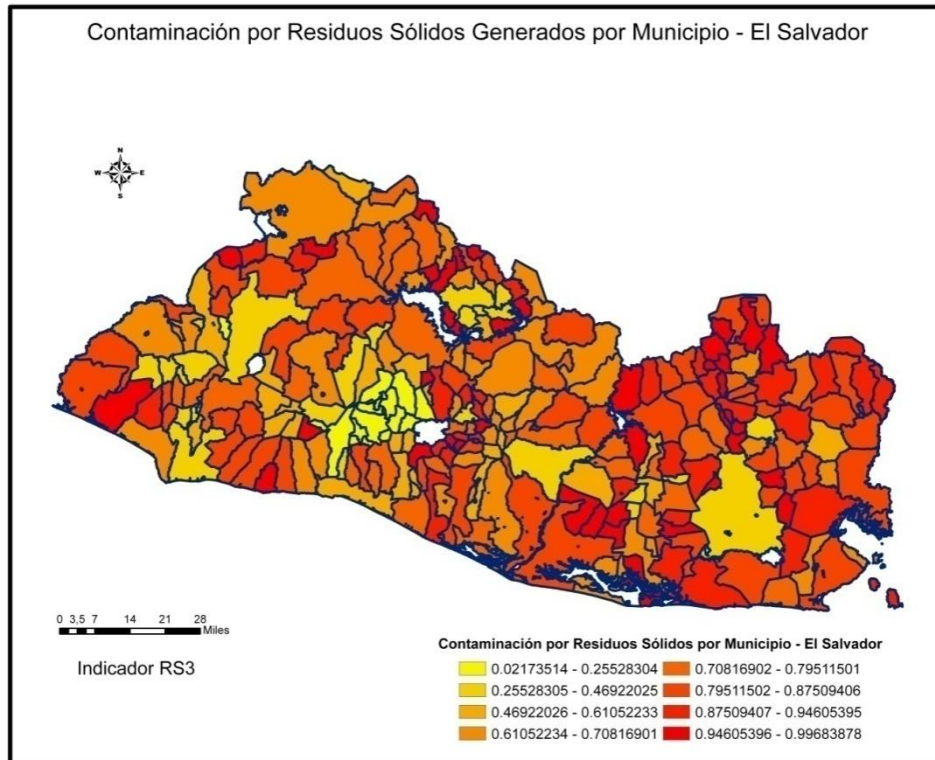
Interpretación: Si para el municipio de Ahuachapán tenemos un valor 0.02207, esto nos indica que en ese municipio el 2.2% de los hogares ha perdido un familiar. El aporte de este indicador lo consideramos negativo por lo que: $\frac{\partial Imo_j}{\partial MO_{1j}} < 0$

TABLA A.5 - DEMOGRÁFIA 1

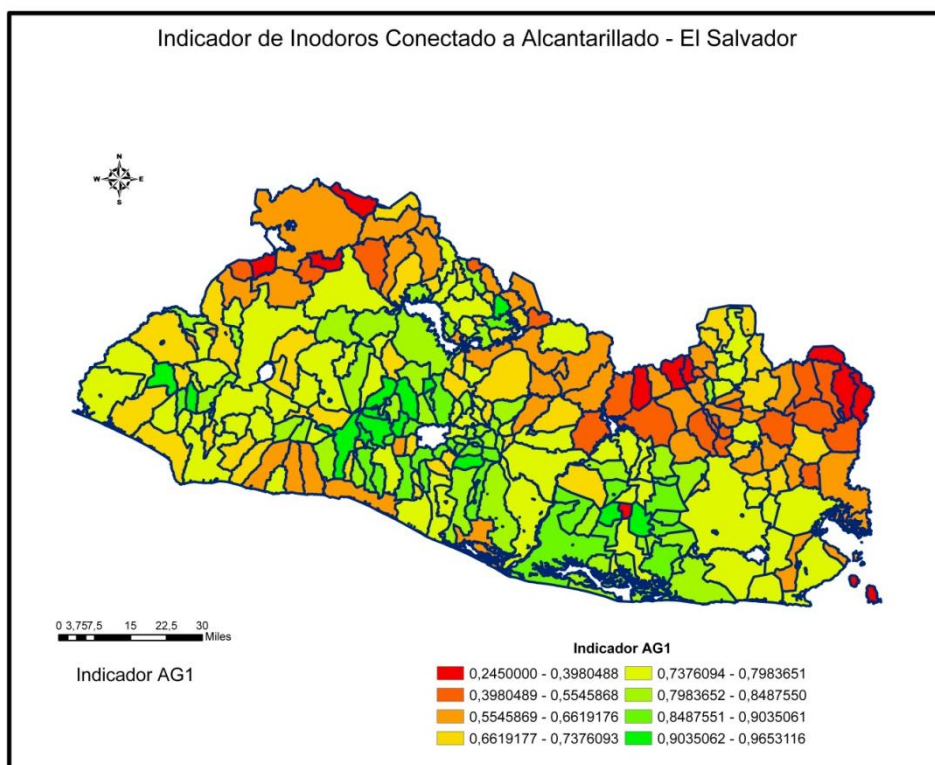
PROYECCIÓN DEMOGRÁFICA DE EL SALVADOR			
Departamento	2000	2007	2010
Ahuachapán	319,781	319,503	392,446
Cabañas	152,842	149,326	160,850
Chalatenango	196,587	192,788	206,890
Cuscatlán	202,951	231,480	222,290
La Libertad	682,093	660,652	880,107
La Paz	292,887	308,087	344,821
La Unión	289,022	238,217	316,715
Morazán	173,501	174,406	184,757
San Miguel	480,276	434,003	599,173
San Salvador	1,985,294	1,567,156	2,357,761
San Vicente	161,104	161,645	180,793
Santa Ana	550,209	523,655	667,392
Sonsonate	450,118	438,960	568,725
Usulután	338,334	344,235	357,942
Total	6,274,999	5,744,113	7,440,662

TABLA A.6 - DEMOGRÁFIA 2

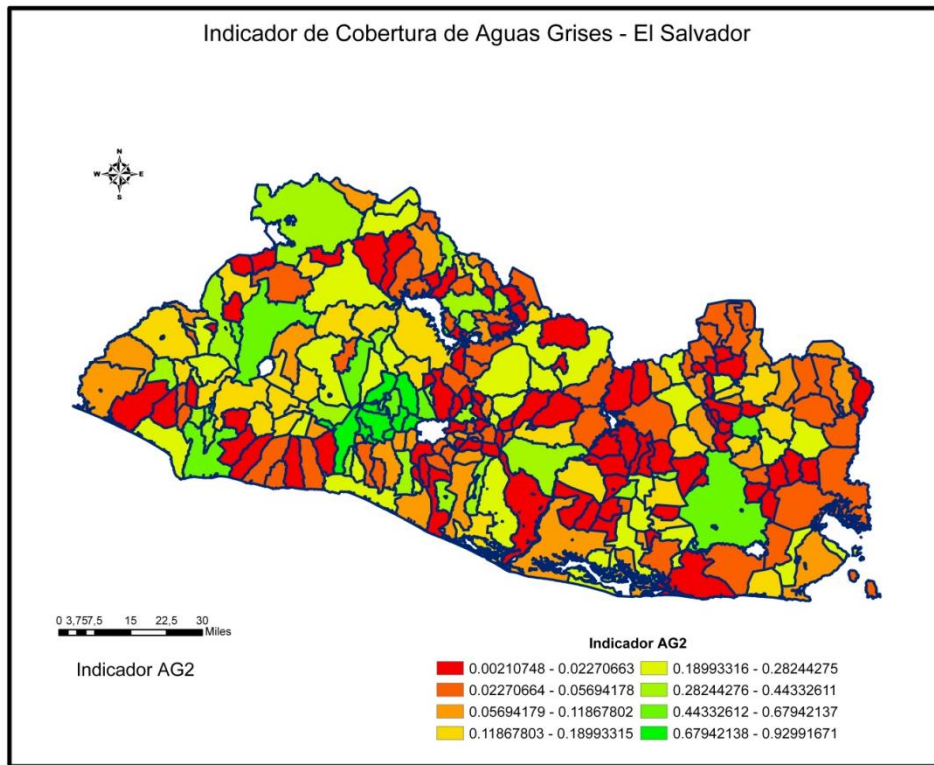
DIVISIÓN POLÍTICO ADMINISTRATIVA			
Departamentos	Superficie	Población	Densidad
Ahuachapán	1,239.60	319,503	258
Cabañas	1,103.51	149,326	135
Chalatenango	2,016.58	192,788	96
Cuscatlán	756.19	231,480	306
La Libertad	1,652.88	660,652	400
La Paz	1,223.61	308,087	252
La Unión	2,074.34	238,217	115
Morazán	1,447.43	174,406	120
San Miguel	2,077.10	434,003	209
San Salvador	886.15	1,567,156	1,768
San Vicente	1,184.02	161,645	137
Santa Ana	2,023.17	523,655	259
Sonsonate	1,225.77	438,960	358
Usulután	2,130.44	344,235	162
Total	21,040.79	5,744,113	273



