



# **Análisis multivariado de datos para la evaluación de la dinámica de la pobreza monetaria en Colombia por departamentos durante el período 2012-2020.**

**Andrés Felipe Plaza Hernández**

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

# Análisis multivariado de datos para la evaluación de la dinámica de la pobreza monetaria en Colombia por departamentos durante el período 2012-2020.

Andrés Felipe Plaza Hernández

Trabajo de grado

Director:

PhD. Jaime Mosquera

Codirector:

MSc. Andrés Felipe Ochoa

Universidad del Valle

Facultad de Ingeniería, Escuela de Estadística

Santiago de Cali, Colombia

2022

## Dedicatoria

Dedico este trabajo de grado a mis padres, quienes me apoyaron incondicionalmente en los momentos más difíciles. Sin ellos este trabajo no sería una realidad.

A mi hermana, quien ha sido voz de sabiduría. A todos los profesores que me guiaron en este largo camino y me dieron muchas oportunidades. A todos ellos mi admiración.

## Agradecimientos

Agradezco al profesor Jaime Mosquera Restrepo, por guiarme con sus aportes académicos y en general durante todo el pregrado, a convertirme en una mejor persona.

Gracias a la Universidad del Valle, a todos los profesores de la Escuela de Estadística, con quienes tuve la gran fortuna de coincidir en este largo camino que es la vida, son todos excelentes personas. También quiero agradecer al profesor Andrés Felipe Ochoa, una excelente persona, con sus conocimientos académicos fue una guía importante para poder graduarme como profesional. Al profesor Rafael Tovar Cuevas, por su orientación durante estos años de estudio. A mis familiares quienes siempre confiaron y fomentaron el deseo de superación y éxito. A mi madre, quien a pesar de la distancia siempre la llevo en mi alma. A mi padre, por haberme forjado como la persona que soy en la actualidad.

Agradezco a mis amigos más cercanos, aquellos quienes estuvieron apoyándome en el curso de mi vida. A todos ellos gracias por su sincera amistad.

## Resumen

La pobreza monetaria comprende la capacidad de adquisición de bienes y servicios básicos necesarios para que un hogar persista. Desde un enfoque académico, el análisis de la pobreza permite el entendimiento y la formulación de políticas sostenibles que apunten a la reducción de este fenómeno. En dicha línea, el presente trabajo de grado caracteriza la dinámica de la pobreza monetaria en 24 departamentos de Colombia entre 2012 y 2020. Para tal caracterización se recurre a tres métodos estadísticos: Análisis Exploratorio de Datos, Análisis Factorial Múltiple y Análisis de Conglomerados. Los datos fueron obtenidos de la página oficial del Departamento Nacional de Estadística, institución que hace seguimiento anual al desarrollo económico y social de los departamentos. Consecuentemente, los resultados del estudio mostraron que los departamentos más pobres están ubicados en las zonas más alejadas del distrito capital, como el Pacífico y el Caribe. También se presentaron departamentos que se han mantenido a través del tiempo en situación de pobreza monetaria; y finalmente, otros han logrado hacer una transición para reducir considerablemente la incidencia de la pobreza monetaria en sus territorios.

**Palabras clave:** Pobreza Monetaria, Colombia, Análisis Cluster, Análisis Factorial Múltiple.

## Abstract

Monetary poverty includes the ability to acquire basic goods and services necessary for a household to persist. From an academic approach, the analysis of poverty allows the understanding and formulation of sustainable policies aimed at reducing this phenomenon. In this line, the present degree work characterizes the dynamics of monetary poverty in 24 departments of Colombia between 2012 and 2020. For such characterization, three statistical methods are used: exploratory data analysis, multiple factor analysis and cluster analysis. The data was obtained from the official website of the National Department of Statistics, which annually monitors the departments. Consequently, the results of the study showed that the poorest departments are located in the areas furthest from the capital district, such as the Pacific and the Caribbean. Departments that have remained in a situation of monetary poverty over time were also presented; and finally, others have managed to make a transition to considerably reduce the incidence of monetary poverty in their territories.

**Keywords:** monetary poverty, cluster analysis, multiple factor analysis.

# Contenido

<b>Resumen</b>	<b>v</b>
<b>Lista de figuras</b>	<b>vii</b>
<b>Lista de tablas</b>	<b>ix</b>
<b>1. Introducción</b>	<b>2</b>
<b>2. Planteamiento del problema</b>	<b>4</b>
2.1. Objetivos . . . . .	5
2.1.1. Objetivo general . . . . .	5
2.1.2. Objetivos específicos . . . . .	5
2.2. Justificación . . . . .	6
<b>3. Antecedentes</b>	<b>8</b>
3.1. Antecedentes sociales . . . . .	8
3.2. Antecedentes estadísticos . . . . .	9
<b>4. Marco Teórico</b>	<b>11</b>
4.1. Marco conceptual . . . . .	11
4.1.1. Cálculo de los indicadores de Pobreza Monetaria . . . . .	12
4.2. Marco teórico estadístico . . . . .	13
4.2.1. Análisis de Componentes Principales - ACP . . . . .	13
4.2.2. Análisis de la proximidad entre los individuos . . . . .	14
4.2.3. Análisis de la proximidad entre las variables . . . . .	16
4.2.4. Representación Simultánea . . . . .	17
4.2.5. Elementos de ayuda a la interpretación . . . . .	18
4.3. Análisis Factorial Múltiple - AFM . . . . .	19
4.3.1. Tabla de datos y notación . . . . .	19
4.3.2. Etapas del AFM . . . . .	20
4.3.3. Ponderación de los grupos de variables . . . . .	21
4.3.4. Representación de los grupos de variables . . . . .	21
4.3.5. Representación de las variables . . . . .	23
4.3.6. Representación de los individuos . . . . .	24
4.3.7. Representación superpuesta . . . . .	25

4.3.8.	Ayudas para la interpretación . . . . .	26
4.4.	Análisis de Conglomerados - AC . . . . .	26
4.4.1.	Similitud entre individuos . . . . .	28
4.4.2.	Similitud entre grupos de individuos . . . . .	29
4.4.3.	Construcción de una jerarquía indiciada . . . . .	29
4.4.4.	Jerarquía y partición . . . . .	30
4.4.5.	Método de Ward . . . . .	31
4.4.6.	Análisis de una jerarquía, elección de una partición . . . . .	33
<b>5.</b>	<b>Metodología</b>	<b>34</b>
5.1.	Estructura de los Datos . . . . .	34
5.2.	Procesamiento de datos . . . . .	35
5.3.	Software . . . . .	38
<b>6.</b>	<b>Resultados</b>	<b>39</b>
6.1.	Análisis exploratorio . . . . .	39
6.1.1.	Pobreza Monetaria o Incidencia . . . . .	39
6.1.2.	Línea o umbral de Pobreza Monetaria . . . . .	41
6.1.3.	Evolución de la Pobreza Monetaria . . . . .	42
6.1.4.	Evolución del número de personas en Pobreza Monetaria . . . . .	44
6.1.5.	Coefficiente de Gini . . . . .	45
6.1.6.	Severidad de la Pobreza Monetaria . . . . .	46
6.1.7.	Evolución de la Brecha de la Pobreza Monetaria . . . . .	48
6.2.	Análisis Factorial Múltiple (AFM) . . . . .	50
6.2.1.	Análisis de factores . . . . .	50
6.2.2.	Análisis global de las variables . . . . .	51
6.2.3.	Contribuciones de las variables . . . . .	51
6.2.4.	La nube de individuos . . . . .	52
6.2.5.	Calidad de la representación . . . . .	53
6.2.6.	Contribuciones de los individuos . . . . .	54
6.2.7.	Análisis grupal . . . . .	55
6.3.	Análisis de Conglomerados - AC . . . . .	58
<b>7.</b>	<b>Conclusiones y recomendaciones</b>	<b>62</b>
7.1.	Conclusiones . . . . .	62
7.2.	Recomendaciones . . . . .	62
<b>A.</b>	<b>Anexo: Nombrar el anexo A de acuerdo con su contenido</b>	<b>64</b>
	<b>Bibliografía</b>	<b>65</b>

# Lista de Figuras

4-1. Representación gráfica de los datos en el espacio original de las variables (gráfico izquierdo) y en el espacio de las componentes principales (gráfico derecho). Fuente: Elaboración propia . . . . .	16
4-2. Círculo de correlación en plano F1 y F2 del ACP. Fuente: Elaboración propia. . . . .	17
4-3. Representación simultánea de individuos y variables en el ACP. Fuente: Elaboración propia. . . . .	18
4-4. Tabla de datos en AFM. Fuente: Escofier y Pagés (1992) . . . . .	19
4-5. Ejemplo hipotético de la representación de los grupos de variables en el AFM. Fuente: Elaboración propia. . . . .	22
4-6. Ejemplo hipotético de representación de las variables sobre los ejes factoriales F1 y F2 en AFM. Fuente: Elaboración propia. . . . .	23
4-7. Ejemplo hipotético de representación de los individuos de variables en el plano F1 y F2 del AFM. Fuente: Elaboración propia. . . . .	24
4-8. Ejemplo hipotético de representación superpuesta de la nube de individuos en el plano F1 y F2 en AFM. Fuente: Elaboración propia. . . . .	25
4-9. Ejemplo de árbol jerárquico (dendrograma). Fuente: Lê y Pagès (2009) . . . .	27
4-10. Salto mínimo y diámetro entre dos grupos de individuos (identificados por símbolos diferentes). Fuente: Lê y Pagès (2009). . . . .	29
4-11. Etapas de la construcción de un árbol jerárquico a partir de seis individuos repartidos sobre un plano. Fuente: Lê y Pagès (2009) . . . . .	30
4-12. Jerarquía y partición. Fuente: Lê y Pagès (2009) . . . . .	31
5-1. Fases en la metodología del análisis de datos . . . . .	34
5-2. Estructura de los indicadores de Pobreza Monetaria por departamentos. Fuente: DANE (2021) . . . . .	35
6-1. Líneas de Pobreza Monetaria Extrema por departamentos en el año 2020 (Cifras en COP). . . . .	42
6-2. Incidencia de la Pobreza Monetaria por departamentos, desde el año 2012 hasta el 2020. . . . .	43
6-3. Incidencia de la Pobreza Monetaria por departamentos, desde el año 2012 hasta el 2020. . . . .	43
6-4. Número de personas en Pobreza Monetaria . . . . .	44
6-5. Número de personas en Pobreza Monetaria. . . . .	45



<b>6-6.</b> Coeficiente de Gini 2012- 2020. . . . .	46
<b>6-7.</b> Severidad de la Pobreza Monetaria. . . . .	47
<b>6-8.</b> Severidad de la Pobreza Monetaria. . . . .	48
<b>6-9.</b> Brecha de la Pobreza Monetaria. . . . .	49
<b>6-10.</b> Brecha de la Pobreza Monetaria. . . . .	49
<b>6-11.</b> Porcentaje de varianza explicado. . . . .	50
<b>6-12.</b> Círculo de correlaciones del AFM, Planos F1 y F2. . . . .	51
<b>6-13.</b> Contribuciones de las variables al primer y segundo eje factorial del AFM . .	52
<b>6-14.</b> Representación global de los departamentos en el primer y segundo plano del AFM. . . . .	52
<b>6-15.</b> Cosenos cuadrados del AFM . . . . .	53
<b>6-16.</b> Contribuciones de los individuos al primer y segundo eje factorial del AFM .	54
<b>6-17.</b> Representación de los departamentos a nivel global y grupal del AFM . . . .	55
<b>6-18.</b> Representación de los grupos del AFM. . . . .	56
<b>6-19.</b> Proyección de las primeras componentes principales de cada grupo del AFM	57
<b>6-20.</b> Número de clúster óptimo . . . . .	59
<b>6-21.</b> Dendrograma sobre las coordenadas globales de los individuos del AFM . . .	59
<b>6-22.</b> Dendrograma sobre las coordenadas parciales de los individuos del AFM . .	60

# Lista de Tablas

5-1. Listado de indicadores de Pobreza Monetaria y etiquetas . . . . .	36
5-2. Estructura final de datos para el AFM. . . . .	37
6-1. Estadísticas descriptivas de los indicadores de Pobreza Monetaria para los años 2012, 2016 y 2020. . . . .	40
6-2. Valores propios del AFM . . . . .	50
6-3. Coeficientes Lg . . . . .	58
6-4. Coeficientes Rv . . . . .	58

# 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. He reconocido todos los pasajes que se han tomado de manera textual o figurativa de textos publicados y no publicados. Ninguna parte de este 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, 25 de marzo del 2022.

---

(Andrés Felipe Plaza Hernández)

# 1. Introducción

La pobreza es un fenómeno social que afecta a la mayoría de los países. Si bien su percepción varía dependiendo cada región, más aún si se tiene en cuenta las coyunturas particulares, se reconoce que la pobreza conlleva a problemáticas comunes tales como: la afectación de derechos humanos, el inaccess a sistemas de salud, la desnutrición, el analfabetismo, la violencia de género, (Laís y Pérez, 2018), entre otros. En sí, se puede afirmar que la pobreza afecta drásticamente la calidad de vida de los individuos y de los grupos sociales; por ello, es un fenómeno de interés tanto para las ciencias económicas como para la estadística.

En dicha línea, para estudiar la pobreza es importante antes que nada describir la naturaleza del fenómeno. De tal modo, cabe mencionar que hay múltiples formas de medir la pobreza. No obstante, en Colombia existen dos metodologías oficiales para analizarla: directa e indirecta. Según el DANE (2018), la primera metodología *“evalúa los resultados de insatisfacción (o privación) que tiene un individuo respecto a ciertas características que se consideran vitales como salud, educación, empleo, entre otras”*. La segunda metodología *“evalúa la capacidad de adquisición de bienes y servicios que tienen los hogares”* (DANE, 2018).

La metodología directa, también conocida como pobreza multidimensional, es medida en Colombia por el DANE con referencia a los datos recogidos en la Encuesta Nacional de Calidad de Vida. Este índice comprende cinco dimensiones tales como: condiciones educativas del hogar, condiciones de la niñez y la juventud, trabajo, salud y acceso a servicios públicos domiciliarios y condiciones de la vivienda para un total de 15 indicadores (DANE, 2018). La encuesta de calidad de vida se aplica anualmente, pero no goza de representatividad departamental ni de las grandes ciudades. Consecuentemente, esto implica una serie de limitaciones para los estudios de la pobreza multidimensional a nivel departamental.

Por su parte, la metodología indirecta es aplicada en el indicador de pobreza monetaria. Esta se mide en Colombia de forma anual, pero, a diferencia de la pobreza multidimensional, si goza de representatividad departamental. Dicho de forma sucinta, la pobreza monetaria identifica a los pobres a través de *“la comparación del ingreso per cápita de la unidad de gasto con el costo monetario de adquirir una canasta de bienes alimentarios y no alimentarios, a través de un umbral mínimo”* (DANE, 2018).

Ahora bien, para comprender la anterior definición a plenitud es necesario aclarar algunos

de sus atributos constitutivos: A) por Unidad de Gasto se entiende *“la persona que atiende sus propios gastos, o al grupo de personas que comparten la vivienda y tienen arreglos para satisfacer en común sus necesidades esenciales”*; B) Por ingreso per cápita de la unidad de gasto se entiende como *“el resultado de dividir el ingreso corriente disponible de la unidad de gasto entre el número de personas que conforman la unidad de gasto”*; C) Y finalmente por umbral mínimo se comprende la línea de pobreza, la cual es *“el costo mínimo necesario por persona para acceder a una canasta básica de bienes y servicios (alimentarios y no alimentarios) en un área geográfica determinada”* (DANE, 2018).

En síntesis, la pobreza monetaria mide si un hogar tiene los ingresos suficientes para acceder a una canasta mínima. Dado que en el presente trabajo de grado interesa caracterizar los departamentos en función de la pobreza, se opta por analizar específicamente esta, a través de los siguientes indicadores: Incidencia de la Pobreza Monetaria, Línea de pobreza monetaria, Coeficiente de Gini, Brecha de la Pobreza, Línea de Pobreza Monetaria Extrema, Incidencia de la Pobreza Monetaria Extrema, Brecha de la Pobreza Extrema, Severidad de la Pobreza. La definición formal de todos estos indicadores son definidos en el capítulo 5.

En dicha línea, el presente trabajo de grado caracteriza la dinámica de la pobreza monetaria en 24 departamentos de los que se tienen datos desde el año 2012 hasta el 2020. Para tal caracterización se recurre a tres métodos estadísticos: análisis exploratorio de datos, Análisis Factorial Múltiple que permite estudiar las relaciones entre los distintos años además de medir su grado de semejanza entre indicadores de pobreza y departamentos y el Análisis de Conglomerados que permite agrupar a los departamentos con similitud en su estructura de pobreza. Los datos fueron obtenidos de la página oficial del Departamento Nacional de Estadística. Consecuentemente, los resultados del estudio mostraron que los departamentos más pobres están ubicados en las zonas mas alejadas del distrito capital, como el pacífico y el Caribe; también se presentaron departamentos que se han mantenido a través del tiempo en situación de pobreza monetaria; y finalmente, otros han logrado hacer una transición para reducir considerablemente la incidencia de la pobreza monetaria en sus territorios.

## 2. Planteamiento del problema

De acuerdo a la Comisión Económica para América Latina (CEPAL), sostienen que *“la pobreza como la insuficiencia de recursos, es la carencia en el flujo de dinero que recibe un hogar y que expresa su capacidad para consumir bienes y servicios”* (CEPAL, 2018). Esta situación genera innumerables consecuencias en la población, entre ellas el aumento de la desigualdad social. Según (Stiglitz, 2016), *“la desigualdad resulta, en un desempeño económico más débil”*, además, *“la redistribución de los ingresos, excepto cuando es muy extrema, no perjudica el crecimiento económico”* (Ostry y Loungani, 2016) y que, por tanto, *“mayor igualdad y mejor crecimiento, son procesos correlacionados”* (ONU, 2018).

En dicha línea *“la desigualdad, no permite sustentar un crecimiento económico a un ritmo que proporcione: aumentar la productividad favoreciendo la armonización de las relaciones entre crecimiento, inversión y empleo, generar empleo de calidad, reducir la pobreza, y revertir los procesos de desigualdad de los ingresos monetarios”* (CEPAL, 2018). Además, *“la alta desigualdad predominante en América latina y el Caribe se refleja en un proceso de concentración de la riqueza que aumenta más cuando se consideran los activos financieros respecto a los activos físicos”*.(CEPAL, 2018)

Para enfrentar el fenómeno de la pobreza, cada región debe evaluar y generar lineamientos políticos locales, que les permitan hacer frente ante esta creciente situación; estos lineamientos se derivan producto del procesamiento de los datos que se convierte en información valiosa para la toma de decisiones. En el contexto colombiano, los datos son recolectados de manera anual y a nivel regional, grandes ciudades y departamentalmente mediante matrices de datos por cada año, estos son procesados por el DANE, principalmente de las Encuestas de Ingresos y Gastos de los hogares Colombianos y la gran Encuesta de Calidad de Vida entre otras. Según el DANE (2018), las fuentes de ingreso de los hogares son: ingresos laborales, arriendo imputado, pensiones, ayudas, renta de la propiedad y otros ingresos. *“En Colombia todas las fuentes de ingreso son inequitativas, unas más que otras, y algunas como las ayudas se aproximan lentamente a la equidad; otras como las pensiones o las rentas de la propiedad han sido crecientemente regresivas en su distribución”* (Acosta y Zúñiga, 2017). *“Esta conceptualización parte de asumir que la pobreza es una situación generada por acciones y omisiones de los estados, y de los distintos agentes económicos y sociales del sistema social”* (CEPAL, 2018).

Lo anterior configura un panorama para evaluar la dinámica de la pobreza en Colombia por departamentos y caracterizarlos en términos de: grupos de departamentos con igual estructura de pobreza, determinar correlaciones entre sus indicadores y evaluar trayectorias presentadas en el tiempo, todo esto en función de sus indicadores de pobreza monetaria. Para ello, metodologías como el Análisis Factorial Múltiple y Análisis de Conglomerados permitirán realizar dicha labor. Este tipo de análisis no se presentan en los boletines que expide el DANE anualmente, solo se presentan análisis univariados y bivariados, motivo por el cual esta investigación enriquece la caracterización y brinda un panorama más global de la pobreza en Colombia para los años 2012 a 2020, que permita generar mejores lineamientos que ayuden a disminuir el fenómeno de la pobreza. Por lo tanto, la pregunta de investigación es:

**¿Cómo es la dinámica de los indicadores de pobreza monetaria por departamentos en Colombia durante el periodo de tiempo 2012 a 2020?.**

## **2.1. Objetivos**

### **2.1.1. Objetivo general**

Caracterizar la dinámica de la pobreza monetaria de 24 departamentos de Colombia durante los años 2012 a 2020.

### **2.1.2. Objetivos específicos**

1. Describir la estructura de correlación que existe entre los indicadores que miden la pobreza monetaria.
2. Caracterizar los cambios temporales en la estructura de la pobreza monetaria para los departamentos de Colombia.
3. Agrupar los departamentos con similares estructuras de pobreza y estudiar sus cambios en el tiempo.

## 2.2. Justificación

En el mundo, desde el año 2015 existe un acuerdo por parte de los países líderes, en el que se contemplan los objetivos de desarrollo sostenible para los próximos 15 años y que todos los países son responsables a contribuir para lograrlo. El primero de estos objetivos es poner fin a toda forma de pobreza. Más de 700 millones de personas, el 10 % de la población mundial, viven en situación de pobreza extrema con dificultades para satisfacer sus necesidades más básicas (Naciones, 2015). El aumento de la desigualdad producto de la pobreza, según Stiglitz (2016) es *“perjudicial para el crecimiento económico y socava la cohesión social, lo que incrementa las tensiones políticas y sociales y, en algunos casos, impulsa la inestabilidad y los conflictos”*.

En el contexto de América Latina, la Comisión Económica para América Latina y el Caribe (CEPAL), organismo adscrito a la Organización de las Naciones Unidas, es el encargado de promover el desarrollo sostenible en la región. Colombia hace parte de los países integrantes de la CEPAL y tiene un acuerdo firmado desde 1963, en el cual sus temas de trabajo son: estudiar la identificación de los problemas del desarrollo que enfrenta la región, protección social, desarrollo rural, entre otros. Bajo este orden de ideas, Colombia cuenta con el DANE, organismo gubernamental que realiza estudio y seguimiento anual a los indicadores de pobreza monetaria. Los datos se reportan nivel departamental y sus publicaciones a través de los boletines de resultados son de carácter univariado. Desde un enfoque multivariado, donde los individuos (departamentos) puedan ser estudiados en función de sus variables(indicadores) que lo describen, es posible aplicar herramientas analíticas como el Análisis Factorial Múltiple y Análisis de Conglomerados. Estas permiten revelar estructuras de relaciones entre los indicadores de pobreza monetaria, factores comunes entre indicadores, posibles relaciones entre los años y generar agrupaciones de departamentos que presenten igual estructura de pobreza monetaria, bajo este análisis se obtiene un panorama mucho más global de esta problemática y con ello se pueden priorizar recursos económicos en los departamentos más vulnerables.

