



Análisis Factorial Múltiple para describir las condiciones de salud sentidas de la población priorizada de la ciudad de Cali en el año 2018

Crhistian Camilo Garcia Altamirano

Universidad del Valle
Facultad de Ingeniería
Escuela de Estadística
Santiago de Cali
2021

Análisis Factorial Múltiple para describir las condiciones de salud sentidas de la población priorizada de la ciudad de Cali en el año 2018

Crhistian Camilo Garcia Altamirano

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

Magister en Estadística

Director:

Ph. D. Jaime Mosquera Restrepo

Universidad del Valle
Facultad de Ingeniería
Escuela de Estadística
Santiago de Cali, Colombia
2021

Dedicatoria

A mis padres Oscar y Mirma,
A mis hermanos Erick Mauricio y Lina Marcela,
A mi abuela Aura.
A la memoria de mi abuela *Esther*

Agradecimientos

Agradezco a la Secretaria de Salud Pública Municipal de Santiago de Cali por facilitar la base de datos para la realización de este ejercicio.

A los profesores de la Escuela de Estadística, por compartir su conocimiento y contribuir a mi crecimiento profesional.

Al profesor Jaime Mosquera por sus múltiples aportes que contribuyeron a perfeccionar esta investigación.

A mi familia, por motivarme a continuar a pesar de las adversidades, por su amor, paciencia y comprensión.

A mis compañeros Andrés Ochoa, Alejandro Delgado, Jefferson Peña, Cristian Garcia, Ernesto Peláez, Rogelio Nuñez y Ronald Galeano por su incondicionalidad y los momentos compartidos.

Y a ti lector, por dedicar unos minutos de tu tiempo para conocer acerca de esta investigación.

Resumen

En este trabajo se analizan, desde una perspectiva multivariada, los datos recogidos en el proyecto: *implementación de la estrategia de atención primaria en salud en el municipio de Santiago de Cali*. Estos datos fueron colectados a través de un formulario tipo encuesta que indaga sobre las condiciones de la vivienda, de los individuos y de su entorno, la cual está dividida en 7 módulos que generan las tablas de datos: vivienda, generalidades, personas menores de 10 años, salud sexual y reproductiva-cáncer, estilos de vida saludable, salud mental y micobacterias.

Las unidades de análisis corresponden a 18 barrios de la ciudad que fueron priorizados por la estrategia *Territorios de Inclusión y Oportunidad* (TIO) de la alcaldía, para estos barrios se calcularon algunos indicadores resumen de la condición de salud y vida de sus habitantes, tales como: porcentaje de hacinamiento, porcentaje de animales domésticos sin vacunas, porcentaje de niños con esquema de vacunación incompleto, porcentaje de gestantes inasistentes a control prenatal, porcentaje de inasistencia a programas de Promoción y Prevención (hipertensión, diabetes, crecimiento y desarrollo, entre otros). Cada subconjunto de indicadores fue consolidado en diferentes tablas multivariadas, la cuales se analizan de forma conjunta a través de la metodología del Análisis Factorial Múltiple (AFM).

El AFM es una técnica estadística de análisis de datos multivariados adaptado al tratamiento de tablas de datos en las que un mismo conjunto de individuos se describe a través de múltiples grupos de variables, con la condición de que las variables de un mismo grupo deben ser de la misma naturaleza: cualitativa o cuantitativa.

Como resultado se caracterizan los barrios de la ciudad, identificando en ellos los principales problemas de salud sentida que los afectan. Complementariamente, el AFM permitió resumir la información colectada en un índice sintético, a través del cual se genera un ranking u ordenamiento de los barrios con mayor afectación en salud en la ciudad. Este índice será de utilidad para la administración municipal, en el sentido de que su aplicación permite priorizar las intervenciones sociales y de salud pública sobre el conjunto de barrios más vulnerables.

Palabras clave: Salud sentida, Atención primaria en salud, Análisis de múltiples tablas, Caracterización multivariante, Índice sintético

Abstract

In this work, from a multivariate perspective the data collected has been analyzed in the project: *implementation of the primary health care strategy in the municipality of Santiago de Cali*. These data were collected through survey forms that inquiries about housing conditions, individuals and their environment, this survey was divided into 7 modules that creates the data tables: housing, generalities, people under 10 years of age, sexual and reproductive health-cancer, healthy lifestyles, mental health, and mycobacteria.

The analysis units belong to 18 neighborhoods of the city, they were prioritized by the strategy *Territories of Inclusion and Opportunity* (TIO) created by the Mayor's office, for these neighborhoods we calculated some summary indicators of the health and life condition of its inhabitants, such as: overcrowding percentage, domestic animals without vaccines, children with vaccination schedule incomplete, pregnant women that are not attending prenatal care, percentage of non-attendance at programs Promotion and Prevention (hypertension, diabetes, growth and development, among others). Every subset of indicators was consolidated in different, multivariate tables, which are jointly analyzed through the Multiple Factor Analysis (MFA) methodology.

The MFA is a statistical technique of multivariate data analysis adapted to the treatment of data tables in which the same set of individuals is described through multiple groups of variables, with the condition that the variables of the same group must be of the same nature: qualitative or quantitative

As a result, the city's neighborhoods are characterized, identifying in them the main felt health problems that affect them. Complementarily, the MFA allowed to summarize the information collected in a synthetic index, through which it is possible to generate a ranking or ordering of the sectors with high health affectations in the city. This index can be useful for the municipal administration, in the sense that its application allows prioritize social and public health interventions over the set of neighborhoods most vulnerable

Keywords: Felt health, Primary health care, Analysis of multiple tables, Multivariate characterization, Synthetic index

Contenido

Lista de figuras	x
Lista de tablas	xi
1. Introducción	1
2. Planteamiento del problema	4
2.1. Definición del problema	4
2.2. Objetivos	7
2.2.1. Objetivo General	7
2.2.2. Objetivos Específicos	7
2.3. Justificación	8
3. Antecedentes	9
3.1. Antecedentes conceptuales	9
3.2. Antecedentes estadísticos	13
4. Marco teórico	17
4.1. Análisis de Componentes Principales - ACP	17
4.2. Análisis Factorial Múltiple	25
4.2.1. La representación de los grupos de variables	27
4.2.2. La representación de las variables	28
4.2.3. La representación de los individuos	30
4.2.4. La representación superpuesta de las J nubes	31
4.2.5. Ayudas a la interpretación en el AFM	32
4.2.6. Elementos suplementarios	33
4.3. Creación de índices basados en métodos factoriales	34
4.3.1. Factor de tamaño	34
4.3.2. Transformación del índice en puntajes no negativos	35
5. Metodología	37
5.1. Definición de la población: criterios de inclusión	38

5.2. Estructura de los datos	38
5.3. Procesamiento de datos	40
5.3.1. Software	44
6. Resultados	45
6.1. Análisis Exploratorio de Datos	45
6.2. Análisis Factorial Múltiple	50
6.2.1. Análisis de Factores	50
6.2.2. Ponderación de los grupos de variables	51
6.2.3. Tipología de los grupos	52
6.2.4. Representación de las variables	54
6.2.5. Representación de los individuos	57
6.2.6. Representación superpuesta de grupos e individuos	59
6.2.7. Proyección de las componentes principales de cada grupo	61
6.2.8. Formulación de un índice sintético	61
7. Conclusiones y recomendaciones	65
7.1. Conclusiones	65
7.2. Recomendaciones	66
Bibliografía	67
A. Anexo A: Formulario de Encuesta	72

Lista de Figuras

4-1.	Representación gráfica del centramiento y escalado de variables	18
4-2.	Ejemplo de un histograma de valores propios	20
4-3.	Representación gráfica de de los datos en a) el espacio original de las variables, b) el espacio de las componentes principales	21
4-4.	Ejemplo de una representación simultánea	24
4-5.	Representación general de la estructura de la tabla de datos en AFM	25
4-6.	Ejemplo de representación de los grupos de variables en el AFM	28
4-7.	Ejemplo de representación de las variables sobre los ejes factoriales del AFM	29
4-8.	Ejemplo de representación de los individuos en el AFM	30
4-9.	Ejemplo de representación superpuesta de la nube de individuos en AFM . .	31
4-10.	Ejemplo de factor <i>tamaño</i>	34
5-1.	Secuencia metodológica del análisis de datos	37
6-1.	Distribución de las variables por grupos	46
6-2.	Gráfico de la matriz de correlación de las variables consolidadas	49
6-3.	Gráfico de sedimentación y valores propios del AFM	50
6-4.	Representación de los grupos en el análisis global	52
6-5.	Circulo de correlación de las variables en el análisis global	54
6-6.	Contribución de las variables a la inercia de los ejes factoriales	55
6-7.	Representación de los territorios en el análisis global	58
6-8.	Contribución de los territorios a las componentes	58
6-9.	Representación superpuesta de algunos territorios	60
6-10.	Contribución de los grupos a las dimensiones	61
6-11.	Índice APS expresado en porcentaje	64

Lista de Tablas

4-1. Ejemplo de identificación de variables que más aportan a la construcción de los ejes factoriales	29
5-1. Resumen de la obertura de Viviendas- Estrategia APS Cali 2018.	39
5-2. Estructura de los grupos de variables de la hoja de datos crudos.	41
5-3. Listado de indicadores resumen de barrio por grupo de variables y etiquetas	42
5-4. Estructura de la tabla de datos final - nivel barrios (grupos activos)	43
5-5. Resumen de la estructura de datos para el AFM	43
6-1. Estadísticas descriptivas de las variables	47
6-2. Valores atípicos en cada grupo de variables	48
6-3. Valores propios y ponderaciones de los ACP separados por grupo	51
6-4. Descomposición de la inercia de las dos primeras dimensiones	51
6-5. Correlaciones para los dos primeros ejes del AFM	52
6-6. Coeficientes Lg	53
6-7. Coeficientes RV	53
6-8. Vector propio, coordenadas, contribuciones y calidad de la representación de las variables	57
6-9. Elementos para el cálculo del índice sintético	62
6-10. Índice sintético	63

Capítulo 1

Introducción

De acuerdo con [Hermida \(2016\)](#), para 1970, la medicina hospitalaria fue criticada porque no contribuía a la solución eficaz de los problemas de salud, sino que parecía más preocupada por aspectos económicos. En consecuencia, [Lalonde \(1974\)](#) propuso un enfoque de atención integral en salud que evite que la población enferme y requiera de servicios hospitalarios. Dicho enfoque está basado en el conjunto de factores, tanto personales como sociales, económicos y ambientales, que determinan el estado de salud de la población, conocidos como determinantes sociales. Además, usa estadísticas de mortalidad para conocer el punto donde las enfermedades son inevitables. También se apoya en la historia para demostrar el cambio de enfermedades infecciosas a enfermedades crónicas, que posteriormente conllevan a la muerte. A partir de lo anterior, la Organización Mundial de la Salud (OMS), la Organización Panamericana de la Salud (OPS) y el Fondo de las Naciones Unidas para la Infancia (UNICEF) organizaron la Conferencia Mundial de Alma Ata, realizada en 1978.

La Conferencia alertó a los gobiernos, profesionales sanitarios y demás personal de la salud pública, a generar acciones para promover y proteger la salud de la población mundial ([OMS, 1978](#)). En Colombia la Atención Primaria en Salud (APS), fue puesta en práctica con el apoyo de la OMS y la OPS durante los años 80 del siglo XX, mediante las Unidades Primarias de Atención (UPA), con una débil articulación intersectorial, participación comunitaria y algunos programas sectoriales de intervención selectivos ([Angulo, 2012](#)).

Años después, a través de la ley 1438 de 2011, el Ministerio de Salud y Protección Social (MSPS) en Colombia formaliza la APS como una estrategia intersectorial para reducir las brechas de inequidad, mejorar el acceso a los servicios de salud desde la salud pública, el diagnóstico, la promoción de la salud, la prevención de la enfermedad y la rehabilitación del paciente en todos los niveles de complejidad. ([MSPS, 2011](#))

Desde el año 2013, en la ciudad de Cali, la estrategia de APS es implementada por la Secretaría de Salud Pública Municipal (SSPM), la cual se ejecuta mediante visitas puerta a

puerta, donde auxiliares de enfermería indagan sobre las condiciones del entorno, de la vivienda y de las condiciones de salud de los habitantes. En este punto, vale la pena anotar que los auxiliares de enfermería no validan clínicamente el estado de salud de las personas y que los datos que se registran son subjetivos, ya que representan la opinión del encuestado sobre su estado de salud. Éste es justamente el concepto de *salud sentida*, que se refiere a la *opinión subjetiva de un individuo sobre su estado de salud*.

Debido a limitaciones económicas, las visitas no se realizan sobre toda la ciudad. Los sectores en los que se realizan las visitas surgen como resultado de un análisis estratégico anual de prioridad, en el cual se tienen en cuenta tres criterios de inclusión: i. Ser Territorio de Inclusión y Oportunidad (TIO) para la administración municipal, ii. Presentar la mayor cantidad de eventos de interés en salud pública, iii. Presentar la mayor cantidad de casos de mortalidad de acuerdo a lo reportado por las estadísticas vitales del Departamento Administrativo Nacional de Estadística (DANE).

Fruto de este análisis, para el año 2018 la administración municipal priorizó un subconjunto de 18 barrios sobre los cuales se implementó la encuesta denominada *Ficha de Diagnóstico y Caracterización de Familias*. La información de la encuesta se utiliza para elaborar un documento sintético de Análisis de la Situación Integral de Salud (ASIS) del Ministerio de Salud y Protección Social, el cual contiene principalmente estadísticas básicas de morbilidad y mortalidad, y algunos análisis de tendencias mediante el análisis de regresión ([MSPS, 2014](#)).

Los datos que conforman la encuesta se subdividen en 7 módulos: vivienda, generalidades, personas menores de 10 años, salud sexual y reproductiva-cáncer, estilos de vida saludable, salud mental y micobacterias. La información colectada en cada uno de estos módulos, puede ser organizada en una tabla de datos, compuesta a su vez por múltiples indicadores o variables. Hasta el momento, esta estructura de datos ha sido analizada a través de estadísticas básicas univariadas y bivariadas, en las cuales se omite la posible interrelación entre múltiples variables y entre múltiples tablas.

Partiendo de lo anterior, el presente trabajo se propone realizar un análisis integral de la información disponible en las 7 tablas de datos, aprovechando la estructura multivariante de los datos. Como técnica de análisis se sugiere el Análisis Factorial Múltiple (AFM), el cual describe la relación de covariación entre J grupos (tablas) de variables, en términos de “*pocas variables aleatorias no observables, llamadas factores*”. ([Díaz y Rivera, 2012](#))

De acuerdo con [Abascal y Calvo \(2002\)](#): “*La estructuración en grupos de variables enriquece el estudio, pues los objetivos no se limitan a obtener una tipología de los individuos determinada por el conjunto de variables, sino que se amplían a la búsqueda de relaciones entre las estructuras en de cada uno de los grupos*”.

Se pretende entonces realizar una aplicación del AFM, a través del cual se analiza el conjunto de individuos (Barrios) mediante las variables establecidas por la estrategia de APS de la ciudad de Cali, para identificar la salubridad en aquellos barrios con alta priorización según la Secretaria de Salud Pública de Santiago de Cali, de tal forma que permita mejorar la planeación de la entidad para mejorar la calidad de vida de la población. Complementariamente, basado en la descomposición obtenida en el AFM, se construye un índice multivariado que refleja el riesgo según la estrategia de APS de los barrios intervenidos.

Para una mejor comprensión, el contenido de este documento ha sido dividido en siete capítulos: esta introducción conforma el Capítulo 1. La definición del problema y sus objetivos se presentan en el Capítulo 2. El Capítulo 3 resume los hallazgos más importantes encontrados en la revisión de antecedentes. El Capítulo 4 contiene el marco teórico. La metodología de análisis de datos se presenta en el Capítulo 5, mientras que los resultados obtenidos se resumen en el Capítulo 6. Finalmente, el Capítulo 7 presenta las conclusiones del presente estudio y las recomendaciones.

Capítulo 2

Planteamiento del problema

Éste capítulo contiene la definición del problema y se establecen los objetivos del presente proyecto. Adicionalmente, se presenta una justificación, desde el punto de vista de la salud de la región y una justificación técnica estadística del desarrollo del proyecto.

2.1. Definición del problema

De acuerdo con la OMS, la Atención Primaria en Salud es *“la asistencia sanitaria esencial accesible a todos los individuos y familias de la comunidad a través de medios aceptables para ellos, con su plena participación y a un costo asequible para la comunidad y el país. Es el núcleo del sistema de salud del país y forma parte integral del desarrollo socioeconómico general de la comunidad”* (Kroegeer y Luna, 1987). Motivados por lo anterior, los estados formulan políticas que integran la APS en los sistemas de salud, en el desarrollo social y económico global de la sociedad. También articulan diferentes sectores con la participación de la sociedad en la planificación, la organización y la prestación de los servicios de salud.

Por ejemplo, en América del Norte, se brindó acceso a médicos generales para aquellas personas que no contaban con seguro médico o no tenían acceso a la prestación de servicios en centros de salud comunitarios. En la ciudad de Nápoles - Italia se realizó la evaluación de la red de los centros de salud comunitaria, en torno a temas como pobreza, enfermedades mentales, riesgos laborales y exclusión social. En centros de APS en Madrid se involucró la participación comunitaria y en Bélgica se diseñaron programas de atención primaria orientada a la comunidad. (Labonté et al., 2009)

En el caso de América Latina, la OPS (2007) mostró que la mayoría de los países de Latinoamérica no ha logrado avances significativos para alcanzar la visión de la APS expuesta en la Declaración de Alma Ata, adoptada hace ya más de 30 años. Además, se han publicado estudios que realizan la caracterización de la salud de la población en países como Costa Rica

y Brasil (Segredo, 2009), en los cuales se indica que se han logrado reducir indicadores como la mortalidad infantil, mortalidad en adultos y reducción de las inequidades en el acceso a los servicios de salud.

En Colombia, Angulo (2012) hace un recorrido histórico sobre la implementación de la APS en el país, mostrando como posible punto de partida la década de los 80's mediante las Unidades Primarias de Atención (UPA), en las cuales realizaban actividades en algunos programas sectoriales de intervención, tales como: asistencia médica, suministro de medicamentos formulados, educación nutricional, vacunación y vigilancia epidemiológica, promoción de la construcción de sistemas de abastecimiento de agua e instalación de filtros caseros, educación para la disposición de basuras, control de zoonosis y mejoramiento de la vivienda, los cuales reflejaron poca articulación intersectorial y participación comunitaria.

Angulo (2012) también hace mención de la puesta en marcha de la Ley 100 de 1990, que descentralizó el Sistema Nacional de Salud (SNS), concediéndole la oportunidad a los alcaldes de integrar los servicios y las acciones intersectoriales para conectar el sector salud con los planes locales de desarrollo. Además introduce cambios de organización y de operación buscando universalizar los servicios de salud con el programa *Salud Básica para Todos* y fomenta el aseguramiento privado y voluntario, fortaleciendo la participación comunitaria. Posteriormente, la Ley 100 de 1993 genera un retroceso en los adelantos alcanzados en la APS integral, pues se la redujo a simple atención clínica y a un paquete básico de acciones en Salud Pública. (Vega et al., 2009)

A nivel local, es posible encontrar resultados de la implementación de la estrategia de atención primaria en salud en ciudades capitales como Bogotá (Angulo, 2012), Barranquilla (Tuesca-Molina et al., 2015), Manizales¹, Cali (SSPM, 2017) y la zona rural de 6 municipios de Santander (Rodríguez et al., 2013).

En Cali, la estrategia de APS se ejecuta desde el año 2013 mediante la implementación de una encuesta que indaga sobre las condiciones de la vivienda y de salud de la población de barrios priorizados por la SSPM. A pesar de que la estructura de la encuesta es multivariada y que esta se encuentra subdividida en siete grupos de variables, el análisis que actualmente se realiza de estos datos se limitan a resúmenes univariantes de indicadores y tablas bivariadas, además de representaciones gráficas, los cuales a pesar de ser pertinentes no brindan un panorama global sobre la salud sentida en los barrios y desconocen la posible existencia de una estructura de correlación entre las variables que conforman la evaluación de la estrategia de APS. Frente a este tipo de estructuras, cabe la posibilidad de desarrollar un análisis más profundo desde una perspectiva multivariante que haga uso de las estructuras de correlación para descubrir otros patrones de comportamiento en el conjunto de datos y

¹disponible en <https://cutt.ly/7j4Laqt>

otras características de los barrios que han sido priorizados.

El análisis que se propone en este trabajo se realizará sobre los datos recolectados por la estrategia de atención primaria en salud en Santiago de Cali, durante el año 2018. Se toma como referencia este año, debido a que a partir de ese año la encuesta tuvo una estabilización; en aplicaciones previas, la encuesta ha tenido múltiples modificaciones.

Para la aplicación de la encuesta se visitaron 18 territorios, entre barrios y corregimientos de la ciudad. La encuesta recoge información acerca de indicadores de interés en salud pública, tales como: hacinamiento en las viviendas, uso de cilindros de gas, inadecuado almacenamiento de alimentos, tenencia responsable de mascotas, criaderos de zancudos en las viviendas, inasistencia a programas de Promoción y Prevención (crecimiento y desarrollo, Programa Ampliado de Inmunización-PAI, agudeza auditiva y visual, planificación familiar, hipertensión, diabetes, toma de citología, mamografía, examen de próstata, entre otros), síntomas depresivos, intento de suicidio, maltrato verbal o físico y violencia sexual por parte de la pareja de la persona, dividido en 7 módulos o secciones.

Luego de consolidar la información a nivel de barrios, la estructura de los datos queda finalmente conformada por 7 tablas de datos, cada una con 18 filas (barrios), la cual coincide con la configuración requerida por el AFM, el cual permite describir la relación de covariación entre J grupos (tablas) de variables, en términos de “pocas variables aleatorias no observables, llamadas factores”. ([Díaz y Rivera, 2012](#)). Adicionalmente, tiene la ventaja de proporcionar un espacio común (es decir, los factores), así como información sobre cómo se relaciona cada barrio a este espacio común (es decir, los factores parciales). ([Abdi et al., 2013](#)).

En el contexto de los datos disponibles, la implementación del AFM contribuirá a los tomadores de decisiones en el sentido de que podrán tener una visión más general del comportamiento de los barrios respecto a cada una de las dimensiones que miden las subtablas. Por ejemplo, podrán identificar si un barrio particular se destaca por tener buenas condiciones de salud en una de las dimensiones evaluadas pero malas condiciones de salud en otras. Estos aspectos ayudaran a redirigir la política pública de forma focalizada centrándose en esos elementos particulares más relevantes.

2.2. Objetivos

2.2.1. Objetivo General

Caracterizar las condiciones de salud sentida por la población priorizada en la ciudad de Cali durante 2018 a través de un Análisis Factorial Múltiple.

2.2.2. Objetivos Específicos

- Analizar desde una perspectiva multivariante, los datos disponibles sobre la estrategia de atención primaria en salud de Santiago de Cali en el año 2018.
- Caracterizar las condiciones de salud sentida de los barrios priorizados en Santiago de Cali.
- Construir un índice de necesidades de salud para priorizar los barrios con mayores problemas de salud pública.

2.3. Justificación

Desde su creación en 1978, la APS se pensó como una estrategia para reducir las brechas de inequidad en la atención en salud de la población y de esta forma contribuir al mejoramiento de su calidad de vida, pues retoma la definición conceptual de salud, incorporada en la Constitución de la OMS: *“el estado de completo bienestar físico, mental y social, y no sólo la ausencia de afecciones o enfermedades”* (Copplesstone, 1991). Por ello se plantearon un conjunto de indicadores mínimos que debían lograr todos los países. Sin embargo, la OMS ha aceptado que aún está lejos de cumplir esos mínimos en muchas partes del mundo (Angulo, 2012).

De acuerdo con SSPM (2017), en 1979 se consideró que la propuesta realizada en la conferencia de Alma Ata era costosa y poco realista; entonces se asumieron formas de APS focalizadas, de acciones específicas, menos costosas y con un fin de supervivencia infantil. Para ello se priorizaron los programas de crecimiento y desarrollo, lactancia materna, inmunización y técnicas de rehidratación oral. Esta estrategia se conoció como Atención Primaria de Salud Selectiva (APS-S), la cual dejó sin bases la participación de las comunidades en la planificación, decisión y control.

Hasta hoy, la Secretaría de Salud Pública Municipal de Cali presenta un informe resumen, en el cual se presentan los resultados de las encuestas aplicadas y el porcentaje de población que logró acceder a los servicios de salud, resaltando que se hace énfasis en la educación a la población sobre los temas que aborda la encuesta.

Desde la perspectiva estadística, este proyecto pretende realizar un aporte metodológico para el análisis de un conjunto de datos que hasta el momento se estaba realizando con estadísticas univariadas o bivariadas, con el riesgo de omitir estructuras de correlación que podrían ser aprovechables para un mejor análisis de las condiciones de salud de la población. La estructura multivariante de los datos y su natural división en subtablas, permite aplicar la técnica estadística Análisis Factorial Múltiple (AFM), bajo la cual se logra una caracterización de los barrios priorizados, dejando en evidencia las similitudes desde la salud pública que hay entre ellos. Complementario a ello, también se pretende realizar un ranking que clasifique los barrios de acuerdo a su nivel de salubridad, mediante la generación de un índice sintético. Éste índice puede ser empleado por la administración municipal para orientar acciones y definir políticas públicas que conlleven al mejoramiento de la calidad de vida de los barrios que se encuentren en situaciones más desfavorables.