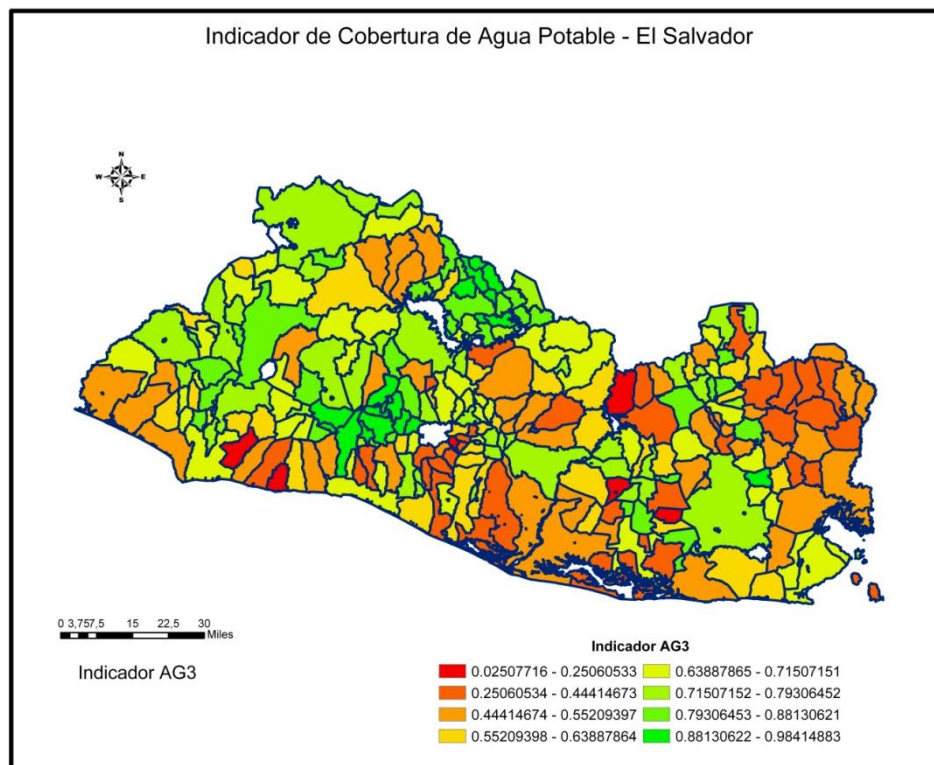
MAPA 2 – AG1



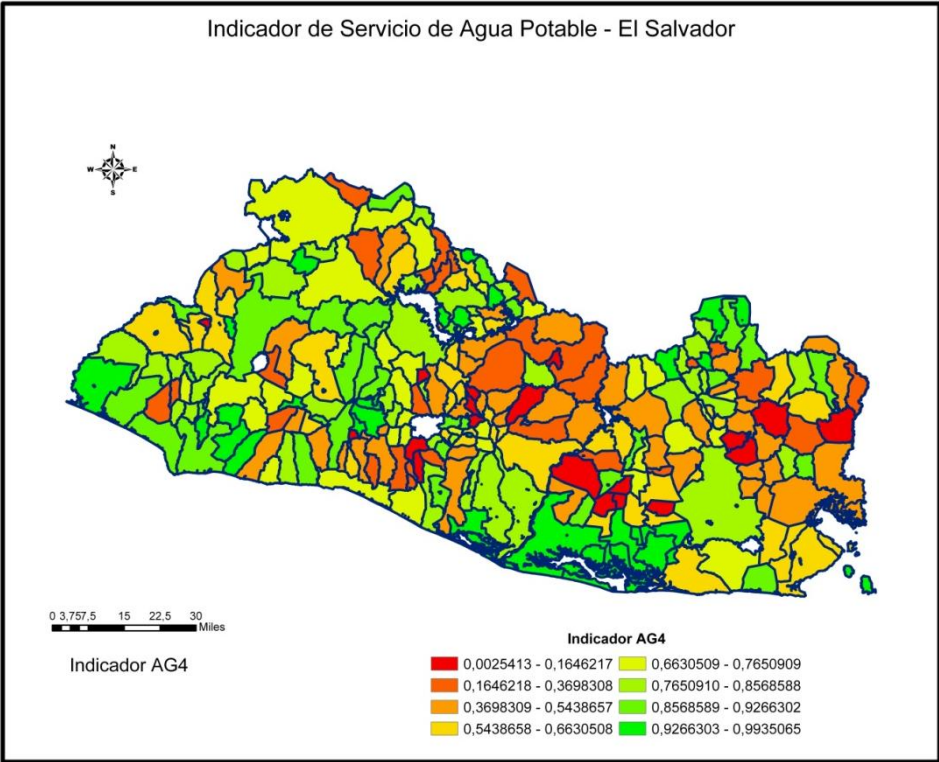
MAPA 3 – AG2



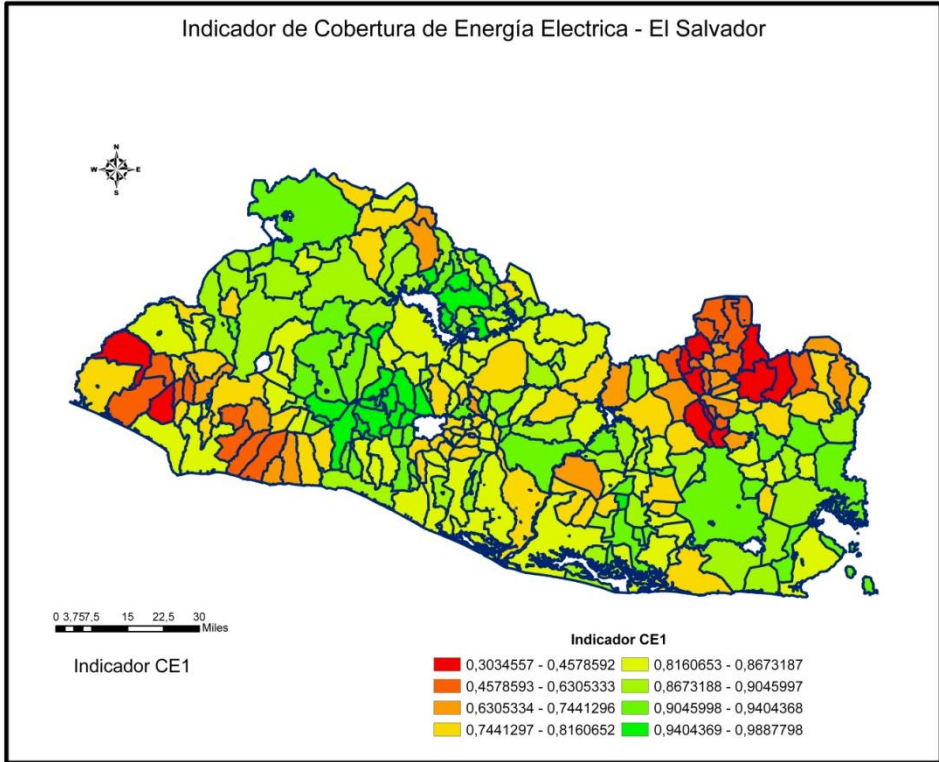
MAPA 4 – AG3



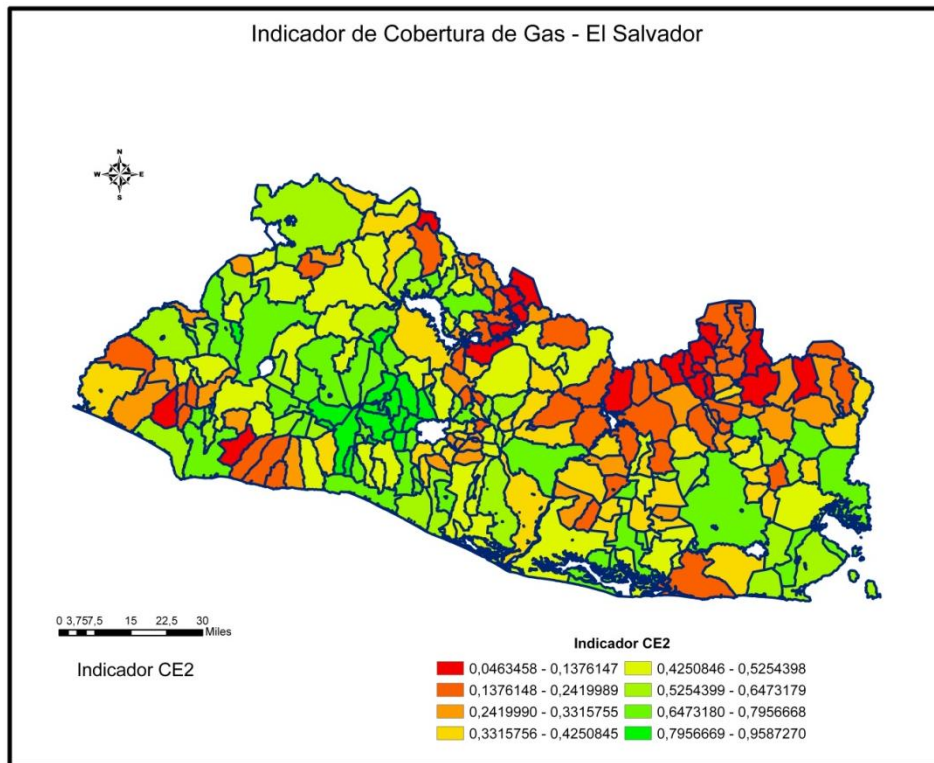
MAPA 5 – AG4



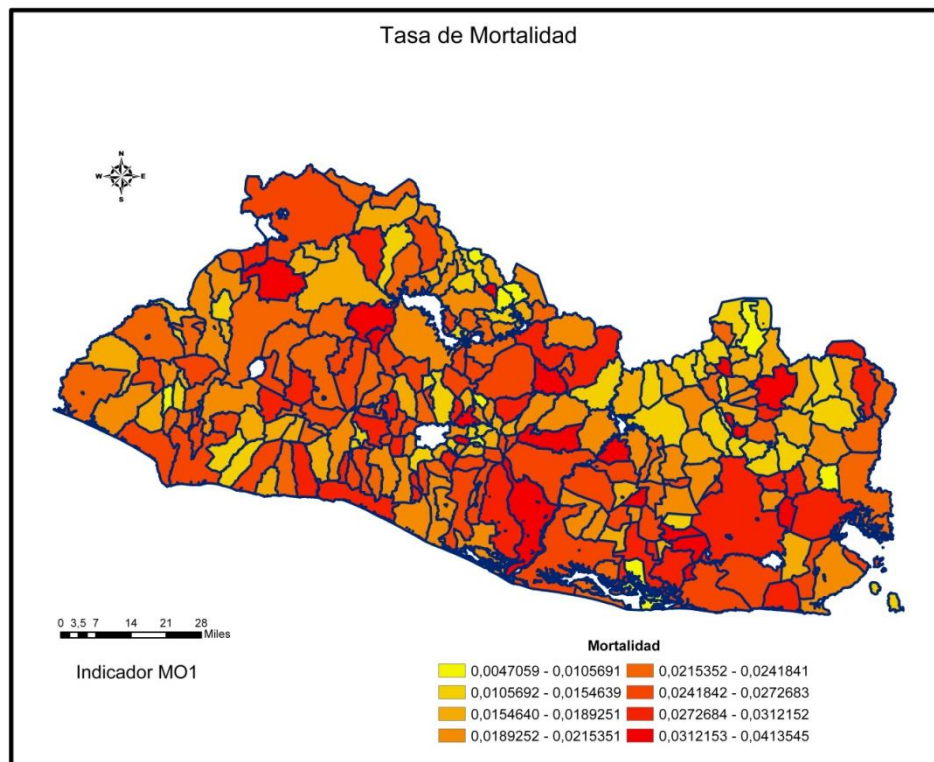
MAPA 6 – CE1



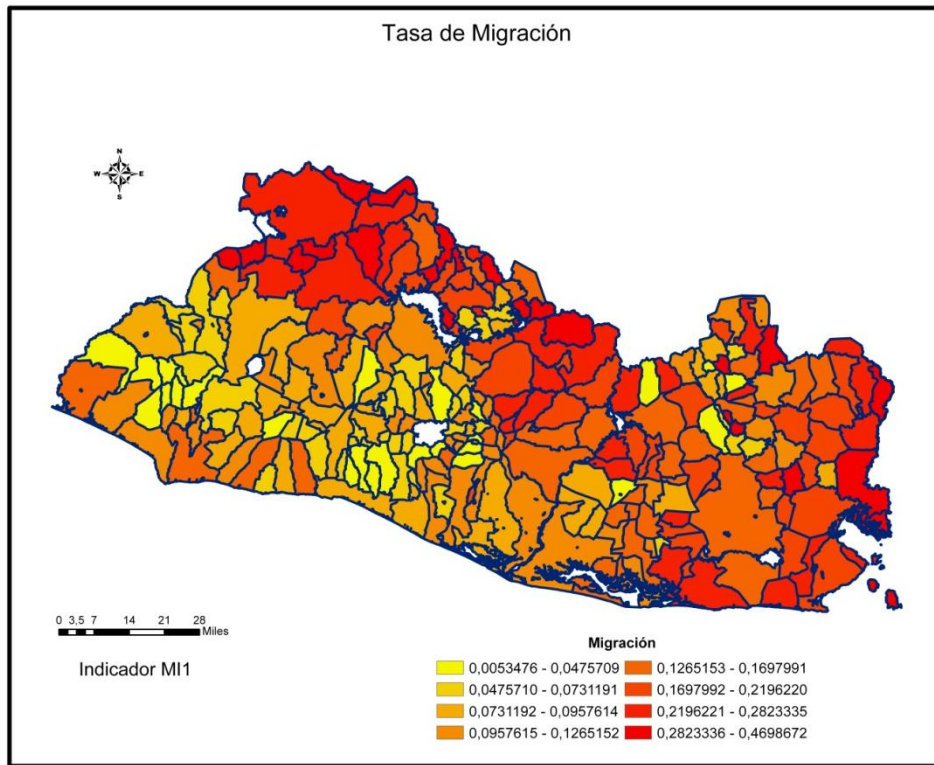
MAPA 7 – CE2



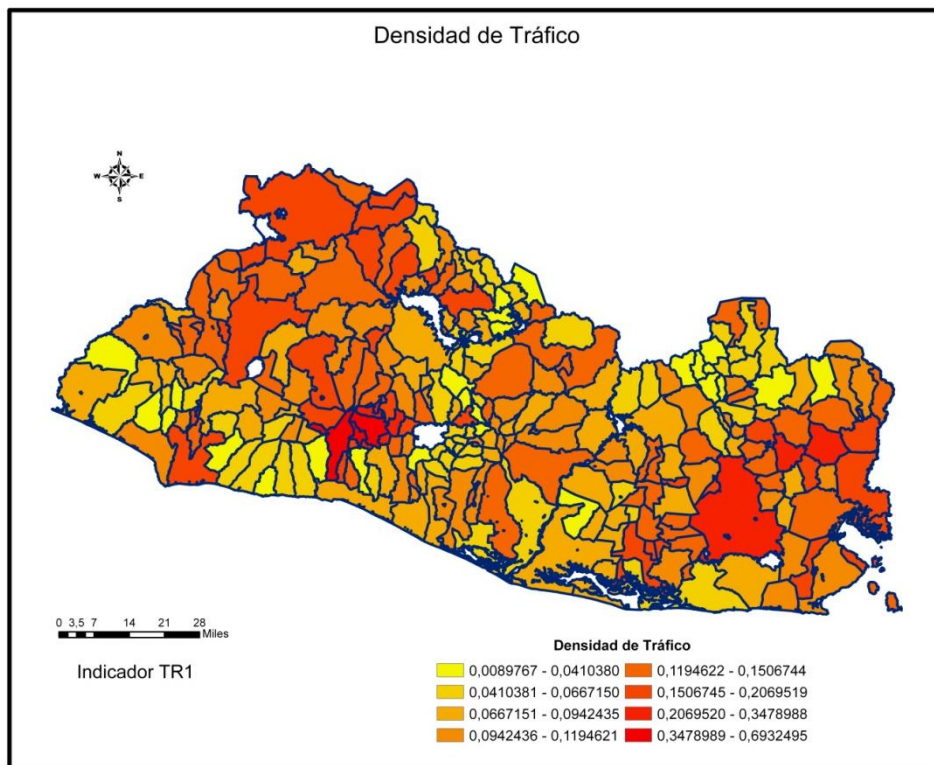
MAPA 8 – MO1



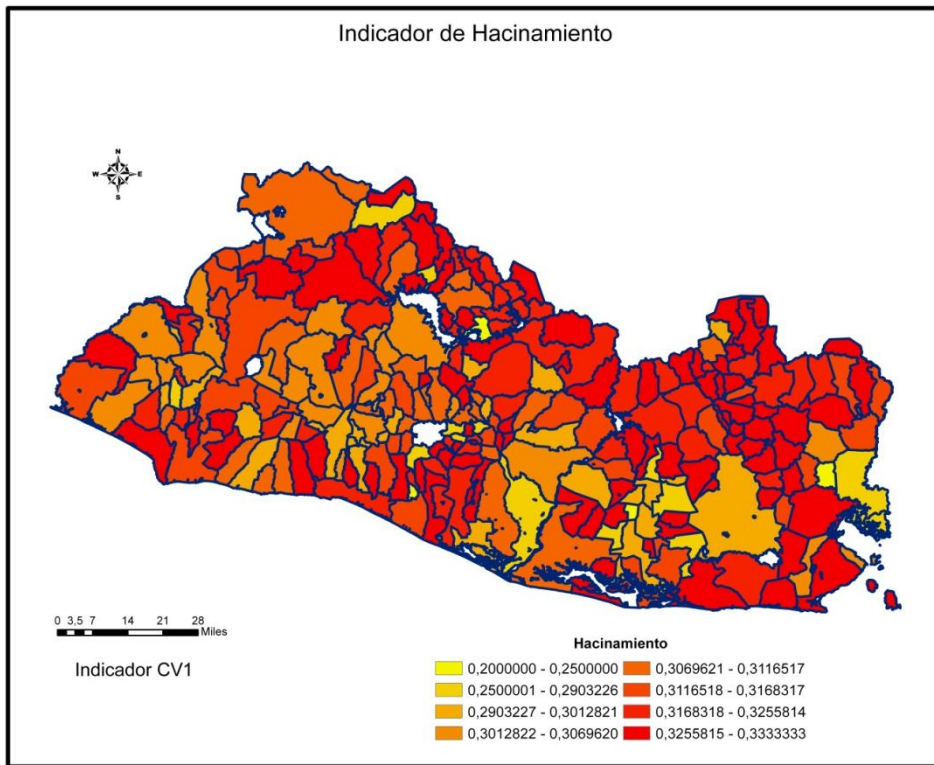
MAPA 9 – MII



MAPA 10 – TR1



MAPA 11 – CV1



MAPA 12 – CV2

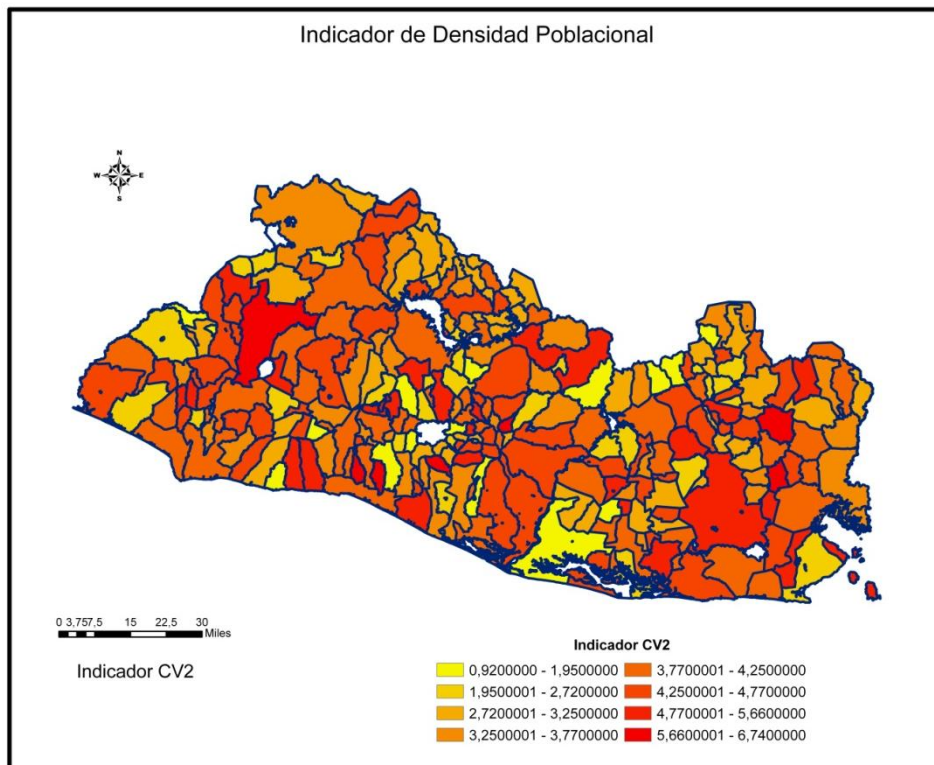


Tabla A.7 – Importancia relativa de cada variable en relación al ISCA, Santa Ana

MUNICIPIO	ISCA	SANTA ANA											
		SANTA ANA	SEBASTIAN CALERILLA	METAPAN	CHALCHUA PA	COATEPEQUE	DE LA FRONTERA	EL CONGO	EL PORVENIR	IA DE LA FRONTERA	MASAHUAT	ANTONIO ROSA	GUACHIPILI
CE3	-0.613	-1.55	-1.7	-0.45	-0.91	-0.05	0.57	-0.61	-0.17	-0.52	1.07	-0.34	0.7
CE2	0.61	1.58	1.75	0.44	0.94	0.06	-0.59	0.64	0.21	0.55	-1.13	0.3	-0.77