Además, con los resultados de estos análisis, *“Los gobiernos pueden ayudar a crear un entorno propicio para generar empleo productivo y oportunidades de empleo para los pobres y los marginados. Pueden formular estrategias y políticas fiscales que estimulen el crecimiento económico de los pobres y, por tanto, reducir la pobreza.”*(Naciones, 2015). En cuanto a la comunidad educativa *“La ciencia sienta las bases de enfoques, soluciones y tecnologías nuevas y sostenibles para hacer frente a los desafíos de reducir la pobreza y lograr el desarrollo sostenible”*(Naciones, 2015). Desde el sector privado y bajo un enfoque económico se *“puede promover oportunidades para los pobres centrándose en los segmentos de la economía donde trabaja la mayoría de estas personas, a saber, las microempresas y las pequeñas empresas, y las que operan en la economía informal”* (Naciones, 2015). Todas estas oportunidades de entendimiento de la pobreza monetaria derivadas del análisis bajo las metodologías estadísticas multivariantes apuntan a un objetivo que es el de mejorar las condiciones de calidad de



vida de los Colombianos.

## 3. Antecedentes

Numerosas son las investigaciones desarrolladas desde varios campos del conocimiento en relación con el estudio de la pobreza. En esta verificación del estado de arte, se tiene como objetivo identificar los estudios que han utilizado metodologías estadísticas para analizar la pobreza. Se encontraron dos trabajos de grado de maestría desarrollados con la metodología *Análisis Factorial Múltiple*, los cuales utilizaron como insumo indicadores económicos y de salud y un artículo investigativo sobre la pobreza monetaria urbana mediante un modelo econométrico. Así mismo se encontró una investigación sobre la pobreza que realizó un *Análisis de Conglomerados* con los datos de pobreza monetaria. Por otro lado, se presentan tres estudios desde el ámbito social. La consulta de estas investigaciones contribuyeron de la siguiente manera a este trabajo de grado:

### 3.1. Antecedentes sociales

Esta investigación desarrollada por Fresneda (2007), tiene como objetivo plantear el indicador de Necesidades Básicas Insatisfechas, como medida complementaria del indicador de Línea de Pobreza. Es decir, una guía para comparar el método directo e indirecto de medición de pobreza que permita generar y complementar las interpretaciones de manera correcta relacionando las dos formas de medir el fenómeno. Al realizar estas comparaciones, se obtienen numerosas conclusiones, una de ellas advierte del uso del indicador NBI, para ciudades pequeñas poblacionalmente debido a que carece de un claro sustento teórico que permita al gobierno basarse en este indicador para la asignación de recursos económicos.

Núñez y Ramírez (2002) investigan acerca de los *determinantes de la pobreza durante los años 1995 y 2000*, con base en los datos de ingresos de los hogares Colombianos extraídos de la Encuesta Nacional de Hogares realizada por el DANE. Primero realiza una descripción sobre la evolución de la pobreza discriminando por zonas urbanas y rurales. Segundo, se plantea una metodología económica donde se desagregan los indicadores micro y macroeconómicos con el fin de establecer los efectos de las variables socio demográficas sobre la probabilidad de ser pobre. Se encontró que en este periodo de tiempo, la pobreza aumento de manera significativa a causa de la falta de empleo y falta de estímulo por parte del gobierno para generar programas de empleo. Además, se estableció que los hogares que tenían más nivel educativo disminuían su probabilidad de estar en situación de pobreza, motivando a la creación de políticas para que la educación tuviera más cobertura en el país. Este trabajo se

relaciona con la investigación planteada, debido a que evalúa desde los datos de Pobreza Monetaria entre ciudades y zonas rurales, las variables socio demográficas como nivel educativo y tamaño del hogar como determinantes claves en la reducción de la pobreza.

## 3.2. Antecedentes estadísticos

Ariza y Retajac (2020), publicaron un artículo investigativo denominado, *Descomposición y determinantes de la pobreza monetaria urbana en Colombia*, con el objetivo de estudiar los cambios a nivel urbano en términos de los indicadores de Pobreza Monetaria y algunas características socio demográficas de los hogares durante el periodo 2002 a 2018. Los datos utilizados provienen de la base de datos de Pobreza Monetaria urbana del DANE. Se utilizaron dos metodologías; la primera para estimar y descomponer el coeficiente de Shapley y la segunda estimar un modelo econométrico logit, con el fin de encontrar los efectos entre las interacciones que tienen la desigualdad y el nivel de ingresos, sobre la pobreza en las ciudades. De acuerdo al modelo econométrico estimado de los principales factores asociados a la pobreza considerando aspectos socio demográficos de los hogares, se concluye que un incremento de un año de escolaridad reduce la probabilidad de ser pobre en un 10 %, además un incremento de un 1 % en la proporción de personas desempleadas en el hogar aumenta la probabilidad de ser pobre en 48 %. Por otro lado, la descomposición del indicador de Shapley, permitió estimar que parte de la variación de los indicadores de pobreza monetaria están relacionados con el incremento de los ingresos en los hogares Colombianos. En esta publicación se destaca el uso del modelo econométrico logit tomando como insumo los principales indicadores de Pobreza Monetaria para posteriormente interpretar los efectos marginales.

Gil y Puerta (2019), desarrollaron un artículo de investigación denominado *Aplicación del algoritmo de agrupamiento de K-medias para la pobreza monetaria en Colombia*, el objetivo del trabajo fue agrupar los departamentos Colombianos que registraron la mayor concentración de pobreza monetaria. Los datos fueron recolectados desde la página del DANE desde el 2008 hasta el 2017 para 24 departamentos. La metodología usada fue el algoritmo de K-medias. Se concluye que los departamentos más pobres fueron el grupo conformado por, Cauca, Chocó, Córdoba, debido a la escasa actividad económica que se desarrolla en ellos. Este trabajo se relaciona con la investigación planteada, ya que utiliza la misma metodología de agrupamiento y los mismos datos de pobreza monetaria, salvo que tienen una ventana de tiempo distinta.

Ruiz y Bonilla (2018), realizaron un estudio longitudinal denominado *Análisis de datos en tres vías para la evaluación de la dinámica económica en Colombia*, este trabajo realiza una caracterización del Producto Interno Bruto de Colombia que se genera por los 33 departamentos que componen la nación y se desagrega en 10 grandes áreas que permiten su construcción. El intervalo de tiempo fue del año 2000 a 2015, considerado como un periodo de crecimiento económico. Se utilizaron métodos estadísticos multivariados como el Análisis

de Componentes Principales y Análisis Factorial Múltiple. Esta investigación contribuyó al entendimiento del sustento de la economía Colombiana, este se basa en la productividad industrial, productividad en explotación de recursos naturales y productividad del sector social y comunitario. Los datos fueron obtenidos a través del Sistema Estadístico Nacional adscrito al Departamento Administrativo Nacional de Estadística. Este trabajo se relaciona con la investigación en curso porque utiliza el Análisis Factorial Múltiple y la misma estructura de datos, salvo que los autores estudiaron los datos que corresponden al producto interno bruto.

Bajo un enfoque teórico y práctico Mendoza (2009) realizó una tesis de maestría denominada *Comparación de los métodos Tucker 3 y Análisis Factorial Múltiple para el análisis de datos en tres vías*, en la primera parte se revisan teóricamente los dos métodos a través de las expresiones matemáticas, contribuciones, inercia y representación de los individuos. Posteriormente, se evalúan los dos métodos con datos que corresponden a variables socio demográficas de 24 departamentos de Colombia en el periodo de tiempo que comprende los años 2004 a 2007. Finalmente, los autores encontraron características en común con los dos métodos estudiados, ventajas y desventajas uno con respecto al otro a nivel teórico. Este trabajo se relaciona con la investigación en curso porque utiliza la misma estructura de datos y el mismo enfoque metodológico, salvo que la autora utilizó datos de salud y condiciones de vida.

## 4. Marco Teórico

En este capítulo, se presenta en primer lugar un marco conceptual de los indicadores de pobreza utilizados en el análisis y la forma de realizar el cálculo de los indicadores de Pobreza Monetaria en Colombia. En segundo lugar, se presentan los fundamentos teóricos de los métodos que son empleados en el análisis de los datos.

### 4.1. Marco conceptual

Según el DANE (2018) se tienen las siguientes definiciones para los indicadores de Pobreza Monetaria en Colombia:

1. **Línea de pobreza Monetaria (LPM):** es el costo mínimo necesario por persona para acceder a una canasta básica de bienes y servicios (alimentarios y no alimentarios) en un área geográfica determinada.
2. **Incidencia de la Pobreza Monetaria (PM):** es el porcentaje de la población de un área geográfica específica que tiene un ingreso mensual per cápita de la unidad de gasto en el hogar debajo de la línea de pobreza monetaria.
3. **Línea de Pobreza Monetaria Extrema (LPME):** es el costo per cápita mensual mínimo necesario para adquirir una canasta de alimentos que satisfaga sus requerimientos calóricos, que son definidos por la FAO (Food and Agriculture Organization).
4. **Incidencia de la Pobreza Monetaria Extrema (IPME):** es el porcentaje de la población de un área geográfica específica que tiene un ingreso mensual per cápita en el hogar debajo de la línea de indigencia (pobreza extrema monetaria).
5. **El coeficiente de Gini (GINI):** es un indicador de desigualdad que mide la concentración de la riqueza en un área geográfica determinada. Este índice toma valores entre cero y uno, donde cero representan igualdad absoluta y uno desigualdad absoluta.
6. **Brecha o intensidad de la pobreza (BP):** es un indicador que mide la cantidad de dinero que le falta a una persona pobre para dejar de estar en esa situación. Es decir, para alcanzar la línea de pobreza. Entre más grande la brecha, más pobre el individuo. La brecha corresponde a la distancia relativa entre el ingreso del pobre y el valor en pesos de la línea de pobreza.

7. **Unidad de Gasto:** la unidad de gasto está compuesta por los miembros del hogar diferentes al servicio doméstico y sus familiares.
8. **Brecha de la pobreza extrema (BPME):** mide la cantidad de dinero que le falta a una persona en situación de pobre extrema para dejar de estar en esa situación, es decir, para alcanzar la línea de pobreza extrema.
9. **Severidad de la pobreza (SPM):** integra al indicador de incidencia y brecha de la pobreza. Las diferencias entre el ingreso per cápita de cada persona pobre con respecto a la línea de pobreza es ponderada para darle mayor importancia a las personas pobres que están más lejos de la media, para incluir el efecto de la desigualdad entre los ingresos de los pobres.

Estos indicadores que componen la Pobreza Monetaria en Colombia son importantes pues permiten desglosar la pobreza en diferentes rubros que la compone.

#### 4.1.1. Cálculo de los indicadores de Pobreza Monetaria

De acuerdo a Foster y Greer (1984), para el cálculo de los indicadores de incidencia, brecha y severidad de la pobreza se plantea la Ecuación 4-1. El siguiente conjunto  $y = (y_1, y_2, \dots, y_n)$ , representan los ingresos per cápita de la unidad de gasto de todos los hogares; la línea de pobreza se define como  $z$ ; a partir de estos dos elementos se plantea la siguiente familia de indicadores de pobreza que se desprenden de la Ecuación 4-1.

$$P_\alpha(y; z) = \frac{1}{n} \sum_{i=1}^q \left( \frac{g_i}{z} \right)^\alpha \quad (4-1)$$

Donde  $g_i$ , representa el nivel de ingreso que le hace falta a un hogar  $i$  para alcanzar el nivel de bienestar normativo determinado por la línea de pobreza  $z$ .  $q$  es el número de personas definidas como pobres, aquellos que su ingreso per cápita de la unidad de gasto está por debajo de la línea de pobreza  $z$ .  $n$  es el número total de personas.

El parámetro  $\alpha \geq 0$ , mientras más grande, mayor será el peso que le asigne a los hogares más pobres. Si  $\alpha \rightarrow \infty$ , la medida de la pobreza dependerá totalmente de la distancia entre el ingreso de la persona más pobre y la línea de pobreza. Cuando  $\alpha = 0$ , se hace referencia a la incidencia, es decir el porcentaje de personas en situación de Pobreza Monetaria en un dominio geográfico determinado. Si  $\alpha = 1$  mide la brecha de la pobreza, considerando que tan pobre es un hogar en términos porcentuales. Cuando  $\alpha = 2$ , se convierte en el índice de severidad de la pobreza, que mide la intensidad de la pobreza, pero en este caso asigna más peso a las personas más pobres, por lo cual también mide la desigualdad entre los pobres.

## 4.2. Marco teórico estadístico

En la literatura se encuentra gran variedad de métodos que permiten estudiar la relación entre las variables que se encuentran en las tablas de datos. Una metodología importante es el Análisis Factorial Múltiple y Análisis de Conglomerados para estudiar la similitud entre individuos. Para cumplir con los objetivos planteados en éste trabajo de grado se presentan las siguientes teorías estadísticas:

### 4.2.1. Análisis de Componentes Principales - ACP

Para entender los métodos multivariantes es de gran importancia comprender el Análisis de Componentes Principales para datos cuantitativos; este es la base para posteriores métodos, como el Análisis Factorial Múltiple para el tratamiento de tablas que contienen datos cuantitativos y cualitativos.

De acuerdo a ?, el Análisis de Componentes Principales ACP es una técnica estadística para la descripción simultánea de variables y similitud entre los individuos. Es también una técnica de reducción de la dimensionalidad de un conjunto de variables continuas, utilizable como paso intermedio para análisis posteriores.

El ACP se puede aplicar a una tabla de datos organizada de tal forma que las columnas representan  $p$  variables cuantitativas y las filas comprenden los  $n$  individuos. Esta matriz comúnmente se centra o se estandariza para poder analizar los datos y su notación es  $X_{n,p}$ , su término general es  $x_{i,j}$ , que corresponde a la  $i$ -ésima observación de la  $j$ -ésima variable con  $i = 1, \dots, n$  y  $j = 1, \dots, p$ .

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \cdots & \cdots & \cdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix}$$

Debido a que las variables pueden expresarse en diferentes unidades de medida y estas escalas afectan el aporte de cada variable a la varianza total los datos, la matriz  $X$  se suele estandarizar por columna de acuerdo a la Ecuación 4-2 .

$$z_{i,j} = \frac{x_{ij} - \bar{x}_j}{\sqrt{s_{xj}}} \quad (4-2)$$

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \quad (4-3)$$

$$s_{xj} = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \quad (4-4)$$

Donde el promedio está determinado por la Ecuación 4-3 y la varianza por la Ecuación 4-4. Esta transformación convierte los datos estandarizados  $z_{i,j}$  con media 0 y varianza 1. Los datos  $z_{i,j}$  transformados en su conjunto se denominan matriz de puntuaciones, denotada como  $Z$ . Realizar un ACP sin la estandarización de los datos, resulta en que las variables y sus escalas con mayor variabilidad absorben la mayoría de varianza, lo cual conduce a resultados inadecuados o altamente influenciados por el efecto de la variabilidad.

La finalidad del ACP es encontrar un espacio, de menor dimensionalidad al de las variables originales  $p$  que maximice la inercia en las primeras componentes factoriales  $\alpha$ , donde las distancias de los individuos proyectados difiera mínimamente de la distancia entre los individuos sin proyectar. Las proximidades entre las variables se miden en términos de correlación y la distancia entre individuos en términos de similitudes.

#### 4.2.2. Análisis de la proximidad entre los individuos

En el espacio de los individuos  $R^p$  sin transformar los datos originales, la distancia euclidiana está determinada por  $d^2(i, i') = \sum_{j=1}^p m_j (x_{ij} - x_{i'j})^2$  entre pares de individuos. Los coeficientes  $m_j$  ponderan la influencia de cada columna  $j$ , con ellos se conforma la matriz diagonal  $M$  conocida como Métrica.

Según Lebart y Piron (1994), el mejor espacio de representación se obtiene explorando un vector director  $u$  en el eje  $\alpha(u_\alpha)$ , sobre el que se maximiza la suma de los cuadrados de las distancias entre todas las parejas de puntos  $(i, i')$   $\max_u = d^2 \sum_i^n (i, i')$ , lo que es equivalente a maximizar sobre  $u$  la suma de cuadrados de las distancias entre cada punto y el centro de gravedad  $\max_u = d^2 \sum_i^n (i, G)$  Esta es la razón por la cual el ACP requiere al menos centrar los datos,  $x_{ij} = r_{ij} - \bar{r}_j$ , equivalente a trasladar el origen a  $G$ .

Cuando existen variables con distintas escalas, se corrigen estas variables de acuerdo a la Ecuación 4-2 y bajo esta transformación (normalización), que ya contiene los pesos de las columnas con métrica,  $M = I$  la matriz idéntica y  $N = \text{diag}(1/n)$  la matriz de pesos de los individuos, se desarrolla el *Análisis de Componentes Normado*, donde  $Z'NZ$  es la matriz de correlaciones que se diagonaliza.

Como consecuencia de la diagonalización de  $Z'NZ$ , se obtiene la matriz  $U'DU$ , donde  $Z'NZ = U'DU$ ;  $U$  es ortogonal ( $U'U = I$ ) que contiene los vectores propios de  $u_\alpha$  y  $D$  es diagonal con los valores propios de  $\lambda_i$  de  $Z'NZ$ . En el ACP normado, la suma de los valores propios  $\lambda_i$  de  $Z'NZ$ , es igual a la varianza, esto es:  $\sum_{i=1}^p \lambda_i = p$ , así cada componente explica una proporción de variabilidad total  $\lambda_1 + \lambda_2 + \dots + \lambda_p = p$ , con  $\lambda_1 > \lambda_2 > \dots > \lambda_p$ .

En el caso del ACP normado la varianza de las variables es igual a 1 y la suma de las varianzas es igual a  $p$ . El valor propio más grande  $\lambda_i$  representa la varianza más grande en



la nube de datos. El vector propio asociado a este valor propio da la dirección a lo largo de la cual puede observarse esta variación. El número de valores propios no nulos se denomina rango de una matriz. Este rango corresponde al número mínimo de variables necesarias para describir los datos sin ninguna pérdida de información.

### Determinación del número correcto de factores

La extracción del número correcto de factores es una de las decisiones más problemáticas del análisis factorial (Cattell, 1966). En general no existe un criterio formal para determinar cuantos componentes principales retener. El criterio de Káiser toma los factores con valores propios superiores a uno (Johan y Marcel, 2016). El cuadrado de la correlación entre una variable y un factor es la proporción de varianza explicada por esa variable. Si se suman todos los cuadrados de los pesos factoriales de las variables en un factor (columna de la matriz factorial) obtenemos el autovalor de ese factor, que expresa la magnitud de varianza explicada por ese factor. El punto de corte de 1 se fija porque las variables están estandarizadas con la varianza igual a 1 y sería inadecuado interpretar un factor que explique menos varianza que la explicada por una variable particular (Kahn, 2006).

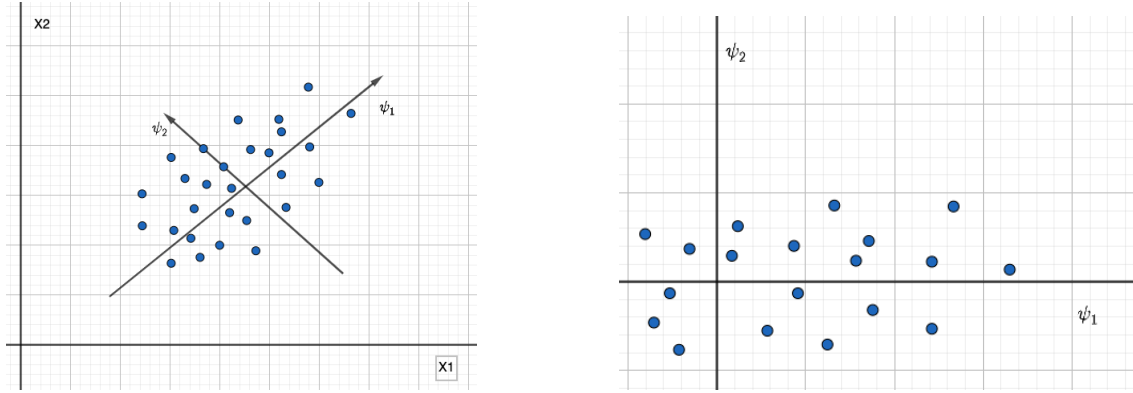
El criterio más usado es el scree test o gráfica de sedimentación, que representa la magnitud de los valores propios. En el eje vertical se representan los valores propios y en el horizontal el número de factores. Sobre la gráfica conformada se traza una línea recta base a la altura de los últimos autovalores (los más pequeños) y aquellos que queden por encima de esa línea base indicarán el número de factores a retener (Cattell, 1966).

### Coordenadas de los individuos

Los  $n$  individuos proyectados sobre el eje canónico  $u_\alpha$  y sus coordenadas quedarán determinados por  $\psi_\alpha = Zu_\alpha$ , que es una combinación lineal de las variables normalizadas  $Z_j$

$$\begin{aligned}\psi_1 &= u_{11}Z_1 + u_{12}Z_2 + \dots + u_{1p}Z_p \\ \psi_2 &= u_{21}Z_1 + u_{22}Z_2 + \dots + u_{2p}Z_p \\ &\vdots \\ \psi_p &= u_{p1}Z_1 + u_{p2}Z_2 + \dots + u_{pp}Z_p\end{aligned}$$

Donde  $\psi_1$  representa la primera componente,  $\psi_2$  la segunda componente y  $\psi_p$  la P-ésima componente. Cada elemento  $u_{i,j}$  es la ponderación o peso que tiene la variable  $Z_j$  sobre la componente  $i$ . Geométricamente, el análisis de componentes principales es una rotación de las coordenadas de los datos, los factores  $\psi_i$  definen las nuevas coordenadas como se ilustra en la Figura 4-1.



**Figura 4-1.:** Representación gráfica de los datos en el espacio original de las variables (gráfico izquierdo) y en el espacio de las componentes principales (gráfico derecho).  
Fuente: Elaboración propia

### 4.2.3. Análisis de la proximidad entre las variables

De acuerdo a Lebart y Salem (2000) una variable puede ser representada como un vector del espacio vectorial denominado  $R^n$ . La distancia entre variables se calcula de acuerdo a la Ecuación 4-5:

$$d^2(j, j') = \sum_{i=1}^n (z_{ij} - z_{ij'})^2 \quad (4-5)$$

Si se desarrolla el binomio se obtiene:

$$d^2(j, j') = \sum_{i=1}^n z_{ij}^2 + \sum_{i=1}^n z_{ij'}^2 - \sum_{i=1}^n z_{ij} z_{ij'} \quad (4-6)$$

teniendo en cuenta que las variables están estandarizadas y la definición de la varianza se obtiene que  $\sum_{i=1}^p z_{ij}^2 = \sum_{i=1}^p z_{ij'}^2 = 1$  e igualmente  $\sum_{i=1}^n z_{ij} z_{ij'} = c_{jj'}$ .

