



# Análisis de tablas múltiples para la caracterización de la relación entre la pobreza y la actividad económica de los departamentos de Colombia en el año 2018

Diana Carolina Arias Sinisterra

Universidad del Valle  
Facultad de Ingeniería, Escuela de Estadística  
Santiago de Cali, Colombia  
2022

# **Análisis de tablas múltiples para la caracterización de la relación entre la pobreza y la actividad económica de los departamentos de Colombia en el año 2018**

**Diana Carolina Arias Sinisterra**

Trabajo de grado presentado como requisito parcial para optar al título de:  
**Estadística**

Director:  
Jaime Mosquera Restrepo,PhD.

Universidad del Valle  
Facultad de Ingeniería, Escuela de Estadística  
Santiago de Cali, Colombia  
2022

«El futuro tiene muchos nombres. Para los débiles es lo inalcanzable. Para los temerosos, lo desconocido. Para los valientes es la oportunidad.»

Victor Hugo

## **Agradecimientos**

A mis padres quienes fueron un apoyo incondicional durante todo este tiempo, a mi hermano quien fue un pilar y una guía en todo este proceso.

De manera especial a mi director de tesis, por haberme brindado todo su conocimiento y haberme guiado para la elaboración de este trabajo de grado.

A la Universidad del Valle, por brindarme experiencias, oportunidades y enriquecerme de conocimientos en muchos ámbitos.

## Resumen

Para estudiar la relación de la pobreza y la actividad económica en las regiones de Colombia, en este trabajo se realiza el análisis simultaneo de los Indicadores de Necesidades Básicas Insatisfechas(NBI) y Producto Interno Bruto (PIB), los cuales se encuentran medidos para las las 33 unidades administrativas en que se divide la nación. El PIB se encuentra segregado en 12 sectores productivos más el rubro por la recaudación de impuestos y el NBI está dividido en 5 componentes. En vista de que los indicadores de pobreza son obtenidos a través del censo poblacional y las cuentas nacionales se obtienen de forma anual, se considera el año 2018 (año del censo más actual) como periodo de referencia para el análisis, ya que se dispone de información de ambos conjuntos de datos. Los datos son obtenidos a través de la plataforma web oficial del Departamento Administrativo Nacional de Estadística (DANE). Como se tienen dos conjuntos de datos, cada agrupación conforma una tabla de datos diferente, por lo tanto se aplican técnicas de Análisis de Múltiples Tablas para el estudio simultaneo de estos dos conjuntos de datos. Por lo tanto, el Análisis Factorial Múltiple (AFM) junto con el método de los mínimos cuadrados parciales (PLS), se encuentra que el conjunto de variables de pobreza y de productividad se encuentran inversamente correlacionadas, es decir, que apunta a que la productividad en las diferentes regiones del país podría contribuir a bajar los altos niveles de pobreza que se presentan en algunas regiones específicas. Adicionalmente a través de la reducción de dimensionalidad del AFM, se toman las dos primeras componentes para realizar un análisis de clúster, en el que se consolidan agrupaciones de departamentos con similares condiciones de actividad económica y de pobreza, en el que adicionalmente se encuentra una relación espacial entre esas agrupaciones de regiones. Finalmente se realiza una breve comparación de los resultados de los datos disponibles para el año 2005 (censo anterior) con los del año 2018, con el objetivo de analizar la existencia de cambios en las relaciones entre el PIB y NBI.

**Palabras clave:** Actividad económica, Pobreza, AFM, PLS, Clúster, PIB, NBI.

## Abstract

In order to study the relationship between poverty and economic activity in the regions of Colombia, this paper analyzes simultaneously the Unsatisfied Basic Needs Indicators (UBN) and Gross Domestic Product (GDP), which are measured for the 33 administrative units into which the nation is divided. The GDP is segregated into 12 productive sectors plus the item for tax collection and the UBN is divided into 5 components. Since the poverty indicators are obtained through the population census and the national accounts are obtained annually, the year 2018 (year of the most current census) is considered as the

reference period for the analysis, since information from both datasets is available. The data are obtained through the official web platform of the National Administrative Department of Statistics (DANE). Since there are two data sets, each grouping forms a different data table, therefore, Multiple Table Analysis techniques are applied for the simultaneous study of these two data sets. Therefore, the Multiple Factor Analysis (MFA) together with the Partial Least Squares (PLS) method, it is found that the set of poverty and productivity variables are inversely correlated, that is, it points to the fact that productivity in the different regions of the country could contribute to lower the high levels of poverty that are present in some specific regions. Additionally, through the dimensionality reduction of the AFM, the first two components are taken to perform a cluster analysis, in which groups of departments with similar conditions of economic activity and poverty are consolidated, in which a spatial relationship between these groups of regions is also found. Finally, a brief comparison of the results of the data available for the year 2005 (previous census) with those of the year 2018 is made, with the objective of analyzing the existence of changes in the relationships between GDP and NBI.

**Keywords:** Economic activity, Poverty, AFM, PLS, Cluster, GDP, NBI

# Contenido

<b>Resumen</b>	v
<b>Lista de Figuras</b>	xI
<b>Lista de Tablas</b>	1
<b>1 Introducción</b>	3
<b>2 Planteamiento del problema</b>	5
2.1 Definición del problema . . . . .	5
2.2 Justificación . . . . .	7
2.3 Objetivos . . . . .	8
2.3.1 Objetivo principal . . . . .	8
2.3.2 Objetivos específicos . . . . .	8
<b>3 Revisión de antecedentes</b>	9
3.1 Análisis de la pobreza y la dinámica económica . . . . .	9
3.2 Análisis de Tablas Múltiples . . . . .	12
<b>4 Marco Teórico</b>	15
4.1 Definiciones económicas y de pobreza . . . . .	15
4.1.1 Producto Interno Bruto (PIB) . . . . .	15
4.1.2 Sectores económicos . . . . .	16
4.1.3 Pobreza . . . . .	21
4.1.4 Necesidades Básicas Insatisfechas (NBI) . . . . .	23
4.2 Análisis de Tablas Múltiples . . . . .	24
4.2.1 Análisis de Componente Principales (ACP) . . . . .	24
4.2.2 Análisis Factorial Múltiple (AFM) . . . . .	30
4.2.3 Método de Mínimos Cuadrados Parciales (PLS) . . . . .	35
4.3 Análisis de Clúster (AC) . . . . .	39
<b>5 Metodología</b>	43
5.1 Descripción de los datos . . . . .	43
5.2 Análisis Estadístico . . . . .	46
5.2.1 Análisis Exploratorio . . . . .	46

5.2.2	Análisis de Tablas Múltiples . . . . .	47
5.2.3	Análisis Clúster . . . . .	49
5.2.4	Comparación de resultados con periodo 2005 . . . . .	50
<b>6</b>	<b>Resultados</b>	<b>51</b>
6.1	Análisis exploratorio de datos . . . . .	51
6.2	Análisis Factorial Múltiple -AFM . . . . .	62
6.3	PLS . . . . .	70
6.4	Integración de Resultados AFM y PLS . . . . .	74
6.5	Análisis de Clúster - AC . . . . .	75
6.6	Comparación con el año 2005 . . . . .	79
<b>7</b>	<b>Conclusiones y recomendaciones</b>	<b>84</b>
7.1	Conclusiones . . . . .	84
7.2	Recomendaciones . . . . .	85
<b>Bibliografía</b>		<b>86</b>

# **Lista de Figuras**

<b>4-1</b>	Matriz de datos $X_{n,p}$ . . . . .	25
<b>4-2</b>	Proximidad entre los individuos - $\mathbb{R}^p$ . . . . .	26
<b>4-3</b>	Círculo de correlaciones . . . . .	27
<b>4-4</b>	Ejemplo representación simultanea-Biplot . . . . .	29
<b>4-5</b>	Representación de la estructura de la tabla de datos en AFM . . . . .	31
<b>4-6</b>	Representación de un dendrograma . . . . .	40
<b>4-7</b>	Representación de las inercias Intra-Clases e Inter-clases . . . . .	41
<b>4-8</b>	Métodos de distancias entre clúster . . . . .	41
<b>4-9</b>	Representación del método del codo . . . . .	42
<b>5-1</b>	Mapa político de Colombia . . . . .	43
<b>5-2</b>	Estructura de los datos . . . . .	45
<b>5-3</b>	Diseño metodológico propuesto . . . . .	46
<b>6-1</b>	Diagrama de cajas de los Indicadores de Pobreza por Departamentos . . . . .	52
<b>6-2</b>	Distribución espacial de los indicadores de pobreza (NBI), año 2018 . . . . .	53
<b>6-3</b>	Gráfico de correlación entre las variables del NBI . . . . .	54
<b>6-4</b>	Diagrama de cajas de las Actividades económicas por Departamentos . . . . .	55
<b>6-5</b>	Distribución espacial de la actividad económica por departamentos, año 2018	57
<b>6-6</b>	Gráfico de correlación entre las variables del PIB . . . . .	58
<b>6-7</b>	Gráfico de dispersión entre los sectores minería y agricultura . . . . .	59
<b>6-8</b>	Gráfico de dispersión variables del PIB y NBI . . . . .	60
<b>6-9</b>	Correlaciones entre las variables NBI y PIB . . . . .	61
<b>6-10</b>	Gráfico de sedimentación y valores propios del AFM . . . . .	63
<b>6-11</b>	Representación de los grupos y su contribución a los ejes factoriales . . . . .	64
<b>6-12</b>	Contribución de las variables a los ejes . . . . .	65
<b>6-13</b>	Contribución de las variables a los ejes . . . . .	67
<b>6-14</b>	Individuos y contribución a los ejes . . . . .	68
<b>6-15</b>	Nube de individuos y representación superpuesta . . . . .	69
<b>6-16</b>	Círculo de correlaciones con método de PLS . . . . .	72
<b>6-17</b>	Nube de individuos con método de PLS . . . . .	72
<b>6-18</b>	Óptimo Número de clústeres . . . . .	76
<b>6-19</b>	Dendrograma Global, NBI y el PIB . . . . .	77
<b>6-20</b>	Ubicación espacial de los Clúster de Departamentos . . . . .	78

---

<b>6-21</b> Comparación de Varianza explicada periodo 2005 y 2018 . . . . .	80
<b>6-22</b> Comparación Circulo de correlaciones periodo 2005 y 2018 . . . . .	81
<b>6-23</b> Comparación ubicación espacial de clústers global periodo 2005 y 2018 . . . .	82
<b>6-24</b> Comparación ubicación espacial de clústers NBI periodo 2005 y 2018 . . . .	82
<b>6-25</b> Comparación ubicación espacial de clústers PIB periodo 2005 y 2018 . . . .	83



# **Lista de Tablas**

<b>4-1</b>	Sectores económicos de desagregación del PIB . . . . .	16
<b>5-1</b>	Sectores productivos de desagregación del PIB en Colombia . . . . .	44
<b>5-2</b>	Indicadores de Necesidades Básicas Insatisfechas . . . . .	45
<b>5-3</b>	Resumen de la estructura de los datos para el desarrollo del AFM . . . . .	47
<b>6-1</b>	Estadísticas descriptivas de los indicadores de pobreza . . . . .	51
<b>6-2</b>	Estadísticas descriptivas de los indicadores del PIB . . . . .	55
<b>6-3</b>	Valores propios y ponderaciones de los ACP de cada conjunto . . . . .	63
<b>6-4</b>	Contribuciones del conjunto del PIB y NBI a las dimensiones . . . . .	64
<b>6-5</b>	Coeficientes Lg . . . . .	65
<b>6-6</b>	Coeficientes RV . . . . .	65
<b>6-7</b>	Coordenadas, Contribuciones y Calidad de representación de las variables . .	67
<b>6-8</b>	Estadístico VIP . . . . .	71
<b>6-9</b>	Correlación entre T y U . . . . .	71
<b>6-10</b>	Valores Predichos para el NBI . . . . .	73



## **Declaración**

Me permito afirmar que he realizado el presente Trabajo de Grado de manera autónoma y con la única ayuda de los medios permitidos y no diferentes a los mencionados en el propio trabajo. Todos los pasajes que se han tomado de manera textual o figurativa de textos publicados y no publicados, los he reconocido. Ninguna parte del presente trabajo se ha empleado en ningún otro tipo de Tesis o Trabajo de Grado.

Igualmente declaro que los datos utilizados en este trabajo están protegidos por las correspondientes cláusulas de confidencialidad.

Santiago de Cali, 14.03.2022

---

Diana Carolina Arias Sinisterra

# 1 Introducción

El Producto Interno Bruto (PIB) es un indicador macroeconómico que mide la actividad económica de un país en un periodo de tiempo determinado. Este indicador, visto desde el enfoque de la producción, es posible desagregarlo por ramas de actividad económica para analizar el aporte de cada sector al crecimiento económico del país. De acuerdo con Bonilla, J. & Ruiz, J. (2018) el PIB es un indicador económico muy importante, dado que por medio de su análisis es posible brindar un panorama general del estado de la economía. Por lo tanto, su importancia radica en la capacidad de determinar si la economía se encuentra en un estado de crecimiento o en contracción, si existe peligro de recesión o de inflación.

Por otra parte, la pobreza, según la Organización de Naciones Unidas (ONU), es una condición caracterizada por una privación de necesidades básicas, es decir, toda aquella persona que no tenga acceso a unas condiciones mínimas que le permitan el desarrollo de una vida básica, pero digna. Por lo anterior, la pobreza de una población, esta constituida por la carencia de una serie de factores, como lo es el acceso a alimentos, agua potable, instalaciones sanitarias, salud, vivienda, educación, así como la información.

Se tiene entonces dos conceptos, los cuales representan las condiciones económicas y de vida de una población. En el caso de Colombia, el cual es un país que administrativamente se divide en 32 departamentos más su distrito capital, cuenta con diversidad de productos y procesos económicos diferenciados e influenciados por diferentes factores según el departamento. Dada la importancia de conocer las condiciones de pobreza de la población y a su vez el crecimiento económico de un país , en Colombia, el Departamento Administrativo Nacional de Estadística (DANE), es la entidad responsable de la planeación, levantamiento, procesamiento, análisis y difusión de las estadísticas oficiales del país. En el desarrollo de una de sus funciones, el DANE realiza un censo poblacional y de vivienda en el que cuenta y caracteriza las personas que residen en el país, así como las viviendas y los hogares del territorio nacional. Es decir, que a través del censo, se obtienen datos sobre el número de habitantes, el como se encuentran distribuidos en el territorio y cuales son las condiciones de vida.

Por lo tanto, el censo permite generar información estadística que sirve para que diferentes organismos ya sean públicos y/o privados del país planeen y tomen decisiones con respecto a temas de desarrollo económico , bienestar social, vivienda, salud, entre otros.

Debido a que para el presente trabajo el interés son las condiciones de pobreza, se tomó el índice de Necesidades Básicas Insatisfechas, el cual es obtenido a través del censo. Este índice se encuentra compuesto por una serie de indicadores con los que se busca determinar si las necesidades básicas de la población en los diferentes departamentos se encuentran cubiertas. Por otra parte, el DANE es la entidad encargada de realizar anualmente el cálculo del PIB para el país, en el cual considera a los 33 sectores en los que se encuentra dividido administrativamente y en el que se desagrega en 12 sectores económicos más el rubro de recaudación de impuestos.

Bonilla, J. & Ruiz, J. (2018), realizaron un análisis multivariante en tres vías (departamentos, sector económico, año), para caracterizar la dinámica económica a partir del PIB. Los resultados reflejaron un continuo incremento de la economía en el que cada departamento ha resultado favorecido. Detectaron que la economía Colombiana se puede caracterizar a través de tres componentes: como la productividad industrial y de servicios , productividad en explotación de recursos naturales y uso del suelo y por último productividad del sector social y comunitario. También se identificó que habían departamentos como Vichada, Vaupes, entre otros, que tenían baja productividad en varios sectores de la economía.

Perez (2005) realizó un análisis espacial de la pobreza Colombiana con base al Índice de Necesidades Básicas Insatisfechas(NBI) y el Índice de Calidad de Vida (QLI). Los resultados mostraron que el NBI de los departamentos no se distribuía aleatoriamente, por el contrario, las necesidades insatisfechas de los departamentos dependen de sus condiciones internas y también la situación de sus vecinos cercanos. Además, encontraron clústers en los que se destaca en número aquellos en los que municipios pobres están rodeados de municipios pobres y municipios ricos están rodeados de municipios ricos.

Dado el panorama en el que se ve un claro crecimiento económico del país se percibe al mismo tiempo que esto no refleja la calidad de vida de su población dado que se encuentra el panorama donde existen departamentos con altos niveles de pobreza. Con el desarrollo del presente trabajo se logra caracterizar la relación existente entre la actividad económica y las condiciones de pobreza de los departamentos de Colombia para el año 2018. Con la aplicación de técnicas multivariadas, específicamente para el Análisis de Múltiples Tablas, donde las técnicas de AFM y PLS, se complementaron en resultados para lograr identificar las relaciones estadísticas entre el conjunto de indicadores del PIB y el conjunto de indicadores de pobreza, en el que se observó que el aporte en productividad de las regiones se ve ciertamente relacionado con las condiciones de vida de su población. Adicionalmente se identificaron departamentos con similares niveles de pobreza y de actividad económica las cuales se veían relacionadas por ser departamentos vecinos.

## **2 Planteamiento del problema**

En este capítulo se presentan las razones que motivan la realización de este trabajo, se explica el panorama económico y de pobreza de Colombia y como estas dos temáticas se ven estrechamente relacionadas, además de declarar los objetivos y alcance del trabajo.

### **2.1. Definición del problema**

Colombia es un país situado en América del sur, que cuenta con una superficie de  $1.141.749 km^2$  y una población aproximada de 48.258.494 de habitantes, según el Departamento Administrativo Nacional de Estadística (DANE,2018). Administrativamente se divide en 32 departamentos descentralizados y un distrito capital (Bogotá); cada uno de los cuales cuenta con una variedad de productos y de procesos económicos diferenciados e influenciados por factores específicos tales como: la geografía de la región, el clima, las políticas del departamento, la calidad del sistema educativo, los cuales generan variabilidad en los bienes y servicios ofrecidos por el departamento.

El crecimiento económico del país puede ser evaluado a través de la evolución temporal del Producto Interno Bruto (PIB), el cual es un indicador económico que refleja el valor monetario de todos los bienes y servicios finales producidos en un determinado periodo de tiempo (Mankiw, 2002). Para el año 2018, según cifras DANE, Colombia presentó un PIB de USD. 331 mil millones y un crecimiento del 2.5 % en el cuarto trimestre de ese año. Este crecimiento ubica al país como una economía en vía de desarrollo.

Estas cifras de crecimiento económico contrastan con las cifras suministradas por el mismo DANE sobre los indicadores de pobreza. Para el año 2018, el 27 % de la población fue clasificada por debajo de la línea de pobreza nacional. Para ese periodo, los departamentos que mostraron un incremento de la pobreza fueron La Guajira y Chocó con 53,7 % y 61,1 %, respectivamente. Estas son regiones ricas en recursos naturales, pero presentan más de la mitad de su población por debajo de la línea de pobreza. En consecuencia, el país cuenta con un panorama contradictorio, puesto que los indicadores muestran que la productividad económica va en aumento, por lo cual se esperaría que la pobreza disminuya. Sin embargo, para algunos departamentos sucede lo contrario, la pobreza se ha incrementado.

Arnedo.V. (2013), realizó un análisis del PIB y de los indicadores de desigualdad,

nivel de vida y pobreza. En su trabajo concluye que el desarrollo de Colombia no se puede medir solo como el aumento del PIB, pues es necesario relacionarlo con el conjunto de variables macroeconómicas y cómo estas influyen en la cobertura de las necesidades básicas de la población, como lo es el nivel de educación, el acceso al servicio de salud y al mercado laboral, las cuales son en general las condiciones que generan bienestar a la sociedad. Los resultados de su investigación indicaron que en el país el crecimiento del PIB no reflejó bienestar humano, lo que evidencia que la desigualdad persiste y que, a pesar de las reducciones, las cifras se encontraban por encima del promedio en América Latina.

Estudiar a detalle esta contradicción es el objetivo del presente trabajo de grado, en el cual se pretende realizar un análisis estadístico simultáneo del conjunto de variables que miden la pobreza y la generación de valor agregado de los departamentos. Para este estudio se utilizan datos públicos disponibles en la plataforma en línea que dispone el DANE, obteniendo las 12 grandes áreas de la actividad económica más el rubro de impuestos, además los indicadores de pobreza, basados en la metodología del Índice de Necesidades Básicas Insatisfechas (NBI). En vista de que los indicadores de pobreza son calculados por el DANE de forma censal, se toma como referencia el año 2018, como el censo más actual para el desarrollo de este estudio. Para analizar la relación entre estos conjuntos de variables de forma simultánea, se propone aplicar los métodos estadísticos Análisis Factorial Múltiple (AFM) y Regresión por mínimos cuadrados parciales (PLS) e integrar sus resultados. La ventaja de utilizar estos métodos es que facilitan la interpretación de los datos, debido a que establecen las relaciones que pueden existir entre los dos conjuntos de variables

## 2.2. Justificación

En los últimos años la productividad en Colombia ha mostrado un crecimiento en sus diversas actividades económicas. Por ejemplo, en agricultura y ganadería, industrias manufactureras, minería, comercio, entre otras. (Bonilla, J. & Ruiz, J.). Sin embargo, indicadores suministrados por el DANE, informan que para el año 2018, el 27% de la población fue clasificada por debajo de la línea de pobreza nacional. Estos dos ámbitos nacionales se están desarrollando con cierta contrariedad, puesto que, pensado de manera intuitiva se esperaría una disminución en la pobreza cuando hay un crecimiento de la economía.

Bajo la consideración anterior, el propósito de este estudio es estudiar la relación entre la actividad económica y la pobreza, a partir de dos grandes conjuntos de variables provenientes de cada una de las actividades económicas que constituyen el esquema de cuentas nacionales (PIB) y los Indicadores de pobreza (NBI) los cuales se encuentran asociados a cada uno de los 32 departamentos del país.

Es importante resaltar que, desde el punto de vista estadístico, realizar este estudio es un desafío, porque en el análisis de tablas de datos de pobreza y riqueza existirán relaciones entre cada subconjunto de variables y a su vez entre los dos grupos de variables.

Para desarrollar el proyecto, se conformará una matriz de datos que contiene 18 variables (13 Actividades Económicas; 5 indicadores de Pobreza) evaluadas para los 32 departamentos y el distrito capital, 33 unidades de observación. Considerando esta estructura de los datos, para analizar la relación de estas variables de forma simultánea se propone aplicar los métodos estadísticos AFM y PLS, los cuales permitirán caracterizar de manera más precisa, dando paso a nuevos estudios o propuestas para investigación de estos factores sociales.

## 2.3. Objetivos

### 2.3.1. Objetivo principal

Caracterizar la relación entre la actividad económica de los departamentos en Colombia y los Indicadores de pobreza para el año 2018.

### 2.3.2. Objetivos específicos

- Realizar un análisis exploratorio de los indicadores de actividad económica y de pobreza para los departamentos de Colombia en el año 2018.
- Analizar de forma simultánea la relación entre las actividades económicas y la pobreza e integrar los resultados.
- Realizar agrupaciones de departamentos con similares niveles de pobreza y de actividad económica.
- Evaluar la existencias de cambios en las relaciones entre PIB y NBI respecto al año 2005

# **3 Revisión de antecedentes**

Diversos estudios se han desarrollado entorno al crecimiento económico de Colombia y otros se han enfocado en estudiar el aumento de la pobreza, pues estos ámbitos son el eje central cuando se trata de visualizar el desarrollo del país. De acuerdo con Valencia(2018) , “ *un aumento en la economía o en el PIB no representa, ni ahora, ni antes ni en el futuro, un aumento en la calidad de vida de los más pobres, de la mayoría de los colombianos* ”. Este columnista expone que, así incrementen cada una de las fuentes de generación de riqueza no habrá disminución en los niveles de pobreza. Esta afirmación se refleja en las cifras que suministra el DANE, lo cual da paso a diferentes estudios sobre este tema.

El desarrollo de este capítulo se presenta en dos partes, la primera corresponde a investigaciones relacionadas con la situación económica y los niveles de pobreza de los departamentos colombianos y la segunda parte corresponde a investigaciones que aplican las metodologías de AFM y PLS para el análisis de tablas múltiples. Dentro de cada sección, los documentos se presentan en orden cronológico ascendente.

## **3.1. Análisis de la pobreza y la dinámica económica**

**Pérez,G. (2005)**, desarrolla un análisis espacial de la pobreza colombiana, con base en el Índice de Necesidades Básicas Insatisfechas (NBI) y el Índice de Calidad de Vida (QLI), información suministrada por el DANE para los periodos 1985 y 1993. Su objetivo fue evaluar la dependencia espacial en las medidas de pobreza del país, tanto a nivel departamental como municipal. La metodología que utilizó el autor para realizar este análisis fue por medio del cálculo del estadístico I de Moran a nivel global y local, el cual verifica la presencia clústers espaciales de pobreza y detecta procesos de difusión de la pobreza.

Los resultados mostraron que el NBI de los departamentos no se distribuía aleatoriamente, por el contrario, las necesidades insatisfechas de los departamentos dependen de sus condiciones internas y también la situación de sus vecinos cercanos. Además, encontraron clusters en los que se destaca en número aquellos en los que municipios pobres están rodeados de municipios pobres y municipios ricos están rodeados de municipios ricos. Finalmente, el autor concluye que conocer el patrón de distribución espacial de la pobreza le permite al gobernante tomar la decisión de focalizar la inversión de esos recursos en zonas

específicas del territorio.

**Arnedo, V. (2013)**, realizó un análisis del PIB y de los indicadores de desigualdad, nivel de vida y pobreza, con el objetivo de determinar el impacto que genera el crecimiento del PIB sobre el desarrollo humano en Colombia, durante el periodo comprendido entre el 2000 y 2010. Realizó un contraste entre el PIB, el coeficiente Gini (el cual es un método que mide la concentración de ingresos entre los habitantes de una región en un periodo de tiempo determinado), el índice de desarrollo humano (IDH) y la línea de pobreza. Con el fin de estudiar el crecimiento de la economía y determinar qué tan desigual es la distribución de los ingresos generados, el autor realizó un análisis de la relación entre la tasa de crecimiento del PIB y la tasa de crecimiento del Coeficiente de GINI.

Según los datos suministrados por el Banco Mundial, durante el periodo de este estudio, el coeficiente de GINI reflejó una tendencia de inequidad existente en el país, pues alcanza niveles que van desde 0,56 en el año 2005 y 2010, hasta 0,61 en el 2002, a pesar de que se observó un crecimiento de la economía.

La comparación entre la tasa de crecimiento del PIB en Colombia y la tasa de crecimiento del Coeficiente GINI, se observó que entre el año 2000 y 2006 hubo un crecimiento acelerado del PIB. El Coeficiente GINI alcanzó el máximo valor en el año 2002 con 0,61, lo que evidenció la fuerte inequidad que existe en el país y la poca realidad que refleja el PIB per cápita. A su vez al analizar el Índice de Desarrollo Humano ajustado a la Desigualdad (IDH-D6), lo cual reflejó la desigualdad del país con respecto al acceso a los servicios de salud, educación y en niveles de ingresos.

Finalmente, el estudio concluye que el desarrollo de Colombia no se puede medir solo como el aumento del PIB, pues es necesario relacionarlo con el conjunto de variables macroeconómicas y cómo estas influyen en las necesidades básicas. Variables como el nivel de educación, el acceso al servicio de salud y al mercado laboral, son en general las que generan bienestar a la sociedad, pues los resultados de la investigación indicaron que en el país el crecimiento del PIB no reflejó bienestar humano, lo que evidencia que la desigualdad persiste y que, a pesar de las reducciones, las cifras se encontraban por encima del promedio en América Latina.

**Aponte, C., Romero, E. & Santa, L. F. (2015)**, ajustaron un modelo de regresión Beta para estimar el Índice de Necesidades Básicas Insatisfechas en la región Andina de Colombia, las variables involucradas en este estudio fueron el NBI, logaritmo de densidad de población(LOGDP), magnitud de la inversión municipal (MIM), generación de recursos propios(GRPI), logaritmo del consumo promedio de energía por habitante(LOGPENR), tasa de afiliación régimen contributivo(TCONTVO), cobertura en educación (COBEDU)y

deficit de vivienda total (DECVT). Esta región alberga el 67.17% de la población de Colombia, lo que corresponde a casi 16 millones de habitantes según el censo de 2005 (DANE). El estudio se hizo con el objetivo de identificar los lugares donde hacía falta intervención por parte de las administraciones estatales, y sugerir una asignación de recursos de acuerdo con el nivel de pobreza presente en cada municipio.

El análisis espacial permitió identificar la interacción del NBI en cada municipio; usando la prueba de Mantel como medida de dependencia espacial, los autores comprobaron estadísticamente que esta variable en el año 2005 estaba correlacionada espacialmente, por lo que se identificó mayor pobreza en las zonas norte y sur de la Región Andina, siendo los municipios Almaguer-Cauca, Chita-Boyacá, San Calixto-Norte de Santander y La Vega-Cauca los que presentaban el mayor NBI.

Para asociar un modelo al patrón de los departamentos, inicialmente se considera el análisis de las funciones de enlace utilizadas en el ajuste de un modelo de regresión beta econométrico y espacial. Los modelos propuestos fueron evaluados con el parámetro de precisión, el criterio BIC y el  $R^2$ .

Finalmente los autores desarrollaron un modelo mixto de regresión beta regresivo autorregresivo espacial que estima el índice de NBI en la Región Andina. Este modelo presenta una buena bondad de ajuste. Según el modelo, el índice de NBI es explicado por las variables cobertura en educación, déficit de vivienda total, densidad poblacional, tasa de afiliación al régimen contributivo, ingreso público per-cápita, generación de recursos propios y consumo promedio de energía por habitante, las cuales posiblemente pueden explicar el NBI sin la necesidad de elaborar un censo a nivel nacional.

