



Uso de variables latentes para evaluar la percepción de riesgo frente a la COVID-19 en Cali y Palmira.

Laura Lucía Domínguez Barrios

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

Uso de variables latentes para evaluar la percepción de riesgo frente a la COVID-19 en Cali y Palmira.

Laura Lucía Domínguez Barrios

Trabajo de Grado presentado como requisito parcial para optar el título de:
Estadística

Director:

José Rafael Tovar Cuevas PhD. MSc

Codirector:

Andrés Felipe Ochoa Muñoz MSc

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

Dedicatoria

A mis padres quienes me han apoyado en el transcurso de mi vida académica, deportiva, y en general, siempre brindándome sus consejos y confianza para realizar cada una de las actividades del día a día.

Laura Lucía Domínguez Barrios

Agradecimientos

Agradezco a la Dr. Diana Caicedo y al grupo de investigación de la Pontificia Universidad Javeriana Cali, por compartir conmigo la información y los datos.

Al profesor José Rafael Tovar por ser mi director de trabajo de grado, por confiar en mí, escucharme, apoyarme, escucharme, aconsejarme y guiarme en el transcurso de la carrera como profesional, en mi vida en general y en la realización de este proyecto.

Al profesor Andrés Ochoa por su acompañamiento en el transcurso de este proyecto y de mi formación, su disposición, su amabilidad, su carisma y sus consejos, siempre estuvo ahí para mí.

A Taina Sánchez y Mónica Morales, por su amistad, amabilidad, compañerismo y por la estadística que nos permitió conocernos. A Jose Larsen por todo.

Al profesor Jaime Mosquera por guiarme y aconsejarme en los últimos años de estudio; a los evaluadores por sus comentarios para mejorar el presente trabajo de grado.

Laura Lucía Domínguez Barrios

Resumen

Los datos obtenidos en el proyecto: “Percepción de riesgo COVID-19 y sus factores relacionados en Cali y Palmira 2020-2021” son la semilla de este proyecto de trabajo de grado, los cuales fueron recolectados mediante un formulario con ítems tipo diferencial semántico que indaga sobre las características sociodemográficas de los individuos, además de tres dimensiones que forman la Percepción de riesgo de contagio.

La muestra de estudio estuvo conformada por 1463 individuos que trabajan y/o residen en las ciudades de Cali o Palmira, para estas ciudades se midieron indicadores resumen de cada una de las características sociodemográficas. Los ítems pertenecientes a cada uno de los subgrupos de percepción de riesgo (Probabilidad de contagio, Severidad y Susceptibilidad) se analizaron por medio de métodos multivariados, con la técnica del Análisis Factorial Múltiple (AFM o MFA en inglés); adicionalmente, se desarrolló una metodología heurística con el objetivo de darle importancia al juicio del experto y finalmente, comparar ambas metodologías.

Por medio de la metodología del AFM y la metodología heurística, se construyeron índices marginales como generales que permitieron resumir la información recolectada. Posteriormente se realizó un Pseudo Error Cuadrático Medio para comparar ambas metodologías y determinar el mejor índice construido entre AFM y Heurística; empleando métodos de reemuestreo y k-means se logró establecer niveles o puntos de corte de cada índice construido, en una escala de bajo, medio y alto. El índice seleccionado como el mejor, fue aquel que presentó un menor PECM. La medición de la variable latente Percepción de Riesgo podría ser de utilidad, debido a que su implementación probablemente sea por parte de la administración municipal y/o los entes de salud pública para llevar a cabo acciones y políticas para orientar a los individuos que se encuentren en situaciones más desfavorables frente a la enfermedad COVID-19 de acuerdo a su percepción de riesgo y a sus características sociodemográficas.

Palabras clave: Percepción, COVID-19, Análisis factorial múltiple, Variable latente, Índice sintético, Índice heurístico

Abstract

The data obtained in the project: “Perception of COVID-19 risk and its related factors in Cali and Palmira 2020-2021” are the seed of this degree project, which were collected through a form with semantic differential type items that inquires about the sociodemographic characteristics of individuals, in addition to three dimensions that form the Perception of contagion risk.

The study sample consisted of 1463 individuals working and/or residing in the cities of Cali or Palmira, for these cities summary indicators of each of the sociodemographic characteristics were measured. The items belonging to each of the risk perception subgroups (Probability of contagion, Severity and Susceptibility) were analyzed by means of multivariate methods, with the Multiple Factor Analysis (MFA) technique; additionally, a heuristic methodology was developed with the objective of giving importance to the expert’s judgment and finally, comparing both methodologies.

By means of the AFM methodology and the heuristic methodology, marginal and general indexes were constructed to summarize the information collected. Subsequently, a Pseudo Mean Squared Error was performed to compare both methodologies and determine the best index constructed between AFM and Heuristics; using re-sampling and k-means methods, it was possible to establish levels or cut-off points for each index constructed, on a scale of low, medium and high. The index selected as the best was the one with the lowest PECM. The measurement of the latent variable Risk Perception could be useful, since its implementation will probably be carried out by the municipal administration and/or public health entities to carry out actions and policies to guide individuals who are in more unfavorable situations in relation to the COVID-19 disease according to their risk perception and their sociodemographic characteristics.

Key words: COVID-19, perception, index, principal component analysis, multiple factorial analysis.

Contenido

Lista de Figuras	x
Lista de Tablas	x
1. Introducción	2
2. Planteamiento del problema	4
2.1. Problema contextual y estadístico	4
2.2. Justificación	6
2.3. Objetivos	7
2.3.1. Objetivo general	7
2.3.2. Objetivos específicos	7
3. Antecedentes	8
3.1. Antecedentes contextuales	8
3.2. Antecedentes estadísticos	9
4. Marco teórico	12
4.1. Marco contextual	12
4.2. Marco teórico estadístico	13
5. Metodología	22
5.1. Obtención de los datos	22
5.2. Fase 1: Construcción de índices	25
5.3. Fase 2: Pseudo Error Cuadrático Medio	28
5.4. Fase 3: Puntos de corte y clasificación de individuos	29
5.5. Presentación de resultados	30
6. Resultados	31
6.1. Análisis exploratorio	31
6.2. Análisis Factorial Múltiple	31
6.2.1. Análisis de Factores	32
6.2.2. Ponderación de los grupos de variables	32
6.2.3. Tipología de los grupos	33
6.2.4. Representación de las variables	34

6.2.5. Representación de los individuos	36
6.2.6. Proyección de ejes parciales de cada grupo	37
6.3. Formulación de un índice sintético	38
6.3.1. Índices usando métodos multivariados de la estadística	38
6.3.2. Índice usando información de expertos	39
6.4. Asociaciones entre variables	41
7. Conclusiones y recomendaciones	43
7.1. Conclusiones	43
7.2. Recomendaciones	44
A. Anexo: Resultados	45
B. Anexo: Cuestionarios	48
C. Anexo: Certificado de ponencia	51
D. Anexo: Código R	53
Bibliografía	55

Lista de Figuras

1-1. COVID-19 en Colombia hasta 01 de enero 2022. Tomada de www.larepublica.co/economia	3
4-1. Factor tamaño. Tomada de Lebart et al. (1995)	16
4-2. Tabla de datos X . Tomada de Escofier & Pagès (1992)	17
5-1. Secuencia metodológica	22
6-1. Porcentaje de varianza	32
6-2. Representación de los grupos	33
6-3. Representación del Círculo de correlación de las variables del AFM	34
6-4. Contribución de las variables cuantitativas	35
6-5. Nube de individuos del AFM	36
6-6. Contribución de los grupos a las dimensiones	37
6-7. Plano factorial simultáneo. M=Mujer, H=Hombre.	41
6-8. Identificación de tipos de individuos (clústers) según sus características sociodemográficas y los niveles del IPRG.	42
A-1. Distribuciones de los índices AFM	45
A-2. Distribuciones de los índices heurísticos	45
A-3. Mediana de cada índice construido en las respectivas ciudades	46

Lista de Tablas

5-1. Definición de variables sociodemográficas	23
5-2. Descripción de variables	23
5-3. Descripción de variables recodificadas	24
5-4. Ponderaciones brindadas por el especialista	27
6-1. Valores propios del AFM	32
6-2. Descomposición de la inercia de las primeras dimensiones	32
6-3. Correlaciones entre las proyecciones del AFM y de los grupos	33
6-4. Coeficientes RV	34
6-5. Coordenadas y contribuciones de las variables	36
6-6. Medidas de los Índices por dimensión	40
6-7. Valores mínimos y máximos por nivel de Índice.	40
6-8. Índice de Percepción de Riesgo Global en Cali y Palmira	41
A-1. Descripción de cluster 1	46
A-3. Descripción de cluster 3	47
A-2. Descripción de cluster 2	47

Declaración

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

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

Santiago de Cali, 09.08.2022

Laura Lucía Domínguez Barrios

1. Introducción

El presente trabajo de grado está inmerso en el proyecto de investigación de profesionales de la Pontificia Universidad Javeriana Cali, en el cual se estudia la percepción de riesgo de contagio de la enfermedad COVID-19 en los municipios de Cali y Palmira del departamento del Valle del Cauca, departamento bastante afectado por la mencionada enfermedad después de Bogotá Distrito Capital y el departamento de Antioquia, como se puede observar en la Figura 1-1. En estos municipios se encontró una alta prevalencia de casos positivos COVID-19, lo cual motivó a los investigadores para estudiar la evaluación que hacen los individuos expuestos, por medio de la construcción de un índice o indicador. A la muestra se le aplicó un cuestionario, donde se obtuvieron respuestas de 814 habitantes de Cali y 649 residentes de la ciudad de Palmira, con el objetivo de conocer aspectos relevantes sobre los imaginarios y las condiciones sociodemográficas de cada persona.