Donde  $c_{jj'}$ , es el coeficiente de correlación entre las dos variables, Se tiene que la distancia entre dos variables está dada por:  $d^2(j, j') = 2(1 - c_{jj'})$

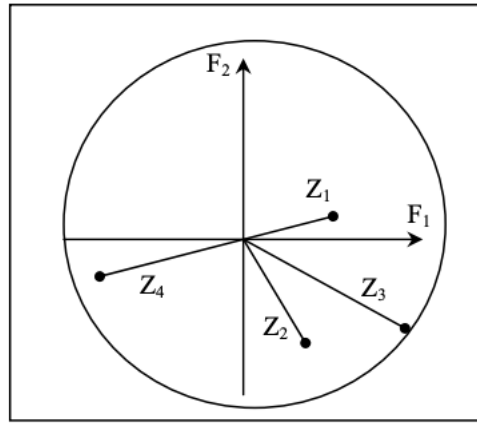
que implica:

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

La Figura 4-2, ilustra geométricamente en el espacio  $R^n$  el coseno del ángulo  $\Theta$  de dos variables  $Z_1$  y  $Z_2$  estandarizadas es el coeficiente de correlación entre las variables, este queda determinado por la Ecuación 4-7. La coordenada de las  $p$  variables están determinadas por

$\varphi_{j\alpha} = \sqrt{\lambda_\alpha} u_{j\alpha}$ , que se desprenden de las relaciones de transición que permiten relacionar el espacio de los individuos con el espacio de las variables.

$$\cos(\theta) = \frac{k' N j}{\|k\| \|k\|} = k' j = \text{cor}(k, j) \quad (4-7)$$



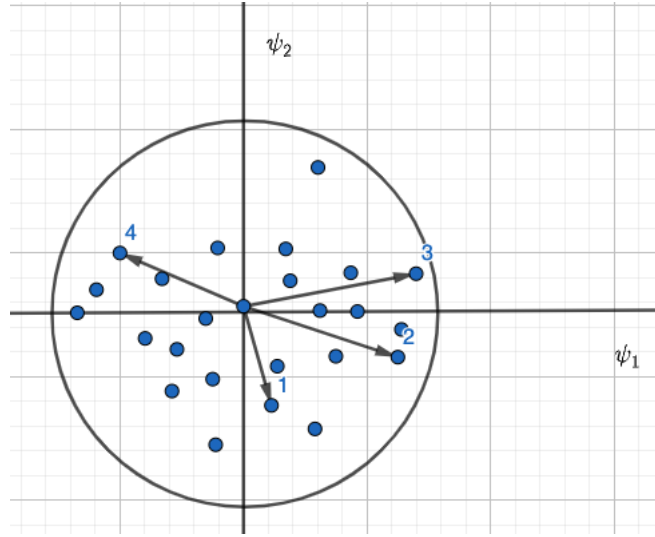
**Figura 4-2.:** Círculo de correlación en plano F1 y F2 del ACP. Fuente: Elaboración propia.

Dos variables fuertemente correlacionadas están muy próximas la una de la otra  $c_{jj'} = 1$  o, por el contrario, tan alejadas como sea posible  $c_{jj'} = -1$  según la relación lineal que se presente. Dos variables ortogonales  $c_{jj'} = 0$  están a una distancia media. Las proximidades entre puntos de variables se interpretan en términos de correlaciones.

#### 4.2.4. Representación Simultánea

De acuerdo a Lebart y Salem (2000), para lograr una representación simultánea de individuos y variables, como se ilustra en la Figura 4-3 es posible lograrlo si se toma la proyección (en la fila suplementaria) del antiguo eje canónico unitario  $e_j$  sobre el nuevo eje  $\alpha$ , cuya coordenada de proyección de  $e_j$  es  $e'_j u_\alpha = u_{j\alpha}$ . La interpretación de los individuos se realiza en términos de similitud de comportamiento con respecto a las variables.

Según Lebart y Salem (2000), la dirección de una variable define zonas para los individuos de un lado, aquellos que toman valores fuertes para esta variable y del lado opuesto, aquellos que toman valores bajos.



**Figura 4-3.:** Representación simultánea de individuos y variables en el ACP. Fuente: Elaboración propia.

En la representación simultánea de individuos y variables, se interpreta las coordenadas de los individuos respecto a la dirección y cercanía de las variables y su asociación a cada componente principal. Esto es una perspectiva deformada del sistema de ejes original teniendo en cuenta las relaciones existentes entre las variables iniciales (Lebart y Salem, 2000).

#### 4.2.5. Elementos de ayuda a la interpretación

Los ejes factoriales permiten obtener en  $(R^p)$  la mejor visualización aproximada de las distancias entre los individuos, y de otra parte, las variables que a su vez califican y caracterizan tanto a los ejes como a los individuos. Las variables fuertemente correlacionadas con un componente principal van a contribuir a la calificación de ese eje. Dicho esto interesa identificar aquellas variables que presenten coordenadas altas y que se sitúen más cerca del círculo de correlación, interpretando las componentes principales en función del reagrupamiento de ciertas variables y la oposición de otras. La correlación es conservada en la proyección según la calidad del ajuste.

Para el ACP es de interés aquellos individuos y variables, con contribuciones y cosenos altos; son puntos excéntricos (más alejados de los demás) de alta calidad responsables en buena parte de la formación de los ejes factoriales. La Ecuación 4-8, evalúa la calidad de la representación, esta presenta una medida de que tan cercano se encuentra un punto a un eje principal.

$$\cos^2(j, \alpha) = \frac{\varphi_{j\alpha}^2}{d^2(j, G)} = \varphi_{j,\alpha}^2 = \text{cor}^2(j, \alpha) \quad (4-8)$$

El cálculo de la contribución de las variables está determinada por la Ecuación 4-9.

$$CT_{\alpha}(j) = \frac{\varphi_{j\alpha}^2}{\lambda_{\alpha}} \quad (4-9)$$

En cuanto a los individuos que más aportan a la construcción de un eje principal se calcula de acuerdo a la Ecuación 4-10 donde  $m_i = 1/n$

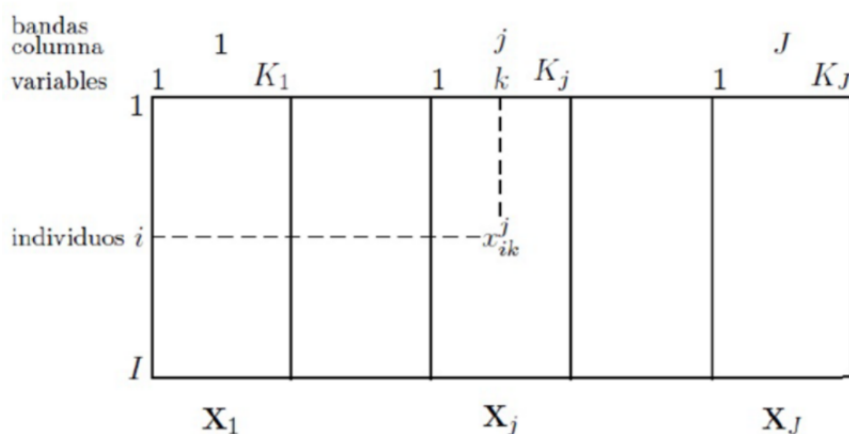
$$CT_{\alpha}(i) = \frac{m_i \psi_{i\alpha}^2}{\lambda_{\alpha}} \quad (4-10)$$

### 4.3. Análisis Factorial Múltiple - AFM

Escofier y Pagés (1992), plantean el AFM como un método factorial adaptado al tratamiento de tablas de datos en las que un mismo conjunto de individuos se describe a través de varios grupos de variables. Los grupos de variables pueden surgir de la utilización conjunta de variables de diferente naturaleza, cuantitativas y cualitativa, del empleo de tablas que provienen de otras dimensiones o del manejo de un mismo conjunto de variables medidas en distintos periodos de tiempo o bajo diferentes condiciones Abascal y Fernandez (2002).

El número de variables en cada grupo puede diferir y también la naturaleza de las variables (cualitativas o cuantitativas) puede variar de un grupo a otro, pero las variables deben ser similares en un grupo determinado. Según el tipo de variables, el AFM puede considerarse como un análisis factorial general, en el que su núcleo se basa en el Análisis de Correspondencias Múltiples - ACM, si las variables son de tipo cualitativo, ó el núcleo se basa en un Análisis de Componentes Principales - ACP si las variables son cuantitativas.

#### 4.3.1. Tabla de datos y notación



**Figura 4-4.:** Tabla de datos en AFM. Fuente: Escofier y Pagés (1992)

La Figura 4-4, ilustra la tabla completa de datos yuxtapuestos, los individuos se denotan con la letra  $i = 1, \dots, I$ ; ellos constituyen la nube  $N_I$  que se encuentra en el espacio  $K$  dimensional  $R^K$ . Las  $K$  variables constituyen la nube  $N_K$  que se encuentra en el espacio  $I$ -dimensional  $R^I$ .

Todos los grupos de variables se definen sobre el mismo conjunto de individuos. Todas las tablas se pueden yuxtaponer en línea y así formar una sola tabla de individuos y variables. El conjunto inicial de varias tablas aparece entonces como una sola tabla estructurada en sub tablas. De acuerdo a la Figura 4-4, se tiene que  $X$  es la tabla completa,  $I$  todos los individuos,  $K$  todas las variables (todos los grupos combinados),  $J$ , el conjunto de sub tablas,  $K_j$  el conjunto de variables del grupo  $j$ ,  $X_j$  la tabla asociada con el grupo  $j$ .

El AFM puede considerarse como la generalización del ACP en el que la influencia de los grupos de variables está equilibrada. El objetivo general del análisis es examinar la existencia de estructuras comunes a todos los grupos.

#### Notación

En adelante se presenta la siguiente notación usada en AFM:

- $X$  : tabla de datos completa.
- $J$  : Número para el grupo de variables.
- $I$  : Número de individuos.
- $K$  : Número de variables.
- $K_j$  : Número de variables del  $j$ -ésimo grupo.
- $X_j$  : Tabla asociada al grupo  $j$

### 4.3.2. Etapas del AFM

Para el desarrollo del Análisis Factorial Múltiple se utiliza la metodología del ACP y se desarrolla bajo dos etapas:

**Etapas I:** Cada grupo de variables es analizado por separado y se obtienen los factores parciales asociados a cada tabla. Es decir, se realiza un ACP (o ACM, dependiendo del tipo de variables que conforman el grupo) a cada una de las tablas  $X_j$ , la cual contiene los valores de las variables del grupo  $j$ ,  $j = 1, 2, \dots, J$ , en el conjunto de individuos.

**Etapas II:** Se realiza un ACP de la tabla global que resulta de yuxtaponer las  $J$  tablas  $X_j$ . En este análisis cada tabla  $X_j$  está ponderada por el inverso del primer valor propio obtenido en el ACP o ACM de la propia tabla,  $\frac{1}{\lambda_j}$ , estos valores conforman la matriz diagonal  $M$ . Esta ponderación mantiene la estructura de cada tabla, ya que todas las variables de un mismo

grupo han recibido la misma ponderación(peso) además consigue equilibrar la influencia de los grupos, ya que la inercia máxima de cada una de las nubes de individuos, definida por los diferentes grupos vale 1 en cualquier dirección (Abascal y Fernandez, 2002). El AFM, considera los espacios  $R^k$ ,  $R^I$ ,  $R^{I^2}$  respectivamente, el espacio de la nube de individuos, variables y grupos.

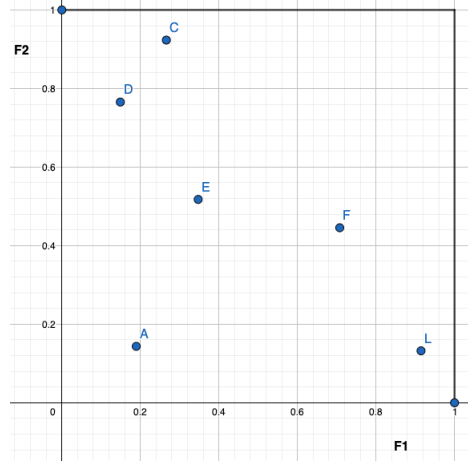
### 4.3.3. Ponderación de los grupos de variables

Para la ponderación del grupo de variables intervienen dos elementos: el número de las variables de cada grupo, entre más variables contenga mayor será la influencia de este y la relación de las variables del grupo.

El peso del grupo de variables está dado por el inverso del primer valor propio, que también es conocido como el cuadrado del primer valor singular  $\frac{1}{\lambda_j}$  obtenido de un ACP o ACM según el tipo de variable aplicado a cada grupo por separado. Dichos pesos se aplican a las componentes principales de los análisis factoriales aplicados a los grupos de variables por separado o a las variables de cada grupo. En ese sentido, la columna correspondiente a la  $K$ -ésima variable del  $J$ -ésimo grupo se define como  $Z_{jk} = X_{jk}$ . Así, se unen las  $J$  tablas ponderadas y se construye la tabla yuxtapuesta global normalizada  $Z$ . De esta manera se logra el equilibrio de cada uno de los grupos, ya que ninguno de ellos determina por si solo la primera componente principal del análisis maximizando la inercia proyectada de todas las variables (Escofier y Pagés, 1982).

### 4.3.4. Representación de los grupos de variables

En el AFM, es de interés estudiar las relaciones entre los grupos de variables a nivel global. Así como se estudia similitud entre individuos o relaciones entre variables, también es posible estudiar la cercanía entre grupos. Dos grupos de variables se parecen globalmente si dos individuos próximos en un grupo también lo son en el otro; por tanto, los individuos también se podrán describir a través de los grupos en el referencial común (González y Olaya, 2009). La Figura 4-5, ilustra un ejemplo hipotético acerca de como los grupos de variables,  $A, D, C, E, F, L$  son representados en el AFM.



**Figura 4-5.:** Ejemplo hipotético de la representación de los grupos de variables en el AFM.  
Fuente: Elaboración propia.

La nube  $N_J$  se conforma por las distancias entre los grupos de variables, las cuales se derivan del producto escalar  $Lg = \langle W_j D, W_l D \rangle = \text{traza}(W_j D W_l D)$ , donde  $W_j = X_j M_j X_j'$  y  $D$  es la matriz diagonal de los pesos de los individuos.

El producto escalar es interpretado como una medida de relación entre los grupos. A partir de este, se construye el coeficiente de correlación vectorial  $RV$  representado en la Ecuación 4-15. Este se obtiene a partir de los coeficientes de correlación lineal entre dos variables cualesquiera. Su valor está comprendido entre 0 y 1, donde cero es la no relación entre variables y uno las nubes que representan a los grupos son homotéticas. Es decir, que los puntos en el plano o el espacio se alinean con un punto fijo y la razón entre sus distancias a este punto es constante. (Abascal y calvo, 2002).

$$RV = \frac{\langle W_j D W_l D \rangle}{\|W_j D\| \|W_l D\|} = \frac{\text{traza}(W_j D W_l D)}{\|W_j D\| \|W_l D\|} \quad (4-11)$$

La norma de  $W_j D$  se expresa como  $\|W_j D\| = \sqrt{\sum \lambda_s^j}$  donde  $\lambda_s^j$  es el  $j$ -ésimo valor propio del análisis de la tabla  $j$ .

Una interpretación de la Figura 4-5 muestra que el grupo  $L$  y  $F$  contribuyen más a la formación del primer eje factorial  $F1$ , mientras que el grupo  $A$  no contribuye mucho en el primer factor  $F1$ . De manera similar para el grupo  $C$  y  $D$  están más relacionados con el segundo factor  $F2$ , mientras que el grupo  $A$  no se relaciona con ningún factor.

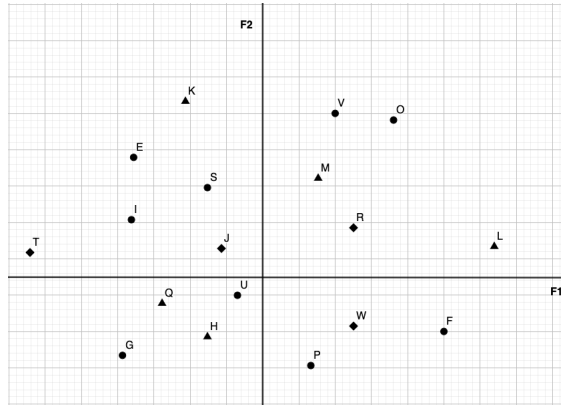


### Factores comunes

De acuerdo a Escofier y Pagés (1992), si la correlación entre grupos y el eje factorial es fuerte, existe un factor común que se calcula a partir del concepto de inercia acumulada de las variables representadas en el grupo respecto de dicho factor, que puede denominarse como el peso del factor en el grupo.

#### 4.3.5. Representación de las variables

Para la representación de las variables se toman las coordenadas de las proyecciones sobre los ejes factoriales calculadas en la tabla completa  $X$  cuando se realiza el ACP, mediante la expresión  $\varphi_s = \sqrt{\lambda_s u_s}$ , que se deriva de las relaciones de transición, Esta permite relacionar los individuos y las variables.

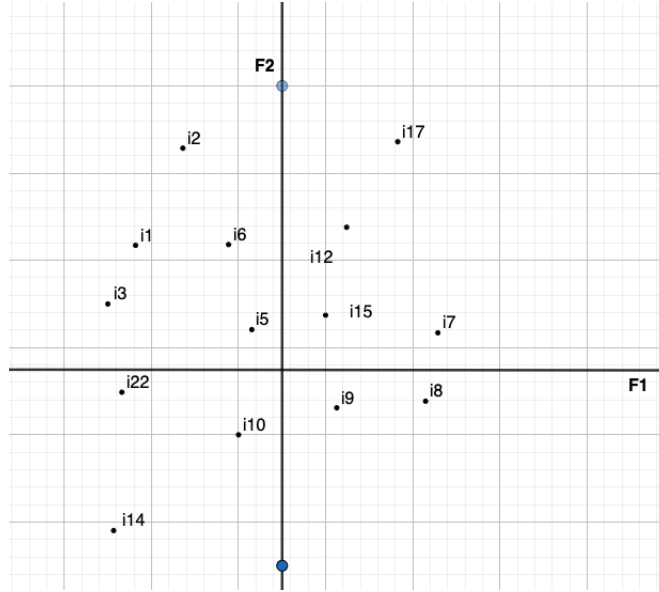


**Figura 4-6.:** Ejemplo hipotético de representación de las variables sobre los ejes factoriales  $F_1$  y  $F_2$  en AFM. Fuente: Elaboración propia.

La Figura 4-6, ilustra un caso artificial de la proyección de las variables por grupo en el primer eje factorial  $F_1$ . Por ejemplo, la variable  $L$ , que pertenece al grupo ▲, está más relacionada con la primera componente  $F_1$ , mientras que la variable  $K$  que también pertenece al grupo ▲, está más relacionada con la segunda componente principal  $F_2$ . Las variables que se encuentran cerca al origen no aportan mucho a la construcción de los ejes principales, como lo son  $J$  del grupo ◆ y  $U$  del grupo ●.

Las componentes principales maximizan la inercia de las proyecciones de todas las variables. La inercia proyectada de cada nube  $N_k^j$  puede ser interpretada como la contribución de un grupo. La ponderación de los grupos por  $\frac{1}{\lambda_j^j}$  equilibra su influencia en el sentido de que la contribución de un grupo a la construcción de un eje está limitada por el valor 1. De acuerdo a esto, ningún grupo puede, por si solo, determinar el primer eje (salvo una situación de simetría excepcional), y un grupo influye sobre tantos más ejes cuanto mayor sea su dimensión.

### 4.3.6. Representación de los individuos



**Figura 4-7.:** Ejemplo hipotético de representación de los individuos de variables en el plano F1 y F2 del AFM. Fuente: Elaboración propia.

Sea  $u_s$ , el eje de inercia de orden  $s$  de la nube de individuos  $N_I$  asociada a la tabla  $X$  en  $R^K$ , esto se deriva de la componente principal  $F$ , por la relación:  $u_s = (1/\sqrt{\lambda_s})X'DF_s$  en la que  $\lambda_s$  es el valor propio de  $WD$  asociado a  $F_s$  ( $W = \sum_j W_j$ ). La proyección de  $N_I^j$  sobre  $u_s$  se escribe de  $F_s^j = \bar{X}_j M u_s = (1/\sqrt{\lambda_s}) \bar{X}_j M X' D F_s = 1/\sqrt{\lambda} W_j D F_s$ .

Las coordenadas de los individuos sobre un eje constituyen la proyección de la nube media o global  $N_I$ . Las  $J$  nubes  $N_I^j$  de individuos, cada una caracterizada por un conjunto  $j$  de variables y denominadas “nubes parciales”, se proyectan como ilustrativas sobre los ejes factoriales. La proyección simultánea de la nube media y de las nubes parciales se denomina “representación simultánea” (Pardo, 2011) .

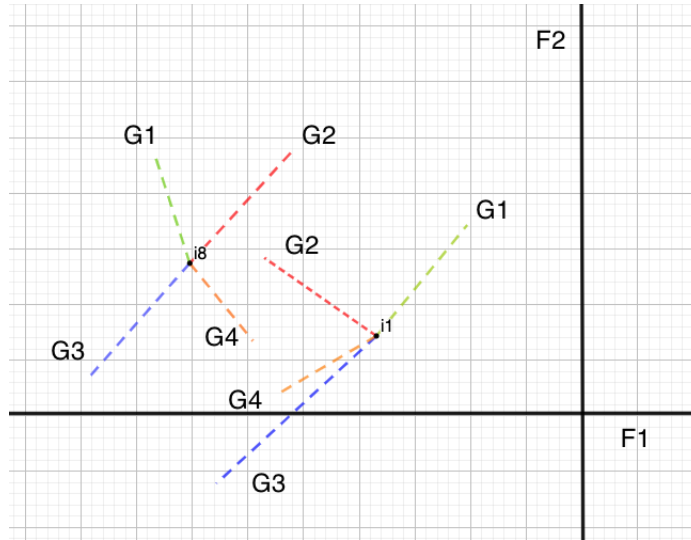
La Figura 4-7, es una representación hipotética que muestra la nube de individuos en el primer y segundo plano factorial  $F1$  y  $F2$ . Una interpretación para el individuo  $i_7$  y  $i_8$  es que estos aportan más a la construcción del primer eje factorial  $F1$ . El individuo  $i_{22}$  también aporta a la construcción de este factor, pero se sitúa en dirección contraria. El individuo  $i_{14}$  es considerado como un individuo excéntrico. Los individuos  $i_5$ ,  $i_{15}$ ,  $i_6$  son considerados como individuos promedio.