**Bonilla, J. & Ruiz, J. (2018)**, realizaron un análisis multivariante en tres vías (departamentos, sector económico, año), para caracterizar la dinámica económica a partir del PIB generado por las 33 unidades administrativas en las que se divide Colombia, en cada una de las fuentes de generación de riquezas, entre el periodo comprendido entre el 2000-2015. Los datos fueron suministrados por el DANE, y conforman una estructura de tres vías: las unidades administrativas, rubros de la actividad económica y tiempo. Los métodos que utilizaron para el análisis de datos fueron el análisis de componentes principales multivía (Multi Way-PCA) y el método Tucker 3.

Para realizar el análisis, los autores corrigieron las variaciones en el PIB debidas a las diferencias poblacionales entre los departamentos, por lo que se utilizó el PIB per cápita, al igual que se corrigió las variaciones en el PIB debidas al incremento inflacionario natural de la economía. Las matrices de datos para cada año se deflactaron a pesos constantes del año 2005, bajo el valor económico del PIB es comparable a través del tiempo de observación.

El análisis del arreglo en tres vías se realiza al desplegar el arreglo tridimensional tanto por filas como por columnas, con el objetivo de disponer de las diferentes perspectivas de la dinámica económica del país. El despliegue por filas permitió comparar las trayectorias de los departamentos, mientras que el despliegue por columnas permitió observar la ubicación relativa global de los departamentos durante el periodo de análisis. Por otra parte, para implementar el método de Tucker 3, es necesario llevar a cabo un arreglo previo a la estructura de los datos.

Finalmente, los resultados reflejaron un continuo incremento de la economía en el que cada departamento ha resultado favorecido. Por medio de los métodos propuestos detectaron que la economía Colombiana se puede caracterizar a través de tres componentes: como la productividad industrial y de servicios, productividad en explotación de recursos naturales y uso del suelo y por último productividad del sector social y comunitario. También se observó que el descenso de la productividad en el sector minero ha sido cubierto por los departamentos con un incremento en su potencial agrícola-ganadero. Por otro lado, identificaron departamentos como Vichada, Vaupés, Guainía, Putumayo, Guaviare y Amazonas, los cuales se observó con condiciones de productividad tanto bajas y fortaleza en su industria de servicios sociales.

## 3.2. Análisis de Tablas Múltiples

**Pagés, J. & Tenenhaus, M. (2001)**, presentan de manera detallada un ejemplo de análisis de relaciones entre variables fisicoquímicas, perfiles sensoriales y juicios hedónicos, aplicando la técnica de Análisis de factores múltiples (AFM) y PLS path modelling. Los autores resaltan ambos métodos, pues la diferencia que hay entre ambos los hace complementarios. Es decir, la principal diferencia entre estos dos enfoques radica en la existencia de un modelo de causalidad que se tiene en cuenta con el PLS path modelling. Este modelo PLS: busca estructuras comunes, teniendo en cuenta los vínculos de causalidad entre los bloques, mientras que todos los grupos juegan el mismo papel en AFM, finalmente el método AFM es puramente descriptivo. Como resultado de estos puntos comunes y diferencias, los dos métodos son altamente complementarios en la construcción de un enfoque descriptivo y de modelado de un problema, que en el caso que propusieron conduce a grupos de juicios hedónicos que pueden explicarse de manera muy simple mediante variables fisicoquímicas y sensoriales.

**Vitelleschi, M.S. & Chavasa, V. (2015)**, caracterizaron 21 variedades de trigo pan, evaluando ocho variables cuantitativas relacionadas a la calidad y al rendimiento. Para ello aplicaron el método multivariado AFM. Se evaluaron las variables: rendimiento, peso, proteína grano, rendimiento harina, gluten húmedo, Alveograma W, Alveograma (de agua) y volumen de panificación. Los ensayos fueron realizados en campo de productores

de las localidades de Corral de Bustos y Cavanagh, por lo tanto, cada localidad representa un ambiente, en ese sentido la matriz de datos resultante es: variedades, características y ambientes.

Por medio del AFM se identificaron las variedades más afectadas por el ambiente y a su vez facilitó estudiar la sensibilidad a los cambios ambientales. Este método multivariado permitió caracterizar a las variedades de trigo pan, sin que hubiera un grupo de variables más predominante que otro. Los autores lograron obtener una representación superpuesta de las variedades de trigo pan vistas a través de cada ambiente, permitiendo analizar qué variedades estaban más afectadas por el entorno. Finalmente concluyeron que las variables más sensibles a los cambios ambientales resultaron ser la alveograma W, gluten húmedo y volumen de panificación.

**Osorio, M., Rodríguez N., & Salamanca, G. (2017)**, presentan un trabajo en el que evalúan y clasifican variedades de miel de las regiones Lara y Yaracuy, a partir de las características fisicoquímicas tales como cromaticidad, PH, acidez libre, azúcares totales, humedad, cenizas, conductividad eléctrica, densidad y contenido de agua. Además, relacionan los atributos sensoriales y propiedades intrínsecas de las mieles para el proceso de clasificación, por medio de métodos multivariados como el análisis de conglomerados, AFM y PLS-PM.

Con la información generada de las propiedades fisicoquímicas y sensoriales obtenidas en el trabajo de caracterización, realizaron un estudio de las relaciones sensoriales y los parámetros fisicoquímicos, mediante técnicas de PLS, de tal forma proporcionaron una metodología que permitió cuantificar las relaciones de las caracterizaciones.

El análisis y criterio sensorial posibilitaron la comprensión y clasificación de grupos de muestras con atributos semejantes de manera independientemente del origen geográfico, condición que se reafirma a través del análisis discriminante, por conglomerados y redes neuronales. Finalmente se concluyó que la relación entre las respuestas y la regresión, reveló tipos de mieles con 50 % dominante ámbar, 40 % ámbar claro y 10 % ámbar oscuro de naturaleza multifloral. Las propiedades fisicoquímicas y atributos sensoriales son conducentes a la diferenciación en virtud a su origen geográfico

**Visbal, D. Martínez, M. & Escoria, R. (2020)**, realizaron un estudio en el que se realiza una caracterización de las universidades públicas Colombianas, mediante el estudio de los indicadores de resultado establecidos en el Índice de Progreso de la Educación Superior (IPES) desarrollado por el Ministerio de Educación Nacional de Colombia (MEN), haciendo uso del método multivariado AFM, y a su vez propusieron la creación de un índice de rendimiento sintético basado en la misma técnica. De manera que el objetivo era generar

una clasificación de las instituciones y determinar las variables sobre las cuales orientar esfuerzos de mejora en su desempeño.

El análisis comparativo del desempeño de las instituciones de educación superior (IES), sugirieron medidas de mejora para cada universidad individual. Por lo tanto, el AFM es una técnica que aportó un mayor grado de información al análisis clásico de Componentes Principales, permitiendo la agrupación de variables de diferente naturaleza, pues se identificaron variables clave para cada IES. Por otra parte, la aplicación de un proceso de agregación para la creación de un índice de desempeño basado en AFM permite eliminar la arbitrariedad en la estimación de la importancia de cada variable.

A través de los antecedentes recopilados en esta sección, se logra construir un panorama del país, tanto en el aspecto de pobreza y la dinámica económica. Donde se puede concluir que Colombia ha presentado incremento de su economía, pero el desarrollo del país no se puede medir solo como el incremento del PIB, pues es necesario relacionarlo con el conjunto de variables macroeconómicas y como estas influyen en las necesidades básicas, como lo es el acceso a servicios públicos, salud, educación y mercado laboral, las cuales son condiciones que están directamente relacionadas con el bienestar de la sociedad.

Los resultados de la investigación de Arnedo.V. (2013), indicaron que en el país el crecimiento del PIB no reflejó bienestar humano, lo que evidenció que la desigualdad persiste y que, a pesar de las reducciones, las cifras se encontraban por encima del promedio en América Latina.

A su vez, los antecedentes con respecto al Análisis de Tablas Múltiples, se puede concluir que los métodos AFM y PLS son altamente complementarias, por lo tanto, el integrar los resultados facilita la interpretación de los datos, debido a que establecen las relaciones que pueden existir entre los dos conjuntos de variables que se desean estudiar.

# **4 Marco Teórico**

En este capítulo se desarrolla un marco teórico metodológico, en el que se presentan los conceptos necesarios para comprender el desarrollo de este estudio. Este apartado está dividido en dos partes, la primera corresponde al marco conceptual en donde se establecen las definiciones económicas del Producto Interno Bruto (PIB) y el sector económico, junto con sus desagregaciones y además la definición correspondiente al índice de necesidades básicas insatisfechas (NBI) y sus respectivas dimensiones en las que se divide. La segunda parte corresponde al metodológico, donde se describirán las técnicas estadísticas que se van a implementar en este trabajo para el análisis de tablas múltiples como lo son el AFM y el PLS, asimismo una descripción de la técnica multivariante análisis de clúster, que será implementada para realizar agrupaciones de departamentos con similares niveles de pobreza y actividad económica.

## **4.1. Definiciones económicas y de pobreza**

A continuación, se presentan todas las definiciones referentes al Producto Interno Bruto (PIB) y su desagregación en los diferentes sectores de la economía colombiana, y los conceptos con respecto a la pobreza y el índice de necesidades básicas insatisfechas (NBI) con sus respectivas dimensiones.

### **4.1.1. Producto Interno Bruto (PIB)**

El PIB es un indicador macroeconómico que mide el valor de mercado de todos los bienes y servicios finales producidos dentro de un país en un periodo determinado. Normalmente, ese intervalo es de un año o de un trimestre (3 meses). Este indicador, mide el flujo de renta y de gasto de la economía de ese periodo. (Mankiw 2002).

#### **Actividad económica**

Proceso o grupo de operaciones que combinan recursos, tales como equipo, mano de obra, técnicas de fabricación e insumos para la producción de bienes y servicios. (DANE,2012)

#### **PIB nominal**

Expresa el valor monetario de la producción de bienes y servicios de demanda final de un país o de una región durante un periodo determinado de tiempo (normalmente un año) con el nivel de precios existente en cada momento. Por tal razón, se dice PIB a precios corrientes.

## PIB real o deflactado

Es el valor de los bienes y servicios destinados a la demanda final a precios constantes; es decir; elimina la distorsión de las variaciones en los precios (tanto de la inflación como de la deflación), tomando los precios del año que se toma como base. El cálculo de este se obtiene mediante la ecuación 4-1.

$$\text{PIB real} = \frac{\text{PIB Nominal}}{\text{Deflactor del PIB}} \quad (4-1)$$

## PIB per cápita

Es un indicador económico que mide la relación entre el nivel de ingresos de un país o región y su población. Este se calcula través del cociente entre el PIB de la región y el total de personas que la habitan.

### 4.1.2. Sectores económicos

Un sector económico es la división de la actividad económica de un determinado territorio. Esto debido al gran número de trabajos existentes, que de acuerdo a sus características han sido clasificados en grupos. El desarrollo de dichos sectores puede ser mayor o menor dependiendo de las fortalezas productivas de un territorio. El DANE, desagrega la actividad económica en 12 sectores, más el rubro de impuestos como se expone en la Tabla 4-1.

Nº	Actividad económica
1	Agricultura, ganadería, caza, silvicultura y pesca
2	Explotación de minas y canteras
3	Industrias manufactureras
4	Suministro de electricidad, gas, vapor y aire acondicionado; distribución de agua; evacuación y tratamiento de aguas residuales, gestión de desechos y actividades de saneamiento ambiental
5	Construcción
6	Comercio al por mayor y al por menor; reparación de vehículos automotores y motocicletas; transporte y almacenamiento; alojamiento y servicios de comida
7	Información y comunicaciones
8	Actividades financieras y de seguros
9	Actividades inmobiliarias
10	Actividades profesionales, científicas y técnicas; actividades de servicios administrativos y de apoyo
11	Administración pública y defensa; planes de seguridad social de afiliación obligatoria; educación; actividades de atención de la salud humana y de servicios sociales
12	Actividades artísticas, de entretenimiento y recreación y otras actividades de servicios; actividades de los hogares individuales en calidad de empleadores; actividades no diferenciadas de los hogares individuales como productores de bienes y servicios para uso propio
13	Impuestos

**Tabla 4-1:** Sectores económicos de desagregación del PIB

A continuación se presentará una breve definición de cada uno de los sectores económicos según el DANE. Para una descripción mas detallada de cada sector se puede dirigir al documento Clasificación Industrial Internacional Uniforme de las actividades económicas (CIIU Rev.4A.C).

### **Agricultura, ganadería, caza, silvicultura y pesca**

Las actividades económicas pertenecientes a este sector son todas aquellas que se encuentran relacionadas con el cultivo de productos agrícolas y la cría de animales, a su vez comprende categorías de agricultura orgánica, el cultivo de plantas genéticamente modificadas y la cría de animales genéticamente modificados. También abarca actividades asociadas a la explotación de animales o de productos derivados de ellos en actividades agropecuarias, así como también la caza ordinaria; la explotación maderera y recolección de productos forestales silvestres distintos de la madera; actividades de pesca y acuicultura que abarca la utilización de los recursos humanos pesqueros marinos, de agua salobre y agua dulce, con el objetivo de recolectar peces, crustáceos, moluscos y otros organismos y productos marinos.

### **Explotación de minas y canteras**

Se consideran actividades pertenecientes a este sector, aquellas que abarcan la extracción de minerales que se encuentran en la naturaleza en estado sólido como lo es el carbón y minerales, líquido como el petróleo y en estado gaseoso (gas natural), se incluye actividades para la preparación de dichos materiales en bruto para su comercialización, como lo es el triturado, molienda, limpieza, secado, selección y concertación de minerales, licuación de gas natural y aglomeración de combustibles sólidos.

### **Industrias manufactureras**

Son actividades que abarcan la transformación física o química de materiales, sustancias o componentes procedentes de la agricultura, la ganadería, la selvicultura, la pesca y la explotación de minas y canteras, es decir que, la alteración, renovación o reconstrucción de productos son considerados usualmente como actividades manufactureras las cuales provienen de unidades como plantas o fábricas que utilizan maquinaria y equipos de manipulación de materiales que funcionan con electricidad, o también productos provenientes de procesos de manufactura manual o que son realizados en el hogar del trabajador, o que se venden en el mismo lugar en el que fueron fabricados, tales como panaderías y sastrerías. Por lo tanto, son considerados productos o procesos manufactureros aquellos que son acabados o listos para consumo o semiacabados cuando son insumo para otra industria manufacturera.

## **Suministro de electricidad, gas, vapor y aire acondicionado**

Las actividades económicas pertenecientes a este sector son aquellas relacionadas con el suministro de energía eléctrica, gas natural , vapor y agua caliente por medio de una estructura de conductos y tuberías de distribución permanente. También se incluye el suministro de estos servicios en parques industriales o bloques de apartamentos. Las diferentes actividades que conforman los servicios públicos de energía eléctrica y gas combustible ya sea generación, transmisión, distribución y comercialización pueden ser realizados por una misma unidad o por unidades diferentes.

## **Construcción**

Comprende todas las actividades corrientes y especializadas relacionadas con la construcción de edificios y obras de ingeniería civil tales como viviendas, edificios de oficinas o edificios públicos y de servicios; y obras de ingeniería civil como carreteras, calles, puentes, túneles, aeropuertos, proyectos de ordenamiento hídrico, entre otros. Estas actividades se pueden llevar a cabo por cuenta propia a cambio de una retribución o por contrato.

## **Comercio al por mayor y al por menor; reparación de vehículos automotores y motocicletas; transporte y almacenamiento; alojamiento y servicios de comida**

Esta actividad económica incluye las ventas tanto al por mayor como al por menor (venta sin transformación) de cualquier tipo de artículos. El comercio al por mayor hace referencia a la reventa de productos a minoristas, usuarios industriales , comerciales y a otros mayoristas; el comercio al por menor consiste en la reventa de mercancías destinados exclusivamente para consumo o uso personal o doméstico. Se considera la venta sin transformación cuando todas sus operaciones se relacionan con el comercio como por ejemplo la selección, clasificación y montaje de productos.

Por otra parte, este sector incluye el mantenimiento y reparación de vehículos automotores, como por ejemplo reparaciones mecánicas, eléctricas, lavado y encerado, montaje y despinchado de llantas, instalación de partes y accesorios que no hacen parte del proceso de fabricación, entre otros.

En cuanto a las actividades de transporte y almacenamiento hace referencia al conjunto organizado de operaciones encargadas de realizar el traslado de personas u objetos, ya sea de forma conjunta o separada, de un lugar a otro , utilizando uno o varios modos de transporte conforme con las autorizaciones expedidas por las entidades competentes a nivel nacional o internacional. Las actividades de alojamiento y servicios de comida, comprende

el aprovisionamiento de hospedaje por un tiempo determinado y provisión de comidas de consumo inmediato.

### **Información y comunicaciones**

Este sector económico incluye la producción y la distribución de información y productos culturales, el suministro de los medios para transmitir o distribuir esos productos, así como de datos o de comunicaciones, actividades de tecnologías de información y el procesamiento de datos y otras actividades de servicios de información.

Las actividades de edición comprenden la transformación de contenidos sujetos a la adquisición de los derechos de autor sobre contenido (productos de información) y la difusión de estos contenidos entre el público, reproduciéndolos y distribuyéndolos directamente u organizando la reproducción y la distribución de este contenido en varias formas. Se incluyen en esta sección todas las formas viables de edición (en impresión, forma electrónica o de audio, en la internet, en forma de productos multimedia, tales como libros de consulta en CD-ROM).

### **Actividades financieras y de seguros**

Este sector incluye actividades de servicio financiero, a su vez incluyendo actividades de seguros; las actividades de apoyo a los servicios financieros; las actividades de sociedades y de cartera y las actividades de fondos y entidades financieras similares. Las actividades de arrendadores, agentes y/o corredores inmobiliarios en operaciones como venta, compra, administración, alquiler, arrendamiento de inmuebles entre otros servicios inmobiliarios que pueden realizarse con bienes propios o arrendados también se consideran parte de este sector Servicios

### **Actividades inmobiliarias**

Este sector, comprende las actividades de arrendadores, agentes y/o corredores inmobiliarios en una o más de las siguientes operaciones: venta o compra, administración, alquiler y/o arrendamiento de inmuebles y la prestación de otros servicios inmobiliarios, tales como la valuación y las consultorías inmobiliarias. Las actividades incluidas en esta sección pueden realizarse con bienes propios o arrendados o a cambio de una retribución o por contrata.

### **Actividades profesionales, científicas y técnicas; actividades de servicios administrativos y de apoyo**

Este sector abarca las actividades profesionales, científicas y técnicas especializadas. Estas actividades requieren un alto nivel de capacitación y ponen a disposición de los usuarios

conocimientos y aptitudes especializados.

Dichas actividades pueden ser por ejemplo de tipo jurídico y de contabilidad, como en la representación de los intereses de la partes sea o no antes tribunales u otros órganos judiciales, realizadas por abogados o bajo la supervisión de los mismos. También se incluyen actividades de administración empresarial en las que se brinda asesoría y asistencia a empresas y otras entidades en temas de gestión empresarial. Prestación de servicios de arquitectura, servicios de ingeniería.

Por otra parte se encuentran actividades de investigación científica y desarrollo las cuales se dividen en tres tipos como la investigación básica, la aplicada y el desarrollo experimental; la primera hace referencia a investigaciones experimentales o teóricas encaminadas a adquirir nuevos conocimientos sobre fenómenos y hechos observables; el segundo tipo hace referencia a la adquisición de conocimientos dirigidos hacia objetivos prácticos y el tercer tipo es basado en conocimientos obtenidos mediante investigaciones dirigidas a producir nuevos materiales, productos y mejorar procesos.

### **Administración pública y defensa; planes de seguridad social de afiliación obligatoria; educación; actividades de atención de la salud humana y de servicios sociales**

Este sector comprende las actividades que están a cargo de la administración pública, entre las que se cuentan las actividades legislativas, ejecutivas y judiciales; actividades tributarias, de defensa nacional, de orden público y seguridad; y las relaciones exteriores y la administración de programas gubernamentales. Se incluyen también las actividades relacionadas con planes de seguridad social de afiliación obligatoria. El factor determinante para clasificar una actividad en esta sección no es la condición jurídica o institucional de las unidades que la realizan, sino la propia naturaleza de la actividad.

Por otra parte, este sector comprende la organización y la prestación de la educación pública o privada en los niveles de primera infancia, preescolar, básica (primaria y bachillerato), media, superior, para el trabajo y el desarrollo humano; dirigido a niños y jóvenes en edad escolar o a adultos, a grupos vulnerables y diversos.

También se considera dentro de este sector la prestación de servicios de atención de la salud humana y de asistencia social que entre sus actividades se destaca los servicios de atención de la salud prestados por profesionales en hospitales y otras entidades; y la asistencia social sin participación de profesionales de la salud, pasando por actividades de atención en instituciones con un componente importante de atención de la salud.

## **Actividades artísticas, de entretenimiento y recreación y otras actividades de servicios; actividades de los hogares individuales en calidad de empleadores; actividades no diferenciadas de los hogares individuales como productores de bienes y servicios para uso propio**

Esta sección comprende una amplia gama de actividades de interés cultural, de entretenimiento y recreación para el público en general, como: la producción y la promoción de actuaciones en directo, espectáculos en vivo, exposiciones, funcionamiento de museos y lugares históricos, juegos de azar y actividades deportivas y recreativas. Con respecto a otras actividades se incluye el de asociaciones, reparaciones de computadores efectos personales y enseres domésticos y una variedad de servicios personales, que no son abarcados en otros sectores.

Las actividades de los hogares como empleadores de personal doméstico, tales como: empleadas domésticas, cocineros, camareros, mayordomos, lavaderos, jardineros, porteros, conductores, vigilantes, institutrices, niñeras, instructores, profesores particulares, secretarias, entre otros.

Por otra parte se tienen las actividades de los hogares como productores de bienes y servicios. Los hogares se clasifican en esta división solo si es imposible identificar una actividad principal entre las actividades que realizan para su propia subsistencia.

## **Impuestos**

Entre las principales fuentes que conforman el PIB, está la recaudación de impuestos (Mankiw 2014), que conforma todas las actividades de prestación de dinero o en especie que exige el Estado para la financiación de sus gastos y para atender requerimientos de interés social. Cabe aclarar que Impuesto no es un sector económico, pues este no agrega valor a la economía, pero se es incluido en el estudio porque teóricamente puede tener relación con un mejoramiento en las condiciones de vida.

### **4.1.3. Pobreza**

Según la Organización de Naciones Unidas (ONU), define la pobreza como la condición caracterizada por una privación severa de necesidades básicas.

La pobreza incluye, entre estas condiciones, el acceso a alimentos, agua potable, instalaciones sanitarias, salud, vivienda, educación, así como la información. De carecer de estas condiciones, la ONU considera dicha condición como estado de pobreza. En otras palabras, toda aquella persona que no tenga acceso a unas condiciones mínimas que permitan el desarrollo de una vida básica, pero digna.

La pobreza puede identificarse de manera directa o indirecta. La primera hace referencia a entrar a evaluar directamente si los hogares han logrado satisfacer sus necesidades básicas, encuestándolos sobre los bienes y servicios que disponen, es decir, que se encuentra relacionado con el bienestar que tiene un individuo respecto a ciertas condiciones que se consideran vitales como lo es la salud, la educación, empleo, entre otras. La segunda forma consiste en medir los recursos del hogar, por medio de sus ingresos y sus gastos, para así estimar si estos son suficientes para que el hogar pueda tener un buen nivel de vida, es decir, que esta relacionado con la capacidad de adquisición de bienes y servicios que tienen los hogares.

Un método para medir la pobreza indirecta es el de Pobreza Monetaria. Esta cifra se compone, a partir de la medición del ingreso per cápita corriente de la unidad de gasto, es decir, el ingreso corriente de la unidad de gasto dividido por el total de integrantes de la misma; luego es comparado con el costo monetario de adquirir una canasta de alimentos, esto en el caso de la pobreza monetaria extrema o indigencia, o con el costo de adquirir una canasta de alimentos y no alimentos mínimos para la subsistencia, en el caso de la pobreza monetaria. Este método no es medido para todos los departamentos de Colombia, no incluye a 7 de ellos como lo son Arauca, Casanare, Vichada, San Andrés, Amazonas, Guainía, Guaviare, Vaupes y Vichada.

Cabe resaltar que existen diferentes métodos directos para medir la pobreza, como lo es el Índice de Pobreza Multidimensional(IPM). Este esta compuesto por 5 dimensiones: condiciones educativas del hogar, condiciones de la niñez y juventud, salud, trabajo, acceso a servicios públicos domiciliarios y condiciones de la vivienda.Los datos reportados por el DANE para este método están clasificados por regiones:Caribe, Oriental, Central, Pacifica(sin incluir al valle), Bogotá, Antioquia y Valle del Cauca.

El método directo más distinguido y utilizado en América Latina es el de las Necesidades Básicas Insatisfechas (NBI), este fue introducido por la Comisión Económica para América Latina y el Caribe (CEPAL), a comienzos de los años ochenta con el objetivo de aprovechar la información de los censos, demográficos y de vivienda, en la caracterización del fenómeno de la pobreza. Debido a que este trabajo está enfocado en caracterizar la relación entre la actividad económica de los “departamentos” en Colombia y los Indicadores de pobreza, el DANE suministra los datos del NBI para los 32 departamentos de Colombia más Bogotá, por tal motivo, se toma este método para el desarrollo de este estudio.

#### **4.1.4. Necesidades Básicas Insatisfechas (NBI)**

“La metodología de NBI busca determinar, con ayuda de algunos indicadores simples, si las necesidades básicas de la población se encuentran cubiertas. Los grupos que no alcancen un umbral mínimo fijado, son clasificados como pobres. Los indicadores simples seleccionados, son: Viviendas inadecuadas, hogares con hacinamiento crítico, viviendas con servicios inadecuados, hogares con alta dependencia económica y hogares con niños en edad escolar que no asisten a la escuela.” (DANE).

Para poder realizar el calculo del NBI, inicialmente se asignan puntajes específicos a cada posibilidad de satisfacción de una necesidad. Es decir, para cada necesidad se fija un indicador de logro, donde los valores presentan la forma mas extrema de carencia hasta la forma más adecuada de satisfacer una necesidad. Seguidamente se establece un límite o “norma mínima”, que indica la mínima calificación a partir de la cual se considera apropiado los satisfactores para cada necesidad.

Después de establecer la “norma mínima”, se construye un indicador de privación, que presenta la distancia entre el indicador de logro y la norma fijada, es decir:

$$P_{ij} = \frac{X_i^* - X_{ij}}{X_i^*} \quad (4-2)$$

donde  $P_{ij}$  es el indicador de la privación  $i$  para el hogar  $j$ ,  $X_{ij}$  es el indicador de logro respecto de la variable  $i$  para el hogar  $j$ , y  $X_i^*$  es la norma mínima. De esta manera, los hogares que se encuentren con carencias les corresponderá un valor positivo en el indicador de privación, los hogares que estén en la norma les corresponderá un valor de cero, y aquellos hogares que estén sobre la norma recibirán un valor negativo.

A continuación, se presenta la definición de cada una de las componentes del NBI según el DANE.

#### **Viviendas inadecuadas**

Viviendas móviles, o ubicadas en refugios naturales o bajo puentes, o sin paredes o con paredes de tela o de materiales de desecho o con pisos de tierra (en zona rural el piso de tierra debe estar asociado a paredes de material semipermanente o perecedero).

#### **Servicios inadecuados**

- o Zona Urbana: Carencia de servicios sanitario o carencia de acueducto y aprovisionamiento de agua de río, nacimiento, carro tanque o de lluvia.

- o Zona rural: Carencia de servicios sanitario y de acueducto que se aprovisionan de agua de río, nacimiento o lluvia.

## **Hacinamiento crítico**

Más de tres personas por cuarto (incluyendo en estos todas las habitaciones con excepción de cocinas, baños y garajes).

## **Inasistencia escolar**

Hogares en donde uno o más niños entre 7 y 11 años de edad, parientes de la persona a cargo del hogar, no asisten a un centro de educación formal.

## **Alta dependencia económica**

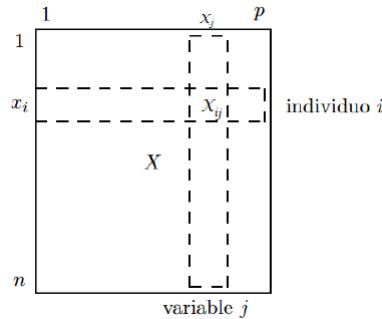
Hogares con más de tres personas por miembro ocupado y cuyo jefe ha aprobado, como máximo, dos años de educación primaria.

## **4.2. Análisis de Tablas Múltiples**

A continuación, se presentan todas las definiciones referentes a las técnicas estadísticas que se aplicaran en el presente trabajo. Inicialmente se presenta el concepto del método Análisis de Componentes Principales (ACP), que aunque no es el método utilizado en este estudio, es la base fundamental de los métodos multivariados, es decir, que sus conceptos pueden ser extendidos a otras técnicas de análisis de datos multivariado.

### **4.2.1. Análisis de Componente Principales (ACP)**

El Análisis de Componentes Principales (ACP), presentado por primera vez por Karl Pearson en 1901 e integrado a la estadística matemática por Harold Hotelling en 1933. Este método busca interpretar o comprender las relaciones entre variables e individuos en una matriz de datos  $X_{n.p}$ . Esta técnica supone que no hay factores comunes, por lo tanto, el objetivo se centra en simplificar la estructura de los datos transformando las variables en unos pocos componentes principales que sean combinaciones lineales de las variables. Es decir, el ACP trata de transformar un conjunto de variables interrelacionadas en un conjunto de variables no correlacionadas llamadas factores, analizando la estructura de dependencia y correlación que existe entre las variables, y así especificando como estas explican parte de la información que contienen los factores.(Diaz,2018).