Capítulo 3

Antecedentes

En éste capítulo se presentan los hallazgos más relevantes identificados en una revisión bibliográfica reciente sobre el tema de análisis multivariado, especialmente Análisis Factorial Múltiple, y sobre los diferentes métodos de análisis que se han empleado para estudiar y medir la salud sentida la población a través de la estrategia de APS. En consecuencia, el capítulo se divide en dos secciones: revisión de antecedentes relacionados con APS y revisión de antecedentes de las aplicaciones del análisis multivariado; al interior de cada subsección los trabajos se presentan en orden cronológico, desde la publicación más antigua a la más reciente.

3.1. Antecedentes conceptuales

- [Kroeger y Luna \(1987\)](#): proponen los aspectos a considerar para implementar la atención primaria en salud. Éstos son: objetivos, estrategias, metodologías básicas, aspectos programáticos, perfil del trabajador básico de la salud, investigación a nivel local, entre otros. Un capítulo de este libro trata sobre la investigación a nivel local, se refiere a algunas técnicas de recolección de información, entre las cuales mencionan: encuesta informal, encuesta formal, historia del caso, grupo focal y la “Autopsia verbal” (estudio de un fallecimiento). En cuanto a las técnicas para procesar los datos, menciona la distribución de frecuencias de las variables exploradas y el cruce de variables.

A pesar de los aportes importantes que hacen en éste libro, no brinda lineamientos de análisis de datos, y cuando los hace no sugiere algún procesamiento multivariado.

- [Cueto \(2004\)](#): Presenta un estudio histórico del rol desempeñado por la OMS y UNICEF en el surgimiento y difusión del concepto de atención primaria de salud a fines de la década de 1970 y principios de la de 1980. El autor analiza el contexto político de estas organizaciones, sus líderes, las metodologías y tecnologías asociadas a la perspectiva de la atención primaria en salud y los debates sobre su significado. Estos

debates llevaron al desarrollo de un enfoque alternativo y más restringido, conocido como atención primaria en salud selectiva.

Este documento configura un referente histórico de la evolución de la estrategia de APS. Sin embargo, no da lineamientos sobre el análisis de datos o como hacer un ejercicio basado en datos colectados por la estrategia de APS.

- [Hutchison et al. \(2011\)](#): Realizan un análisis de las políticas implementadas en Canadá con relación a la atención primaria en salud durante las décadas de 1980 y 1990, que conllevó a un nuevo entorno político en la década del 2000. Se examina la reforma de la atención primaria en salud, basándose en información descriptiva de literatura publicada y gris (también denominada como literatura no convencional, literatura semipublicada o literatura invisible) y en entrevistas semiestructuradas a personas informados de la atención primaria de salud en Canadá. Se encontró que la atención primaria en salud de Canadá ha entrado en un período de cambio potencialmente transformador. Las iniciativas clave incluyen el apoyo a equipos multidisciplinarios de atención primaria en salud, prácticas y redes grupales, inscripción de pacientes con un proveedor de atención primaria, incentivos financieros y esquemas de pago combinado, desarrollo de mecanismos de gobernanza de la atención primaria en salud, expansión de proveedores de atención primaria en salud, implementación de la historia clínica electrónica, formación y apoyo para la mejora de la calidad.

Este documento configura un referente político de la estrategia de APS, en el cual no se presentan análisis de datos y las decisiones se toman basados en fuentes secundarias o de acuerdo a criterios de expertos.

- [Rasella et al. \(2014\)](#): Evalúan el impacto de la implementación reciente del Programa de Salud de la Familia (PSF) en Brasil, el programa de atención primaria de salud más grande del mundo, sobre la mortalidad por enfermedades cardíacas y cerebrovasculares entre 2000 y 2009. Éste trabajo se soporta en un estudio de datos longitudinales sobre 1622 municipios, sobre los cuales ajustan un modelo de regresión binomial negativo con estructura de efectos fijos. La cobertura anual de PSF y la cobertura media de PSF en años anteriores se utilizaron como principales variables independientes y se clasificaron como: ninguna (0 %), incipiente (< 30 %), intermedia (30 – 69 %), o consolidada (70 %). Las tasas de mortalidad estandarizadas por edad y por causas en el grupo de enfermedades cerebrovasculares, isquémicas y otras formas de enfermedades cardíacas, son variables dependientes calculadas para todos los municipios cada año. Ellas representan el 40 % de todas las muertes de estos grupos durante el período de estudio. La cobertura de PSF se asoció negativamente con las tasas de mortalidad de enfermedades cerebrovasculares y cardíacas en modelos ajustados y no ajustados para factores de

confusión sociales y económicos. La cobertura del PSF no tuvo ningún efecto sobre la tasa de mortalidad por accidentes, utilizada como control. Se concluye que programas como el PSF en Brasil, que actúan a través de la prevención, la atención y el seguimiento de las enfermedades cardiovasculares pueden contribuir a reducir la morbilidad y la mortalidad por enfermedades cardiovasculares en un país en desarrollo como Brasil.

De este trabajo se destaca la intención de los autores de abordar el análisis de los datos desde una perspectiva multivariada apoyada con modelos de regresión. Este tipo de análisis no son habituales en el procesamiento de datos por las entidades públicas encargadas de dirigir estudios similares.

- [Alfageeh et al. \(2017\)](#): examinan los factores que influyen en el acceso y la utilización de los centros de atención primaria de salud en las zonas urbanas y rurales de la provincia de Riyadh, del Reino de Arabia Saudita, a través de la aplicación de una encuesta diseñada para identificar las barreras y facilitadores de acceso a la APS en 5 áreas rurales y 5 áreas urbanas de la provincia de Riyadh, seleccionadas de acuerdo a la densidad de población. La encuesta incluyó 50 preguntas, divididas en 11 secciones, que evalúan una amplia gama de factores relacionados con el acceso y la experiencia de los encuestados en los servicios de atención primaria en salud (Características socio-demográficas, estado de salud, uso de servicios, factores organizacionales, factores financieros, medicamentos, pruebas, referencias, atención dental, comunicación doctor-paciente, promoción y prevención de la salud y satisfacción). Los encuestados fueron mayores de 18 años y residentes saudíes. Se obtuvieron un total de 935 respuestas con 52,9 % ($n = 495$) de áreas urbanas y el 47,1 % restante ($n = 440$) de áreas rurales de la provincia de Riyadh. Para el análisis de los datos se realizaron pruebas *chi-cuadrado* entre pares de variables dentro de cada una de las secciones de la encuesta. Se destaca que hay altos niveles de satisfacción entre los pacientes entre todos los servicios.

En relación a las diferencias entre los encuestados urbanos y rurales, los hallazgos indicaron que hubo variaciones significativas en relación a: nivel educativo, ingresos mensuales, investigaciones médicas, realización de análisis de sangre a tiempo, horas extra de apertura, distancia, limpieza y prevención de la salud. Estos hallazgos tienen implicaciones para la política y planificación de los Centros de Atención Primaria en Salud, reducción las desigualdades en la atención de la salud entre la población rural y urbana, además de contribuir a la reducción de la carga de morbilidad crónica en la provincia de Riyadh.

En este documento es posible identificar que los datos tienen una estructura multivariante. Sin embargo, los análisis solo consideran cruces de pares de variables donde evalúan la independencia de cada variable en el área rural y en el área urbana.

- [Jeffreys et al. \(2020\)](#): Investigan los cambios en los determinantes socio-demográficos en el acceso gratuito y de bajo costo a la atención primaria en salud utilizando datos de la Encuesta de salud de Aotearoa, Nueva Zelanda (1996/97–2016/17). Se utilizó regresión logística para estimar los Odds Ratios (OR) de la asociación entre los factores sociodemográficos y el acceso a la APS de costo cero y la APS de bajo costo por separado, y la asociación entre el acceso a la APS para maoríes y no maoríes, por diferentes niveles de medidas socioeconómicas. Los análisis del acceso a la APS sin tarifas se restringieron a las personas que habían consultado a un médico de cabecera en los 12 meses anteriores y se compararon las que informaron haber pagado \$0 con las que informaron haber pagado más de \$0.

La investigación sobre el acceso a la APS de bajo costo se limitó a aquellas personas que habían consultado a un médico de cabecera en los 12 meses anteriores y se comparó a las que informaron haber pagado entre \$1 y \$20 con las que informaron haber pagado más de \$20, es decir, no incluyó a las personas que pagaron \$0. Concluyen que la estrategia de atención primaria en salud, disminuye los copagos en las personas que acceden a los servicios de atención primaria en salud, donde los más bajos fueron para las personas de pueblos del Pacífico y aquellos que viven en las áreas más desfavorecidas. Aunque los maoríes tenían más probabilidades que los no maoríes de acceder a cuidados de bajo costo en todos los niveles de privación, esto era menos evidente en las zonas más desfavorecidas que en las zonas más ricas. Sugieren que las políticas futuras deberían trabajar para hacer que la APS sea más asequible.

En este documento se destaca el uso de la regresión logística para evaluar la asociación entre los factores sociodemográficos, el costo de acceso a los servicios de salud y el lugar de residencia de las personas. Los autores reconocen que sus análisis son limitados, por la elección de preguntas y la cantidad de datos faltantes en las encuestas, dejando la sensación de que quizá el método de análisis presentado no sea el más adecuado.

Otros trabajos como los realizados por [Ahmed et al. \(2017\)](#) y [Tuesca-Molina et al. \(2015\)](#), presentan aplicaciones de la atención primaria en salud desde la perspectiva de la medicina cardiovascular y la enfermería, respectivamente, en los cuales hacen uso de estadísticas descriptivas básicas. Además, en Colombia, se encuentran los resultados de la implementación de la estrategia de atención primaria en salud en Bogotá ([Angulo, 2012](#)), Manizales¹, Cali ([SSPM, 2017](#)) y la zona rural en 6 municipios de Santander ([Rodríguez et al., 2013](#)) a través de estadísticas descriptivas básicas y evaluación de tendencia mediante regresión.

¹disponible en <https://cutt.ly/7j4Laqt>

3.2. Antecedentes estadísticos

- [Castaño \(2004\)](#): Recapitula los elementos estadísticos propuestos por [Castaño et al. \(1998\)](#) para la construcción del Índice de Calidad de Vida (ICV) de los hogares en la ciudad de Medellín. Éste índice es un resumen de las particularidades de la vivienda y de las personas que viven en el hogar: aspectos demográficos, calidad de la vivienda, servicios en la vivienda, seguridad social y capital humano.

A diferencia del Índice de Necesidades Básicas Insatisfechas (NBI), que capta principalmente condiciones de desarrollo en la infraestructura urbana o la línea de pobreza que no considera ingresos no monetarios, como el trueque, o el ingreso real, relacionado con el acceso a subsidios del gobierno; el índice propuesto busca mejorar la comprensión y la medición de las dimensiones del bienestar, combinando variables de acceso físico: características de la vivienda, acceso a servicios públicos y variables de capital humano, seleccionadas de una encuesta aplicada por Planeación Metropolitana de Medellín, en el año 2001.

Para la construcción del ICV, se cuantificaron las variables cualitativas relacionadas con las condiciones de vida mediante escalamiento óptimo ([W. F. Kuhfeld et al., 1985](#); [Saporta, 1983](#); [Van de Geer, 1993](#); [Young, 1981](#); [Young et al., 1978, 1985](#)), el cual asignó un peso positivo a cada una de las categorías de las variables. Una vez cuantificadas, se empleó el Análisis de Componentes Principales (ACP) para obtener el índice mediante la primera componente principal.

- [Carrasco y Arredondo \(2006\)](#): hacen un diagnóstico de la situación de América Latina y el Caribe Hispano en relación a las políticas públicas en los años 1990 y 2002. Estos autores realizan un AFM con variables de los informes del Programa de las Naciones Unidas para el Desarrollo (PNUD), agrupándolas en 5 grupos de variables: Salud, Desarrollo, Educación, esfuerzo público y un grupo ilustrativo (que contiene las variables: tasa de desempleo urbano, líneas telefónicas por cada 1000 habitantes. y crecimiento del consumo privado por habitante a precios corrientes del año 1995).

En el primer plano factorial, se observa que Haití está más alejado del resto de países y que sus condiciones empeoran con el pasar de los años, similarmente ocurre con México, Panamá, Venezuela y Brasil, mientras que en Bolivia, El Salvador y Chile ocurre lo contrario, pues se observa un desplazamiento hacia la derecha para 2002 en el primer factor, explicado por los grupos Salud, Educación y Desarrollo. Argentina y Perú muestran un pequeño avance con el pasar de los años. Entre sus conclusiones, destacan que el AFM permite al analista y al actor político la toma de decisiones con menor incertidumbre ya que aporta información adicional en comparación al Análisis

de Componentes Principales (ACP) y que las trayectorias de los individuos (países) permiten estudiar la evolución de los resultados obtenidos una vez aplicadas determinadas políticas públicas.

- [Avella et al. \(2010\)](#): Presenta los desarrollos teóricos del Analisis Factorial Múltiple (AFM) ([Escofier y Pagès, 1992](#); [Escoufier, 1973](#)) y el Análisis de componentes principales para datos cualitativos (PRINQUAL) ([W. Kuhfeld, 1990](#); [Michailidis y de Leeuw, 1998](#); [Tenenhaus, 1977](#)), el cual hace necesaria la cuantificación de las variables cualitativas mediante iteraciones de mínimos cuadrados alternantes (ALS, por sus siglas en ingles). Posteriormente realizan un ACP sobre la matriz de datos que tiene las variables óptimamente cuantificadas a través de mínimos cuadrados alternantes y cuantificación óptima (ALSOS, Alternating Least Squares y Optimal Scaling) ([Young, 1981](#)).

Construye un índice sintético basado en los resultados de ambas metodologías, el cual se conforma por la primera componente principal en cada caso, reescalada a valores de 0 a 100 mediante transformaciones lineales o puntajes no negativos ([DNP, 2000](#)) para facilitar su interpretación. También comparan el proceso de construcción de los índices sintéticos por ambos métodos, a través del índice de calidad de vida obtenido sobre la base de datos de la encuesta de hogares de la ciudad de Bogotá, realizada durante el 2003, la cual está descrita por variables ordinales agrupadas en 4 grupos: Acceso y calidad de los servicios, capital humano presente y potencial, tamaño y composición del hogar, calidad de la vivienda. De forma complementaria, también proponen un índice de competitividad para las Pymes, siguiendo la misma metodología del índice de calidad de vida.

Entre sus conclusiones destacan que los métodos difieren principalmente en el tratamiento de las variables ordinales del PRINQUAL y la estructuración de las variables en grupos temáticos del AFM. En términos generales los autores indican que no es posible definir cual método es mejor, ya que las diferencias dependen del contexto en el que se apliquen (definición y clasificación de las variables nominales u ordinales).

- [DNP \(2015\)](#): El Departamento de Planeación Nacional de Colombia (DNP) construye un indicador para medir el desarrollo territorial de municipios y departamentos. Éste indicador resume el comportamiento de variables de tipo social: % población alfabetada, % de personas sin NBI 2005 resto, % de viviendas con servicios de energía, % de personas sin NBI 2005 cabecera, % de viviendas con acueducto, % asistencia escolar, Viviendas por hogar con personas presentes, % de población en cabecera y % de viviendas con alcantarillado, y variables financieras: Ingresos tributarios per cápita (\$ corrientes), Inversión pública municipal per cápita, % de no dependencia de las transferencias. En éste trabajo se usa el Análisis de Componentes Principales para la

construcción del índice, mientras que la conformación de las tipologías fue realizada aplicando el Análisis de Conglomerados. Finalmente, la primera componente principal es reescalada utilizando el método propuesto en [DNP \(2000\)](#) para facilitar su interpretación.

- [Ivaldi et al. \(2016\)](#): Realizan un estudio sobre la medición del bienestar en 27 países de la Unión Europea, evaluando 12 dimensiones: salud, educación y entrenamiento, equilibrio entre el trabajo y la vida, bienestar económico, relaciones sociales, política e instituciones, seguridad, bienestar subjetivo, paisaje y patrimonio cultural, ambiente, investigación e innovación y calidad de los servicios.

Los autores construyen un índice multidimensional de bienestar mediante el análisis de factores, cuyo objetivo es detectar un número estadísticamente suficiente de variables (se utilizaron 146 variables) para representar la mayor parte de la varianza explicada.

Las variables seleccionadas se establecieron según 3 criterios: varianza explicada, gráfico de valores propios y variables cuyo valor propio es mayor a 1. Una vez seleccionadas las variables de acuerdo a los criterios mencionados, utilizan como índice el puntaje factorial resultante del análisis factorial en los indicadores parciales, los cuales estandarizan para eliminar el efecto de la unidad de medida.

El ordenamiento de los países se realizó en función de sus parámetros de acuerdo a los lineamientos de [Carstairs y Morris \(1990\)](#), pues permite mantener las características discriminatorias de la distribución ([Carstairs, 2001](#)). Los valores de corte para definir las clases (subgrupos de países con similar condición) fueron $\pm\frac{2}{3}\sigma$, $\pm\frac{4}{3}\sigma$, 0 y el valor promedio de la distribución de las puntuaciones de los factores.

- [Ramosa y Morinsonb \(2017\)](#): estudia la calidad de vida de los niños de la costa Caribe colombiana en 2010. Para ello emplea indicadores demográficos, socioeconómicos, de salud, educativos y de nutrición mediante el Análisis Factorial Múltiple (AFM). Posteriormente clasifica los departamentos a través de la clasificación jerárquica aglomerativa a través del método de Ward.

Las variables fueron extraídas de fuentes secundarias: Departamento Administrativo Nacional de Estadística; Asociación Probienestar de la Familia Colombiana; Ministerio de Educación y Departamento Nacional de Planeación y se consolidaron en una sola base de datos. En el análisis global, los factores: salud, nutrición y demografía representaron el primer eje principal, mientras que el segundo estuvo representado por los factores: educación y socioeconómico.

Se evidenció que los niños del departamento de La Guajira tienen las condiciones más desfavorables de calidad de vida, los departamentos de Bolívar, Córdoba, Sucre, Magdalena y Cesar conformaron el grupo medio en cuanto a calidad de vida, seguido del departamento del Atlántico, que tuvo los niveles mas altos de calidad de vida infantil.

Cómo síntesis de la presente revisión bibliográfica, se concluye que pese a que la estrategia de APS se planteó hace más de 40 años, aún no se tiene claridad sobre su forma de operar y tampoco sobre la manera de realizar el análisis estadístico de la información que colecta, pues los análisis estadísticos realizados difieren. Algunos autores realizan análisis de tendencias vía regresión y otros únicamente presentan estadísticas descriptivas univariadas y bivariadas. Con relación al análisis de datos para describir situaciones asociadas a políticas públicas, se hace uso de técnicas estadísticas de análisis multivariado de datos, y a partir de ellas surge la construcción de índices sintéticos que permiten clasificar a los individuos de acuerdo a las variables analizadas.

En esta revisión no se encontraron antecedentes relacionados con el análisis estadístico multivariante de datos provenientes de la evaluación del las condiciones de salud sentida de la población. Esto destaca la necesidad de emplear análisis como los propuestos en este trabajo de grado para el estudio de la salud sentida de la población.

Capítulo 4

Marco teórico

En éste capítulo se presentan los fundamentos teóricos que sustentan lo métodos de análisis multivariante que serán empleados en el análisis de los datos de este proyecto. Inicialmente se presenta el Análisis de Componentes Principales (ACP), que aunque no es el método empleado en este trabajo, es la base fundamental de los métodos multivariados y sus conceptos pueden ser extendidos a otras técnicas de análisis de datos multivariados. Posteriormente se presenta de forma detallada la técnica Análisis Factorial Múltiple (AFM), que será empleada en el análisis de los datos. Una descripción más detallada de estos métodos puede ser consultada en [Lebart et al. \(2000\)](#) y [Escofier y Pagès \(1992\)](#) respectivamente.

4.1. Análisis de Componentes Principales - ACP

Un conjunto de datos generalmente contiene características observadas mediante a variables (descriptores). Debido a la dificultad de representar esos datos en el espacio original, es conveniente encontrar un espacio con menos dimensiones en el que sea posible representar las observaciones, perdiendo un mínimo de información. Idealmente, un espacio apropiado tendría solo una o dos dimensiones ([Gonze, 2007](#)).

De acuerdo a [Lebart et al. \(2000\)](#) el Análisis de Componentes Principales - ACP se considera como una técnica de representación (descripción) de los datos que se utiliza sin referencia a hipótesis ni a un modelo estadístico en particular.

El ACP se aplica a una matriz de datos donde las columnas representan las mediciones de las p variables *cuantitativas* y las filas representan a los individuos u objetos de observación. El objetivo del ACP es reducir un número de p variables en un número mas pequeño p' de nuevas variables llamadas *factores*(o *componentes principales*, los cuales son combinación lineal de las variables originales que dan máxima información sin ser redundantes). En otras palabras, el ACP descubre cuales variables conforman un subconjunto coherente (correlacionado) que

esta relativamente no correlacionado con otro mediante una proyección adecuada de los datos (Gonze, 2007).

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

En ésta matriz X , cada elemento x_{ij} representa la “medida” (*puntaje*) de la variable j para el individuo i . Debido a que las variables pueden expresarse en diferentes unidades de medida y que los efectos de escala afectan sustancialmente el aporte de cada variable a la varianza total, los datos de la matriz X suelen ser estandarizados por columna de la siguiente manera:

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

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

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$$

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

Esta transformación asegura que la media y la varianza de los datos estandarizados z_{ij} son iguales a 0 y 1 respectivamente, con lo cual se da una nivelación en la escala de los datos. Los datos transformados se denominan *puntuación z* y denotaremos $\mathbf{Z}$ la matriz de puntuaciones z .

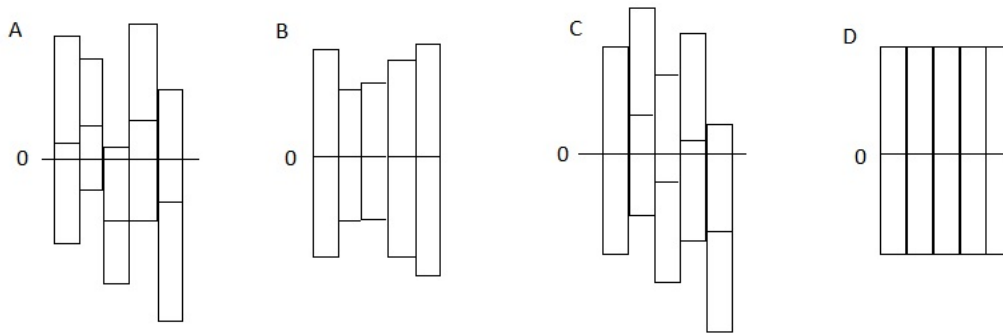


Figura 4-1: Representación gráfica del centramiento y escalado de variables

La Figura 4-1 representa el procesamiento de datos previo a la ejecución del ACP, los datos para cada variable están representados por una barra con diferente amplitud y centro. La figura A muestra los datos originales con diferentes amplitudes y diferentes centros, la figura B es el resultado después de solo realizar el centramiento de las variables (restar la media

de la variable), la figura C es el resultado de escalar los datos únicamente (dividir por la desviación estándar) y la figura D es el resultado de centrar y escalar los datos.

Realizar un ACP sin considerar la estandarización de los datos puede tener el inconveniente de que aquellas variables que tengan mayor variabilidad pueden tener un mayor peso en la conformación de los factores. Para evitar ese fenómeno la literatura sugiere estandarizar las variables.

En el ACP, las proximidades entre las variables se miden en términos de la correlación y la distancia entre individuos se mide en términos de *similitudes* (*semejanzas*). Lo esencial del ACP consiste en la búsqueda de un nuevo espacio, preferiblemente de menor dimensión al del espacio de las variables, en el cual las distancias entre las observaciones proyectadas difieran lo mínimo posible de las distancias entre las observaciones en el espacio original.

Análisis de la proximidad entre los individuos

En el espacio de la representación original de los individuos $\mathbf{R}^p$, la distancia euclidiana entre parejas de individuos se cuantifica como $d^2(i, i') = \sum_{j=1}^p m_j (x_{ij} - x_{i'j})^2$, donde los coeficientes m_j ponderan la influencia de cada columna j , con ellos se conforma la matriz diagonal M , conocida como métrica.

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') ($\max_u = \sum_i^n d^2(i, i')$), que resulta equivalente a maximizar sobre u la suma de cuadrados de las distancias entre cada punto y el centro de gravedad: ($\max_u = \sum_i^n d^2(i, G)$).

El Análisis de Componentes Principales normado se desarrolla diagonalizando la matriz de correlación $Z'NZ$, donde $N = \text{diag}(1/n)$ es la matriz que contiene el peso de los individuos.