### 4.3.7. Representación superpuesta

Para lograr la representación simultánea en el AFM, es necesario tener en cuenta las siguientes consideraciones:

1. La nube  $N_I^j$  está en  $R^{K_j}$  subespacio de  $R^K$ .
2.  $R^K = \oplus_j R^{K_j}$  (suma directa de los subespacios ortogonales dos a dos).
3.  $M_j$  es la métrica en  $R^K$ , submatriz con las  $k_j$  filas y las  $K_j$  columnas de  $M$ .
4. Las coordenadas de los puntos  $N_I^j$  son las filas de  $X_j$ , entonces las coordenadas de estos puntos en  $R^K$  son las filas de la matriz  $\tilde{X}_j$  definida como:  

$$\tilde{X}_j = [0 \cdots 0 X_j 0 \cdots 0]$$
5.  $X = \sum_j \tilde{X}_j$ , entonces con el fin de hacer coincidir  $F_s(i)$  con el punto medio de los  $J$  puntos parciales  $F_s^j(i)$ , las filas de  $\tilde{X}_j$ , denominadas filas parciales, se proyectan como puntos suplementarios amplificadas por  $J$ . En efecto  $X = 1/J \left( J \sum_j \tilde{X}_j \right)$ .
6. La unión de las  $J$  nubes parciales forma la nube  $N_I^j$ .
7. La inercia de la nube  $N_I^j$  se puede expresar como la  $I_{intra} + I_{entre}$  de subnubes de  $N_i^J$ . La inercia de la nube media  $N_I$  del AFM corresponde a la  $I_{entre}$ , cuando las coordenadas parciales se amplifican por  $J$  (Escofier y Pagés, 1982).



**Figura 4-8.:** Ejemplo hipotético de representación superpuesta de la nube de individuos en el plano F1 y F2 en AFM. Fuente: Elaboración propia.

La Figura 4-8, ilustra un ejemplo hipotético de representación superpuesta de dos individuos  $i1$  e  $i8$  en cuatro grupos de variables. El punto conformado por cada individuo representa el

centro de gravedad de todos los grupos de variables. La representación de la nube global se obtiene superponiendo las cuatro nubes generadas por los cuatro grupos que resultan de la proyección de cada grupo a su subespacio. Así el individuo  $i1$  se encuentra mejor representado en el grupo  $G4$  que en individuo  $i8$  en el mismo grupo  $G4$ .

#### 4.3.8. Ayudas para la interpretación

**Contribuciones** Las contribuciones en el AFM, se pueden calcular para cada individuo, variables o grupos de variables expresadas mediante porcentajes en términos de la inercia que contribuye en la construcción de cada eje factorial, mediante las siguientes expresiones:

- **Contribuciones de las variables:** asumiendo que los pesos de las variables del  $j$  ésimo grupo,  $w_i$  es igual al inverso del primer valor propio, son los mismos, la contribución de la variable a la formación del  $\lambda$  ésimo eje factorial está determinada por:

$$CTR_{k,s} = \frac{\varphi_s^2}{\lambda_j^I} \quad (4-12)$$

$$h = 1, 2, \dots, m$$

- **Contribuciones de los individuos:** si se asume que el peso es  $w_i = 1/n$  la contribución de un individuo a la formación del  $\alpha$  ésimo eje factorial, se define como:

$$CTR_{i,s} = \frac{m_i(F_s^j)^2}{\lambda_s} \quad (4-13)$$

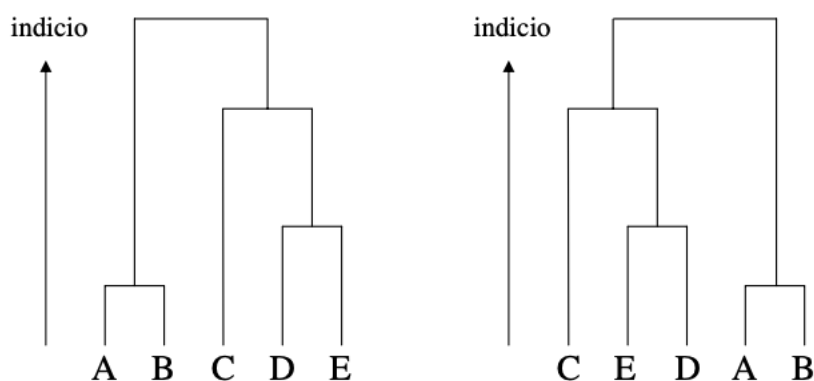
- **Contribuciones de los grupos:** es la sumatoria de las contribuciones de las variables del  $j$ -ésimo grupo de variables en el  $\alpha$ -ésimo eje factorial, definido como:

$$CTR_{j,s} = \sum_{k=1}^{K_j} ctr_{k,s} \quad (4-14)$$

### 4.4. Análisis de Conglomerados - AC

Los métodos del análisis multivariado de datos principalmente proporcionan representaciones sintéticas de objetos (estos objetos son esencialmente individuos, variables o modalidades de variables cualitativas) correspondientes a las filas y a las columnas de una tabla numérica. El modo emblemático de la representación de un conjunto de objetos es la nube de puntos (cada punto es un objeto) evolucionando en un espacio euclidiano (pudiendo reducirse a un plano).

El término euclidiano significa que las distancias entre puntos se interpretan en términos de similitud para los individuos y en términos de correlación para las variables cuantitativas. Los métodos como el ACP y el AFM proporcionan representaciones euclidianas. Otra forma de representación de un conjunto de objetos, que pone en evidencia los parentescos entre ellos (similitudes o correlaciones) es el árbol jerárquico; se conoce también como una jerarquía y, más precisamente, una jerarquía indexada donde la altura a la cual los objetos se reagrupan se interpreta (se utiliza el término dendrograma). La utilización de este árbol es intuitiva: dos objetos son más parecidos cuanto más, para ir de uno a otro, no es necesario subir alto en el árbol. La utilización de este árbol es intuitiva: dos objetos son más parecidos cuanto más, para ir de uno a otro, no es necesario subir alto en el árbol. La Figura 4-8 ilustra con un ejemplo utilizando similitudes entre cinco objetos (A, B, C, D, E) sintéticas.



**Figura 4-9.:** Ejemplo de árbol jerárquico (dendrograma). Fuente: Lê y Pagès (2009)

De la Figura 4-9 se puede interpretar, que los objetos A y B se parecen más que los objetos D y E; el objeto C se parece más al conjunto de dos objetos D, E que al conjunto A, B. No se modifica la estructura de un árbol efectuando simetrías como lo muestran ambas representaciones de la Figura 4-9. Las proximidades laterales (por ejemplo entre B y C Figura 4-9 (izquierda) entre los objetos no se interpretan. Se trata de construir un árbol jerárquico (y no un plano factorial) para visualizar las similitudes entre objetos, que es un modo de estudiar la variabilidad contenida en la tabla. Esta problemática es la misma que en el análisis factorial: solo el modo de representación difiere entre ambos métodos. Los algoritmos que construyen los árboles son reagrupados bajo el término (clasificación jerárquica). Estos son numerosos: los más utilizados proceden de manera ascendente (reagrupando primero los objetos más semejantes y a continuación los grupos así constituídos) y son reagrupados bajo el término (Clasificación Jerárquica Ascendente). El algoritmo más utilizado es el algoritmo de *Ward*.

#### 4.4.1. Similitud entre individuos

En el caso de una tabla que cruza  $I$  individuos y  $K$  variables cuantitativas en término general  $x_{ik}$  (valor del individuo  $i$  para la variable  $k$ ), el ACP normado sitúa primero los  $I$  individuos  $i$  en el espacio (vectorial)  $\mathbb{R}^K$  y utiliza, para medir la similitud entre dos individuos  $i$  y  $l$ , la distancia (euclidiana) usual en  $\mathbb{R}^K$ , es decir, anotando  $\bar{x}_k$  la media de la variable  $k$ :

$$d^2(i, l) = \sum_{k=1}^K (x_{ik} - x_{lk})^2 \quad (4-15)$$

$$d^2(i, l) = \sum_{k=1}^K \left( \frac{x_{ik} - \bar{x}_k}{s_k} - \frac{x_{lk} - \bar{x}_k}{s_k} \right)^2 \quad (4-16)$$

$$d^2(i, l) = \sum_{k=1}^K \frac{1}{s_k} (x_{ik} - x_{lk})^2 \quad (4-17)$$

En la Ecuación 4-15, la más general, los datos  $x_{ik}$  han sido previamente centrados y reducidos. En las dos otras fórmulas, ponemos de manifiesto explícitamente el centrado y la reducción, lo que será comentado más tarde. Cuando se define la distancia  $d$  entre individuos de un espacio (aquí  $\mathbb{R}^K$ ), se dice que se provee este espacio de la distancia  $d$  (conocido también como métrica y se habla entonces de espacio métrico) porque en este espacio todos los cálculos relacionados a la noción de distancia deben ser hechos con esta distancia  $d$ . La función  $d$  de  $I \times I$  en  $\mathbb{R}^+$  definida en el ACP posee todas las propiedades matemáticas deseables, comenzando por la de una distancia (en el sentido matemático del término) sea:

$$\begin{cases} d(i, l) = 0 \leftrightarrow i = l, \\ d(i, l) = d(l, i), \\ d(i, l) \leq d(i, j) + d(j, l) \text{ (desigualdad triangular)} \end{cases}$$

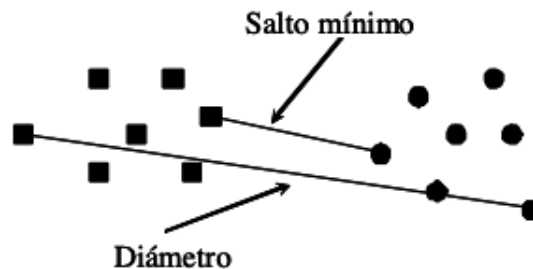
Se trata además de una distancia euclidiana, es decir, que permite definir una noción de ángulo y de proyección ortogonal. Esta última noción estando en el centro del análisis factorial, todo método factorial necesariamente utiliza una distancia euclidiana. Salvo una necesidad muy fuerte implicada por los datos, es recomendable utilizar una distancia euclidiana, ya que permite poner en marcha conjuntamente una clasificación y un análisis factorial.

### Otras Distancias euclidianas

Existe una infinidad de distancias. Las más clásicas y fáciles de interpretar consisten en partir de la distancia usual representada en la Ecuación 4-15 y en otorgar un peso a cada dimensión. Por ejemplo, en ACP normado, se puede también considerar que los datos son solamente centrados y que la distancia utilizada asigna a cada variable un peso inverso a su desviación acorde a la Ecuación 4-16. Las Ecuaciones 4-14 a 4-16 ilustran el hecho de que, en presencia de una distancia euclidiana cualquiera, se puede con la distancia usual que se muestra en la Ecuación 4-15 con una transformación de los datos.

#### 4.4.2. Similitud entre grupos de individuos

Para construir un árbol jerárquico es necesario definir una distancia o una disimilitud entre grupos de individuos. Existen varias posibilidades: se mencionarán sólo las más importantes. Sean dos grupos de individuos  $A$  y  $B$ . El salto mínimo entre  $A$  y  $B$  es el más pequeño de las distancias entre un elemento de  $A$  y un elemento de  $B$ . El diámetro entre  $A$  y  $B$  es el más grande de las distancias entre un elemento de  $A$  y un elemento de  $B$ . La Figura 4-9 ilustra estas definiciones. El mayor interés de las definiciones precedentes es que son aplicables a todas las distancias o disimilitudes.



**Figura 4-10.:** Salto mínimo y diámetro entre dos grupos de individuos (identificados por símbolos diferentes). Fuente: Lê y Pagès (2009).

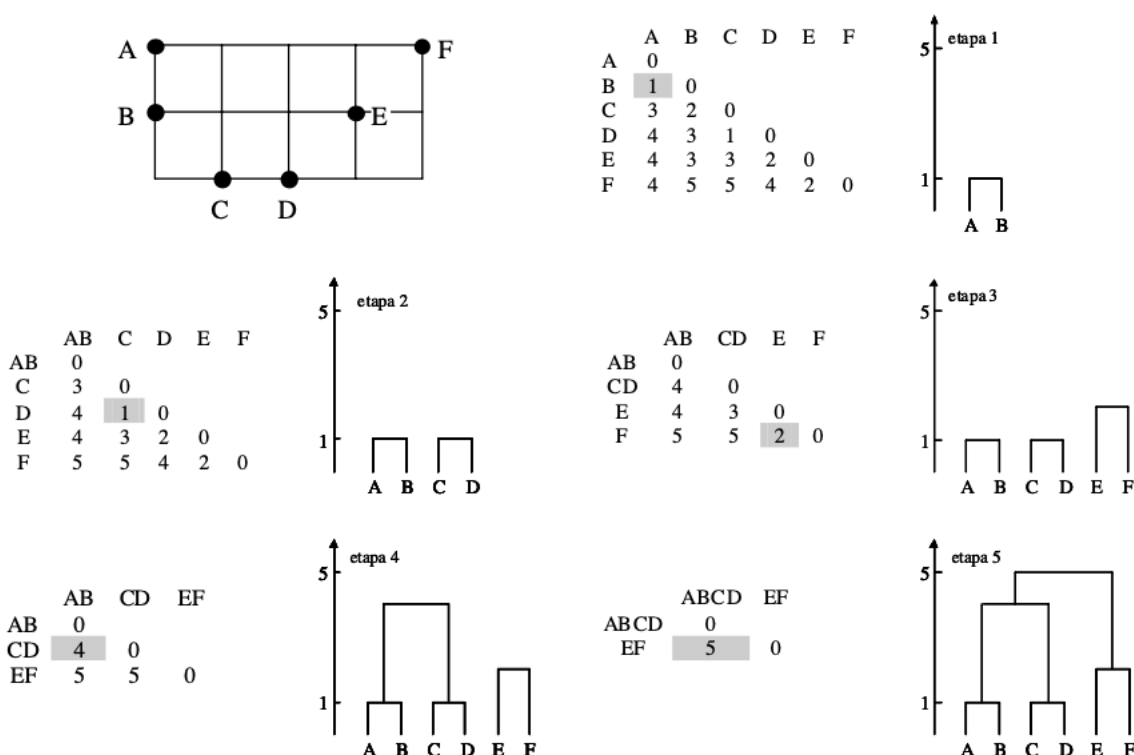
#### 4.4.3. Construcción de una jerarquía indiciada

##### Algoritmo clásico de construcción ascendente

El punto de partida es una matriz de disimilitudes  $D$  (estas disimilitudes pueden ser distancias euclidianas) entre individuos donde el término general  $d(i, l)$  es la disimilitud entre los individuos  $i$  y  $l$ . Esta matriz es simétrica y contiene ceros sobre la primera diagonal: solamente es necesario una mitad. Por convenio, la parte triangular inferior. Se toman los individuos  $i$  y  $l$  más *similares* o *próximos*: se conforma un nuevo elemento  $(i, l)$ : este grupo de individuos será fijo en adelante. El valor  $d(i, l)$  es el índice de la agregación entre  $i$  y  $l$ .

Este valor es utilizado para definir la altura a la cual las ramas del árbol que corresponden a  $i$  y a  $l$  se juntan.

Se actualiza la matriz  $D$  suprimiendo las filas y las columnas que corresponden a los individuos  $i$  y  $l$  y creando una nueva fila y una nueva columna para el grupo  $(i, l)$  que se completa con disimilitudes entre este grupo y cada uno de los individuos restantes. Se obtiene la matriz  $D_1$  en la cual se busca la pareja de los elementos más próximos. Estos elementos son agregados hasta que ya no se encuentren más para agregar.



**Figura 4-11.:** Etapas de la construcción de un árbol jerárquico a partir de seis individuos repartidos sobre un plano. Fuente: Lê y Pagès (2009)

La Figura 4-10, muestra las etapas en un ejemplo hipotético de la aplicación del algoritmo a un pequeño conjunto de datos que contiene seis individuos repartidos sobre un plano. Por razones de facilidad de cálculo, se trabaja con la distancia Euclídea y el recálculo de las distancias según el diámetro.

#### 4.4.4. Jerarquía y partición

Los puntos donde se reúnen las ramas que corresponden a los elementos que se reagrupan son llamados *nudos*. También se les conoce como *bifurcación* para expresar una descripción descendente del árbol. Los individuos que hay que clasificar son llamados *nudos terminales*.



Con  $I$  individuos, hay  $(I - 1)$  nudos a menudo numerados de  $I + 1$  a  $2xI$ . Por orden de aparición en la construcción del árbol (los  $I$  primeros números son reservados a los nudos terminales; no obstante, en ciertos software, la numeración de los nudos no toma en consideración los nudos terminales). Trazando una línea horizontal con un índice dado, definimos una partición (se dice que se corta el árbol). Sobre la Figura 4-11, el nivel de corte  $A$  define una partición en dos clases  $(1, 2, 3, 4)$  y  $(5, 6, 7, 8)$ ; El nivel de corte  $B$  define una partición más fina en cuatro clases  $(1, 2)$ ,  $(3, 4)$ ,  $(5, 6)$  y  $(7, 8)$ . Por construcción, estas particiones son encajadas: cada clase del nivel  $B$  es incluida en la misma clase del nivel  $A$ . Resulta así, que un árbol jerárquico puede ser visto como una continuación de particiones encajadas, yendo de la más fina (en la cual cada individuo constituye una clase) a la más ordinaria (en la cual sólo hay una clase).

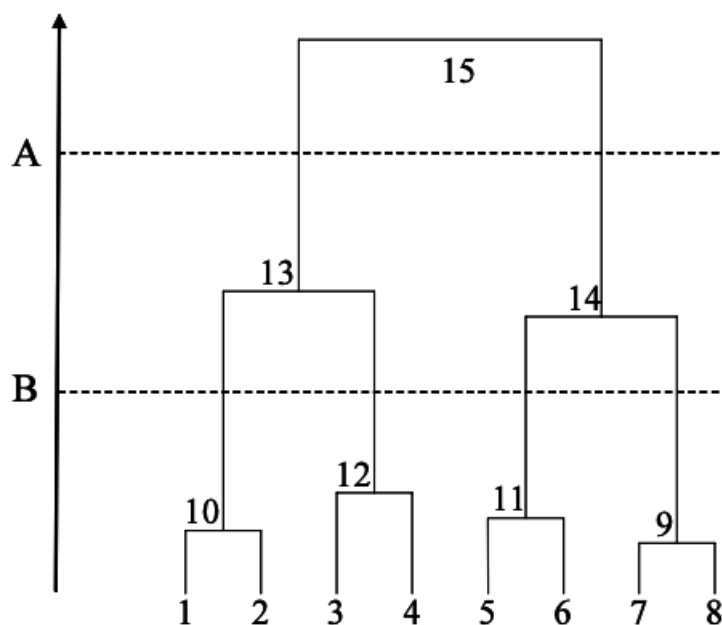


Figura 4-12.: Jerarquía y partición. Fuente: Lê y Pagès (2009)

#### 4.4.5. Método de Ward

Este método se aplica a individuos situados en un espacio euclidiano. Es el caso más frecuente, una tabla en la cual un conjunto de individuos es descrito por un conjunto de variables. Cuando los datos son cuantitativos, estudiamos la nube  $N_I$  (nube de individuos) evolucionando en  $\mathbb{R}^K$ . Este método, ascendente consiste a cada paso en reagrupar dos elementos (individuos aislados o clases ya formadas) maximizando la calidad de la partición obtenida.

### Calidad de una partición

Una buena partición es tal que:

- Dentro de una clase los individuos son homogéneos (variabilidad intraclase débil).
- De una clase a otra los individuos son diferentes (variabilidad interclases elevada).
- Si los individuos están en un espacio euclidiano, el teorema de *Huygens* proporciona un marco de análisis bien adaptado al estudio de una partición. Este teorema descompone la inercia total (de la nube de los individuos) en dos partes:
  1. La inercia intraclase, fundada sobre la diferencia entre cada punto y el centro de gravedad de la clase a la cual pertenece;
  2. La inercia interclases, fundada sobre la diferencia entre cada centro de gravedad de una clase y el centro de gravedad general. De modo general se escribe:

Inercia total = Inercia interclases + Inercia intraclase.

El teorema de Huygens representado en la Ecuación 4-17, sí considera a los individuos descritos por una sola variable cuantitativa  $y$ , se tiene la ecuación del análisis de varianza a un factor. Con  $I$  individuos (de mismo peso) repartidos en  $Q$  clases entonces:  $y_{iq}$  es el valor para  $y$  del  $i$  individuo de la clase  $q$ ;  $\bar{y}_q$  la media de  $y$  para los individuos de la clase  $q$ ;  $I_q$  el número de individuos de la clase  $q$ ;  $\bar{y}$  la media general de  $y$ .

$$\sum_{q=1}^Q \sum_{i=1}^{I_q} (y_{iq} - \bar{y})^2 = \sum_{q=1}^Q I_q (\bar{y}_q - \bar{y})^2 + \sum_{q=1}^Q \sum_{i=1}^{I_q} (y_{iq} - \bar{y}_q)^2 \quad (4-18)$$

En general, disponemos de  $K$  variables cuantitativas y la nube de los individuos evolucionando en  $\mathbb{R}^K$ . Las dimensiones de  $\mathbb{R}^K$  siendo ortogonales, el teorema de Huygens se obtiene sumando las inercias a lo largo de cada dimensión. Sea,  $y_{ikq}$  el valor para la variable  $k$  del individuo  $i$  de la clase  $q$ :

$$\sum_{q=1}^Q \sum_{i=1}^{I_q} \sum_{k=1}^K (y_{ikq} - \bar{y}_k)^2 = \sum_{q=1}^Q \sum_{k=1}^K I_q (\bar{y}_{qk} - \bar{y}_k)^2 + \sum_{q=1}^Q \sum_{i=1}^{I_q} \sum_{k=1}^K (y_{ikq} - \bar{y}_{qk})^2 \quad (4-19)$$

Inercia total = inercia interclases + inercia intraclase.

Si se adopta esta descomposición como marco de análisis, se mide la variabilidad por la inercia, entonces en la búsqueda de una buena partición es indiferente minimizar la variabilidad intraclase o maximizar la variabilidad interclases (ya que la variabilidad total es fijada por los datos). De ello resulta que la calidad de una partición puede ser medida por:

$$\frac{Inercia_{interClases}}{Inercia_{Total}} \quad (4-20)$$

Este cociente indica la parte de variabilidad total expresada por la partición. A menudo es multiplicado por 100 para poder ser enunciado en término de porcentaje. Este porcentaje debe ser comparado con el número de individuos y con el número de clases. En efecto, aumentando el número de clases, podemos encontrar una partición que presenta un porcentaje (de inercia expresada) tan elevado como se desea.