**Figura 4-1:** Matriz de datos  $X_{n,p}$

La Figura 4-1 representa la estructura de la Matriz  $X$ , donde cada elemento  $x_{ij}$  es la medida de la variable  $j$  para el individuo  $i$ . El espacio  $\mathbb{R}^p$  es el espacio asociado a los individuos y  $\mathbb{R}^n$  es el espacio asociado a las variables.

Antes de aplicar el método del ACP se debe hacer un preprocessamiento de los datos. Es decir, en vista de que las variables se pueden expresar en diferentes unidades de medida, generalmente se realiza la estandarización (centrar respecto a la media y escalar por la desviación estándar) de la matriz de datos. De esta manera se evita que los efectos de escala afecten el aporte de cada variable a la varianza total, pues aquellas variables con mayor varianza van a influir más en la conformación de los factores. Por lo tanto, la estandarización de la matriz  $X$  se hace por columna de la siguiente manera:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sqrt{s_{xj}}}$$

Donde  $\bar{x}_j$  y  $s_{xj}$  son la media y la varianza de la variable  $x_j$ . Es decir:

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \quad s_{xj} = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$$

Con esta transformación de la matriz de datos, se garantiza que la media y la varianza de los datos  $z_{ij}$  son iguales a 0 y 1 respectivamente. De esta forma los datos quedan a una misma escala. Se denota entonces  $Z$  a la matriz de puntuaciones  $z$ , la cual representa los datos transformados.

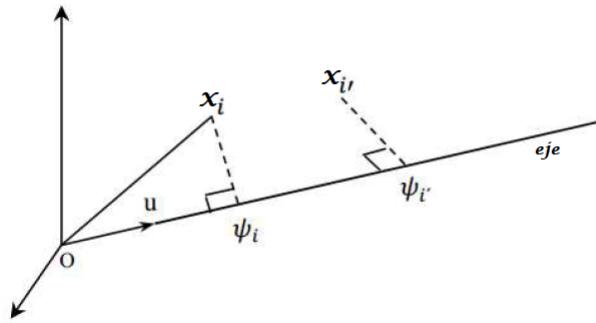
### Análisis de la nube de individuos- $\mathbb{R}^p$

Dado que en el espacio de la representación original de los individuos  $\mathbb{R}^p$ , la distancia euclíadiana entre parejas de individuos esta dada por:

$$d^2(i, i') = \sum_{j=1}^p m_j (x_{ij} - x_{i'j})^2$$

donde, los coeficientes  $m_j$  ponderan la influencia de cada columna  $j$ , de esta manera se conforma la matriz diagonal  $M$ , conocida como métrica.

Se pretende buscar un subespacio vectorial de una dimensión, es decir, una recta que pasando por el origen, realice el mejor ajuste posible de la nube de puntos.



**Figura 4-2:** Proximidad entre los individuos -  $\mathbb{R}^p$

Por lo tanto, es necesario definir el vector director unitario  $u$  en el eje  $\alpha(u_\alpha)$ , en el que se maximiza la suma de cuadrados de las distancias entre todas las parejas de puntos  $(i, i')$  ( $\max_u = \sum_i^n d^2(i, i')$ ). Esto resulta equivalente a maximizar sobre  $u$  la suma de cuadrados de las distancias entre cada punto y el centro de gravedad: ( $\max_u = \sum_i^n d^2(i, G)$ ).

Dicho lo anterior, el análisis de los individuos se realiza una traslación y rotación alrededor del origen, donde se diagonaliza la matriz de correlaciones  $Z'NZ$ , donde  $N = \text{diag}(1/n)$  es la matriz de los pesos  $p_i$  asociados a los individuos y  $Z$  la matriz de datos estandarizados. Las coordenadas de los individuos asociados al espacio  $\mathbb{R}^p$  está dado por:

$$\psi_\alpha = Zu_\alpha \quad (4-3)$$

Es decir, que las coordenadas de los  $N$  individuos proyectados sobre el eje canónico  $u_\alpha$  son los componentes del factor descrito en la ecuación 4-3, lo cual es una combinación lineal de las variables iniciales  $Z_j$ .

$$\psi_1 = u_{11}Z_1 + u_{12}Z_2 + \dots + u_{1p}Z_p \quad (\text{Primera Componente})$$

$$\psi_2 = u_{21}Z_1 + u_{22}Z_2 + \dots + u_{2p}Z_p \quad (\text{Segunda Componente})$$

...

$$\psi_p = u_{p1}Z_1 + u_{p2}Z_2 + \dots + u_{pp}Z_p \quad (P - \text{ésima Componente})$$

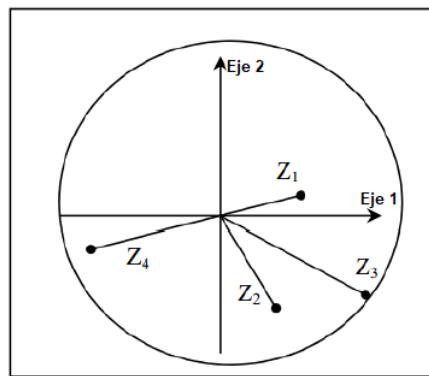
Por lo cual, cada elemento  $u_{ij}$  es la ponderación (peso) que tiene la variable  $Z_j$  sobre la componente  $i$ .

Debido a que la nube de individuos tuvo una translación y rotación alrededor del origen, se tiene que:

1.  $E(\psi_\alpha) = 0$
2.  $V(\psi_\alpha) = \psi'_\alpha N \psi_\alpha = \lambda_\alpha$ , esto bajo  $u'_\alpha u_\alpha = 1$ , donde  $\lambda_\alpha$  es el valor propio asociado a la inercia de la nube de individuos proyectada sobre el eje  $\alpha$ . Lo cual indica que cada componente explica una proporción de la variabilidad total.
3.  $cov(\psi_i, \psi_j) = 0, \forall i \neq j$
4. Si la matriz  $Z'NZ$  es simétrica los vectores propios  $u$  son ortogonales
5. Los valores propios son decrecientes ( $\lambda_1 > \lambda_2 > \dots > \lambda_p$ ), el valor propio más grande ( $\lambda_1$ ) representa la varianza más grande en la nube de datos. Por tanto, el vector propio asociado a ese valor propio da la dirección a lo largo de la cual puede observar esa variación.

### Análisis de la nube de variables- $\mathbb{R}^n$

En el espacio  $\mathbb{R}^n$  el análisis de los puntos variables se hace con referencia al origen 0, con lo cual están a una distancia 1 del origen. Es decir, que todas las variables se encuentran en una hiperesfera de radio 1 centrada en el origen. Por lo anterior, el interés se enfoca en encontrar un plano factorial que proporcione la máxima información de las correlaciones entre variable llamado círculo de correlaciones.



**Figura 4-3:** Círculo de correlaciones

De acuerdo con Lebart et al. (1995), la distancia entre variables se calcula de la siguiente manera:

$$d^2(j, j') = \sum_{j=1}^n (z_{ij} - z_{ij'})^2$$

es decir,

$$d^2(j, j') = \sum_{j=1}^n z_{ij}^2 + \sum_{j=1}^n z_{ij'}^2 - 2 \sum_{j=1}^n z_{ij} z_{ij'}$$

Dado que las variables están estandarizadas y teniendo en cuenta la definición previa sobre la varianza se tiene entonces que:

$$\sum_{j=1}^n z_{ij}^2 = \sum_{j=1}^n z_{ij'}^2 = 1 \quad \text{e igualmente} \quad \sum_{j=1}^n z_{ij} z_{ij'} = c_{jj'}$$

Donde  $c_{jj'}$  es el coeficiente de correlación entre las dos variables. Por lo tanto, la distancia entre dos variables esta dada por:

$$d^2(j, j') = 2(1 - c_{jj'})$$

esto implica que,

$$0 \leq d^2(j, j') \leq 4$$

En el coseno del ángulo de dos vectores variable en el espacio  $\mathbb{R}^n$ , es el coeficiente de correlación entre esas dos variables ( $c_{jj'} = \cos(jj')$ ), como dichas variables están a una distancia de 1 del origen, el coseno del ángulo es el producto escalar entre los vectores.

Se tiene entonces, que dos variables fuertemente correlacionadas estan muy próximas una de la otra ( $c_{jj'} = 1$ ) o por el contrario tan alejadas como sea posible ( $c_{jj'} = -1$ ). Dos variables ortogonales ( $c_{jj'} = 0$ ) están a una distancia media.

En forma semejante a  $\mathbb{R}^p$  se trata de maximizar la información contenida en la proyección, es decir, el cuadrado de la longitud del vector  $X'v$ . El vector  $v$  esta dado por  $v_\alpha = \frac{1}{\sqrt{\lambda_\alpha}} X u_\alpha$  y por las relaciones de transición permiten relacionar el espacio de los individuos con el de las variables, por lo que a través de ellas se obtienen las coordenadas de las  $p$  variables que están dada por:

$$\varphi_\alpha = \sqrt{\lambda_\alpha} u_\alpha$$

## Selección del número de componentes

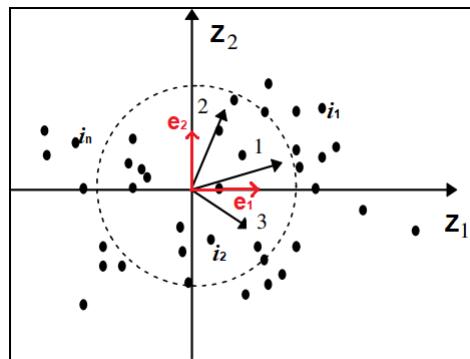
Dado que con el objetivo del ACP es simplificar la estructura de los datos transformando las  $J$  variables en un espacio de dimensión  $q$ , es decir, sustituir las  $J$  variables originales por  $q$  nuevas variables que resuman óptimamente la información. Para decidir el número adecuado de componentes que permita condensar de mejor manera la estructura de datos se sugieren algunos criterios para determinar el número  $q$  de componentes a retener, dentro de los cuales se destacan los siguientes:

- Criterio de Kaiser: De acuerdo con Kaiser(1958), este consiste en eliminar todas las componentes que sus valores sean menores a la unidad. Según este criterio, no tiene sentido agregar una componente que explique menos varianza que la contenida en una variable.
- Contraste de caída: Propuesto por Cattel(1965), este consiste en graficar los valores propios de forma descendiente, el cual se le conoce como “gráfico de sedimentación”. Se seleccionan aquellas componentes en que la curva converge en una linea horizontal, para lo cual se identifica el punto de inflexión en la curva, punto en el cual la varianza explicada por las componentes no aporta mucho.
- Criterio del porcentaje acumulado: Consiste en tomar los valores propios que acumulen un porcentaje de varianza explicada mínimo aceptable (por ejemplo entre el 70 y 80 %).

## Representación simultanea

Se debe tener en cuenta que en el ACP las dos nubes(individuos y variables) están definidas en espacios diferentes, con orígenes y bases distintas. Para la nube de individuos , el origen esta dado por la gravedad de los individuos, que es de dimensión  $p$  y se tiene a  $u_\alpha$  como la correspondiente base hallada. Por otra parte, para la nube de variables de variables se tiene que el origen de coordenadas es de dimensión  $n$  y los ejes factoriales están dados por  $v_\alpha$ .

Es posible representar las direcciones definidas por cada variable activa sobre la base de los ejes factoriales  $u_\alpha$ . Es decir, representar en  $\mathbb{R}^p$  las direcciones dadas por las variables de origen sobre el plano factorial de la nube de individuos.



**Figura 4-4:** Ejemplo representación simultanea-Biplot

La representación entre los individuos y las variables puede apreciarse en la Figura 4-4. La proyección del antiguo eje canónico en  $\mathbb{R}^p$ ,  $e_j$  sobre el nuevo eje  $u_\alpha$ , cuya coordenada vale  $e' u_\alpha = u\alpha$ .

De acuerdo con Lebart et al. (2000), la dirección de una variable define zonas para los individuos: de un lado, aquellos que toman valores fuertes para esta variable y, al lado opuesto, aquellos que toman valores bajos.

## Elementos para la interpretación

Los ejes factoriales permiten obtener la mejor visualización aproximada (en el sentido de los mínimos cuadrados) de las distancias entre los individuos de una parte y entre las variables de otra parte. Para interpretar los factores es necesario apreciar correctamente esta aproximación.(Lebart et al.2000)

- El cálculo de la contribución de las variables está dado por:

$$CTR(j, \alpha) = \frac{\varphi_{j,\alpha}^2}{\lambda_j} = u_{j\alpha}^2$$

- Para detectar los individuos que más contribuyen a la formación de los ejes factoriales, se calcula la contribución de la siguiente manera:

$$CTR(i, \alpha) = \frac{m_i \psi_{i\alpha}^2}{\lambda_\alpha} \quad \text{donde} \quad m_i = \frac{1}{n}$$

- Teniendo en cuenta que el objetivo es tener una idea de que tan próximo está un punto respecto al plano factorial de proyección. La proximidad al plano factorial de proyección se mide mediante cosenos cuadrados de cada punto en los ejes factoriales.

\* Individuos:

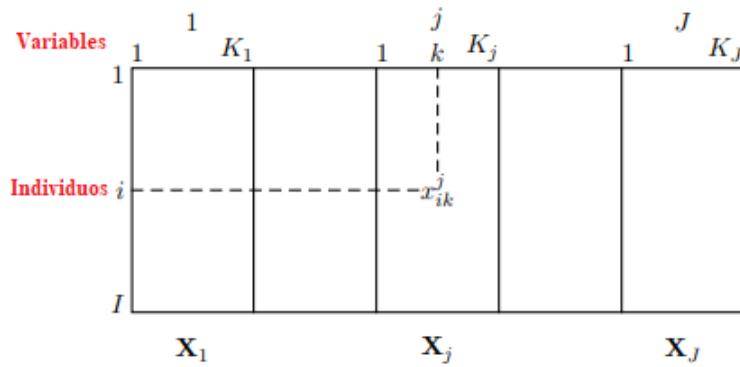
$$\cos^2(i, \alpha) = \frac{\psi_{i\alpha}^2}{d^2(i, G)}$$

\* Variables:

$$\cos^2(j, \alpha) = \varphi_{j\alpha}^2 = cor^2(j, \alpha)$$

### 4.2.2. Análisis Factorial Múltiple (AFM)

El AFM es un método desarrollado por B. Escofier y J. Pagés (1992), que permite el análisis simultáneo de varios grupos de variables medidas sobre el mismo conjunto de individuos equilibrando la influencia de cada grupo. Los grupos de variables pueden ser diferentes, incluso en la naturaleza y número de variables que los componen. La única condición es que las variables que integran un grupo sean de la misma naturaleza, cuantitativa o cualitativa. Un ejemplo de un arreglo de datos con grupos de diferente naturaleza se presenta en la Figura 4-5, donde las matrices  $X_1, \dots, X_j, \dots, X_J$  forman subgrupos de variables, cada uno conformado por distinto número y tipo de variables.



**Figura 4-5:** Representación de la estructura de la tabla de datos en AFM

Se debe tener en cuenta que, los métodos tradicionales, como el Análisis en Componentes Principales (ACP), Análisis Factorial de Correspondencias (AFC), y Análisis de Correspondencias Múltiples (ACM), sólo permiten estudiar las relaciones entre las variables de la misma naturaleza, todas cuantitativas o todas cualitativas, lo que restringe su utilización. Sin embargo, el AFM proporciona una visión más enriquecedora del fenómeno a estudiar, pues se presentan casos en los que se dispone de gran cantidad de información que es interesante relacionar, y por medio de este método es posible realizar relaciones entre variables de diferente naturaleza.

El objetivo del AFM es, como en cualquier análisis factorial, poner de manifiesto los principales factores de variabilidad de los individuos, pero en este caso, estando estos últimos descritos, de manera equilibrada, por los diversos grupos de variables. Es decir el AFM, es un análisis en componentes principales (ACP) ponderado, el cual busca representar una nube de puntos en un espacio de menor dimensión, donde se pueda conservar de la mejor manera posible la distancia entre individuos.

Como se mencionó anteriormente el AFM se basa en la metodología de un ACP, de manera que este se desarrolla en dos etapas:

1. Se realiza un análisis por separado con cada grupo de variables en donde se obtienen los factores parciales. Es decir, que se ejecuta un ACP (o un ACM), esto dependiendo al tipo de variable con el que está conformado el grupo) a cada una de las tablas  $X_j$ , donde se contiene los valores de las variables del grupo  $j$ , donde  $j = 1, 2, \dots, J$ , en el conjunto de individuos.
2. Se lleva a cabo un ACP de la tabla global, la cual es resultado de yuxtaponer las  $J$  tablas  $X_j$ . En este análisis cada tabla  $X_j$  es ponderada por el inverso del primer valor propio obtenido en el ACP (o ACM) de la propia tabla,  $1/\lambda_j$ , lo cual estos valores

conforman la matriz diagonal  $M$ . Esta ponderación mantiene la estructura de cada tabla puesto que todas las variables han recibido la misma ponderación, pero de esta forma logra equilibrar la influencia de los grupos, puesto que la inercia máxima de cada una de las nubes de individuos, definida por los distintos grupos, vale 1 en cualquier dirección. (Abascal Calvo, 2002)

De acuerdo con Escofier Pagés(1992), el AFM considera tres espacios de análisis de los datos:

- $R^K$ : Espacio de la nube de individuos
- $R^I$ : Espacio de la nube de variables
- $R^{I^2}$ : Espacio de la nube de grupos de variables

Se tiene entonces que el conjunto inicial de matrices múltiples aparece entonces como una única tabla estructurada en subtablas. Es decir, a cada grupo de variables le corresponde una tabla, de tal forma que cada tabla puede ser yuxtapuesta y formar así una única tabla que cruza individuos y variables. Siguiendo la estructura presentada en la Figura 4-5 se denotará entonces a  $X$  como la tabla completa (matriz de datos general),  $I$  al número de individuos,  $K$  es el número de variables (todos los grupos combinados),  $J$  es el número de conjuntos o grupos de variables,  $K_j$  es el número de variables del  $j$ -ésimo grupo de variables y  $X_j$  es la tabla asociada al grupo  $j$ .

Inicialmente se obtiene el peso de cada variable del grupo  $j$ , el cual se divide por  $\lambda_j^1$  (este representa el primer valor propio del ACP o ACM separado del grupo  $j$ ). El objetivo de esta ponderación es equilibrar el papel de los grupos en todos los aspectos del análisis. Seguidamente se realiza un ACP a la tabla general (estandarizada y ponderada), donde se obtienen los valores y vectores propios que dan origen al análisis general donde la inercia de cada factor está formado proporcionalmente por las variables activas de cada grupo.

### **Representación de los grupos de variables**

En el estudio de varios grupos de variables, uno de los objetivos es comparar de forma global los grupos. Para ello, se entra a hablar sobre el espacio  $R^{I^2}$ , puesto que, base de esta representación puede contribuir a la interpretación del ACP de la tabla completa  $X$ , pero teniendo en cuenta que esta tiene su propia optimización.

Para poder estudiar los grupos de variables, se construye una nube de puntos a la cual se le denotará como  $N_j$  y este estará conformado por las distancias entre los grupos de variables, donde se derivan del producto escalar  $Lg(j, l) = \langle W_j D, W_l D \rangle = \text{traza}(W_j D W_l D)$ , donde  $W_j D = X_j M_j X'_j D$ , siendo  $X_j$  la matriz de datos asociada al grupo  $j$  y  $D$  es la

matriz diagonal de los pesos de los individuos. Es entonces  $Lg$  un coeficiente que mide la dimensionalidad (número de factores de inercia considerables) de cada grupo. Estos valores toman el valor de 0 cuando no existe relación entre los grupos y no tienen límite superior.

A partir del producto escalar el cual se interpreta como una medida de relación entre los grupos, se desarrolla el coeficiente de correlación vectorial RV, el cual se obtiene a partir de los coeficientes de correlación lineal entre dos variables cualesquiera y su valor está comprendido entre 0 (no existe relación entre las variables de los dos grupos considerados) y 1 (las nubes que representan a los grupos son homotéticas). Entonces para dos grupos cualesquiera  $j$  y  $l$  se tiene :

$$RV = \frac{\langle W_j D, W_l D \rangle}{\|W_j D\| \|W_l D\|} = \frac{\text{traza}(W_j D W_l D)}{\|W_j D\| \|W_l D\|}$$

donde,  $W_j D = X_j X'_j D$ , siendo  $X_j$  la matriz asociada al grupo  $j$  de variables y  $D$  la matriz diagonal de los pesos de los individuos. La norma de  $W_j D$  se expresa como  $\|W_j D\| = \sqrt{\sum_s (\lambda_s^j)^2}$ , donde  $\lambda_s^j$  es el  $j$ -ésimo valor propio del análisis de la tabla  $j$ . El numerador representa los productos escalares entre los representantes de los grupos.

Las coordenadas con valores altos en los grupos, indican que el factor se encuentra asociado a las variables de los grupos, de manera que cuando las coordenadas tienen valores bajos indican poca relación del factor con las variables de los grupos.

### Representación de las variables

La representación se obtiene directamente del ACP de la tabla completa  $X$  la cual se ha ponderado previamente. Se utilizan las coordenadas de las proyecciones sobre los ejes factoriales. Estas coordenadas se pueden calcular, por medio de las relaciones de transición del ACP, por lo tanto el cálculo está dado por :

$$\varphi_s = \sqrt{\lambda_s} u_s; \quad s = 1, 2, \dots, K$$

De acuerdo con Escofier y Pagés (1992), las componentes principales maximizan la inercia de las proyecciones de todas las variables. La inercia proyectada por cada nube  $N_k^j$  puede ser interpretada como la contribución que hace cada grupo. Dada la ponderación que se hace por  $1/\lambda_j$  se equilibra la influencia en el sentido de la contribución de un grupo a la construcción de un eje se encuentra acotada por 1. Por ello se tiene entonces que, ningún grupo puede, por si solo, determinar el primer eje (excepto si se presenta una situación de simetría excepcional), por otra parte, entre más dimensionalidad tenga un grupo así mismo influye en los ejes.

### Representación de los individuos

Sea  $u_s$  el eje de inercia de orden  $s$  de la nube de individuos  $N_I$  la cual se encuentra asociada a la tabla  $X$  en el espacio  $R^k$ . Este se deduce de la componente principal  $F$ , por la relación  $U_s = (1/\sqrt{\lambda_s})X'DF_s$ , donde  $\lambda_s$  es el valor propio de  $WD$  asociada a  $F_s(W = \sum_j)W_j$ . La proyección de  $N_I^j$  sobre  $u_s$  de manera que:

$$F_j^j = \bar{X}_j Mu_s = (1/\sqrt{\lambda_s})\bar{X}'_j MX'DF_s = 1/\sqrt{\lambda_s}W_j DF_s$$

### Representación superpuesta de las $J$ nubes definidas por cada grupo de variables

Para la representación simultanea en el AFM , se debe tener en cuenta que las  $J$  nubes  $N_I^j$  están situadas en el espacio  $R^K$  y por ello se debe tener en cuenta ciertas consideraciones:

- La nube  $N_I^j$  está en  $R^{K_j}$ , subespacio de  $R^K$
- $R^K = \bigoplus_j R^{K_j}$ , es la suma directa de los subespacios ortogonales dos a dos.
- $M_j$  es la métrica en  $R^{K_j}$ , submatriz con las  $K_j$  filas y las  $k_j$  columnas de  $M$
- Las coordenadas de los puntos de  $N_I^j$  son las filas de  $X_j$ , entonces las coordenadas de estos puntos en  $R^k$  son las filas de la matriz  $\bar{X}_j$ , la cual se encuentra definida como  $\bar{X}_j = [0..,0X_j,0..,0]$
- $X = \sum_j \bar{X}_j$ , con el objetivo de hacer coincidir  $F_s(i)$  con el punto medio de los  $J$  puntos parciales  $F_s^j(s)$ , las filas de  $\bar{X}_j$  se proyectan como puntos suplementarios ampliados por  $J$ .
- La unión de las  $J$  nubes parciales forman la nube  $N_I^j$
- La inercia de la nube  $N_I^j$  se puede expresar como *InerciaIntra + InerciaEntre* subnubes  $N_i^J$ . La inercia de la nube media  $N_I$  del AFM corresponde a las *InerciaEntre*, cuando las coordenadas parciales se aplican por  $J$ .

### Elementos para la interpretación del AFM

Para identificar el concepto implícito que representan los ejes factoriales del AFM y comprender las relaciones entre grupos, variables e individuos, se calcula las contribuciones de cada uno de estos a los ejes factoriales.

- El cálculo de la contribución de las variables a una dimensión está dado por:

$$CTR_{k,s} = \frac{\varphi_s^2}{\lambda_J^1}$$

La contribución de la K-ésima variable en la S-ésima dimensión, teniendo en cuenta que las variables de un mismo grupo comparten el mismo peso ( $1/\lambda_j^1$ )

- El calculo de la contribución de las individuos a una dimensión está dado por:

$$CTR_{i,s} = \frac{m_i(F_s^j)^2}{\lambda_s}$$

Donde  $m_i$  es el peso del individuo y  $F_s^j$  la coordenada del i-ésimo individuo en el s-ésimo eje factorial.

- El calculo de la contribución de un grupo de variables a una dimensión está definida como la sumatoria de las contribuciones del J-ésimo grupo de variables en el S-ésimo eje factorial, es decir:

$$CTR_{j,s} = \sum_{k=1}^{K_j} CTR_{k,s}$$

### **Elementos suplementarios**

Los elementos suplementarios intervienen de manera posterior del análisis activo para enriquecer y caracterizar aún más los ejes. Estos elementos al estar situados en los planos factoriales activos tienen que pasar por las mismas transformaciones es decir estar estandarizados.

Se pueden presentar situaciones, en los que hay individuos que aportan mucho a la inercia, éstos individuos no permiten visualizar las relaciones presentes en las variables y el resto de individuos, por lo tanto, dentro del análisis activo este tipo de individuos puede considerarse como datos atípicos. De tal forma, para visualizar de mejor manera las relaciones, se toman estos individuos como suplementarios. Para ello, se introducen los centros de gravedad de estas clases, estos no influyen sobre las representaciones de los individuos activos: se calcula su proyección en la nube  $N_I^*$  y en las diferentes nubes  $N_I^j$ .

### **4.2.3. Método de Mínimos Cuadrados Parciales (PLS)**

El estadístico Herman Wold introdujo el método de regresión PLS aproximadamente en el año 1975, para el tratamiento de cadenas de matrices y aplicaciones en econometría. Este es un método de modelado de datos multivariantes, versátil para el análisis de múltiples relaciones entre uno o más conjuntos de variables medidas en algunos objetos. La idea básica del PLS es la de reducción de la dimensión en regresión múltiple, con la garantía de que las primeras componentes ortogonales mejoran la predicción.

Se ha demostrado que este método es una buena alternativa a los métodos más antiguos de

regresión lineal porque es “robusto”. Es decir, el modelo matemático no es muy alterado cuando se toman en cuenta nuevas muestras. La suposición básica de la regresión PLS es que el sistema depende de un número pequeño de variables instrumentales llamadas variables latentes. Este concepto es similar al de componentes principales. Las variables latentes son estimadas como combinaciones lineales de las variables observadas. En los modelos PLS, se establece una representación de la matriz X en término de dichas variables latentes.

De acuerdo con Peña (2016), la aplicación de este método presenta unos factores importantes que motivan su utilización:

- El modelo de regresión PLS es un potente método de regresión lineal , el cual tiene como propiedad considerar la multicolinealidad en las variables explicativas y acepta gran número de variables.
- El modelo resultante, predice las respuestas a partir de un conjunto de variables que se encuentran linealmente dependientes  $x_1, x_2, \dots, x_n$
- En el desarrollo del modelo se obtienen unas componentes PLS que son utilizadas en la regresión.
- La complejidad del modelo es determinada por el número de componentes PLS y este puede ser optimizado para mejorar el rendimiento de la predicción.

Cuando se emplea el método de PLS con una única variable respuesta  $y$ , se desarrolla de la siguiente manera:

1. Se realiza el cálculo de la primera componente PLS como la variable latente que tiene la máxima covarianza entre los scores y la respuesta  $y$
2. Seguidamente, la información de la componente que fue calculada es eliminada de las variables explicativas. Esto es una proyección del espacio de las variables respuesta sobre un hiperplano el cual es ortogonal a la dirección de la componente encontrada. Por otra parte la matriz residual resultante del proceso tiene el mismo número de variables que la matriz original formada por las variables explicativas, pero la dimensionalidad intrínseca se reduce a uno.
3. A partir de la matriz residual resultante del proceso, se calcula la siguiente componente.
4. El proceso se sigue haciendo hasta que en el modelamiento de la respuesta  $y$  se observe un mejoramiento.

En el método de PLS con una sola variable respuesta recibe el nombre de PLS1. Por otro lado, cuando hay más de una variable respuesta la regresión PLS recibe el nombre de PLS2. Para el desarrollo de este estudio que se tienen dos conjuntos de datos, se utilizará el método de PLS2, por tal motivo, se hará contexto de como se desarrolla este método.