MO1	0.582	0.33	-0.92	0.67	-0.08	0.1	0.26	0.61	-1.27	-0.19	-0.96	1.1	-0.61
TR1	0.55	1.31	0.82	0.89	0.35	-0.47	0.56	0.2	-0.16	0.55	-0.21	0.7	0
AG2	0.517	1.88	1.79	0.8	0.8	-0.28	-0.68	0.2	-0.67	0.14	0.19	-0.73	-0.67
RS2	0.503	1.59	2.17	0.35	0.75	-0.55	-1.12	0.84	-0.56	-0.35	-0.93	-0.9	-1.17
CE1	0.485	0.64	0.81	0.71	0.54	0.06	0.65	0.21	-0.17	0.59	0.08	0.54	0.63
RS4	0.455	1.49	0.12	1.21	0.57	0.12	-0.74	-0.34	-1.48	-0.48	-3.2	-0.74	-0.74
AG3	0.347	1.01	1.27	0.49	0.85	-0.55	-0.14	0.35	0.52	0.06	0.74	0.42	0.95
MI1	0.308	-0.53	-0.56	1.11	-0.8	-0.56	1.93	-0.46	-0.87	0.04	2.83	1.82	3.02
AG4	0.296	0.93	1.14	0.15	-0.04	-0.43	0.67	-1.33	0	-0.8	1.27	0.21	0.84
CV2	0.27	2.45	0.86	-0.43	0.63	0.21	-1.05	1.21	0.86	1.06	0.17	-1.05	-1.05
CV1	0.199	-0.22	0.3	-0.34	-0.78	-0.39	-0.02	-0.55	-0.02	-0.11	0.98	-0.02	-0.02
RS1	-0.14	0.5	-0.31	-0.43	-0.3	-0.31	-0.21	-0.05	1.04	0	3.32	-0.5	-0.22
AG1	-0.068	0.24	0.54	-0.84	-0.06	-0.01	-1.63	0.02	0.38	-0.67	-2.24	-2.44	-3.03