Como resultado de la diagonalización se obtiene la matriz UDU' , donde U contiene los vectores propios u_α y D es una matriz diagonal que contiene los valores propios λ_i de $Z'NZ$. La suma de los valores propios es igual a la varianza total (de hecho, la suma de los valores propios de una matriz dada es igual a la suma de los elementos diagonales de esta matriz (Gonze, 2007)). 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 :

$$\sum_{i=1}^p \lambda_i = p$$

El valor propio más grande λ_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 observar esta

variación. Los valores propios restantes, ordenados de forma decreciente ($\lambda_1 > \lambda_2 > \lambda_3 > \dots$) dan las mayores variaciones en direcciones ortogonales a los anteriores.

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.

Selección del número de componentes principales

No hay criterios estrictamente formales para la determinación de componentes principales a retener (Díaz y Rivera, 2012). En un análisis normado, la suma de las inercias es igual al número de variables y por lo tanto la inercia media vale 1. Cada componente es una combinación particular de las variables de origen, interesarán en general aquellos ejes cuya inercia sea “notablemente” superior a 1 (Lebart et al., 2000). De acuerdo a Kaiser (1958) es irrazonable agregar una componente que explique menos varianza que una variable.

Cattell (1965) propuso el *contraste de caída*. Este criterio consiste en graficar los valores propios de forma decreciente (gráfico de sedimentación), seleccionando aquellas componentes en que la curva converge a una línea horizontal, punto en el cual la varianza explicada por las componentes no aporta mucho más.

Porcentaje de la varianza explicada. Consiste en tomar los valores propios que acumulen un porcentaje mínimo aceptable de varianza explicada (por ejemplo entre 70 y 80 %) (Soto y Schuschny, 2009).

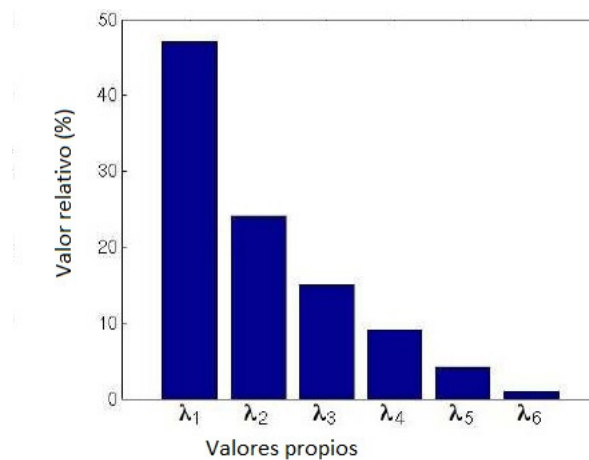


Figura 4-2: Ejemplo de un histograma de valores propios

El anterior histograma (Figura 4-2) muestra los valores propios, ordenados según su valor relativo (en %). En este ejemplo ficticio, podemos observar que tomando los dos primeros

valores (es decir, los dos primeros componentes) es suficiente para describir aproximadamente el 70 % de los datos.

Coordenadas de los individuos

Las coordenadas de los n individuos proyectados sobre el eje “canónico” u_α son los componentes del vector (factor) $\psi_\alpha = Zu_\alpha$ el cual es una combinación lineal de las variables iniciales.

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

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

...

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

Donde cada elemento u_{ij} es la ponderación (peso) que tiene la variable Z_j sobre la componente i .

Debido a que la nube de individuos está centrada sobre el centro de gravedad, la media del factor es nula, esto es $E(\psi_\alpha) = 0$ y la varianza es $V(\psi_\alpha) = \psi'_\alpha N \psi_\alpha = \lambda_\alpha$, lo cual implica que cada componente explica una proporción de la variabilidad total.

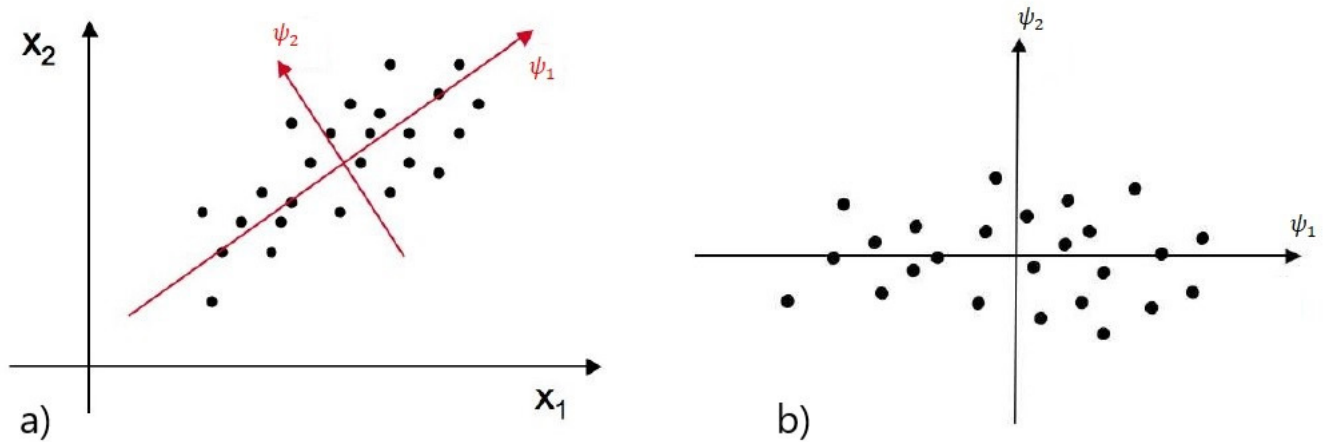


Figura 4-3: Representación gráfica de de los datos en a) el espacio original de las variables, b) el espacio de las componentes principales

Geoméricamente, el ACP puede verse como una rotación de las coordenadas de nuestros datos, los factores ψ_i definen las nuevas coordenadas. (Figura 4-3)

Análisis de la proximidad entre las variables

Estar interesado en las variables es considerar la tabla como una yuxtaposición de columnas. Cada variable está asociada con una secuencia de n números. Según este punto de vista, una variable se puede representar como un vector del espacio vectorial denominado $\mathbf{R}^n$.

De acuerdo a [Lebart et al. \(2000\)](#) La distancia entre variables se calcula de la siguiente forma:

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

desarrollando el binomio se obtiene:

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

Teniendo en cuenta que las variables están estandarizadas y la definición previa de 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 entonces que la distancia entre dos variables esta dada por:

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

Que implica:

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

Geométricamente, en el espacio $\mathbf{R}^n$, el coseno del ángulo de dos vectores variable es el coeficiente de correlación entre esas dos variables ($c_{jj'} = \cos(j, j')$). Si estas dos variables están a una distancia 1 del origen (si están estandarizadas), el coseno es su producto escalar.

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 sea la relación lineal que las liga directa o inversa.

Dos variables ortogonales ($c_{jj'} = 0$) están a una distancia media.

Las proximidades entre puntos variable se interpretan entonces en términos de correlaciones.

Coordenadas de las variables

Las *relaciones de transición* permiten relacionar el espacio de los individuos con el de las variables, a través de ellas se obtiene que las coordenadas de las p variables están dadas por $\varphi_\alpha = \sqrt{\lambda_\alpha} u_\alpha$.

Representación simultánea

Puesto que la nube de individuos y la nube de variables se definen en espacios diferentes, es imposible situarlas en un mismo espacio. Sin embargo, si se considera la dirección de las variables en $\mathbf{R}^p$ es posible realizar la representación simultánea en éste espacio.

Previo a la transformación de la tabla de datos, se contaba con dos sistemas de ejes ([Lebart et al., 2000](#)):

- Los antiguos ejes unitarios $(e_1, e_2, \dots, e_j, \dots, e_p)$ correspondientes a las p variables al comienzo del análisis donde: $e_j = (0, 0, \dots, 1, 0, \dots, 0)$ $j = 1, \dots, p$ es el sistema de ejes de referencia para las coordenadas iniciales de los individuos.
- los nuevos ejes unitarios u_α ($\alpha = 1, \dots, p$), que constituyen los ejes factoriales.

Como se observa en la Figura 4-4 la representación simultánea es la proyección (en fila suplementaria) del eje unitario e_j sobre el nuevo eje α . La coordenada de la proyección de e_j sobre u_α es $e'_j u_\alpha = u_{\alpha j}$.

Es posible entonces comparar las posiciones relativas de dos individuos respecto al conjunto de variables, o dos variables con respecto al conjunto de los individuos.

Según [Lebart et al. \(2000\)](#) la dirección de una variable define zonas para los individuos: de un lado, aquellos que toman valores fuertes para esta variable y al lado opuesto, aquellos que toman valores bajos.

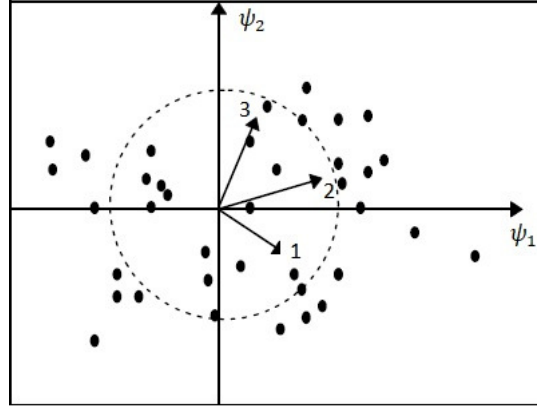


Figura 4-4: Ejemplo de una representación simultánea

En la representación simultanea se interpreta la posición relativa de dos individuos respecto conjunto de variables o dos variables respecto al conjunto de individuos. Lo que se llamará *representación simultánea* es “el aplastamiento” de la referencia ortonormada de los ejes de origen sobre el plano factorial de la nube de individuos. Se dispone así de una perspectiva deformada del sistema de ejes original teniendo en cuenta las relaciones existentes entre las variables iniciales (Lebart et al., 2000).

Ayudas a la interpretación

Los componentes principales permiten obtener la mejor visualización aproximada (en el sentido de los mínimos cuadrados) de las distancias entre los individuos de una parte y entre las variables de otra parte. Para interpretar los componentes es necesario apreciar correctamente esta aproximación.

Las variables fuertemente correlacionadas con un componente principal van a contribuir a la definición de ese eje. En éste sentido, interesarán entonces aquellas variables que presentan las coordenadas más fuertes (son aquellas que se sitúan más cerca del círculo de correlación) y se interpretarán los componentes principales en función del reagrupamiento de ciertas variables y la oposición de otras (Lebart et al., 2000). El calculo de la contribución de las variables está dado por:

$$CTR_{\alpha}(j) = \frac{\varphi_{j\alpha}^2}{\lambda_{\alpha}} = u_{j\alpha}^2$$

Para determinar que individuos aportan más a la construcción de un componente principal se calcula su contribución de la siguiente forma:

$$CTR_{\alpha}(i) = \frac{m_i \psi_{i\alpha}^2}{\lambda_{\alpha}}; \text{ donde } m_i = \frac{1}{n}$$

Los individuos que contribuyen más a la determinación del componente principal son los más excéntricos (están mas alejados de los demás) y la lectura de la gráfica es suficiente para

identificarlos en este caso.

La calidad de la representación presenta una idea acerca de que tan cercano se encuentra un punto a un componente principal. Su cálculo se realiza mediante los cosenos cuadrados entre cada punto y el componente principal de la siguiente forma:

Individuos:

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

Variables:

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

4.2. Análisis Factorial Múltiple

El Análisis Factorial Múltiple (AFM), desarrollado por [Escofier y Pagès \(1992\)](#), es un método factorial adaptado al tratamiento de tablas de datos en las que un mismo conjunto de individuos se describe a través de múltiples grupos de variables. Los grupos suelen surgir cuando se analizan simultáneamente variables de diferente naturaleza: cuantitativas, cualitativas, de un arreglo de datos en tres dimensiones o del manejo de un mismo conjunto de variables medidas en distintas condiciones o modalidades. ([Abascal y Calvo, 2002](#))

Un ejemplo de un arreglo de datos en múltiples subgrupos se presenta en la siguiente imagen:

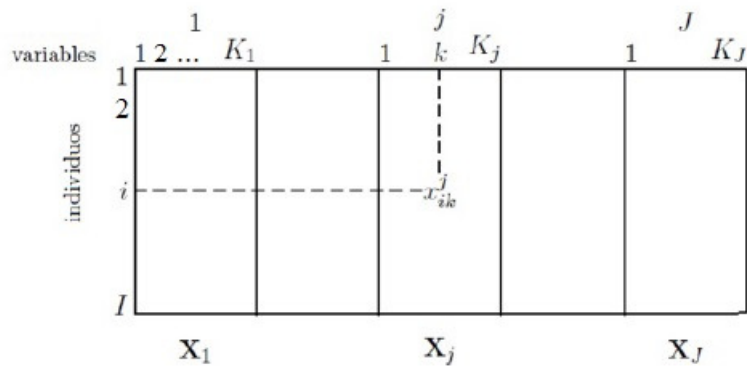


Figura 4-5: Representación general de la estructura de la tabla de datos en AFM

En la Figura 4-5 las matrices $X_1, \dots, X_j, \dots, X_J$ conforman subgrupos de variables, cada uno conformado por distinto número y tipo de variables.

El AFM puede considerarse 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 todas, o a parte, de las nubes parciales.

El AFM se fundamenta en la metodología empleada en el ACP y se ejecuta en 2 etapas:

1. Cada grupo de variables es analizada por separado y se obtienen los factores parciales. Es decir, realiza un ACP (o ACM, dependiendo del tipo de variables que conforman el grupo) a cada una de las tablas $\mathbf{X}_j$, la cual contiene los valores de las variables del grupo j , $j = 1, 2, \dots, J$, en el conjunto de individuos.
2. Realiza un ACP de la tabla general, la cual resulta de yuxtaponer las J tablas $\mathbf{X}_j$. En este análisis cada tabla $\mathbf{X}_j$ es ponderada por el inverso del primer valor propio obtenido en el ACP (o ACM) de la propia tabla, $1/\lambda_j^1$, estos valores conforman la matriz diagonal $\mathbf{M}$. Esta ponderación mantiene la estructura de cada tabla, ya que todas las variables de un mismo grupo han recibido el mismo 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 Calvo, 2002).

El AFM, considera tres espacios de análisis de datos:

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

A cada grupo de variables le corresponde una tabla, de manera que todas las tablas pueden ser yuxtapuestas y formar así una única tabla que cruza individuos y variables (Escofier y Pagès, 1992). Siguiendo la notación de la Figura 4-5, denotaremos:

- X : Tabla completa.
- I : Número de individuos.
- K : Número de variables (incluye todos los grupos)
- J : Número de conjuntos o grupos de variables.
- K_j : Número de variables del j -ésimo grupo de variables.
- X_j la tabla asociada al grupo j .

Ponderación de los grupos de variables

Inicialmente se debe obtener el peso de cada variable del grupo j , por lo cual se divide por λ_j^1 (denotando por λ_j^1 el primer valor propio del ACP o ACM¹ separado del grupo j). Esta ponderación se realiza con el objetivo de equilibrar el papel de los grupos en todos los aspectos del análisis, ya que ninguno determina por si solo la primera componente principal del análisis. (Escofier y Pagès, 1992).

Posteriormente, se realiza un ACP de la tabla general (estandarizada y ponderada), se obtienen los valores y vectores propios que dan lugar al análisis general en el que la inercia de cada factor está conformada proporcionalmente por las variables activas de cada grupo. Estas inercias se utilizan como coordenadas para graficar una “interestructura” que pone de manifiesto la tipo de los grupos (González et al., 2009).

4.2.1. La representación de los grupos de variables

Uno de los objetivos del estudio de varios grupos de variables es comparar globalmente los grupos. Se introduce el espacio R^{I^2} base de esta representación que puede aparecer como una ayuda a la interpretación del ACP de la tabla completa X , pero que posee su propia optimización.

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 se interpreta como una medida de relación entre los grupos. A partir de éste producto escalar, se construye el coeficiente de correlación vectorial RV . Éste coeficiente se obtiene a partir de los coeficientes de correlación lineal entre dos variables cualesquiera. Su valor está comprendido entre 0 (no existe relación entre las variables) y 1 (las nubes que representan a los grupos son homotéticas²) (Abascal y Calvo, 2002)

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

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

¹según la naturaleza de las variables

²Transformación del plano o del espacio, en la que los puntos correspondientes se alinean con un punto fijo y la razón entre sus distancias a este punto es constante

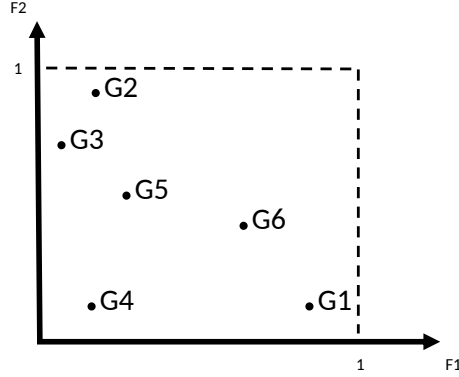


Figura 4-6: Ejemplo de representación de los grupos de variables en el AFM

La Figura 4-6 representa un ejemplo ficticio sobre la representación gráfica de los grupos de variables, muestra que el primer factor se debe principalmente a los grupos 1 y 6, mientras que el segundo factor está más relacionado con los grupos 2 y 3. El grupo 4 no está relacionado con ninguno de los dos ejes factoriales.

De manera sintética, los valores altos de las coordenadas en los grupos indican que el factor está asociado a las variables de los grupos. Por el contrario, valores bajos de las coordenadas indican poca relación del factor con las variables de los grupos.

Factores comunes

De acuerdo a [Escofier y Pagès \(1992\)](#) si la correlación entre grupos (todos o algunos de ellos) 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.

4.2.2. La representación de las variables

Esta representación se obtiene directamente del ACP de la tabla completa X (Previamente ponderada). Para ello, se utilizan las coordenadas de las proyecciones sobre los ejes factoriales. Dichas coordenadas también pueden calcularse, utilizando las relaciones de transición del ACP.

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

La representación de las variables puede considerarse a la vez:

- Como una ayuda a la interpretación de la representación de la nube de individuos.
- Como una representación óptima de las correlaciones entre variables.

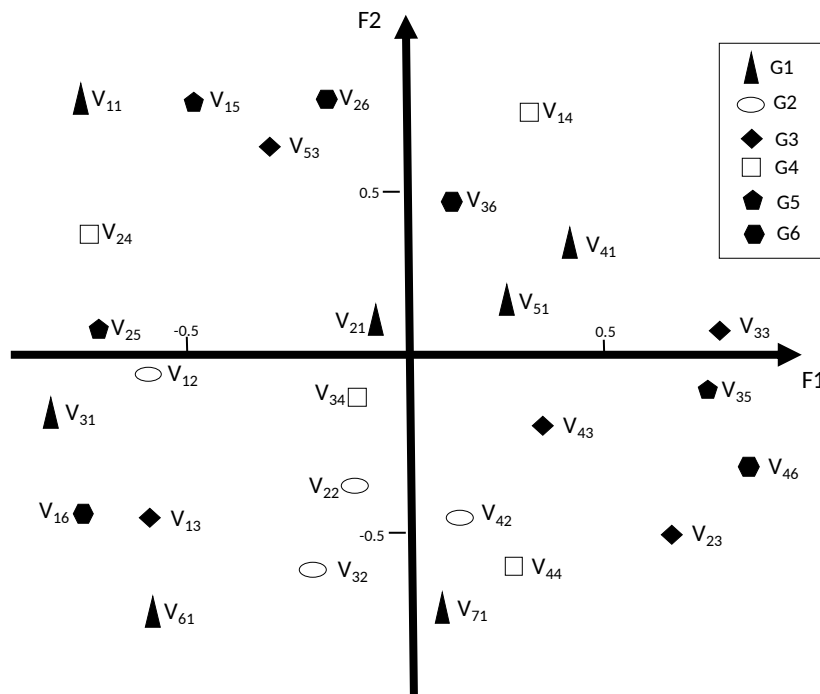


Figura 4-7: Ejemplo de representación de las variables sobre los ejes factoriales del AFM

La Figura 4-7 es una representación de la proyección de las variables en el primer plano factorial. Basados en esta representación, en la Tabla 4-1 se identifican las variables de cada grupo que más se relacionan con los factores.

Factores	G1	G2	G3	G4	G5	G6
Factor 1	V_{31}, V_{11}	V_{12}	V_{33}	V_{24}	V_{25}, V_{35}	V_{46}, V_{16}
Factor 2	V_{71}	V_{42}, V_{22}, V_{32}	V_{53}	V_{14}	V_{15}	V_{26}

Tabla 4-1: Ejemplo de identificación de variables que más aportan a la construcción de los ejes factoriales

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 $1/\lambda_j^1$) 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, Surge de nuevo, la idea según la cual:

- Ningún grupo puede, por sí sólo, determinar el primer eje (salvo una situación de simetría excepcional),
- Un grupo influye sobre tantos más ejes cuanto mayor sea su dimensión.

4.2.3. La representación de los individuos

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 . Este se deduce de la componente principal F , por la relación: $u_s = (1/\sqrt{\lambda_s})X'DF_s$ en la que λ_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:

$$F_s^j = \tilde{X}_j M u_s = (1/\sqrt{\lambda_s}) \tilde{X}_j M X' D F_s = 1/\sqrt{\lambda_s}) W_j D F_s$$

Las J nubes de individuos, llamadas “nubes parciales”, se proyectan como ilustrativas sobre los ejes.

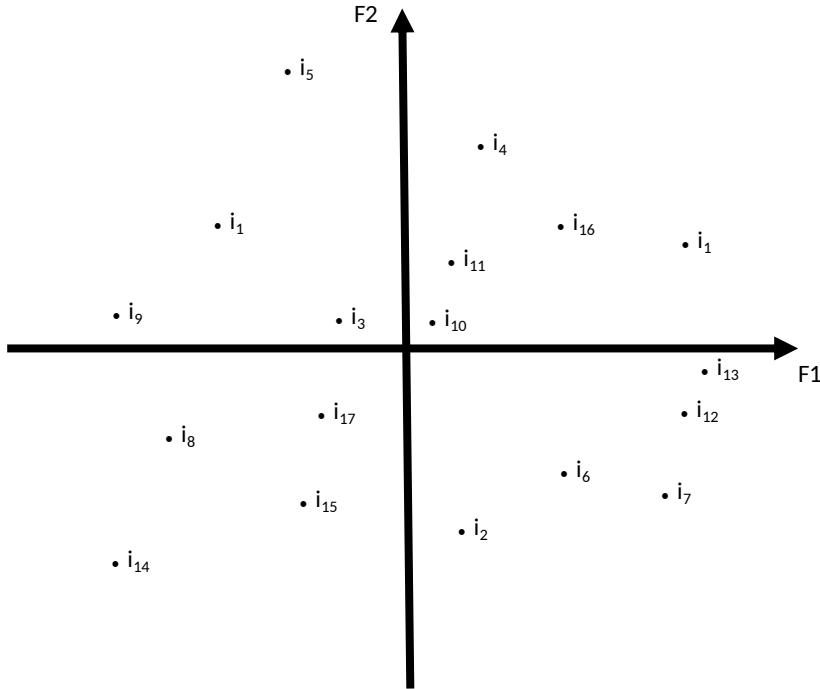


Figura 4-8: Ejemplo de representación de los individuos en el AFM

La Figura 4-8 es una representación gráfica ficticia que ilustra como se verían los individuos en el primer plano factorial. Se observa que el individuo i_4 es el más excéntrico para los dos ejes, los individuos i_{12} e i_{13} van en la dirección del primer factor (contrariamente al individuo i_9), mientras que el individuo i_5 aporta más a la construcción del segundo factor. Individuos como i_3, i_{10}, i_{11} e i_{17} son considerados “individuos promedio” ya que su aporte a la construcción del primer plano factorial es muy débil.

4.2.4. La representación superpuesta de las J nubes definidas por cada grupo de variables

Como todas las nubes N_I^j están situadas en R^K , es posible obtener una representación simultánea por proyección sobre un mismo subespacio. Para obtenerla, se tienen en cuenta las siguientes consideraciones ([Pardo, 2011](#)):

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

$$\tilde{X}_j = [0 \cdots 0 X_j 0 \cdots 0]$$

- $X = \sum_j \tilde{X}_j$, entonces, con el fin de hacer coincidir $F_s(i)$ con el punto medio de los J puntos parciales $F_s^j(s)$, las filas de $\tilde{X}_j$ se proyectan como puntos suplementarios amplificadas por J .
- La unión de las J nubes parciales forma la nube N_I^J .
- La inercia de la nube N_I^J se puede expresar como $InerciaIntra + InerciaEntre$ de las subnubes de $N : i^J$. La inercia de la nube media N_I del AFM corresponde a la $InerciaEntre$ cuando las coordenadas parciales se amplifican por J [Escofier y Pagès \(1992\)](#).