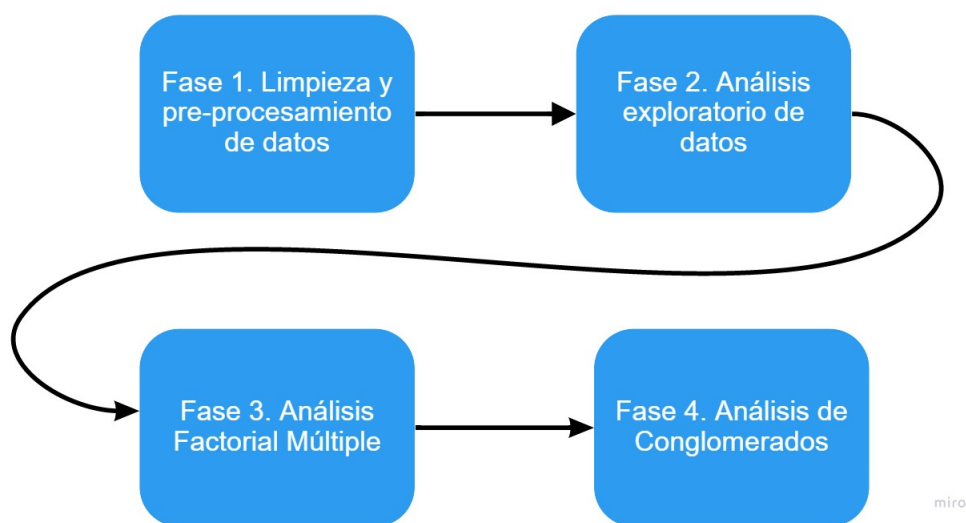
#### 4.4.6. Análisis de una jerarquía, elección de una partición

Aunque construida de modo ascendente, una jerarquía es generalmente analizada de modo descendente. El objetivo es dar una visualización de la variabilidad de los datos o, de otro punto de vista, del conjunto de las similitudes entre los individuos. El término *nudo* evoca más la similitud de dos clases (óptica de la construcción ascendente) que una subdivisión en dos clases; de ahí el término de *bifurcación* utilizado algunas veces en una investigación descendente. Con la agregación por la inercia, el nivel de un nudo, visto de modo descendente, determina la cantidad que se gana (en inercia interclases o en disminución de inercia intraclase) separando ambas clases que reagrupa. Entonces se percibe que una jerarquía es muy útil para razonar la elección de una partición. Concretamente, se debe tener en cuenta:

1. El aspecto general del árbol conformado.
2. Los niveles de los nudos, para cuantificar el punto precedente. Cada irregularidad en este decrecimiento sugiere un nivel de corte.
3. El número de clases que no debe ser demasiado elevado en cuyo caso el interés sintético del procedimiento disminuye.
4. La interpretabilidad de las clases: aunque corresponde a una ganancia de inercia interclases apreciable, no se debe retener una subdivisión que no puede ser interpretada; del mismo modo, se debe retener una subdivisión claramente interpretable aunque corresponde a una ganancia de inercia poco elevada.

## 5. Metodología





En este capítulo, con el fin de lograr el cumplimiento de los objetivos propuestos y dar solución a la pregunta de investigación, se plantea desarrollar la siguiente metodología en cuatro fases que resume la Figura 5-1.



**Figura 5-1.:** Fases en la metodología del análisis de datos

### 5.1. Estructura de los Datos

Los datos que se utilizan en este trabajo de grado, son obtenidos desde la página web del Departamento Nacional de Estadística. En la tabla de datos se encuentran 23 departamentos y Bogotá D.C como unidad administrativa. No se presentan resultados de los demás departamentos. La estructura de los datos se encuentra conformada como se ilustra en La Figura 5-2.

Incidencia de la Pobreza Monetaria									
Incidencia de la Pobreza Monetaria 23 Departamentos y Bogotá D.C. 2012 - 2020									
   									
Cifras en Porcentaje									
Departamento	Año								
	2012	2013	2014	2015	2016	2017	2018	2019	2020
Antioquia	35.3	31.8	31.7	31.1	29.8	29.0	28.3	29.8	34.0
Atlántico	38.0	36.0	32.6	30.0	29.2	28.7	27.8	27.3	40.2
Bogotá D.C.	27.0	24.3	23.3	25.1	25.9	28.4	26.4	27.2	40.1
Bolívar	50.8	48.5	46.7	46.8	48.0	44.3	43.4	46.0	52.7
Boyacá	42.7	45.7	44.3	42.9	40.5	36.5	34.2	35.7	39.8
Caldas	43.0	39.4	37.1	36.3	36.2	33.3	29.6	28.7	30.7
Cauquetá	51.4	50.6	47.3	51.2	46.0	45.3	48.5	48.8	43.9
Cauca	68.5	65.0	63.0	61.0	59.7	57.2	58.8	59.6	55.6
Cesar	51.7	50.4	47.2	48.6	48.5	46.8	48.9	51.7	58.3
Chocó	73.9	69.0	72.3	70.7	67.2	66.1	68.6	68.4	64.6
Córdoba	66.4	57.8	54.2	55.4	53.8	53.2	52.6	54.2	59.4
Cundinamarca	26.8	21.7	19.9	19.6	20.1	17.9	18.2	20.4	27.0
Huila	53.6	54.3	52.8	53.3	53.3	44.1	45.8	51.2	55.6
La Guajira	61.9	59.5	57.0	57.2	56.1	55.6	57.2	61.8	66.3
Magdalena	58.3	56.1	54.6	51.7	55.9	54.7	51.7	53.5	59.8
Meta	37.3	35.0	31.4	30.2	32.5	32.7	32.6	32.7	40.0
Nariño	60.1	57.1	53.5	50.2	54.1	48.9	50.4	51.0	49.9
Norte de Santander	44.2	43.6	44.6	45.9	46.2	45.1	47.4	51.9	56.3
Quindío	47.1	43.7	40.2	40.5	38.9	35.3	32.1	33.0	38.3
Risaralda	38.4	39.4	34.9	33.7	30.8	27.9	28.7	28.7	35.8
Santander	30.6	28.5	27.8	26.9	29.4	29.7	30.9	31.1	38.8
Sucre	57.1	52.6	49.4	50.5	52.1	47.5	46.2	50.3	51.4
Tolima	48.8	42.1	40.2	40.5	39.7	37.6	39.1	39.3	45.7
Valle del Cauca	31.8	32.1	27.5	26.6	27.5	26.0	24.7	24.0	34.5
Total Nacional	40.8	38.3	36.3	36.1	36.2	35.2	34.7	35.7	42.5

**Figura 5-2.:** Estructura de los indicadores de Pobreza Monetaria por departamentos. Fuente: DANE (2021)

La Figura 5-2, ilustra la tabla de datos de la Pobreza Monetaria. En las columnas de la tabla se encuentran los años que comprenden una ventana de tiempo entre el 2012 al 2020 y en las filas 23 departamentos y Bogotá D.C. como unidad administrativa. Los demás indicadores como Severidad, Brecha, Coeficiente de Gini e Ingreso por Unidad del Gasto se encuentran en tablas similares. Para cada indicador existe una tabla de datos como se ilustra en la Figura 5-2. De la anterior estructura de datos se desprenden las estadísticas descriptivas que se encuentran en el Capítulo 6 de este trabajo.

## 5.2. Procesamiento de datos

Para realizar el AFM con las tablas de datos de los distintos indicadores, se conformaron grupos por cada dos años para un total de cinco grupos. La Tabla 5-1 representa los cinco grupos conformados por los indicadores de pobreza durante un intervalo de tiempo de 10 años comprendido entre los años 2012 a 2020.

Grupo	Indicador	Etiqueta
G_1: año 2012	$V_{1,1}$ : % pobreza monetaria	PM2012
	$V_{1,2}$ : % pobreza monetaria extrema	PME2012
	$V_{1,3}$ : % coeficiente de gini	GINI2012
	$V_{1,4}$ : Ingreso por unidad del gasto	IPUG2012
	$V_{1,5}$ : linea de pobreza	LP2012
	$V_{1,6}$ : linea de pobreza extrema	LPE2012
	$V_{1,7}$ : % brecha de pobreza	BP2012
	$V_{1,8}$ : % brecha de pobreza extrema	BPE2012
	$V_{1,9}$ : % severidad de la pobreza	SP2012
	$V_{1,10}$ : % severidad de la pobreza extrema	SPE2012
G_2: año 2014	$V_{2,1}$ : % pobreza monetaria	PM2014
	$V_{2,2}$ : % pobreza monetaria extrema	PME2014
	$V_{2,3}$ : % coeficiente de gini	GINI2014
	$V_{2,4}$ : Ingreso por unidad del gasto	IPUG2014
	$V_{2,5}$ : linea de pobreza	LP2014
	$V_{2,6}$ : linea de pobreza extrema	LPE2014
	$V_{2,7}$ : % brecha de pobreza	BP2014
	$V_{2,8}$ : % brecha de pobreza extrema	BPE2014
	$V_{2,9}$ : % severidad de la pobreza	SP2014
	$V_{2,10}$ : % severidad de la pobreza extrema	SPE2014
G_3: año 2016	$V_{3,1}$ : % pobreza monetaria	PM2016
	$V_{3,2}$ : % pobreza monetaria extrema	PME2016
	$V_{3,3}$ : % coeficiente de gini	GINI2016
	$V_{3,4}$ : Ingreso por unidad del gasto	IPUG2016
	$V_{3,5}$ : linea de pobreza	LP2016
	$V_{3,6}$ : linea de pobreza extrema	LPE2016
	$V_{3,7}$ : % brecha de pobreza	BP2016
	$V_{3,8}$ : % brecha de pobreza extrema	BPE2016
	$V_{3,9}$ : % severidad de la pobreza	SP2016
	$V_{3,10}$ : % severidad de la pobreza extrema	SPE2016
G_4: año 2018	$V_{4,1}$ : % pobreza monetaria	PM2018
	$V_{4,2}$ : % pobreza monetaria extrema	PME2018
	$V_{4,3}$ : % coeficiente de gini	GINI2018
	$V_{4,4}$ : Ingreso por unidad del gasto	IPUG2018
	$V_{4,5}$ : linea de pobreza	LP2018
	$V_{4,6}$ : linea de pobreza extrema	LPE2018
	$V_{4,7}$ : % brecha de pobreza	BP2018
	$V_{4,8}$ : % brecha de pobreza extrema	BPE2018
	$V_{4,9}$ : % severidad de la pobreza	SP2018
	$V_{4,10}$ : % severidad de la pobreza extrema	SPE2018
G_5: año 2020	$V_{5,1}$ : % pobreza monetaria	PM2020
	$V_{5,2}$ : % pobreza monetaria extrema	PME2020
	$V_{5,3}$ : % coeficiente de gini	GINI2020
	$V_{5,4}$ : Ingreso por unidad del gasto	IPUG2020
	$V_{5,5}$ : linea de pobreza	LP2020
	$V_{5,6}$ : linea de pobreza extrema	LPE2020
	$V_{5,7}$ : brecha de pobreza	BP2020
	$V_{5,8}$ : brecha de pobreza extrema	BPE2020
	$V_{5,9}$ : severidad de la pobreza	SP2020
	$V_{5,10}$ : severidad de la pobreza extrema	SPE2020

**Tabla 5-1.:** Listado de indicadores de Pobreza Monetaria y etiquetas

En la **Fase 1**, *limpieza y preprocesamiento* se verifican las condiciones de los datos, búsqueda de valores faltantes, posibles errores de digitación y se revisan las métricas de las variables.

Se toma la decisión de tomar todas las variables como activas y cuya unidad de medida se encuentra en porcentaje (%). Las variables IPUG y LP que se encuentran en pesos Colombianos, para todos los análisis son deflactadas al año base 2020 para hacer posible las comparaciones en el tiempo.

En la **Fase 2**, *análisis exploratorio de datos* se toman los datos de los años 2012, 2016 y 2020 como año inicial, intermedio y final respectivamente, con el fin de resumir el comportamiento de todos los años. En este análisis se realiza una caracterización descriptiva de los departamentos en función de sus indicadores de Pobreza Monetaria. Se calcula promedio, mínimo y máximo, mediana, varianza y coeficiente de variación. Además, se realizan gráficas de la evolución de la incidencia, severidad, brecha y líneas de la Pobreza Monetaria.

En la **Fase 3**, el *Análisis Factorial Múltiple* se desarrolla sobre la tabla de datos consolidada (Tabla 5-2) y estandarizada para eliminar el efecto de las escalas. Esta se compone de 24 departamentos con cinco grupos, cada uno con diez variables para un total de 50 variables. Esta tabla contiene información que apunta hacia el desarrollo de los objetivos de este estudio.

<i>Departamento</i>	<i>Años</i>				
	G1:(2012)	G2:(2014)	G3:(2016)	G4:(2018)	G5:(2020)
$D_1$	$V_{11} \cdots V_{10}$	$V_{21} \cdots V_{210}$	$V_{31} \cdots V_{310}$	$V_{41} \cdots V_{410}$	$V_{51} \cdots V_{510}$
$D_2$	% ... %	% ... %	% ... %	% ... %	% ... %
$\vdots$	$\vdots \vdots$	$\vdots \vdots$	$\vdots \vdots$	$\vdots \vdots$	$\vdots \vdots$
$D_{24}$	% ... %	% ... %	% ... %	% ... %	% ... %

**Tabla 5-2.:** Estructura final de datos para el AFM.

De acuerdo a los resultados del AFM, se retienen las dos primeras componentes principales mediante el criterio de varianza acumulada, que agrupa aproximadamente el 85 % de variabilidad del conjunto de datos. La representación de las variables se realizó mediante el círculo de correlaciones, y el gráfico de los individuos mediante la proyección de estos. Se evaluaron los cosenos cuadrados como medida de calidad de la representación y las contribuciones de las variables e individuos a los ejes factoriales. También se calcularon los coeficientes Lg y RV

Finalmente, en la **Fase 4** *Análisis de conglomerados*, se desarrolló un análisis tomando como insumo las coordenadas globales de los individuos del AFM para agrupar departamentos con similitud en su estructura de Pobreza Monetaria durante los años 2012 a 2020. Con el objetivo de evaluar los cambios en la estructura de pobreza presentados en la ventana de tiempo que comprende los años 2012 a 2020 por los departamentos, se realizó otro análisis clúster sobre las coordenadas parciales de los individuos del AFM. Los resultados se muestran en dos dendrogramas obtenidos, uno para las coordenadas globales y otro para las coordenadas parciales de los individuos del AFM. Se utilizó el paquete *NbClust*, éste calcula 30

índices, estos son: “kl”, “ch”, “scott”, “marriot”, “trcovw”, “tracew”, “friedman”, “rubin”, “cindex”, “db”, “silhouette”, “duda”, “pseudot2”, “beale”, “ratkowsky”, “ball”, “ptbiseria”, “gap”, “frey”, “mcclain”, “gamma”, “gplus”, “tau”, “dunn”, “hubert”, “sdindex”, “dindex”, “sdbw”, y propone al usuario el mejor esquema de conglomerado a partir de los diferentes resultados obtenidos al variar todas las combinaciones de número de conglomerados, medidas de distancia y métodos de conglomerado. Se utilizaron 5 grupos sugeridos por el paquete para los análisis posteriores.

### 5.3. Software

Se implementó el software estadístico R en su versión 4.0.3 de distribución libre para el procesamiento de los datos. Se utilizaron las librerías *FactoMine*, *FactoExtra* para la extracción de todos los resultados del Análisis Factorial Múltiple y el análisis de Conglomerados.



## 6. Resultados

En este capítulo se presentan los resultados de las Fases 3 y 4 planteadas en la metodología, que implica una caracterización de los departamentos en función de sus indicadores de pobreza monetaria en Colombia durante el periodo 2012 a 2020. Primero, se realiza un Análisis Exploratorio de Datos. Segundo, un Análisis Factorial Múltiple y por último un Análisis de Conglomerados sobre las coordenadas de los individuos(departamentos) obtenidas mediante el AFM. Los indicadores IPUG, LPM y LPME, por estar medidos en pesos colombianos(COP), se encuentran deflactados para el año base 2020, lo que permite su comparación en el tiempo.

### 6.1. Análisis exploratorio

En la Tabla 6-1, se encuentran las estadísticas descriptivas para los distintos indicadores que miden la Pobreza Monetaria en Colombia; Coeficiente de Gini, Severidad, Brecha, Incidencia, Ingreso per cápita de la unidad de Gasto de 24 departamentos Colombianos durante los años 2012, 2016 y 2020. Es decir, al inicio, intermedio y final del periodo de evaluación.

#### 6.1.1. Pobreza Monetaria o Incidencia

Los resultados muestran que en promedio para los 24 departamentos el indicador de Incidencia de Pobreza Monetaria durante el año 2012 se situó en 47.95 %, en el año 2016 estuvo en 42.56 % y para el año 2020 fue de 46.61 %. Esto refleja una reducción de 5.14 % de la Pobreza Monetaria desde el año 2012 hasta el año 2016, mientras que desde el año 2016 hasta el año 2020 se incrementó un 4.05 %.

También se observa que durante los años 2012 y 2016 se presentaron valores máximos en la Pobreza Monetaria, que corresponden al 73.90 % y 67.20 % respectivamente de la población del departamento de Chocó. De acuerdo a las proyecciones realizadas por el DANE (2015), Chocó para el año 2016 tenía una población total de 505.016 habitantes. Esto significa que aproximadamente 338.360 personas tenían un ingreso per cápita de la unidad de gasto por debajo de la línea de pobreza, que para este departamento, en el año 2016, fue de \$287.690. Como consecuencia esto les impide tener acceso a una canasta básica de bienes y servicios básicos necesarios para desarrollar un estándar de vida adecuado.

	Mínimo	P25	Mediana	Media	P75	Máximo	Desviación	C.V
PM.2012	26.80	37.83	47.95	47.70	57.40	73.90	13.21	0.28
PM.2016	20.10	30.55	43.25	42.56	53.42	67.20	12.59	0.30
PM.2020	27.00	38.67	44.80	46.61	55.78	66.30	11.20	0.24
PME.2012	3.70	9.00	12.80	15.64	18.18	43.40	9.95	0.64
PME.2016	4.00	7.57	11.55	13.20	15.05	38.50	8.48	0.64
PME.2020	7.20	12.25	15.15	17.49	22.13	39.70	7.99	0.46
Gini.2012	0.46	0.49	0.51	0.52	0.53	0.62	0.04	0.07
Gini.2016	0.43	0.47	0.48	0.49	0.51	0.57	0.04	0.07
Gini.2020	0.46	0.48	0.51	0.51	0.52	0.56	0.03	0.06
SP.2012	4.60	7.60	10.35	11.49	13.12	30.00	6.02	0.52
SP.2016	3.60	6.00	8.25	9.23	10.55	25.20	4.91	0.53
SP.2020	6.80	8.78	10.85	11.84	13.70	22.00	3.99	0.34
SPE.2012	0.70	1.67	2.25	3.16	3.25	14.00	2.87	0.91
SPE.2016	0.40	1.30	1.90	2.65	2.82	12.40	2.47	0.93
SPE.2020	1.40	3.03	4.20	4.41	5.12	9.80	2.02	0.46
BP.2012	6.80	8.78	10.85	11.84	13.70	22.00	3.99	0.34
BP.2016	7.20	11.05	15.95	16.90	19.85	36.90	7.23	0.43
BP.2020	11.10	16.00	18.50	19.88	23.70	33.70	6.03	0.30
BPE.2012	1.30	3.40	4.20	5.80	6.35	21.50	4.61	0.79
BPE.2016	1.00	2.88	3.50	4.84	5.22	18.80	3.85	0.80
BPE.2020	3.10	4.97	6.35	7.12	7.98	16.40	3.35	0.47
IPUG.2012	307462	406629	513572	545273	639799	1159515	190768	0.35
IPUG.2016	331903	422233	537691	569772	663258	1192335	191439	0.34
IPUG.2020	300775	394195	457316	515367	619653	1002293	166634	0.32
LP.2012	262275	283588	305291	309992	321121	457888	40682	0.13
LP.2016	266420	287092	305769	314843	327503	462040	40682	0.13
LP.2020	252842	274129	294319	304803	320511	455036	42958	0.14
LPE.2012	121627	125342	129664	133027	135978	174122	11396	0.09
LPE.2016	131147	134628	137923	141698	143964	181790	11276	0.08
LPE.2020	127121	131542	134077	138534	143371	178604	11419	0.08

**Tabla 6-1.:** Estadísticas descriptivas de los indicadores de Pobreza Monetaria para los años 2012, 2016 y 2020.

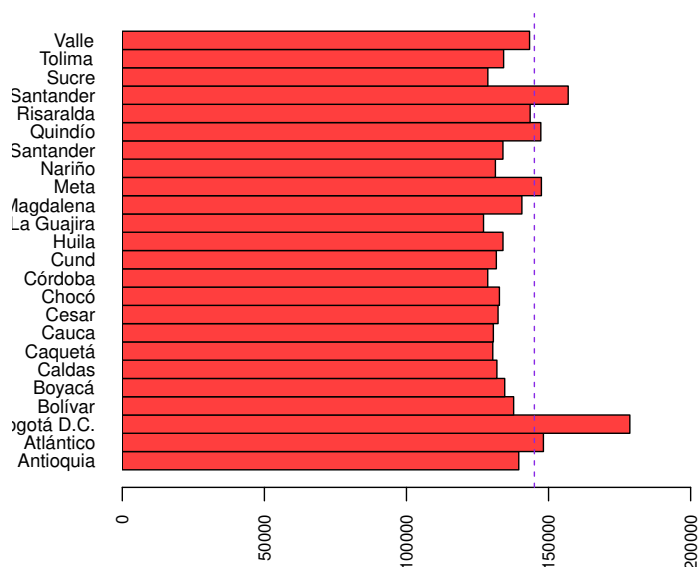
El promedio departamental para el indicador de Pobreza Monetaria, desde el año 2012 hasta el año 2016 presentó una reducción, pasando de 47 % a 42 %. Sin embargo, para el año 2020 se presenta un aumento que lo sitúa en el 45 %, durante este año la Pobreza Monetaria más alta fue para el departamento de La Guajira con el 66.33 % de sus habitantes y el registro más bajo fue para Cundinamarca con el 27 % de su población en esta situación. El indicador

de Pobreza Monetaria Extrema promedio para los 24 departamentos se entiende como la población media que tiene un ingreso per cápita del hogar por debajo de la Línea de Pobreza Extrema. Este indicador se situó en 15.64 % en el año 2012, 13.20 % en el año 2016 y para el año 2020 fue de 17.49 %. El departamento del Chocó en el año 2012 alcanzo un máximo con 43.40 %, en el año 2016 registro un porcentaje de 38.50 % y en el año 2020 el departamento de la Guajira paso a ocupar el primer lugar con 39.70 % siendo este el mayor registro durante el año 2020.

### 6.1.2. Línea o umbral de Pobreza Monetaria

La línea o umbral de pobreza es el costo mínimo por persona de una canasta básica de bienes y servicios (alimentarios y no alimentarios). Durante el año 2012 el umbral promedio departamental se ajustó en \$309.992. Durante este año la Línea de Pobreza más alta fue para Bogotá con \$457.888 y la más baja correspondió al departamento de La Guajira con \$262.275. En el año 2016 el umbral promedio departamental fue de \$314.843. La Línea de pobreza más alta fue para Bogotá con \$462.040 y la más baja para La Guajira con \$266.420. Finalmente en el año 2020 el umbral promedio estuvo en \$304.803. La Línea de Pobreza más alta durante el año 2020 fue para Bogotá D.C con \$455.036 y la más baja fue para La Guajira con \$252.842. En los tres momentos del tiempo, se observa como Bogotá se ha mantenido con la línea de pobreza más alta y la Guajira con la menor. Esta línea se interpreta de la siguiente manera. Una familia de tres personas en la Guajira es considerada en situación de pobreza monetaria durante el año 2020, si el ingreso total del hogar se encuentra debajo de \$758.526. Mientras que una familia de tres personas en el distrito capital durante el año 2020 estaría en situación de Pobreza Monetaria si el ingreso del hogar se encuentra por debajo de \$1.365.108, esta varía de acuerdo al dominio geográfico y desarrollo económico de cada departamento.