Tabla A.8 – Importancia relativa de cada variable en relación al ISCA, Sonsonate

MUNICIPIO	ISCA	SONSONATE															
		ACAUTLA	ARMENIA	CALUCO	CUISNAHUAT	ISABEL	IZALCO	JUAYUA	NAHUZALCO	NAHULINGO	SALCOATITAN	ANTONIO BEL MONTE	SAN JULIAN	CATARINA	MASAHUAT	DOMINGO DE GUZMAN	SONSONATE
RS1	-0.14	0.7	-0.45	1.43	-0.66	-0.73	0.1	-0.6	-0.81	0.35	1.83	-0.82	-0.8	-0.93	-0.26	0.03	-0.18
RS2	0.5	0.13	0.59	-0.53	-0.56	-0.65	-0.1	1.35	-0.15	0.5	1.39	1.24	-0.45	-0.13	-0.4	1.29	2.36
RS4	0.46	0.82	1.31	1.41	0.39	0.04	1.69	0.04	1.98	0.79	-1.77	1.7	0.04	0.04	1.41	1.53	1.41
AG1	-0.07	-0.37	0.27	0.33	-0.38	-0.11	0.36	0.24	0.76	-0.21	0.41	0.4	0.34	1.32	0.33	0.46	1.11
AG2	0.52	0.47	0.19	-0.7	-0.74	-0.66	0.02	0.6	-0.03	1.36	0.25	1.49	0	-0.74	-0.47	1.79	2.69
AG3	0.35	-0.68	0.21	-0.31	-3.41	-0.72	0.44	0.88	0.91	0.44	0.02	1.13	-0.36	-0.08	0.34	0.22	1.68
AG4	0.3	1.02	0.2	1.21	1.37	1.37	0.15	0.84	0.67	1.32	1.09	0.7	0.2	0.24	0.82	1.11	0.7
CE1	0.49	0.06	0.34	-1.81	-2.4	-2.06	-0.32	-0.05	-0.88	-0.09	-0.58	0.15	-0.72	-1.9	-0.44	0.23	1.03
CE2	0.61	0.63	1.18	-0.76	-1.52	-1.18	0.3	0.29	-0.58	0.91	0.35	1.32	0.31	-1.07	-1	1.28	1.92
CE3	-0.61	-0.64	-1.14	0.81	1.52	1.19	-0.26	-0.26	0.6	-0.85	-0.28	-1.26	-0.31	1.11	0.99	-1.24	-1.95
MO1	0.58	0.51	1.07	0.42	-1.34	-1.51	-0.28	0.79	0.17	0.69	-0.61	0.11	-0.5	-1.24	0.2	0.87	0.45
MI1	0.31	-0.44	-0.68	-1.03	-0.69	-0.41	-0.89	-1.02	-1.32	-0.29	-1.25	-0.59	-0.54	-1.3	-0.45	-0.13	-0.57
TR1	0.55	-0.02	-0.05	-0.66	-1.14	-0.9	-0.31	-0.04	-0.84	0.24	-0.25	0.06	-0.36	-1.25	-0.99	0.96	1.66
CV1	0.2	0.98	-0.22	0.98	0.98	-0.49	-0.05	-0.73	-1.04	0.98	0.11	0.36	-1.05	-1.49	-0.49	-0.16	0.53
CV2	0.27	0.04	-0.98	0.95	-0.09	0.56	0.32	0.59	0.63	-0.25	0.65	-1.19	0.72	1.23	0.98	0.21	-0.71

Tabla A.9 – Importancia relativa de cada variable en relación al ISCA, Chalatenango

MUNICIPIO	ISCA	CHALATENANGO																																			
		AGUA CALIENTE	ARCATAO	AZACUALPA	CITALA	COMALAPA	QUEZALTEPE	CHALATENANGO	NOMBRE DE	EL CARRIZAL	EL PARAISO	LA LAGUNA	LA PALMA	LA REINA	LAS VUELTAS	NOMBRE DE	JESUS NUEVA	CONCEPCION NUEVA	TRINIDAD	OJOS DE	AGUA	POTONICO	ANTONIO DE	ANTONIO	LOS SAN	FERNANDO	FRANCISCO	FRANCISCO	SAN	IGNACIO	SAN ISIDRO	LABRADOR	CANCASQUE	LAS FLORES	DEL	DE	SAN RAFAEL
RS1	-0.1	-0.3	-0.5	0.2	-0.1	-0.2	-0.3	-0.6	-0.2	-0.2	-0.2	-0.2	0.3	-0.3	-0.5	0.1	-0.2	-0.5	-0.1	-0.6	-0.5	-0.6	-0.2	-1.1	-0.6	-0.1	-0.5	-0.5	-0.8	-0.2	-0.4	-0.5	-0.5	-0.7			
RS2	0.5	-0.2	0.2	-1.2	0.7	-1.2	0.4	1.7	0.4	-1.1	0.1	-0.7	0.3	-0.1	0.2	0.4	-0.1	-1	-0.6	0.8	-1.1	1.5	-1.1	1.3	-0.3	-0.1	1.8	-1.1	1.4	-1.2	1.3	-1.1	-1.2	-0.2			
RS4	0.5	0.2	-1.3	-1.3	1	-0.1	0.2	0.2	-0.1	-0.1	-0	-0.1	-0.1	0.2	-0.1	-0.7	1.4	-0.1	-0.3	0.2	-0.1	0.2	-0.1	0.2	0.2	1	-0.1	-1.3	0.2	-1.3	0.2	-1.3	-1.3	0.2			
AG1	-0.1	-2.1	-0.5	0.7	-2.7	-0.3	0.2	0.3	0.2	-1.4	0.8	0.7	-1	-0.9	-0.2	-1.4	0.2	-0.6	-1.1	0.7	-0.1	-0.2	-1	0.3	-0.6	-0.3	1	-0.1	1.4	-0.6	0.4	0	0.4	-0.2			
AG2	0.5	-0.7	-0.6	-0.6	-0.3	0.1	-0.6	1.1	0.8	-0.3	-0.6	0.7	0.5	-0.7	-0.7	-0.1	0.5	-0.7	-0.6	-0.5	-0.7	-0.5	-0.6	1.9	-0.3	0.3	-0.6	-0.7	0.8	-0.7	1.1	-0.7	-0.7	-0.5			
AG3	0.3	-0.9	0.7	0.9	0.6	1.2	1.7	1.2	1.3	1.6	1.1	1.4	0.4	-0.9	1.4	1.3	-0.1	1	1.7	1.1	0.9	1.5	-0.4	1.9	-0.6	0.6	1.8	0.6	1.3	0.3	0.8	0.6	-0.4	-1			
AG4	0.3	-1.7	-1.3	1.2	-1.4	-1.5	0	0.6	-1.5	1.2	0.7	-0	0.5	-0.7	0.3	0.2	0.4	0.8	1	0.4	1.1	0.7	0.8	-0.6	0.4	0.9	-0.7	-0.2	1.3	1	1.2	-1.7	-1.3	-0			
CE1	0.5	-0.3	0.1	1	-0.2	0.7	1	1.1	0.4	0.8	0.6	0.3	-0.3	0.5	0.6	0.6	0.6	-0.1	0.8	1	0.6	0.4	-1.3	0.9	-1.2	0.1	0.6	0.3	0.5	0.7	0.7	1	0.9	0.6			
CE2	0.6	0	-1.4	-0.1	-0.2	-0.4	0.3	1.2	0.3	-1	0.9	-0.6	-0.1	-0.2	-1.3	-0.8	0.3	-1.7	-0.8	-0.9	-1.6	-1.2	-1.4	0.9	-1.1	0.1	-1.3	-1.6	-1	-0.7	0.1	1	0.4	0.7			
CE3	-0.6	-0	1.3	-0.1	0.1	0.4	-0.3	-1.2	-0.3	1	-0.8	0.6	0.1	0.2	1.2	0.7	-0.3	1.7	0.8	0.9	1.6	1.2	1.4	-0.9	1	-0.1	1.3	1.6	1	0.5	-0.1	-0.9	-0.5	-0.7			
MO1	0.6	1	-0.1	0.5	0.3	-1	-1.1	-0	-0.7	-2.1	0.1	-1.6	-0.6	-1.1	2.1	0.2	-0.7	-2.2	-0.8	0.3	-0.9	-0.1	-0.2	0	0.6	0.1	-1.3	-0.2	-2.2	-1.7	0.1	-0.8	-0.3	0			
MI1	0.3	2.1	0.2	1.8	1.7	1.9	0.9	0.7	2.2	1	0.7	0.2	1.5	1.1	0.5	3.2	1.4	0.2	2.5	-1.1	1.8	1.2	0.6	0.4	0.1	1.9	-0.9	-0.6	-0.6	0.5	-1	1.9	0.9	0.4			
TR1	0.6	1.4	-1.1	-0.3	0.4	-0.4	0	0.9	0	-0.5	0.1	-0.7	0.8	0.3	-0.8	0.3	0.6	-0.8	-0.9	-0.5	-0.5	-0.9	-0.7	0.4	-0.8	0.8	-1.1	-1.2	-1.1	-0.9	-0.2	0.8	0.4	0.7			
CV1	0.2	1	1	1	-0.5	0.1	-0.3	-0.4	1	0.1	1	1	-2.2	0.1	0.1	0.1	1	0.1	0.1	-3.8	0.1	0.1	1	0.1	1	1	0.1	0.1	0.1	1	-1.8	1	-0.4				
CV2	0.3	0.6	-0.7	-0.7	-0.5	-0.7	-0.3	0.7	-0.7	-0.6	0.5	-0.7	0.8	-0.3	-0.6	0.4	0.2	-0.6	0.3	0.2	-0.7	-0.6	-0.9	0.5	-0.7	0.9	-0.7	-0.8	0.6	-1	-0.1	-0.6	0.2	-0.8			

Tabla A.10 – Importancia relativa de cada variable en relación al ISCA, La Libertad