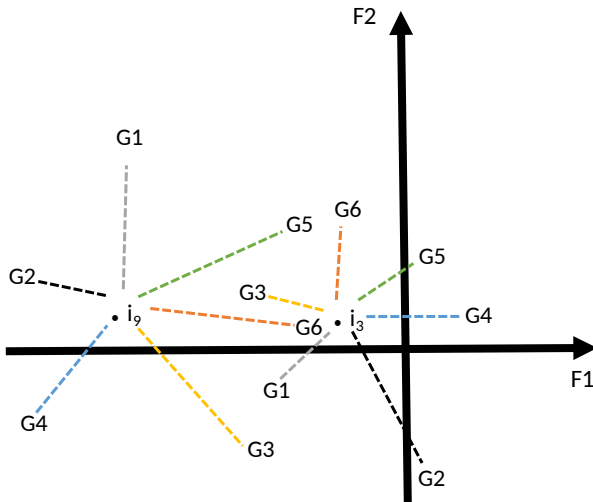


Figura 4-9: Ejemplo de representación superpuesta de la nube de individuos en AFM

La Figura 4-9 es un ejemplo de la representación superpuesta que por razones de claridad se limita a dos individuos, el individuo i_9 está mejor representado por los grupos G5 y G6, mientras que para estos mismos grupos de variables el individuo i_3 tiene un aporte más bajo.

4.2.5. Ayudas a la interpretación en el AFM

Al igual que en el ACP, para identificar el concepto subyacente que representan los ejes factoriales del AFM y comprender las relaciones entre individuos, variables y grupos, se calcula la contribución de cada uno de estos objetivos de análisis a los ejes factoriales.

Contribución de los individuos a una dimensión

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

Donde m_i y F_s^j son, respectivamente el peso del i -ésimo individuo (usualmente $\frac{1}{n}$) y la coordenada del i -ésimo individuo en el s -ésimo eje factorial.

Contribución de las variables a una dimensión

Al igual que con los individuos, se puede encontrar la importancia de las variables para una dimensión dada. Formalmente, la contribución de la k -ésima variable a la s -ésima dimensión está dada por

$$ctr_{k,s} = \frac{\varphi_s^2}{\lambda_j^1}$$

Se debe recordar que todas las variables de un mismo grupo comparten el mismo peso $\frac{1}{\lambda_j^1}$.

Contribución de un grupo de variables a una dimensión

Está definida como la sumatoria de las contribuciones del j -ésimo grupo de variables en el s -ésimo eje factorial. Formalmente

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

Cabe resaltar que la contribución toma valores entre 0 y 1. Para un eje factorial dado, la suma de las contribuciones de todos los individuos (o variables) es igual a 1. Individuos (o variables) con altas contribuciones y con coordenadas factoriales con diferentes signos pueden entonces contrastarse para ayudar a interpretar el componente (Abdi et al., 2013).

4.2.6. Elementos suplementarios

Como en ACP, también es posible proyectar elementos suplementarios en un AFM. Los elementos suplementarios permiten explorar asociaciones con los elementos activos o enriquecer los análisis. Los elementos que participan en el análisis se denominan activos.

Individuos suplementarios

Con frecuencia, es necesario reemplazar el estudio de cada individuo por el de las clases de individuos que tienen una característica en común. Para ello, se introducen como elementos suplementarios los centros de gravedad de estas clases. Estos individuos no influyen sobre las representaciones de los individuos activos: se calcula su proyección en la nube N_I^* y en las diferentes nubes N_I^j .

Grupos de variables suplementarias

Si existe un grupo adicional que se pueda utilizar como una referencia, es interesante compararlo con otros grupos, con todos los medios disponibles para estos últimos, sin que el grupo haya influido sobre la nube media y sobre los resultados del análisis. La mayor parte de los cálculos (pero no todos) efectuados sobre los grupos activos se aplican a un grupo suplementario (Escofier y Pagès, 1992):

- Normalización de la nube N_K^j : es necesario normalizarla de la misma forma sobreponderando las variables del grupo para comparar la nube asociada al grupo suplementario con las otras nubes.
- Proyección de las componentes principales del grupo: a través de ella es posible comparar la forma general de la nube N_K^j con la de la nube media N_K y la de las nubes asociadas a otros grupos de variables.
- Proyección de los $W_j D$: La coordenada de un elemento suplementario $W_j D$ sobre el eje de rango s concuerda con la medida de la relación entre el s -ésimo factor y el grupo j .

Selección del número de ejes factoriales

Como se mencionó anteriormente, el AFM se puede considerar una generalización del ACP, por lo que la selección del número de ejes factoriales a retener pueden seguir los mismos criterios de varianza explicada o valores propios superiores a 1 en el ACP.

4.3. Creación de índices basados en métodos factoriales

Según [Soto y Schuschny \(2009\)](#) un índice es “una función de una o más variables, que conjuntamente ‘miden’ una característica o atributo de los individuos en estudio” y “se construyen con el objetivo de medir el desempeño de una unidad de análisis en un área o tema determinado”. Para construir el índice a través de métodos de análisis factorial se debe satisfacer el factor “tamaño” que se describe a continuación.

4.3.1. Factor de tamaño

La primera componente principal surge de la alta correlación existente entre todas las variables, formando un haz de vectores muy homogéneo. La primera componente correspondería aproximadamente a la bisectriz de este haz y estaría muy correlacionada con todas las variables originales; es por esto que a esta componente se le denomina factor de tamaño (Figura 4-10). ([Aluja et al., 1999](#)).

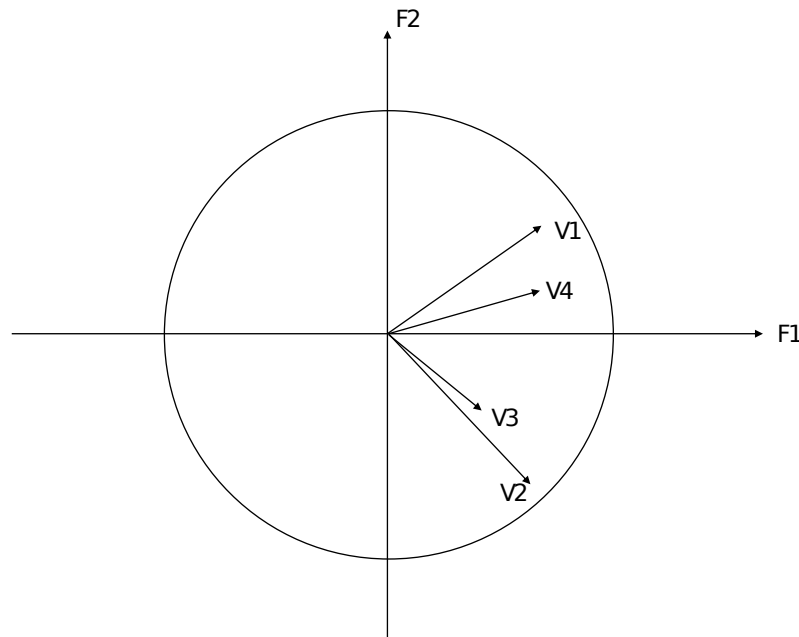


Figura 4-10: Ejemplo de factor *tamaño*

Lo anterior significa que cuando todas las variables están correlacionadas positivamente o negativamente entre ellas, se sitúan del mismo lado del primer eje factorial. En ese caso, con la primera componente principal es posible formar un índice estadístico que sirve para diferenciar los individuos con valores altos de los bajos en el grueso de las variables de

análisis. Por ejemplo, si un individuo toma un valor alto en una variable, es muy probable que su valor sea alto también para todas las otras variables. El índice puede considerarse así: $\psi_s = Zu_s$, donde Z es la matriz de datos estandarizada y u_s es el vector propio asociado al eje s .

4.3.2. Transformación del índice en puntajes no negativos

Generalmente, los índices construidos bajo enfoque de métodos factoriales, se representan en una escala no natural y difícil de interpretar. En este caso, los valores del índice suelen ser transformados linealmente a una escala normada de acuerdo con las siguientes expresiones (Avella et al., 2010):

$$Indice = a\psi_1(i) + b, \quad \text{donde} \quad \psi_1(i) = Comp1$$

$$a = \frac{100}{Max(\psi_1(i)) - Min(\psi_1(i))}$$

$$b = -a \cdot Min(\psi_1(i))$$

La transformación de escala permitirá al analista tener una escala de referencia de los individuos más característicos cercanos al 100 %, los individuos promedios al rededor del 50 % y los individuos menos característicos cercanos al 0 %.

Otra forma de presentar la primera componente principal es utilizando el vector propio asociado a la primera componente principal mediante el siguiente procedimiento (Avella et al., 2010):

- Conseguir una nueva variable que se denomina puntaje-factor denotada por P_{kj} , la nueva variable se calcula restando a cada valor de la variable (original) estandarizada su mínimo, este resultado se multiplica por la ponderación de la variable en la construcción de la componente del AFM u_{kj} :

$$P_{kj} = |X_k - min(X_k)|\dot{u}_{jk}$$

- Encontrar los máximos de cada nueva variable o puntaje-factor:

$$max(P_{kj})$$

- Obtener el puntaje-factor máximo, que es la suma de los máximos de cada nueva variable o puntaje-factor:

$$sumax_J = \sum_{j \in J, k \in K_j} max(P_{kj})$$

- Obtener los puntajes de cada variable:

$$p_{kj} = \frac{P_{kj}}{sumax_J} 100$$

- Construir el índice para el individuo i , de acuerdo a la siguiente ecuación:

$$Indice(i) = \sum_{j \in J, k \in K_j} p_{kj}$$

Capítulo 5

Metodología

En el presente capítulo se describe la metodología empleada en este proyecto para lograr el cumplimiento de los objetivos propuestos. A grandes rasgos, el análisis del problema objeto de estudio se abordó a través de la definición de 7 etapas, las cuales se resumen gráficamente en la Figura 5-1.



Figura 5-1: Secuencia metodológica del análisis de datos

Antes de describir cada una de las fases metodológicas, se detallan los criterios de inclusión y exclusión considerados para la definición de la población objetivo de estudio (Sección 5.1) y la estructura de los datos disponibles (Sección 5.2)

5.1. Definición de la población: criterios de inclusión

La población objeto de estudio la conforman todos los habitantes de la ciudad de Cali y las viviendas de habitación son las unidades muestrales. Debido a limitaciones económicas, no es posible que la estrategia de APS visite a todos los habitantes de la ciudad, ya que de acuerdo a [DAPM \(2017\)](#), para 2018 Cali tenía una población proyectada de 2.445.405 habitantes y 670.917 viviendas. Por ello, la SSPM de Cali selecciona los barrios a intervenir de acuerdo a tres criterios de inclusión:

- Ser Territorio de Inclusión y Oportunidad (**TIO**) para la administración municipal.
- Presentar la mayor cantidad de eventos de interés en salud pública, las cuales se notifican mediante la plataforma **SIVIGILA**.
- Presentar la mayor cantidad de casos de mortalidad, de acuerdo a lo reportado por las estadísticas vitales del **DANE**

Estos criterios y la posterior selección de barrios son definidos por la autoridad sanitaria en salud del municipio. Siguiendo estos criterios, para el 2018 se seleccionaron 18 barrios (16 en la zona urbana y 2 corregimientos de la zona rural). La selección de estos barrios también considera la división de la ciudad en las 5 zonas definidas por la redes de salud pública o Empresas Sociales del Estado (ESE).

Para recolectar la información requerida, se empleó la *ficha de diagnóstico y caracterización de familias* (Anexo A), desarrollada por la SSPM de Cali en conjunto con las ESE del municipio. Inicialmente se planea una visita a todas las viviendas del barrio, a manera de censo. Sin embargo, no todas las visitas tienen el éxito esperado, en algunos casos la vivienda es visitada pero por múltiples factores no se tienen respuesta. La Tabla **5-1** muestra la división de los barrios incluidos por ESE y algunas estadísticas del número total de viviendas y la cobertura de la encuesta. En total se visitaron 17117 viviendas y 28923 habitantes.

5.2. Estructura de los datos

La *ficha de diagnóstico y caracterización de familias* está dividida en siete *módulos* de preguntas o grupos de variables: i. Vivienda, ii. Generalidades, iii. Personas menores de 10 años, iv. Salud sexual y reproductiva-cáncer, v. Estilos de vida saludable, vi. Salud mental y vii. Micobacterias (los módulos iv, v y vi son exclusivos para personas mayores de 10 años). Los datos de cada uno de estos módulos son registrados en diferentes tablas. A cada vivienda se le asigna un código único que sirve como identificador de la misma en cada tabla. Mediante éste código, es posible conocer el número de personas que habitan la vivienda. Se debe mencionar que al comparar la información de las viviendas reportadas por los equipos operativos

ESE	Barrio	Viviendas con encuesta	Censo de viviendas ¹	% Cobertura	Personas encuestadas
Ladera	Cto. Cascajal	332	780	42.56	578
	Cto. El Hormiguero	713	1088	65.44	1020
	Terrón Colorado	2834	5287	52.54	4518
	Prados del Sur	760	1315	57.26	1353
	El Calvario	213	113	185.84	327
	San Pascual	411	296	138.51	567
Oriente	Alfonso Bonilla Aragón	1540	4384	35.10	2960
	Ciudad Córdoba	1204	5912	20.23	2132
	El Diamante	943	1611	58.41	1641
	El Poblado I	1091	2450	44.37	1910
Centro	El Hoyo	24	94	25.53	42
	El Piloto	75	145	51.72	118
	Nueva Floresta	1999	3316	60.28	2388
	Sucre	342	945	35.77	393
Norte	San Luis	1229	1744	70.41	3081
	San Luis II	1160	2188	53.02	2084
	Urbanización Calimio	1069	1800	59.39	2342
Suroriente	Ciudad 2000	1178	3510	33.45	1469
Total		17117	36978	46.04	28923

¹ Fuente: Catastro municipal

Tabla 5-1: Resumen de la obertura de Viviendas- Estrategia APS Cali 2018.

que realizaron la implementación de la *ficha* en los territorios, se encontró que en los barrios El Calvario y San Pascual hay más viviendas que las reportadas por la oficina de Catastro municipal (Tabla 5-1)

Cada tabla puede variar su estructura, debido al cambio de la unidad objetivo de población (Viviendas o personas) y a condiciones particulares como el rango de edad y el sexo del habitante. Por ejemplo, una mujer en edad fértil requiere ser evaluada respecto a condiciones específicas tales como: planificación familiar, gestación, toma de citología, entre otras. De otro lado, un hombre mayor de 50 años, requiere de la evaluación de aspectos relacionados con enfermedades de la próstata y el colon.

El primer módulo de la ficha: *Vivienda*, recoge la información sobre las condiciones de la vivienda. Este modulo presenta un total de 17117 registros. Los módulos ii a vii conforman tablas, en la cuales las observaciones corresponden a los habitantes. Dado que una vivienda suele estar habitada por varias personas, se dispone de una tabla de datos con 29326 registros que conforman el módulo ii. *Generalidades*, la tabla asociada al módulo iii. está conformada por 3225 registros correspondientes a *menores de 10 años* y las tablas de los módulos iv, v, vi y vii contienen 26101 registros que corresponden a los grupos de variables *salud sexual y reproductiva, estilos de vida saludable, salud mental y micobacterias*. Para una mejor comprensión de la metodología de análisis de datos, a estas primeras tablas de datos se llamarán Tablas de datos crudos.

5.3. Procesamiento de datos

En la *Fase 1. limpieza y preprocesamiento de datos*, se identificaron variables con muy baja variabilidad en la respuesta (por ejemplo, casi todos los habitantes tienen servicios de acueducto, pocos niños con signos de malnutrición e inasistentes al programa de salud oral, pocas personas con síntomas respiratorios o diagnóstico de tuberculosis). Sobre estas variables se toma la decisión estratégica de excluirlas de las tablas de datos.

En la *Fase 2. Selección y filtro de variables* se verifica la consistencia de la información recolectada. Es decir, se valida que los campos diligenciados para cada vivienda o persona sean coherentes (por ejemplo, no deben aparecer registros de hombres con toma de citología, entre otras situaciones detectables).

Cabe resaltar que al ser un ejercicio realizado mediante una encuesta, las respuestas son “autoreportes” de las personas encuestadas y evaluaciones subjetivas a criterio del encuestador en las preguntas relacionadas con las condiciones físicas de la vivienda.

Para la realización del presente análisis de datos, las variables de la ficha se agrupan de acuerdo a 5 dimensiones activas comunes de evaluación: G_1 . *Vivienda*, G_2 . *Menores de 10 años*, G_3 . *Salud sexual y reproductiva*, G_4 . *Estilos de vida saludable*, G_5 . *Salud mental*. Particularmente los grupos *salud sexual y reproductiva*, *estilos de vida saludable* y *salud mental* lo conforman variables que se miden sobre el conjunto de habitantes con edad superior a los 10 años. La Tabla 5-2 resume la conformación de los grupos activos.

La variable vacunación incompleta para la edad (**PAI**) en el grupo *menores de 10 años* se tratará como suplementaria. Esto se debe a que vacunación es un programa “sin barreras”, es decir, que la asistencia a vacunación está garantizada y no requiere de ningún trámite o gestión adicional de parte del gobierno municipal. Igualmente se incluyen como variables suplementarias algunas variables del grupo *Generalidades de la población*, tales como: *sexo*, *edad promedio*, *régimen de afiliación* y *etnia*. Se decide incluir estas variables como suplementarias con la intención de analizar la relación con respecto a las variables que miden la salud sentida sin que ellas influyan en la construcción del índice que se desarrolla en la *Fase 6*, el cual es construido con indicadores de salud.

Como se mencionó anteriormente, las variables de la *ficha de diagnóstico y caracterización de familias* se miden sobre diferentes entidades (Viviendas y personas). Sin embargo, el objetivo de este trabajo es realizar una caracterización y priorización de los barrios. Para tal fin, los datos crudos son transformados en la *Fase 3. Consolidación de datos (nivel barrio)*. En esta fase se calculan medidas de frecuencia relativa (%) para cada una de las variables de la tabla de datos crudos, considerando para cada indicador la dirección de mayor riesgo

Grupo (nombre tabla)	Descripción de las Variables (datos primarios)
Vivienda (Vivienda)	Hacinamiento
	Uso de cilindro de gas
	Inadecuado almacenamiento de alimentos
	Inadecuado almacenamiento de agua para consumo humano
	Animales domésticos sin vacunar
	Criaderos de zancudos en la vivienda
	Presencia de roedores en la vivienda
Menores de 10 años (Men10a)	Inasistencia al programa de Crecimiento y Desarrollo
	Inasistencia a los controles de agudeza visual
	Inasistencia a los controles de agudeza auditiva
	Inasistencia al programa de Salud Oral
Salud sexual y reproductiva (SSR)	No uso de método anticonceptivo
	Inasistencia a la toma de citología
	Inasistencia a la toma de mamografía
	Inasistencia a la toma del examen de próstata
	Ha presentado deposiciones con sangre
	Inasistencia al Programa Joven
	Mujeres en estado de embarazo
Estilos de vida saludable (EVS)	Inasistencia a Consulta Odontológica
	Personas con diagnostico de diabetes
	Personas con diagnostico de hipertensión
	Colesterol o trigliceridos alto
	Sobrepeso
	Glucosa alta
Salud mental (sal.ment)	Cambios en el apetito o en el sueño
	Sentimientos depresivos
	Intento de suicidio
	Maltrato Verbal o físico por parte de la pareja
	Violencia sexual
PAI (PAI)	Menores de 10 años sin esquema de vacunación completo
Generalidades (generalidades)	Hombres
	Mujeres
	Afiliación al régimen de salud subsidiado
	Afiliación al régimen de salud contributivo
	Afiliación al régimen de salud especial
	Sin Afiliación a un régimen de salud
	Raizal
	Afrodescendiente
	Gitano
	Palenquero
	Otro
	Ninguno
	Edad Promedio

Tabla 5-2: Estructura de los grupos de variables de la hoja de datos crudos.

de insalubridad sobre la población. Es decir que para todos los indicadores se cumple que el riesgo de insalubridad es mayor cuando el indicador porcentual se incrementa.

La Tabla **5-3**, presenta la estructura final de la tabla de datos activos consolidada, indicando la pertenencia de cada variable a los grupos particulares, además de las etiquetas empleadas para la representación gráfica y su pertenencia a cada uno de los grupos de variables.

Grupo	indicador por barrio (consolidado)	Etiqueta
G_1 : Vivienda (Vivienda)	V_{11} : % de viviendas con hacinamiento	hacinam
	V_{12} : % de viviendas con cilindro de gas	cilindr.gas
	V_{13} : % de viviendas con inad. almac. de alim.	no_alm.alim
	V_{14} : % de viviendas con inad. almac. de agua	alm.agua
	V_{15} : % de viviendas con anim.domes.sin.vac	anim.domes.sin.vac
	V_{16} : % de viviendas con criad.zancudos	criad.zancudos
	V_{17} : % de viviendas con presenc.roedores	presenc.roedores
G_2 : Menores de 10 años (Men10a)	V_{21} : % de menores inastentes a Crec. y Desar.	cydno
	V_{22} : % de menores inastentes a agudeza vis.	cntrlvisno
	V_{23} : % de menores inasistentes a agudeza aud.	cntrlaudno
	V_{24} : % de menores inastentes a salud oral	soralno
G_3 : Salud sexual y reproductiva (SSR)	V_{31} : % de mujeres que no utilizan anticonceptivos	plnfno
	V_{32} : % de mujeres entre 10 y 69 años sin citología	citolno
	V_{33} : % de mujeres mayores de 50 años sin mamografía	mamogno
	V_{34} : % de hombres mayores de 50 años sin examen de próstata	prostno
	V_{35} : % de personas que han presentado deposiciones con sangre	altcolonsi
	V_{36} : % de personas inasistentes al programa joven	prjvnno
	V_{37} : % de mujeres en estado de embarazo	gstssi
G_4 : Estilos de vida saludable (EVS)	V_{41} : % de personas inasistentes a consulta odontológica	cons.odont
	V_{42} : % de personas con diabetes	diabeticos
	V_{43} : % de personas con hipertensión	hipert
	V_{44} : % de personas con coelsterol y trigliceridos altos	colesterol_alto
	V_{45} : % de personas con sobrepeso	sobrepeso
	V_{46} : % de personas con glucosa alta	glucosa
G_5 : Salud mental (sal.ment)	V_{51} : % de personas con cambios en el apetito o el sueño	camb.apet
	V_{52} : % de personas con depresión	depresion
	V_{53} : % de personas con intento de suicidio	int.suicidio
	V_{54} : % de personas con maltrato verbal o físico	malt.verb.fis
	V_{56} : % de personas victimas de violencia sexual	viol.sex
PAI (PAI)	V_{61} : % de menores de 10 años sin vacunas completas para la edad	PAI
Generalidades (generalidades)	V_{71} : % de Hombres	hombres
	V_{72} : % de Mujeres	mujeres
	V_{73} : % de personas afiliadas al régimen de salud subsidiado	subs
	V_{74} : % de personas afiliadas al régimen de salud contributivo	cont
	V_{75} : % de personas afiliadas al régimen de salud especial	esp
	V_{76} : % de personas sin Afiliación a un régimen de salud	afregno
	V_{77} : % de personas con etnia Raizal	raizal
	V_{78} : % de personas con etnia Afrodescendiente	afrod
	V_{79} : % de personas con etnia Gitano	gitano
	V_{710} : % de personas con etnia Palenquero	palenq
	V_{711} : % de personas con etnia Otro	otro
	V_{712} : % de personas con etnia Ninguno	ninguno
	V_{713} : Edad Promedio	edadpr

Tabla 5-3: Listado de indicadores resumen de barrio por grupo de variables y etiquetas

Así, la tabla de datos final la conforman los indicadores resumen de las variables para cada barrio, en una estructura como la que se presenta en la Tabla 5-4.

Barrio	G_1 : Vivienda			Personas											
				G_2 : Menores de 10 años			Mayores de 10 años								
	G_3 : SSR						G_4 : EVS			G_5 : SM					
	V_{11}	$\cdots$	V_{18}	V_{21}	$\cdots$	V_{24}	V_{31}	$\cdots$	V_{37}	V_{41}	$\cdots$	V_{46}	V_{51}	$\cdots$	V_{55}
B_1	%	$\cdots$	%	%	$\cdots$	%	%	$\cdots$	%	%	$\cdots$	%	%	$\cdots$	%
B_2	%	$\cdots$	%	%	$\cdots$	%	%	$\cdots$	%	%	$\cdots$	%	%	$\cdots$	%
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
B_{18}	%	$\cdots$	%	%	$\cdots$	%	%	$\cdots$	%	%	$\cdots$	%	%	$\cdots$	%

Tabla 5-4: Estructura de la tabla de datos final - nivel barrios (grupos activos)

La Tabla 5-4 contiene la información relacionada con los objetivos del presente estudio. A partir de este punto los análisis estadísticos que se plantean son desarrollados sobre los datos de la Tabla 5-4. Como una aproximación a las variables y al fenómeno estudio, en la *Fase 4* se realiza un *análisis exploratorio de datos*. En este análisis, las variables son visualizadas a través de gráficos boxplot y de correlación. El análisis por Zona no es realizado, debido a la pequeña cantidad de datos disponibles para cada subdivisión. Por ejemplo, la Zona Suroriente solo tiene 1 barrio asociado y la zona norte 3 barrios.

La Fase 5. *Análisis Factorial Múltiple*, es el punto central de la metodológica propuesta. Este análisis se realiza sobre la tabla de datos consolidada (Tabla 5-4). La estructura del análisis se describe a continuación en la Tabla 5-5.