En la Figura 6-1, se encuentran las líneas de Pobreza Monetaria Extrema o indigencia calculadas para el año 2020. Los valores situados por arriba del promedio departamental representado por la línea vertical punteada corresponden a los departamentos de Santander, Quindío, Meta y Atlántico. El umbral más extremo corresponde a la unidad administrativa Bogotá. Por otro lado, las situadas por debajo del promedio nacional fueron para los departamentos de Valle, Risaralda, Antioquia, Tolima, Sucre, Nariño, Caldas, Boyacá, Bolívar y Quindío. Si se toma un hogar de tres personas en el Valle del Cauca, el ingreso total del hogar deberá ser de \$435.012 para alcanzar el umbral mínimo de pobreza, para no estar en condición de indigencia.

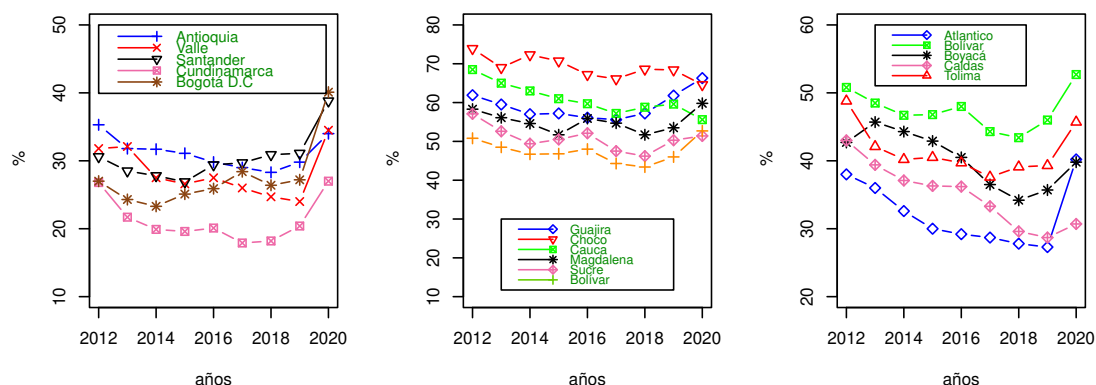


**Figura 6-1.:** Líneas de Pobreza Monetaria Extrema por departamentos en el año 2020 (Cifras en COP).

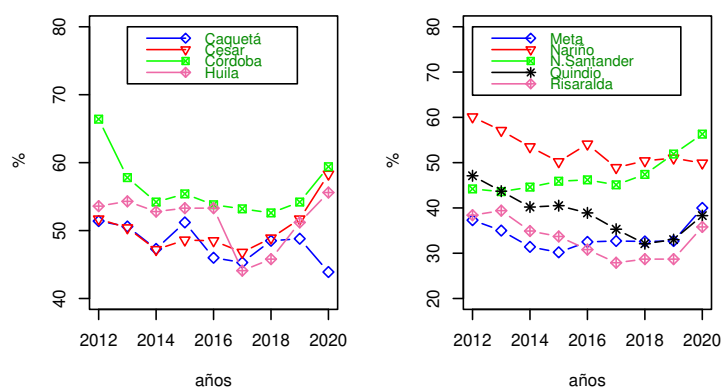
### 6.1.3. Evolución de la Pobreza Monetaria

A continuación, de manera gráfica se presenta la evolución de la Pobreza Monetaria de 24 departamentos evaluados durante una ventana de tiempo que comprende desde el año 2012 hasta el año 2020.

En la Figura 6-2, la primera gráfica corresponde al grupo de departamentos menos pobres, donde se resalta el comportamiento de Cundinamarca por su bajo nivel de incidencia de pobreza. Por otro lado, todos los departamentos de este grupo, desde el año 2019 al 2020, cerraron con un aumento fuerte de pobreza monetaria. En el segundo gráfico, se encuentran los departamentos más pobres; El Chocó presenta niveles del 75 % de pobreza en el año 2012. Para el año 2020 logro reducir su pobreza empatando con el departamento de La Guajira; sin embargo, estos porcentajes siguen muy elevados. En el tercer grupo, se observa que todos los departamentos desde el año 2012 hasta el año 2018 lograron reducir su pobreza, se destaca el comportamiento de Atlántico con niveles bajos de pobreza, sin embargo, todos los departamentos desde el año 2019 al 2020 presentaron un aumento fuerte en la pobreza inclusive superando el año 2012.



**Figura 6-2.:** Incidencia de la Pobreza Monetaria por departamentos, desde el año 2012 hasta el 2020.

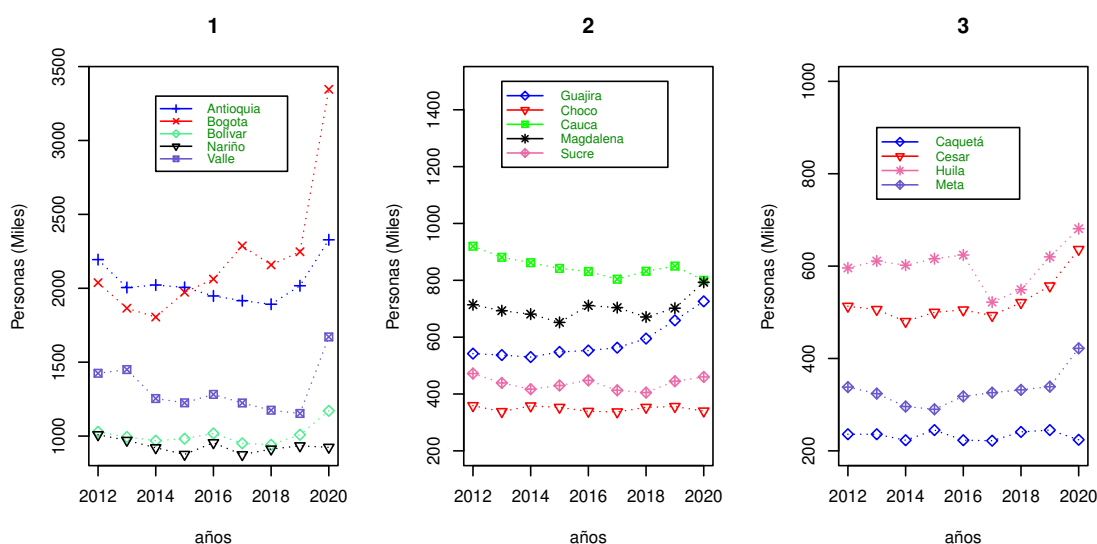


**Figura 6-3.:** Incidencia de la Pobreza Monetaria por departamentos, desde el año 2012 hasta el 2020.

Acorde a la Figura 6-3, de manera general se observa que todos los departamentos de este grupo en el año 2012 tenían más del 50 % de su población en situación de pobreza monetaria. En el primer gráfico se observa que Córdoba presento un comportamiento muy marcado en el año 2012 con al rededor del 70 % de su población en condición de pobreza monetaria. Al finalizar el año 2020 se observa una tendencia al incremento de este indicador para la mayoría de departamentos de este grupo.

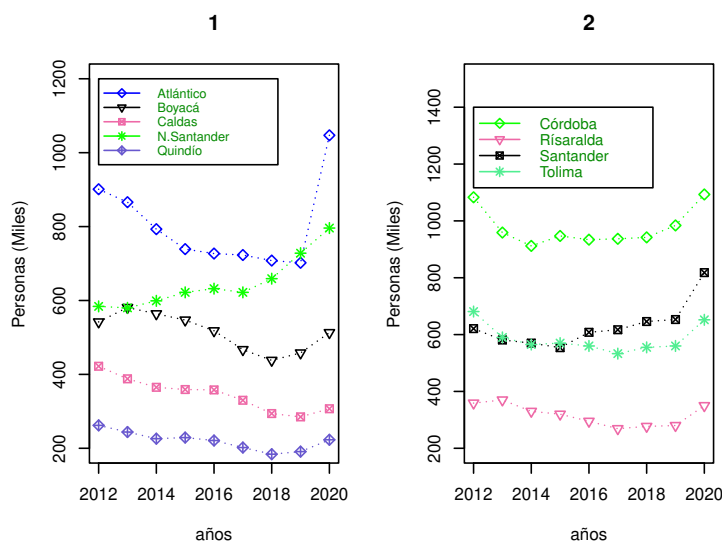
### 6.1.4. Evolución del número de personas en Pobreza Monetaria

A continuación y de manera gráfica, se representa la evolución del número de personas en situación de pobreza monetaria a nivel departamental durante los años 2012 a 2020.



**Figura 6-4.:** Número de personas en Pobreza Monetaria

Acorde a la Figura 6-4 primer gráfico, se destaca el cambio en la pendiente para el Distrito Capital Bogotá desde el año 2019 hasta el 2020. El 46 % de su población, es decir, aproximadamente 3.500.000 de personas, se encontraban en situación de pobreza monetaria. En el segundo gráfico se observa desde el año 2012 hasta el 2020 el departamento de Cauca viene disminuyendo el número de personas en pobreza monetaria. El comportamiento de Chocó es muy nivelado desde durante los años 2012 hasta 2020. La Guajira ha presentado desde el año 2012 hasta el 2020 una tendencia creciente a seguir aumentando su población en situación de pobreza monetaria. En el tercer gráfico los departamentos Cesar, Huila y Meta desde el año 2012 presentan una tendencia al aumento de su población en situación de Pobreza Monetaria. Por otro lado, el departamento del Caquetá presenta un comportamiento muy nivelado en el tiempo.

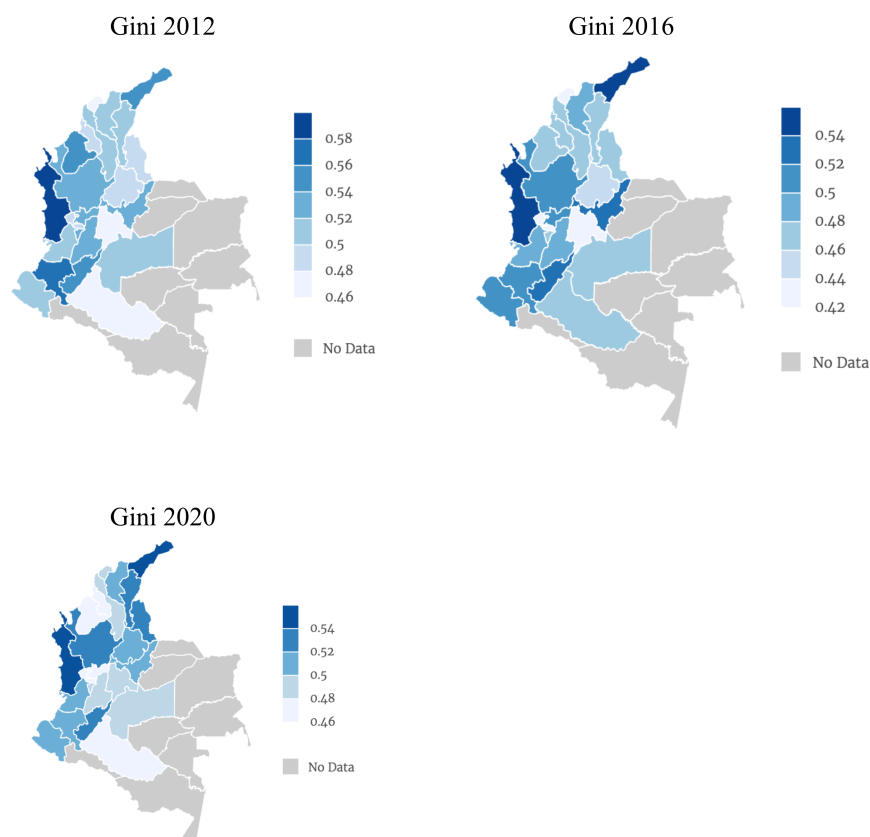


**Figura 6-5.:** Número de personas en Pobreza Monetaria.

De acuerdo a la Figura 6-5, el primer gráfico muestra una disminución desde el año 2012 hasta el año 2018 de la población en pobreza monetaria en los departamentos Boyacá, Caldas, Quindío y Atlántico. El segundo gráfico, de manera general los departamentos de Córdoba, Risaralda, Santander y Tolima presentaron un comportamiento muy constante. Desde el año 2019, el número de personas en pobreza monetaria, aumenta en todos los departamentos.

### 6.1.5. Coeficiente de Gini

De acuerdo a la Tabla 6-1, el Coeficiente de Gini durante el año 2012 determinó a Chocó como el departamento de mayor desigualdad en el ingreso de sus habitantes, mientras que el de menor desigualdad la registro el departamento de Atlántico. En el año 2016 la mayor desigualdad en el ingreso fue de nuevo para Chocó y la menor desigualdad fue para Cundinamarca. Finalmente, para el año 2020 la mayor desigualdad fue para Chocó y la menor para Risaralda. También se observa que el promedio departamental del coeficiente de Gini para los años 2012, 2016 y 2020 fue de 0.52, 0.49, 0.51 respectivamente. Esto refleja que en Colombia durante estos años la desigualdad departamental en la distribución del ingreso de sus habitantes se ha mantenido en condiciones muy similares.



**Figura 6-6.:** Coeficiente de Gini 2012- 2020.

En La Figura 6-6, representa la evolución del coeficiente Gini mediante tres mapas de calor. Los departamentos de Córdoba, Caldas, Risaralda, Quindío, Tolima, Córdoba y Sucre, desde el año 2012 al 2020 han logrado mejorar su distribución en el ingreso. Se observa como el departamento de Chocó y La Guajira, han mantenido la alta desigualdad del ingreso en estos tres instantes de tiempo.

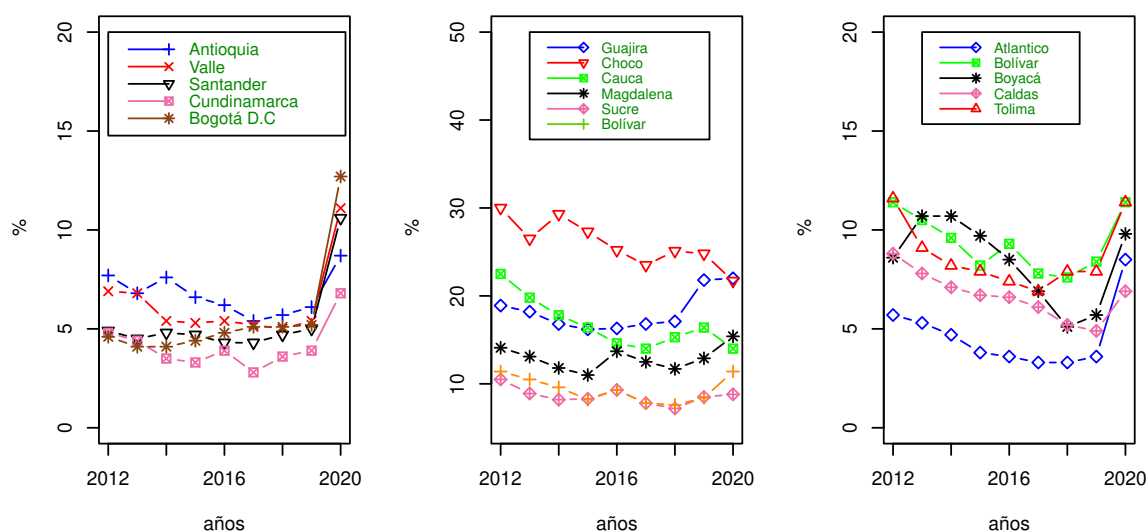
### 6.1.6. Severidad de la Pobreza Monetaria

La severidad es una medida complementaria que permite evidenciar la distribución de los ingresos entre los pobres y pobres extremos, aspecto que no captura la incidencia o la brecha. Aunque es de difícil interpretación, el indicador tiene la propiedad de incrementarse a medida que los pobres sean más pobres. Por lo tanto, la utilidad de la medida radica en su habilidad por ordenar las distribuciones de pobreza, es decir, establecer *rankings* de lugares donde la pobreza se vive de manera más severa (Ravallion, 1992).

Acorde a la Figura 6-7, el primer gráfico destaca a Cundinamarca como el departamento



con menor severidad, mientras que la unidad administrativa Bogotá D.C, es donde se vive la pobreza con mayor condición de desigualdad. Esto quiere decir que en Bogotá se vive la pobreza en condiciones más severas que en el Valle del Cauca y Santander. En el segundo gráfico, se destaca el departamento de Chocó por su alto porcentaje de severidad durante el año 2012 y como este ha disminuido en el año 2020. El departamento de La Guajira muestra una reducción hasta el año 2018; sin embargo, durante el cambio de los años 2019 y 2020 presento un aumento igualándose con el Chocó. En el tercer gráfico encontramos que Atlántico resulto el de menor severidad dentro de este grupo. Todos los departamentos presentan un comportamiento a reducir la severidad hasta el año 2019, donde sus pendientes pronunciadas indican un repunte en el aumento de la severidad de la pobreza para el año 2020.



**Figura 6-7.:** Severidad de la Pobreza Monetaria.

Acorde a la Figura 6-8, el primer gráfico representa a Córdoba como el departamento de mayor severidad durante el año 2012 seguido de Huila y Cesar. Todos los departamentos de este grupo apuntan a reducir la severidad hasta el año 2019. Durante el año 2020, estos departamentos cierran con unas condiciones mucho más severas que el año inicial 2012. En el segundo gráfico se observa una tendencia de todos los departamentos a disminuir este indicador hasta el año 2019. Sin embargo, en el año 2020 incrementan su severidad.

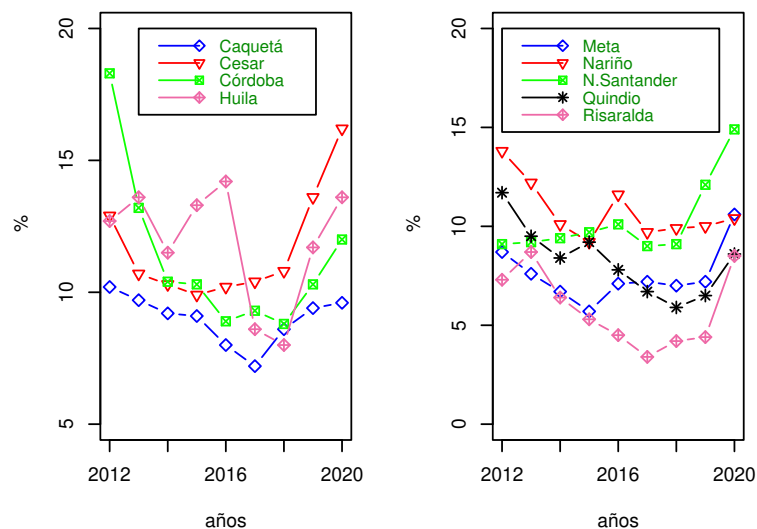
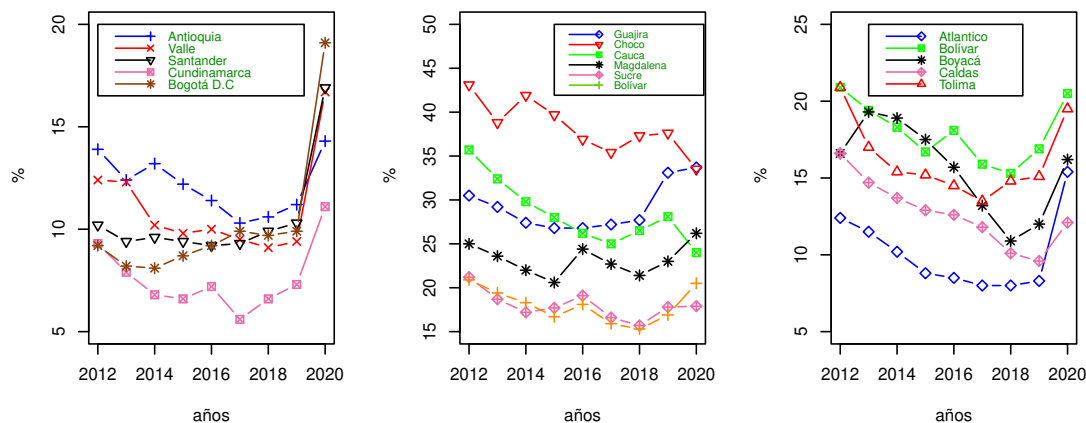


Figura 6-8.: Severidad de la Pobreza Monetaria.

### 6.1.7. Evolución de la Brecha de la Pobreza Monetaria

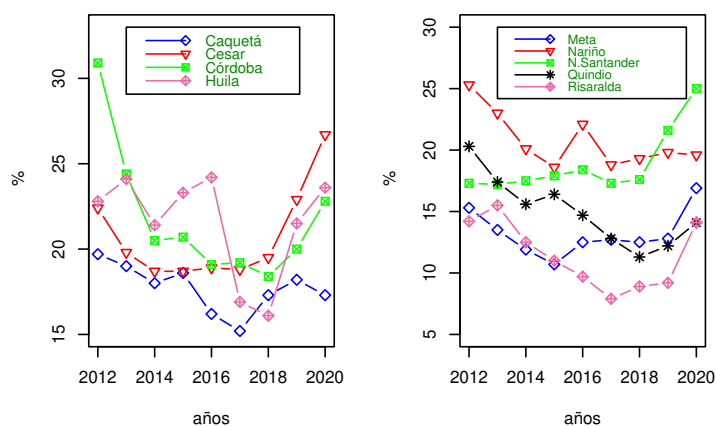
La brecha mide la cantidad de dinero que le falta a una persona pobre para superar esta condición (DANE, 2018).

La Figura 6-9, el primer gráfico indica que Cundinamarca en el año 2020 presentó la brecha más baja entre los departamentos de este grupo, es decir, que en promedio la población de Cundinamarca requiere de 11.1 % de su línea de pobreza calculada en 269.745\$ para superar esta condición. Esto es equivalente a 29.671\$ por persona. Por otro lado, la brecha más alta se observa para la unidad administrativa Bogotá, con 19.1 %, que equivale de acuerdo a su línea de pobreza calculada en 455.036\$ a 86.456\$ por persona. En el segundo gráfico se observa una tendencia de todos los departamentos desde el año 2012 hasta el 2019 a reducir las brechas. Sin embargo, para el año 2020 la Guajira con una brecha de 33.7. %, es la brecha más alta calculada para todos los departamentos y que coincide con el valor máximo que se muestra en la **Tabla 6-1**. Esto significa que para una persona en este departamento salir de la condición de pobreza necesitaría en promedio de 33.7. % de su línea calculada en 252.842\$ lo que equivale a 83.437\$ por persona. En el tercer gráfico destaca Bolívar con mayor brecha de pobreza, todos los departamentos de este grupo terminan el año 2020 con una brecha hacia el alza.