		LA LIBERTAD																					
MUNICIPIO	ISCA	ANTIGUO CUSCATLAN	CIUDAD ARCE	COLON	COMASAGUA	CHILTIUPAN	HUIZUCAR	JAYAQUE	JICALAPA	LA LIBERTAD	NUEVO CUSCATLAN	SANTA TECLA	QUEZALTEPEQUE	SACACOYO	SAN JOSE VILLANUEVA	SAN JUAN OPICO	SAN MATIAS	SAN PABLO TACACHICO	TAMANIQUE	TALNIQUE	TEOTEPEQUE	TEPECOYO	ZARAGOZA
RS1	-0.14	2.47	-0.54	-0.6	-0.65	-0.47	-1.04	-0.43	-0.46	-1.09	-0.27	2.09	-0.2	-0.46	0.74	-1.07	-0.37	-0.54	-0.53	-0.5	-0.57	2.06	-0.76
RS2	0.503	2.98	-0.06	1.56	0.43	-0.49	-0.35	0.48	-1.14	0.62	2.33	2.71	1.21	0.75	-0.05	0.28	-0.58	-0.42	0.41	-1.1	-0.64	0.64	1.62
RS4	0.455	1.29	1.96	1.29	1.29	0.6	0.04	1.24	0.6	0.04	0.04	-0.54	1.95	1.42	-1.12	0.04	0.04	0.04	1.29	0.6	1.29	-2.39	1.29
AG1	-0.068	1.58	0.14	-0.25	0.58	-0.74	1.07	0.51	0.41	-0.66	-0.34	1.72	0.63	-0.35	0.66	0.21	0.49	0.57	-0.87	0.38	-0.49	0.54	-0.05
AG2	0.517	3.86	0.19	1.31	-0.69	-0.66	-0.58	0.28	-0.6	0.37	-0.31	3.74	1.66	-0.11	0.26	0.4	-0.63	0.05	-0.6	-0.31	-0.61	-0.12	1.01
AG3	0.347	1.8	0.91	1.39	-0.64	-0.29	-1.12	0.37	-2.8	0.24	0.84	1.57	0.73	0.68	-1.18	0.5	0.38	0.14	-1.05	1.01	-1.5	0.33	0.31
AG4	0.296	1.28	0.19	0.8	-0.82	0.62	-0.64	-0.57	0.17	0.43	-2.07	1.08	1.07	1.16	-0.14	0.07	0.26	0.95	1.01	0.04	-0.47	-1.62	-0.81
CE1	0.485	1.3	0.71	1.12	-0.34	-1.17	0.14	0.24	-0.85	0.58	0.89	1.17	0.91	0.61	0.69	0.72	0.64	0.7	-0.47	0.41	-1.94	-0.33	0.89
CE2	0.61	1.26	1.44	2.08	-0.38	-0.8	-0.05	0.93	-1.31	1.05	1.5	1.78	1.43	1.58	1.3	1.17	0.01	0.22	-0.02	0.73	-1.02	0.56	1.71
CE3	-0.613	-2.22	-1.39	-2.01	0.39	0.81	0.06	-0.9	1.3	-1.01	-1.54	-2.07	-1.38	-1.52	-1.32	-1.14	0.01	-0.22	0.03	-0.72	1.04	-0.55	-1.75
MO1	0.582	-1.11	0.91	0.64	-0.6	0.13	0.45	-0.46	-0.8	0.97	-1.4	-0.04	0.87	0.17	0.47	0.35	0.57	0.29	0.91	0.08	0.71	-0.77	1.32
MI1	0.308	-0.79	-0.65	-0.76	-0.99	-0.59	-1.18	-0.98	-0.96	-0.59	-0.99	-0.75	-0.66	-0.42	-1.11	-0.47	0.27	0.88	-0.1	-1.01	0.04	-1.32	-0.84
TR1	0.55	8.34	-0.03	0.97	-1	-0.81	-0.94	-0.15	-1.1	-0.42	0.7	5.83	0.44	-0.34	0.21	0.92	0.07	0.09	-0.71	-0.47	-0.76	-0.74	1.1
CV1	0.199	-1	-0.68	-0.58	0.98	-0.26	-1.15	-0.65	-0.26	0.11	-0.93	-0.91	-0.37	-0.11	0.98	-0.66	0.98	-0.76	0.98	0.11	-1.27	0.3	-0.64
CV2	0.27	0.02	0.34	-0.24	-0.14	1.33	0.23	0.87	-1.95	0.38	0.61	0.24	-0.25	0.77	2.66	0.56	0.64	0.21	1.54	-1.72	-0.61	-0.67	0.44

Tabla A.11 – Importancia relativa de cada variable en relación al ISCA, San Salvador

SAN SALVADOR																				
MUNICIPIO	ISCA	AGUILARES	APOPA	AYUTUXTEPEQ UE	CUSCATANCING O	EL PAISNAL	GUAZAPA	ILOPANGO	MEJICANOS	NEJAPA	PANCHIMALCO	ROSARIO DE MORA	SAN MARCOS	SAN MARTIN	SAN SALVADOR	SANTIAGO TEXACUANGOS	SANTO TOMAS	SOYAPANGO	TONACATEPEQ UE	CIUDAD DELGADO
RS1	-0.14	0.99	-0.45	-0.77	-0.25	1.93	-0.22	2.99	0.64	0.12	1.03	-0.21	0.3	4.94	3.76	0.68	3.62	-0.08	-0.31	-0.22
RS2	0.503	1.69	2.39	2.39	2.96	-0.52	0.25	2.79	2.94	0.77	-0.05	-0.27	2.87	2.01	3.02	0.79	0.83	2.96	2.35	2.06
RS4	0.455	-2.12	0.01	0.24	0.24	-2.32	0.24	-1.06	-0.19	0.24	0.24	0.24	0.74	-0.65	-0.24	-1.01	-3.35	0.93	0.24	1.58
AG1	-0.068	0.19	1.46	0.66	1.11	0.68	-0.19	1.1	1.29	0.39	0.08	1.12	0.97	0.68	1.56	-0.19	-0.58	1.68	1.66	0.72
AG2	0.517	1.15	3.41	2.58	3.41	-0.13	0.43	3.36	3.57	0.57	-0.22	-0.57	2.85	1.72	3.95	-0.26	-0.36	3.96	3.28	2.21
AG3	0.347	0.57	1.44	1.65	1.63	0.14	1.31	1.71	1.79	-0.76	-0.61	0.02	1.34	0.72	1.91	0.18	-0.42	1.86	1.1	1.39
AG4	0.296	1.01	0.9	1.25	1.07	0.95	0.26	0.19	1.21	1	-0.43	-1.34	-0.37	-1.2	1.2	0.11	0.44	0.61	0.24	0.7
CE1	0.485	1	0.99	1.14	1.24	0.47	0.37	1.22	1.25	0.57	0.29	0.22	1.19	0.97	1.29	0.85	0.77	1.28	0.99	1.1
CE2	0.61	1.78	2.11	2.07	2.28	0.76	1.14	2.26	2.15	1	0.18	0.05	2.17	1.96	1.87	1.44	1.68	2.31	1.98	2.04
CE3	-0.613	-1.71	-2.07	-2.11	-2.24	-0.7	-1.09	-2.22	-2.23	-0.99	-0.16	-0.04	-2.15	-1.88	-2.26	-1.44	-1.6	-2.27	-1.91	-2.01
MO1	0.582	2.32	0.39	-0.13	0.32	2.25	0.66	0.7	0.64	0.11	-0.09	0.27	0.7	-0.01	1.52	-0.1	0.95	0.77	-0.91	1.08
MI1	0.308	0.65	-0.76	-0.89	-0.8	-0.31	-0.29	-0.48	-0.71	-1.16	-1.27	-1.26	-0.79	-0.63	-0.56	-1.04	-0.95	-0.42	-1	-0.79
TR1	0.55	1.12	0.33	2.34	0.71	-0.07	0.09	1.16	3.43	0.34	-0.29	-0.63	2.36	0.49	4.71	-0.12	0.54	1.53	0.1	0.73
CV1	0.199	-0.6	-0.53	-1.05	-0.36	0.36	-0.73	-0.48	-0.41	-0.22	0.03	0.98	-1.1	-0.28	-0.72	-0.12	-0.16	-1.04	0	-0.28
CV2	0.27	0	-1.19	1.52	1.05	0.86	-0.16	0.26	1.73	-0.61	-2.49	2.54	-0.29	-1.46	0.42	-1.83	-1.31	0.4	-2.1	1.2

Tabla A.12 – Importancia relativa de cada variable en relación al ISCA, Cuscatlán

MUNICIPIO	ISCA	CUSCATLAN																					
		CANDELARIA	COJUTEPEQUE	EL CARMEN	EL ROSARIO	MONTE SAN JUAN	DE	BARTOLOME	SAN	CRISTOBAL	SAN JOSE	GUAYABAL	SAN PEDRO	PERULAPAN	SAN RAFAEL	CEDROS	SAN RAMON	SANTA CRUZ	ANALQUITO	SANTA CRUZ	MICHAPA	SUCHITOTO	TENANCINGO
RS1	-0.1	0.62	3.71	-0.6	0.87	-0.5	-1	-0.1	-0.5	-0.4	0.14	-0.2	-0.5	-0.2	-0	-0.3	-0.7						
RS2	0.5	-0.3	1.64	-0.9	-0.7	-1	-1.2	0.78	-1.2	0.51	-0.8	-0.2	-0.3	-1	0.29	-0.2	-0.4						
RS4	0.46	-0.9	-1	0.32	-1.3	-0.1	-0.4	-0.1	-0.4	0.69	0.19	0.22	0.32	-0.4	0.71	0.32	0.32						
AG1	-0.1	0.5	0.09	0.44	0.25	0.84	1.43	1.14	0.74	0.71	0.98	-0.2	0.62	0.69	0.28	0.68	-0.1						
AG2	0.52	-0.6	1.07	-0.7	-0.7	-0.7	-0.7	1.03	-0.7	-0	-0.7	-0.2	-0.6	-0.4	-0.7	0.07	-0.6						
AG3	0.35	-0.4	0.57	0.4	0.11	0.07	-1.3	-0.1	-0.3	-0.9	0.35	0.09	-2	-1.9	0.19	0.59	0.1						
AG4	0.3	0.93	-0.6	-2.3	-2.4	-0.2	-2.4	0.66	-2.4	0.42	-0.6	-0.4	0.98	0.26	-0.2	0.59	0.08						
CE1	0.49	-0	0.71	-0.6	-0.1	-0.3	0.44	0.59	-0.2	0.12	-0.1	0.73	0.31	-0.1	0.3	0.11	0.27						
CE2	0.61	0	1.52	-0.5	-1	-0.9	-0.3	1.21	-1.1	-0	0.05	0.9	-0.2	-0.3	0.67	-0.2	-0.6						
CE3	-0.6	0.02	-1.5	0.44	1.02	0.94	0.25	-1.2	1.09	0.02	-0	-0.8	0.26	0.25	-0.7	0.23	0.6						
MO1	0.58	-0.1	1.67	-0.4	-1.9	-0.5	-1.7	0.88	-1	0.53	-1	-0.4	-1.4	2.45	1.24	-0.1	0.65						
MI1	0.31	-0.8	-0.3	-1.2	-0.2	-1	-0.4	-0.7	-0.9	-1	-1.2	-0.2	-0.7	-0.4	-1.1	-0.3	-0.7						
1	0.55	-0.7	0.83	-1	-0.9	-1.1	-0.5	0.02	-1	-0.4	-0.7	0.36	-0.8	-0.7	-0.4	-0.5	-1						
CV1	0.2	-2.2	-0.4	-0.4	-0.4	0.98	0.98	-0.4	-1.8	-0.4	-0.3	-0.9	0.98	-1.8	-0.6	-0.6	0.98						
CV2	0.27	-2.3	0.07	1.13	0.86	0.46	-1.6	0.46	0.46	1.36	1.2	0.18	1.57	-0.6	-0.1	0.12	-1.1						