Buscando cuantificar la percepción de riesgo de los individuos estudiados, se desarrollaron dos propuestas metodológicas para la construcción de un índice que clasifique individuos de acuerdo con el grado de percepción; a partir de las preguntas del cuestionario (Anexo B) de los grupos de Probabilidad de Contagio, Severidad y Susceptibilidad.

La construcción del índice se basó en dos metodologías estadísticas, que permiten tener diferentes perspectivas de abordar el problema por medio de las variables latentes, las cuales no son medibles directamente; la adecuada selección del índice con mejor desempeño de una u otra metodología, se realiza desde una aproximación al error cuadrático medio vía bootstrap.

Posteriormente, con el fin de asociar dicho nivel de percepción de riesgo con cada una de las características sociodemográficas, se realizó un Análisis de Correspondencias Múltiples (ACM) y análisis de clustering, por medio de una clasificación jerárquica para agrupar individuos similares.

Para mayor comprensión, el contenido del presente trabajo de grado se dividió en siete capítulos: El Capítulo 1 está constituido por la Introducción, El Capítulo 2 está conformado por el planteamiento del problema, objetivos y justificación, El Capítulo 3 recopila las publicaciones más importantes encontradas en la literatura, El Capítulo 4 abarca el marco teórico, tanto conceptual alusivo al problema, como estadístico. El Capítulo 5 expone la metodología del análisis de los datos, en el Capítulo 6 se presentan los resultados obtenidos

desarrollados paso a paso de acuerdo a cada metodología empleada, por último, el Capítulo 7 exhibe las conclusiones, recomendaciones y líneas futuras. Se encuentran tres anexos, los cuales son de: Resultados, portada de los cuestionarios en ambas ciudades y certificado de ponencia de este proyecto.

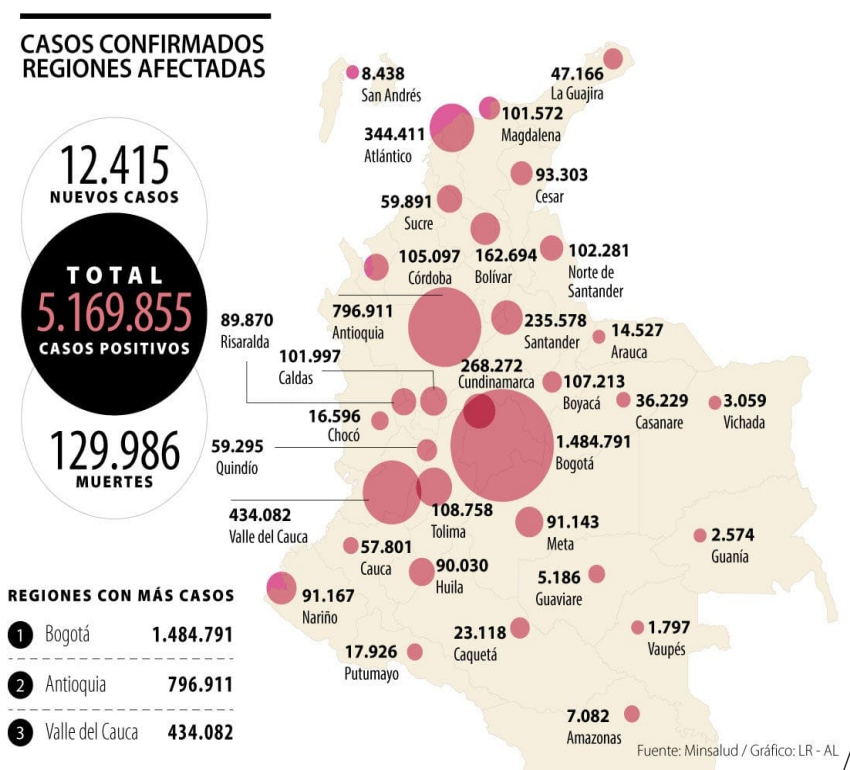


Figura 1-1.: COVID-19 en Colombia hasta 01 de enero 2022. Tomada de www.larepublica.co/economia

2. Planteamiento del problema

2.1. Problema contextual y estadístico

Según la Organización Panamericana de la Salud (2020), la enfermedad del coronavirus (COVID-19) fue ocasionada por el virus SARS-CoV-2 presentándose por primera vez como casos de neumonía en el mes de diciembre del año 2019 en la ciudad de Wuhan (China). Como consecuencia, se generó una epidemia mundial que se ha propagado rápidamente, exhibiendo efectos en el área social, económica y de salud pública.

En Colombia, el Ministerio de Salud y Protección Social expidió en marzo del 2020 la Resolución 385, titulada como Declaración de Emergencia Sanitaria, mod. 407 y 450, donde se toma como medida preventiva la cuarentena a nivel nacional. Sin embargo, hasta el 01 de enero de 2022 la enfermedad ha desencadenado 289927842 casos confirmados y 5436401 de muertes a nivel mundial. En Colombia, la cifra es de 129986 de muertes y 5169855 casos confirmados. En el Departamento del Valle del Cauca se hallan hasta el día 01 de enero de 2022, 434082 casos confirmados y 13291 muertes Johns Hopkins University (2022). Particularmente las ciudades de Cali y Palmira presentaron la mayor cantidad de casos confirmados de la enfermedad COVID-19, con 276492 y 22572 respectivamente; de igual forma el total de muertes en estas ciudades, es de 7242 y 1102 respectivamente Secretaría de Salud del Valle del Cauca (2022).

Debido a que no existe un tratamiento para la enfermedad, la OMS (2020a) recomienda como medidas de prevención guardar distanciamiento físico de mínimo un metro, usar tapabocas, lavado de manos, usar hidroalcohólico, rehusar al máximo tocarse los ojos, nariz y boca; evitando asistir a lugares concurridos. De acuerdo con la OMS (2020b), la tecnología creó una infodemia¹, que conlleva al exceso de la difusión de información errónea, generando temor, desconcierto, preocupación y pánico a través de las redes sociales; al que contribuyeron los diferentes medios de información social, generando una desconfianza y angustia en la comunidad en general por la crisis económica que fue aflorando a raíz del confinamiento; tornando escéptica a la población frente a los entes gubernamentales, implicando directamente el factor de la salud mental de la población, suscitando estrés,

¹Es un acrónimo de "información" y "epidemia" que se refiere a una propagación rápida, tanto de la información exacta e inexacta sobre algo, tal como una enfermedad.

depresión, frustración, enojo y cambiando por ende, la conciencia de los ciudadanos.

La percepción de riesgo frente alguna situación por parte de los individuos, influencia en su comportamiento preventivo y/o escepticismo, por lo cual, se estudió la percepción de riesgo frente a la la enfermedad COVID-19 en las ciudades de Cali y Palmira, donde los individuos realizaron una evaluación de los peligros de la enfermedad mencionada (Duan et al. 2020).

Es necesario para los mandatarios y entes de la salud pública, conocer cómo los habitantes y/o trabajadores de las ciudades de Cali y Palmira perciben el riesgo frente a la enfermedad COVID-19 e indagar si estos individuos tienen la capacidad de asumir comportamientos preventivos frente a la enfermedad, y así lograr implementar medidas preventivas más eficaces, puesto que aún no se tiene un índice de vacunación relativamente alto; hasta el primero de enero de 2022, el porcentaje de población completamente vacunada en Colombia fue de 55.7 %, recibieron dosis de refuerzo 6.5 % y al menos una dosis 75.4 %.

La Universidad Javeriana (Cali, Colombia) en dirección de las profesoras Diana Caicedo y Sheila Gómez, construyeron un instrumento de medición con reactivos, tipo diferencial semántico con el objetivo es identificar el nivel de percepción de riesgo que la población tiene frente a las posibilidades sufrir de esta enfermedad y los factores que puedan influenciar dicha percepción (Ver Anexo B) y factores que pueden estar asociados de diferentes aspectos del individuo, a partir de sus respuestas durante el pico de la pandemia enero-abril del 2021, por medio de 84 preguntas. El instrumento fue aplicado a través de las redes sociales, alcaldías y Empresas Sociales del Estado (ESE) en las ciudades de Cali y Palmira de manera separada. Como criterio de inclusión se tuvo en cuenta que la población fuera mayores de 18 años, trabajadores y/o residentes de cada una de las ciudades. En la ciudad de Cali se obtuvieron 814 observaciones y 649 en la ciudad de Palmira mediante un marco muestral, donde se utilizaron las bases de datos suministradas por las alcaldías y gobernaciones de cada ciudad; el formulario fue enviado a las personas que estaban en dichas bases de datos, con la expectativa de completar el formulario, sin embargo su diligenciamiento es aleatorio.

Los factores de interés epidemiológicos de la Percepción de riesgo de contagio fueron:

- Probabilidad de contagio.
- Percepción de riesgo: Severidad.
- Percepción de riesgo: Susceptibilidad.

De acuerdo con el instrumento construido (Anexo B); la Percepción de riesgo frente a la COVID-19 y las dimensiones que la componen, se consideran variables latentes en la estadística, puesto que no son medibles directamente. Según Silva (1997), las variables

latentes son un conjunto de variables relacionadas, que se desean operacionalizar para lograr cuantificarlas mediante la construcción de un indicador o índice. Por tal motivo, el problema estadístico de este trabajo es obtener un índice de percepción de riesgo global que sintetice y exprese la mejor aproximación de forma numérica de las variables de interés para el estudio (Hernández Sampieri et al. 2014).

2.2. Justificación

La enfermedad COVID-19 se catalogó como una amenaza global a nivel nacional y mundial, donde las consecuencias que conlleva para la población contagiada conllevan a una gran carga tanto social como económica. El controlar la multiplicación de la enfermedad se tomaron múltiples medidas preventivas, sin embargo el cuidado de cada individuo está relacionado con su percepción de riesgo de contagio frente a la enfermedad (Rohrmann 2008).