### **Algoritmo PLS2**

Se aplica el método de PLS2, cuando se tiene que explicar un conjunto de variables  $y_1, y_2, \dots, y_q$ , mediante un conjunto de variables explicativas  $x_1, x_2, \dots, x_p$ . Por lo tanto, este algoritmo consiste en aplicar el método de los componentes principales de un conjunto de variables  $x_1, x_2, \dots, x_p$ , bajo la condición de que estas componentes principales sean también lo más explicativas del conjunto de variables  $y_1, y_2, \dots, y_q$ . (Peña,2016)

Con la aplicación de PLS2 se logran dos objetivos:

1. Se consigue maximizar la correlación entre los conjuntos de variables  $x_1, x_2, \dots, x_p$  y  $y_1, y_2, \dots, y_q$ .
2. Por medio del hiperplano, aproximar lo mejor posible los espacios vectoriales generados por los conjuntos de variables  $x_1, x_2, \dots, x_p$  y  $y_1, y_2, \dots, y_q$ .

Para desarrollar el algoritmo PLS2, se construyen las matrices  $X$  e  $Y$ , donde  $X$  tiene columnas los vectores  $x_i$  para  $i = 1, 2, \dots, p$  y donde  $Y$  tiene columnas los vectores  $y_k$  para  $k = 1, 2, \dots, q$ . Cada vector  $x_i$  y cada vector  $y_k$  pertenece al espacio vectorial  $\mathbb{R}^n$  y las matriz  $X$  pertenece al espacio vectorial  $\mathbb{R}_{n \times p}$  y  $Y$  al  $\mathbb{R}_{n \times q}$ . Este algoritmo consiste en hacer proyecciones simultáneas en ambos espacios sobre hiperplanos de dimensiones reducidas. De esta forma , las coordenadas de los puntos en esos hiperplanos forman los elementos de las matrices  $T$  y  $U$ .

Por lo anterior, se presenta el PLS2 por etapas, de esta forma se da una visión general de como se desarrolla este algoritmo. Para iniciar se considera el conjunto de variables  $(x_1, x_2, \dots, x_p)$  y el conjunto de variables respuesta  $(y_1, y_2, \dots, y_q)$ , seguidamente a partir de estos conjuntos se construyen sus respectivas matrices  $X \in \mathbb{R}_{n \times p}$  y  $Y \in \mathbb{R}_{n \times q}$ .

1. Se construyen las matrices  $X_0$  y  $Y_0$  que están conformadas por las variables centradas y reducidas de las variables predictoras y respuesta.
2. Después se construye una combinación lineal  $u_1$  de las columnas  $Y_0$  y una combinación lineal  $t_1$  de las columnas de  $X_0$  de tal modo que se maximice la  $Cov(u_1, t_1)$ . Así se tienen dos nuevas variables  $u_1$  y  $t_1$  que están lo más correlacionadas posible y que a su vez están resumiendo lo mejor que se puede las información que contiene  $X_0$  y  $Y_0$ .
3. **Determinación de las componentes**

Para construir la primera componente, se busca que esta sea combinación lineal de  $X_0$  denominada  $t_1$  y otra componente que sea combinación lineal de las columnas de  $Y_0$ , denominada  $u_1$ , por lo tanto se tiene que :

$$t_1 = X_0 w_1, \quad u_1 = Y_0 c_1$$

Donde  $w_1$  y  $c_1$  tiene una norma euclíadiana 1. De las dos combinaciones lineales que se construyeron se debe obtener una maximización de la  $\text{cov}(t_1, u_1)$ .

Luego se pasa a construir la regresión lineal simple del conjunto de variables explicativas y el conjunto de variables a explicar sobre la componente  $t_1$ , es decir, que queda de la siguiente manera:

$$X_0 = t_1 p_1^T + X_1$$

$$Y_0 = t_1 q_1^T + Y_1$$

Donde,

$$p_1 = \frac{X_0^T t_1}{t_1^T t_1} \quad q_1 = \frac{Y_0^T t_1}{t_1^T t_1}$$

Entonces,  $p_1$  es el vector de los coeficientes de regresión sobre  $t_1$  para cada variable original independiente de  $x_i$  y  $q_1$  es el vector de coeficientes de regresión de  $t_1$  para la variable original dependiente  $y_k$ .

#### 4. Determinación de la segunda componente

La etapa 3 se repite, pero esta vez reemplazando las matrices  $X_0$  y  $Y_0$ , por las nuevas matrices residuales  $X_1$  y  $Y_1$ ; así se pasa a obtener dos nuevas componentes  $t_2$  y  $u_2$  las cuales maximizan la  $\text{Cov}(u_2, t_2)$ . Por lo tanto, a partir de estas componentes se obtiene por regresión lineal simple:

$$X_1 = t_2 P_2^T + X_2$$

$$Y_1 = t_2 q_2^T + Y_2$$

Por consiguiente se deduce que:

$$X_0 = t_1 P_1^T + t_2 P_2^T + X_2$$

$$Y_0 = t_1 q_1^T + t_2 q_2^T + Y_2$$

De tal forma estas etapas se repiten hasta que las componentes  $t_1, t_2, \dots, t_h$  expliquen suficientemente a  $Y_0$ . De esta forma se tendría entonces:

$$Y_0 = t_1 q_1^T + t_2 q_2^T + \dots + t_h q_h^T + Y_h$$

- **Propiedades de las componentes**

1.  $p_h^T w_h = 1$
2.  $q_h = b_h c_h$
3.  $t_h^T X_h$

El objetivo de la regresión PLS2 es maximizar la covarianza entre los datos x-sxore y los y-scores y como el problema de maximización no es unico , es necesario realizar una restriccción en los vectores scores, por lo tanto,  $\|t\| = \|u\| = 1$ .

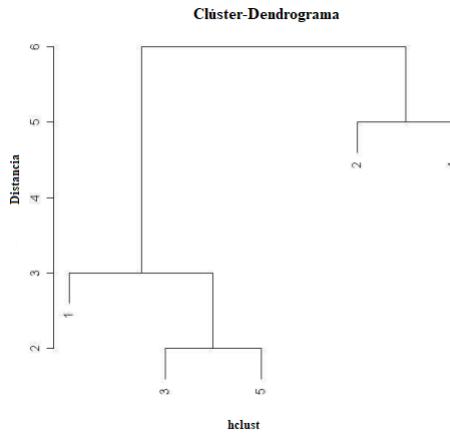
### 4.3. Análisis de Clúster (AC)

El análisis de conglomerados, también conocido como analisis de clúster, es una técnica multivariante cuya idea básica es clasificar objetos formando grupos (clúster) que sean lo más homogéneos posible dentro de sí mismos y heterogéneos entre sí. Surge ante la necesidad de diseñar una estrategia que permita definir grupos de objetos homogéneos. Este agrupamiento se basa en la idea de distancia o similitud entre las observaciones. La obtención de dichos clústers depende del criterio o distancia considerados, es decir, que el número de clústers depende de lo que consideremos como similar.

Los algoritmos para la formación de clúster, se agrupan en dos categorías:

- **Algoritmo de partición:** Este método consiste en dividir el conjunto de de observaciones en  $k$  grupos (clúster), donde  $k$  es definido inicialmente por el usuario o este es determinado por algún criterio de convergencia.
- **Algoritmos jerárquicos:** Este tipo de métodos tiene como fin agrupar clúster para formar uno nuevo o separar alguno ya existente para dar origen a otros dos, y así sucesivamente se va desarrollando el proceso de agrupamiento, donde se minimice o se maximice alguna distancia o medida de similitud. Los métodos jerárquicos se dividen en aglomerativos y disociativos.
  1. Método asociativos o aglomerativos: Se inicia el análisis con tantos grupos como individuos haya. De manera que, a partir de esas unidades iniciales se van formando agrupaciones , de manera ascendente, hasta que al final del proceso todos los casos están en un mismo grupo o conglomerado.
  2. Métodos disociativos: A diferencia del método anterior, este inicia con todos los casos del análisis en un solo conglomerado, y a partir de ese grupo inicial, en pasos sucesivos se van dividiendo y así se van formando grupos más pequeños.Por lo tanto, al final del proceso se tienen tantas agrupaciones como casos tratados.

En cualquiera de los casos, de ambos métodos se deriva un dendrograma, el cual es un gráfico que ilustra como se van desarrollando las subdivisiones o agrupamientos (como se ilustra en la Figura 4-6). Los individuos similares se conectan mediante enlaces cuya posición en el diagrama está determinada por el nivel de similitud/disimilitud entre los individuos.



**Figura 4-6:** Representación de un dendrograma

### Distancias entre conglomerados

De acuerdo con De la Fuente (2011), hay diferentes maneras de medir la distancia entre clúster que producen diferentes agrupaciones y a su vez diferentes dendrogramas. Exactamente, no existe un criterio para elegir cual de los algoritmos es mejor, pues la decisión es arbitraria y depende más del método que refleje mejor los propósitos específicos de cada estudio.

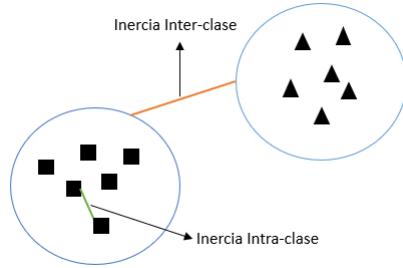
El objetivo es entonces, formar grupos de manera que la distancia entre ellos sea grande (Inercia Inter-clases) y que la distancia entre los individuos dentro de cada grupo sea pequeña (Inercia Intra-clases). Es decir formar grupos (clúster) que sean lo más homogéneos posible dentro de sí mismos y heterogéneos entre sí.

- Método del vecino más cercano: El objetivo es agrupar los casos que se encuentren a menor distancia. Se denota como

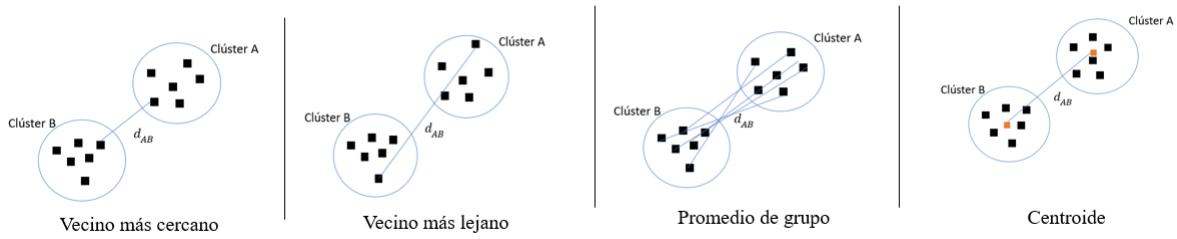
$$d(A, B) = \min d(i, j); i \in A, j \in B$$

- Método del vecino más lejano: En este método, se procede a agrupar los casos que se encuentren a mayor distancia. Se denota como

$$d(A, B) = \max d(i, j); i \in A, j \in B$$



**Figura 4-7:** Representación de las inercias Intra-Clases e Inter-clases



**Figura 4-8:** Métodos de distancias entre clúster

- Promedio de grupo: Se denota como

$$d(A, B) = \frac{1}{n_A n_B} \sum d(i, j); i \in A, j \in B$$

- Centroide: Con este método la distancia que hay entre dos grupos esta dada por sus centros de gravedad (centroides). Por lo tanto, para iniciar su ejecución se debe calcular el centro de gravedad de cada conglomerado para después agrupar aquellos conglomerados cuya distancia sea mínima.

$$d(A, B) = d(\bar{x}_A, \bar{x}_B)$$

- Método de Ward: También conocido como perdida de la inercia mínima. En este método se enfoca en unir los casos buscando minimizar la varianza dentro de cada grupo. Se inicia calculando la media de todas las variables en cada conglomerado, después se realiza el calculo de la distancia entre cada caso y la media del conglomerado, sumando después las distancias entre todos los casos. Finalmente se agrupan los clúster que generan menos aumento en la suma de las distancias dentro de cada grupo. Esté método crea grupos homogéneos y con tamaños similares.

El análisis de cluster se puede desarrollar por etapas:

- La primera es la selección de variables (cualitativas o cuantitativas).
- Seguidamente, la selección de la medida de asociación, esta puede ser de distancia o una de similitud. Si es por distancia (por ejemplo, distancia euclídea), los grupos formados tendrán individuos parecidos, por lo que sus distancias entre ellos será pequeña. Si se selecciona una medida de similaridad (como el coeficiente de correlación), los grupos que se formen tendrán individuos con similaridad alta.
- Luego, se escoge la técnica de cluster a aplicar (vecinos más cercanos, centroide, método ward)
- Finalmente se validan los resultados

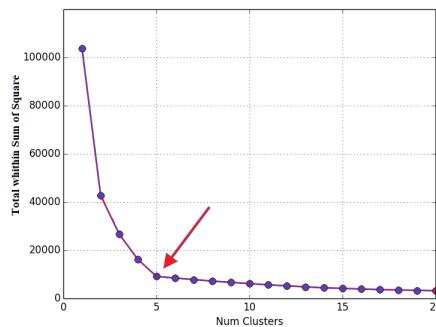
### Selección del número óptimo de clúster

En el momento de aplicar algún método de clúster se presentan un problema y es el de elegir un número óptimo de clúster. A pesar de que no existe un criterio objetivo para la selección del número de clúster, se cuenta con diferentes métodos de ayudan a elegir un apropiado número de grupos. Para este estudio se eligió el método del codo.

### Método del codo

Este método emplea los valores de la inercia obtenidos al aplicar K-means a diferentes números de clúster. Donde la inercia es la suma de las distancias al cuadrado de individuo u objeto del clúster a su centroide.

$$\text{Inercia} = \sum_{i=0}^N ||x_i - \mu||^2$$



**Figura 4-9:** Representación del método del codo

Después de obtener los valores de kmeans de 1 a N clústeres, como se puede observar en la Figura 4-9 , se presenta en una gráfica lineal la inercia respecto al número de clúster. Donde en ella se puede apreciar un cambio brusco en la evolución de la inercia , donde la linea hace un quiebre en forma de brazo o de codo. Por lo tanto el punto donde se observa ese cambio, es el que indicara el número óptimo de clúster.

# 5 Metodología

En este capítulo se presenta la metodología propuesta para dar cumplimiento a los objetivos planteados de este trabajo, para analizar la relación entre la actividad económica de los departamentos en Colombia y los Indicadores de pobreza para el año 2018. Inicialmente, se da una descripción de los datos, se definen los individuos, las variables y las condiciones del estudio, a su vez se brinda detalle de la fuente de los datos. Luego, se describe el análisis estadístico para este estudio, el cual se divide en dos fases, la primera, es un análisis exploratorio de los datos y la segunda es la implementación de los métodos multivariantes, AFM y PLS, seguidamente de un análisis de clúster y una comparación de resultados con respecto al año 2005.

## 5.1. Descripción de los datos

En la estructura para la conformación de la matriz de datos, los individuos corresponden a las 33 unidades administrativas en las que se divide Colombia. Esto es, 32 departamentos y el distrito capital (Ver Figura 5-1), estas serán las unidades de observación.



Figura 5-1: Mapa político de Colombia

Las variables están agrupadas en dos grandes conjuntos; el primero comprende a los 12 sectores productivos del PIB más el rubro por la recaudación de impuestos, y el segundo conjunto son los indicadores de pobreza que se encuentran segregados en 5 dimensiones, los cuales corresponden al porcentaje de la población que para cada departamento presenta Necesidades Básicas Insatisfechas (NBI). La estructura de estos dos conjuntos se expone en la Tablas **5-1** y **5-2**.

<b>Abreviatura</b>	<b>Actividad económica</b>
Agri	Agricultura, ganadería, caza, silvicultura y pesca
Min	Explotación de minas y canteras
Manu	Industrias manufactureras
Elec	Suministro de electricidad, gas y agua
Cons	Construcción
Come	Comercio ,reparación , restaurantes y alojamiento
Comu	Información y comunicaciones
Fina	Actividades financieras y de seguros
Inmo	Actividades inmobiliarias
Prof	Actividades profesionales, científicas y técnicas; actividades de servicios administrativos y de apoyo
Serv.Soc	Actividades de servicios sociales, comunales y personales
Recre	Actividades de entretenimiento y recreación
Impu	Impuestos

**Tabla 5-1:** Sectores productivos de desagregación del PIB en Colombia

Se debe tener en cuenta que, debido a las diferencias del tamaño de la población entre los departamentos, se hace un ajuste al PIB a través del cociente sobre el número de habitantes, obteniendo con ello el PIB per cápita para los departamentos. Por otra parte, el conjunto de variables que corresponde a los indicadores de pobreza son calculados a través del censo poblacional y está conformado como se expone en la Tabla **5-2**.

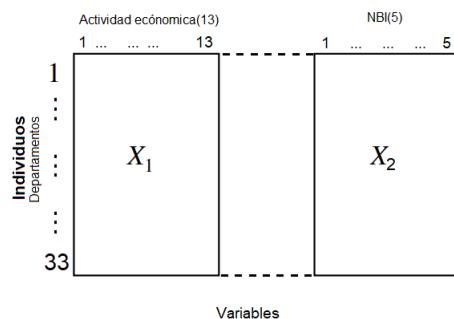
En vista de que los indicadores de pobreza son obtenidos a través del censo poblacional y las cuentas nacionales se obtienen de forma anual, se considera el año 2018 (año del último censo) como periodo de referencia para el análisis, ya que en este periodo se dispone de información de ambos conjuntos de datos.

Abreviatura	Indicador
Vivienda	Componente Vivienda
Servicios	Componente Servicios
Hacinamiento	Componente Hacinamiento
Inasistencia	Componente Inasistencia Escolar
Depen.economica	Componente Dependencia Económica

**Tabla 5-2:** Indicadores de Necesidades Básicas Insatisfechas

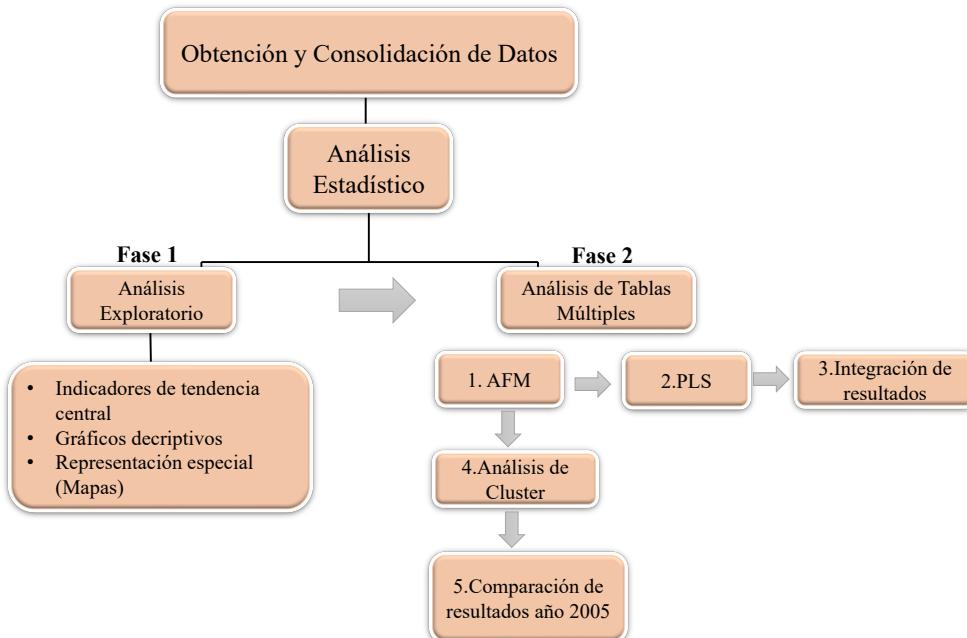
La información para la construcción de la matriz de datos para este estudio proviene de la plataforma web oficial del DANE en Colombia (<http://www.dane.gov.co/index.php/estadisticas-por-tema>), la cual brinda información de los departamentos para el año de interés. El conjunto de variables de las 12 actividades económicas y el rubro impuestos, se encuentra en la sección Cuentas Nacionales en el ítem PIB departamental y se descarga el archivo «Resultados por actividad económica». Así mismo, la página del DANE dispone de los 5 indicadores de pobreza a nivel departamental, ubicados en la sección «Pobreza y condiciones de vida »en el ítem Necesidades básicas insatisfechas (NBI), se descarga el archivo denominado «Indicador de necesidades Básicas Insatisfechas(NBI) censo nacional de población y vivienda 2018».

Estos datos se pueden organizar en un arreglo de dos tablas, tal como se presenta en la Figura 5-2, donde la matriz  $X_1$  está compuesta por todas las variables que miden la actividad económica de las 33 unidades administrativas (Departamentos) de Colombia y la matriz  $X_2$  la componen las variables del índice de necesidades básicas insatisfechas (NBI). El análisis simultáneo de estos datos, en estadística recibe el nombre de análisis de múltiples tablas que es la estrategia metodológica que se seguirá en este proyecto de grado.

**Figura 5-2:** Estructura de los datos

## 5.2. Análisis Estadístico

El análisis estadístico para este estudio está dividido en dos fases, la primera corresponde al análisis exploratorio de los datos y la segunda al análisis de tablas múltiples que se divide en 5 ítems, los cuales son: aplicar las técnicas multivariadas AFM y PLS, luego se realiza una integración de resultados de las dos técnicas. Del método AFM se desprende el cuarto y quinto ítem correspondientes al análisis de cluster y a la comparación de los resultados con respecto al año 2005. A continuación se presenta en la Figura 5-3 las fases de desarrollo de este trabajo.



**Figura 5-3:** Diseño metodológico propuesto

### 5.2.1. Análisis Exploratorio

En esta fase se realiza un análisis exploratorio en cada subconjunto de datos, en el cual se estudia la variabilidad del conjunto, a través de indicadores de tendencia central, de dispersión y de asociación (promedios, varianzas, medianas, coeficientes de correlación), gráficos descriptivos (Boxplot), esto con la intención de identificar casos atípicos, posibles errores y valores faltantes que se puedan presentar en el conjunto de datos. También se realiza un análisis de la correlación entre el conjunto de variables NBI y el PIB con el objetivo de visualizar las relaciones de los dos conjuntos de datos. Para las diferencias entre departamentos se harán representaciones espaciales por medio de mapas y así, evaluar la fortaleza productiva de un determinado sector económico dependiendo a la región en la que

se ubique, al igual que se tendrán mapas para visualizar los indicadores de pobreza y su distribución por regiones.

### 5.2.2. Análisis de Tablas Múltiples

Después de realizar los respectivos análisis exploratorios de datos a cada una de las tablas, el interés se orienta estudiar el nivel de asociación entre las variables de actividad económica y condiciones de pobreza en los departamentos. Para ello, se realizará la fusión de las dos tablas en una sola tabla conjunta, es decir una matriz de 33 individuos (departamentos y distrito capital) con 18 variables (13 indicadores del PIB y 5 del NBI), sobre la cual se realizará el análisis a través de las técnicas AFM y PLS.

#### Análisis Factorial Múltiple (AFM)

Este método permite el análisis simultáneo de varios grupos de variables medidas sobre el mismo conjunto de individuos, equilibrando la influencia de cada grupo. Por ello, para aplicar este método se realizará la conformación de una tabla de datos, la cual está compuesta por los dos conjuntos antes mencionados, es decir, el valor agregado de cada una de las áreas de la actividad económica y los indicadores de pobreza, las cuales estarán ponderadas por el inverso de la raíz cuadrada del primer valor propio de cada tabla de datos. A continuación se presenta en la Tabla 5-3 el resumen de la estructura de los datos para el desarrollo del AFM.

Individuos	Grupos de variables	Variables activas		Individuos Suplementarios
		1.PIB	2.NBI	
32	2	13	5	1

**Tabla 5-3:** Resumen de la estructura de los datos para el desarrollo del AFM

Como se mencionó en el apartado 5.1, se cuenta con 32 departamentos más el distrito capital (Bogotá), por lo que la nube de individuos estaría conformada por 33 individuos. Resultados previos como los identificaron Bonilla y Ruiz, muestran que Bogotá es el centro de desarrollo económico de la nación, con un muy alto aporte al PIB, lo cual convierte en un registro muy atípico, comparado con el resto de departamentos, por lo tanto, no permitiría visualizar de manera adecuada a los demás individuos y variables. Por tal motivo se toma la decisión de dejar este individuo como suplementario en el análisis del AFM.

El AFM se realiza con el paquete FactoMineR y factoextra del software estadístico R a través de la función MFA. Se debe tener en cuenta que en vista de que las variables se pueden expresar en diferentes unidades de medida, generalmente se realiza la estandarización (centrar respecto a la media y escalar por la desviación estándar) de la matriz de datos. De

esta manera se evita que los efectos de escala afecten el aporte de cada variable a la varianza total, pues aquellas variables con mayor varianza van a influir más en la conformación de los factores. Por lo anterior, entre los argumentos la función MFA se especifica que cada conjunto de datos debe ser estandarizado.

Para una mejor interpretación de los datos, se apoyará el análisis por medio del círculo de correlaciones, el cual proporcionará la máxima información de las correlaciones entre variables; la nube de individuos facilitará visualizar la contribución y calidad de representación de los departamentos sobre el plano factorial. Además por medio de la representación simultánea se analizará la dirección que tomen las variables así se definirán las zonas para los individuos, es decir, que a un lado estarán aquellos que tomen valores fuertes para la variables y al lado opuesto aquellos individuos que tomen valores bajos.

Aplicando este método, se expondrá los principales factores de variabilidad de los departamentos, estando estos últimos descritos, de manera equilibrada, por los diversos grupos de variables. Se estudiará las relaciones entre los grupos, medir su grado de semejanza, relaciones entre las variables del grupo de actividad económica y las del grupo de indicadores de pobreza, a su vez visualizar semejanzas entre los departamentos a través de los diferentes grupos de variables.

## **PLS**

Complementariamente se empleará el método de PLS, el cual generaliza y combina características del método de Análisis de Componentes Principales y Análisis de Regresión Múltiple. Esta técnica se utiliza cuando se tiene una estructura de múltiples variables predictoras y una o múltiples variables de respuestas de datos. Además, una de las bondades que ofrece este método es la reducción de dimensionalidad de manera que facilita la interpretación del comportamiento del conjunto de variables.

Para este estudio, se tiene dos conjuntos de datos, por lo tanto, se construyen las respectivas matrices, una de ellas será aquella que contenga las variables predictoras que para este caso se seleccionó al conjunto del PIB y para la matriz de variables de respuesta se tomó al conjunto del NBI.

Ya teniendo identificado la matriz de datos predictora y de respuesta, se realiza el análisis PLS con el paquete plsdepot del software estadístico R a través de la función plsreg2, la cual realiza una regresión de mínimos cuadrados parciales para el caso multivariante, es decir, más de una variable de respuesta. Los datos son estandarizados. Para mejor interpretación del análisis se utilizará el círculo de correlaciones y la nube de individuos, además de ver la importancia de las variables a la proyección de los datos.

### **Integración de resultados**

Cada uno de los métodos aplicados para el análisis de tablas múltiples arroja resultados que dan un panorama del comportamiento de los datos y sus relaciones. Por lo tanto, el objetivo de este punto es integrar cualitativamente los hallazgos de cada técnica y de esta forma fortalecer el análisis del estudio. Es decir, a partir de los rasgos generales que arroje cada método se destacará los puntos más relevantes.

#### **5.2.3. Análisis Clúster**

Se plantea la agrupación de departamentos que desde las perspectivas de la actividad económica y de las condiciones de pobrezas, permitan identificar subgrupos con similares condiciones, los cuales a futuro puedan recibir tratamiento de política administrativa y de inversión pública similar.

Para realizar el análisis clúster, se utilizan las proyecciones globales y parciales de los departamentos sobre los primeros componentes del AFM construido en la anterior fase. Para la retención de ejes factoriales se utilizó el criterio de porcentaje de varianza acumulada, donde se obtuvo que los dos primeros ejes factoriales retienen el 67.3% de la inercia contenida en los datos. Es decir, que con la información resumida en las coordenadas de las dos primeras componentes del AFM se construye la matriz de datos para poder realizar el análisis de clúster. Cabe resaltar que el AFM arroja unas coordenadas globales y a su vez se tiene el análisis factorial de cada conjunto es decir coordenadas para el PIB y para el NBI, por lo tanto, el análisis de cluster se realizará con las coordenadas globales y parciales de los conjuntos.

El análisis de clúster sobre las tres nuevas matrices, se realiza con el paquete factoextra del software estadístico R. Inicialmente se realiza un análisis del número óptimo de clúster, para cada conjunto de datos, por medio de la técnica del codo, la cual se ejecuta un gráfico que visualiza el valor adecuado del número de grupos, esto a través de la función fviz\_nbclust. Cabe recordar que este método emplea los valores de la inercia obtenidos al aplicar K-means a diferentes números de clúster. Donde la inercia es la suma de las distancias al cuadrado de individuo u objeto del clúster a su centroide. Dicho lo anterior, se especifica en la función el argumento "wss" que hace referencia a el total dentro de la suma del cuadrado. Seguidamente se ejecuta la visualización de los grupos a través de un dendograma, por medio de la función hclust, la cual se le especifica construir el dendrograma con el método ward, pues este método se enfoca en unir los casos buscando minimizar la varianza dentro de cada grupo. Se presenta el dendrograma global, y los parciales (PIB y NBI). Finalmente los clúster formados se visualizan en mapas para poder observar como se distribuyen esos grupos de forma espacial y así tener un panorama más claro sobre los grupos conformados.

#### **5.2.4. Comparación de resultados con periodo 2005**

Finalmente, con intención de verificar cambios en el tiempo, se realizará el AFM con datos del periodo 2005 (censo anterior), de tal manera poder comparar estos resultados con el AFM del periodo 2018. En este punto se presenta para ambos periodos la gráfica de varianza explicada, el círculos de correlación y mapas del análisis espacial de los clusters que se generaron en los dos conjuntos de variables(es decir, que para construir los mapas, se aplica el análisis de clúster a los datos del año 2005). Esto con el objetivo de visualizar que tanto ha cambiado el comportamiento de las variables e individuos en las dos estancias de tiempo. Es decir, entre los dos periodos a comparar hay una diferencia de 13 años por lo que la situación de productividad y pobreza del país podrían ser diferentes, por tanto contrastar los resultados de la situación más actual (periodo 2018) con respecto al periodo anterior, podría dar un panorama de que indicadores de pobreza han disminuido o aumentado, como ha sido la evolución de cada departamento, que sectores económicos se han fortalecido y como se presenta la relación de ambos conjuntos de datos en ambos periodos de tiempo.