Individuos	Grupos de variables	Variables activas					Variables suplem.	
		G_1	G_2	G_3	G_4	G_5	G_6	G_7
18	7	7	4	7	6	5	1	13

Tabla 5-5: Resumen de la estructura de datos para el AFM

Como se mencionó anteriormente, la cobertura en vacunación es de interés para la SSPM, sin embargo, a diferencia de los demás programas involucrados en la estrategia de APS, la vacunación no requiere citas previas o protocolos de atención, por lo cual la variable **PAI** se utilizará como variable suplementaria en el análisis. Igualmente sucede con las generalidades (*edad*, *sexo*, *afiliación a Sistema General de Seguridad Social de Salud* y la *etnia*). De esta forma, en total se cuenta con 43 variables: 29 activas y 14 suplementarias.

El AFM se realiza entonces sobre 18 barrios y 29 variables activas ($I < K$). Según [Jolliffe y Cadima \(2016\)](#):

“Cuando $n < p$, es el número de individuos observados, en lugar del número de variables, es lo que generalmente determina el rango de la matriz. El rango de la matriz de datos centrada en columnas X^ (o su contraparte estandarizada Z) debe ser igual al rango de la matriz de covarianza (o correlación). La implicación práctica de esto es que solo hay r valores propios distintos de cero; por lo tanto, r componentes principales explican toda la variabilidad en el conjunto de datos. **Nada impide el uso de ACP en tales contextos.** Los componentes principales se pueden determinar como de costumbre, ya sea mediante una descomposición singular de la matriz de datos (centrada) o los autovectores/valores de la matriz de covarianza (o correlación)”.*

Además de transformar las variables en porcentajes, estas también fueron estandarizadas posteriormente para nivelar rangos de variación. Para la retención de ejes factoriales se utilizó el criterio de porcentaje de varianza acumulada, se obtuvo que los dos primeros ejes factoriales retienen el 60.88 % de la inercia contenida en los datos. La representación de los individuos se realizó a través del gráfico de proyección de individuos y la representación de variables mediante el círculo de correlación. Se midieron las contribuciones y la calidad de la representación tanto de los individuos como de las variables, así como los coeficientes Lg y RV con los que se evalúa la relación entre los grupos de variables.

Posteriormente, luego de verificar sobre el primer eje factorial el cumplimiento del *factor de tamaño*, se procede a la *Fase 6. Construcción del índice sintético* utilizando el primer eje factorial del AFM. Este índice es normalizado de acuerdo al método propuesto por [Avella et al. \(2010\)](#) descrito en la Sección 4.3.

Finalmente, éste índice sintético es empleado en la *Fase 7. Priorización de los barrios* para generar un *ranking* de barrios, considerando que este índice resumirá las condiciones de salud sentidas de estos territorios. Como resultado, se tendrá entonces que a mayor valor del índice mayor será la insalubridad del barrio, razón por la cual los resultados de esta fase pueden ser empleados por la administración pública como criterio de priorización de recursos y de estrategias de intervención.

5.3.1. Software

Se empleó el software estadístico R versión 4.0.3, para el realizar el procesamiento y análisis de las tablas, con ayuda de los paquetes *FactoMineR* y *factoextra* se obtuvieron las salidas gráficas y las tablas para la interpretación de los resultados obtenidos con el Análisis Factorial Múltiple (AFM).

Capítulo 6

Resultados

En éste capítulo se procede a describir los resultados obtenidos en el desarrollo de las fases 4-6, relacionadas con el análisis de datos, de acuerdo con la metodología planteada en la Sección 5. Se inicia con un análisis descriptivo de las variables consolidadas por barrio, haciendo uso de herramientas gráficas para observar la distribución marginal de cada variable y la estructura de correlación entre ellas, finalmente se presentan los resultados obtenidos en el Análisis Factorial Múltiple (AFM), la construcción del índice sintético y la priorización de los barrios.

6.1. Análisis Exploratorio de Datos

La Figura 6-1 y la Tabla 6-1 resumen el comportamiento de las variables. En promedio, el 3.21 % de las viviendas en los barrios bajo estudio presentan condición de hacinamiento, el 13.38 % usa cilindros de gas, el 9.41 % realiza inadecuado almacenamiento de alimentos, en el 4.87 % se encontraron criaderos de zancudos, en el 7.94 % se evidenció presencia de roedores y el 22.01 % de los animales domésticos no tiene vacunas.

Dentro del grupo de variables asociadas a las condiciones de salud en *menores de 10 años*, la mayor inasistencia promedio se observó en el programa de Crecimiento y Desarrollo. De acuerdo con los datos, el 12.60 % de niños no asisten a este programa, seguido de la inasistencia al programa de Salud Oral con 11.74 %. Además se observa que el porcentaje de inasistencia promedio a vacunación es del 6.59 %, mientras que la inasistencia a los controles de agudeza visual y auditiva es muy baja (prácticamente todos los niños asistente al programa). Cabe resaltar que el control de la agudeza visual y auditiva se realiza en niños hasta los cinco años mientras que el programa de Crecimiento y Desarrollo realiza controles a los niños hasta los 10 años.

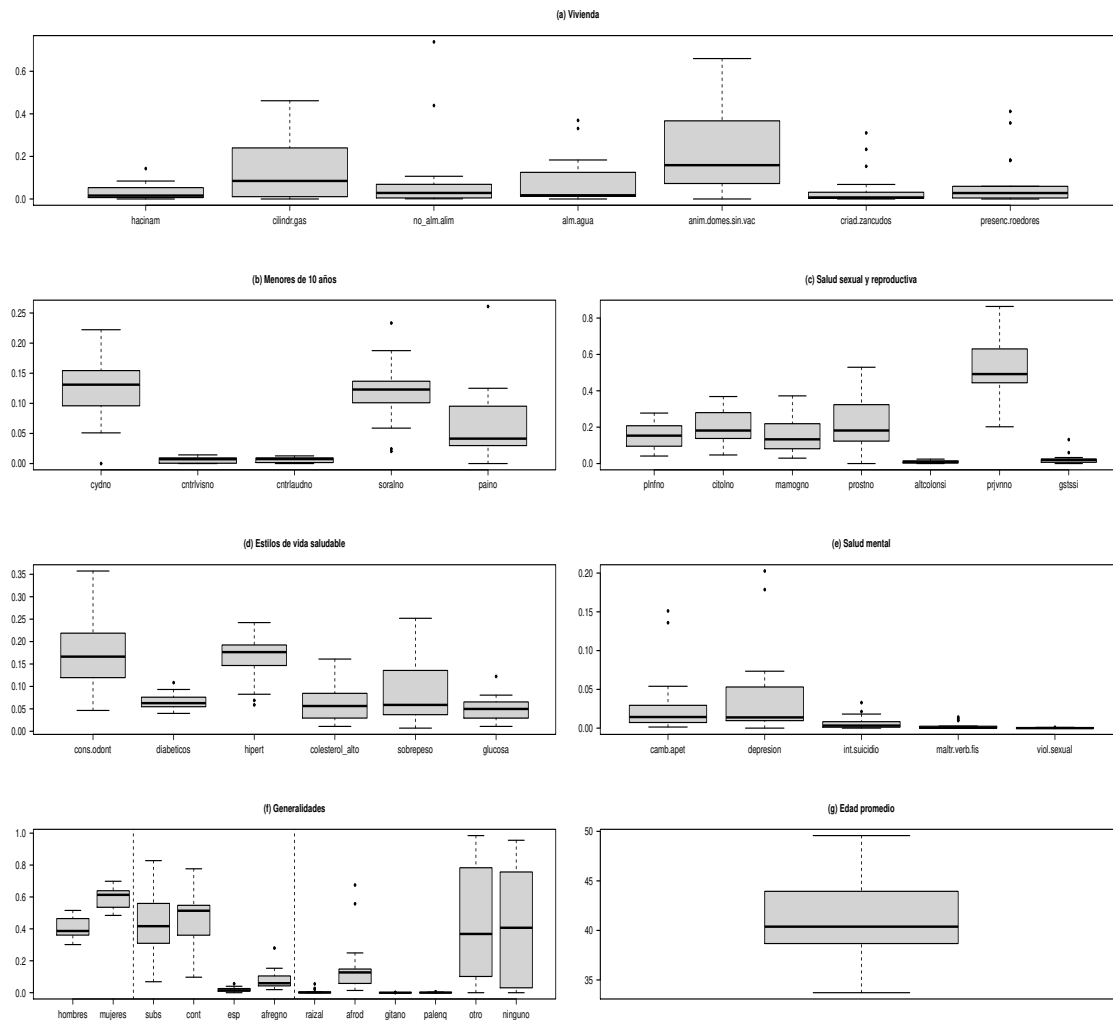


Figura 6-1: Distribución de las variables por grupos

En el grupo de variables relacionadas a *salud sexual y reproductiva*, se observa un promedio de 51.94 % de inasistencia al programa joven y 20 % de las mujeres no asisten a la toma de la citología. Similar cantidad, 20 %, de los hombres no asiste al examen de próstata. La insistencia a la mamografía se presentan en promedio para el 15.6 % de las mujeres mayores de 50 años, mientras que el 15.3 % de las mujeres en edad fértil manifiesta no emplear un método anticonceptivo.

Se encontraron pocos casos de personas que manifestaron haber realizado *deposiciones con sangre* en el ultimo año y de *mujeres en estado de embarazo*. En los barrios El Calvario y El Piloto se encontró que el 6 % y el 13 % de las mujeres en edad fértil se encontraban en estado de embarazo al momento de la visita.

En el grupo de variables asociadas a *estilos de vida saludable*, se encontró que en promedio, el 17.5 % de las personas no asiste a consulta odontológica, el 16.3 % de las personas mayores de 10 años son hipertensas, el 6.48 % de las personas son diabéticas y el promedio de personas con sobrepeso es de 8.59 %.

En cuanto al grupo de variables de *salud mental*, se observa que la variable *depresión* tiene un rango de variabilidad mayor que las demás, el promedio de personas con depresión es de 4 %. Además, se observa que son muy pocos los casos encontrados de *maltrato verbal o físico, violencia sexual e intento de suicidio*.

Grupo	Variable	Mínimo	Cuartil 1	Mediana	Promedio	Cuartil 3	Máximo	Desv.Est
Vivienda	hacinam	0.000	0.006	0.015	0.032	0.050	0.142	0.036
	cilindr.gas	0.000	0.012	0.084	0.133	0.232	0.46	0.141
	no.alm.alim	0.000	0.007	0.028	0.094	0.063	0.738	0.189
	alm.agua	0.000	0.012	0.017	0.082	0.123	0.369	0.113
	anim.domes.sin.vac	0.000	0.079	0.158	0.220	0.345	0.659	0.191
	criad.zancudos	0.000	0.002	0.007	0.048	0.030	0.310	0.090
	presenc.roedores	0.000	0.005	0.027	0.079	0.059	0.411	0.124
Menores de 10 años	cydno	0.000	0.09	0.130	0.126	0.153	0.222	0.053
	cntrlvisno	0.000	0.001	0.007	0.005	0.009	0.014	0.004
	cntrlaudno	0.000	0.002	0.007	0.006	0.009	0.012	0.004
	soralno	0.020	0.101	0.122	0.117	0.135	0.233	0.052
Salud sexual y reproductiva	plnfno	0.041	0.101	0.153	0.156	0.204	0.277	0.069
	citolno	0.047	0.139	0.181	0.201	0.269	0.368	0.099
	mamogno	0.029	0.083	0.133	0.153	0.207	0.372	0.096
	prostno	0.000	0.126	0.181	0.204	0.296	0.529	0.136
	altcolonsi	0.000	0.003	0.007	0.008	0.014	0.024	0.007
	prjvnno	0.202	0.450	0.491	0.519	0.626	0.863	0.186
	gstssi	0.000	0.007	0.019	0.024	0.024	0.131	0.030
Estilos de vida saludable	cons.odont	0.046	0.120	0.166	0.175	0.218	0.357	0.084
	diabeticos	0.039	0.055	0.062	0.066	0.075	0.108	0.018
	hipert	0.058	0.147	0.176	0.163	0.190	0.242	0.051
	colesterol_alto	0.010	0.029	0.056	0.064	0.082	0.161	0.042
	sobrepeso	0.007	0.037	0.058	0.085	0.124	0.251	0.072
	glucosa	0.010	0.029	0.049	0.049	0.064	0.122	0.027
Salud mental	camb.apet	0.001	0.007	0.014	0.030	0.028	0.151	0.043
	depresion	0.000	0.010	0.013	0.040	0.047	0.202	0.058
	int.suicidio	0.000	0.001	0.003	0.006	0.007	0.032	0.008
	maltr.verb.fis	0.000	0.000	0.000	0.002	0.002	0.014	0.004
	viol.sexual	0.000	0.000	0.000	0.000	0.000	0.001	0.0003
PAI	paino	0.000	0.030	0.041	0.065	0.091	0.260	0.060
Generalidades	hombres	0.301	0.363	0.386	0.401	0.457	0.515	0.065
	mujeres	0.484	0.543	0.613	0.598	0.636	0.698	0.065
	subs	0.068	0.327	0.416	0.421	0.550	0.827	0.191
	cont	0.097	0.365	0.513	0.479	0.546	0.776	0.166
	esp	0.000	0.009	0.014	0.019	0.025	0.056	0.014
	afregno	0.019	0.043	0.059	0.080	0.097	0.279	0.061
	raizal	0.000	0.000	0.001	0.006	0.004	0.054	0.014
	afrod	0.013	0.059	0.127	0.158	0.147	0.674	0.061
	gitano	0.000	0.000	0.000	0.000	0.000	0.001	0.0004
	palenq	0.000	0.000	0.000	0.001	0.001	0.005	0.001
	otro	0.000	0.105	0.368	0.423	0.756	0.984	0.361
	ninguno	0.000	0.037	0.406	0.409	0.756	0.955	0.378
	edadpr	33.71	38.74	40.38	41.14	43.73	49.57	4.11

Tabla 6-1: Estadísticas descriptivas de las variables

La Tabla 6-2 muestra los valores de las variables consolidadas de algunos barrios que se muestran como atípicos en la Figura 6-1. El barrio El Calvario presenta una condición extrema de altos niveles para las variables: *Hacinamiento*, *Inadecuado almacenamiento de alimentos*, *criadero de zancudos* y *presencia de roedores*. Los corregimientos de Cascajal y Hormiguero presentan condiciones atípicas, puesto que los valores de todas las variables del grupo *vivienda*, excepto *Hacinamiento* e *Inadecuado almacenamiento de alimentos*, son bastante altos en comparación con los demás barrios, mientras que San Pascual tiene los valores más altos para las variables *Inadecuado almacenamiento de alimentos* y *presencia de roedores*.

Barrio	hacinam	no_alm.alim	alm.agua	criad.zancudos	presenc.roedores
CORREGIMIENTO CASCAJAL	0.05	0.03	0.33	0.15	0.18
CORREGIMIENTO EL HORMIGUERO	0.08	0.07	0.37	0.31	0.41
EL CALVARIO	0.14	0.74	0.01	0.23	0.36
SAN PASCUAL	0.06	0.44	0.01	0.07	0.18
	cydno	soralno	paino		
CIUDAD 2000	0.00	0.02	0.02		
CORREGIMIENTO CASCAJAL	0.22	0.23	0.10		
SAN LUIS II	0.05	0.02	0.02		
SUCRE	0.17	0.15	0.26		
	diabeticos	hipert	glucosa		
CORREGIMIENTO CASCAJAL	0.09	0.18	0.12		
EL CALVARIO	0.04	0.07	0.01		
EL HOYO	0.09	0.06	0.03		
EL POBLADO I	0.11	0.24	0.06		
	camb.apet	depresion	int.suicidio	maltr.verb.fis	viol.sexual
CORREGIMIENTO CASCAJAL	0.14	0.18	0.02	0.009	0.00
CORREGIMIENTO EL HORMIGUERO	0.15	0.20	0.03	0.014	0.001
NUEVA FLORESTA	0.01	0.01	0.00	0.00	0.0004

Tabla 6-2: Valores atípicos en cada grupo de variables

Con relación al grupo de variables *menores de 10 años*, en el barrio Ciudad 2000 no se encontraron menores de 10 años inasistentes al programa de Crecimiento y Desarrollo y muy pocos casos de inasistencia a salud oral, contrario a lo ocurrido en el Corregimiento Cascajal, donde cerca del 25 % de los menores no asisten a este programa. El barrio Sucre presentó el porcentaje más alto de inasistencia al programa de vacunación.

Con respecto al grupo de variables *estilos de vida saludable*, el barrio El Poblado I tiene el mayor porcentaje de personas *diabéticas*, El Hoyo y El Calvario tienen menor porcentaje de personas *hipertensas* y en el corregimiento Cascajal se encontró el mayor porcentaje de personas con valores de *glucosa* alta.

Los corregimientos de Cascajal y Hormiguero presentan los porcentajes más altos en las variables *cambios en el apetito*, *depresión*, *maltrato verbal o físico*, en ésta última el barrio Prados del Sur presentó un 1.18 % de casos reportados, mientras que el mayor porcentaje de casos de *violencia sexual* se encontró en el corregimiento El Hormiguero.

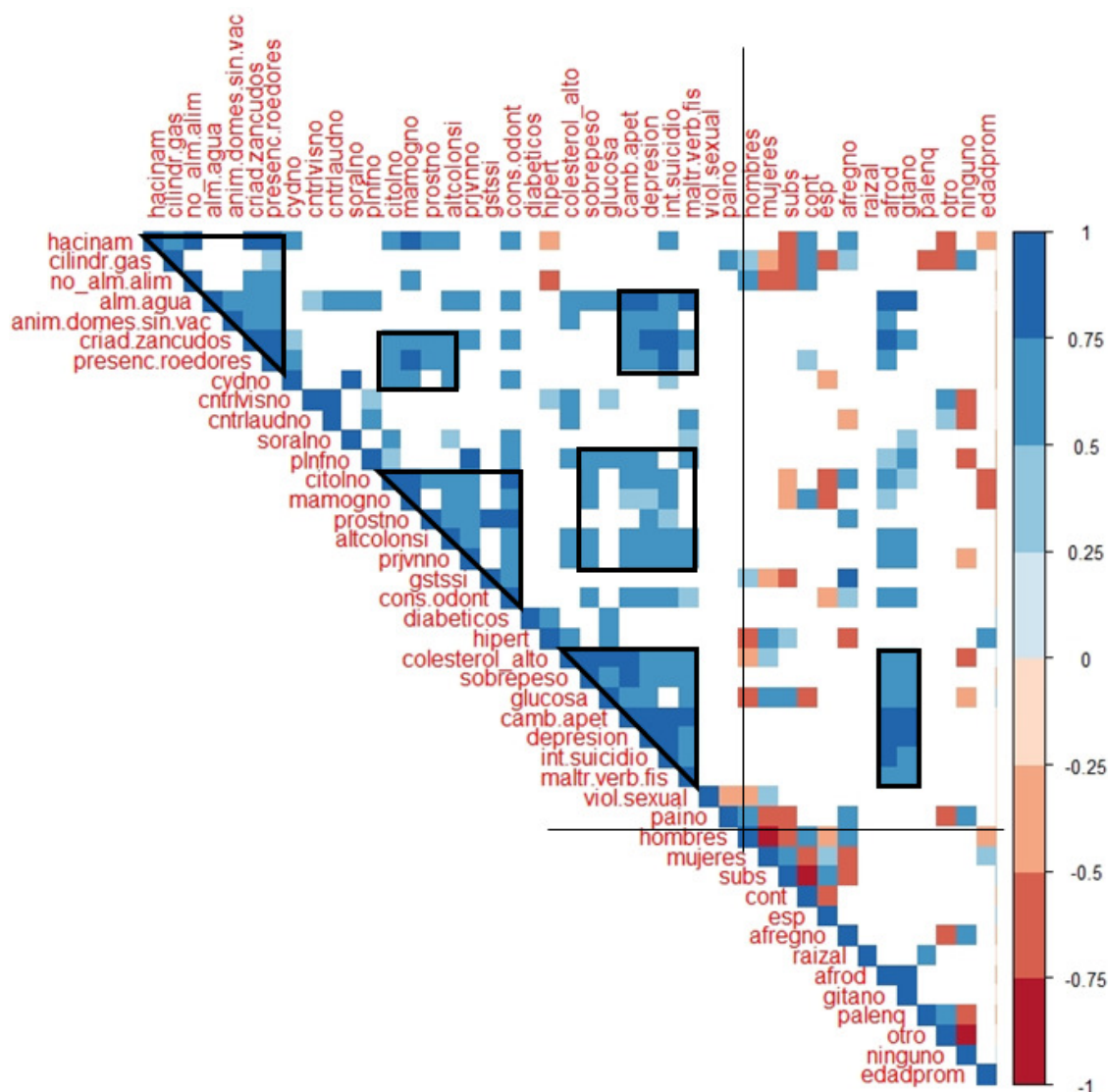


Figura 6-2: Gráfico de la matriz de correlación de las variables consolidadas

La Figura 6-2 es una representación de la estructura de correlación entre todos los pares de variables. Para efectos de simplificar la representación, aquellas correlaciones no significativas ($Valor-p > 0.05$) no presentan un color asignado. En ésta figura se destaca que dentro de los grupos *vivienda*, *salud sexual y reproductiva* y *estilos de vida saludable* sus variables presentan una correlación superior a 0.5.

Las variables del grupo *estilos de vida saludable*: *colesterol alto*, *sobrepeso* y *glucosa* están correlacionadas con las variables del grupo *salud mental*, al igual que las variables *inasistencia al examen de próstata*, *inasistencia al programa joven* y *deposiciones con sangre*.

Las etnias *afrodescendiente* y *gitano* presentan correlación positiva con las variables *coleste-*

rol alto, sobrepeso, glucosa alta, cambios en el apetito, depresión, intento de suicidio. Esto significa que existe una tendencia a incrementarse el riesgo de padecer esta clase de enfermedades para los individuos que pertenecen a estos grupos étnicos. Particularmente se observa que los indicadores de salud mental están correlacionadas positivamente con las variables del grupo *vivienda:inadecuado almacenamiento de agua, presencia de criaderos de zancudos, presencia de roedores* y con las variables *Deposiciones con sangre, inasistencia a programa joven, colesterol alto y sobrepeso*. Desde lo cual se deduce que condiciones de pobreza y mal condición de la vivienda se relaciona con un mayor riesgo de enfermedades de salud mental y estilo de vida saludable.

6.2. Análisis Factorial Múltiple

Se realiza el análisis de la tabla completa de datos mediante el Análisis Factorial Múltiple, teniendo en cuenta que para cada variable el porcentaje consolidado a nivel de barrio se calculó sobre la modalidad que indica una mayor condición de riesgo para la salud pública y que las tablas se conformaron con las frecuencias relativas de estas modalidades para cada barrio.

6.2.1. Análisis de Factores

Inicialmente se define el número de dimensiones a ser analizados, a través del análisis de los valores propios que representan la inercia o la variación explicada por cada dimensión.

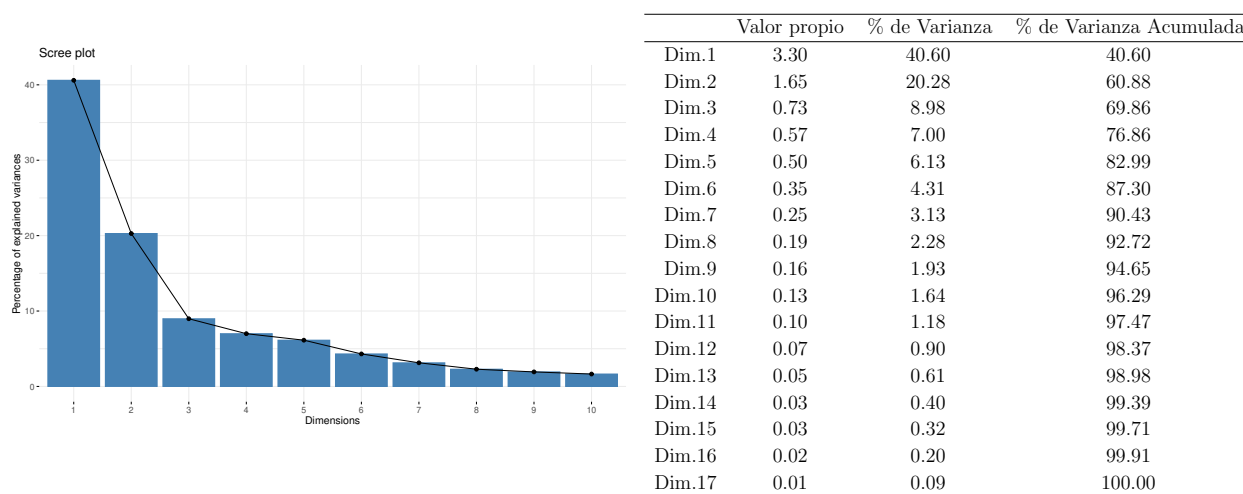


Figura 6-3: Gráfico de sedimentación y valores propios del AFM

En la Figura 6-3 se observa que se obtuvo que los dos primeros ejes factoriales acumulan el 60.88 % de la variabilidad de la tabla de datos, resaltando el primer eje factorial, que aporta

un 40.6 % de la inercia total y el segundo un 20.28 %. Como se observa, la variabilidad recogida en el primer plano factorial es suficientemente grande y en consecuencia, el análisis de datos se centrará en el estudio de los barrios, de las variables y de los grupos de variables sobre este primer plano.