Hasta el momento se encontraron pocos estudios relacionados en las ciudades de Cali y Palmira; como en Caicedo et al. (2021a) y Caicedo et al. (2021b), donde realizaron un estudio descriptivo de la percepción del riesgo de contagio por COVID-19 y las variables que influyen en la misma. Por otro lado, se destaca un estudio a nivel mundial, publicado en Reino Unido como lo presentan en Dryhurst et al. (2020), entre otros publicados en Argentina, Cuba, Estados Unidos.

Este proyecto de investigación está orientado a desarrollar una metodología con buen desempeño, para construir un índice sintético que exprese el nivel de percepción de riesgo de contagio de la enfermedad mencionada, además de los factores relacionados de la población objeto de estudio. Se usaron metodologías de variables latentes, puesto que el estudio tiene en cuenta constructos de conciencia sobre la enfermedad COVID-19, los cuales no son observables directamente; entendiendo como constructos o dimensiones, cada uno de los factores que conforman la percepción de riesgo, los cuales son: Probabilidad de contagio, Severidad y Susceptibilidad. El análisis se desarrolló partiendo del uso de los métodos multivariados.

Este índice podría ser de utilidad, debido a que su implementación probablemente sea por parte de la administración municipal y/o los entes de salud pública para llevar a cabo acciones y políticas para orientar a los individuos que se encuentren en situaciones más desfavorables frente a la enfermedad COVID-19 de acuerdo a su percepción de riesgo y a sus características sociodemográficas.

2.3. Objetivos

2.3.1. Objetivo general

Proponer una metodología estadística para construir un índice que mida la percepción de riesgo frente a la enfermedad COVID-19 en los habitantes de Cali y Palmira.

2.3.2. Objetivos específicos

1. Construir un índice para cada una de las dimensiones identificadas por los investigadores en salud pública usando dos aproximaciones desde la estadística.
2. Evaluar empíricamente el desempeño estadístico de los índices propuestos.
3. Agrupar los ciudadanos con características sociodemográficas similares y de acuerdo al nivel de percepción de riesgo frente a la COVID-19.

Pregunta de investigación

Según los índices construidos, ¿Cuál es la percepción de riesgo frente a la enfermedad COVID-19 en los individuos de las ciudades de Cali y Palmira?

3. Antecedentes

Existe una variedad de investigaciones entorno a la enfermedad COVID-19; sin embargo se encontraron muy pocos estudios alusivos a la percepción de riesgo de contagio en Colombia. Este capítulo se fraccionó en dos secciones: Citas de investigaciones halladas referentes a la percepción de riesgo y citas de investigaciones donde emplearon métodos de la estadística multivariada para trabajar con variables latentes.

3.1. Antecedentes contextuales

A nivel internacional se encontró una gran cantidad de artículos relacionados a la percepción del riesgo de contagio por la enfermedad mencionada, como en Fujii et al. (2021), donde estudian la percepción pública, entendida por la susceptibilidad y la severidad percibida, al igual que los comportamientos de prevención de la enfermedad como tal; se realiza el estudio de manera transversal en diferentes países, entre ellos China y otros; la encuesta se dirige a personas mayores de edad y recogieron los datos en el lapso de un mes en el año 2020, obteniendo 5945 registros. Para el desarrollo del análisis, utilizan estadísticas descriptivas y un modelo de regresión logística, teniendo en cuenta las siguientes variables, condiciones de vida, probabilidad percibida, severidad percibida, efectividad percibida y sentimiento de ansiedad y características personales de la población en estudio.

En Colombia se encontraron dos publicaciones, Caicedo et al. (2021a) y Caicedo et al. (2021b), donde se lleva a cabo un estudio transversal en dos ciudades de Colombia, recolectando los datos por medio de una encuesta durante finales del mes de diciembre del 2020 hasta finales de enero del año 2021. Realizaron un análisis estadístico descriptivo. En primer lugar se caracteriza la población de acuerdo a factores socioeconómicos y de equidad, como sexo, étnia, estrato socioeconómico, ingreso, nivel de estudios, entre otros; y por último, se presenta el análisis sobre las variables relacionadas a la enfermedad COVID-19, tales como Percepción de riesgo de contagio, Conocimiento, Susceptibilidad, Severidad y Medidas de prevención (cumplimiento e incumplimiento) medidas en escalas tipo Likert y Diferencial semántico; demostrando por medio de porcentajes, el comportamiento y conciencia de los encuestados frente a la enfermedad, en Cali y Palmira, Valle del Cauca.

3.2. Antecedentes estadísticos

Análisis de Componentes Principales (ACP)

Xian et al. (2016) estudian la excelencia de la formación en la educación superior, la cual se evaluó mediante un sistema de índices (SET). Los autores obtuvieron los datos en el semestre de otoño del año 2012, de la Universidad de Correos y Telecomunicaciones de Chongqing, ubicada cerca de Chongqing en China. El SET está influenciado en una gran proporción por las actitudes, tanto de los educadores como de alumnos, y es por ello que desean determinar si el sistema de índices presenta multicolinealidad o si en realidad es confiable. Para llevar a cabo el estudio, la escala de medida que emplean es de tipo Likert, y se realizaron un Análisis de Componentes Principales (ACP) y un Análisis Cluster, clasificando cursos dictados en el semestre mencionado por medio de dendrogramas, de acuerdo a la calificación que se obtuvo en el SET, y Análisis discriminado de Fisher para probar la confiabilidad de la evaluación del SET, puesto que se determina la agrupación de los resultados de una nueva muestra.

Análisis de Correspondencias Múltiples (ACM)

Varela et al. (2018) estudian las costumbres de salud relacionadas con el sobrepeso y la obesidad de 239 pacientes entre 8 a 12 años de edad; establecieron la aleatoriedad de la muestra, seleccionando los menores de diferentes estratos socioeconómicos en Cali, Valle del Cauca, Colombia. A través de nueve índices, los autores sintetizan las variables del estudio, planteadas en el “Cuestionario de Hábitos de Salud relacionados con el Sobrepeso/Obesidad Infantil (CHS-SO)”, contando con treinta preguntas, las cuales son medidas con escalamiento tipo Likert. Construyen los índices de acuerdo a la información obtenida de acuerdo a las respuestas afirmativas de la encuesta realizada y el criterio del experto en el tema le asigna un valor numérico, para así darle el respectivo peso a los ítems y poder realizar la suma de estos; obteniendo valores cuantificables que, entre mayor sea el puntaje, representa un estado de salud saludable para el niño. Realizaron un análisis de la relación de cada índice construido con variables sociodemográficas; y se empleó la unión de un Análisis de Correspondencias Múltiples (ACM) y un Análisis de Conglomerados (ANC), para encontrar los tipos de Costumbres de Salud de los menores.

Análisis Factorial Múltiple (AFM)

Visbal-Cadavid et al. (2020) construyeron un índice sintético de Progreso de Educación Superior (IPES) basado en el Análisis Factorial Múltiple (AFM), donde clasificaron 32 universidades públicas de Colombia (IES) de acuerdo a su desempeño en el año 2018, mediante 16 variables en cuatro grupos activos y uno suplementario. Para el análisis utilizaron variables relacionadas con cada uno de los grupos, los cuales fueron: Calidad, Logro, Acceso y Continuidad de estudios.

Analizaron el porcentaje de varianza acumulado de cada grupo por separado y de forma global, luego deciden calcular el índice con las cargas factoriales de las primeras cuatro dimensiones del AFM, dado que recoge el 85.76 % de varianza acumulada en los datos.

Realizaron una transformación exponencial para lograr que los puntajes factoriales presenten un mínimo de cero, de tal forma que $f(F_i)$ es $1 + \frac{k-1}{2}e^{F_i}$ si $F_i < 0$ y $k - \frac{k-1}{2}e^{-F_i}$ en otro caso; donde k es el parámetro que representa el máximo que pueden tomar los valores transformados. El índice lo calcularon como $\sum_{i=1}^p f(F_i) \times \frac{\text{Varianza explicada por } F_i}{\text{Varianza total}}$ y por medio de este, lograron determinar el rendimiento de cada una de las universidades públicas de Colombia.

Tovar et al. (2021) midieron el placer hacia la lectura independiente en 752 estudiantes universitarios activos y matriculados en algún programa académico de una Universidad pública de Costa Rica por medio de la construcción de índices con ACP y AFM. Donde obtuvieron los datos por medio de un cuestionario con 18 ítems y nueve grupos.

Plantearon calcular un índice por AFM con respuestas de tipo ordinal, sumando la ponderación de los pesos de cada una de las variables y sus respectivas respuestas para cada individuo,

$$Indice_i = \sum_{j=1}^m P_j X_j + \sum_{j=1}^l \sum_{h=1}^k P_{hj} X_j$$

donde, denotaron m como las respuestas de las variables cuantitativas y dicotómicas y l las respuestas de las variables ordinales; P_j como la suma de las ponderaciones de los pesos de cada variable; X_j la variable; P_{hj} la suma de las ponderaciones de los pesos de cada variable ordinal. Adicionalmente mencionan que debió realizarse una transformación para lograr interpretar los resultados en términos de porcentaje,

$$Indice_i = \frac{Indice_{obs} - Indice_{min}}{Indice_{max} - Indice_{min}} * 100.$$

Debido a que los autores construyeron diferentes índices, desarrollaron un error de estimación de forma heurística, donde por medio de bootstrap repitieron 100 veces los índices construidos para cada individuo y se calculaba la mediana de dichos índices, donde por último, calcularon una mediana general. Luego, calcularon un error de estimación, como la diferencia entre la mediana del índice y la mediana general de las medianas obtenida anteriormente.