Tabla A.13 – Importancia relativa de cada variable en relación al ISCA, Cabañas

MUNICIPIO	CABAÑAS									
	ISCA	CINQUERA	GUACOTECTI	ILOBASCO	JUTIAPA	SAN ISIDRO	SENSUNTEPEQUE	TEJUTEPEQUE	VICTORIA	DOLORES
RS1	-0.14	-0.39	0.05	0.16	-0.59	-0.39	-0.11	0.52	-0.65	-0.43
RS2	0.503	0.14	-0.28	0.14	-0.66	-0.27	0.36	0.32	-0.57	0.19
RS4	0.455	0.36	0.39	-0.38	0.96	0.85	0.96	1.36	0.96	-0.03
AG1	-0.068	0.01	-0.79	-0.13	-0.96	-0.74	-0.96	0.15	0.16	-1.02
AG2	0.517	-0.73	-0.67	0.38	-0.57	0.36	0.33	-0.63	-0.71	-0.5
AG3	0.347	0.58	0.61	-0.54	-1.19	-0.47	0.19	-0.02	0.16	-0.27
AG4	0.296	-0.12	-1.83	-1.27	-0.45	0.61	-1.4	-0.91	-0.91	-1.38
CE1	0.485	0.21	0.49	-0.21	0.06	0.55	0.13	0.1	0.19	-0.53
CE2	0.61	-1.13	0.07	0.11	-1.54	-0.3	0.03	-0.31	-1.08	-1
CE3	-0.613	1.1	-0.01	-0.08	1.52	0.27	-0.03	0.31	1.06	1
MO1	0.582	0.6	0.37	0.48	0.13	2	1.19	0.78	-0.11	-1.46
MI1	0.308	-1.01	0.75	0.39	0.76	1.64	1.33	-0.18	1.88	0.89
TR1	0.55	-0.82	0.44	0.23	-0.6	0.15	0.3	-0.3	-0.72	-0.37
CV1	0.199	0.35	0.98	0.42	0.98	-0.93	0.35	-0.93	0.98	0.98
CV2	0.27	-1.43	-1.55	0.64	-0.43	-0.14	1.15	-1.81	-0.31	-1.93

Tabla A.14 – Importancia relativa de cada variable en relación al ISCA, San Vicente

MUNICIPIO	ISCA	SAN VICENTE												
		APASTEPEQUE	GUADALUPE	CAYETANO	SANTA CLARA	SANTO DOMINGO	ESTEBAN	SAN ILDEFONSO	SAN LORENZO	SAN SEBASTIAN	SAN VICENTE	TECOLUCA	TEPETITAN	VERAPAZ
RS1	-0.14	-0.57	-0.5	0.95	0.35	0.02	-0.26	-0.22	-0.3	-0.89	-0.55	-0.5	-0.36	-0.13
RS2	0.503	-0.31	0.6	-0.03	-0.55	0.32	0.35	-0.45	0.3	0.63	1.11	-0.55	0.56	0.26
RS4	0.455	0.84	1.67	-1.38	-0.37	-0.52	0.45	0.18	0.84	0.84	0.84	0.84	0.85	-0.07
AG1	-0.068	0.16	1.22	-0.15	-0.24	0.77	-0.63	-1.36	0.4	0.64	0.38	0.49	0.75	0.74
AG2	0.517	-0.02	-0.67	-0.59	-0.73	0.57	-0.74	-0.59	-0.7	0.47	1.24	-0.65	-0.7	-0.56
AG3	0.347	-0.68	0.7	0.51	-1.14	-0.18	-0.52	-0.35	-0.55	-0.65	0.72	-0.71	0.62	0.58
AG4	0.296	-1.12	0.23	-0.15	-0.39	-0.01	-2.31	-0.4	-0.5	-0.53	-0.19	0.81	-0.04	0.49
CE1	0.485	0.52	0.24	0.7	-0.21	0.31	0.05	0.01	0.75	0.6	0.73	-0.25	-0.02	0.33
CE2	0.61	-0.32	0.13	0.36	-1.06	0.79	-0.3	-0.91	-0.18	0.44	1.12	-0.42	0.48	0.02
CE3	-0.613	0.35	-0.12	-0.35	1.06	-0.8	0.27	0.86	0.1	-0.46	-1.1	0.42	-0.38	-0.01
MO1	0.582	1.66	0.52	0.69	-0.29	-0.8	-0.04	-0.16	-0.67	1.26	0.89	1.64	0.62	-0.68
MI1	0.308	0.23	-0.19	0.66	0.41	0.33	1.26	0.09	1.06	0.97	0.02	-0.42	-0.27	0.04
TR1	0.55	0.1	-0.35	-0.16	0.08	-0.01	-0.22	-0.37	-0.2	-0.03	0.6	-0.72	-0.11	-0.3
CV1	0.199	-0.93	-0.3	0.98	-0.11	-0.11	-0.34	-0.11	0.98	0.05	-0.64	-1.75	0.98	-0.11
CV2	0.27	0.72	0.65	-0.63	-0.2	-0.39	-1.44	0.29	2.26	0.89	0.85	0.63	0.45	0.98

Tabla A.15 – Importancia relativa de cada variable en relación al ISCA, La Paz

MUNICIPIO	ISCA	LA PAZ																									
		CUYULTITAN	EL ROSARIO	JERUSALEN	MERCEDES LA CEIBA	OLOCUILTA	PARAISO DE OSORIO	ANTONIO	SAN EMIGDIO	FRANCISCO	SAN JUAN	NONJALCO	SAN JUAN TALPA	SAN JUAN TEPEZONTES	SAN LUIS TALPA	SAN MIGUEL TEPEZONTES	SAN PEDRO	MASAHUAT	SAN PEDRO	NONJALCO	SAN RAFAEL	OBRAJUELO	MARIA	SANTIAGO	NONJALCO	TAPALHUACA	ZACATECOLU
RS1	-0.1	-0.3	-0.4	-1	-0.5	-0.1	-0.2	0.31	-0.9	-0.2	-0.5	-0.3	-0.2	-0.3	-0.7	-1	-0.5	-0.4	0.15	-0.1	-0.2	-0.6	0.54				
RS2	0.5	0.71	0.3	-0.9	-0.9	0.54	-1.2	0.03	-1.2	-1.1	0.52	0.48	-0.9	-0.3	-0.7	-0.7	0.15	0.4	-0.7	-0.6	-0.5	0.27	-0.2				
RS4	0.46	0.51	1.11	-0.6	-0.6	-0.1	1.63	-0.2	-0.6	-0.6	1.44	0.53	-0.6	1.81	-0.6	0.42	0.42	1.1	0.13	-0.2	0.02	0.42	-0.8				
AG1	-0.1	0.29	0.38	1.09	0.49	0.96	1.08	-0.1	0.96	0.98	0.92	0.03	0.04	0.29	1.13	0.28	1.51	0.7	1.16	0.72	1.24	0.56	-0.7				
AG2	0.52	-0.7	0.97	-0.6	-0.6	0.88	-0.7	-0.6	-0.2	-0.7	0.99	0.3	-0.6	-0.4	-0.6	-0.7	-0.2	0.5	-0.7	-0.2	-0.7	0.6	-0.1				
AG3	0.35	0.43	0.3	0.08	-0.4	0.67	-1.9	-1.2	-2.5	-1.8	-0.1	0.73	-1.5	-0	-1.1	-1.3	-0.1	0.6	-0.3	-0.4	-1.6	-1.1	-1.2				
AG4	0.3	-1.3	0.91	-0.4	1.3	-1.6	0.74	-1.4	-0.1	-2.1	0.8	-1.4	1.11	0.43	0.83	0.57	-0.9	0.8	-0.3	-0.4	-2	0.53	1.14				
CE1	0.49	0.88	0.57	-0.3	-0	0.77	0	0.36	-0.1	-0.2	0.55	0.61	-0.2	0.36	-0.4	0.2	-0.2	0.8	-0.3	0.28	-0.4	0.24	0.35				
CE2	0.61	1.19	1.06	-0.6	-0.9	1.29	-0.5	-0.8	-0.7	-0.4	0.37	1.29	-0.5	0.81	0.02	0.47	-0.5	0.7	-0.8	0.05	-0.2	0.72	0.14				
CE3	-0.6	-1.2	-1.1	0.52	0.79	-1.3	0.51	0.71	0.64	0.41	-0.4	-1.3	0.43	-0.8	-0	-0.5	0.49	-0.7	0.81	-0.1	0.13	-0.7	-0.2				
MO1	0.58	1.22	-0.3	-2	0.03	-0.2	-0.3	0.13	-1.1	-1.1	0.67	1.05	1.53	-0.2	-0.8	-0.2	0.83	0	-0.6	0.54	0.54	0.93	-0.4				
MI1	0.31	-0.9	-0.8	-1.2	-0.7	-1	-0.4	-0.5	-0.4	-1.2	-0.4	-0.5	-0.2	-0.7	-1.3	-0.5	-1.1	0.5	-1.3	-0.5	-0.5	-0.6	-0.8				
TR1	0.55	0.53	-0.1	0.04	0.14	0.18	-0.8	-0.8	-0.7	-1	-0.1	-0	-0.7	-0.3	-0.8	-0.3	-0.7	0.3	-0.6	-0.1	-0.4	0.27	-0.7				
CV1	0.2	0.11	0.16	0.16	0.98	0.6	0.98	0.16	0.16	-1.8	0.98	-0.3	0.98	-0	0.98	0.98	0.98	1	0.16	0.21	0.98	-0.5	-0.9				
CV2	0.27	0.19	-1	-0.7	-1.2	0	0.64	2.86	-1.5	-0.6	-2.3	0.68	-0.6	1.2	-0.7	-0.3	1.05	0.2	0.65	-0.2	-0.4	0.97	0.35				