**Figura 6-9.:** Brecha de la Pobreza Monetaria.

La Figura 6-10, en el primer gráfico se observa una tendencia a la reducción de la brecha de todos los departamentos desde el año 2012 hasta el año 2018. Se destaca el departamento de Córdoba por su elevada brecha hacia el año 2012. En el segundo gráfico se muestra como el comportamiento de la brecha para estos departamentos no ha presentado muchos cambios.



**Figura 6-10.:** Brecha de la Pobreza Monetaria.

## 6.2. Análisis Factorial Múltiple (AFM)

### 6.2.1. Análisis de factores

En el Análisis Factorial Múltiple es importante determinar que factores se van a mantener para realizar los análisis posteriores. Estos factores se determinan mediante los valores propios que se encuentran calculados en la Tabla 6-2, que representan el porcentaje de variabilidad explicada (inercia) por el conjunto de datos.

	Valor propio	porcentaje de varianza	porcentaje de varianza acumulada
comp 1	4.68	67.25	67.25
comp 2	1.40	20.08	87.33
comp 3	0.29	4.16	91.49
comp 4	0.22	3.22	94.70

Tabla 6-2.: Valores propios del AFM

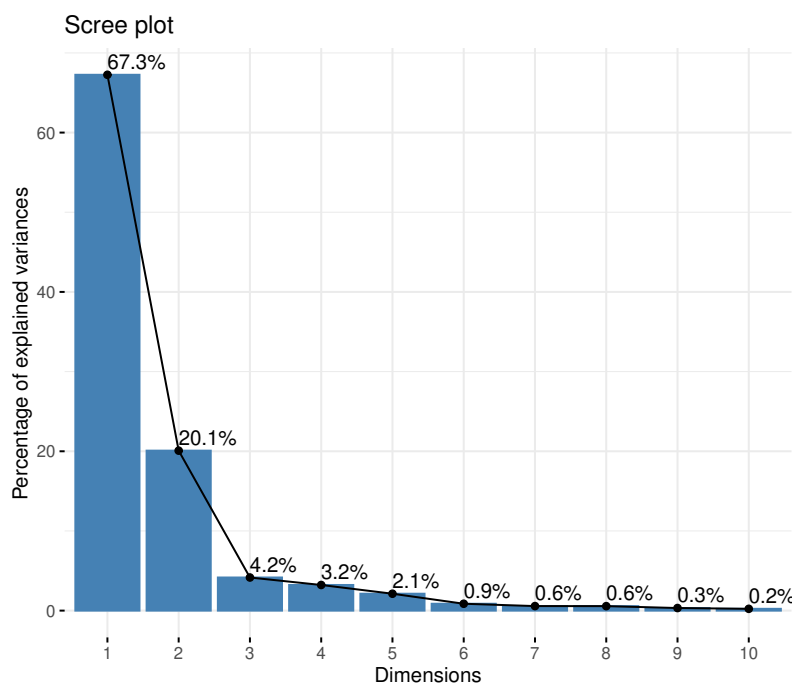
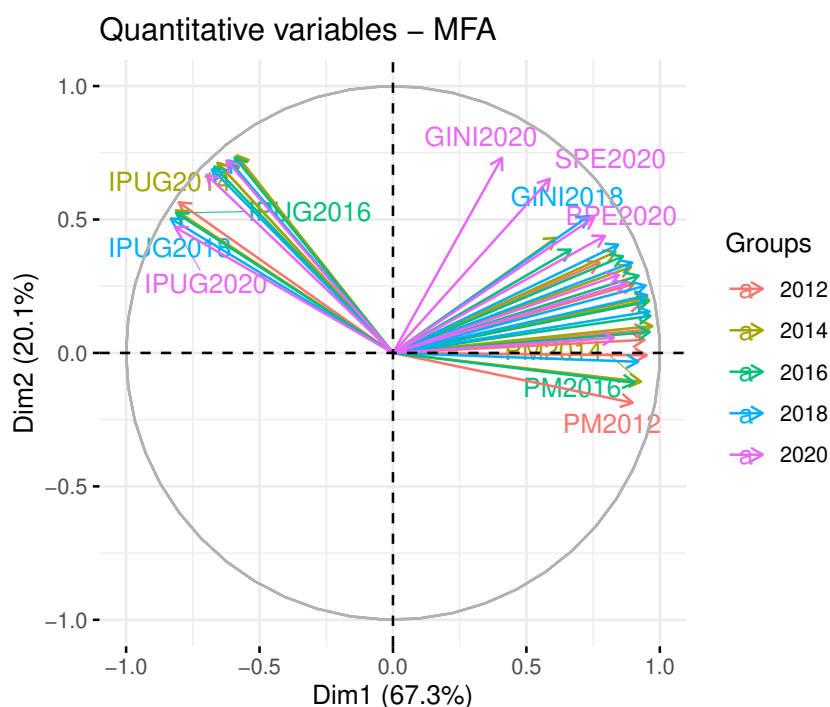


Figura 6-11.: Porcentaje de varianza explicado.

Acorde a la Figura 6-11, se observa el porcentaje de inercia explicada para los ejes factoriales conformados. De acuerdo al conjunto de datos, para el primer y segundo eje factorial se obtuvo un 67.3% y 20.1% de varianza explicada respectivamente. Estos dos ejes factoriales acumulan el 87.4% de variabilidad por los indicadores de pobreza monetaria durante los años 2012 a 2020, motivo por el cual se mantendrán para los análisis posteriores.

### 6.2.2. Análisis global de las variables



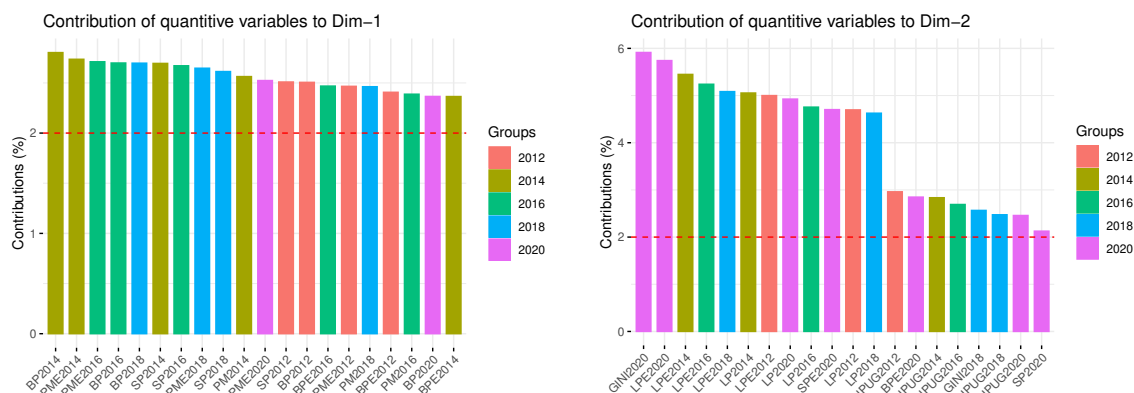
**Figura 6-12.:** Círculo de correlaciones del AFM, Planos F1 y F2.

Sobre la Figura 6-12, se observa que la mayoría de las variables que miden la pobreza monetaria se encuentran más relacionadas con la primera componente principal y unas pocas variables con la segunda componente principal. La dirección de las variables asociadas con el primer factor evidencia altas correlaciones positivas con este componente principal que está relacionada con la pobreza extrema. También se observa que los indicadores Brecha, Pobreza Monetaria y Severidad se relacionan más durante los años 2012 a 2018. La segunda componente principal se encuentra más relacionada con las variables que miden la desigualdad, el umbral extremo de pobreza y el ingreso para los años 2016 a 2020. Por otro lado, se presenta una relación inversa entre la Pobreza Monetaria y el Ingreso por unidad del gasto. El primer eje principal se puede denominar como un eje de Pobreza Monetaria Extrema. El segundo eje principal como un eje que está ligado a la desigualdad y el ingreso.

### 6.2.3. Contribuciones de las variables

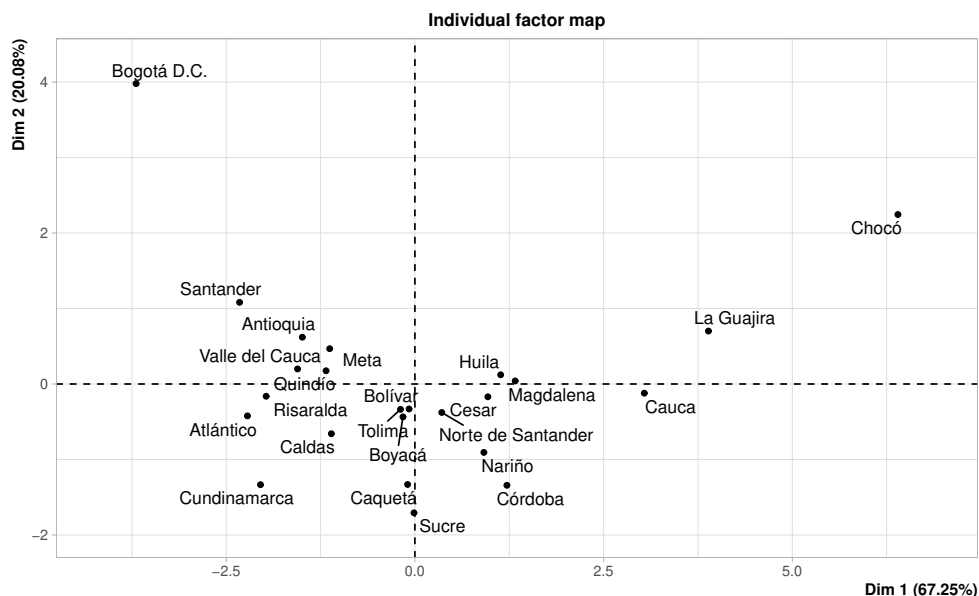
En la Figura 6-13, la primera gráfica ilustra que la brecha en el año 2014 junto con la Pobreza Monetaria Extrema y la severidad en los años 2012 y 2016 son las contribuciones que más aportan a la construcción de la primera componente principal. En la segunda gráfica se observa que el coeficiente de Gini para el año 2020 y la línea de pobreza extrema de los años

2014, 2016 y 2018 aportaron más a la construcción de la segunda componente principal. En cuanto a la primera dimensión, se observa que esta quedó conformada en mayor parte por las brechas y severidades extremas de los años 2014, 2016 y 2018. La segunda dimensión presentó mayor asociación a los indicadores que miden la desigualdad, el ingreso y el umbral extremo de pobreza del año 2020.



**Figura 6-13.:** Contribuciones de las variables al primer y segundo eje factorial del AFM

#### 6.2.4. La nube de individuos



**Figura 6-14.:** Representación global de los departamentos en el primer y segundo plano del AFM.

Acorde a la Figura 6-14 y la denominación realizada de los ejes principales, se tiene que los departamentos situados en el primer cuadrante son los de mayores condiciones de po-

breza extrema y desigualdad. Como se observa, el departamento de **Chocó** es el de mayor condición de pobreza y desigualdad extrema. Por su parte la **Guajira** es un departamento con alta condición de pobreza extrema, pero no tan desigual como lo fue Chocó. De manera análoga en el segundo cuadrante se tienen los departamentos en menor condición de pobreza extrema y desigualdad de pobreza. Se observa como el distrito capital **Bogotá** por su alta coordenada en la segunda componente resulta como el de menor condición de pobreza, pero el de mayor desigualdad en el ingreso de sus habitantes. Departamentos como **Santander**, **Antioquia**, **Meta** y **Valle del Cauca** por su cercanía presentan condiciones similares de pobreza y desigualdad. En el tercer cuadrante se observa que **Cundinamarca** es un departamento que no presenta condiciones de pobreza tan extremas y ni condiciones de desigualdad elevadas. **Atlántico**, presenta menores condiciones de pobreza, a pesar de estar cercano a Cundinamarca, se vive la desigualdad de manera más severa. En el cuarto cuadrante, El departamento de **Córdoba** se sitúa en una condición pobreza extrema pero hay menos desigualdad en sus habitantes.

6.2.5. Calidad de la representación

En la Figura 6-15 se observa la calidad de la representación o cosenos cuadrados, donde los departamentos de Chocó, Santander, La Guajira, Cauca y Bogotá como unidad administrativa quedaron mejor representados en las primeras dos componentes principales conformadas.

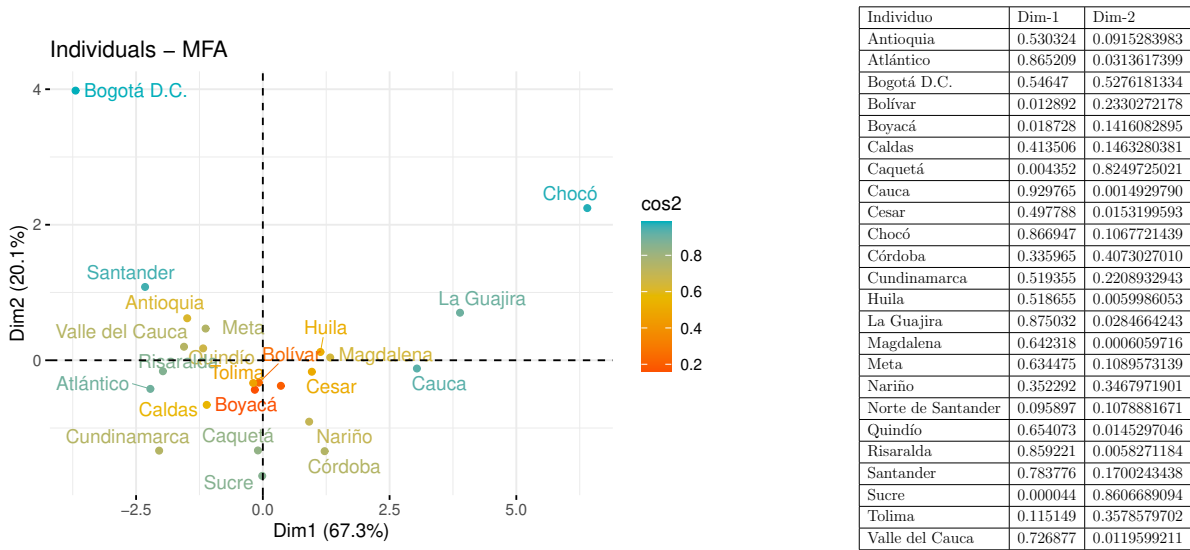
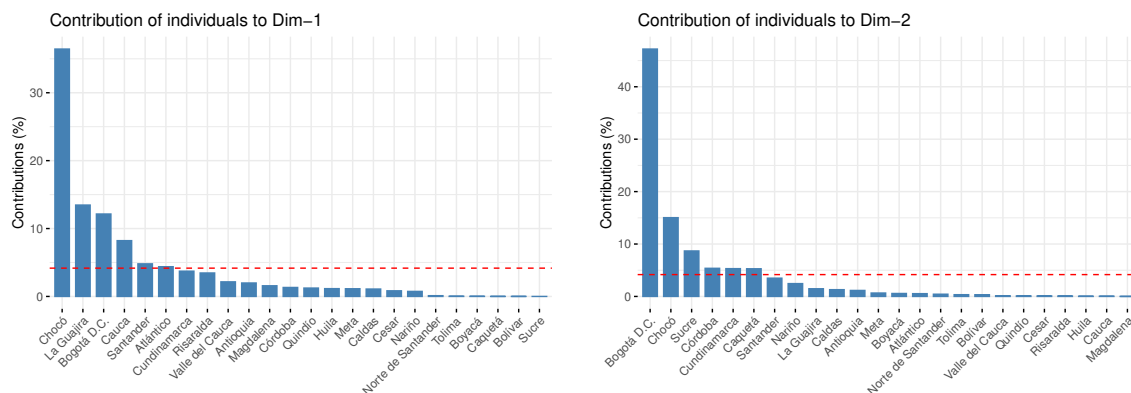


Figura 6-15.: Cosenos cuadrados del AFM

### 6.2.6. Contribuciones de los individuos

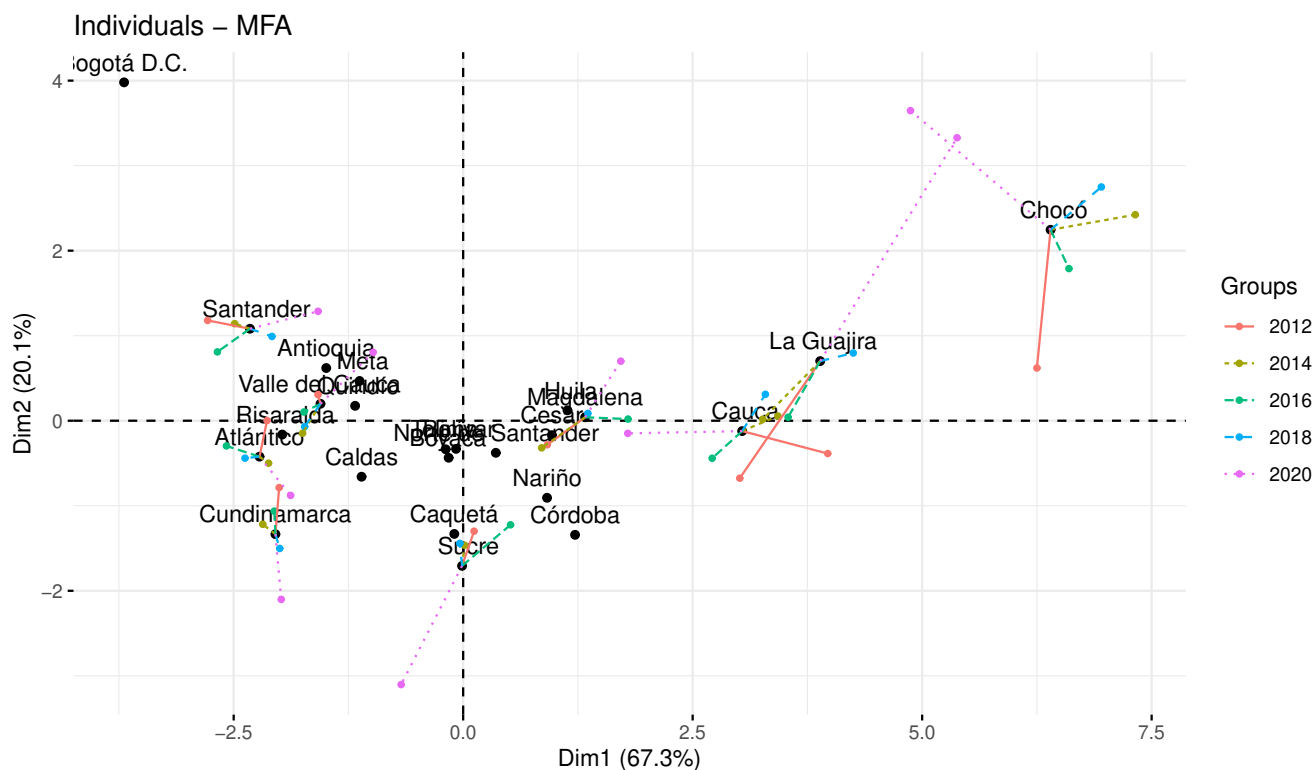


**Figura 6-16.:** Contribuciones de los individuos al primer y segundo eje factorial del AFM

Acorde a la Figura 6-16, los departamentos que resultan ser los individuos excéntricos de la nube que se muestra en la Figura 6-14, son los que más contribuyen a la formación de la primera y segunda componente principal porque recogen la mayoría de la inercia explicada por los ejes factoriales. Para la primera componente, el grupo de Chocó, la Guajira, Bogotá y Cauca. En la segunda componente el grupo de Bogotá, Chocó, Sucre y Córdoba.



### 6.2.7. Análisis grupal



**Figura 6-17.:** Representación de los departamentos a nivel global y grupal del AFM

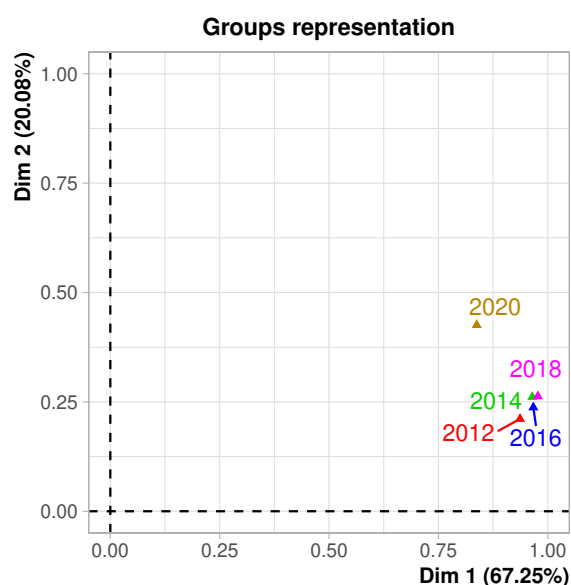
Acorde a la Figura 6-17, permite observar los individuos parciales que resultan de realizar el análisis con cada uno de los grupos. El punto conformado por cada individuo representa el baricentro o centro de gravedad de todos los grupos de variables. La representación de la nube global, que se obtiene superponiendo las cinco nubes generadas por los cinco grupos, resultan de la proyección de cada grupo a su subespacio y permite observar las trayectorias presentadas por los departamentos entre los años 2012 a 2020. Dicho esto podemos caracterizar los departamentos de la siguiente manera:

**Chocó** se interpreta, en general, que ha sido un departamento con mucha condición de pobreza extrema y desigualdad. Se observa en el año 2020 un desplazamiento en la segunda componente principal y una pequeña reducción en la primera, interpretado como un aumento en la desigualdad y una pequeña reducción en la pobreza extrema.

La **Guajira** se representa como un departamento con mucha condición de pobreza extrema, pero no tan desigual como lo ha sido Chocó. La trayectoria en el año 2020 evidencia un fuerte incremento en las coordenadas de las dos componentes principales, que se interpreta como un gran aumento en las condiciones de pobreza extrema y desigualdad en sus habitantes. **Sucre** ha sido un departamento que no ha presentado condiciones de pobreza

muy extremas ni desigualdad, su trayectoria durante el año 2020 apunta a reducir más la pobreza y desigualdad.