En Fetjah et al. (2022) investigaron el impacto del biocarbón de desecho de banano en el crecimiento de paspalum en la costa empleando datos bioquímicos y fisiológicos obtenidos durante tres estaciones climáticas en el sur de Marruecos, a una profundidad de 0-20 cm. Puesto que el conjunto de datos contiene variables cuantitativos y cualitativas organizadas

en grupos, realizaron un Análisis Factorial Múltiple (AFM) para valorar el efecto de la adición de enmiendas comestibles en los cambios fisiológicos estacionales y el análisis bioquímico del paspalum cultivados en dos fuentes de agua diferentes. Donde se descubrió que las mezclas de compost y biocarbón mejoraron las características fisiológicas, anatómicas y bioquímicas de paspalum en condiciones de maceta y campo durante las tres estaciones. Además, el biocarbón de desecho de banano mantuvo la estabilidad del estado hídrico de la hoja.

En Matar-Khalil et al. (2021) realizaron un estudio observacional transversal con el objetivo de validar un instrumento para medir la percepción de riesgo de contagio de COVID-19 (PCR-C19) por medio de cuatro dimensiones: Vulnerabilidad cognitiva, vulnerabilidad emocional, gravedad y , conductas de riesgo-protección; 40 ítems con diferencial semántico de cinco opciones de respuesta en 2350 individuos de 22 departamentos de Colombia. Realizaron un muestreo estratificado, proporcional y aleatorio.

Para el análisis de los datos construyeron un índice de forma heurística llamado $I_{PCR-C19}$ donde se divide la suma de cada uno de los puntajes promedios $\bar{P}P_{Di}$ correspondientes a la i -ésima dimensión del instrumento con $i = 1, \dots, n$ sobre $5n$ como se puede observar en la siguiente expresión

$$I_{PCR-C19} = \frac{\bar{P}P_{D1} + \dots + \bar{P}P_{Dn}}{5n},$$

donde el índice está en un rango de 0 a 1. Los autores generaron cinco categorías del índice con particiones de amplitud igual mediante la selección de modelos desde los criterios de BIC y AIC, hallando la distribución Weibull con el mejor ajuste. Pudieron llegar a concluir que el formulario PCR-C19 es válido y confiable para ser aplicado en otra ciudad, departamento o País.

De acuerdo a la revisión de los antecedentes, existen diferentes metodologías para la construcción e interpretación contextual de índices para variables latentes, los más útiles en estos casos son el ACP y AFM, permitiendo trabajar tanto con una amplia cantidad de variables como con una gran cantidad de individuos. También se hallaron estudios donde construyen índices desde metodologías heurísticas o empíricas.

4. Marco teórico

A continuación se expone el marco teórico empleado en este proyecto, el cual está constituido en dos secciones, la primera abarca los conceptos a tener en cuenta para comprender los términos del COVID-19 y percepción. Y la segunda, contiene las definiciones de los técnicas estadísticas útiles para desarrollar el estudio.

4.1. Marco contextual

Enfermedad por coronavirus (COVID-19)

Los coronavirus (CoV) son parte de la parentela de virus que originan múltiples consecuencias, tal como acontece con el coronavirus generador del síndrome respiratorio de Oriente Medio (MERS-CoV) y el que produce el síndrome respiratorio agudo severo 2 (SARS-CoV-2). El reciente coronavirus conocido como 2019-nCoV o COVID-19 se localizó por primera vez en Wuhan (China) en diciembre de 2019 (OPS 2020).

Etiología

El origen del COVID-19 es desconocido, sin embargo se han hecho investigaciones donde se concluye que es poco probable que el SARS-CoV-2 surgiera de maniobras en laboratorio de un coronavirus similar al SARS-CoV relacionado como lo expresan Andersen et al. (2020). De igual forma, la información hereditaria muestra que el SARS-CoV-2 no proviene de ninguna línea de virus previamente utilizada. Se ha tratado de explicar con muchas hipótesis la causa del origen del SARS-CoV-2.

Percepción de riesgo

La Percepción del Riesgo se define como las apreciaciones subjetivas de la población con respecto a una amenaza o peligro. En el despliegue de este proyecto, se utilizaron las siguientes nociones como lo indican Davis et al. (2003) y Caicedo et al. (2021 *a*):

- **Equidad:** La equidad o inequidad están relacionadas a varios factores característicos del individuo, como lo son: Género, Etnia, Estrato socioeconómico, Barrio y/o Comuna, Régimen de seguridad social, Salario y Ocupación. Por medio de la encuesta, se recogen las características mencionadas de cada encuestado para evaluar los riesgos percibidos.
- **Percepción de riesgo:** Se refiere al nivel de conciencia del individuo, visto desde tres dimensiones fundamentales para el estudio, tales como lo son:

- Probabilidad de contagio: Valoración subjetiva del individuo de la probabilidad de contraer la enfermedad COVID-19 según su cotidianidad.
- Susceptibilidad: Vulnerabilidad del individuo a enfermarse de la enfermedad COVID-19
- Severidad: Gravedad subjetiva de las consecuencias de contraer la enfermedad COVID-19 de cada individuo.

4.2. Marco teórico estadístico

Prueba Chi-cuadrado

Mendivelso & Rodríguez (2018) definen como prueba no paramétrica, la Chi-cuadrado; la cual se emplea para determinar asociaciones entre variables de una tabla de contingencia. La mencionada prueba presenta un contraste de hipótesis, donde la hipótesis nula expresa la no existencia de asociación entre las variables y la hipótesis alterna, la existencia de asociación entre las variables. El estadístico de prueba mide la diferencia entre una distribución de frecuencias esperadas (E_i) y una observada (O_i), expresado en la siguiente ecuación:

$$\chi^2 = \sum_{i=1}^k \left[\frac{(O_i - E_i)^2}{E_i} \right].$$

Variable latente

De acuerdo a Silva (1997) una variable latente, sintética o factor, es un conjunto de variables relacionadas entre sí, que se desean sintetizar para concluir los estudios, dichas variables se operacionalizan para darles un sentido representativo de forma matemática o cuantitativa, esto es lo que se denomina como indicador o índice. Por lo tanto, la Percepción de riesgo frente a la COVID-19 y las dimensiones que la conforman (Probabilidad de contagio, Severidad, Susceptibilidad) son consideradas variables latentes.

Escalas para la construcción de ítems

Existen tres métodos utilizados con frecuencia para llevar a cabo la medición por escalas de las variables latentes constituidas por actitudes (Hernández Sampieri et al. 2014):

- La **escala Likert** consiste en la reacción del encuestado ante un conjunto de ítems presentados en forma de afirmaciones o juicios, donde cada ítem tiene un valor para luego generar la suma de todas las puntuaciones del encuestado.
- El **diferencial semántico** es una escala de medición ordinal, sin embargo es común trabajarla como de intervalo. Y consiste en la calificación del encuestado ante una serie

de adjetivos extremos, siendo la reacción del individuo aquella que más lo identifica con su actitud.

- El **escalograma de Guttman** es similar a la escala Likert, solo que difieren en que algunos ítems indican en mayor medida la fuerza o intensidad de la actitud.

Índice o Indicador

En Silva (1997) se define el indicador, como una construcción teórica concebida para ser aplicada a la población o muestra y la intención es la cuantificación de algún concepto concurrente a la misma. El término de índice es empleado en dos casos, para referirse a una medida resumen como en nuestro caso, Índice de Percepción de riesgo y el otro caso, para dimensionar otra variable construida con otras.

Validación criterio del índice

Silva (1997) expresa en su libro, que una variable latente o sintética, debe ser fiable y válida para ser útil. Donde es fiable, cuando la variable mide en realidad lo que se desea medir en reiteradas ocasiones, el autor lo relaciona con la estabilidad que presenta el instrumento o la variable sintética; para medir dicha fiabilidad, se emplea el coeficiente kappa, el cual es un coeficiente de concordancia de los ítems. Y es válida, cuando la variable latente o sintética mide realmente lo que se pretende medir; dentro de la validez, están presentes tres tipos, los cuales son: Validez de aspecto, Validez de criterio, Validez de contenido, Validez predictiva y Validez por construcción.

Análisis Multivariado

La estadística multivariada es muy útil cuando se tiene gran información sobre cada individuo, así se pueden estudiar las interacciones entre todas las variables. Se presenta de forma teórica el Análisis de Componentes Principales (ACP) y el Análisis Factorial Múltiple (AFM).

Análisis de Componentes Principales (ACP)

El Análisis de Componentes Principales (principal component analysis) o PCA, se utiliza en el análisis de tablas de individuos representados por filas y las columnas, por valores numéricos; es útil para determinar el comportamiento de los individuos de acuerdo al conjunto de variables en estudio. Además, el objetivo principal del ACP, radica en la reducción de las dimensiones del conjunto inicial de variables, perdiendo muy poca información, es decir, proyectar los n individuos sobre el espacio de dimensión q , menor que

p . Trejos et al. (2014).

Existen dos formas de trabajar el ACP, cuando se centra la nube de los individuos, interpretando las relaciones de las variables con la correlación lineal y entre los individuos, por medio de las distancias; o no se centra dicha nube. Por otro lado, para llevar a cabo esta técnica de análisis, se debe cumplir con ciertos requerimientos, los cuales son: Continuidad en las variables, supuesto de linealidad en las variables, el número n de individuos debe ser mayor que el número p y la base de datos no permite trabajar con datos faltantes Ochoa Muñoz (2018).

Se trabaja con la matriz de datos $X_{n,p}$, donde n es la cantidad de individuos, representando las filas como $x_1, \dots, x_n$ y p , el número de variables, las cuales representan las columnas de la matriz de datos, denotadas como $x^1, \dots, x^p$.