Tabla A.16 – Importancia relativa de cada variable en relación al ISCA, Usulután

MUNICIPIO	ISCA	USULUTAN																						
		ALEGRIA	BERLIN	CALIFORNIA	CONCEPCION BATRES	EL TRIUNFO	EREGUAYQUIN	ESTANZUELAS	JIQUILISCO	JUCUAPA	JUCUAPAN	MERCEDES UMAÑA	NUEVA GRANADA	OZATLAN	PUERTO EL TRIUNFO	SAN AGUSTIN	SAN BUENA VENTURA	SAN DIONISIO	SANTA ELENA	FRANCISCO JAVIER	SANTA MARIA	SANTIAGO DE MARIA	TECAPAN	USULUTAN
RS1	-0.1	5.78	-0.2	-0.3	0.09	0.16	4.67	-0	1.15	-0.6	-0.8	-0.7	-0.5	0.84	-1.1	-0.2	-0	-0.5	0.22	-0.2	-0.1	-0.3	-0.2	0.14
RS2	0.5	-0.8	0.79	1.3	-0.7	0.57	-0.5	0.01	-0.7	0.54	-0.9	-0.3	-1.1	-1.2	0.11	-1.2	-0.3	-1.2	-0.3	-1.1	0.29	1.43	-1.1	0.17
RS4	0.46	-3.5	1.53	0.27	-0.7	-0.9	-1.4	0.75	1.42	0.27	0.27	0.27	-0.7	-1.3	0.27	-0.7	-0.4	-0.7	-1	-0.7	0.27	0.27	-0.7	1.21
AG1	-0.1	1	0.06	-2.5	1.06	0.22	0.66	0.29	1.09	0.5	0.68	0.84	0.5	1.25	0.83	0.97	0.47	0.7	1.43	0.81	0.71	0.32	1.31	0.47
AG2	0.52	-0.7	-0	-0.6	-0.6	-0.7	-0.7	-0.7	-0.4	0.63	-0.7	-0.7	-0.7	-0.7	0.58	-0.7	-0.7	-0.4	0.3	-0.7	-0.4	1.09	-0.7	0.55
AG3	0.35	-2.6	-0.1	1.11	-1.5	-0.1	0.41	0.56	-0.9	0.7	-0.5	0.48	0.37	-0.9	-1.1	-0.5	0.49	-1.8	0.87	-0.2	-1.7	0.64	-1.6	0.14
AG4	0.3	1.06	-1.9	-2.5	1.24	1	1.34	-0.3	1.36	0.24	-0.2	-1.2	0.09	-0	1.34	-0.5	0.4	1.33	-0	1.07	1.24	-1.9	-1.9	1.2
CE1	0.49	0.03	-0.7	1.15	0.23	0.29	0.69	0.73	0.25	0.33	-0.1	0.53	0.44	0.39	0.62	-0.5	0.46	0.25	0.46	-0.2	0.9	0.8	-0.3	0.83
CE2	0.61	-1	-0.2	0.07	0.04	0.25	0.3	-0.5	0.04	0.15	-1.1	-0.5	-1.2	-0.1	1.21	-0.7	-0.4	0.25	-0.2	-1.2	1.15	0.98	-0.5	1.22
CE3	-0.6	1.06	0.24	-0.1	0	-0.3	-0.3	0.46	-0	-0.2	1.09	0.5	1.16	0.11	-1.2	0.69	0.41	-0.3	0.16	1.18	-1.1	-1	0.46	-1.2
MO1	0.58	-0.8	0.79	-0.3	1.5	0.36	-0.8	2.67	0.78	0.24	0.62	0.69	-0.1	-0.5	0.31	-0.2	-0.2	-1.8	0.5	-0.4	1.5	1.7	0.67	1.15
MI1	0.31	-1.1	-0.5	0.01	1.28	0.48	-1	1.47	-0.2	-0.5	1.29	1.13	0.56	-0.4	-0	-0.5	0.42	-0.4	-0.1	-0.6	-0.4	-0.1	0.25	0.05
TR1	0.55	-0.8	-0.3	-0.6	0.14	0.4	0.89	0.17	-0.4	0.45	-0.7	-0.2	-0.2	-0.4	-0.4	-0.9	0.47	-0.7	0.21	-0.5	0.84	0.33	-0.3	1
CV1	0.2	0.98	-0.9	-0.9	-0.2	0.98	-0.2	0.11	-0.5	-0.9	0.98	-0	0.98	-1.8	0.98	0.98	-1.8	-0.2	-0.9	0.98	0.98	-1.2	0.98	-0.9
CV2	0.27	1.83	0.2	1.13	1.34	0.83	0.67	-1.1	-2.4	0.53	0.51	0.96	-1.3	0.28	0.54	-0.7	-0.4	-1.5	-0.5	-0.6	-0.2	-0.4	-1.7	0.22

Tabla A.17 – Importancia relativa de cada variable en relación al ISCA, San Miguel

MUNICIPIO	SAN MIGUEL																					
	ISCA	MORAZAN																				
	CAROLIN	CIUDAD BARRIO	COMACAF	CHAPELT E	CHINAME	CHIRILAG	EL TRANSIT	LOLOTIQ	MONCAG	NUEVA GUADALU	NUEVA EDEN D	SAN JIL	QUELEF	SAN ANTONI	SAN GERARD	SAN JOR	DE LA PEÑA	SAN MIG	SAN RAF	ORIENT	SESOR	ULUAZA
RS1	-0.14	1.17	0.29	-0.21	-0.95	-0.53	1.1	-0.56	0.72	0.72	0.54	-0.48	-0.5	0.84	-0.28	1.75	-0.29	0.9	-0.21	-0.56	0.1	
RS2	0.503	-0.37	-0.44	-1.13	-0.24	-0.19	-0.5	-0.4	-0.88	-0.81	1.09	-1.17	-1.15	-0.72	-0.72	-0.73	-0.53	1.48	-1.14	-0.63	-0.4	
RS4	0.455	-2.35	-1.23	-1.1	-0.43	1.86	-1.61	-0.43	-1.07	-1.06	-0.68	-1.1	-1.1	-1.98	0.39	-2.18	0.46	0.13	-1.1	-0.43	-0.82	
AG1	-0.068	-2.44	-0.73	-0.36	-0.81	1.06	0.36	0.84	0.87	0.85	0.99	-1.96	0.06	-2.41	-2.93	1.05	-0.7	0.37	1.27	-1.73	0.24	
AG2	0.517	0.36	0.37	-0.6	0.21	0.05	-0.53	0.24	-0.7	-0.68	0.76	-0.73	-0.7	-0.57	-0.74	-0.73	-0.4	1.8	-0.08	-0.58	-0.73	
AG3	0.347	0.6	1.14	1.46	0.07	-1.25	-0.3	1.01	-0.27	-0.54	1.32	-2.16	-0.53	-0.14	-1.54	-2.7	-0.56	0.55	-0.23	-1.23	0.32	
AG4	0.296	1.19	0.76	-0.94	0.2	-0.02	0.18	1.18	-0.45	-0.78	-1.04	-0.78	-0.12	0.63	0.44	-1.99	0.36	0.78	-0.03	-0.62	-0.71	
CE1	0.485	-2	-0.63	0.12	-0.03	-0.34	0.5	0.1	0.15	0.47	0.74	-0.68	0.46	-2.83	0.43	-0.48	0.35	0.87	0.18	-0.41	0.77	
CE2	0.61	-1.37	-0.75	-0.32	-0.27	-0.33	-0.37	0.27	-1.07	0.62	0.79	-1.38	0.54	-1.77	-1.12	-0.53	-1.23	1.57	-0.15	-1.12	-0.13	
CE3	-0.613	1.4	0.75	0.32	0.28	0.35	0.36	-0.24	1.06	-0.6	-0.74	1.3	-0.56	1.77	1.11	0.52	1.24	-1.54	0.18	1.13	0.08	
MO1	0.582	-1.2	-0.11	1.28	-0.6	-0.4	0.49	2	-0.19	-0.4	0.46	-0.58	0.41	-1.15	-1.16	-1.2	-0.95	1.08	3.16	-1.6	-0.38	
MI1	0.308	-0.37	0.67	1.01	-0.04	-0.75	0.14	0.8	-0.1	0.6	0.04	1.03	0.75	-0.77	-1.16	1.14	1.5	-0.02	0.9	0.28	-0.12	
TR1	0.55	-1.02	-0.04	0.27	0.21	-0.26	-0.19	0.35	-0.4	0.17	0.99	-0.71	0.73	-1.25	-0.59	0.39	-0.06	1.8	0.28	-0.54	0.16	
CV1	0.199	0.98	0.21	0.23	0.23	-1.75	0.11	-1.75	0.98	0.98	-1.15	0.23	0.98	0.98	0.98	0.98	0.23	-1.1	0.98	0.23	0.98	
CV2	0.27	-2.7	0.87	-0.35	1.61	-0.56	0.2	0.08	0.48	-1.09	0.85	-0.86	-0.47	0.86	-0.01	0.52	-2.08	1.2	-0.71	0.09	-0.35	