# 6 Resultados

En este capítulo se presenta en dos fases el análisis estadístico, la primera corresponde al análisis exploratorio de los datos, en la que se presentan los resultados descriptivos de los dos conjuntos de variables, PIB y NBI para cada uno de los departamentos en Colombia para el año 2018. En los que se visualizarán medidas de tendencia central y se presentarán mapas para observar el comportamiento de forma espacial de cada componente de pobreza y de actividad económica en las unidades administrativas del país. La segunda fase corresponde al análisis de tablas múltiples a través de las técnicas multivariadas AFM y PLS, seguidamente de un análisis de clúster y la comparación de resultados con respecto al año 2005.

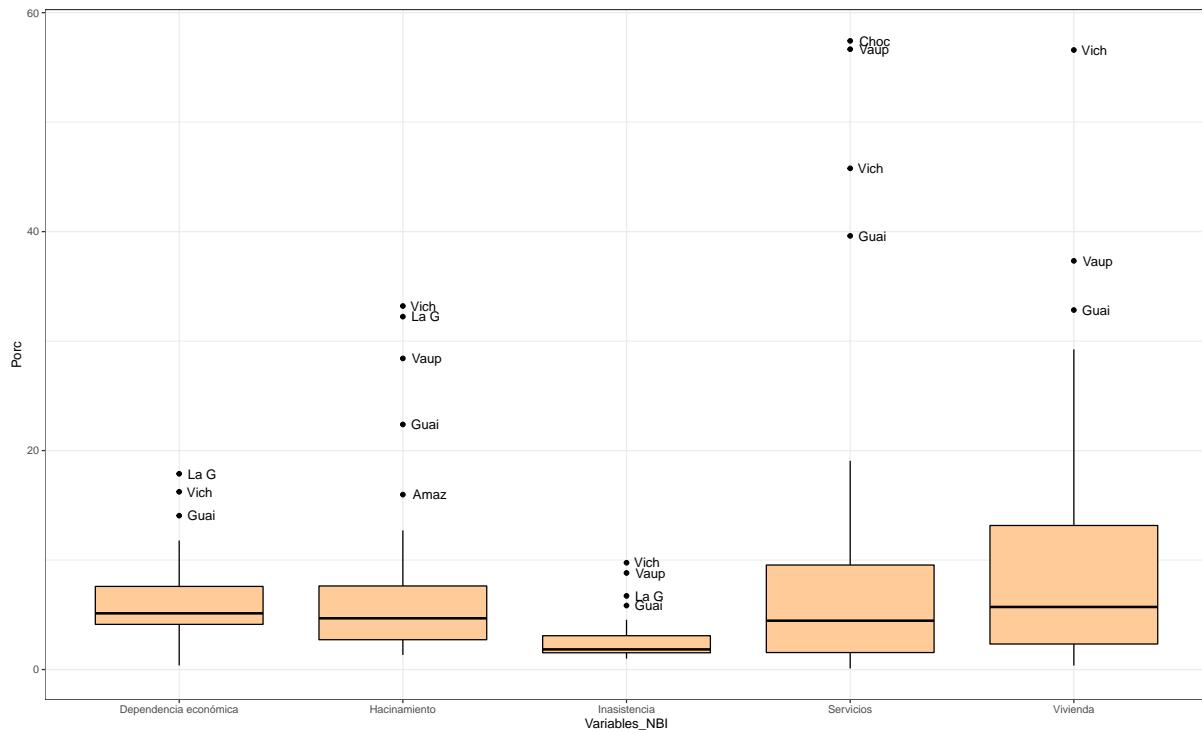
## 6.1. Análisis exploratorio de datos

Se realizó un análisis exploratorio dentro del conjunto de indicadores de pobreza y de actividad económica. El primero se mide a través de la insuficiencia en la satisfacción de necesidades básicas (%NBI) y el segundo está compuesto por los 12 sectores económicos en los que esta desagregado según el DANE. Este primer acercamiento con los datos, permite dar un panorama sobre su comportamiento..

Indicador	Promedio	CV %	Mín.	Q1	Mediana	Q3	Máx.
Vivienda	10.71	118.84	0.36	2.33	5.72	13.16	56.59
Servicios	10.36	150.07	0.10	1.55	4.47	9.55	57.42
Hacinamiento	7.85	109.29	1.34	2.73	4.68	7.64	33.21
Inasistencia escolar	2.77	75.50	1.01	1.53	1.84	3.09	9.76
Dependencia económica	6.38	62.71	0.39	4.13	5.14	7.60	17.88

**Tabla 6-1:** Estadísticas descriptivas de los indicadores de pobreza

La Tabla 6-1, contienen el conjunto de estadísticos descriptivos de las componentes del NBI. En términos generales se presentan altos niveles de variabilidad relativa ( $CV > 55\%$ ), a causa de algunos departamentos que presentan índices muy altos en ciertas componentes de pobreza, por lo que es evidente la alta desigualdad social que presenta el país. En la Figura 6-1 se identifican algunos departamentos con problemas críticos de insatisfacción de necesidades en su población.

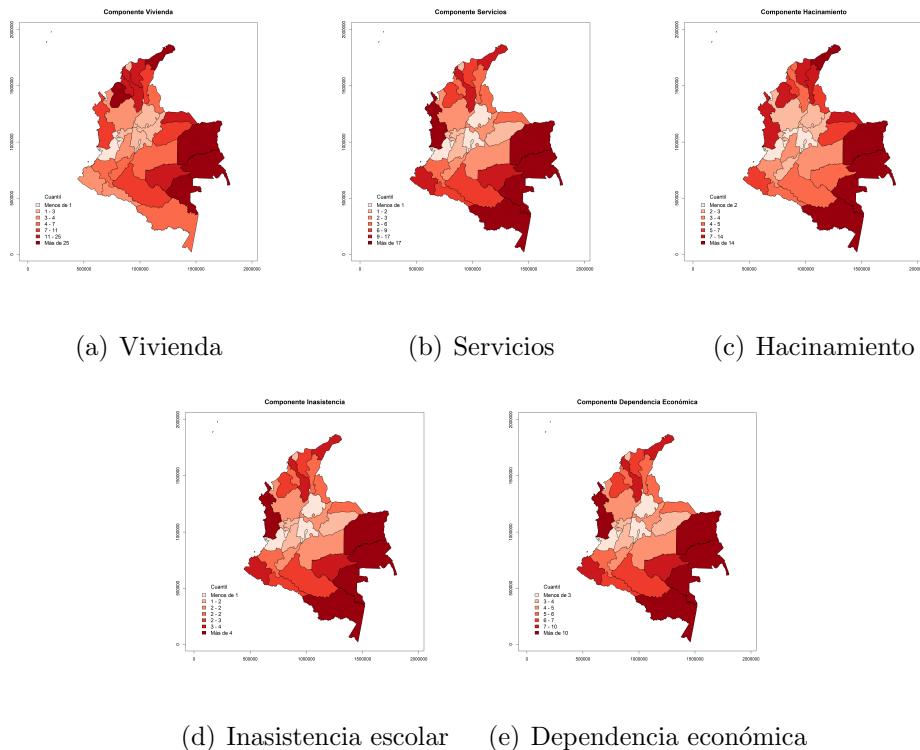


**Figura 6-1:** Diagrama de cajas de los Indicadores de Pobreza por Departamentos

En el caso de los departamentos Chocó y Vaupés, su población presenta en un 57.42 % y 56.66 %, respectivamente, problemas de acceso a los servicios públicos domiciliarios, seguidamente se tiene a los departamentos Vichada y Guainía. Respecto al hacinamiento se identifican a Vichada y la Guajira como los departamentos de mayor insuficiencia en esta necesidad. Este par de departamentos comparten con Vaupés y Guainía los porcentajes más críticos de inasistencia escolar en su población. Adicionalmente es La Guajira el departamento con más extrema situación de dependencia económica y el departamento que de manera general tiende a presentar alta insuficiencia en cada componente.

La Figura 6-2 representan la distribución espacial de cada uno de los indicadores de pobreza del NBI. En esta figura la intensidad del color rojo, indica que los problemas que caracteriza ese indicador son bastante altos en el departamento. En el componente vivienda representado en el mapa (a), los departamentos que presenta problemas de viviendas inadecuadas es Vichada, Guanía, Vaupés, La Guajira y Córdoba, donde los 3 primeros departamentos mencionados se encuentran muy cerca espacialmente.

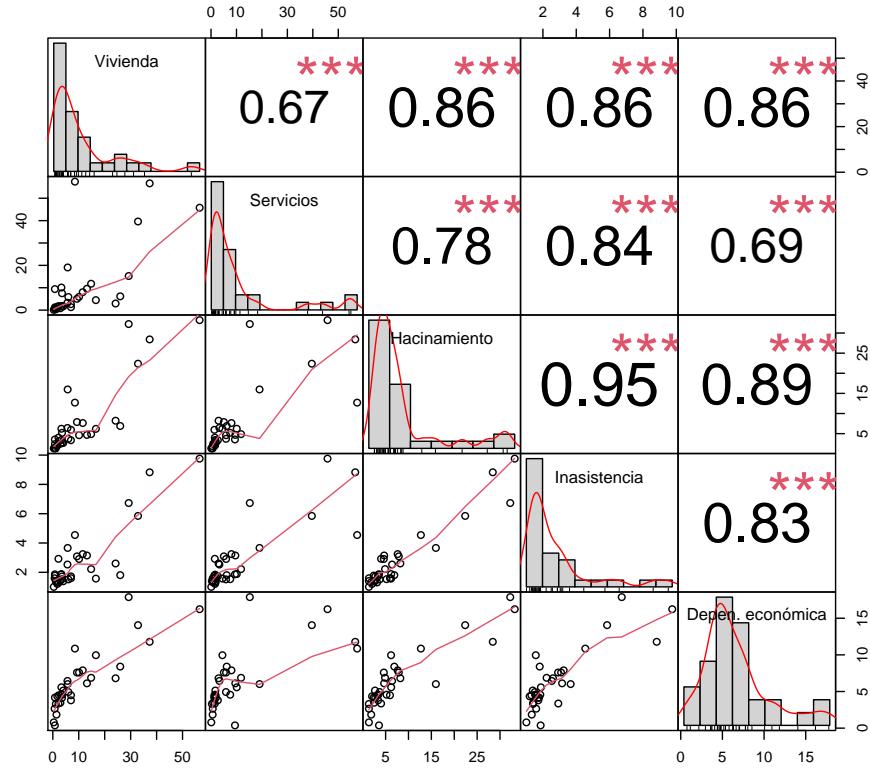
Por otro lado, en el mapa (b) se presenta el componente de servicios, en el que los departamentos del Choco y Vaupés están presentando mayor problema en este componente, con unas tasas de 57.42 % y 56.66 % respectivamente, seguido de estas regiones, se encuentran los departamentos Vichada y Guainía. Por lo tanto, solo teniendo en cuenta estos dos indicadores, se observa cierta tendencia de problemas en algunas zonas.



**Figura 6-2:** Distribución espacial de los indicadores de pobreza (NBI), año 2018

Los mapas (c) y (d) representan las componentes hacinamiento e inasistencia escolar, siguiendo la misma lógica de interpretación de los mapas, se puede observar que en el indicador de hacinamiento, los departamentos de La Guajira y Vichada son los que están presentando altos niveles de insuficiencia en esta componente, con unas tasas de 32.23 % y 33.2 % respectivamente, seguido de estas regiones se encuentran a una escala cercana Vaupés, Guainía, Amazonas y Chocó. Con respecto al indicador de Inasistencia escolar, se observa que Vichada y Vaupés tienen mayor valor en esta componente, por lo cual es un gran problema en esos departamentos. La pobreza es más critica en la costa Chocuana, la costa Norte y la región Llanos-Amazonas.

En el mapa (e) se puede observar el quinto indicador del Índice de Necesidades Básicas insatisfechas (NBI), el cual corresponde a dependencia económica, en este caso, los departamentos que presentaron este indicador alto fueron La Guajira y Vichada, con 17.88 % y 16.24 % respectivamente, seguidos a estas regiones están Gainía, Vaupés y Chocó. En general, al realizar un análisis global del comportamiento de los cinco indicadores del NBI, se puede decir que los departamentos Vichada, Guainía, La Guajira y Vaupés, son las regiones que presentaron las tasas más altas en cada uno de los componentes. Por lo tanto, según estos resultados estos departamentos son los que muestran mayor condición de pobreza en Colombia.



**Figura 6-3:** Gráfico de correlación entre las variables del NBI

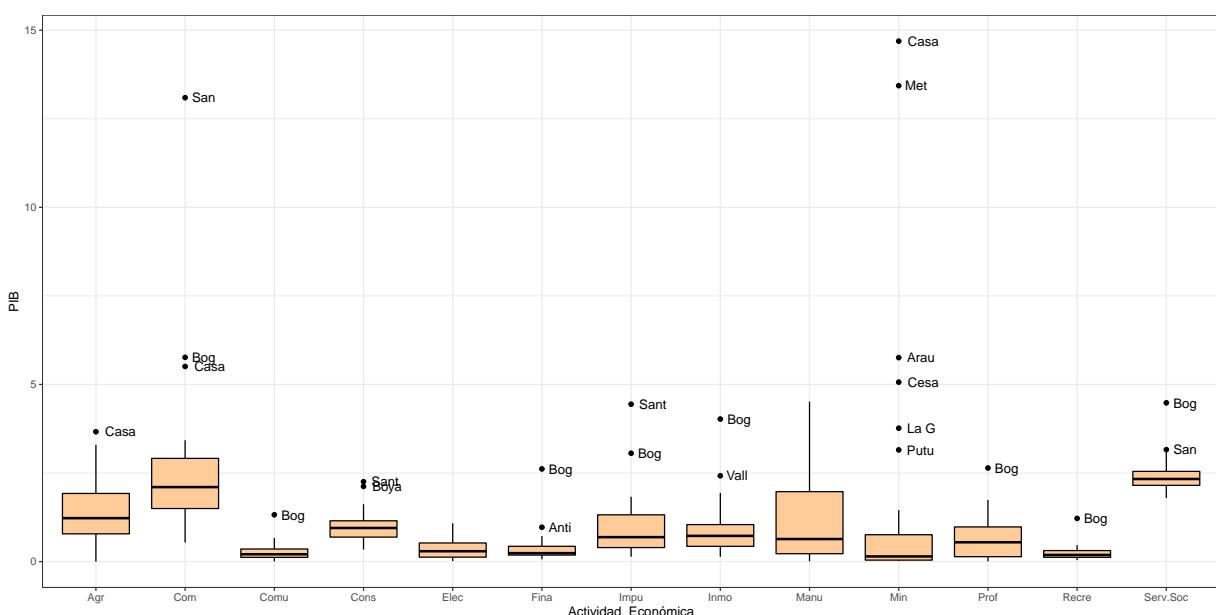
Por otra parte, la Figura 6-3, presenta la estructura de asociación de los indicadores de pobreza, donde se presenta una fuerte asociación lineal, en todos los casos positiva y con coeficientes de correlación estadísticamente significativos, entre los pares de variables. Lo que significa lo altamente relacionadas que se encuentran las condiciones de pobreza y que el incremento de una de las condiciones implica que aumente otra. Hay pares de variables que presentan correlaciones más altas que otras, como es el caso entre Inasistencia Escolar y Hacinamiento, donde puede ser más propenso que menores de edad no asistan a un centro de educación formal cuando cabe la posibilidad que la persona a cargo del hogar deba responder por integrantes de su núcleo familiar, lo cual implica el porque la correlación de hacinamiento es bastante alta con la condición de dependencia económica.

Adicionalmente se puede observar que hay una cierta tendencia en los datos en los que gran parte de ellos se concentran en una nube y son unos cuantos puntos que quedan alejados de los demás, por lo que posiblemente esos puntos que quedan apartados pueden ser departamentos como el Chocó, Vichada, Guainía ,Vaupes y La Guajira, que como se pudo analizar en las anteriores Figuras estos tendían a ser las regiones que presentaban altos niveles en los indicadores de pobreza.

Por otra parte, se tiene el comportamiento del PIB, desagregado en los 12 sectores de actividad económica más el rubro por recaudación de impuestos, en el que se debe tener en cuenta el ajuste que se realizó a nivel per cápita. Para ello, se presenta en la Tabla 6-2 y en la Figura 6-4 las estadísticas descriptivas asociadas.

Act.econ.	Promedio	CV %	Min.	Q1	Mediana	Q3	Max.
Agri	1.35	64.58	0.002	0.78	1.23	1.92	3.67
Min	1.59	218.42	0.015	0.042	0.15	0.76	14.69
Manu	1.17	104.03	0.010	0.22	0.64	1.97	4.52
Elec	0.36	80.83	0.014	0.13	0.29	0.53	1.08
Cons	1,00	43.55	0.341	0.69	0.95	1.15	2.26
Com.	2.59	83.98	0.539	1.50	2.10	2.91	13.09
Comu.	0.27	91.20	0.005	0.12	0.21	0.36	1.32
Finan	0.40	110.27	0.065	0.18	0.24	0.43	2.62
Inmo	0.88	84.69	0.138	0.43	0.73	1.05	4.02
Prof	0.63	94.89	0.007	0.14	0.55	0.98	2.64
Serv.Soc	2.43	19,03	1.793	2.15	2.33	2.55	4.48
Recre	0.24	87.94	0.043	0.12	0.19	0.31	1.22
Impu	0.99	90.66	0.138	0.40	0.69	1.32	4.44

**Tabla 6-2:** Estadísticas descriptivas de los indicadores del PIB



**Figura 6-4:** Diagrama de cajas de las Actividades económicas por Departamentos

Con respecto a los resultados de la Tabla 6-2, los altos coeficientes de variación en cada una de las componentes indican que el país presenta una estructura económica altamente variable entre regiones, pues algunos departamentos sobresalen en su capacidad de desarrollo en cierta actividad económica, lo cual puede estar influenciada por factores específicos tales como: la geografía de la región, el clima, las políticas del departamento, la calidad del sistema educativo y demás rasgos que pueden generar variabilidad en los bienes y servicios ofrecidos por el departamento.

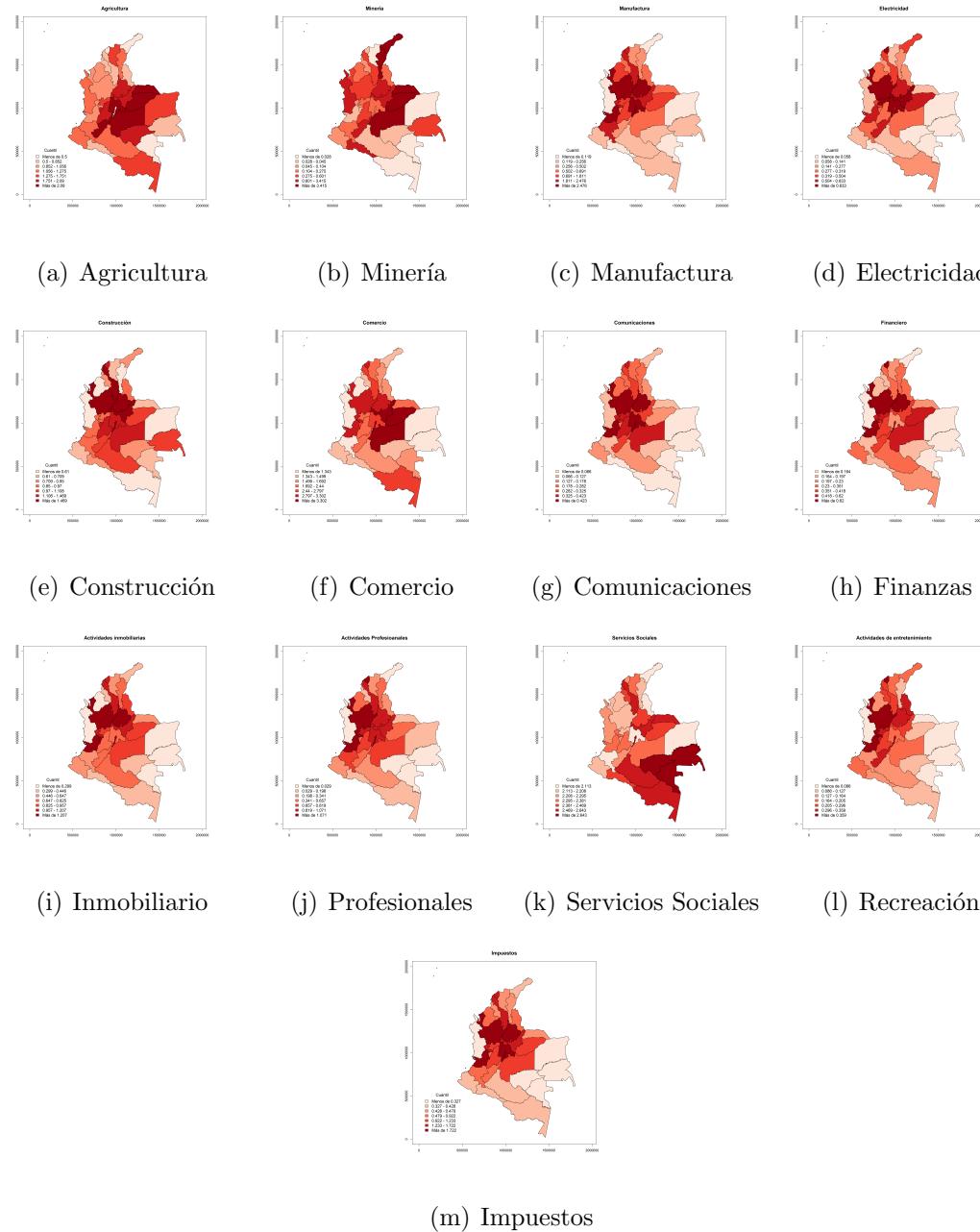
Uno de los sectores que se puede destacar por su elevado coeficiente de variación es el sector minero, pues este presenta un comportamiento altamente asimétrico, en el que aproximadamente el 50 % de los departamentos presenta un nivel casi nulo de valor agregado. Sin embargo, departamentos tales como Casanare y Meta presentan elevadas cifras de valor agregado per cápita en este sector, derivados posiblemente de la explotación petrolera. Puesto que el petróleo se constituye como la actividad que más ingresos genera para el departamento del casanare, además de existen explotaciones mineras de oro, magnesio, fosforo y níquel. Cesar y La Guajira presentan también elevadas cifras de valor agregado per cápita en este sector, siendo departamentos con grandes centros de explotación del carbón y sal. Por otra parte se tiene el departamento del Putumayo donde presenta un alto valor agregado en el sector minero, siendo la extracción de oro y de cobre una de las actividades que aportan a la economía de la región.

En el sector de la manufactura se destaca la fortaleza de los departamentos de Santander y Cundinamarca, seguido de los departamentos de Antioquia y Valle del Cauca. Por otra parte en el sector del comercio sobresalen San Andrés y Bogotá, pues los sectores más fuertes de la economía de San Andrés, están radicados en el comercio de importación de bienes de electrodomésticos textiles con destino al mercado nacional, además de que representa como uno de los primeros destinos turísticos en el Caribe. Por parte de Bogotá, según estudios del observatorio regional de la capital, informa que es considerada como el centro de negocios más grande del país, a su vez también destaca en el sector de actividades financieras y de seguros.

Casanare y Arauca destacan por su fortaleza en el sector de la Agricultura y la Ganadería. Arauca se concentra en la cría de vacunos y su producción agrícola se destina principalmente a satisfacer la demanda local. Por parte de Casanare, su economía se basa en la ganadería extensiva para la producción de carne la cual se cuenta como principal actividad económica del departamento. Cabe destacar que el coeficiente de variación más bajo está dado en el sector de servicios sociales, el cual a su vez cuenta con el mayor nivel medio de valor agregado per cápita , sin además de que cuenta con el de mayor valor agregado neto nacional.

Las fortalezas productivas de los departamentos en cada sector de la economía pueden representarse espacialmente, tal como se observa en los mapas de representación agrupados

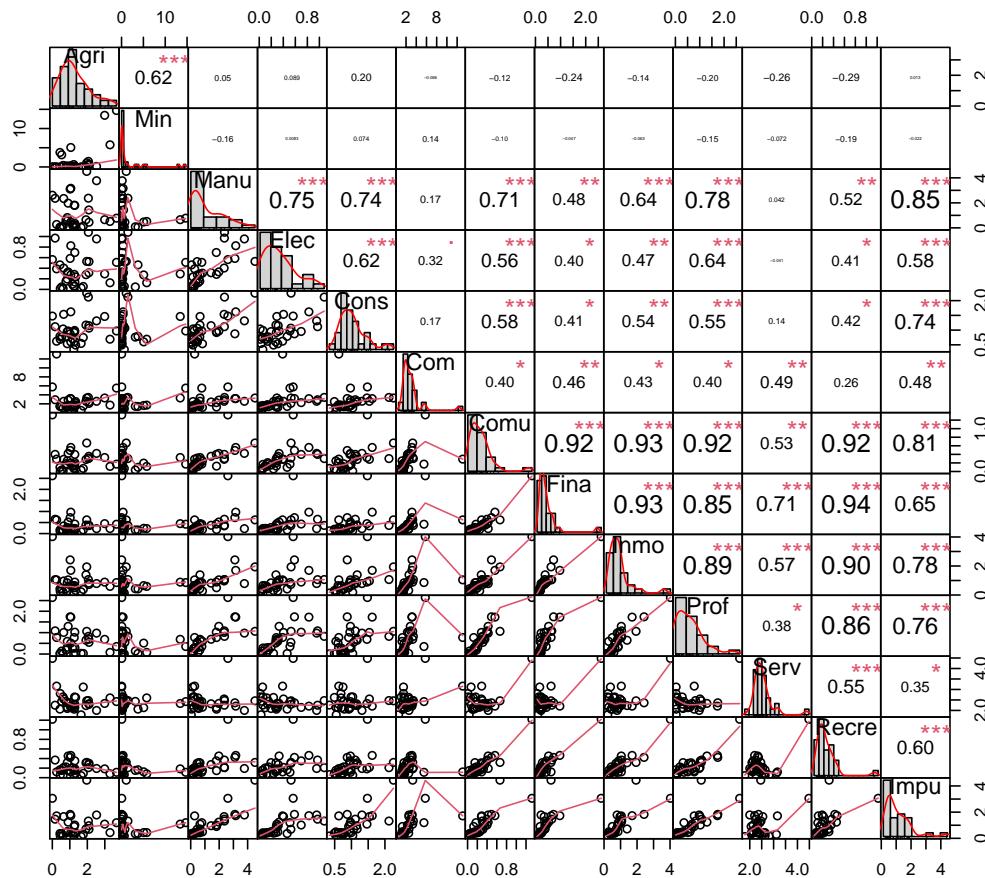
por cuartiles que se presentan en la Figura 6-5.



**Figura 6-5:** Distribución espacial de la actividad económica por departamentos, año 2018

Los mapas dan un panorama del comportamiento de cada departamento dependiendo del sector que se observe, a su vez se puede dar un acercamiento de posibles agrupaciones de algunas regiones. Es decir, se puede percibir que hay ciertas regiones las cuales por estar espacialmente cerca son fuertes en algunas actividades económicas. Como sucede por ejemplo en el sector de la agricultura y minería con los departamentos de Arauca, Casanare y Meta.

Casos contrarios en los que hay departamentos que no son tan fuertes en algunos sectores y que espacialmente se encuentran cerca, si se observa el mapa (h) el cual hace referencia al sector financiero se percibe que regiones como Vichada, Guanía y Vaupés entran dentro del cuantil que abarca el menor valor agregado en este sector. Se puede observar que Bogotá y departamentos como Antioquia, Valle del Cauca, son regiones que a nivel general presentan valores agregados altos en gran parte de los sectores económicos.

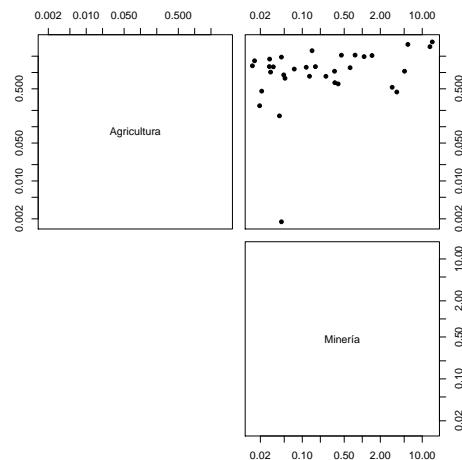


**Figura 6-6:** Gráfico de correlación entre las variables del PIB

Por otra parte, en la Figura 6-6 se presentan las correlaciones en el conjunto de variables de actividad económica, en la que cabe resaltar que cada una de estas variables se encuentran en la misma escala (Millones de pesos por habitante). Cabe resaltar que en cada rama de la actividad económica presentan diferentes rangos de variación entre los departamentos. Es decir, como se había analizado con las estadísticas descriptivas y los mapas espaciales hay departamentos que son más fuertes en ciertas actividades por lo que se presentan cifras muy grandes o lo contrario muy pequeñas.