La matriz a diagonalizar, es la matriz de correlaciones $Z'NZ$, la cual es simétrica y positiva, puesto que se desea maximizar la correlación entre los nuevos ejes y las variables originales; X ha sido estandarizada y es reconocida como Z , $N_{n \times n}$ es la matriz de pesos de los individuos.

Teniendo en cuenta las definiciones del ACP en Trejos et al. (2014) y González Rojas (2014):

Para maximizar la forma cuadrática $u'Z'NZu$ se utiliza el langragiano $L(u)$, garantizando que en el subespacio H se garantice un valor λ máximo.

$$\begin{aligned} L(u) &= u'Z'NZu - \lambda(u'u - 1) \\ L(u) &= u'Z'NZu - \lambda u'u + \lambda \\ \frac{\partial L}{\partial u} &= 2Z'NZu - 2\lambda u = 2(Z'NZu - \lambda u) \end{aligned}$$

Igualando a cero se tiene el siguiente sistema de valores y vectores propios, donde que $Z'NZ = \text{cor}(x_j, x_{j'})$:

$$Z'NZu = \lambda u$$

- **Distancia:** Se calculan entre dos puntos i y l , como se expresa en el teorema de Pitágoras

$$d^2(i, l) = \sum_j m_j (x_{ij} - x_{lj})^2$$

- **Inercia:** Se refiere a la varianza y/o dispersión subyacente en la nube de los individuos con el centro de gravedad g , la cual se calcula como $I_n = \sum_{i=1}^n p_i d^2(i, g)$
- **Vectores principales:** Denotados por $u_1, \dots, u_q$, son el resultado de ordenar de forma descendente los valores propios $\lambda_1, \dots, \lambda_q$.

- **Componentes principales:** Son las combinaciones lineales de las variables y están relacionadas a los ejes principales.

Factor tamaño

El factor tamaño o efecto talla como lo define Escofier & Pagès (1992) se presenta cuando todas las variables están correlacionadas positivamente entre ellas, situándose en el mismo lado del primer eje factorial, formando ángulos agudos. Esta característica permite el cálculo de índices estadísticos, a partir de la primera componente principal ψ_1 .

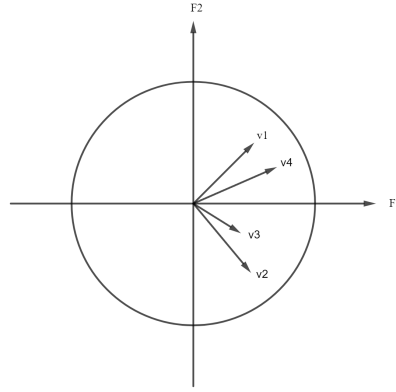


Figura 4-1.: Factor tamaño. Tomada de Lebart et al. (1995)

Análisis de Correspondencias Múltiples (ACM)

Para analizar tablas de individuos descritos por variables categóricas y para estudiar las asociaciones entre diferentes modalidades de las variables en estudio se emplea el ACM. Este análisis parte de una tabla disyuntiva completa Z de n individuos y p categóricas o modalidades de s preguntas según Pardo & Cabarcas (2001).

La tabla disyuntiva completa asigna 1 a la presencia de la característica en la variable y 0 la ausencia de la característica, de tal forma $Z_{i.} = \sum_{j=1}^p Z_{ij} = s$ y $Z_{.j} = \sum_{i=1}^n Z_{ij}$ corresponde a la suma de los individuos que indican presentar la modalidad j de la pregunta q , por lo tanto la suma de las filas y las columnas generan la frecuencia total z de la tabla Z , así $z = \sum_{i=1}^n \sum_{j=1}^p Z_{ij} = ns$.

Teniendo en cuenta que el ACM busca describir los individuos de acuerdo a su proximidad (similitud entre i e i'), se construyen las distancias ji-cuadrado, las cuales también se pueden construir entre modalidades, se realizan para la tabla disyuntiva completa Z , dichas

distancias están dadas por:

$$d^2(i, i') = \sum_{j=1}^p \left(\frac{z_{ij}}{s\sqrt{\frac{z_{.j}}{ns}}} - \frac{z_{i'j}}{s\sqrt{\frac{z_{.j}}{ns}}} \right)^2 = ns \sum_{j=1}^p \frac{1}{z_{.j}} (z_{ij} - z_{i'j})^2$$

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

donde se consideran dos individuos próximos, si han seleccionado las mismas modalidades j y dos modalidades de variables diferentes coinciden, si son seleccionados por los mismos individuos. Mientras que las modalidades con baja frecuencia se alejará de las demás, al igual que los individuos excéntricos. Y como objetivo geométrico en ACM, se busca un nuevo sistema de ejes ortogonales u_x , los que se proyecte la inercia de la nube de individuos, tal que los primeros ejes concentren la mayor parte de la misma y en forma decreciente, la inercia está dada por $I = \sum_q^s I_q = \frac{p}{s} - 1 = \sum_{j=1}^p I_j = \sum_{q=1}^s I_q$.

Análisis Factorial Múltiple (AFM)

El AFM desarrollado por Escofier & Pagès (1992), es una metodología basada en el análisis de varios grupos de variables de un mismo conjunto de individuos, donde cada grupo de variables está representado por una tabla. El AFM tiene en cuenta los tres espacios para llevar a cabo los análisis: R^K : Nube de individuos, R^I : Nube de variables y R^{I^2} : Nube de grupos de variables.

El AFM se aplica a bases de datos que cuentan con variables cualitativas y cuantitativas, adicionalmente, las variables pertenecen a grupos estrictamente del mismo tipo de variable (cualitativa o cuantitativa). Por lo tanto, se generan tablas para cada grupo.

Los datos se expresan en una matriz X , la cual es una tabla completada por subtablas (grupos de variables), cruzando de esta forma, los individuos y las variables (cualitativas y cuantitativas).

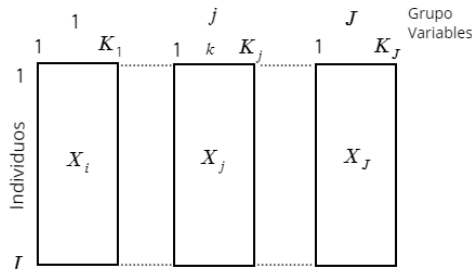


Figura 4-2.: Tabla de datos X . Tomada de Escofier & Pagès (1992)

La notación que se expresa en la Figura 4-2, es la seguida por Escofier & Pagès (1992) donde X : Tabla completa de datos, X_j : Tabla del grupo j , I : Conjunto de individuos, K : Conjunto de variables de todos los grupos, J : Conjunto de subtablas, K_j : Conjunto de variables del j -ésimo grupo.

El AFM tiene busca una visión general de los datos y las posibles relaciones entre los individuos como la relación entre las variables. En Escofier & Pagès (1992) sugieren realizar tanto el AFM, como análisis separados para encontrar relaciones entre grupos. Por ello, debe encontrarse el peso de cada variable en el grupo j , diviendolo por el primer valor propio del AFM separado del grupo j , denotada por λ_j^1 .

Según Abascal & Landaluce (2002), Escofier & Pagès (1992), el AFM presenta dos pasos fundamentados en el ACP, los cuales son: El análisis separado y El análisis global, quienes se describen a continuación

1. **El análisis separado** se fundamenta en asignarle la nube de individuos con los valores de las variables de X_j correspondientes a cada uno de los j -ésimos grupos con $j = 1, \dots, J$; se debe tener en cuenta el tipo de variable (cuantitativa o cualitativa), puesto que esto influye en el análisis a realizar (ACP o ACM respectivamente).
2. **El análisis global** se obtiene al realizar un ACP de la tabla global, donde se asigna una ponderación a cada grupo X_j para lograr equilibrarlos. La ponderación de cada grupo se obtiene del inverso del primer valor propio del grupo correspondiente en el ACP de la propia tabla o ACM de acuerdo al tipo de la variable, es decir, $\frac{1}{\lambda_j}$. Dicha ponderación sostiene la estructura de cada tabla, debido a que la inercia máxima en cada una de las nubes de individuos es uno en cualquier dirección.

Bootstrap o remuestreo

La técnica de Bootstrap es empleada para remuestrear diferentes submuestras de los datos originales, siendo independientes e idénticamente distribuidas. Da la posibilidad de realizar estimaciones puntuales. El paso inicial para llevar a cabo esta técnica, es escoger el estimador de interés $\hat{\theta}$ y luego, se obtiene una muestra de n individuos con reemplazo y m variables aleatorias idénticamente distribuidas. Posteriormente, se itera t veces el paso anterior y de esta forma, se establece la *muestra bootstrap* (Solanas & Olivera 1992).

Cluster

El Análisis de Clúster o Conglomerados es conocido como un método de clasificación no supervisada, donde se busca formar agrupamientos de observaciones similares entre sí, pero diferente de las demás; el algoritmo produce una jerarquía de particiones por medio de

dendrogramas con $n - 1$ particiones (Pardo & Del Campo 2007).

El concepto de distancia d , se define si x, y, z son tres objetos donde x e y se agrupan en un solo elemento denotado como h ; la distancia de salto mínimo es buscar la distancia menor entre los elementos del grupo h a z se evalúa a través de la siguiente expresión

$$d(h, z) = \min \{d(x, z), d(y, z)\}.$$

La distancia de salto máximo consiste en tomar la distancia más grande de los elementos de h a z por medio de la expresión

$$d(h, z) = \max \{d(x, z), d(y, z)\}.$$

La distancia media es la más empleada para hallar la distancia entre dos objetos x e y reagrupados en h , de tal forma que

$$d(h, z) = \frac{d(x, z) + d(y, z)}{2}.$$

El algoritmo de clasificación ascendente jerárquica se desarrolla mediante las siguientes etapas:

1. Se cuenta con n individuos para clasificar.
2. Se forma la matriz de distancias entre los n individuos y se buscan los dos más cercanos, los cuales se agregan en un individuo nuevo; generando en esta primera etapa $n - 1$ clases.
3. Se forma de nuevo la matriz de distancias una vez hecha la agregación anterior, calculando las distancias entre el nuevo individuo agregado y los individuos restantes, contando con $n - 1$ individuos a clasificar.
4. Se buscan dos individuos más cercanos y se agregan, generando una segunda partición con $n - 2$ clases.
- m. Se calculan las nuevas distancias y se itera de nuevo el algoritmo hasta tener un grupo con todos los individuos, representando la última partición.