6.2.2. Ponderación de los grupos de variables

El AFM equilibra la influencia de los grupos de variables dando a cada variable dando a cada variable un peso. Este peso debe ser el mismo para todas las variables de un mismo grupo, con el fin de conservar la estructura interna de cada grupo. [Escofier y Pagès \(1992\)](#). El peso de las variables de un grupo corresponde al inverso del primer valor propio del ACP de ese grupo. Así, los primeros valores propios obtenidos por los ACP separados de cada grupo, sus inversos y el número de variables activas por grupo, se presentan en la Tabla **6-3**

Grupo	Cantidad de variables	λ_1^j	$1/\lambda_1^j$
Vivienda	7	4.24	0.23
Menores de 10 años	4	2.57	0.38
Salud sexual y reproductiva	7	3.99	0.25
Estilos de vida saludable	6	3.36	0.29
Salud mental	5	3.59	0.27

Tabla 6-3: Valores propios y ponderaciones de los ACP separados por grupo

En la Tabla **6-3** se observa un peso similar para todas las tablas, exceptuando el grupo de variables relacionadas con *menores de 10 años* cuyo peso es ligeramente superior. Esto se puede presentar debido a una más baja estructura de correlación entre las variables que conforman este grupo o menor variabilidad de las variables.

	Dimensión 1	Dimensión 2
Inercia total	3.30	1.65
Vivienda	0.64	0.47
Menores de 10 años	0.47	0.30
Salud sexual y reproductiva	0.77	0.27
Estilos de vida saludable	0.61	0.54
Salud mental	0.80	0.06

Tabla 6-4: Descomposición de la inercia de las dos primeras dimensiones para los cinco grupos de variables

De acuerdo con la Tabla **6-4**, para la primera dimensión, las inercias de las variables de los grupos salud mental y salud sexual y reproductiva están más próximas del valor 1, es decir, esta primera dimensión está muy relacionada con estos dos grupos. La segunda dimensión

presenta un más notable relación con el grupo de variables asociada a los estilos de vida saludable, aunque su inercia es tan solo de 0.54.

6.2.3. Tipología de los grupos

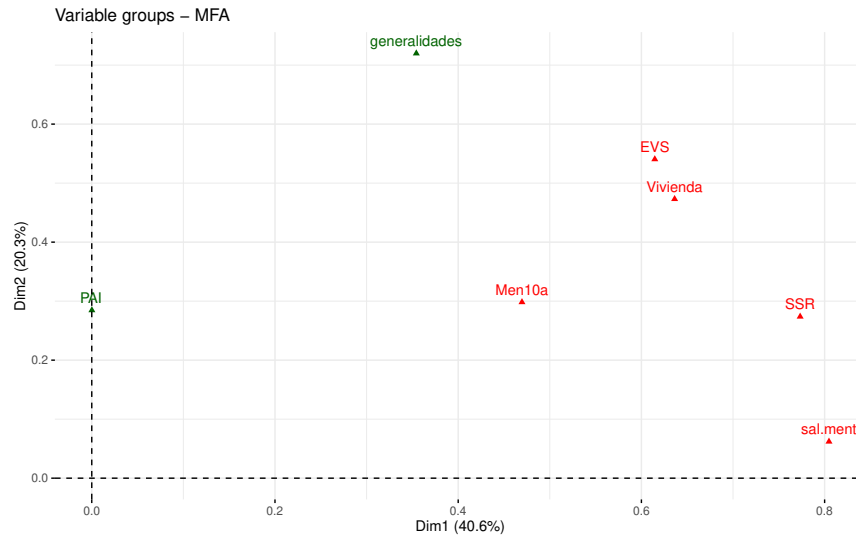


Figura 6-4: Representación de los grupos en el análisis global

La Figura 6-4 ilustra la correlación entre grupos y dimensiones. El grupo de variables del módulo de salud mental está mejor representado en la primera dimensión, pues es el que tiene la coordenada más alta. Por su parte, la variable vacunación incompleta para la edad (PAI) presenta una coordenada cercana a cero a lo largo del eje 1, mostrando que está poco relacionada con la primera dimensión.

	Dimensión 1	Dimensión 2
Vivienda	0.84	0.81
Menores de 10 años	0.69	0.69
Salud sexual y reproductiva	0.89	0.74
Estilos de vida saludable	0.84	0.81
Salud mental	0.90	0.51

Tabla 6-5: Correlaciones para los dos primeros ejes del AFM, entre las proyecciones de la nube global y las de cada una de las cinco nubes asociadas a un solo grupo

La Tabla 6-5 muestra que los grupos de variables salud mental, salud sexual y reproductiva, vivienda y estilos de vida saludable tienen una correlación superior a 0.80 con la primera dimensión, mientras que la correlación del grupo de variables asociadas a menores de 10

años es de 0.69. En la segunda dimensión, cuatro de los cinco grupos presentan correlaciones mayores a 0.70, pues la correlación del grupo salud mental es de 0.51.

Los grupos de estilos de vida saludable y vivienda tienen las coordenadas más altas en la segunda dimensión, que indican una contribución más alta. La variable PAI, tratado como ilustrativa, está más relacionada a esta dimensión.

	Vivienda	Men10a	SSR	EVS	sal.ment	PAI	generalidades	MFA
Vivienda	1.15	0.28	0.57	0.40	0.65	0.17	0.65	0.93
Men10a	0.28	1.23	0.37	0.37	0.22	0.30	0.42	0.75
SSR	0.57	0.37	1.19	0.62	0.52	0.09	0.57	0.99
EVS	0.40	0.37	0.62	1.23	0.55	0.10	0.62	0.96
sal.ment	0.65	0.22	0.52	0.55	1.06	0.07	0.42	0.91
PAI	0.17	0.30	0.09	0.10	0.07	1.00	0.50	0.22
generalidades	0.65	0.42	0.57	0.62	0.42	0.50	1.45	0.81
MFA	0.93	0.75	0.99	0.96	0.91	0.22	0.81	1.38

Tabla 6-6: Coeficientes Lg

De acuerdo con [Corrales \(2010\)](#) “el coeficiente Lg mide la dimensionalidad de cada grupo. El coeficiente vale cero cuando los grupos no están relacionados y no tiene límite superior”. En la Tabla 6-6 se puede apreciar que éstos coeficientes son cercanos a cero, por tanto, se puede concluir que los grupos no son similares. Lo anterior cobra sentido si recordamos que las líneas de atención en salud pública que hacen parte de la ficha de caracterización, evalúan aspectos particulares para diferentes cursos de vida.

	Vivienda	Men10a	SSR	EVS	sal.ment	PAI	generalidades	MFA
Vivienda	1.00	0.24	0.49	0.34	0.59	0.16	0.51	0.74
Men10a	0.24	1.00	0.30	0.30	0.19	0.27	0.31	0.58
SSR	0.49	0.30	1.00	0.51	0.46	0.08	0.43	0.77
EVS	0.34	0.30	0.51	1.00	0.48	0.09	0.47	0.74
sal.ment	0.59	0.19	0.46	0.48	1.00	0.07	0.34	0.75
PAI	0.16	0.27	0.08	0.09	0.07	1.00	0.41	0.19
generalidades	0.51	0.31	0.43	0.47	0.34	0.41	1.00	0.58
MFA	0.74	0.58	0.77	0.74	0.75	0.19	0.58	1.00

Tabla 6-7: Coeficientes RV

Según [Abascal y Calvo \(2002\)](#) “el coeficiente RV se consigue a partir de los coeficientes de correlación lineal entre dos variables cualesquiera, por ello este coeficiente mide la relación de los grupos teniendo en cuenta las variables de cada grupo y su valor está comprendido entre 0 y 1”.

Teniendo en cuenta que los resultados de la Tabla **6-7**, en donde todos los coeficientes son menores a 0.6, se infiere que existe una relación débil entre las variables que pertenecen a diferente grupos.

6.2.4. Representación de las variables

Las reglas de interpretación de estos primeros resultados son idénticas a las de un ACP. Teniendo en cuenta los porcentajes de inercia (40.6 % para la dimensión 1; 20.28 % para la dimensión 2).

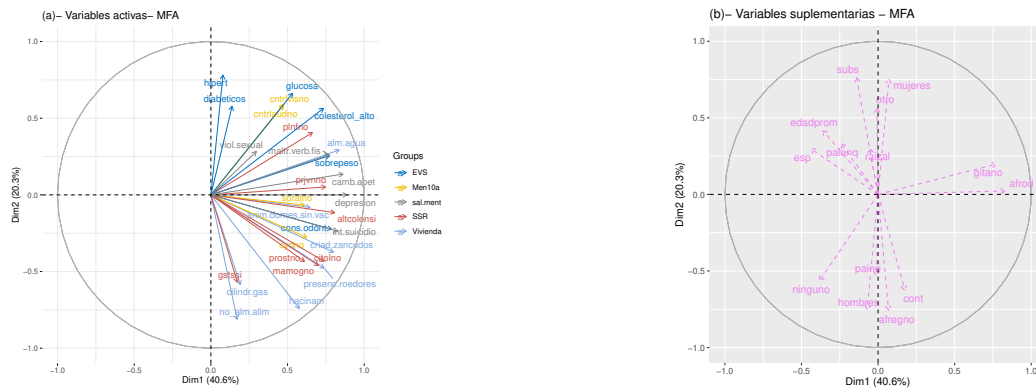


Figura 6-5: Circulo de correlación de las variables en el análisis global

La Figura **6-5** presenta la correlación entre las variables y las dimensiones, se observa que todas las variables activas (a), tienen correlación positiva con la primera dimensión. Por lo tanto se cumple la condición de factor tamaño, necesaria para formar un índice estadístico resumen del “riesgo autoreportado”. Adicionalmente, se observa que en el primer cuadrante se encuentran las variables del grupo estilos de vida saludable, a excepción de inasistencia a consulta odontológica, junto con la inasistencia a programas de planificación, del joven, de control auditivo y visual, violencia sexual, maltrato verbal o físico e inadecuado almacenamiento de agua para consumo.

En el cuarto cuadrante, se ubican la mayor parte de las variables de los grupos *Salud mental*, *Salud sexual y reproductiva*, *Vivienda* y *Menores de 10 años*. En (b) se observa que las variables suplementarias *gitano* y *afrodescendiente* se encuentran en la misma dirección que las variables activas, indicando que las condiciones de salud sentida en estos grupos poblacionales son más desfavorables. La edad promedio se encuentra en dirección opuesta a las condiciones de la vivienda, mostrando que los barrios donde las condiciones de vivienda son más desfavorables tienen una estructura poblacional más joven que en aquellos barrios con mejores condiciones de vivienda.

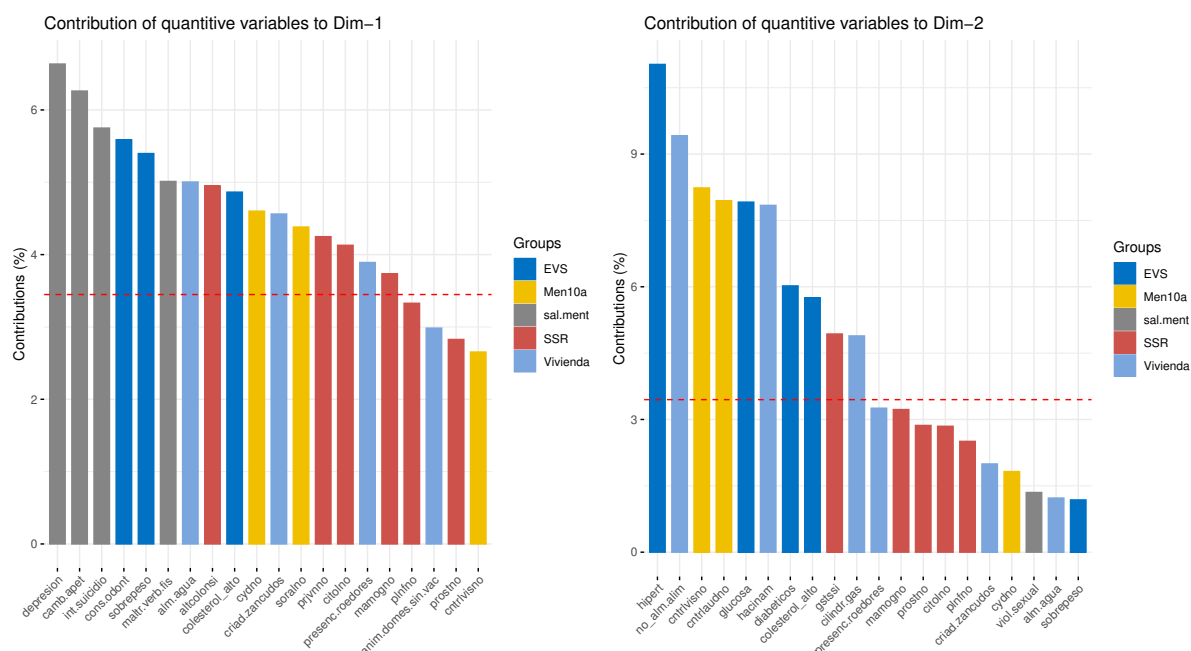


Figura 6-6: Contribución de las variables a la inercia de los ejes factoriales

La contribución de las variables a las dimensiones se muestra en la Figura 6-6, se observa que los grupos de variables que más aportan a la construcción de la primera dimensión son:

- **Salud mental:** depresión, cambios en el apetito, intento de suicidio y maltrato verbal o físico.
- **Estilos de vida saludable:** sobrepeso, inasistencia a consulta odontológica y colesterol alto.
- **Vivienda:** almacenamiento de agua para consumo, criaderos de zancudos y presencia de roedores.
- **Salud sexual y reproductiva - cáncer:** alteraciones del colon, inasistencia a programa joven e inasistencia a la toma de la citología.
- **Menores de 10 años:** inasistencia a los programas de crecimiento y desarrollo y salud oral.

En cuanto la segunda dimensión, se observa que la contribución de las variables del grupo de salud mental no parece ser muy notable. Mientras que para los demás grupos, las variables con mayor contribución a esta dimensión son:

- **Estilos de vida saludable:** hipertensión, glucosa alta, diabetes y colesterol alto.
- **Vivienda:** inadecuado almacenamiento de alimentos y hacinamiento.

- **Menores de 10 años:** inasistencia a los controles de agudeza visual y auditiva.
- **Salud sexual y reproductiva - cáncer:** Gestantes.

Revisando la estructura de cada dimensión y las variables asociadas, no se encuentra una referencia que permita “bautizar” o asignar un nombre al constructo que representan las dimensiones. La primera dimensión esta representada por las variables de los grupos salud mental y salud sexual y reproductiva, mientras que la segunda dimensión la representan las variables de los grupos estilos de vida saludable y Vivienda.

La Tabla **6-8** presentan los dos primeros vectores propios, las coordenadas, las contribuciones y la calidad de la representación de las variables en las dos primeras dimensiones. Se observa que la variable *inadecuado almacenamiento de agua* es la que mayor contribución tiene al primer eje factorial en el grupo de variables *vivienda*, además de ser la mejor representada dentro del grupo. Igualmente sucede con las variables *inasistencia a Crecimiento y Desarrollo*, *deposiciones con sangre*, *inasistencia a consulta odontológica* y *depresión* en sus respectivos grupos.

En cuanto a la segunda dimensión, en el grupo de variables *vivienda* destaca *inadecuado almacenamiento de alimentos* con la mayor contribución y calidad de la representación. *Inasistencia a control de agudeza visual*, *gestantes*, *hipertensión* y *violencia sexual* son las que mayor contribución y mejor calidad de la representación tienen en sus grupos.

De acuerdo con lo observado anteriormente, se deduce que en la población de estudio, la variación se encuentra explicada principalmente en la manifestación de las personas de situaciones de depresión, cambios de apetito, intentos de suicidio, maltrato verbal o físico, hipertensión, sobrepeso, colesterol y glucosa altos, embarazadas, inasistencia a los programas de: crecimiento y desarrollo, controles de agudeza visual y auditiva, toma de citología, mamografía y programa joven, además de viviendas con criaderos de zancudos e inadecuado almacenamiento de agua y alimentos.

Variable	Vectores propios		Coordenadas		Contrib		Calidad de la rep.	
	V_1	V_2	Dim.1	Dim.2	Dim.1	Dim.2	Dim.1	Dim.2
hacinam	0.32	-0.58	0.58	-0.74	2.38	7.84	0.33	0.55
cilindr.gas	0.11	-0.46	0.19	-0.58	0.26	4.89	0.04	0.34
no_alm.alim	0.09	-0.63	0.17	-0.81	0.21	9.41	0.03	0.66
alm.agua	0.46	0.23	0.84	0.29	5.00	1.22	0.70	0.09
anim.domes.sin.vac	0.36	-0.06	0.65	-0.08	2.98	0.09	0.42	0.01
criad.zancudos	0.44	-0.29	0.80	-0.37	4.56	2.00	0.64	0.14
presenc.roedores	0.41	-0.37	0.74	-0.48	3.89	3.25	0.55	0.23
cydno	0.34	-0.22	0.63	-0.28	4.60	1.82	0.39	0.08
cntrlvisno	0.26	0.46	0.48	0.59	2.65	8.23	0.23	0.35
cntrlaudno	0.26	0.45	0.47	0.58	2.60	7.95	0.22	0.34
soralno	0.34	-0.05	0.61	-0.07	4.38	0.10	0.37	0.00
plnfno	0.36	0.32	0.66	0.41	3.33	2.51	0.44	0.16
citolno	0.41	-0.34	0.74	-0.43	4.13	2.85	0.54	0.19
mamogno	0.39	-0.36	0.70	-0.46	3.74	3.22	0.49	0.21
prostno	0.34	-0.34	0.61	-0.43	2.83	2.86	0.37	0.19
altcolonsi	0.44	-0.09	0.81	-0.12	4.95	0.20	0.65	0.01
prjvnno	0.41	0.04	0.75	0.05	4.25	0.04	0.56	0.00
gstssi	0.10	-0.44	0.17	-0.57	0.23	4.93	0.03	0.32
cons.odont	0.43	-0.18	0.79	-0.22	5.59	0.91	0.62	0.05
diabeticos	0.08	0.45	0.14	0.58	0.17	6.02	0.02	0.33
hipert	0.04	0.61	0.08	0.78	0.06	11.03	0.01	0.61
colesterol.alto	0.40	0.44	0.73	0.56	4.86	5.75	0.54	0.32
sobrepeso	0.43	0.20	0.77	0.26	5.40	1.18	0.60	0.07
glucosa	0.29	0.52	0.53	0.66	2.56	7.91	0.28	0.44
camb.apet	0.47	0.11	0.86	0.14	6.26	0.31	0.74	0.02
depresion	0.49	0.00	0.89	0.00	6.63	0.00	0.79	0.00
int.suicidio	0.45	-0.18	0.83	-0.24	5.75	0.94	0.68	0.06
maltr.verb.fis	0.42	0.20	0.77	0.26	5.01	1.16	0.59	0.07
viol.sexual	0.16	0.22	0.30	0.28	0.74	1.35	0.09	0.08

Tabla 6-8: Vector propio, coordenadas, contribuciones y calidad de la representación de las variables

6.2.5. Representación de los individuos

Para la interpretación de los barrios (individuos) se tendrá en cuenta que aquellos que tengan perfiles similares están cerca unos de otros en el mapa de factores. Como una guía para interpretar este plano factorial, recordaremos que la primera dimensión se asocia con los riesgos de salud mental y salud sexual y reproductiva en los territorios, mientras que la segunda dimensión se asocia principalmente con los riesgos de estilos de vida saludable y vivienda. La Figura 6-7 presenta el primer plano factorial y la Figura 6-8 muestra la contribución de los barrios en la construcción de éste primer plano factorial.

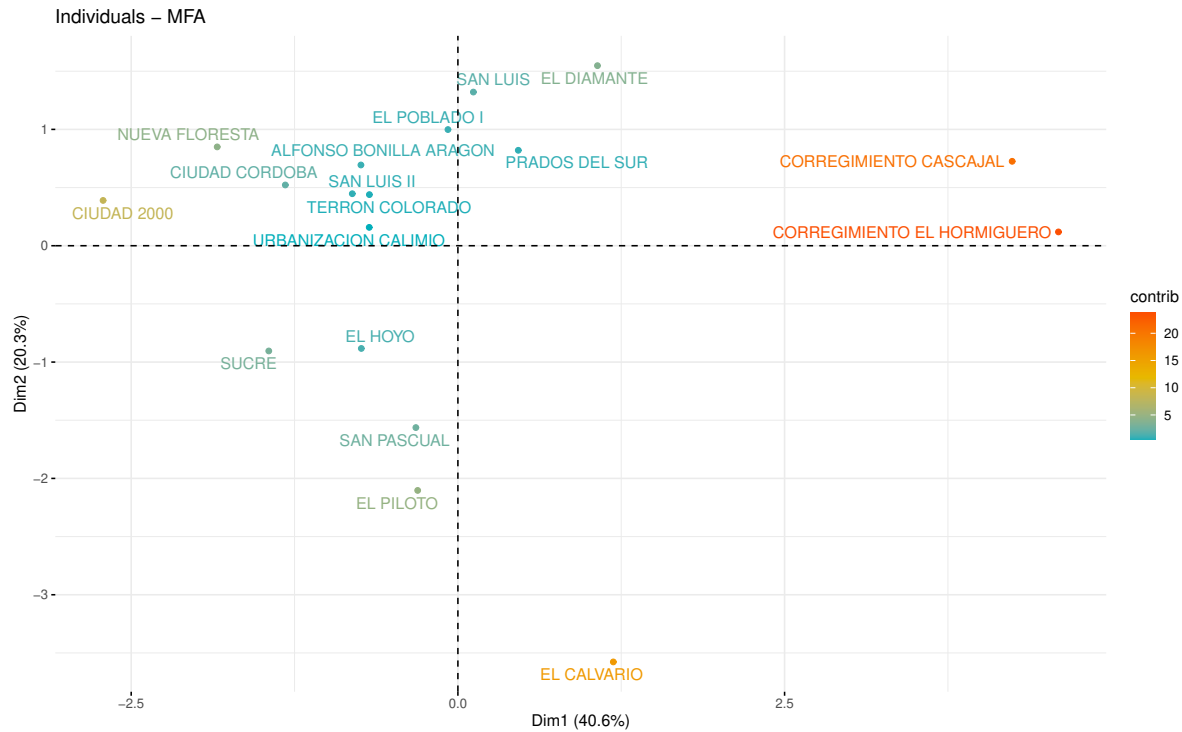


Figura 6-7: Representación de los territorios en el análisis global

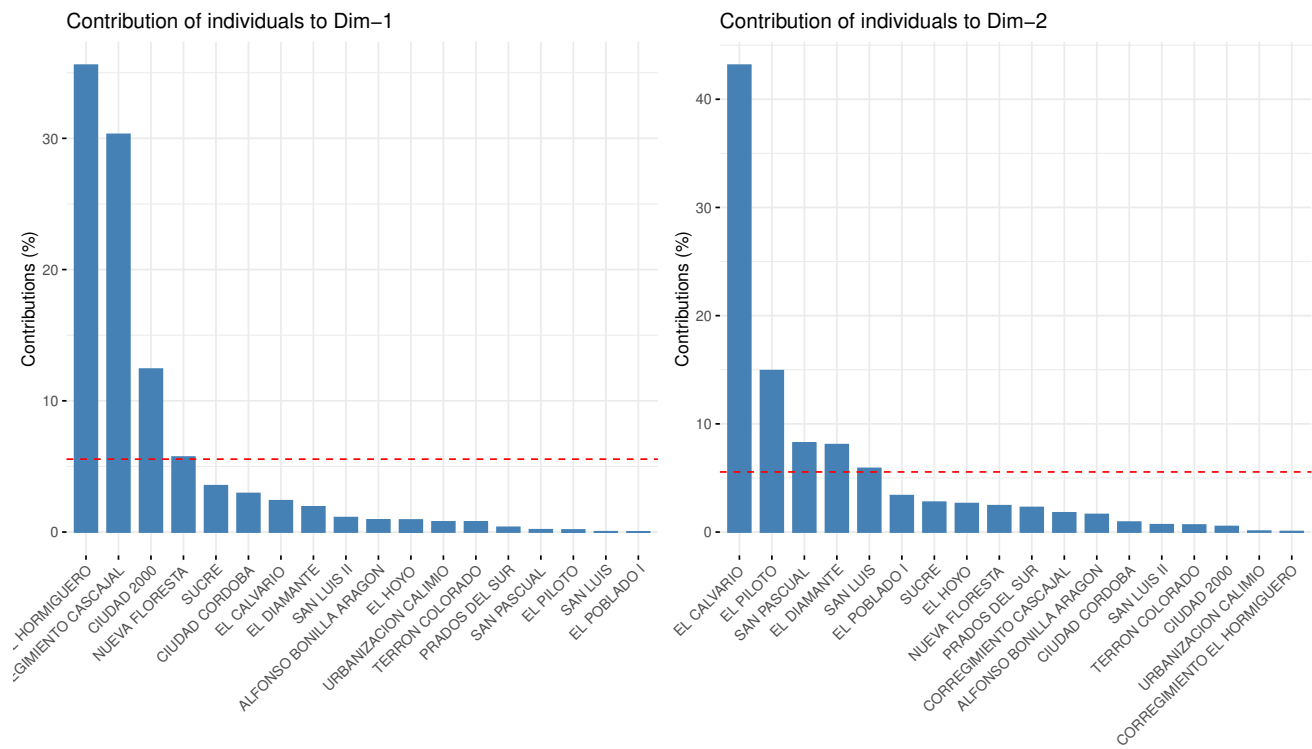


Figura 6-8: Contribución de los territorios a las componentes

De acuerdo a la Figura 6-8, los corregimientos de El Hormiguero, Cascajal y el barrio Ciudad 2000 son los barrios que presentan mayor contribución a la conformación de la primera dimensión. Particularmente estos barrios son los más atípicos respecto a salud mental y sexual y reproductiva, con la diferencia de que Ciudad 2000 es un barrio de bajo perfil de riesgo, mientras que El Hormiguero y Cascajal son de alto perfil de riesgo. Los corregimientos de El Hormiguero y Cascajal presentan en mayor porcentaje situaciones de cambio de apetito, depresión, intento de suicidio, embarazadas, deposiciones con sangre e inasistencia a programas de: planificación familiar, toma de citología, examen de próstata y programa joven. El barrio Ciudad 2000 presenta las frecuencias relativas más bajas para estas variables en el conjunto de datos.