Se puede destacar que existe una fuerte relación entre el valor agregado del sector de manufactura y el recaudo de impuestos. Por otra parte se destaca los altos coeficientes de correlación entre el sector de comunicaciones con actividades relacionadas con finanzas y seguros, inmobiliarias, actividades profesionales, actividades de recreación y entretenimiento y también su fuerte relación con el recaudo de impuestos. A su vez se puede observar como el sector financiero y de seguros presenta un fuerte relación con el sector inmobiliario y recreación. En el caso del recaudo de impuestos, se observa particularmente su fuerte asociación con la manufactura y su nivel bajo de asociación actividades de impacto social y nula relación con las actividades de la agricultura y la ganadería. Una situación extraña se observa en la relación observada entre el recaudo de impuestos y el valor agregado de la explotación minera y agricultura, cuya correlación se presenta como muy débil, tomando incluso un valor negativo. Esto podría deberse a que los impuestos de estos sectores económicos son tributados al gobierno central quien recauda y puede retornar a los departamentos través de proyectos de regalías que generalmente impactan la educación, la infraestructura y los servicios sociales.

Por otra parte, en el caso de los sectores económicos como agricultura y minería presentan coeficientes de correlación muy pequeños con respecto a los demás sectores, pero entre ellas presentan un coeficiente de correlación significativo, por lo que resulta interesante analizar su comportamiento. Por medio de un gráfico de dispersión, representado en la Figura 6-7 se presenta como se comportan los datos entre estas dos variables. La nube de datos ciertamente no se encuentra tan dispersa, pero se observa un punto muy alejado en ella, identificando ese punto como Bogotá dado que estas son actividades que no son fuertes de la capital.



**Figura 6-7:** Gráfico de dispersión entre los sectores minería y agricultura

Para dar un inicio al análisis multivariado, se realiza inicialmente un análisis de correlaciones

entre algunas variables del conjunto del PIB y las variables del NBI, donde se puede observar en la Figura 6-8 el gráfico de dispersión entre los dos conjuntos. Cabe aclarar que la selección de variables del PIB para realizar el gráfico de dispersión se hizo con base al análisis realizado en el gráfico de correlaciones de la Figura 6-6 donde las variables agricultura,minería,comercio, servicios sociales y recreación obtuvieron coeficientes de correlación bajos con los demás sectores.

Se puede destacar que hay una correlación negativa entre los dos conjuntos. Por ejemplo, si se observa la variable vivienda con respecto a cada unas de las variables del PIB, se pude ver una tendencia donde la variable del PIB tiene un valor muy alto y la variable del NBI tiene un valor bajo, por lo que se dice que tienen una relación inversa.Es decir que da indicios de que la productividad económica contribuye a disminuir los índices de pobreza. Hay algunas nubes de puntos donde se marca más o se puede apreciar de mejor manera esa relación negativa entre los dos conjuntos de variables, y otras nubes no tan definidas como sucede en el caso de la variable Construcción y vivienda , donde la nube de puntos se encuentra un poco más dispersa.

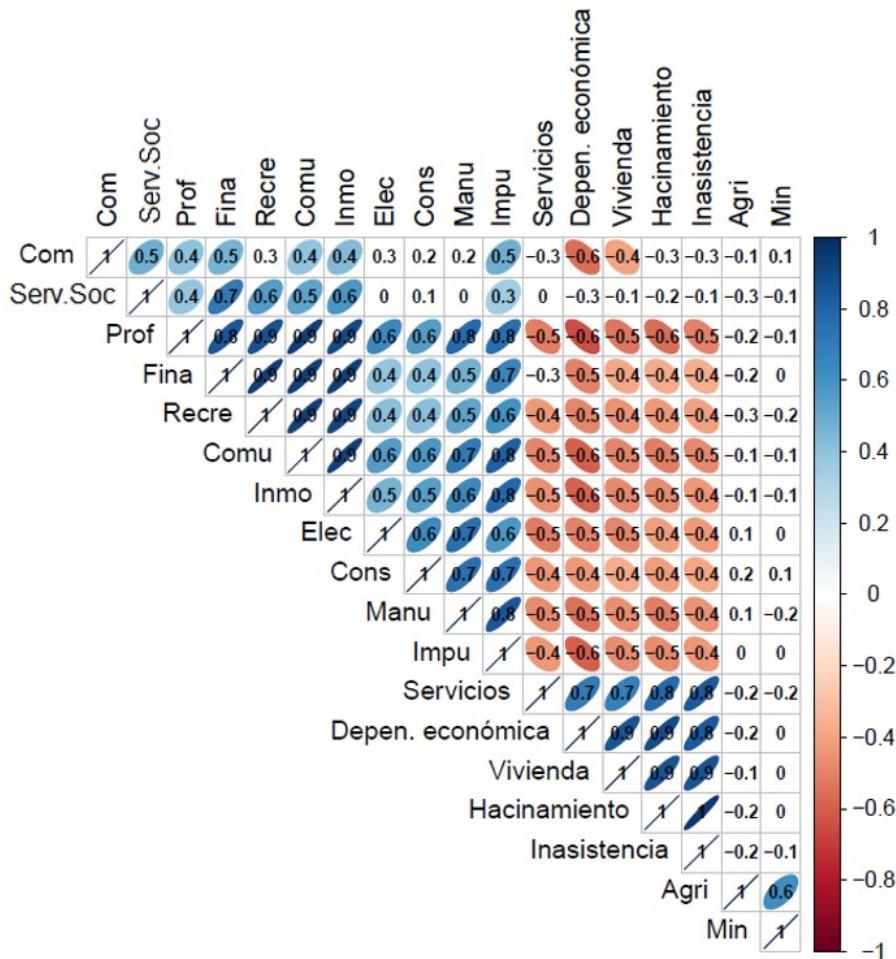


**Figura 6-8:** Gráfico de dispersión variables del PIB y NBI

Cabe recordar que los impuestos son una contribución que las personas naturales y jurídicas están obligadas a pagar al gobierno. Por lo tanto, por medio de los impuestos el Estado puede brindar a la población determinados bienes y servicios que están orientados a incremental el bienestar social. Por lo tanto, se destaca la relación entre la variable impuestos con respecto a las variables de pobreza, donde se puede observar una correlación negativa, es decir, que los impuestos están contribuyendo a la disminución de los índices de pobreza.

La Figura 6-9 es la representación de las estructura de correlaciones entre todas las variables del PIB y el NBI. Para efectos de enriquecer esta representación, aquellas correlaciones que no sean significativas no presentan un color asignado, además se realizaron agrupaciones entre todas las variables, con el objetivo de identificar que variables tenían una relación más fuerte. En el que se destaca un grupo aislado formado entre las variables minería y agricultura que no tuvieron relación fuerte ni con los demás sectores económicos ni con los indicadores de pobreza.

Además por medio de esta representación se puede ver con mayor claridad las fuertes relaciones que se destacaban en los análisis anteriores, pues acá es posible ver los diferentes grupos y su alta correlación, como se destaca la agrupación entre las variables de los sectores económicos como actividades profesionales, Financieras y seguros, comunicaciones, inmobiliarias y recreación. por otra parte la agrupación conformada por actividades eléctricas, de construcción y manufactura.



**Figura 6-9:** Correlaciones entre las variables NBI y PIB

Por otra parte, se puede observar como se agruparon todas las correlaciones entre los dos conjuntos de variables, presentándose una correlación negativa moderada.

Finalmente con el análisis exploratorio se logró dar un panorama sobre el comportamiento de los dos conjuntos de datos de este estudio de manera individual como conjunto y a su vez una visión de como se relacionan las variables del PIB y el NBI, se pudo destacar a su vez las fortalezas de algunos departamentos en los diferentes sectores de la economía,

Se pudo observar que hay una relación negativa entre los dos conjuntos de variables, por lo que encamina a estudiar a detalle esta relación, el cual es el objetivo del presente trabajo, dando paso a realizar un análisis estadístico simultaneo con las técnicas de AFM y PLS.

Se destaca a Bogotá como una de las regiones que mayores aportes hacen al PIB en diferentes sectores económicos, por tal motivo, resulta ser un individuo que dentro del análisis factorial aportaría mucho a la inercia, por lo que no permitiría visualizar las relaciones presentes en las variables y el resto de individuos (departamentos), por lo que se decide dejar a Bogotá como un individuo suplementario en el análisis multivariado.

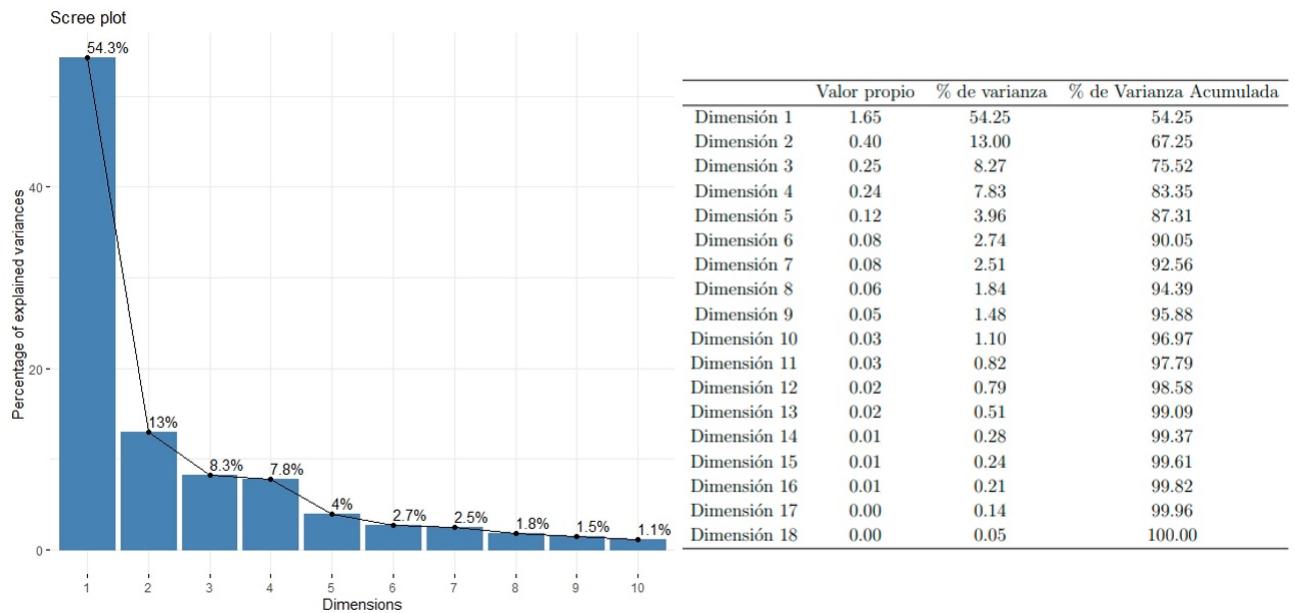
## 6.2. Análisis Factorial Múltiple -AFM

Se realiza el análisis de la tabla de datos completa, mediante el Análisis Factorial Múltiple, teniendo en cuenta que la tabla global está compuesta por 13 variables del conjunto de los sectores económicos del PIB y 5 variables por parte del conjunto de los indicadores de pobreza del NBI, para 32 individuos (departamentos) activos en el análisis y un individuo suplementario (Bogotá).

### Análisis de Factores

Inicialmente se define el número de dimensiones a ser analizados, esto se hace a través del análisis de los valores propios, los cuales representan la inercia o variación explicada para cada dimensión. Por lo tanto, consiste en tomar los valores propios que acumulen un porcentaje de varianza explicada mínimo aceptable.

En la Figura 6-10, se puede observar que los dos primeros ejes factoriales acumulan el 67.25 % de la variabilidad de la tabla de datos, donde se destaca el primer eje factorial el cual aporta el 54.25 % de la inercia total y el segundo eje aporta el el 13 % al plano factorial. Dado que la variabilidad recogida en el primer plano factorial es suficientemente grande, el análisis de datos se centrará en estudiar los grupos, individuos y variables sobre este primer plano factorial.

**Figura 6-10:** Gráfico de sedimentación y valores propios del AFM

### Ponderación de los grupos de variables en el AFM

El AFM es un análisis de componentes principales ponderado. Es decir, equilibra la influencia de los grupos de variables, dando a cada variable un peso. El mismo peso dentro del grupo y peso diferencial entre grupos, de esta forma se conserva la estructura interna de cada grupo. Por tanto la tabla del PIB y NBI son ponderadas por el inverso del primer valor propio obtenido en el ACP individual de cada tabla. Dichos valores propios y ponderaciones de los ACP de cada grupo son presentados en la Tabla 6-3.

Grupo	Cantidad de variables	$\lambda_j^1$	$1/\lambda_j^1$
PIB	13	6.91	0.144
NBI	5	4.29	0.233

**Tabla 6-3:** Valores propios y ponderaciones de los ACP de cada conjunto

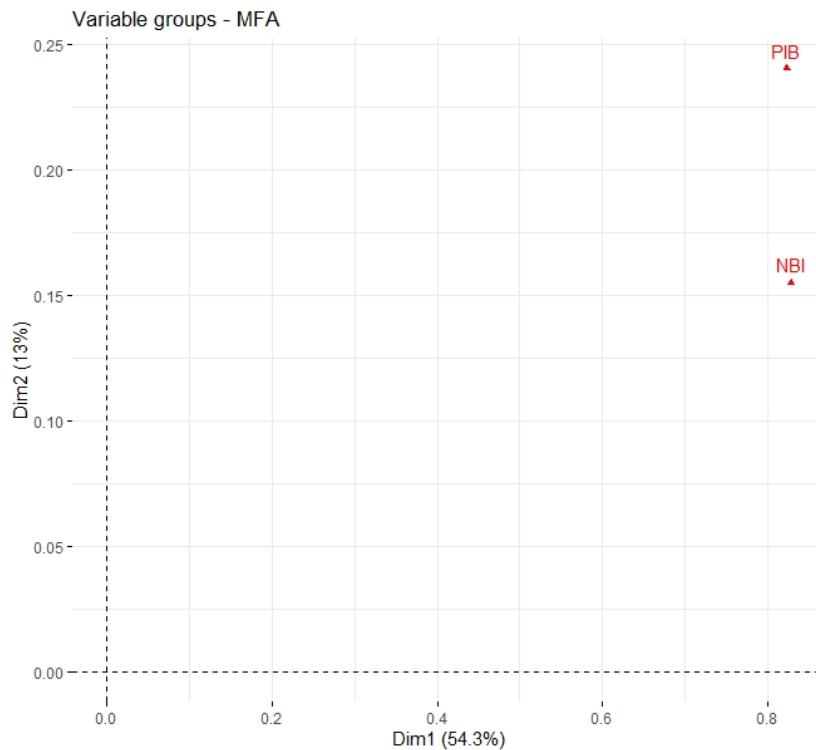
En la Tabla 6-4 se presenta las contribuciones que hace cada grupo a la construcción de los ejes factoriales. Se puede decir que para la primera dimensión se presenta una mayor correlación con el conjunto de las variables de pobreza. Por otra parte, la segunda dimensión presenta una mayor contribución por parte de las variables del PIB. Cabe mencionar que las contribuciones son similares, esto indica que el PIB y el NBI están fuertemente relacionados y aportan similarmente a caracterizar a los departamentos. Esta relación de los grupos y la construcción de los ejes se puede detallar en el punto de Tipología de los grupos.

	Dim.1	ctr	cos2	Dim.2	ctr	cos2
<b>PIB</b>	0.82	49.83	0.59	0.24	60.77	0.05
<b>NBI</b>	0.83	50.17	0.68	0.16	39.23	0.02

**Tabla 6-4:** Contribuciones del conjunto del PIB y NBI a las dimensiones

### Tipología de los grupos

En la Figura 6-11 se presenta la correlación entre los grupos y como estos contribuyen a las dimensiones. Como se puede observar la posición de los grupos sobre su contribución a la primera dimensión no se percibe una diferencia tan grande entre los dos conjuntos dado que ambos se encuentran en una coordenada muy similar son respecto al primer eje. Por otra parte, para la dimensión dos , el conjunto del PIB es el que mayor contribución está dando sobre ese eje factorial.



**Figura 6-11:** Representación de los grupos y su contribución a los ejes factoriales

Para estudiar los grupos de variables se cuenta con el Coeficiente Lg, el cual mide la dimensionalidad de cada grupo. Por lo tanto, el coeficiente vale cero cuando no existe relación entre los grupos y este no tiene límite superior. Dicho lo anterior, se presenta la Tabla 6-5, donde el coeficiente entre los dos conjuntos da indicios de que estos dos grupos tienen una relación moderada.

	PIB	NBI	MFA
PIB	1.15	0.44	0.96
NBI	0.44	1.01	0.88
MFA	0.96	0.88	1.11

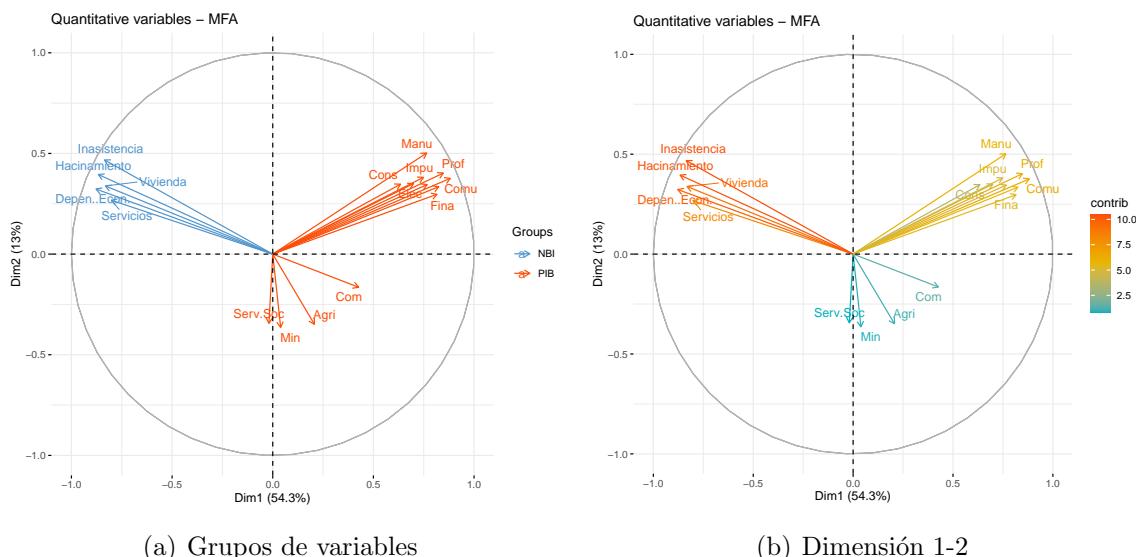
**Tabla 6-5:** Coeficientes Lg

Por otra parte, para continuar detallando la relación entre los grupos, se presenta el coeficiente RV, el cual se obtienen a partir de los coeficientes de correlación entre dos variables cualesquiera, donde este valor esta comprendido entre 0 y 1. Según los resultados de la Tabla 6-6, existe una relación moderada entre las variables que pertenecen a diferentes grupos.

	PIB	NBI	MFA
PIB	1.00	0.41	0.85
NBI	0.41	1.00	0.83
MFA	0.85	0.83	1.00

**Tabla 6-6:** Coeficientes RV

### Representación de las variables

**Figura 6-12:** Contribución de las variables a los ejes

La interpretación de la nube de variables y de individuos se hace de igual forma a como se realiza con un ACP. Por ello teniendo en cuenta que la varibildad recogida en el primer plano factorial fue lo suficientemente grande con una inercia del 67.25 %.

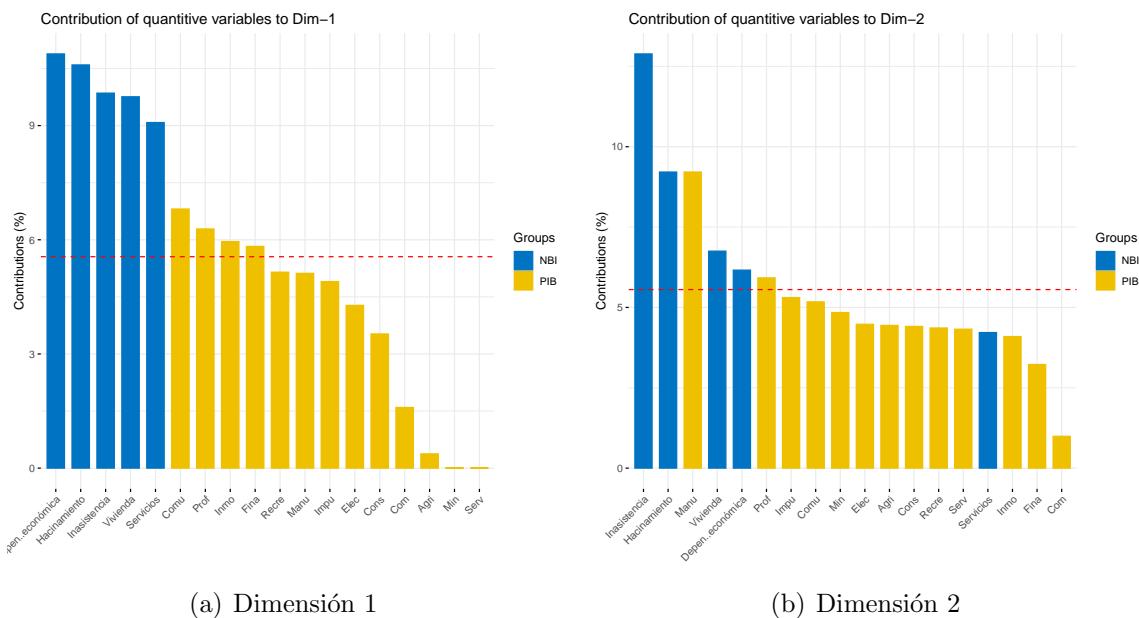
En la Figura 6-12 se presenta el círculo de correlaciones, el cual se puede observar la estructura de relacionan de las variables de los respectivos grupos. Por lo tanto, la dimensión uno puede entenderse como una característica que mide en dirección positiva la actividad económica de los departamentos, mientras que en su dirección negativa mide la pobreza de la población. Estas vías de relación contraria entre productividad y pobreza sugieren que mayor productividad en los diferentes sectores económicos como lo son en manufactura y el sector profesional contribuye a disminuir los indicadores de pobreza en la población colombiana.

La dimensión dos, presenta en su dirección positiva los problemas específicos de inasistencia escolar y en dirección negativa la actividad económica de comercio. Se observa en los diferentes planos que la productividad del sector de agricultura y de la minería no presentan correlación con los indicadores de pobreza, de lo cual se deduce que si bien estas actividades pueden generar beneficios económicos para la nación, no contribuyen a modificar las condiciones de pobreza, es decir que presentan valor agregado económico pero no impacto social.

Otra condición particular que se deduce desde este análisis es sobre la inasistencia en servicios públicos domiciliarios con las variables de productividad, según lo cual la existencia de mayor productividad en las regiones no presenta un impacto sobre sobre esta falencia en la población, siendo posiblemente el mejoramiento de esta condición una tarea del estado y no de la dinámica productiva de sus departamentos.

La Tabla 6-7 presenta las dos primeras coordenadas, contribuciones y la calidad de representación de las variables en las dos primeras dimensiones. Se destaca que las variable Hacinamiento y Dependencia económica son las que más contribuyen a la primera dimensión, además de que son las que se encuentran mejor representadas entre el grupo de las variables de pobreza. Por otra parte, la variable inasistencia escolar es la que a nivel general mayor contribución hace a la segunda dimensión, del conjunto de las variables de productividad el sector de manufactura y el sector profesional son las que más contribuyen a este eje. Lo anterior induce a realizar una relación entre inasistencia escolar con la variable del sector profesional, dado que ambas están asociadas a la educación, por lo tanto el incremento de la variable profesional podría contribuir a la disminución de la componente de inasistencia escolar.

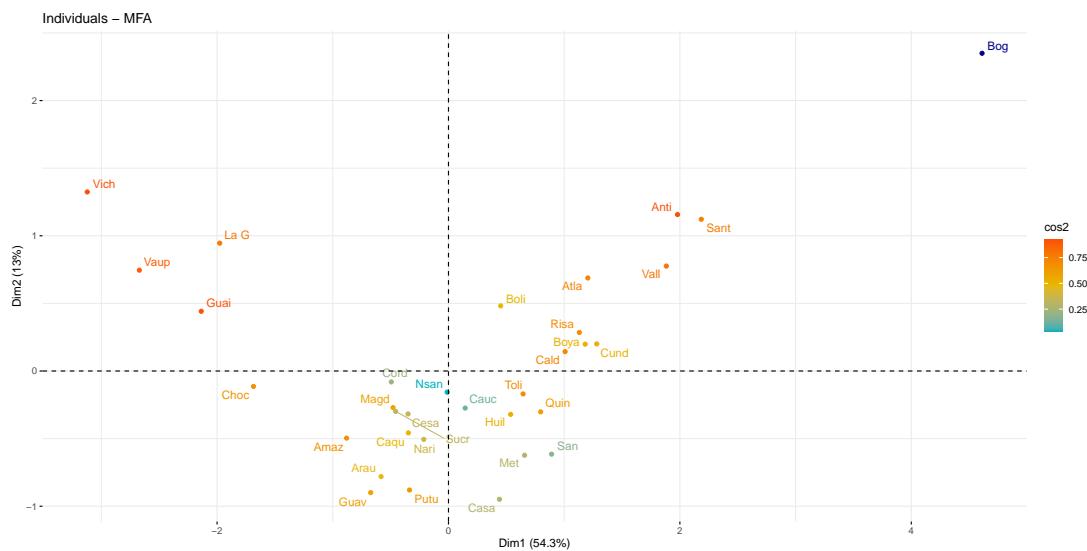
Variable	Coordenadas		Contribución		Calidad. repres.	
	Dim.1	Dim.2	Dim.1	Dim.2	Dim.1	Dim.2
Agri	0.21	-0.35	0.37	4.44	0.04	0.12
Min	0.04	-0.36	0.01	4.84	0.00	0.13
Manu	0.76	0.50	5.12	9.21	0.58	0.25
Elec	0.70	0.35	4.28	4.48	0.49	0.12
Cons	0.63	0.35	3.53	4.41	0.40	0.12
Com	0.43	-0.16	1.59	0.99	0.18	0.03
Comu	0.88	0.38	6.81	5.17	0.78	0.14
Fina	0.82	0.30	5.83	3.22	0.67	0.09
Inmo	0.82	0.33	5.95	4.09	0.68	0.11
Prof	0.85	0.40	6.28	5.92	0.72	0.16
Serv.Soc	-0.02	-0.34	0.00	4.32	0.00	0.12
Recre	0.77	0.35	5.15	4.36	0.59	0.12
Impu	0.75	0.38	4.90	5.31	0.56	0.15
Vivienda	-0.83	0.34	9.76	6.75	0.69	0.11
Servicios	-0.80	0.27	9.08	4.22	0.64	0.07
Hacinamiento	-0.87	0.40	10.59	9.21	0.75	0.16
Inasistencia	-0.84	0.47	9.85	12.89	0.70	0.22
Depen.económica	-0.88	0.32	10.88	6.16	0.77	0.10

**Tabla 6-7:** Coordenadas, Contribuciones y Calidad de representación de las variables**Figura 6-13:** Contribución de las variables a los ejes

La Figura 6-13, representa el aporte que hacen las variables a la construcción de los ejes factoriales, por lo que se puede detallar de mejor manera que grupo y que variables están contribuyendo en mayor y menor medida. Para la dimensión uno se destaca el grupo de NBI,

ya que todas las variables del grupo de pobreza están contribuyendo en gran medida, de las variables de productividad se destaca al sector comunicaciones y actividades profesionales. Para la dimensión dos, la variable de pobreza Inasistencia escolar es la que más contribuye seguidamente de hacinamiento y manufactura que aportan en igual medida.

### Representación de los individuos



**Figura 6-14:** Individuos y contribución a los ejes

La Figura 6-14 representa el comportamiento de los departamentos. Los departamentos que se ubican en los cuadrantes I y IV, representan los departamentos de alto y mediano nivel de productividad y bajos niveles de pobreza.

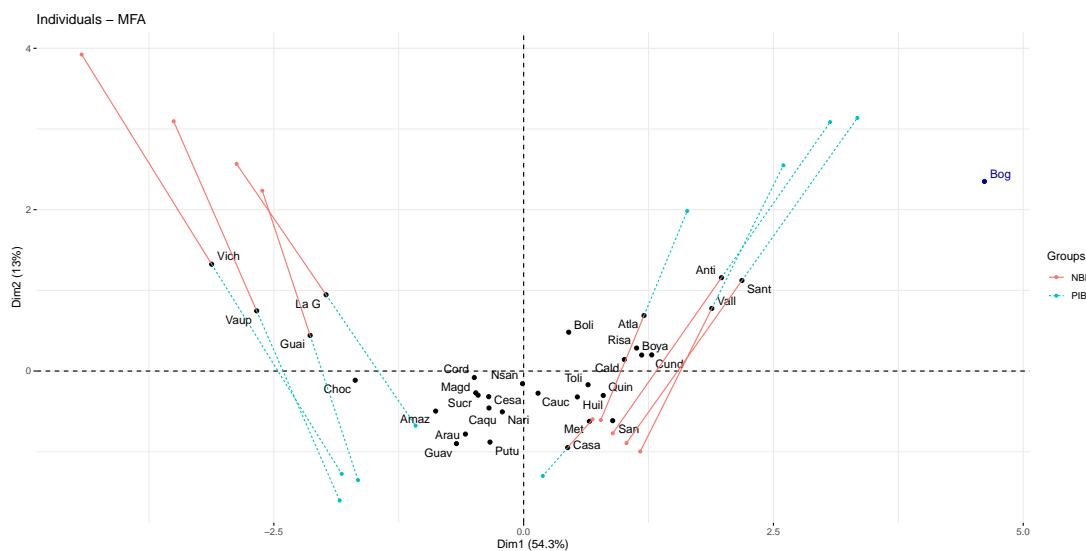
Se observa que los departamentos de Antioquia, Santander y Atlántico destacan en el sector de la Manufactura. Valle del Cauca además de estar relacionado con dicho sector, se ve altamente asociado con el sector profesional e Inmobiliario. Se puede observar como San Andrés se encuentra altamente relacionado con el sector comercial. Por su parte, departamentos como Boyaca y Cundinamarca presentan algún nivel de desarrollo en el sector financiero. Risaralda se encuentra más asociado con el sector de construcción. Departamentos como Cauca, Tolima y Quindío se encuentran relacionados con el sector Agricultor y comercial. Casanare y Meta se encuentran altamente relacionados con el sector Minero y Agricultor, pues estos departamentos se destacan en la explotación de hidrocarburos y ganadería.