Valores test

Según Pardo & Del Campo (2007) para lograr la caracterización de las clases para variables continuas, es necesario contar con X como variable continua, $\bar{X}$ como la media general, $\bar{X}_k$ como la media de la variable X de la clase k y la varianza de la variable en la clase, denotado por $s_k^2(X) = \frac{(n - n_k)s^2(X)}{(n - 1)n_k}$; por lo tanto los valores test se obtienen de la expresión

$$t_k = \frac{\bar{X}_k - \bar{X}}{s_k(X)}$$

donde $s^2(X)$ es la varianza de la variable X en todo el conjunto de datos, n el número de individuos clasificados y n_k el número de individuos dentro de la clase k . Por lo tanto, $|t(X_1)_{k=k}| > 2$ indica que la variable X_1 es relevante en la clase k .

Método de k -means

El algoritmo k -means, llamado también como nubes dinámicas o centros móviles, tiene como objetivo clasificar los individuos. Cuenta con dos métodos diferentes para llevar a cabo dicha técnica, los cuales son propuestos por: Forgy en 1966, donde se clasifican a los individuos y posteriormente, se calculan los centros; y McQueen en 1967, donde cada que se clasifica un individuo, se recalcula los centros (Trejos et al. 2014).

En este algoritmo tiene relevancia la distancia Euclídea, puesto que con esta se construyen los individuos promedios. El método k -means consiste en escoger un K que representa las clases para clasificar a los individuos, posteriormente se calculan los centros de gravedad para los K escogidos $k = \{1, \dots, K\}$, en términos matemáticos sería $g_k = \frac{1}{|C_k|} \sum_{x_i \in C_k} x_i$ y se calcula la inercia al interior de cada una de las clases $W = \frac{1}{n} \sum_{k=1}^K \sum_{i \in C_k} \|x_i - g_k\|^2$ y el método obtiene estabilidad cuando el W calculado supera el máximo de iteraciones (Trejos et al. 2014).

Construcción de índices

En el AFM se construyen los índices por medio de los pesos relativos de las variables y la primera componente principal ya que contiene toda la variación de los datos y permite diferenciar los puntajes de los individuos, así Becerra (2010): $\psi_1(i)$ es la primera componente principal, $a = \frac{100}{\max(\psi_1(i)) - \min(\psi_1(i))}$ y $b = -a * \min(\psi_1(i))$

$$Indice(i) = a\psi_1(i) + b$$

El índice construido estará en una escala de 0 a 100% para ser interpretado. Existe otra metodología para construir índices propuesta en Becerra (2010), donde se tienen variables cualitativas y cuantitativas, obteniendo mediante la suma ponderada de los pesos relativos de cada variable y la respuesta a cada una de las m variables cuantitativas y dicotómicas, y las l variables ordinales,

$$Indice_i = \sum_{j=1}^m \beta_j X_j + \sum_{j=1}^l \sum_{h=1}^k \beta_{hj} X_j$$

Donde,

- $Indice_i$: Índice del individuo i en concreto.
- β_j : Suma ponderada de los pesos relativos de cada j variable cuantitativa y/o dicotómica.
- X_j : Variable.
- m : Respuesta a cada una de las m variables cuantitativas.
- l : Respuesta a cada una de las variables cualitativas.
- β_{hj} : Suma ponderada de los pesos relativos de cada variable cualitativa.

Para poder realizar interpretaciones con los índices contruidos y los resultados obtenidos, se transforman en escalas de 0 a 100 %, de la siguiente manera:

$$Indice_i = \frac{Indice_{obs} - Indice_{min}}{Indice_{max} - Indice_{min}} \times 100$$

5. Metodología

Este capítulo especifica la metodología empleada en el trabajo de grado cumpliendo con los objetivos establecidos. El problema se analizó desde el desarrollo de las etapas expuestas en la Figura 5-1,

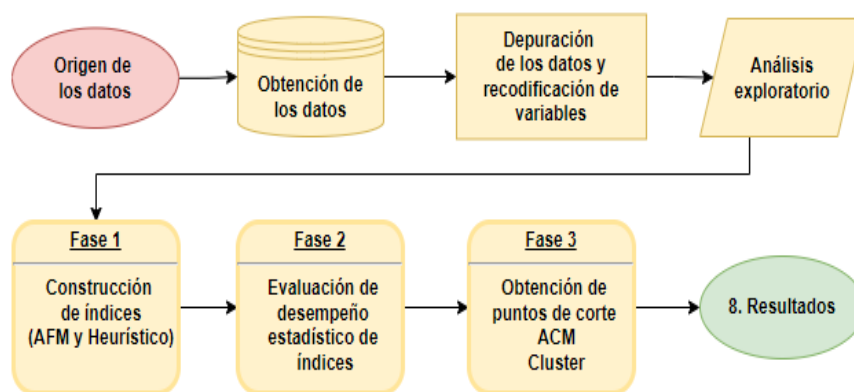


Figura 5-1.: Secuencia metodológica

Origen de los datos

La población objeto de estudio la conforman todos los habitantes mayores de edad de las ciudades de Cali y Palmira, ubicadas en el departamento del Valle del Cauca, Colombia. Los criterios de inclusión fueron: Ser mayor de 18 años, trabajar o residir en la ciudad de Cali y/o Palmira, y autorización de tratamiento de datos.

5.1. Obtención de los datos

Se empleó el instrumento realizado tipo formulario por Caicedo et al. (2021a) (Ver Anexo B) en las ciudades de Cali y Palmira, siendo enviado y publicado por medio de correo electrónico, plataformas digitales, redes sociales de la Pontificia Universidad Javeriana Cali y con la colaboración de las ESE de cada municipio. El formulario para la obtención de los datos estuvo habilitado para ser respondido una sola vez, en el periodo de diciembre del

2020 a enero del 2021. Se obtuvieron 814 observaciones de Cali y 649 de Palmira.

En la Tabla 5-1 se describen las variables sociodemográficas que conforman la Equidad, las cuales fueron útiles para conocer si existía asociación con el índice construido y cada una de las dimensiones que lo conformaron. El ingreso se categorizó de acuerdo a intervalos del salario mínimo legal vigente (SMLV).

Tabla 5-1.: Definición de variables sociodemográficas

Variable	Definición
Sexo	Femenino o Masculino
Edad	18-29 años, 30-59 años o 60 años o más.
Etnia	Mestizo/Blanco, Negro/Raizal/Afrodescendiente o Indígena.
Área de residencia	Urbana, Rural o no sabe.
Municipio	Cali, Palmira
Estrato	1-2, 3-4 o 5-6.
Nivel educativo	Especialización, maestría, doctorado; universitario, bachillerato completo/técnico o bachillerato incompleto.
Ingresos	Sin ingresos o menos a 1 SMLV, 1 a 3 SMLV o más de 3 SMLV.
Ocupación	Jubilado/Pensionado, Desempleado/Hogar/Trabajador informal, Independiente, Estudiante o Empleado.

Instrumento de medición: El instrumento de medición está construido en varias secciones, las cuales son las dimensiones como se observa en la Tabla 5-2, donde las opciones de respuesta de preguntas p46 a p49 y p51 a p54 corresponden a 1=Extremadamente improbable y 7=Extremadamente probable; las preguntas de $p50_1$ a $p50_3$ cuentan con opciones de respuesta de 1=Fuertemente en desacuerdo y 5=Fuertemente en acuerdo; las preguntas de p56 a p60 presentan opciones de respuesta de 1=Nada afectado y 7= Muy afectado.

Tabla 5-2.: Descripción de variables

Dimensión	Pregunta	Descripción	Respuesta
Probabilidad de contagio (X_1)	p46	Probabilidad de afectarse en los próximos 2 meses de COVID-19	1 a 7
	p47	Probabilidad de familiares/amigos sean diagnosticados con COVID-19 en los próximos 2 meses	
	p48	Sea diagnosticado con COVID-19 en 2 meses en comparación con personas de su edad	
	p49	Probabilidad de la mayoría de la población se contagie de COVID-19	1 a 5
	p50 ₁	El coronavirus / COVID-19 afectará a muchas personas en el país	
	p50 ₂	Probablemente me enfermaré con el coronavirus / COVID-19	
Percepción de riesgo: Severidad (X_2)	p50 ₃	Enfermarse con el coronavirus / COVID-19 puede ser grave	1 a 5
	p51	Probabilidad de enfermarse gravemente por COVID-19	1 a 7
	p52	Probabilidad de enfermarse gravemente si usted fuera diagnosticado con COVID-19	
	p53	Probabilidad de morir al ser diagnosticado con COVID-19	
	p54	Probabilidad de sufrir una enfermedad grave por COVID-19 comparado con personas de su edad	
Percepción de riesgo: Susceptibilidad (X_3)	p55	Sufre de una o de las siguientes enfermedades: Asma; EPOC; Cáncer; Presión Alta; Diabetes; Insuficiencia Renal; Enfermedad Autoinmune; Desnutrición; VIH; Tuberculosis; Obesidad; Fumador; Enfermedad Cardíaca	1=Si, 0=No
	p56	Por su condición de salud, probabilidad de afectarse de COVID-19	1 a 7
	p57	Por su edad, probabilidad de afectarse por COVID-19	1 a 7
	p58	Por su salud actual, probabilidad de afectarse por COVID-19 en comparación con personas de su edad	1 a 7
	p60	Gravedad para usted de contraer la enfermedad COVID-19	1 a 7

Depuración de los datos

Para realizar la depuración de las bases de datos se utilizó el software IBM SPSS versión 26. Se codificaron las variables con el objetivo de detectar todas las inconsistencias en los mismos. Finalmente se consolidó la base de datos general (Cali y Palmira) con 1443 individuos; 797 registros de la ciudad de Cali y 646 de Palmira.