En cuanto a la dimensión dos, los barrios El Calvario, El Piloto, San Pascual, El Diamante y San Luis son los barrios que más contribuyen a la conformación de esta dimensión. Los barrios El Diamante y San Luis se caracterizan por presentar las frecuencias relativas más altas en cuanto a hipertensión, sobrepeso, colesterol y glucosa alta, los barrios El Calvario, El Piloto y San Pascual presentan los mayores riesgos en términos de vivienda y vacunación incompleta para la edad.

Los barrios Nueva Floresta, Sucre, Ciudad Córdoba, Terrón Colorado, Prados del Sur, San Luis II, Alfonso Bonilla Aragón, El Hoyo y Urbanización Calimio son barrios de perfil promedio en las primeras dos dimensiones.

6.2.6. Representación superpuesta de grupos e individuos

La Figura 6-7 es una representación global del perfil de riesgo de todos los territorios, dicha representación tiene en cuenta el aporte de todos los grupos de variables. Sin embargo, los barrios pueden tener comportamientos diferenciales dentro de cada grupo de variables. Para observar el comportamiento de los barrios respecto a los grupos, se opta por realizar la representación superpuesta. En dicha representación se presenta la inercia general del barrio y su comportamiento particular respecto a cada uno de los grupos de variables.

Para la interpretación de la representación superpuesta se seleccionan los barrios que en la Figura 6-7 se destacaron como los más característicos en las primeras dos dimensiones. El corregimiento el Hormiguero ha sido percibido como el más riesgoso desde el punto de la salud mental y vivienda. Por el contrario, para el mismo grupo de variables y según este mismo criterio, el barrio Ciudad 2000 ha sido percibido como el menos riesgoso.

La Figura 6-9 muestra que el barrio El Calvario presenta las condiciones más extremas en cuanto a condiciones de vivienda (es el barrio con mayor porcentaje de hacinamiento, inadecuado almacenamiento de alimentos, uso de cilindros de gas. Además presenta valores

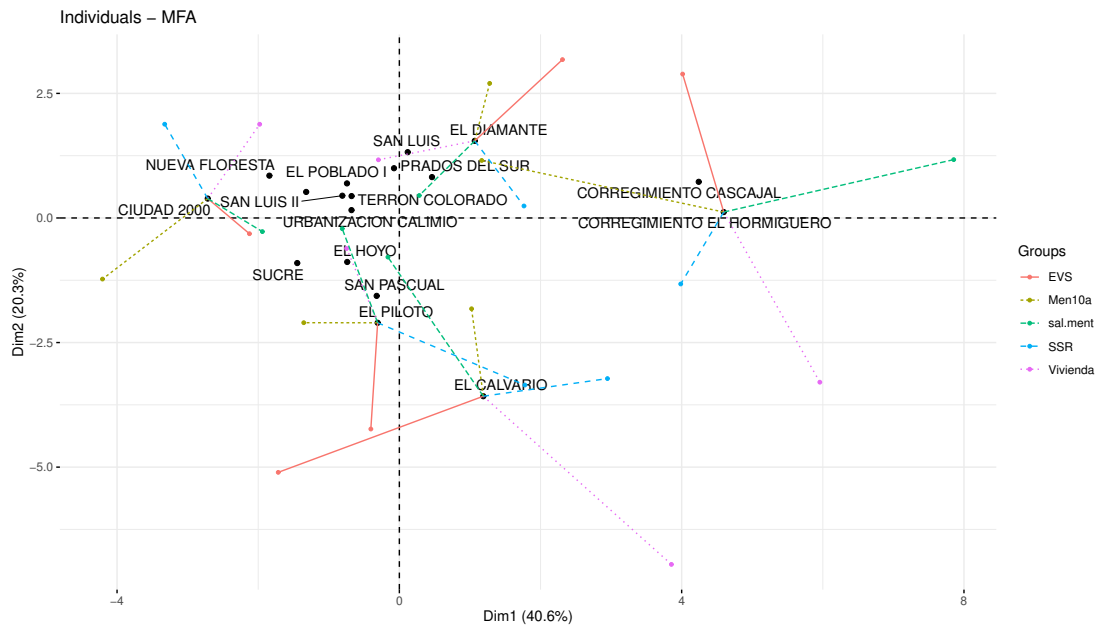


Figura 6-9: Representación superpuesta de los barrios El Diamante, Ciudad 2000, El Piloto, El Calvario y el corregimiento El Hormiguero

altos en cuanto a porcentaje de animales domésticos sin vacunar, porcentaje de criaderos de zancudos y porcentaje de presencia de roedores). Mientras que en el grupo de variables de estilos de vida saludable presenta frecuencias relativas bajas en las variables porcentaje de hipertensos, porcentaje de diabéticos, porcentaje de personas con colesterol alto, porcentaje de personas con sobrepeso y porcentaje de personas con valores de glucosa altos.

Para el grupo de variables asociadas a salud sexual y reproductiva, este barrio presenta los valores más altos en cuanto a porcentaje de mujeres mayores de 50 años sin toma de mamografía, porcentaje de hombres mayores de 50 años sin toma de examen de próstata, además, es uno de los barrios con mayor porcentaje de inasistencia a la toma de la citología y al programa joven. En cambio, toma valores más promedio en el grupo de variables asociado a los menores de 10 años.

El barrio El Diamante está más influenciado por los grupos estilos de vida saludable y menores de 10 años, es decir, que es el barrio con las condiciones de salubridad más desfavorables en cuanto a variables como porcentaje de diabéticos, porcentaje de personas con valores de glucosa alta, porcentaje de inasistencia a control de agudeza visual y auditiva, mientras que para los grupos de salud sexual y reproductiva, salud mental y vivienda toma valores promedio.

Anteriormente se mencionó que el barrio Ciudad 2000 es un barrio de bajo perfil de riesgo, esto se observa en la dirección del aporte del grupo de variables de menores de 10 años. Los

grupos de variables estilos de vida saludable y salud mental toman valores promedio.

6.2.7. Proyección de las componentes principales de cada grupo

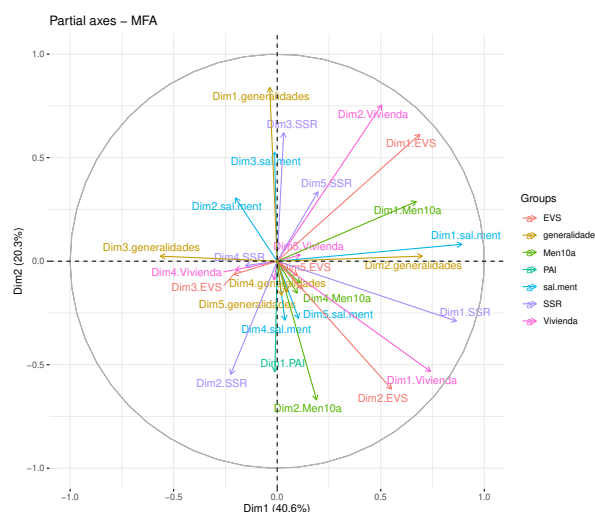


Figura 6-10: Contribución de los grupos a las dimensiones

Como se puede observar en la Figura 6-10, la Dimensión 1 global del AFM tiene una alta correlación con la primera dimensión de cada uno de los grupos de variables. Esto significa que las estructuras de mayor variabilidad de cada uno de los grupos de variables también están reflejadas en la estructura de mayor variabilidad del AFM global.

Respecto a la Dimensión 2 del AFM global, esta refleja principalmente la estructura de variabilidad de los grupos de variables vivienda, menores de 10 años y salud sexual y reproductiva.

A pesar de que al interior de las tablas existen otras estructuras adicionales de variación, se identifica en el plano de contribución de las dimensiones que esas otras estructuras (Dim.3, Dim.4 y Dim.5) no tienen mayor relación con las dos estructuras principales del AFM.

6.2.8. Formulación de un índice sintético

Luego de realizar el AFM general, los resultados de éste análisis serán empleados para construir un índice sintético. Como se pudo observar en la Tabla 6-8, los coeficientes del primer vector propio son todos positivos, lo cual significa que existe un factor tamaño que permite construir un índice de riesgo basado en la estrategia de Atención Primaria en Salud (APS), que corresponde a la primera componente principal. La primera componente del análisis global, se calcula a través de la siguiente expresión

$$\begin{aligned}
F_1 = & 0.31hacinam + 0.10cilindr.gas + 0.09no_alm.alim + 0.46alm.agua \\
& + 0.35anim.domes.sin.vac + 0.44criad.zancudos + 0.40presenc.roedores + 0.34cydno \\
& + 0.26cntrlvino + 0.25cntrlaudno + 0.33soralno + 0.36plnfno + 0.40citolno \\
& + 0.38mamogno + 0.33prostno + 0.44altcolonsi + 0.41prjvnno + 0.09gstssi \\
& + 0.43cons.odont + 0.07diabeticos + 0.04hipert + 0.40colesterol_alto + 0.42sobrepeso \\
& + 0.29glucosa + 0.47camb.apet + 0.48depresion + 0.45int.suicidio + 0.42maltr.verb.fis \\
& + 0.16viol.sexual
\end{aligned}$$

Para llevar a cabo el método de [Avella et al. \(2010\)](#) se deben considerar los siguientes elementos:

Inicialmente, se deben calcular medidas resumen tales como el promedio y la desviación estándar a las variables activas con el fin de estandarizarlas, una vez estandarizadas, se encuentra el valor mínimo de cada una de ellas. Luego, el primer vector propio del análisis global $\hat{u}_{jk}$ es el peso de las variables. Con los anteriores elementos se calcula el puntaje-factor $P_{kj} = |Z_k - \min(Z_k)|\hat{u}_{jk}$ y se extraen los valores máximos para cada nueva variable o puntaje-factor. La Tabla **6-9** presenta estos elementos.

Variable	$\bar{X}_{kj}$	$\sigma_{X_{kj}}$	$\min(Z_{kj})$	$\max(P_{kj})$	Variable	$\bar{X}_{kj}$	$\sigma_{X_{kj}}$	$\min(Z_{kj})$	$\max(P_{kj})$
hacinam	0.03	0.04	-0.88	1.24	altcolonsi	0.01	0.01	-1.22	1.51
cilindr.gas	0.13	0.14	-0.95	0.34	prjvnno	0.52	0.19	-1.70	1.46
no_alm.alim	0.09	0.19	-0.49	0.37	gstssi	0.02	0.03	-0.81	0.41
alm.agua	0.08	0.11	-0.73	1.50	cons.odont	0.18	0.08	-1.54	1.60
anim.domes.sin.vac	0.22	0.19	-1.15	1.23	diabeticos	0.07	0.02	-1.47	0.29
criad.zancudos	0.05	0.09	-0.54	1.51	hipert	0.16	0.05	-2.03	0.16
presenc.roedores	0.08	0.12	-0.64	1.35	colesterol_alto	0.06	0.04	-1.28	1.43
cydno	0.13	0.05	-2.37	1.44	sobrepeso	0.09	0.07	-1.09	1.45
cntrlvino	0.01	0.00	-1.31	0.83	glucosa	0.05	0.03	-1.38	1.17
cntrlaudno	0.01	0.00	-1.36	0.74	camb.apet	0.03	0.04	-0.67	1.65
soralno	0.12	0.05	-1.86	1.37	depresion	0.04	0.06	-0.70	1.70
plnfno	0.16	0.07	-1.64	1.23	int.suicidio	0.01	0.01	-0.78	1.66
citolno	0.20	0.10	-1.55	1.31	maltr.verb.fis	0.00	0.00	-0.58	1.36
mamogno	0.15	0.10	-1.28	1.36	viol.sexual	0.00	0.00	-0.66	0.49
prostno	0.20	0.14	-1.49	1.30					

Tabla 6-9: Elementos para el cálculo del índice sintético

Para este caso, P_{kj} es una matriz de 18 filas y 29 columnas. El puntaje-factor máximo es $sumax_J = 33.46068$. Con él se calculan los puntajes de cada variable de la forma $p_{kj} = P_{kj}/sumax_J$. Finalmente, el valor del índice para cada territorio se obtiene mediante la suma por fila de p_{kj} . Dicho calculo se obtiene mediante $IR_{APS}(i) = \sum_{j \in J, k \in K_j} p_{kj}$

Puesto que las modalidades de las variables se seleccionaron de tal forma que midieran un posible riesgo en salud, se construye un índice sintético que permita evaluar el “riesgo auto-reportado” en cada uno de los barrios a partir de la metodología que propuso [Avella et al. \(2010\)](#). Por ende, este mostrará que a mayor valor, mayor es el riesgo identificado mediante la estrategia de Atención Primaria en Salud en el barrio. Además, se adiciona la distribución porcentual del índice generado.

En la Tabla **6-10** se presenta el calculo de la primera dimensión del AFM global para cada uno de los barrios (F_1 APS), la transformación de ella en un índice sintético (IR_{APS}) y normalización del mismo en escala porcentual ($IR_{APS}(\%)$). La lectura de los tres índices de la Tabla **6-10** es equivalente, los mayores valores se asocian a mayor riesgo y los menores a menor riesgo.

ESE	Territorio	F_1 APS	IR_{APS}	$IR_{APS}(\%)$
Ladera	Cto. Cascajal	4.24	76.53	12.97
	Cto. El Hormiguero	4.60	82.74	14.02
	Terrón Colorado	-0.68	25.23	4.28
	Prados del Sur	0.46	36.20	6.14
	El Calvario	1.19	46.97	7.96
	San Pascual	-0.32	30.83	5.22
Oriente	Alfonso Bonilla Aragón	-0.74	24.02	4.07
	Ciudad Córdoba	-1.32	18.33	3.11
	El Diamante	1.07	43.17	7.32
	El Poblado I	-0.08	31.37	5.32
Centro	El Hoyo	-0.74	25.36	4.30
	El Piloto	-0.31	30.55	5.18
	Nueva Floresta	-1.84	13.20	2.24
	Sucre	-1.45	17.12	2.90
Norte	San Luis	0.12	32.04	5.43
	San Luis II	-0.81	25.60	4.34
	Urbanización Calimio	-0.68	25.17	4.26
Surorient	Ciudad 2000	-2.72	5.68	0.96

Tabla 6-10: Índice sintético para la estrategia de Atención Primaria en Salud en Cali



Figura 6-11: Índice APS expresado en porcentaje

La Figura **6-11** muestra la representación porcentual del índice sintético construido, se observa que los corregimientos de El Hormiguero y Cascajal presentaron los valores más altos, indicando que son los barrios donde se encuentra la población más vulnerable desde la perspectiva de la estrategia de Atención Primaria en Salud para el año 2018, situación contraria con los barrios Ciudad 2000 y Nueva Floresta que obtuvieron valores del índice más bajos, mostrando una mejor condición de salud sentida en sus habitantes.

Se debe recordar que los barrios con los que se realizó este estudio fueron priorizados por parte de la Secretaría de Salud Pública Municipal de Santiago de Cali, que no refleja la situación de toda la población de la ciudad.

Capítulo 7

Conclusiones y recomendaciones

7.1. Conclusiones

El presente trabajo se ha desarrollado con los datos contenidos en la encuesta realizada por la Secretaría de Salud Pública Municipal de Santiago de Cali en el marco del proyecto *“Implementación de la estrategia de Atención Primaria en Salud en el municipio de Cali”* en el año 2018. La encuesta esta dividida en 7 grupos de variables, los cuales fueron consolidados a nivel de barrio, obteniendo indicadores de riesgo de enfermedad para cada variable en cada barrio. Los datos consolidados fueron analizados a través de un Análisis Factorial Múltiple (AFM), a partir del cual se logro identificar que las dos primeras dimensiones explican el 60.88 % de la variabilidad de la tabla de datos. La primera dimensión es explicada por los grupos de variables salud mental y salud sexual y reproductiva mientras que la segunda dimensión es explicada por los grupos de estilos de vida saludable y vivienda.

De forma general se evidencia que la mayor variabilidad en las condiciones de salud de la población se presenta en situaciones como depresión, cambios de apetito, intentos de suicidio, maltrato verbal o físico, hipertensión, sobrepeso, colesterol y glucosa altos, embarazadas, inasistencia a los programas de: crecimiento y desarrollo, controles de agudeza visual y auditiva, toma de citología, mamografía y programa joven, además de viviendas con criaderos de zancudos e inadecuado almacenamiento de agua y alimentos.

Los corregimientos El Hormiguero, Cascajal y el barrio Ciudad 2000 son los barrios que presentan mayor contribución a la conformación de la primera dimensión, es decir que son los barrios más atípicos respecto a salud mental y salud sexual y reproductiva. En contraste, el barrio Ciudad 2000 es un barrio de bajo perfil de riesgo de salud general, mientras que los mayores riesgos de salud se observan en los corregimientos El Hormiguero y Cascajal en cuanto a la salud sentida respecto de este par de grupos de variables.

Adicionalmente se construyó el índice sintético denominado “índice de riesgo basado en Aten-

ción Primaria en Salud IR_{APS} ” que facilitó la priorización de barrios respecto a su situación de riesgo. En esta priorización, se concluye que los barrios con mayores niveles de riesgo fueron los corregimientos de El Hormiguero y Cascajal, mientras que los barrios con menor nivel de riesgo son Ciudad 2000 y Nueva Floresta.

A pesar de que el presente trabajo ha sido desarrollado bajo un caso de estudio particular, la metodología que aquí se presenta puede replicarse en futuras aplicaciones de la misma encuesta sobre otro conjunto de barrios priorizados o incluso extender la construcción del índice para evaluar un barrio que no haya sido incluido en este listado de 18 barrios priorizados. Por ejemplo, un barrio que se evalúa posteriormente, puede ser proyectado en el AFM y luego transformar su Dimensión 1 para obtener el índice de riesgo para este nuevo barrio.

En este punto vale la pena recordar que los resultados que aquí se han obtenido son resultados de riesgo relativo, ya que en primera instancia las variables no miden la condición de salud real sino las condiciones de salud sentidas evidenciadas por los mismos habitantes, lo cual es una evaluación subjetiva de la salud pública en la ciudad. Adicionalmente, el indicador de riesgo que se ha construido solo ha tenido como referencia los 18 barrios que fueron priorizados, es probable que la evaluación de toda la ciudad el índice pueda cambiar.

7.2. Recomendaciones

En el desarrollo de este trabajo se encontraron algunas limitaciones sobre la calidad de los datos. En este sentido, se recomienda mejorar la captura de información, pues los datos encontrados en este trabajo con relación a estadísticas generales como el número de personas por familia, son muy distantes a la información publicada por el Departamento Administrativo de Planeación Municipal (DAPM) de la Alcaldía de Santiago de Cali.

Adicionalmente, en la base de datos no se registraron muchos casos de personas diagnosticadas con tuberculosis, contrario a lo reportado por [López Pérez \(2017\)](#), donde reportan 1188 casos de tuberculosis en Cali durante 2017. Esto conlleva a mejorar las capacitaciones que se le brinda al personal que realiza la encuesta sobre este tema, con el fin de identificar correctamente los casos. Lo anterior implicó no incluir el grupo de variables del módulo *micobacterias* en el AFM.

También se encontró que no todas las preguntas para los mayores de 10 años fueron realizadas, pues el módulo de salud mental tenía menos registros que los demás que le aplican a esta población. Además, se sugiere incluir la variable nacionalidad, ya que en el campo *observaciones* de la *ficha* se encontró población venezolana que no puede acceder a todos los servicios de salud debido a su estado de afiliación. Finalmente, se debe mencionar que otras alternativas para realizar los análisis, pueden ser el Análisis Factorial Múltiple para Tablas de Contingencia (AFMTC) propuesto por [Bécue-Bertaut y Pagès \(2004\)](#) o STATIS.

Referencias

- Abascal, E., y Calvo, M. L. (2002). Análisis factorial múltiple como técnica de estudio de la estabilidad de los resultados de un análisis de componentes principales. *Qüestió: quaderns d'estadística i investigació operativa*, 26(1), 109–122.
- Abdi, H., Williams, L. J., y Valentin, D. (2013). Multiple factor analysis: principal component analysis for multitable and multiblock data sets. *Wiley Interdisciplinary reviews: computational statistics*, 5(2), 149–179.
- Ahmed, S., Chowdhury, M. A. H., Khan, M. A., Huq, N. L., y Naheed, A. (2017). Access to primary health care for acute vascular events in rural low income settings: a mixed methods study. *BMC health services research*, 17(1), 47–53. Descargado de <https://doi.org/10.1186/s12913-017-1987-8> doi: 10.1186/s12913-017-1987-8
- Alfaqeeh, G., Cook, E. J., Randhawa, G., y Ali, N. (2017). Access and utilisation of primary health care services comparing urban and rural areas of riyyadh providence, kingdom of saudi arabia. *BMC health services research*, 17(1), 106–118. Descargado de <https://doi.org/10.1186/s12913-017-1983-z> doi: 10.1186/s12913-017-1983-z
- Aluja, B., Morineau, A., y cols. (1999). *Aprender de los datos: el análisis de componentes principales: una aproximación desde el data mining* (1.^a ed.). Flor del Viento Ediciones S.A.
- Angulo, E. (2012). *Atención primaria en salud: un camino hacia la equidad* (Reporte técnico). Secretaría Distrital de Salud. Alcaldía de Bogotá.
- Avella, B., Oliva, M., y cols. (2010). *Comparación del análisis factorial múltiple (afm) y del análisis en componentes principales para datos cualitativos (prinqual), en la construcción de índices* (Tesis Doctoral no publicada). Universidad Nacional de Colombia.
- Bécue-Bertaut, M., y Pagès, J. (2004). A principal axes method for comparing contingency tables: Mfact. *Computational Statistics & Data Analysis*, 45(3), 481–503.
- Carrasco, S., y Arredondo, J. O. (2006). Aportación del afm al análisis de las políticas públicas. una visión de américa latina y el caribe, 1990-2002. *Estudios Económicos de Desarrollo Internacional*, 6(2), 92–108.

- Carstairs, V. (2001). Factores socioeconómicos a nivel de área y su relación con la salud espacial. *Epidemiology*, 1(9), 51–68.
- Carstairs, V., y Morris, R. (1990). Deprivation and health in scotland. *Health bulletin*, 48(4), 162–175.
- Castaño, E. (2004). Elementos para una metodología del estudio de la calidad de vida. *La Sociología en sus Escenarios*(9). Descargado de <https://revistas.udea.edu.co/index.php/ceo/article/view/1665>
- Castaño, E., Correa, C., y Salazar, B. (1998). La construcción de un indicador de calidad de vida para la ciudad de medellín. *Misión Social*.
- Cattell, R. B. (1965). Factor analysis: An introduction to essentials i. the purpose and underlying models. *Biometrics*, 21(1), 190–215.
- Copplestone, J. F. (1991). Qué es la salud? *Foro mundial de la salud 1991*, 12(4), 455–457.
- Corrales, M. L. (2010). Combinación del análisis factorial múltiple y el análisis armónico cualitativo en la descripción de datos longitudinales cualitativos [Tesis de Maestría]. *Departamento de Estadística. Universidad Nacional de Colombia*.
- Cueto, M. (2004). The origins of primary health care and selective primary health care. *American journal of public health*, 94(11), 1864–1874.
- DAPM, C. (2017). *Cali en cifras 2017* (Reporte técnico). Departamento Administrativo de Planeación Municipal. Alcaldía de Santiago de Cali. Descargado de <https://www.cali.gov.co/planeacion/publicaciones/137803/documentos-cali-en-cifras>
- Díaz, L., y Rivera, M. A. M. (2012). *Análisis estadístico de datos multivariados* (1.^a ed.). Universidad Nacional de Colombia.
- DNP. (2000). *Actualización del modelo de jerarquización y categorización de los departamentos y municipios año 2000*. (Reporte técnico). Departamento Nacional de Planeación, Subdirección Territorial y de Inversión Pública Dirección de Inversiones y Finanzas Públicas.
- DNP. (2015). *Modelo de jerarquización y categorización de los departamentos y municipios* (Reporte técnico). Departamento Nacional de Planeación, Subdirección Territorial y de Inversión Pública Dirección de Inversiones y Finanzas Públicas.
- Escoufier, B., y Pagès, J. (1992). Análisis factoriales simples y múltiples: objetivos, métodos e interpretación. *España: Universidad del País Vasco*.
- Escoufier, Y. (1973). Le traitement des variables vectorielles. *Biometrics*, 10(29), 751–760.