Los departamentos restantes (Cuadrantes II y III) presentan bajos niveles de productividad y de forma simultanea preocupantes niveles de pobreza. Los casos más críticos los

presentan Vichada, La Guajira, Vaupés y Guainía, principalmente por condiciones críticas en inasistencia escolar, hacinamiento y vivienda, mientras que Chocó presenta problemas con abastecimiento de servicios públicos. Los departamentos de Magdalena, Córdoba y Sucre presentan problemas de vivienda adecuada y alta dependencia económica.

Por otra parte, se puede observar a Bogotá, como individuo suplementario en el análisis factorial. Como estaba previsto, Bogotá se encuentra muy aislado de los demás departamentos, esto debido a la alta productividad que aporta en diferentes sectores económicos al PIB, como lo es principalmente en el sector financiero, profesional y comercial.

### Representación superpuesta de grupos e individuos



**Figura 6-15:** Nube de individuos y representación superpuesta

Una herramienta adicional que suministra el Análisis Factorial Múltiple la constituye la posibilidad de realizar representaciones parciales de los individuos en el espacio factorial de cada subconjunto tabla, lo cual en este caso permite evaluar el efecto que presenta una subtabla en la ubicación del departamento en el plano factorial conciliado. La Figura 6-15, es entonces una forma de poder visualizar el comportamiento de los departamentos dentro de cada grupo (NBI y PIB).

Se observa por ejemplo que la Guajira se encuentra en una condición de productividad muy cercana al de los departamentos de mediana productividad, sin embargo sus condiciones de pobreza son muy extremas en comparación con estos departamentos. Por otra parte departamentos como Antioqueña, Valle del Cauca, Santander y Atlántico indican que sus

altos niveles de productividad reflejan a su vez bajos niveles de pobreza. Por su parte Vaupés, Vichada y Guainía tienden a presentar críticas cifras en las diferentes componentes de pobreza y su vez no representar ser departamentos que reflejen una alta productividad en alguno de los sectores de la economía.

### 6.3. PLS

Se emplea el algoritmo de PLS2, el cual generaliza y combina características del método de Análisis de Componentes Principales y Análisis de Regresión Múltiple, esta técnica generalmente se utiliza cuando se tienen gran cantidad de predictores, como sucede en este estudio, pues se cuenta con 18 variables. Para este método se debe seleccionar el conjunto de datos que represente la matriz de las variables predictoras, para este caso se toma al conjunto del NBI y el otro conjunto de variables será la matriz conformada por variables de respuesta, en este caso las variables del PIB.

Se construye la combinación lineal  $u_1$  de las columnas  $Y_0$  (Matriz del conjunto del PIB estandarizada) y la combinación lineal  $t_1$  de las columnas de  $X_0$  (matriz estandarizada del conjunto NBI) de tal modo que se maximice la  $Cov(u_1, t_1)$ . Así se tienen dos nuevas variables  $u_1$  y  $t_1$  que están lo más correlacionadas posible y que a su vez están resumiendo lo mejor que se puede la información que contiene  $X_0$  y  $Y_0$ . Lo anterior corresponde al proceso para la determinación de la primera componente, de manera que para la determinación de la segunda componente esta vez reemplazando las matrices  $X_0$  y  $Y_0$ , por las nuevas matrices residuales  $Y_1$  y  $X_1$ ; se pasa a obtener dos nuevas componentes  $t_2$  y  $u_2$  las cuales maximizan la  $Cov(u_2, t_2)$ . Se presentan a continuación los resultados del análisis de PLS2.

El PLS presenta un estadístico importante para la interpretación sobre las variables predictoras, el cual resumen la contribución de una variable al modelo, es decir, la importancia de la variable en la proyección, nombrado entonces como VIP. El VIP representa el valor de cada variable explicativa en el ajuste del modelo tanto para el conjunto de variables respuesta o dependientes como para las variables predictoras.

De acuerdo con Marquez (2017), los valores (VIP) permiten clasificar las variables  $x_j$  en función del poder explicativo que tiene sobre  $y_j$  sobre el espacio de proyección de la nube de puntos. Las variables que arrojan valores grandes en el VIP ( $> 1$ ) son consideradas las que mayor aporte hacen a las construcciones de  $Y$ . En la Tabla 6-8 se presenta los valores del VIP para cada variable del PIB las cuales son en este análisis las variables predictoras, donde se destaca que los sectores económicos Comunicaciones, Finanzas y seguros, Inmobiliario y sector profesional son las variables que mayor contribución hacen a la construcción de  $Y$ , pero de manera general las demás variables tienen valores cercanos a uno, lo que indican que todas las predictoras están aportando en el análisis.

	t1	t2
Agri	0.53	0.91
Min	0.18	0.40
Manu	1.05	1.04
Elec	1.02	0.96
Cons	0.87	0.88
Com	0.79	0.90
Comu	1.34	1.25
Fina	1.23	1.14
Inmo	1.24	1.15
Prof	1.29	1.19
Serv.Soc	0.12	0.59
Recre	1.24	1.19
Impu	1.04	1.03

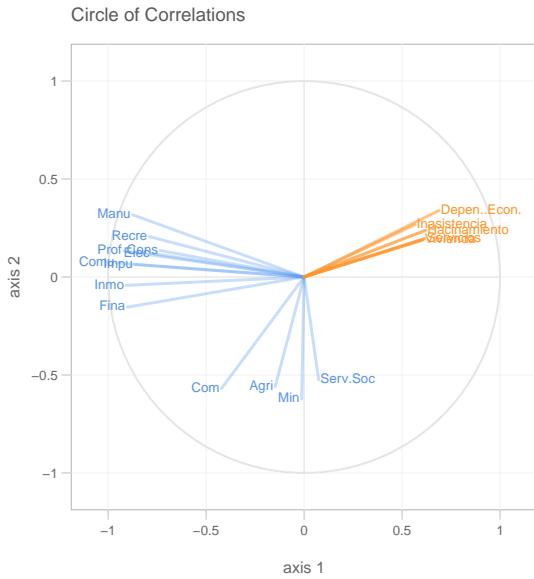
**Tabla 6-8:** Estadístico VIP

En la Tabla **6-9** se presentan las correlaciones de las nuevas variables construidas, las cuales están resumiendo de la mejor manera la información de la matriz de las variables predictoras y la matriz de variables respuesta. Por lo que se destaca que la correlación entre  $t_1$  y  $u_1$  es moderada alta, lo que indica que el conjunto del PIB y el NBI se encuentran relacionados.

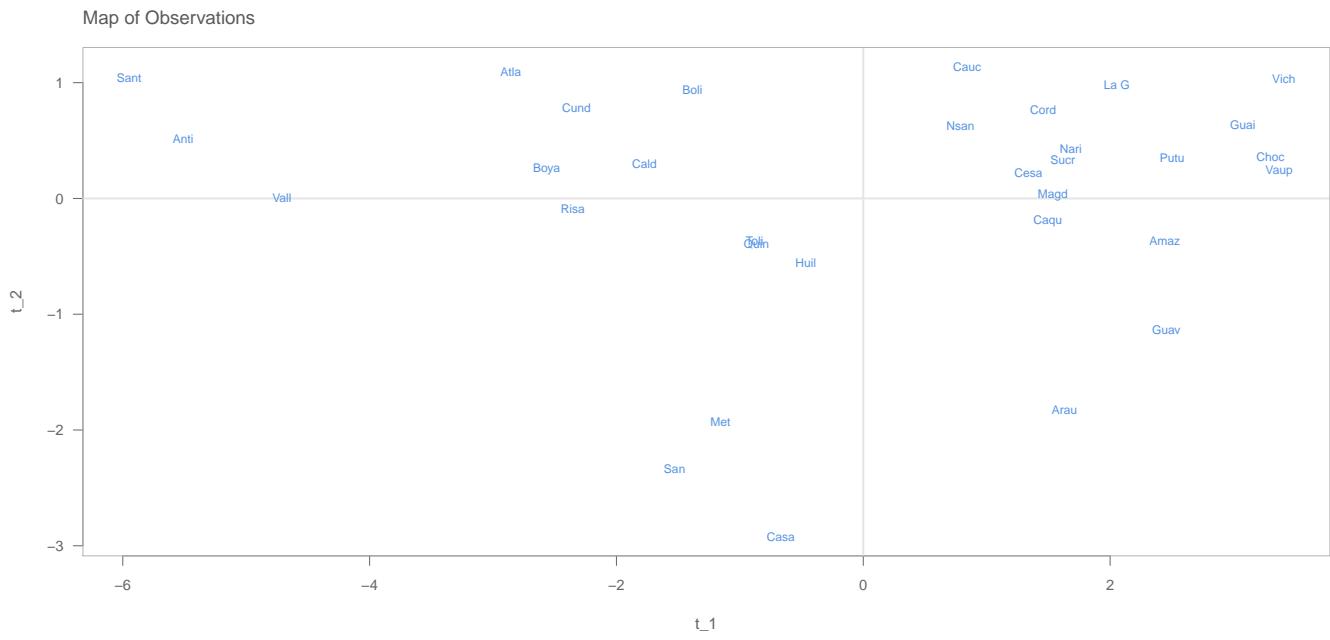
Correlación T_U		
	u1	u2
t1	0.67	5.48e-18
t2	0.27	0.37

**Tabla 6-9:** Correlación entre T y U

Por otra parte, la Figura **6-16** presenta el círculo de correlaciones, el cual es construido a partir de las dos primeras componentes del PLS. En el que se puede observar el comportamiento de las variables, donde el conjunto de variables de pobreza se encuentra inversamente correlacionadas con el conjunto de variables de productividad por lo que da indicios de que a mayor crecimiento de las actividades económicas contribuiría a la disminución de las necesidades básicas que no se están cubriendo en la población. Cabe destacar que las variables Agricultura, Minería, Servicios sociales y Comercio, son sectores que generan un aporte económico a la nación pero no representan un impacto social, es decir, estos sectores no están contribuyendo a la disminución de los índices de pobreza.



**Figura 6-16:** Círculo de correlaciones con método de PLS



**Figura 6-17:** Nube de individuos con método de PLS

La Figura 6-17 presenta el comportamiento de los departamentos, se debe tener en cuenta que se encuentra en el espacio de las variables predictoras (PIB), es decir, el plano está construido a partir de las variables  $t_1$  y  $t_2$ . Donde se destaca que los departamentos como Santander, Antioquia y Valle del Cauca se encuentran muy alejados de los demás.

departamentos, por lo que indica sus altos niveles de productividad en comparación a los otros departamentos, lo cual implica que se presenten a su vez menores índices de pobreza. Cabe destacar que se puede observar como en el plano se forman ciertos grupos de departamentos, en el cuadrante I, está comprendido por todas las regiones que presentan extremas condiciones de pobreza, como lo es Vichada, Vaupes, La Guajira, Guainía, entre otros. Por otra parte, se forma un grupo de departamentos con productividad media alta como los son Cundinamarca, Caldas, Boyaca Y Atlántico.

La Tabla 6-10 presenta valores predichos con el modelo construido a partir del método de PLS, estos valores se presentan con el objetivo de visualizar como la pobreza se ve impactada por la productividad económica, es decir, que por medio de estos valores se da un panorama de comparación con respecto a los valores que se predijeron y los valores originales. Se escogieron departamentos que presentaron altos niveles de pobreza y departamentos con altos niveles de productividad.

Por lo anterior, los departamentos como Vichada, Vaupes, Guainía y la Guajira, presentaron valores muy altos en cada uno de los indicadores de pobreza, por lo que indica que los niveles de productividad de estas regiones no son altos y esto se daba a las extremas condiciones de pobreza. Es decir, una población que no tenga condiciones para suplir sus necesidades básicas, tendrá ciertamente el efecto de no tener condiciones para ser altamente productivos. Caso contrario con departamentos como Antioquia, Valle del Cauca , Santander y Cundinamarca, son departamentos con bajos niveles de pobreza, en los que presentan altos niveles de productividad en varios sectores de la economía, por lo que hace creer que al estar las necesidades básicas de la población adecuadamente cubiertas provocará el efecto de una crecimiento económico. Los valores predichos para estos departamentos dan negativos, lo que indica que la alta productividad de estas regiones contribuya a la disminución de los niveles de pobreza.

	Vivienda	Servicios	Hacinamiento	Inasistencia	Depen.económica
<b>Anti</b>	<b>-4.48</b>	<b>-8.63</b>	<b>-2.46</b>	0.54	1.38
<b>Atla</b>	4.98	3.20	4.29	2.11	4.97
<b>Cund</b>	5.86	4.28	4.78	2.18	5.13
<b>Choc</b>	21.93	24.27	15.71	4.57	10.53
<b>La G</b>	19.62	21.44	14.35	4.33	10.03
<b>Vall</b>	<b>-3.26</b>	<b>-7.14</b>	<b>-1.81</b>	0.64	1.56
<b>Guai</b>	<b>21.92</b>	<b>24.27</b>	<b>15.81</b>	<b>4.62</b>	<b>10.66</b>
<b>Vaup</b>	<b>21.84</b>	<b>24.15</b>	<b>15.60</b>	<b>4.53</b>	<b>10.44</b>
<b>Vich</b>	<b>23.86</b>	<b>26.72</b>	<b>17.30</b>	<b>4.99</b>	<b>11.53</b>
<b>Sant</b>	<b>-4.53</b>	<b>-8.66</b>	<b>-2.29</b>	0.64	1.62

**Tabla 6-10:** Valores Predichos para el NBI

## 6.4. Integración de Resultados AFM y PLS

Las dos técnicas aplicadas para realizar el Análisis de Múltiples Tablas, son ciertamente parecidas pero tienen objetivos distintos. El AFM por su parte busca maximizar la varianza que está contenida en la tabla completa de datos, mientras que PLS busca maximizar la covarianza entre una tabla y otra, es decir, que permite conocer como se comporta una variable, en este caso conjunto de variables, en función de lo que hace otro conjunto de variables.

Una de las propiedades que tiene el AFM es que por medio de los coeficientes LG y RV se puede analizar que tanto se relacionan los grupos involucrados en el estudio, los resultados arrojados por esta técnica indicaron una correlación moderada. Donde del coeficiente Lg y RV obtuvieron valores de 0.44 y 0.41 respectivamente lo que indica a nivel general una correlación moderada entre los conjuntos NBI y PIB. Por otra parte, teniendo en cuenta los objetivos de la técnica de PLS, se obtuvo que la correlación entre los dos conjuntos de variables se maximizara, dando como resultado una correlación de 0.67 para el primer eje y 0,37 para el segundo eje. Por lo que los resultados del PLS dan soporte para indicar que estos conjuntos de variables están relacionados.

Por otra parte, los resultados del comportamiento de las variables en los círculos de correlación en ambos métodos, dan un fiel indicador la correlación negativa que se presenta con los dos conjuntos de estudio. En lo que también se destaca que las variables de Minería, Agricultura y Servicios sociales no representan un impacto social, es decir, son sectores con un alto aporte a la economía nacional pero a nivel de contribuir a reducir los niveles de pobreza estos no tienen mayor aporte.

Un gran aporte que hace el análisis por PLS, es su estadístico VIP, pues por medio de este se pudo afirmar que las variables del PIB tienen un gran aporte en el modelo para la explicación de las variables respuesta que en este caso son el NBI. Es decir, desde el modelo que se desarrolla con PLS, se busca ver como la productividad influye en las condiciones de pobreza en la población. La cual es la interpretación que se detectó desde las variables del NBI predichas, donde se observó que habían departamentos que presentaban cifras negativas, lo que podría significar que los altos niveles de productividad contribuyen en gran medida a que haya una disminución de los niveles de pobreza o en caso contrario aquellos departamentos que presentan altos niveles de pobreza afecte directamente a que no se presente un adecuado desarrollo económico en la región. Lo cual se integra bastante bien con los resultados de la representación superpuesta por parte del AFM, que en casos como La Guajira presenta condiciones de productividad cercana a departamentos de mediana productividad pero se ve afectado por las extremas condiciones de pobreza.

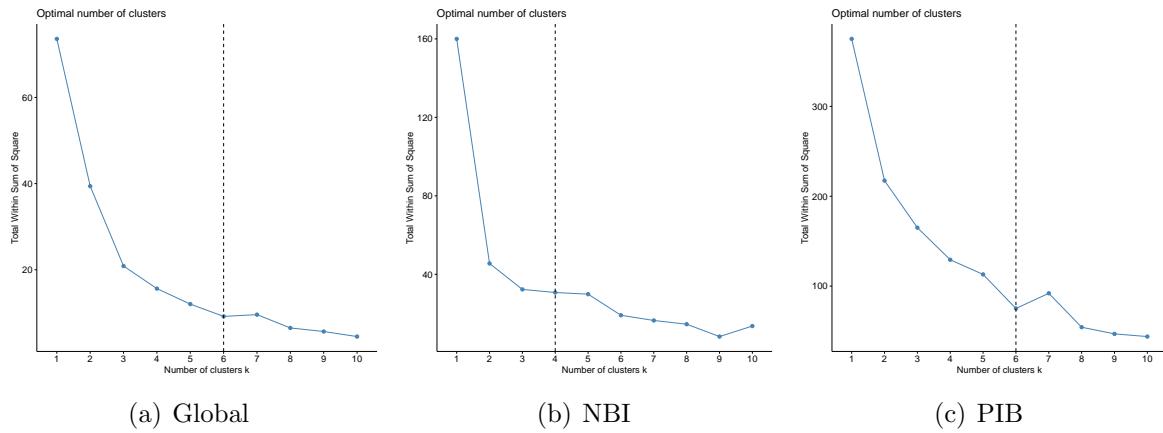
## 6.5. Análisis de Clúster - AC

El objetivo de esta fase es realizar agrupaciones de departamentos que desde las perspectivas de la actividad económica y de las condiciones de pobrezas, permitan identificar subgrupos de departamentos con similares condiciones, que a futuro puedan recibir tratamiento de política administrativa y de inversión pública similar. Por tanto, para realizar el análisis de cluster se parte bajo los dos primeros componentes del AFM construido en la anterior fase. Dado que la retención de ejes factoriales se obtuvo que los dos primeros ejes factoriales retienen el 67.3 % de la inercia contenida en los datos. Es decir, que con la información resumida en las coordenadas de las dos primeras componentes del AFM se construye la matriz de datos para realizar el análisis de clúster. Cabe resaltar que el AFM arroja unas coordenadas globales y a su vez se tiene el análisis factorial de cada conjunto es decir coordenadas para el PIB y para el NBI, por lo tanto, se presentan los resultados bajo las coordenadas globales y las parciales.

Se debe tener en cuenta que, exactamente no existe un criterio para elegir cual de los métodos es mejor para identificar el número de clúster óptimo, por lo que la decisión termina siendo arbitraria y su vez depende más de cual método refleje mejor los propósitos del estudio. En este caso se eligió el método del codo, el cual da una visión del comportamiento de las inercias, pero cabe aclarar que por motivos de una mejor visibilización de las agrupaciones se escogió un número de clúster mayor al que sugería el método, es decir, que este método se tomó como punto de referencia para tomar la decisión que mejor ilustrara las agrupaciones. Finalmente por medio de los dendrogramas se podrá detallar de mejor forma el corte de los grupos.

Inicialmente se presenta en la Figura 6-18 el método del codo, el cual presenta en una gráfica lineal la inercia respecto al número de clúster, en ella se puede apreciar un cambio brusco en la evolución de la inercia , donde la linea hace un quiebre en forma de brazo o de codo. Por lo tanto, para el análisis Global aprecia un quiebre en 3, lo que estaría sugiriendo que el óptimo número de clúster es 3, pero como se argumentaba anteriormente se decide tomar 6 cluster. Para el caso del NBI el quiebre se da en 2, pero dado que no resulta tan ilustrativo dos agrupaciones, se toman para este caso 4 clúster. En el caso del PIB el quiebre se da entre 2 y 3, pero para una mejor representación se toman 6 clúster.

En la Figura 6-19 se presentan los dendrogramas para los conjuntos del NBI, PIB y el conjunto global. Por medio de este gráfico se ilustra el desarrollo de las subdivisiones o agrupamientos de los departamentos. Los departamentos se van conectando mediante enlaces donde la posición en el diagrama está determinada por el nivel de similitud entre las regiones. En el caso del dendrograma Global se presenta un primer grupo conformado por los departamentos de Vichada, Chocó, La Guajira, Guainía y Vaupés, que de manera general estas regiones comprenden fortalezas similares en ciertas actividades económicas, a su vez presentan altos puntajes en los indicadores de pobreza, por lo que está



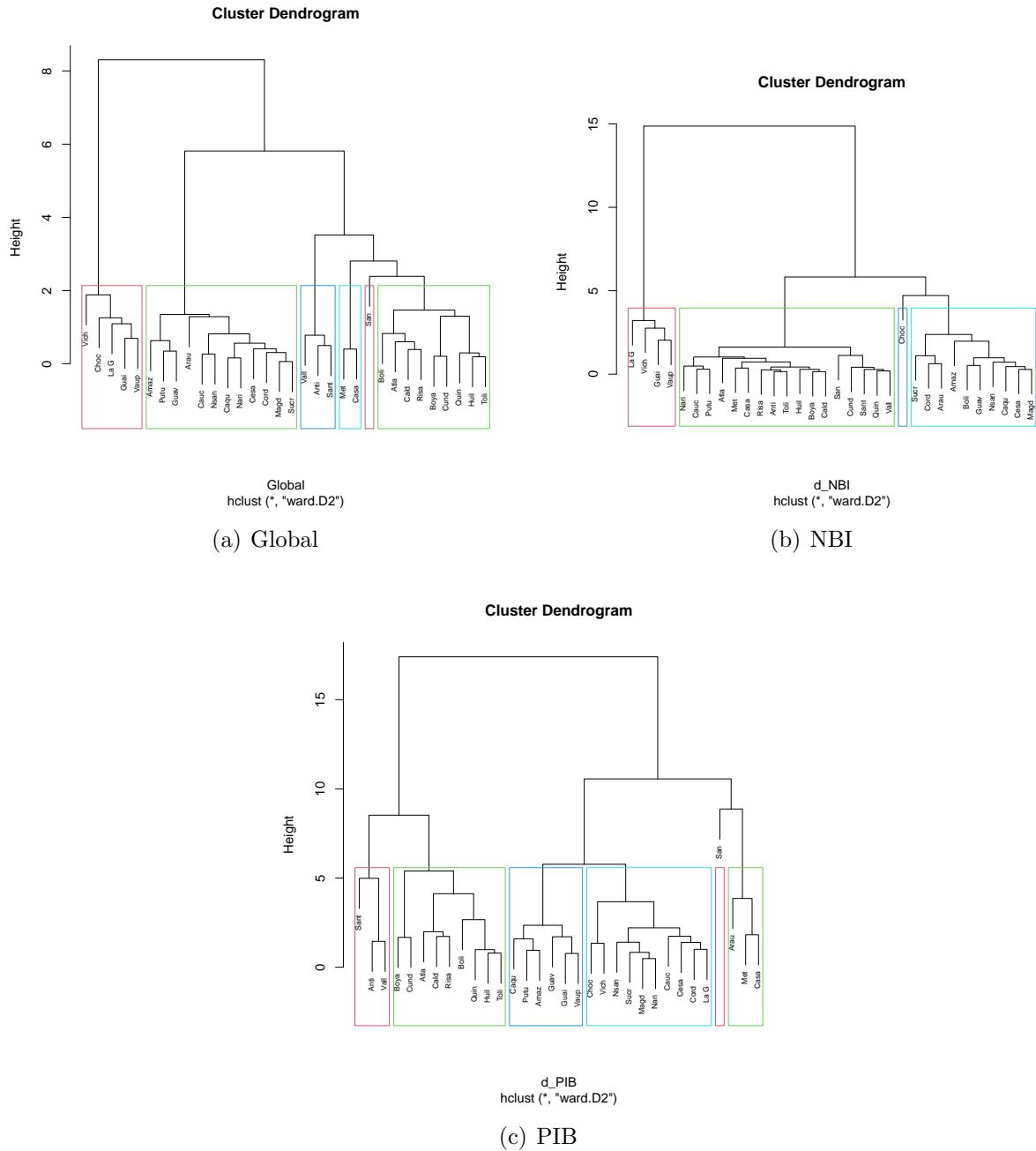
**Figura 6-18:** Óptimo Número de clústers

agrupación representa la alta similitud que tienen estos departamentos en los conjuntos de variables estudiados. Por otra parte, se tiene un grupo conformado por Valle del cauca, Antioquia y Santander, lo cual resulta esperado, dado que estos departamentos presentan elevadas cifras de valor agregado per capita en sectores como manufactura, actividades financieras y de seguros y sector inmobiliario. Otro grupo a destacar es el formado por los departamentos del Meta y Casanare los cuales fueron las regiones que destacaron por tener los valores agregados más altos en el sector minero, a su vez presentaron condiciones similares en indicadores de pobreza como problemas de vivienda y hacinamiento.

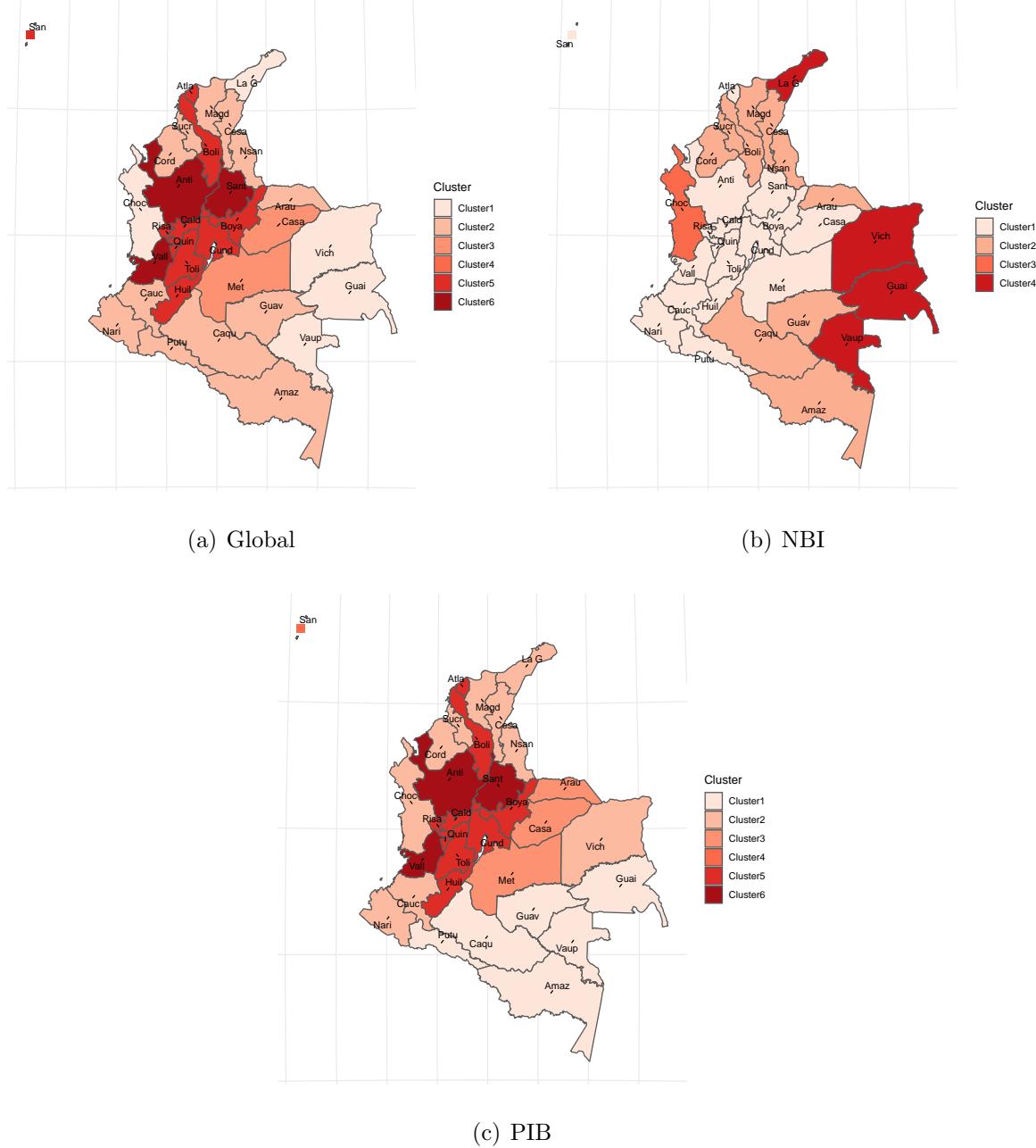
Se destaca la cualidad del dendrograma de poder visualizar por medio de los enlaces y subdivisiones que se forman, el poder identificar similitudes y diferencias representativas como es el caso del dendrograma del NBI, donde el departamento del Chocó se formó como un clúster único, y la ramificación en la que se encuentra indica lo diferente que resulta con otros departamentos y esto puede que se deba a la influencia que pueda presentar su fuerte problema con el abastecimiento de servicios públicos. En el grupo conformado por La guajira, Vichada, Guainia y Vaupes, se puede observar que dentro de este cluster se forma como un subgrupo conformado por Guania y Vaupes, los cuales se encuentran en una misma ramificación lo que indica que presentan unas condiciones muy similares en comparación con los otros dos departamentos pertenecientes al clúster.