MUNICIPIO	ISCA	ARAMBALA	CACAOPELA	CORINTO	CHILANGA	DELICIAS DE CONCEPCION	EL DIVISADERO	EL ROSARIO	GUALOCOCTI	GUATAJIAGUA	JOATECA	JOCOAITIQUE	JOCORO	LOLOTIQUILLO	MEANGUERA	OSCICALA	PERQUIN	SAN CARLOS	SAN FERNANDO	FRANCISCO	SAN ISIDRO	SAN SIMON	SENSEMBRA	SOCIEDAD	TOROLA	YAMABAL	YOLOAIQUIN
RS1	-0.14	-0.41	1.52	-0.57	-0.29	3.09	-0.3	-0.22	-0.46	0.37	-0.46	-0.36	-0.16	-0.56	0.84	-0.47	-0.58	0.54	-1.25	-0.56	-0.48	-0.47	-0.47	-0.57	-1.3	-0.48	-1.09
RS2	0.5	-1.01	-0.7	-0.09	-0.74	-0.44	-0.64	-1.12	-1	-0.24	-1.15	-0.53	-0.01	-0.36	0.06	-0.06	-0.37	0.29	-1.15	1.19	-1.16	-0.88	-1.17	-0.8	-1.1	-1.19	-0.47
RS4	0.46	-1.36	-0.16	-0.16	-2.44	-2.86	-0.16	-0.06	-0.06	-1.39	-0.06	0.84	-0.08	-0.16	-1.23	1.39	-0.16	-0.83	-0.06	-0.16	-0.06	-0.06	-0.06	-0.16	-0.1	-1.36	-0.16
AG1	-0.07	0.41	-0.46	-0.75	-1.1	-1.22	-0.6	-1.05	0.59	-1.86	-0.13	0.28	-0.59	-0.68	0.48	0.12	-0.12	-0.18	-0.41	0.17	-2	-0.87	-1.95	-1.27	-0.7	-1.51	-1.86
AG2	0.52	-0.57	-0.14	-0.44	-0.68	-0.03	-0.16	-0.74	-0.7	-0.42	-0.32	-0.73	0.25	-0.73	-0.71	0.5	-0.52	0.13	-0.57	2.15	-0.7	-0.26	-0.72	-0.08	-0.6	-0.75	-0.4
AG3	0.35	-1.29	-1.4	-1.18	0.33	-0.36	0.1	0.33	1.25	-0.52	-0.24	0.88	-1.01	-0.86	-0.05	1.22	0.72	-0.66	0.29	0.99	0.71	0.54	-0.79	-1.29	-0.9	-1.13	0.48
AG4	0.3	1.09	-1.05	0.05	0.66	-0.41	-2.07	1.2	0.75	-0.85	1.06	-1.43	-0.87	-0.37	-0.98	0.9	1.34	-2.43	0.8	-0.93	-1.55	0.68	-0.05	-1.86	0.81	0.34	0.41
CE1	0.49	-2.06	-1.86	-2.86	-1.06	-0.57	0.23	-0.96	-2.13	-2.92	-3.95	-1.16	0.67	-0.14	-1.63	-0.94	-2.48	-0.58	-2.46	0.54	-1.96	-3.49	-1.84	-0.41	-3.4	-2.77	-0.21
CE2	0.61	-1.18	-1.62	-0.78	-1.07	-0.46	0.27	-1.22	-1.69	-1.29	-1.52	-1	0.47	-0.68	-1.24	-0.83	-1.15	-0.24	-1.41	0.96	-1.78	-1.65	-1.15	-0.53	-1.7	-0.88	-0.88
CE3	-0.61	1.2	1.62	0.9	1.08	0.43	-0.25	1.21	1.69	1.32	1.54	0.95	-0.45	0.69	1.21	0.83	1.11	0.27	1.42	-0.94	1.77	1.67	1.12	0.52	1.76	1.01	0.81
MO1	0.58	-2.76	1.84	-0.54	-0.57	-0.25	-1.09	2.28	-2.02	-1.35	-0.73	-0.86	-1.24	1.32	-0.61	-0.37	-1.44	-1.59	0.06	0.1	-0.35	-0.02	1.63	-0.89	-1.2	-1.48	0.24
MI1	0.31	1.06	-0.38	-0.09	0.08	1.17	0.01	1.88	-0.64	-1.18	1.98	-0.87	-0.32	0.69	0.1	-1.24	-0.49	-0.82	0.58	-0.43	-1.57	-0.56	1.83	0.65	-0.7	-0.9	0.07
TR1	0.55	-0.64	-0.96	-0.46	-0.67	0.52	0.46	-0.84	-1.21	-0.85	-0.72	-0.78	1.63	-0.21	-0.91	0	0.44	0.37	-0.65	0.9	-1.39	-1.13	-0.12	0.2	-1.2	-0.31	-0.18
CV1	0.2	0.63	0.18	0.21	0.98	-0.22	0.98	0.98	0.98	-0.22	0.98	0.98	0.98	0.98	0.42	0.98	0.98	0.98	-1.38	0.98	0.5	0.98	0.98	0.98	-0.3	0.98	0.98
CV2	0.27	-0.86	-0.57	0.86	1.03	0.84	-0.57	-0.57	-1	0.52	0.38	0.51	0.33	-0.37	-1.02	-1.06	-0.18	-0.67	-1.95	0.82	-1.19	0.38	0.53	2.62	-0.7	-0.67	-1

Tabla A.18 – Importancia relativa de cada variable en relación al ISCA, Morazán

Tabla A.19 – Importancia relativa de cada variable en relación al ISCA, La Unión

		LA UNION																	
MUNICIPIO	ISCA	ANAMOROS	BOLIVAR	CONCEPCION DE ORIENTE	CONCHAGUA	EL CARMEN	EL SAUCE	INTIPUCA	LA UNION	LISLIQUE	NUEVA ESPARTA	PASQUINA	POLOROS	SAN ALEJO	SAN JOSE	SANTA ROSA DE LIMA	YAYANTIQUE	YUCUAQUIN	MEANGUERA DEL GOLFO
RS1	-0.14	-0.42	-0.21	0.36	-0.58	-0.15	2.58	-0.1	1.87	-1.25	-0.09	0.14	-0.41	-0.62	0.78	-0.27	-0.47	0.04	1.08
RS2	0.503	-0.51	-0.43	-0.84	-0.37	-0.88	-0.54	-0.18	0.11	-0.88	-0.6	-0.41	-0.71	-0.7	0.11	0.63	-1.18	-0.5	-0.97
RS4	0.455	1.07	0.2	-0.43	0.32	1.17	-2.69	1	-1.23	0.32	-0.32	0.19	0.96	0.32	-1.16	0.32	0.06	0.49	-1.56
AG1	-0.068	-2.12	-1.17	-3.21	0.09	0.15	-1.44	0.47	-0.53	-1.31	-2.13	-1.23	-3.53	0.12	-1.71	-0.3	0.18	-0.29	-2.98
AG2	0.517	-0.53	-0.69	-0.65	-0.23	-0.62	-0.6	-0.11	0.39	-0.55	-0.52	-0.6	-0.42	-0.55	-0.69	0.54	-0.71	-0.71	-0.64
AG3	0.347	-1.72	-1.57	-0.49	0.12	-0.76	-1.68	-0.07	0.31	-1.32	-1.34	-0.87	-0.83	-1.06	-1.27	-0.85	0.35	-0.03	-1.27
AG4	0.296	-0.18	0.69	-1.31	-0.11	-0.04	-1.94	1.02	-0.31	0.77	0.93	-0.65	-0.67	-0.37	0.92	-1.33	-0.83	-0.88	1.25
CE1	0.485	-0.43	0.62	-0.35	0.28	0.63	0.46	0.39	0.84	-2.54	-0.3	0.77	-0.61	0.58	0.28	0.67	-0.05	0.28	0.81
CE2	0.61	-0.63	0.47	-0.38	0.72	0.61	-0.28	0.41	1.23	-1.58	-0.71	1.25	-0.98	0.13	-0.3	1.37	-0.27	-0.96	0.48
CE3	-0.613	0.66	-0.43	0.39	-0.68	-0.58	0.25	-0.4	-1.23	1.58	0.68	-1.21	0.99	-0.12	0.25	-1.34	0.28	0.93	-0.49
MO1	0.582	-1.08	-0.24	0.36	-0.45	-0.7	0.06	1.01	0.49	-1.13	-0.54	0.15	1.17	1.39	-2.05	-0.24	1.69	0.94	-1.02
MI1	0.308	0.7	0.48	3.81	0.44	0.82	1.56	1.21	1.22	0.27	0.91	2.08	1.58	0.68	-0.55	0.64	0.54	1.88	1.85
TR1	0.55	0.48	1.34	0.42	0.16	0.12	0.82	0.17	0.7	-1	0.05	1.15	0.14	0.32	-0.09	1.59	-0.42	-0.24	0.23
CV1	0.199	0.98	-0.19	-0.22	0.98	0.98	-0.19	0.98	-0.68	0.11	-0.19	-2.21	0.98	0.98		-0.76	-0.19	0.98	0.98
CV2	0.27	-0.74	0.74	-0.71	-1.05	0.51	-0.04	0.46	1.01	1.59	-0.1	-0.14	0.24	0.17	0.24	0.16	1.1	2.34	1.01

BIBLIOGRAFÍA.

Referencias.

- [1] Aluja, T. y Morineau, A. (1999). Aprender de los datos: El análisis de componentes principales, EUB, Barcelona.
- [2] Batista, J. M. y Martínez, M. R. (1989). Análisis Multivariante, ESADE.
- [3] Bosque, J. (2000). Sistema de información geográfico. Madrid: Ediciones Rialp.
- [4] Castro, J. M. (2002). Indicadores de desarrollo sostenible urbano. Una aplicación para Andalucía. Tesis doctoral. Málaga: Universidad de Málaga.
- [5] Cuadras, C. M. (1991). Métodos de Análisis Multivariante, Ed. Universitaria de Barcelona (3ª.Ed), Primera edición EUNIBAR, 1981.
- [6] Dallas E. Johnson. (2000). Métodos Multivariados Aplicados al Análisis de Datos. International Thomson Editores.
- [7] Daniel Peña. (2002). Análisis de Datos Multivariado. Editorial: Concepción Fernández, Madrid. Editorial: Mc. Graw Hill.
- [8] Dillon, W., Goldstein, M. (1984). Multivariate Analysis, New York, Wiley.
- [9] Eaton, M. L. (1983). Multivariate Statistics, New York, Wiley.
- [10] Escobar, L. A. (2004). Índices de calidad ambiental urbana: un modelo general y caso aplicado a Cali, Colombia. Universidad de Alcalá de Henares (mimeo).

- [11] Flury, B. (1988). *Common Principal Components and Related Multivariate Methods*, Wiley.
- [12] Jambu, M. (1991). *Applied Multivariate Data Analysis*, Springer.
- [13] Jackson, J. E. (1991). *A User's guide to principal components*, Wiley.
- [14] Joliffe, I. T. (1986). *Principal Components Analysis*, Springer-Verlag.
- [15] Krzanowski, W. J. y Marriot, F. H. C. (1995), *Multivariate Analysis: part I y II* – Edward Arnold, London.
- [16] Laura E. Pla. (1986). *Análisis multivariado: Método de Componentes Principales*, Programa Regional de Desarrollo Científico y Tecnológico. Secretaria General de la Organización de los Estados Americanos. Monografía No.27. Coro, Falcón, VENEZUELA.
- [17] Quiroga, R. (2001). *Indicadores de sostenibilidad ambiental y de desarrollo sostenible: estado del arte y perspectivas*. Santiago: CEPAL.

Manuales y Revistas:

- [18] *Agenda Local 21 (El Programa 21)*. (1991). Naciones Unidas sobre el Medio Ambiente y el Desarrollo.
- [19] EUROSTAT. (1998). “Indicadores de desarrollo sostenible. Estudio piloto según la metodología de la Comisión de Desarrollo Sostenible de las Naciones Unidas”. Luxemburgo: Oficina de Publicaciones.

- [20] MMA. (1996). Indicadores ambientales. Una propuesta para España. Madrid: Dirección General de Calidad y Evaluación Ambiental, Ministerio de Medio Ambiente.
- [21] Revista **Eure** (Vol. XXXII, N° 96), pp. 73-98. Santiago de Chile, agosto de 2006
- [22] UNEP/PNUMA. (2001). “Informe sobre los indicadores ambientales y de la sustentabilidad en América Latina y el Caribe”. *XIII Reunión del Foro de Ministros de Medio Ambiente de América Latina y el Caribe*. Nueva York: UNEP/LACIG. XIII/Inf.4.