- González, V. M., Ávila, E., y Olaya, J. (2009). Análisis de los espacios de representación del statis y del afm en el estudio de movilidad biográfica en bogotá 1993. *Revista Colombiana de Estadística*, 32(1), 1-15.
- Gonze, D. (2007). *Principal components analysis* [Tutorial].
- Hermida, C. (2016). Origen de la atención primaria de salud. *El Tiempo, Diario de Cuenca, Ecuador*. Descargado de <https://www.eltiempo.com.ec/noticias/columnistas/1/origen-de-la-atencion-primaria-de-salud>
- Hutchison, B., Levesque, J.-F., Strumpf, E., y Coyle, N. (2011). Primary health care in canada: systems in motion. *The Milbank Quarterly*, 89(2), 256–288.
- Ivaldi, E., Bonatti, G., y Soliani, R. (2016). The construction of a synthetic index comparing multidimensional well-being in the european union. *Social Indicators Research*, 125(2), 397–430.
- Jeffreys, M., Lopez, M. I., Russell, L., Smiler, K., Ellison-Loschmann, L., Thomson, M., ... others (2020). Equity in access to zero-fees and low-cost primary health care in aotearoa new zealand: Results from repeated waves of the new zealand health survey, 1996-2016. *Health Policy*, 124(11), 1272–1279.
- Jolliffe, I. T., y Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065). Descargado de <http://dx.doi.org/10.1098/rsta.2015.0202>
- Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3), 187–200.
- Kroeger, A., y Luna, R. (1987). *Atención primaria de salud: principios y métodos*. Organización Panamericana de la Salud.
- Kuhfeld, W. (1990). *Algorithms for the prinqual and transreg procedures* (Reporte técnico). SAS technical report R-108, SAS Institute, Cary, NC.
- Kuhfeld, W. F., Sarle, W. S., y Young, F. W. (1985). Methods of generating model estimates in the prinqual macro. En *Sas users group international conference proceedings: Sugi* (Vol. 10, pp. 962–971).
- Labonté, R., Sanders, D., Baum, F., Schaay, N., Packer, C., Laplante, D., ... others (2009). Aplicación, efectividad y contexto político de la atención primaria integral de salud: resultados preliminares de una revisión de la literatura mundial. *Revista Gerencia y Políticas de Salud*, 8(16), 14–29.

- Lalonde, M. (1974). *A new perspective on the health of Canadians: a working document*. Ottawa.
- Lebart, L., PIRON, A., Lebart, M., Morineau, A., y Piron, M. (2000). *Statistique exploratoire multidimensionnelle*. Editorial Dunod.
- López Pérez, M. P. (2017). *Informe de evento tuberculosis, colombia* (Reporte técnico). Instituto Nacional de Salud.
- Michailidis, G., y de Leeuw, J. (1998). The gif system of descriptive multivariate analysis. *Statistical Science*, 13(4), 307–336. Descargado de <https://www.jstor.org/stable/2676814>
- MSPS. (2011). *Ley 1438 de 2011*. Ministerio de Salud y Protección Social. República de Colombia.
- MSPS. (2014). *Guía conceptual y metodológica para la construcción del asis de las entidades territoriales* (Reporte técnico). Ministerio de Salud y Protección Social. República de Colombia.
- OMS. (1978). Declaración de alma-ata.
- OPS. (2007). *Renovación de la atención primaria de salud en las américas: documento de posición de la organización panamericana de salud/organización mundial de la salud (ops/oms)*. Washington, DC: OPS.
- 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*.
- Ramosa, M. C. M., y Morinsonb, M. V. (2017). Análisis factorial múltiple: una mirada a la calidad de vida de la población infantil en la región caribe colombiana. *Revista Colombiana de Estadística*.
- Rasella, D., Harhay, M. O., Pamponet, M. L., Aquino, R., y Barreto, M. L. (2014). Impact of primary health care on mortality from heart and cerebrovascular diseases in brazil: a nationwide analysis of longitudinal data. *British Medical Journal Publishing Group*, 349. Descargado de <https://doi.org/10.1136/bmj.g4014>
- Rodríguez, L., Acosta, N., y Ruiz, M. (2013). Evaluación del desempeño de servicios de atención primaria en salud: experiencia en municipios rurales en santander, colombia. *Revista de Salud Pública*, 15, 183–195.

- Saporta, G. (1983). *Multidimensional data analysis and quantification of categorical variables*. North-Holland. Descargado de <https://hal-cnam.archives-ouvertes.fr/hal-02514149>
- Segredo, A. M. (2009). Caracterización del sistema de dirección en la atención primaria de salud. *Revista Cubana de Salud Pública*, 35(4), 78–109.
- Soto, H., y Schuschny, A. R. (2009). *Guía metodológica: diseño de indicadores compuestos de desarrollo sostenible* (Documento de proyecto). Cepal. Descargado de <https://repositorio.cepal.org/handle/11362/3661>
- SSPM. (2017). *Implementación de la estrategia de atención primaria en salud (APS) en la zona urbana y rural del municipio de Santiago de Cali* (Reporte técnico). Secretaria de Salud Pública de Santiago de Cali.
- Tenenhaus, M. (1977). Analyse en composantes principales d'un ensemble de variables nominales ou numériques. *Revue de statistique appliquée*, 25(2), 39–56.
- Tuesca-Molina, R. d. J., Segura Barrios, I., Bruno Rubio, V., Jiménez Hamburgén, A., Maldonado Mendoza, L., y cols. (2015). Atención primaria en salud. una mirada desde los profesionales de enfermería: Barreras, conocimientos y actividades. barranquilla-colombia. *Revista Científica Salud Uninorte*, 31(2).
- Van de Geer, J. P. (1993). *Multivariate analysis of categorical data* (Vol. 2). Sage.
- Vega, Acosta, Mosquera, y Restrepo. (2009). Atención primaria integral de salud. estrategia para la transformación del sistema de salud y el logro de la equidad en salud. Bogotá: Secretaría Distrital de Salud, Pontificia Universidad Javeriana.
- Young, F. W. (1981). Quantitative analysis of qualitative data. *Psychometrika*, 46(4), 357–388.
- Young, F. W., Takane, Y., y de Leeuw, J. (1978). The principal components of mixed measurement level multivariate data: An alternating least squares method with optimal scaling features. *Psychometrika*, 43(2), 279–281.
- Young, F. W., Takane, Y., y de Leeuw, J. (1985). Proc prinqual-preliminary specifications. *non published manuscript, The University of North Carolina Psychometric Laboratory, Chapel Hill NC*.

Apéndice A

Anexo A: Formulario de Encuesta

 ALCALDÍA DE SANTIAGO DE CALI SECRETARÍA DE SALUD SERVICIO DE SALUD PÚBLICA	SISTEMAS DE GESTIÓN Y CONTROL INTEGRADOS (SISTEDA, SGC y MECI) ESTRATEGIA DE ATENCIÓN PRIMARIA EN SALUD FICHA DE DIAGNÓSTICO Y CARACTERIZACIÓN DE FAMILIAS	REVISIÓN	
		VERSIÓN	
		FECHA DE ENTRADA EN VIGENCIA	


FICHA DE DIAGNÓSTICO Y CARACTERIZACIÓN DE FAMILIAS
DIGITADO

INFORMACIÓN GENERAL									
Código de Vivienda			Nombre de quien realiza la visita						
No. de Familias		Total de Personas		Empresa Social del Estado					
Fecha Visita	Día	Mes	Año	Persona que atiende la visita					
Barrio o Sector									
Dirección						Teléfono fijo	Celular		
¿Tiene algún familiar en condición de habitante de calle?				Si	No	Nombre completo:			
Tipo de ID	No. Documento ID								

No.	Nombre 1	Nombre 2	Apellido 1	Apellido 2	Fecha de Nacimiento			Edad	Familia No.
					Día	Mes	Año		
1									
2									
3									
4									
5									
6									
7									
8									
9									
10									

Observaciones:

--

Firma de quien atiende la visita	
----------------------------------	--

R.C:Registro Civil **T.I:**Tarjeta de Identidad **C.C:**Cédula de Ciudadanía **C.E:**Cédula de Extranjería **I.D:**Numero único de identificación personal
A: Adulto sin identificación **M:** Menor sin identificación

 ALCALDÍA DE SANTIAGO DE CALI DESARROLLO SOCIAL SERVICIO DE SALUD PÚBLICA	SISTEMAS DE GESTIÓN Y CONTROL INTEGRADOS (SISTEDA, SGC y MECI) ESTRATEGIA DE ATENCIÓN PRIMARIA EN SALUD FICHA DE DIAGNÓSTICO Y CARACTERIZACIÓN DE FAMILIAS	REVISION	
		VERSIÓN	
		FECHA DE ENTRADA EN VIGENCIA	



CONDICIONES DEL ENTORNO											
1. ¿Hay actividad económica en la vivienda?		Si ☐ No ☐		¿Cuál?							
2. ¿Cuántas personas de la vivienda participan en la actividad económica?											
3. ¿Cuánto tiempo se ha desarrollado la actividad económica en la vivienda?				Menos de 1 año ☐		1-4 años ☐		5-9 años ☐		Mas de 10 años ☐	
4. ¿Hay hacinamiento en la vivienda?		Si ☐ No ☐		Especifique:		Duermen más de 3 personas en una habitación ☐		Duermen padres con hijos ☐			
5. ¿Hay cilindros de gas?		Si ☐ No ☐		Especifique:		Buen estado ☐		Oxidados ☐		Con Fugas ☐	
						Sin Soporte ☐		Intemperie ☐		Cercano a fuentes de calor ☐	
6. ¿La higiene de la vivienda es adecuada?				Si ☐ No ☐							
7. ¿Las condiciones de la vivienda son adecuadas?		Si ☐ No ☐		Especifique:		Poca ventilación ☐		Poca iluminación ☐		Pisos de difícil lavado ☐	
						Paredes con humedades ☐		Techos de material inadecuado ☐			
8. ¿En que condiciones se encuentra la letrina o sanitario?		Buena ☐		Regular ☐		Mala ☐		Especifique		Sin tapa ☐	
										Mal estado ☐	
										Sucio ☐	
9. ¿Realiza separación de residuos?				Si ☐ No ☐							
10. ¿Dónde dispone los residuos sólidos?				Recipiente ☐		Bolsa ☐		Campo abierto ☐		Fuente hídrica ☐	
11. ¿Las superficies de la cocina (mesones, pisos, paredes, techos) son de fácil limpieza y desinfección?								Si ☐ No ☐			
12. ¿Los alimentos se encuentran almacenados en lugar cerrado y apropiado, evitando el acceso a insectos, roedores y otros animales?										Si ☐ No ☐	
13. ¿Tiene acceso permanente al agua potable en la vivienda?		Si ☐ No ☐		14. ¿Hay almacenamiento de agua para consumo humano?		Si ☐ No ☐		Especifique:		15. ¿Hay alcantarillado?	
								Buen Estado ☐		Sin Tapa ☐	
								Sucio ☐			
16. ¿Hay fuentes de contaminación cerca de la vivienda?		Si ☐ No ☐		17. ¿Fuman dentro de la vivienda?		Si ☐ No ☐		18. ¿Hay humo dentro de la vivienda o estufa dentro de la habitación?		Si ☐ No ☐	
19. ¿Hay animales domésticos en la vivienda?		Si ☐ No ☐		Especifique:		Perros ☐		Gatos ☐		Palomas ☐	
20. ¿Los animales domésticos han sido vacunados en los últimos 12 meses? (Antirrábica)		Si ☐ No ☐		Perros ☐		Gatos ☐		21. ¿Los animales domésticos están esterilizados?		Si ☐ No ☐	
										Perros ☐	
										Gatos ☐	
22. ¿Los animales domésticos han sido desparasitados en los últimos 3 meses?										Si ☐ No ☐ N.A ☐	
23. ¿Hay criaderos de zancudos?		Si ☐ No ☐		24. ¿Hay presencia de roedores y/o insectos en la vivienda?						Si ☐ No ☐	

	SISTEMAS DE GESTIÓN Y CONTROL INTEGRADOS (SISTEDA, SGC y MECI) ESTRATEGIA DE ATENCIÓN PRIMARIA EN SALUD FICHA DE DIAGNÓSTICO Y CARACTERIZACIÓN DE FAMILIAS CONDICIONES DE SALUD EN HOMBRES Y MUJERES MAYORES DE 10 AÑOS	REVISIÓN	
		VERSIÓN	
		FECHA DE ENTRADA EN VIGENCIA	


CONDICIONES DE SALUD EN HOMBRES Y MUJERES MAYORES DE 10 AÑOS

Código de Vivienda				Familia No.				Persona No.						
Nombre 1				Nombre 2										
Apellido 1				Apellido 2										
Fecha de nacimiento	Día	Mes	Año	Edad	Sexo	M ☐ F ☐	Tipo ID	No. Doc. ID						
Víctima del conflicto	Si ☐ No ☐	Etnia	Afiliación al SGSSS		Si ☐ No ☐	Régimen	C ☐ S ☐ E ☐	EPS						
Encuestado por el SISBEN	Si ☐ No ☐	Discapacidad	Si ☐ No ☐	Tipo de Discapacidad	1	2	3	4	Tiene registro de localización	Si ☐ No ☐	Género	1	2	
Peso	Talla	IMC	¿Estudia actualmente?	Si ☐ No ☐	Nivel de escolaridad	1	2	3	4	5	6	7	8	9

SALUD SEXUAL Y REPRODUCTIVA - CÁNCER

PREGUNTAS	SI	NO	NA	ESPECIFIQUE	MOTIVO DE INASISTENCIA
1. ¿Asiste a consulta médica de programa joven? (10-29 años)	☒				
2. ¿Esta embarazada? (10 - 54 años)				¿Cuántas semanas tiene?	
3. ¿El embarazo es deseado?					
4. ¿Asiste a programa de Control Prenatal?				Número de controles:	
5. ¿Ha tenido signos de alarma en la gestación? (mareos, vómito, sangrado, zumbido, otros)					
6. ¿Ha tenido embarazos de alto riesgo?					
7. ¿Desea quedar embarazada en los próximos 6 meses? (10 - 54 años)					
8. ¿Tiene hijos?	☒			¿Cuántos?:	
9. ¿Fecha de última menstruación? (10 - 54 años)				Día	Mes
10. ¿Utiliza algún método anticonceptivo? (10 - 54 años)				¿Cuál?:	1 2 3 4 5 6 7
11. ¿En los últimos 12 meses ha tenido control médico o de enfermería para el método que utiliza?					
12. ¿Utiliza el preservativo masculino? (condón)	☒				
13. ¿Le han tomado la citología en los últimos 12 meses (25 - 69 años / <25 con VSA) ?					
14. ¿Le han tomado la prueba de VPH en los últimos 5 años (30 - 65 años)?					
15. ¿Se realiza mensualmente el autoexamen de seno?					
16. ¿Le han tomado la mamografía en los últimos 2 años? (>50 años)					
17. ¿Le han tomado el examen de próstata en los últimos 5 años? (>50 años)	☒				
18. ¿Ha presentado deposiciones con sangre? (>35 años)	☒				

TIPO ID	RC:Registro Civil TI:Tarjeta de Identidad CC:Cédula de Ciudadanía CE:Cédula de Extranjería NUIP:Numero único de identificación personal A: Adulto sin identificación M: Menor sin identificación
ETNIA	1. Indígena 2. Raizal 3. Afrodescendiente 4. Gitano (ROM) 5. Palenquero 6. Otro 7. No se reconoce
TIPO DISCAPACIDAD	1. Física 2. Sensorial 3. Cognitiva 4. Psicosocial o mental
GÉNERO	1. Heterosexual 2. Comunidad LGTBIQ
NIVEL DE ESCOLARIDAD	1. Ninguno 2. Preescolar 3. Primaria Completa 4. Primaria Incompleta 5. Secundaria Completa 6. Secundaria Incompleta 7. Técnico o Tecnólogo 8. Profesional 9. Posgrado
MÉTODO ANTICONCEPTIVO	1. Hormonal Oral 2. Hormonal Inyectable 3. Quirúrgico 4. DIU 5. Natural 6. Implante Subdérmico 7. Barrera
MOTIVO DE INASISTENCIA	1. Desconocimiento 2. Falta de Recursos y/o tiempo 3. Atención no satisfactoria en IPS 4. Problemas de Afiliación 5. Creencias Religiosas o Culturales 6. Desinterés 7. Otro motivo

	SISTEMAS DE GESTIÓN Y CONTROL INTEGRADOS (SISTEDA, SGC y MECI) ESTRATEGIA DE ATENCIÓN PRIMARIA EN SALUD FICHA DE DIAGNÓSTICO Y CARACTERIZACIÓN DE FAMILIAS CONDICIONES DE SALUD EN HOMBRES Y MUJERES MAYORES DE 10 AÑOS	REVISION	
		VERSIÓN	
		FECHA DE ENTRADA EN VIGENCIA	



ESTILOS DE VIDA SALUDABLE

PREGUNTAS	SI	NO	ESPECIFIQUE	MOTIVO DE INASISTENCIA
19. ¿Presenta algún malestar en los dientes o en alguna otra parte de la boca?			¿Cuál?	
20. ¿Ha asistido a consulta odontológica en los últimos 12 meses?				
21. ¿Tiene alguna de estas enfermedades diagnosticadas?			Cáncer ☐ Cardiopatías ☐ VIH ☐ Lupus ☐ Mentales ☐	
22. ¿Es Diabético?			¿En control? Si ☐ No ☐	
23. ¿Es Hipertenso?			¿En control? Si ☐ No ☐	
24. ¿Toma medicación para hipertensión?				
25. ¿Con que frecuencia consume frutas y/o verduras?			¿Todos los días? ☐ ¿No Todos los días? ☐	
26. ¿Le han encontrado valores de glucosa altos?				
27. ¿Le han encontrado valores de colesterol y/o triglicéridos altos?				
28. ¿Realiza al menos 30 minutos diarios de actividad física?				
29. ¿Usted consume algún producto derivado del tabaco? (Cigarrillo, Pipa, Tabaco en polvo, Tabaco para mascar)				
30. ¿Se observan signos de sobrepeso/obesidad?				
31. ¿Le han diagnosticado diabetes a alguno de sus familiares?			Padres ☐ Abuelos ☐ Hermanos ☐ Tíos ☐ Hijos ☐	
32. Perímetro de la Cintura (CM)				
33. Presión Arterial (mmHg)				

SALUD MENTAL

PREGUNTAS	SI	NO	ESPECIFIQUE	MOTIVO DE INASISTENCIA
34. ¿En las últimas 4 semanas ha tenido cambios en su apetito o sueño?				
35. ¿En las últimas 4 semanas se ha sentido triste, tenso(a), aburrido(a) o con poco interés en las cosas que antes disfrutaba?				
36. ¿En las últimas 4 semanas ha pensado en hacerse daño o en acabar con su vida?				
37. ¿En el último año, le ha parecido a su familia, amigos, médico o alguien cercano que usted ha estado bebiendo demasiado?				
38. ¿Siente miedo de que su pareja, novio(a), compañero(a) o expareja lo(la) maltrate o agreda verbal o físicamente?				
39. ¿En el último año ha sido obligado(a) a tener relaciones sexuales, a que toquen su cuerpo, a tocar el cuerpo de otro o a presenciar actos sexuales en contra de su voluntad?				

MICOBACTERIAS

PREGUNTAS	SI	NO	N.A.	ESPECIFIQUE	MOTIVO DE INASISTENCIA
40. ¿Ha estado enfermo de Hansen o Lepra?					
41. ¿Durante los últimos días ha notado alguna mancha en su piel que no le pique que no le duela o que no sienta esa área de la piel?					
42. ¿Tiene Tuberculosis?					
43. ¿Está en Tratamiento?				¿Dónde?	
44. ¿Ha presentado tos por mas de 15 días seguidos en el último mes?					
Si la respuesta anterior es afirmativa, continuar con las siguientes preguntas					
45. ¿Ha tenido pérdida de peso y/o apetito?					
46. ¿Ha tenido fiebre?					
47. ¿Ha tenido sudoración?					

MOTIVO DE INASISTENCIA

1. Desconocimiento 2. Falta de Recursos y/o tiempo 3. Atención no satisfactoria en IPS
4. Problemas de Afiliación 5. Creencias Religiosas o Culturales 6. Desinterés 7. Otro motivo

 ALCALDIA DE SANTIAGO DE CALI DESARROLLO SOCIAL SERVICIO DE SALUD PÚBLICA	SISTEMAS DE GESTIÓN Y CONTROL INTEGRADOS (SISTEDA, SGC y MECI) ESTRATEGIA DE ATENCIÓN PRIMARIA EN SALUD FICHA DE DIAGNÓSTICO Y CARACTERIZACIÓN DE FAMILIAS CONDICIONES DE SALUD EN NIÑOS HASTA LOS 9 AÑOS, 11 MESES, 29 DÍAS	REVISION	
		VERSIÓN	
		FECHA DE ENTRADA EN VIGENCIA	



CONDICIONES DE SALUD EN NIÑOS HASTA LOS 9 AÑOS, 11 MESES, 29 DÍAS

Código de Vivienda											Familia No.											Persona No.										
Nombre 1											Nombre 2																					
Apellido 1											Apellido 2																					
Fecha de nacimiento	Día	Mes	Año	Edad	Sexo	M	F	Tipo ID	No. Doc. ID																							
Victima del conflicto	Si	No	Etnia	Afiliación al SGSSS	Si	No	Régimen	C	S	E	EPS	Encuestado por el SISBEN										Si	No									
¿Discapacidad?	Si	No	Tipo	1	2	3	4	¿Tiene Registro de Localización?	Si	No	Peso	Talla	IMC																			

PREGUNTAS	SI	NO	NA	ESPECIFIQUE	MOTIVO DE INASISTENCIA
1. ¿Asiste a CDI/Jardín Infantil? (6 meses - 5 años)				Público Privado	
2. ¿Asiste a escuela o colegio? (>5 años)				Público Privado	
3. ¿Asiste a programa de crecimiento y desarrollo?				Fecha última consulta	Día Mes Año
4. ¿Ha asistido a control de salud oral en los últimos 12 meses? (>6 meses)					
5. ¿Tiene vacunas completas para la edad? (Esquema PAI)					
6. ¿Está recibiendo lactancia materna exclusiva? (<6 meses)					
7. ¿Ya inició alimentación complementaria? (6 a 12 meses)					
8. ¿Tiene signos de malnutrición? (Cinta CIMDER) (<5 años)				Rojo Amarillo Verde	
9. ¿Actualmente se encuentra en seguimiento por diagnóstico de desnutrición?					
10. ¿Está inscrito en algún programa de alimentación complementaria?				¿Cuál?	
11. ¿Presenta algún signo de alarma? (<1 año)				¿Cuál?	1 2 3 4
12. ¿Se observan señales de alerta relacionadas con maltrato infantil?					
13. ¿Le han realizado valoración auditiva? (<1 año)					
14. ¿Le han realizado valoración visual? (<1 año)					
15. ¿Ha asistido a controles de agudeza visual? (1 a 5 años)					
16. ¿Ha asistido a controles de agudeza auditiva? (1 a 5 años)					
17. ¿Presenta algún tipo de dificultad en el lenguaje y/o comportamiento?					
18. ¿Ha estado enfermo de Hansen o Lepra?					
19. ¿Durante los últimos días ha notado alguna mancha en su piel que no le pique, no le duela o que no sienta esa área de la piel?					
20. ¿Tiene Tuberculosis?					
21. ¿Está en tratamiento?					
22. ¿Ha presentado tos por mas de 21 días seguidos en el último mes?					
Si la respuesta ANTERIOR es AFIRMATIVA, continuar con las siguientes preguntas					
23. ¿Ha tenido pérdida de apetito?					
24. ¿Ha tenido pérdida de peso o no aumenta de peso?					
25. ¿Ha presentado episodios de fiebre intermitente?					
26. ¿Ha tenido aislamiento al juego?					
27. ¿Ha estado irritable?					

TIPO ID	RC-Registro Civil	TI-Tarjeta de Identidad	NUIP- Número Único de Identificación Personal	M-Menor Sin Identificación
ETNIA	1. Indígena	2. Raizal	3. Afrodescendiente	4. Gitano (ROM)
TIPO DISCAPACIDAD	1. Física	2. Sensorial	3. Cognitiva	4. Psicosocial o mental
SIGNOS DE ALARMA	1.No puede beber o tomar del pecho 2.Vomita todo lo que ingiere 3.Ha tenido convulsiones 4.Está somnoliento o inconsciente			
MOTIVO DE INASISTENCIA	1. Desconocimiento 2. Falta de Recursos y/o tiempo 3. Atención no satisfactoria en IPS 4. Problemas de Afiliación 5. Creencias Religiosas o Culturales 6. Desinterés 7. Otro motivo			