Por parte del dendrograma del PIB, en el que se formaron seis agrupaciones, se destaca que los grupos formados comparten actividades económicas en los que son fuertes, como lo es el caso del clúster formado solo por el departamento de San Andrés, el cual destacó en la actividad económica de comercio donde presentó un valor agregado bastante alto, por lo que la ramificación en la que se encuentra en del dendrograma representa lo diferente que es con respecto a los demás departamentos. El clúster conformado por los departamentos de

Arauca, Meta y Casanare, estos departamentos como se había mencionado antes presentan fortaleza en las actividades económicas de la minería y la agricultura, pero como se puede observar Meta y Casanare se encuentran en una misma ramificación y esto se debe a que estas regiones presentan valores agregados muy similares en los sectores económicos mencionados.



**Figura 6-19:** Dendrorama Global, NBI y el PIB



**Figura 6-20:** Ubicación espacial de los Clúster de Departamentos

En la Figura 6-20 se presentan los mapas correspondientes a los grupos que se formaron en los dendrogramas, esto con el objetivo de visualizar como se encuentran distribuidos espacialmente los clúster. Cabe aclarar que para la construcción de los mapas se tomó la primera coordenada de los individuos (departamentos) del AFM, seguidamente por cada clúster que se formó, se sumó el valor de cada coordenada y se dividió por la cantidad de individuos pertenecientes a ese grupo. Esto se hizo con el objetivo de dar un puntaje a los

individuos de cada grupo y de esta manera poder visualizar en un mapa de calor, cuales grupos de departamentos resultan con puntajes altos y cuales con puntajes bajos. Es decir, entre más rojo sea el color asignado, quiere decir que su puntuación fue alta, y entre más claro sea la tonalidad será la puntuación más baja.

Como se puede observar el mapa de cluster del NBI cuenta con los cuatro clúster que se presentaron en el dendrograma, donde la agrupación de los departamentos Vichada, Guainía, Vaupes y La Guajira, presentaron la puntuación más alta, lo cual tiene sentido, dado que estos departamentos presentaron fuertes problemas en vivienda, en abastecimiento de servicios públicos, hacinamiento, inasistencia escolar y dependencia económica. Seguidamente el segundo cluster con el mayor puntaje está dado por el Chocó el cual presento tener fuertes problemas con el abastecimiento de servicios públicos. De manera general se puede observar como los diferentes grupos que se formaron y que presentan condiciones similares de pobreza se encuentran espacialmente cerca, como en el caso de los departamentos que presentaron una puntuación baja como lo fue el primer cluster, conformado por departamentos que presentan ciertamente valores agregados altos en varios de los sectores económicos. Por lo que se puede percibir que si hay una fuerte influencia en la espacialidad.

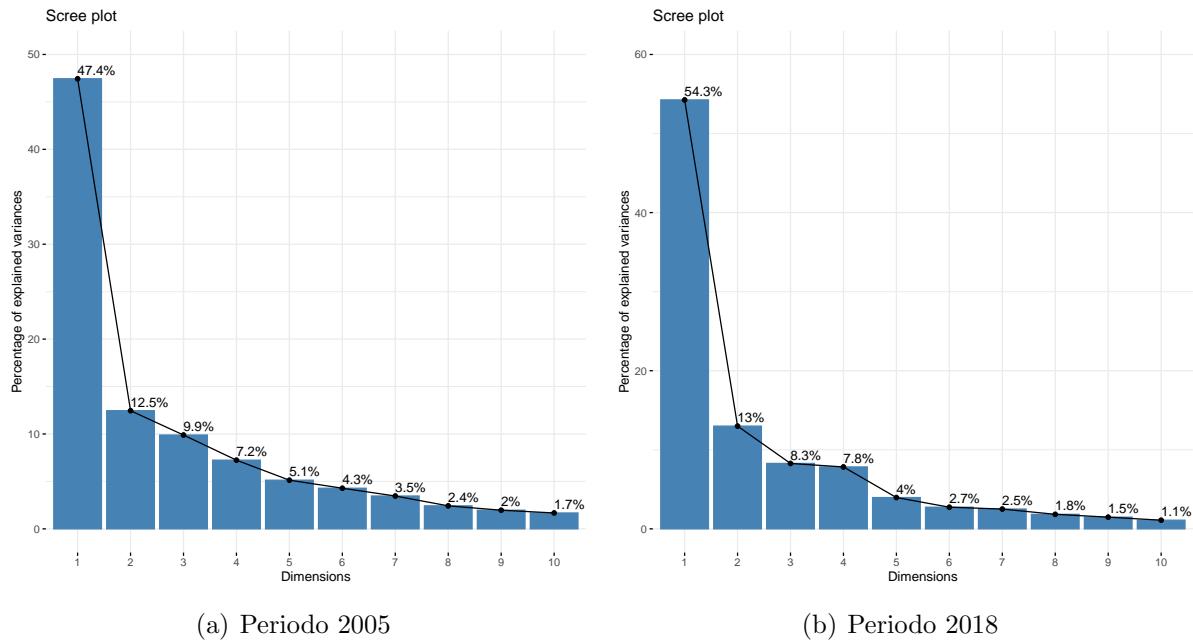
Con respecto a la distribución espacial de los clúster del PIB, se puede destacar que las actividades económicas en las que son fuertes los departamentos se ven ciertamente influenciadas por los departamentos que se tengan como vecinos, tal es el caso del clúster conformado por Arauca, Casanare y Meta, los cuales una de sus principales actividades económicas es la ganadería al igual que la extracción de petróleo. Por otra parte se encuentra el clúster 6 el cual obtuvo la puntuación más alta, y está conformado por los departamentos de Valle del Cauca, Antioquia y Santander, los cuales son regiones con alta productividad en varios sectores económicos, lo cual se llega a la misma deducción sobre como el componente espacial influye en el comportamiento y condiciones de los departamentos.

## 6.6. Comparación con el año 2005

Se debe recordar que los indicadores de pobreza son obtenidos a través del censo poblacional y las cuentas nacionales se obtienen de forma anual y se consideró el año 2018 como periodo referencia para realizar este estudio, dado que este periodo presenta el último censo realizado. Por lo tanto, el censo anterior fue desarrollado en el año 2005. El objetivo de esta sección es realizar una comparación descriptiva de que tanto ha cambiado el comportamiento de las variables de pobreza y de actividad económica en los departamentos del país en las dos estancias de tiempo.

Dicho lo anterior, se realiza un AFM a la tabla de datos compuesta por el conjunto

de variables del PIB y el NBI para el periodo 2005. De ese análisis factorial se presenta el gráfico de sedimentación para visualizar el porcentaje de varianza explicada que acumulan sus ejes factoriales y así comparar con el periodo 2018.



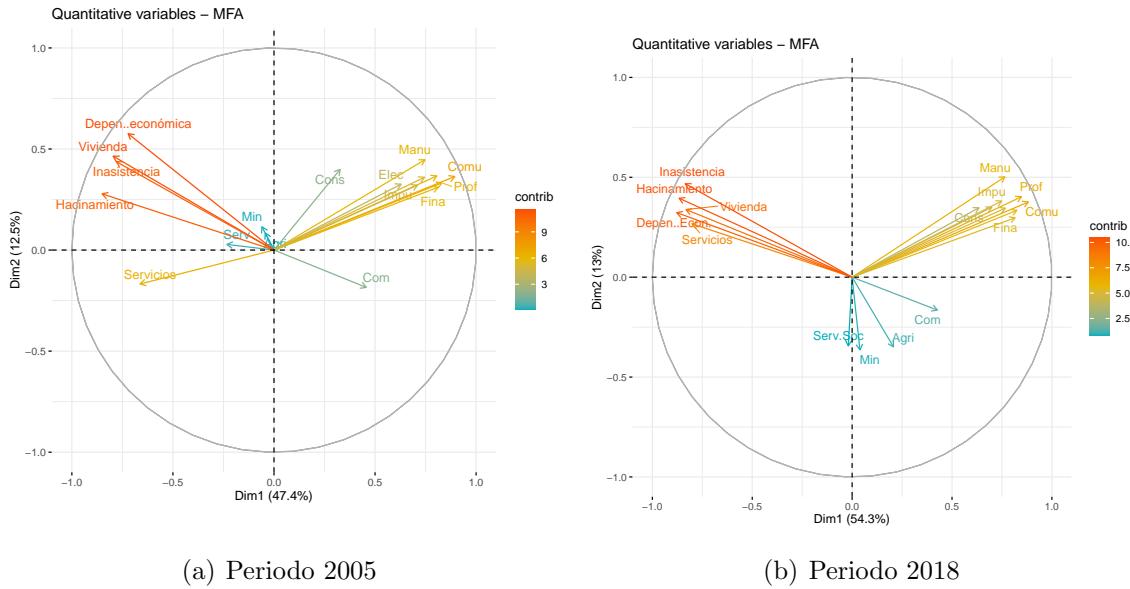
**Figura 6-21:** Comparación de Varianza explicada periodo 2005 y 2018

En la Figura 6-21, se puede observar que los dos primeros ejes factoriales para el periodo 2005 acumulan el 59.9 % de la variabilidad de la tabla de datos, donde se destaca el primer eje factorial el cual aporta el 47.4 % de la inercia total y el segundo eje aporta el el 12.5% al plano factorial. Que en rasgos generales la sedimentación para ambos periodos no se encuentra tan diferente dado que ambos casos el primer plano factorial es el que más recoge la variabilidad de los datos.

Por otra parte, se presenta el circulo de correlaciones para ambos periodos en la Figura 6-22, con el objetivo de observar el comportamiento de las variables y sus contribuciones a los ejes. Por lo anterior, Minería, Agricultura, comercio y servicios, para ambos periodos se siguen presentando como variables que tienen un aporte económico pero no contribuyen a modificar las condiciones de pobreza, es decir, no generan algún impacto social.

Para la comparación de los periodos 2005 y 2018 se realizó un análisis de cluster para el periodo 2005 de la misma forma en que se realizó para el periodo 2018, es decir que se realizó a partir del AFM, por lo tanto se realiza el análisis para las coordenadas globales y para cada uno de los conjuntos de datos (PIB y NBI).

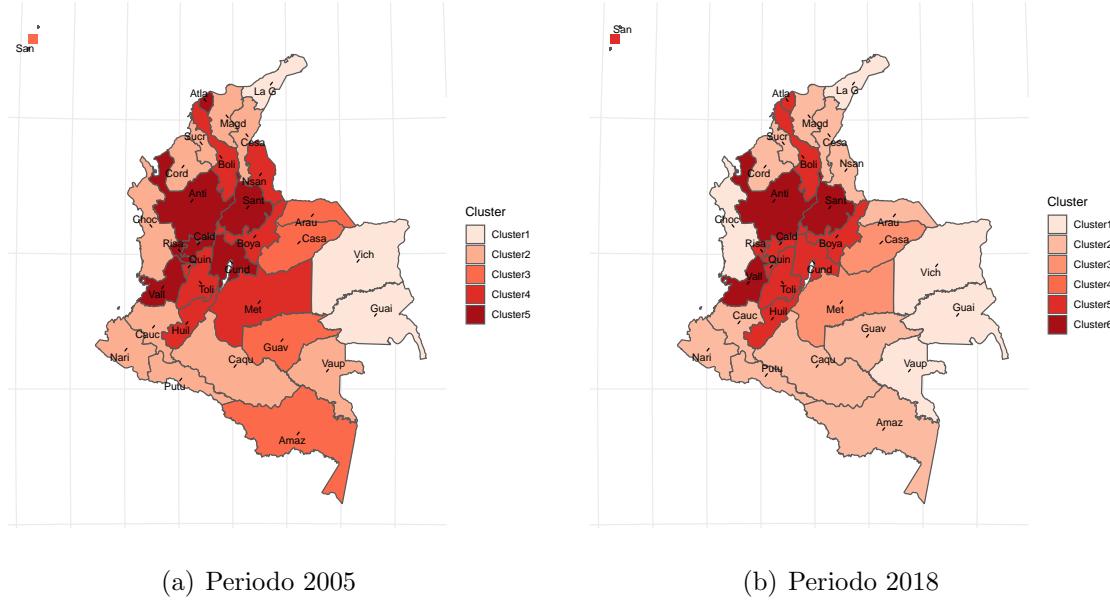
En la Figura 6-23 se presenta los clusters conformados para ambos periodos, en el



**Figura 6-22:** Comparación Círculo de correlaciones periodo 2005 y 2018

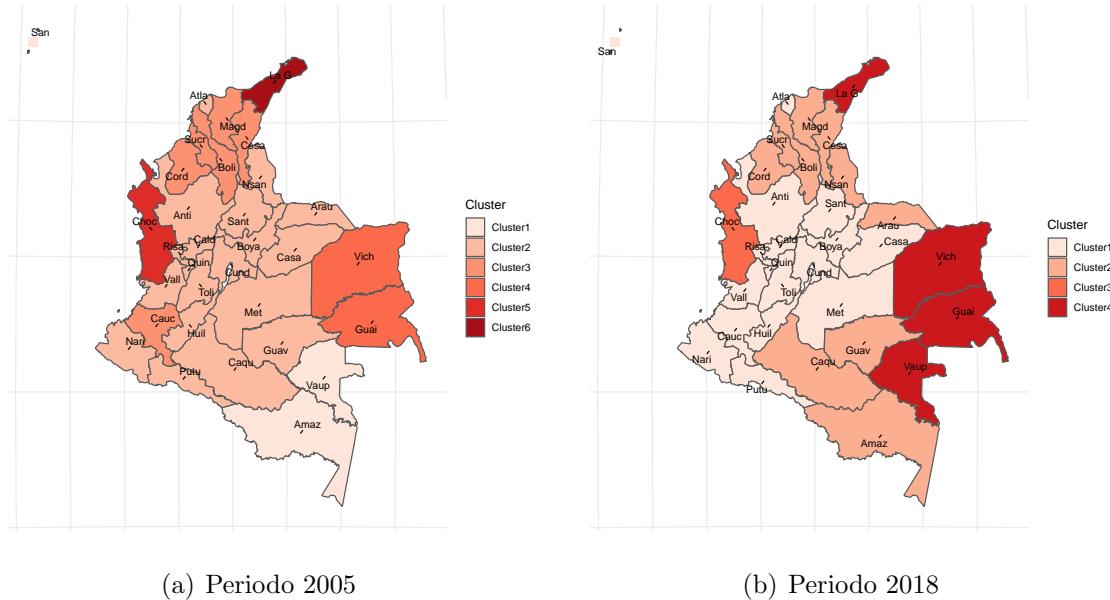
conjunto de coordenadas globales, en lo que se destaca que, para el año 2005 se conformaron 5 clúster, donde el cluster 5 que es el que presenta mayor puntuación se encuentra conformado por los departamentos de Antioquia,Santander, Cundinamarca, Valle del cauca, Risaralda y Cordoba,que a diferencia del año 2018 el cluster que mayor puntuación obtuvo se conformó solo por los departamentos de Valle del cauca, Antioquia y Santander, dado que los departamentos por los que se diferencian estos cluster, se deba más por parte de la productividad, es decir, que estas regiones no vayan a un mismo ritmo de crecimiento económico por lo que se presente las diferencias y pertenezcan a otra agrupación. De manera general se puede apreciar una mayor uniformidad espacial en las agrupaciones que se formaron para el periodo 2018 en comparación con el 2005. Es decir, para el 2018 se presentan grupos de departamentos con similares condiciones tanto de pobreza como de actividad económica que espacialmente se encuentran fronterizos(vecinos).

En la Figura 6-24 se presentan los clúster conformados para el conjunto de datos del NBI, donde se puede observar diferencias representativas entre el periodo del 2005 y 2018. En el 2005 el departamento de la Guajira conformaba el clúster con mayor puntuación, pues para este periodo esta región era la que presentaba mayores problemas en vivienda, así mismo con el clúster conformado por el chocó donde este departamento presentó un valor muy elevado por problemas de abastecimiento de los servicios públicos. Para el año 2018 el departamento del Chocó volvió a conformar un clúster por el mismo problema, que a pesar de que la cifra disminuyó se sigue presentando como la necesidad más fuerte de la región en comparación a los demás departamentos. Departamentos como Vichada y Guainía siguen presentando problemas fuertes en inasistencia escolar y dependencia económica, además de

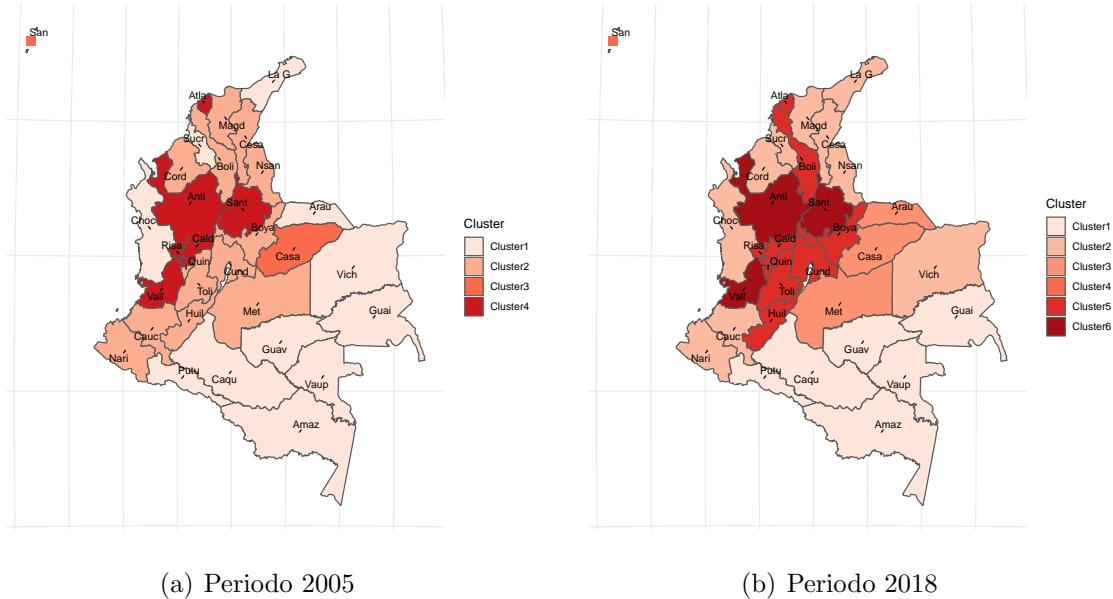


**Figura 6-23:** Comparación ubicación espacial de clústers global periodo 2005 y 2018

que incrementó el problema de vivienda y el de abastecimiento de servicios públicos.



**Figura 6-24:** Comparación ubicación espacial de clústers NBI periodo 2005 y 2018



**Figura 6-25:** Comparación ubicación espacial de clústers PIB periodo 2005 y 2018

La Figura 6-25 presenta los clúster conformados para el conjunto de datos del PIB, la diferencia se destaca en que se logra visualizar que algunos departamentos tuvieron un crecimiento económico, dando paso a nuevas agrupaciones que se presentan para el periodo 2018, como lo fue en el caso del cluster conformado por los departamentos del Huila, Cundinamarca, Boyaca, Bolívar y Risaralda, que para el año 2005 pertenecían a una agrupación más grande. Los departamentos como Casanare, Meta y Arauca para el periodo 2005 se habían agrupado en clúster diferentes, para el 2018 se agruparon en un mismo clúster, esto debido al incremento del valor agregado a actividades en las que comparten fortaleza como lo son en el sector agrario y el sector minero.

De manera general, en esta sección se logró ver los diferentes cambios que se presentaron en el comportamiento de las variables y en la conformación de Clúster, donde se puede destacar que hay departamentos que siguen presentando o han aumentado los indicadores de pobreza, como lo es en el caso del Chocó, La guajira, Vichada, Vaupes y Guiania. Se destaca que los departamentos como Antioquia, Valle del cauca y Santander, se mantienen como regiones de alta productividad. Además de identificar cambios entre los dos períodos, se puede decir que la influencia espacial si afecta el comportamiento de los departamentos, es decir, que el incremento de la actividad económica y la disminución o incremento de los indicadores de pobreza puede verse influenciada por los departamentos vecinos.

# **7 Conclusiones y recomendaciones**

## **7.1. Conclusiones**

Mediante este estudio fue posible realizar la caracterización conjunta de los departamentos respecto a sus condiciones de productividad económica e identificar la relación existente con la de pobreza. Las estadísticas de crecimiento económico de Colombia en los últimos años posicionan al país como una economía emergente, sin embargo el análisis interno deja ver condiciones preocupantes de desigualdad en el crecimiento. Este análisis permitió observar que el crecimiento económico del país se está focalizando en algunos departamentos estratégicamente ubicados espacialmente, conectados entre sí, como lo es la Región Andina, de manera que han podido fortalecer su sector industrial, de comercio y de servicios, obteniendo con ello mayor productividad y bajos niveles de pobreza en su población.

Por otra parte, el estudio deja ver un conjunto de departamentos con menos capacidad productiva, que fundamenta su economía en la agricultura y la ganadería pero conviven con altos niveles de pobreza. También refleja como algunos departamentos con productividad media se ve altamente afectada por sus problemas extremos de pobreza. Tal caso es el de La Guajira quien a nivel general muestra tener una economía con una productividad media, pero se ve afectada por los altos niveles de pobreza.

Los clúster resultantes pueden resultar de interés para dirigir las políticas públicas y evaluar el efecto que sobre la pobreza puede llegar a tener el desarrollo de la productividad de las regiones. En el estudio se pudo observar que la productividad minera no contribuye al mejoramiento de las condiciones sociales de la comunidad, al menos no para el año de referencia del estudio. Probablemente se hace necesario dar una revisión a las políticas de reasignación de estos recursos en el beneficio del global de la comunidad. Por medio de estas agrupaciones se logró identificar que la condición espacial en la que se encuentren los departamentos puede afectar su desarrollo tanto en crecimiento económico como las condiciones de vida de la población de la región.

También se observó que algunas de la necesidad básicas de la población pueden llegar a ser mitigadas en la medida en que se fortalece la industria y el sector profesional, lo cual requiere de un proceso de capacitación del recurso humano, favorecimiento de la innovación de base tecnológica y ubicación de recursos de inversión. Como es en el caso de la inasistencia

escolar que puede ser mitigada en planes de educación que puede ser mitigada por el sector profesional. Por otra parte, la dificultad de acceso que presentan algunos departamentos, los problemas de comunicación y transporte deberán ser disminuidos para contribuir a la disminución de los niveles de pobreza que presentan algunas regiones.

Por medio de la comparación que se realizó con el periodo 2005 con respecto al año 2018, se logra ver un panorama de como evolucionaron los departamentos tanto en la productividad económica y en sus condiciones de pobreza. Por lo que se concluye que a pesar de la diferencia de tiempo de ambos periodos, se sigue presentando la tendencia de algunos departamentos con sus extremas condiciones de pobreza, a pesar de que si se observa un claro crecimiento de la economía a nivel nacional, donde cada sector económico incremento su valor agregado al aporte del PIB. Por lo que se destaca que el crecimiento económico no refleja proporcionalmente una buena calidad de vida de su población, pero es posible que esta contribuya a reducir los niveles de pobreza que se presentan en ciertas regiones. Con la identificación de los departamentos con altos niveles de pobreza y a su vez baja productividad en gran parte de los sectores económicos para ambos periodos, como lo fueron Vichada, Vaupes, Guainía, La Guajira y Chocó, son regiones que se han mantenido en iguales condiciones a pesar de la diferencia de tiempo, por tanto, son departamentos que deben ser priorizados en planes de mejoramiento en sus condiciones de calidad de vida de su población.

## **7.2. Recomendaciones**

En el desarrollo de este trabajo se logra caracterizar la relación existente entre el conjunto de variables de pobreza y de productividad, pero el análisis se podría enriquecer más, aprovechando y explotando las propiedades de técnicas como lo es el AFM para el Análisis de Múltiples Tablas, por lo tanto, se podrían agregar nuevas variables que aporten a la caracterización de los departamentos y así identificar patrones más detallados sobre los altos niveles de pobreza en ciertas regiones. Un ejemplo podría ser incorporar variables que se encuentren relacionadas con el conflicto armado que presenta el país, se es de conocimiento que este problema afecta a ciertos departamentos, por lo que sería interesante identificar que tanto afecta el conflicto armado a que la región no tenga una alta productividad y a su vez refleje altos niveles de pobreza. Otras variables que se podrían incluir serían la morbilidad y la mortalidad, las cuales se esperaría que dieran un panorama sobre el estado de salud y bienestar de la población. Es un estudio que puede seguir enriqueciéndose y de esta forma permita dar un panorama sobre las condiciones de los departamentos y así el gobierno tome la decisión de focalizar inversión de recursos en zonas específicas que contribuyan al mejoramiento de la calidad de vida de la población.

# Bibliografía

- [1] Abigail, O., Quintero, M. L., and Aurea, H. (2011). Evolución del concepto de pobreza y el enfoque multidimensional para su estudio. *Quivera. Revista de Estudios Territoriales*, 13:207–219.
- [2] Aponte, C., Aroca, E., and Guzman, L. (2016). Análisis de datos espaciales del Índice de Necesidades Básicas Insatisfechas en la Región Andina. *Perspectiva Geográfica*, 20:391–418.
- [3] Arnedo, V. M. (2013). Crecimiento económico y desarrollo humano en Colombia (2000 – 2010). *Revista de economía del caribe*, 2106:127–143.
- [4] Bonilla, J. and Ruiz, J. (2018). Análisis multivariante de datos en tres vías para caracterizar la dinámica económica de los departamentos en Colombia en el periodo 2000-2015. *Trabajo de grado-Univesidad del valle*.
- [5] Corrales, M. (2010). Combinación del análisis factorial múltiple y el análisis armónico cualitativo en la descripción de datos longitudinales cualitativos. *Trabajo de grado-Universidad Nacional de Colombia*, 8(1):11–32.
- [6] DANE (2005). Boletín NBI, 2005. Technical report.
- [7] DANE (2012). Resolución Número 066. Technical report.
- [8] DANE (2019a). Boletín técnico pobreza multidimensional en Colombia. Technical report.
- [9] DANE (2019b). Necesidades basicas. Technical report.
- [10] DANE (2019c). Pobreza monetaria en Colombia, año 2018. Technical report.
- [11] DANE (2020). Clasificación Industrial Internacional Uniforme de todas las actividades económicas-Revisón 4 adaptada para Colombia. *Clasificación Industrial Internacional Uniforme*, pages 2–693.
- [12] De la Fuente, S. (2008). Análisis de Conglomerados. *Universidad Autonoma de Madrid*, page 14.

- [13] Elena, A. F., Karmele, F. A., and M. Isabel, L. C. (2001). Técnicas Factoriales de Análisis de Tablas Múltiples: Nuevos Desarrollos Empíricos. *Documentos de Trabajo BILTOKI*, (6):1.
- [14] Escofier, B. and Pagés, J. (1992). *Análisis factoriales simples y múltiples: objetivos, métodos e interpretación*. Universidad del País Vasco, España.
- [15] Garcia, C. (2021). Análisis Factorial Múltiple para describir las condiciones de salud sentidas de la población priorizada de la ciudad de Cali en el año 2018. *Trabajo de grado-Universidad del valle*.
- [16] González Rojas, V. M. (2014). *Análisis conjunto de múltiples tablas de datos mixtos mediante PLS*. PhD thesis.
- [17] Lebart, L., Morineau, A., and Piron, M. (1995). *Statistique exploratoire multidimensionnelle*.
- [18] Mancero, X. and Feres, J. C. (2001). El metodo de las necesidades basicas insatisfechas (NBI) y sus aplicaciones en America Latina. *CEPAL - Serie estudios estadísticos y prospectivos*, 8(7):52.
- [19] Mankiw, N. G. (1998). *Principios de Ecocomía*.
- [20] Marín, J. (2009). Análisis de cluster y árboles de clasificación. In *Springer*, pages 1–21.
- [21] Marquez, C. (2017). Modelo de Regresión PLS. *Trabajo de grado-Universidad de sevilla*. [Morales] Morales, F. C. Pobreza.
- [23] Osorio, M., Rodríguez, N., and Salamanca, G. (2017). Técnicas de análisis multivariado, modelamiento factorial múltiple y PLS-PATH para estudio y clasificación de tipos de mieles venezolanas de los estados de Lara y Yaracuy. *Ciencia en Desarrollo*, 8(1):129–143.
- [24] Pagès, J. and Tenenhaus, M. (2001). Multiple factor analysis combined with PLS path modelling. Application to the analysis of relationships between physicochemical variables, sensory profiles and hedonic judgements. *Chemometrics and Intelligent Laboratory Systems*, 58:261–273.
- [25] Pardo, C. (2011). *Métodos en ejes principales para tablas de contingencia con estructuras de partición en filas y columnas*. PhD thesis, Universidad NACIONAL de Colombia.
- [26] Peña, C. (2015). Regresión por Mínimos Cuadrados Parciales PLS Aplicada a Datos.
- [27] Pérez, G. (2005). Dimensión espacial de la pobreza en Colombia. *Revista ESPE - Ensayos Sobre Política Económica, Banco de la República*, (48):234–293.

- [28] Valdéz, D. (2010). Regresión por Mínimos Cuadrados Parciales. *Revista Varianza*, 7:18–22.
- [29] Visbal-Cadavid, D., Martínez-Gómez, M., and Escoria-Caballero, R. (2020). Exploring university performance through multiple factor analysis: A case study. *Sustainability (Switzerland)*, 12(3):1–24.
- [30] Vtelleschi, M. S. and Chavasa, V. (2016). Análisis factorial múltiple para la caracterización de variedades de trigo pan en diferentes ambientes. *Revista FABICIB*, 19:113–120.