Recodificación de las variables

En la Tabla 5-2 se demuestra la estructura del instrumento de acuerdo a cada grupo o dimensión y las preguntas que se emplearon para conformar los mismos. La mayoría de las preguntas se encuentran en escala Likert de 1 a 5 y 1 a 7. Debido a la existencia de una pregunta con opción múltiple, se procedió de la siguiente manera.

Percepción de Susceptibilidad: La pregunta p_{55} quería conocer la(s) enfermedades que posee el individuo i donde las opciones de respuesta eran binarias, por lo cual se procedió a transformar esta variable en x_{91} , la cual contiene la cantidad de enfermedades o condiciones de salud del participante i con $i = 1, \dots, I$, puesto que se realizó la suma de enfermedades individuo; entre más enfermedades cuente, más susceptible será a la COVID-19. Para este caso, se evidenciaron individuos hasta con 13 enfermedades.

$$x_{i91} = Asma_i + EPOC_i + Cáncer_i + PresiónAlta_i + Diabetes_i + InsuficienciaRenal_i + \\ EnfermedadCardiaca_i + EnfermedadAutoinmune_i + Desnutrición_i + VIH_i + \\ Tuberculosis_i + Obesidad_i + Fumador_i$$

Una vez se recodificaron las variables, se genera la base de datos codificada con la cual se trabajó en el capítulo 6 (ver Tabla 5-3).

Tabla 5-3.: Descripción de variables recodificadas

Dimensión	Codificación	Descripción	Respuesta
Probabilidad de contagio (X_1)	x71	Probabilidad de afectarse en los próximos 2 meses de COVID-19	1 a 7
	x72	Probabilidad de familiares/amigos sean diagnosticados con COVID-19 en los próximos 2 meses	
	x73	Sea diagnosticado con COVID-19 en 2 meses en comparación con personas de su edad	
	x74	Probabilidad de la mayoría de la población se contagie de COVID-19	1 a 5
	x75	El coronavirus / COVID-19 afectará a muchas personas en el país	
	x76	Probablemente me enfermaré con el coronavirus / COVID-19	
	x77	Enfermarse con el coronavirus / COVID-19 puede ser grave	
Percepción de riesgo: Severidad (X_2)	x81	Probabilidad de enfermarse gravemente por COVID-19	1 a 7
	x82	Probabilidad de enfermarse gravemente si usted fuera diagnosticado con COVID-19	
	x83	Probabilidad de morir al ser diagnosticado con COVID-19	
	x84	Probabilidad de sufrir una enfermedad grave por COVID-19 comparado con personas de su edad	
Percepción de riesgo: Susceptibilidad (X_3)	x91	Cantidad de enfermedades o condiciones de salud del participante	0 a 13
	x92	Por su condición de salud, probabilidad de afectarse de COVID-19	1 a 7
	x93	Por su edad, probabilidad de afectarse por COVID-19	1 a 7
	x94	Por su salud actual, probabilidad de afectarse por COVID-19 en comparación con personas de su edad	1 a 7
	x95	Gravedad para usted de contraer la enfermedad COVID-19	1 a 7

Análisis multivariado

Se realizó el Análisis Factorial Múltiple por medio del software estadístico R versión 4.1.1 y las librerías *factoextra* (Kassambara & Mundt 2020) y *FactoMineR* (Lê et al. 2008). El análisis se estructuró con 16 variables activas (7 variables en X_7 , 4 variables en X_8 y 5 variables en X_9).

Por lo tanto el AFM se desarrolla sobre los 1443 individuos y las 16 variables activas de la base de datos general, una vez fue depurada. Para determinar el número de ejes factoriales, se utilizó del porcentaje de varianza acumulada, donde se recogió el 51.94 % en las dos primeras componentes. Las representaciones de los grupos y los individuos se realizaron gráficamente, donde los grupos se presentaron mediante el plano factorial y los individuos por la proyección de individuos, las variables se representaron en el círculo de correlaciones, en el cual se verificó el factor tamaño, necesario para la construcción del índice.

Se estimaron las correlaciones de los grupos en los dos primeros ejes, los coeficientes RV y las contribuciones de las variables, para valorar las relaciones entre los grupos y los aportes que las variables realizan a cada eje factorial o dimensión.

5.2. Fase 1: Construcción de índices

Índice usando AFM

Para construir el Índice de Percepción de Riesgo Global frente a la COVID-19 (IPRG), empleando el Análisis Factorial Múltiple (AFM) se deben llevar a cabo los siguientes pasos:

1. Una vez obtenido el archivo de datos, realiza la tipificación de variables
2. Defina cada subtabla (grupo o dimensión) con sus respectivas variables (factores)
3. Construye la matriz de correlación y se observa si todas las correlaciones son positivas
 - Si cumple, entonces continua al paso siguiente
 - Si no cumple, se podría reescalar las variables o realizar una selección de variables como lo sugiere Pardo (2020)
4. Realiza el Análisis Factorial Múltiple (AFM) con K factores con $k = 1, \dots, K$ e I individuos con $i = 1, \dots, I$ por medio de la función *MFA* de la librería *factoextra* desarrollada por Lê et al. (2008) y confirma el cumplimiento de la característica factor tamaño por medio del círculo de correlación
5. Teniendo en cuenta que el AFM realiza un ACP general con la tabla completa de datos y por separado, un ACP o ACM de acuerdo a las variables de cada subtabla, se tiene

- Análisis del conjunto: Utilizando todo el archivo de datos completo
 - Análisis separado:
 - Realiza un ACP al archivo de datos de la dimensión Probabilidad de contagio de 1443×7
 - Realiza un ACP al archivo de datos de la dimensión Severidad de 1443×4
 - Realiza un ACP al archivo de datos de la dimensión Susceptibilidad de 1443×5
6. Obtenga del análisis en conjunto la primera coordenada y el primer valor propio y obtenga del análisis en separado la primera coordenada y el primer valor propio de cada j -ésima dimensión o subtabla $j = 1, \dots, J$.
 7. Obtenga el vector propio de acuerdo al análisis que corresponda en el paso anterior, mediante la división de la primera coordenada sobre la raíz cuadrada del primer valor propio.
 8. Calcule los pesos relativos β_k y β_{jk} para cada k -ésimo factor y para cada k -ésimo factor de la j -ésima dimensión respectivamente, mediante la división de cada elemento del vector propio sobre la suma del vector propio obtenido en el paso 7. Puede verificar que la suma de los pesos relativos de la unidad
 9. Calcule el índice del análisis del conjunto para cada individuo i , es decir, la Percepción de Riesgo frente a la COVID-19 de la siguiente manera

$$IPRG_i = \sum_{k=1}^K \beta_k X_{ik}$$

donde X_{ik} es el valor de cada respuesta del individuo i en la respectiva variable o factor k (con $k = 1, \dots, 16$ presentes en la Tabla **5-3**)

10. Calcule el índice del análisis del separado para el individuo i , es decir, para cada j -ésima dimensión con $j = 1, \dots, J$ de la siguiente manera

$$Índice_{ij} = \sum_{k=1}^{K_j} \beta_{jk} X_{ijk}$$

donde X_{ijk} es el valor de cada respuesta del individuo i en la respectiva variable o factor k (con $k = 1, \dots, K_j$ de la j -ésima dimensión)

Índice usando información de expertos

Se propone la construcción de índices por medio de una metodología heurística, la cual tiene en cuenta la información del experto o especialista. Esta metodología se realizó con el objetivo de comparar ambos procedimientos. Para obtener el valor del índice se deben llevar a cabo los siguientes pasos:

1. Identifique uno o más especialistas en el tema de investigación
2. Pida al especialista una ponderación de relevancia de 1 a 10, siendo 10 el valor de mayor importancia de cada una de las dimensiones incluidas y posteriormente, a cada uno de los factores incluidos en las respectivas dimensiones o subtablas.
3. Organiza las ponderaciones asignadas por el especialista (ver Tabla 5-4): p_j para la j -ésima dimensión, con $j = 1, \dots, J$ y p_{jk} para el k -ésimo factor de la j -ésima dimensión con $k = 1, \dots, K$.
4. Obtenga la ponderación W_j de las x_j dimensiones, se calcula el peso relativo de cada p_j sobre el total de estos, de forma que $W_j = \frac{p_j}{\lambda}$, donde $\lambda = \sum_{j=1}^J p_j$.
5. Obtenga la ponderación π_{jk} de los x_{jk} factores, se calcula el peso relativo de cada p_{jk} sobre el total de estos en la j -ésima dimensión y se multiplica por el peso W_j , de forma que $\pi_{jk} = \frac{p_{jk}}{\lambda_j^*} W_j$, donde $\lambda_j^* = \sum_{k=1}^{K_j} p_{jk}$. Finalmente se conforma la Tabla 5-4

Tabla 5-4.: Ponderaciones brindadas por el especialista

Dimensión	Variable	p_{1k}	π_{1k}	Dimensión	Variable	p_{2k}	π_{2k}	Dimensión	Variable	p_{3k}	π_{3k}
Probabilidad de contagio $p_1 = 9$ $W_1 = 0,375$	x_{71}	9	0.061	Severidad $p_2 = 8$ $W_2 = 0,333$	x_{81}	7	0.080	Susceptibilidad $p_3 = 7$ $W_3 = 0,292$	x_{91}	8	0.065
	x_{72}	8	0.055		x_{82}	8	0.092		x_{92}	8	0.065
	x_{73}	9	0.061		x_{83}	8	0.055		x_{93}	7	0.057
	x_{74}	7	0.048		x_{84}	6	0.069		x_{94}	6	0.049
	x_{75}	6	0.041						x_{95}	7	0.057
	x_{76}	8	0.055								
	x_{77}	8	0.055								

6. Calcule el Índice de Percepción de Riesgo Global (IPRG) para cada individuo i con $i = 1, \dots, I$

$$IPRG_i = \sum_{j=1}^J \sum_{k=1}^{K_j} \pi_{jk} X_{ijk}$$

donde π_{jk} : Peso del factor k -ésimo en la dimensión j -ésima y X_{ijk} : Respuesta del individuo i en el k -ésimo factor de la j -ésima dimensión.