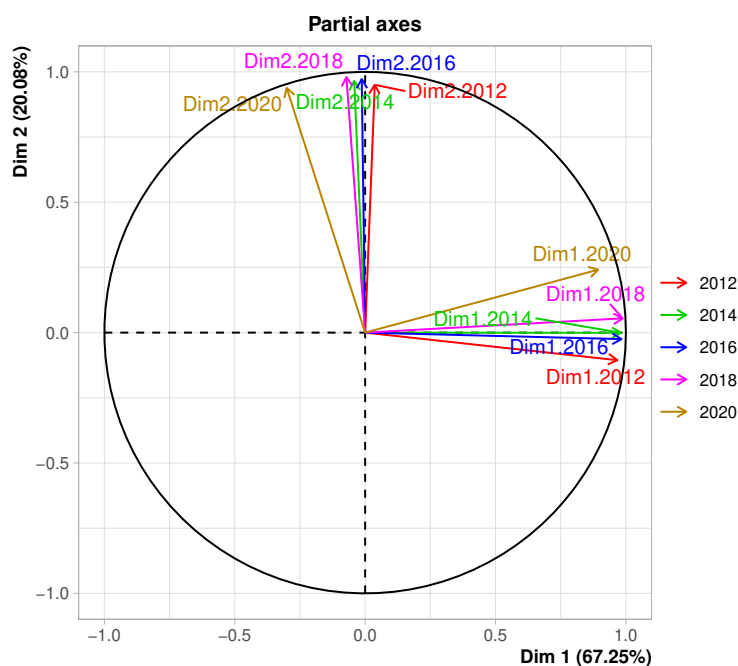
**Cundinamarca** ha sido en el tiempo un departamento con menos condición de pobreza y desigualdad, no presenta desplazamientos muy pronunciados en sus trayectorias salvo en el año 2020, que se observa una reducción en la desigualdad. Este poco movimiento en sus trayectorias refleja que en el tiempo ha permanecido en condiciones similares de pobreza y desigualdad. **Atlántico** muestra que se ha mantenido durante el tiempo con condiciones de pobreza igual, salvo en el año 2020 donde se observa una reducción tanto en su condición de pobreza como desigualdad. **Santander** ha presentado poca condición de pobreza de manera general; sin embargo, se observa un leve incremento en la desigualdad y la pobreza durante el año 2020. **Cauca** un departamento que desde el año 2012 muestra una condición de pobreza muy extrema, en el año 2020 logra reducirla y mantiene su condición de desigualdad. Se destaca que Cauca en el año 2018 había tenido condiciones más fuertes de pobreza y desigualdad. **Nariño** en general ha sido un departamento con condición de pobreza pero con una desigualdad muy baja. Su baja coordenada en la primera y segunda componente lo sitúa en el año 2020 como un departamento en menor condición de pobreza y desigualdad en sus habitantes. **Norte de Santander**, muestra en el año 2012 condiciones de pobreza y desigualdad mucho menores que su baricentro y se observa que su trayectoria en el año 2020 presentó un fuerte desplazamiento en las dos componentes principales, interpretado como un gran aumento en la condición de pobreza extrema y la desigualdad entre sus habitantes.



**Figura 6-18.:** Representación de los grupos del AFM.

Sobre la Figura 6-18, se observa la representación de los grupos activos. Las altas coordenadas de los cinco grupos sobre el primer componente principal, evidencian una fuerte aporte

de inercia dentro de las sub nubes conformadas. Dicho esto se representa la evolución de la estructura de la pobreza monetaria en los distintos años identificando similitudes o diferencias. Se presentó un agrupamiento entre los años 2014 a 2018 que sugiere estructuras similares entre los departamentos. El año 2020 se observa alejado de todos los grupos, lo cual indica un cambio en la estructura de la pobreza monetaria a nivel departamental y su desplazamiento en la segunda componente principal evidencia un aumento de manera general en las condiciones de desigualdad.



**Figura 6-19.:** Proyección de las primeras componentes principales de cada grupo del AFM

De acuerdo a la Figura 6-19, se observa la relación entre los ejes principales globales del AFM y los que se obtienen de analizar cada grupo por separado. Los grupos de los años 2012 a 2018 están estrechamente relacionados con la primera y segunda dimensión del AFM. La dimensión del grupo 2020 se muestra rotada de la primera y segunda componente principal que puede ser interpretada como un cambio en la estructura de relación con respecto a los demás.

	2012	2014	2016	2018	2020	MFA
2012	1.05	0.99	0.94	0.96	0.77	1.01
2014	0.99	1.08	1.02	1.00	0.82	1.05
2016	0.94	1.02	1.07	0.99	0.87	1.04
2018	0.96	1.00	0.99	1.08	0.94	1.06
2020	0.77	0.82	0.87	0.94	1.18	0.98
MFA	1.01	1.05	1.04	1.06	0.98	1.10

**Tabla 6-3.:** Coeficientes  $Lg$

En la Tabla 6-3, se presentan los coeficientes  $Lg$  calculados. Su magnitud brinda información con respecto a la dimensionalidad y mide el grado de similitud o deformación entre los conjuntos de variables (Bossio, 2010). Se observa que estos coeficientes en general están cercanos a uno. Significa que los grupos de variables que miden la Pobreza Monetaria en Colombia tuvieron una dinámica similar entre los años 2012 a 2018. En el año 2020 se observa una disminución en los coeficientes  $Lg$  con respecto a los demás años que puede ser interpretada como un pequeño cambio en la similitud de éste grupo con respecto a los demás.

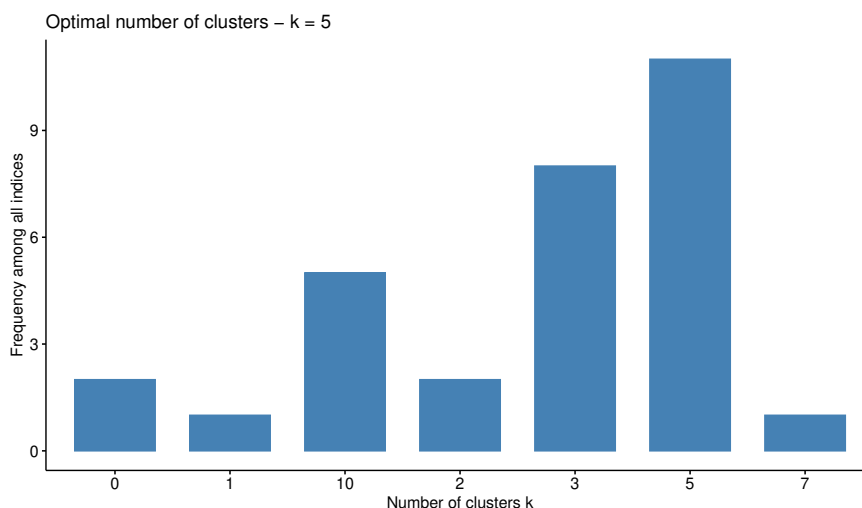
	2012	2014	2016	2018	2020	MFA
2012	1.00	0.93	0.89	0.90	0.69	0.94
2014	0.93	1.00	0.94	0.93	0.73	0.96
2016	0.89	0.94	1.00	0.93	0.78	0.96
2018	0.90	0.93	0.93	1.00	0.83	0.98
2020	0.69	0.73	0.78	0.83	1.00	0.86
MFA	0.94	0.96	0.96	0.98	0.86	1.00

**Tabla 6-4.:** Coeficientes  $Rv$

De acuerdo a E. Abascal y Landaluce (2002), los coeficientes  $Rv$  se obtienen 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 uno (las nubes que representan a los grupos son homotéticas). En la Tabla 6-3, se presentan los coeficientes  $Rv$  donde se observa una alta relación entre las variables de los distintos grupos conformados que miden la Pobreza Monetaria en Colombia.

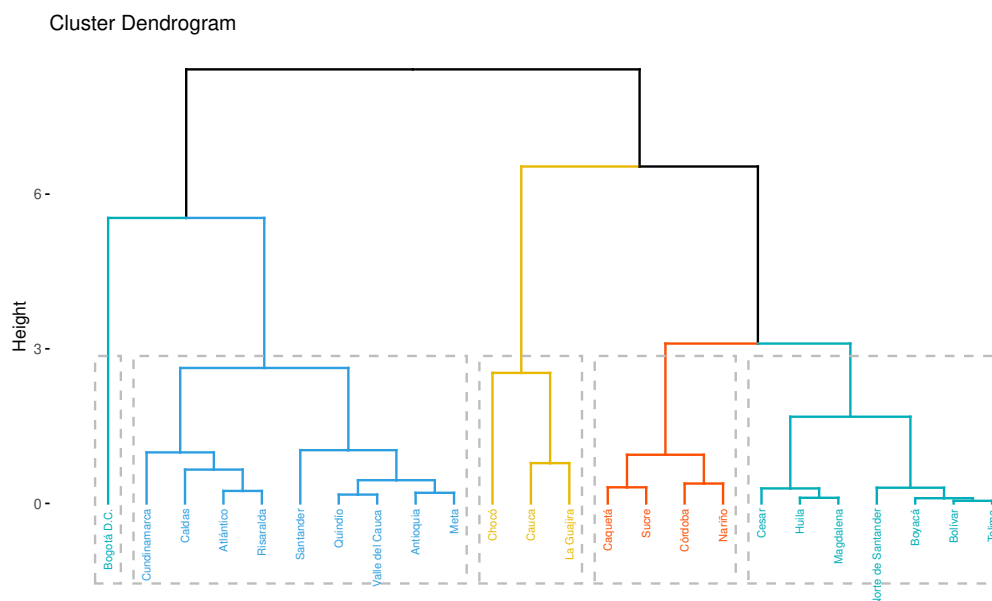
### 6.3. Análisis de Conglomerados - AC

Para el agrupamiento de los departamentos de acuerdo a su similitud, se realizó una clasificación jerárquica por el método *Ward*.



**Figura 6-20.:** Número de clúster óptimo

La Figura 6-20, muestra los distintos resultados del cálculo de 30 índices que se explicaron en la metodología para determinar el mejor número de clústeres. Como se observa el número óptimo fue de 5 clústeres de acuerdo a los índices.



**Figura 6-21.:** Dendrograma sobre las coordenadas globales de los individuos del AFM

Acorde a la Figura 6-21, se traza el corte en cinco grupos conformados mediante la clasificación jerárquica por el método de *Ward*, utilizando las coordenadas globales de los individuos que resultaron del AFM para realizar el agrupamiento. Estos grupos identifican similitud en las condiciones de pobreza a nivel departamental donde encontramos que:

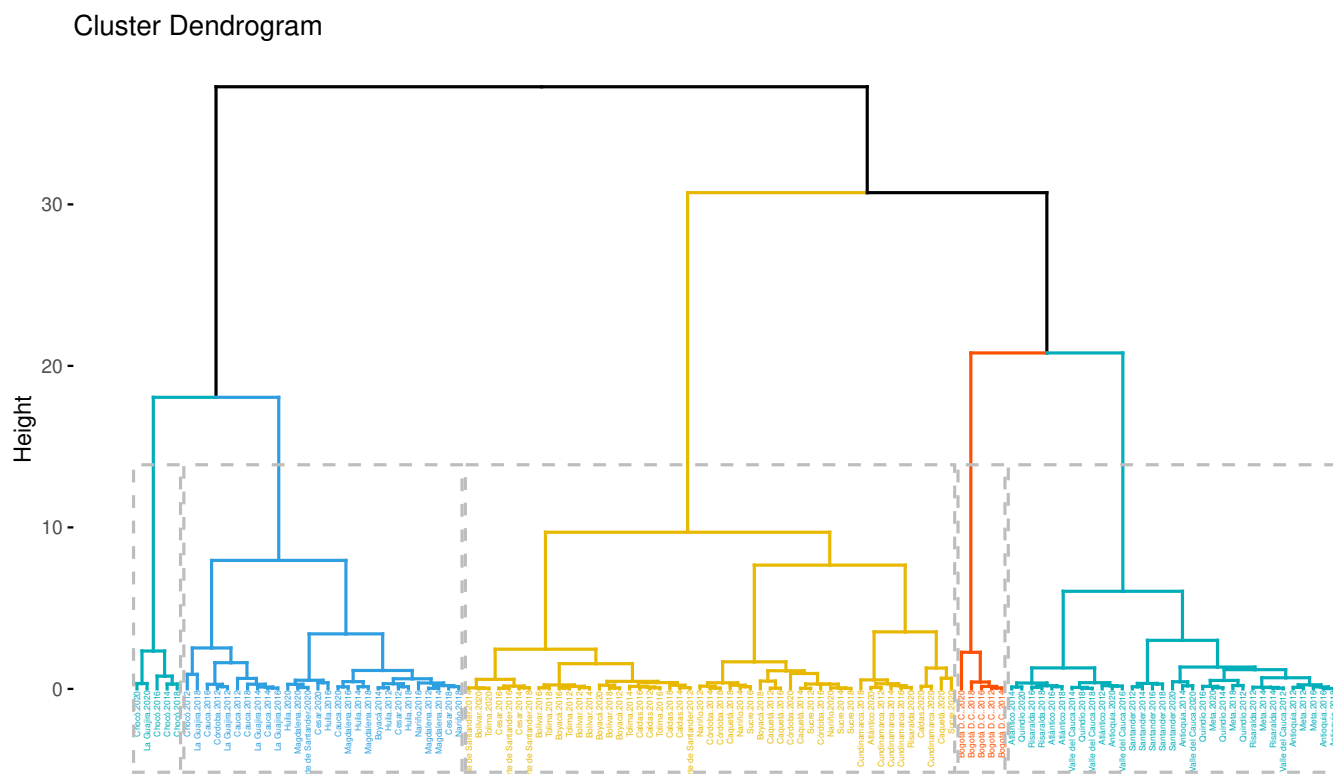
El *grupo 1*, quedo determinado solo por Bogotá. De los resultados del AFM se sabe que esta unidad administrativa resulta ser un individuo excéntrico, motivo por el cual cuando se calculan las distancias no se encuentra asociación alguna entre clústeres.

El *grupo 2*, conformado por los departamentos de Cundinamarca, Caldas, Atlántico, Risaralda, Santander, Quindío, Valle del Cauca, Antioquia y Meta. Estos son los departamentos que conforman la región administrativa de Colombia y de planificación especial. Esta región en términos económicos es la más importante del país.

El *grupo 3*, constituido por los departamentos de Chocó, la Guajira y Cauca. Estos son los de mayor condición de pobreza y desigualdad extrema. Se encuentran situados al rededor de la zona costera Pacífico y Atlántica Colombiana y en ellos se destaca su baja productividad económica.

El *grupo 4*, quedo conformado por los departamentos de Caquetá, Sucre, Córdoba, Nariño. La ganadería y la minería son sectores en común de la economía que comparten. Este grupo se destaca por su baja condición de pobreza extrema y baja condición de desigualdad.

En el *grupo 5*, encontramos los departamentos de Cesar, Huila, Magdalena, Norte de Santander, Boyacá, Bolívar y Tolima. Estos departamentos tienen en común su principal actividad económica, que es el sector agropecuario y minero. Presentan bajas condiciones de pobreza extrema y desigualdad.



**Figura 6-22.:** Dendrograma sobre las coordenadas parciales de los individuos del AFM

La Figura 6-22, presenta los resultados de aplicar el algoritmo de conglomerados por el método de *Ward*, sobre las coordenadas de los individuos parciales que se obtienen en el AFM. Como resultado se obtiene lo siguiente:

En el *grupo 1*, determinado por los departamentos que describen mayores condiciones de Pobreza Extrema como lo son Chocó y la Guajira. Se destaca el departamento de la Guajira por ingresar en el año 2020 a compartir altas condiciones de desigualdad similares a las del Chocó en los años 2016 a 2020.

En el *grupo 2*, constituido con los departamentos que describen condiciones de pobreza menos extremas, se destaca la transición que ha hecho el departamento del Cauca en el año 2020 pues dejó de conformar el grupo de mayor condición Pobreza Monetaria Extrema y paso a compartir similitudes en su pobreza comparables con los departamentos de Huila, Magdalena, Cauca y Boyacá.

En el *grupo 3*, se consolidan los departamentos Nariño, Sucre, Córdoba y Caquetá, que presentan condiciones similares de pobreza en los años 2020 con departamentos como Cundinamarca y Atlántico.

En el *grupo 4*, se muestra como Bogotá D.C. siempre ha mantenido durante los años 2012 a 2020, condiciones similares de pobreza y desigualdad que no se comparan con las de ningún departamento.

En el *grupo 5*, determinado por los departamentos que describen menos condiciones de Pobreza Monetaria y desigualdad, muestra que durante los años 2012 a 2020 los departamentos han mantenido en condiciones similares de Pobreza Monetaria y no se observa que ningún departamento ingrese o salga del grupo.

## 7. Conclusiones y recomendaciones

### 7.1. Conclusiones

1. En general se evidencia una leve reducción de la pobreza monetaria en Colombia a nivel departamental desde el año 2012 hasta el año 2018, que puede estar asociada a las estrategias promovidas desde el gobierno central para mitigar la pobreza como: *Familias en Acción*, *Colombia Mayor*, *Familias En Su Tierra* y desde organismos externos como El Programa de las Naciones Unidas para el Desarrollo con su política social *Pacto por la equidad*. Sin embargo, se observa la alta persistencia de la desigualdad y las crecientes condiciones desfavorables de pobreza extrema que se registraron en los departamentos de Chocó, Cauca y la Guajira y que se agudizan desde el año 2018 hasta el año 2020.
2. Los resultados mostraron que los departamentos situados en la región Pacífica y del Caribe, describen mayores condiciones de pobreza extrema y desigualdad en el ingreso, por su escasa actividad económica y bajo producto interno bruto (PIB). Por el contrario, la mayoría de departamentos que describen condiciones más bajas de pobreza y desigualdad se encuentran en la región Andina.
3. Se presentaron cambios en la estructura de correlación entre todas las variables con respecto al año 2020. El AFM mostró que las variables incidencia extrema, brecha extrema y severidad de la pobreza, se relacionan durante los años 2012 a 2018, y que estas se asocian más con la primera componente principal. Unas pocas variables como coeficiente de Gini y el ingreso con la segunda componente principal conformada. Este cambio guarda relación con respecto a la emergencia sanitaria por la pandemia Covid-19 que atravesó el país, como consecuencia inmediata de la desaceleración en el modelo económico.

### 7.2. Recomendaciones

- Este estudio presenta la oportunidad de evaluar cuál ha sido el impacto de las políticas implementadas por el gobierno a nivel departamental para mitigar la pobreza y como se refleja sobre los indicadores de pobreza monetaria.
- Desde el año 2018 el Departamento Nacional de Estadística empezó la recolección de datos de Pobreza Multidimensional por departamentos de manera anual. Se recomienda



---

utilizar estos nuevos datos para evaluar desde la metodología directa, como lo es la dinámica de la Pobreza Multidimensional a nivel departamental en Colombia.

**A. Anexo: Nombrar el anexo A de acuerdo con su contenido**

# Referencias

- Abascal, y Fernandez, C. (2002). Análisis factorial múltiple como técnica de estudio de los resultados de un análisis de componentes principales. *Madrid..*
- Abascal, E., y Landaluce, M. (2002). Análisis factorial multiple como técnica de estudio de la estabilidad de los resultados de un análisis en componentes principales. *Questiío*, 26, 109-122.
- Acosta, O. L., y Zúñiga, F. A. (2017). El ingreso y su distribución en colombia. *Metodologías de la CEPAL*.
- Ariza, J. F., y Retajac, A. (2020, 5). Descomposición y determinantes de la pobreza monetaria urbana en colombia. un estudio a nivel de ciudades. *Estudios Gerenciales*, 36, 167-176. Descargado de <https://www.icesi.edu.co/revistas/index.php/> doi: 10.18046/j.estger.2020.155.3345
- Bossio, M. L. C. (2010). Combinación del análisis factorial múltiple y el análisis armónico cualitativo en la descripción de datos longitudinales cualitativos. *Tesis de maestria. Universidad Nacional de Colombia*.
- Cattell, R. B. (1966). The Scree Plot Test for the Number of Factors. *Multivariate Behavioral Research*.
- CEPAL. (2018). Medición de la pobreza por ingresos: actualización metodológica y resultados. *Metodologías de la CEPAL*, N.2. Descargado de <http://hdl.handle.net/11362/44314>
- DANE. (2015). Proyecciones de población municipal y en resguardos indígenas. Descargado de <https://www.dane.gov.co/>
- DANE. (2018). Pobreza monetaria por departamentos en Colombia. *Boletín Técnico de Pobreza Monetaria Departamental*, 29(10), 211- 233.
- DANE. (2021). Pobreza monetaria por departamentos en Colombia. *Anexo pobreza monetaria departamental*.
- Escofier, B., y Pagés, J. (1982). Comparaison de groupes de variables définies sur le même ensembles définies sur le même ensemble d individus, Technical report. *Institut National de Recherche en Informatique, París.*, 69–97.

- Escofier, B., y Pagés, J. (1992). Análisis factoriales simples y múltiples. *Objetivos, métodos e interpretación.*, Universidad del País Vasco, Bilbao.
- Foster, J., y Greer, J. (1984). A class of decomposable povertymeasures. , 52, N.º 3.
- Fresneda, O. (2007). La medida de necesidades básicas insatisfechas (NBI) como instrumento de medición de la pobreza y focalización de programas. *Serie Estudios y Perspectivas (Bogotá)*, 18. Descargado de <http://hdl.handle.net/11362/4816>
- Gil, V., y Puerta, I. (2019). Aplicación del algoritmo de agrupamiento de k-medias para la pobreza monetaria en colombia.
- González, V. M., y Olaya, J. (2009). Análisis de los espacios de representacion del statis y del afm en el estudio de movilidad biografica en bogota 1993. *Revista Colombiana de Estadística.*, 32, 1-15.
- Johan, B., y Marcel, v. (2016). An Empirical Kaiser Criterion. *Psychological Methods*. Descargado de <http://dx.doi.org/10.1037/met0000074>
- Kahn, J. H. (2006). Factor analysis in Counseling Psychology research, training and practice. *The Counseling Psychologist*, 34, 1-36.
- Laís, A., y Pérez, R. (2018). Panorama Social de América Latina. *Cepal*, 3P.
- Lebart, L., y Piron, M. (1994). Statisquique exploratoire multimensionnelle. *Dunod*, 56, París.
- Lebart, L., y Salem, A. (2000). Análisis estadístico de datos y textos. *Madrid: Milenio.*
- Lê, S., y Pagès, J. (2009). Análisis de datos con R. *Escuela Colombiana de ingeniería.*, 32, 1-15.
- Mendoza, G. (2009). Comparación de los métodos tucker 3 y análisis factorial múltiple para el análisis de datos tres vías.
- Naciones, U. (2015). Fin de la pobreza: porque es importante. , 1. Descargado de <https://www.un.org/sustainabledevelopment/es/poverty/>
- Núñez, J., y Ramírez, J. J. C. (2002). Determinantes de la pobreza en Colombia. Años recientes. *Serie Estudios y Perspectivas (Bogotá)*, 1. Descargado de <http://hdl.handle.net/11362/4789>
- ONU. (2018). La ineficiencia de la desigualdad. *Metodologías de la CEPAL*, N.4.
- Ostry, J., y Loungani, P. (2016). Neoliberalism: oversold? *Finance and Development*, 53(2). Descargado de <http://hdl.handle.net/11362/44314>

- Pardo, C. E. (2011). Métodos en ejes principales para tablas de contingencia con estructuras de partición en filas y columnas. *Departamento de Estadística. Universidad Nacional de Colombia..*
- Ravallion, M. (1992). Growth and Redistribution Components in Poverty measures: A decomposition with applications to Brazil and India. *Journal of Development Economics.*, 38, 275-295.
- Ruiz, J., y Bonilla, J. (2018). Análisis de datos en tres vías para la evaluación de la dinámica económica de los departamentos en colombia durante el periodo 2000-2015. Descargado de <https://www.researchgate.net/>
- Stiglitz, J. (2016). Inequality and economic growth Rethinking Capitalism. , 1.