7. Calcule los índices de cada j -ésima dimensión para cada individuo i con $i = 1, \dots, I$, donde

$$Índice_{ij} = \sum_{k=1}^{K_j} \frac{p_{jk}}{\lambda_j^*} X_{ijk}$$

Transformación de índices

Una vez contruidos los índices por medio de las dos metodologías planteadas anteriormente (empleando AFM y usando información de expertos), el índice se transforma para cada individuo i con $i = 1, \dots, I$ en un rango de 0 a 100 para facilitar su interpretabilidad y posibilitar la comparación; mediante la siguiente expresión

$$\theta_i = \frac{\text{Índice}_i - \min(\text{Índice})}{\max(\text{Índice}) - \min(\text{Índice})} \times 100$$

donde Índice_i : Valor del índice en el individuo i , $\min(\text{Índice})$: Valor mínimo del índice construido de los I individuos y $\max(\text{Índice})$: Máximo del índice del total de los I individuos.

5.3. Fase 2: Pseudo Error Cuadrático Medio

Para comparar el desempeño estadístico de los índices contruidos, se propone una metodología que aproxime un Pseudo Error Cuadrático Medio (*PECM*) para cada índice. Para obtener el valor del PECM se deben llevar a cabo los siguientes pasos:

1. Aplique la función índice a cada una de las respuestas del i -ésimo individuo ($I=1443$).
2. Obtenga el histograma de los valores del índice asociado a cada uno de los individuos de la muestra. Evalúe el nivel de asimetría de la distribución (puede ser solo visual) y obtenga:
 $\hat{\theta}$ =Mediana de la distribución si la forma es asimétrica. $\hat{\theta}$ =Media aritmética si la distribución es simétrica
3. Obtenga r ($r = 10000$) muestras con reemplazo de la muestra original (bootstrapping)
4. Para cada una de las r muestras, obtenga el $\hat{\theta}$ establecido en el numeral 2.
5. Obtenga el histograma de los r $\hat{\theta}$ del paso anterior y evaluar el nivel de asimetría
 - Si la distribución es simétrica, asuma $E[\hat{\theta}] = \mu$
 - Si la distribución es asimétrica, asuma $E[\hat{\theta}] = \tilde{\mu}$ (Mediana de la distribución de las medianas)
6. Calcule el Pseudo Error Cuadrático Medio (PECM) de la siguiente forma:
 - Distribución de los $\hat{\theta}$ con forma simétrica:

$$PECM = Var(\hat{\theta}) + \left(E[\hat{\theta}] - \mu \right)^2$$

- Distribución de los $\hat{\theta}$ con forma asimétrica:

$$PECM = Var(\hat{\theta}) + \left(E[\hat{\theta}] - \tilde{\mu} \right)^2$$

7. Se puede asumir que el estimador es insesgado si, $E[\hat{\theta}] \approx \sum_{i=1}^r \frac{\hat{\theta}_i}{r}$; es decir que, el sesgo es aproximadamente cero.
8. El criterio para seleccionar el mejor índice, es aquel que presente el menor valor de $PECM$.

5.4. Fase 3: Puntos de corte y clasificación de individuos

Puntos de corte

Una vez seleccionado el IPRG y los índices parciales con menor valor de $PECM$, se procedió a establecer dos valores (puntos de corte) que permitieran construir tres niveles de expresión de la variable latente (Percepción de Riesgo); Alto, medio y bajo. Para esto se utilizó el algoritmo k-means y métodos de remuestreo. Se parte de la base conceptual que el algoritmo k-means es un método de agrupamiento para encontrar y clasificar k grupos de elementos que comparten características semejantes, separados de los otros grupos con los que no comparten características. Los pasos llevados a cabo en esta parte del trabajo son los siguientes:

1. Usando el archivo de datos que contiene los valores estandarizados del índice seleccionado, aplique el algoritmo k-means por medio de la función *kmeans* del paquete *stats*, el cual es parte de R Core Team (2022).
2. Obtenga los valores límite (superior para el nivel más bajo e inferior para el nivel más alto de expresión de la variable) y los valores de las $k = 3$ intra varianzas arrojados por el procedimiento
3. Repita $r = 1000$ veces el paso anterior y se obtenga una matriz de k valores de puntos de corte con sus respectivas intravarianzas
4. Identifique en la matriz anterior, la menor intravarianza y sus puntos de corte asociados
5. Las categorías de clasificación del índice estarán en una escala ordinal, al formar un nivel de expresión alto usando el valor del punto de corte inferior en el cluster con mayores valores del índice. El nivel bajo de expresión estará delimitado por el intervalo de 0 hasta el mínimo observado en el cluster inmediatamente inferior al cluster formado por los elementos de expresión alta. En el medio de los dos puntos de corte se ubica el cluster de elementos con expresión media de la variable latente.

Asociación entre variables

Para caracterizar cada índice construido de los grupos o subtablas (Tabla **5-2**) con los factores socioeconómicos y socioculturales (Tabla **5-1**) se realiza un clustering para establecer las relaciones entre las variables por medio del algoritmo k -means y se clasifican los índices en Alto, Medio y Bajo.

Con las variables definidas en la Tabla **5-1**, se construyó tablas de contingencia con el objetivo de evaluar asociaciones. Se utilizó pruebas Chi cuadrado y exacta de Fisher (en casos que fue necesario), asumiendo el 5 % de nivel de significancia. Posteriormente, se realizó un ACM con las variables significativas de la tabla cruzada resultante con las pruebas mencionadas y luego un análisis de conglomerados, por medio de un cluster jerárquico para lograr clasificar los individuos similares de acuerdo a sus categorías.

5.5. Presentación de resultados

Se construye un tablero interactivo en Microsoft Power BI con los niveles de los índices seleccionados con mejor desempeño y descriptivas de la base de datos estudiada (Ver tablero en [Power BI](#)).

6. Resultados

Este capítulo está conformado por cuatro secciones: El análisis exploratorio, Análisis Factorial Múltiple (AFM), la construcción de índices por medio de dos metodologías y la asociación entre variables.

6.1. Análisis exploratorio

La muestra objeto de estudio se conformó de 1443 trabajadores y/o residentes en las ciudades de Cali y Palmira, de los cuales, 965 (66.9 %) son mujeres, 474 (32.8 %) son hombres y el resto no deseó responder. Las edades de los encuestados se encontraron en un rango intercuartílico de 28 y 46 años, con una mediana de 36 años. El nivel educativo más frecuente fue pregrado y/o posgrado (n=999; 69.2 %), seguido de técnico, bachillerato o menos (n=444; 30.7 %). El 1.80 % (n=26) son migrantes con estabilidad laboral y 0.97 % (n=14) son migrantes sin estabilidad laboral.

En la ciudad de Cali se encontró que la edad más frecuente está entre 30 y 59 años con un 62.3 % (n=501) de los encuestados, 33.4 % (n=266) son mujeres y 66.4 % (n=529) son hombres, y más del 37 % de los encuestados cuentan con educación universitaria, especialización, maestría y/o doctorado, además presentaron 49.2 % (n=392) y 35.5 % (n=283) estratos socioeconómicos entre 2-3 y 4-5 respectivamente.

Así como en Cali, en la ciudad de Palmira, también se presentó que la edad más frecuente fue entre 30 y 59 años con el 65.2 % (n=421), de igual manera, el sexo predominante fue mujeres con un 67.7 % (n=436) y el resto hombres; más del 49 % de los encuestados fueron empleados, aproximadamente el 80 % de los encuestados se encontraron en estratos 1-2 ó 3-4, coincidiendo con niveles de educación bajo, donde el 39 % (n=252) tuvieron técnico, bachillerato completo o menos.

6.2. Análisis Factorial Múltiple

Se examina la tabla de datos con las variables seleccionadas para la construcción de los índices mediante el Análisis Factorial Múltiple (AFM).

6.2.1. Análisis de Factores

De la Tabla 6-1 se observa que los dos primeros ejes recogen el 51.94 % y con los tres primeros ejes se recoge 60.75 % de varianza acumulada, generándose 16 componentes con tres grupos activos denotados como Probabilidad de contagio, Severidad y Susceptibilidad. Por esta razón, se analizan los dos primeros componentes en el transcurso de este documento.

Tabla 6-1.: Valores propios del AFM

Componente	Valor propio	% de varianza	% de varianza acumulada
comp 1	1.95	37.35	37.35
comp 2	0.76	14.58	51.94
comp 3	0.46	8.81	60.75
comp 4	0.31	5.99	66.73
comp 5	0.28	5.35	72.08
comp 6	0.24	4.68	76.76
comp 7	0.22	4.19	80.95
comp 8	0.19	3.70	84.65
comp 9	0.16	3.06	87.71
comp 10	0.14	2.72	90.43
comp 11	0.12	2.23	92.66
comp 12	0.11	2.08	94.74
comp 13	0.08	1.59	96.33
comp 14	0.07	1.37	97.70
comp 15	0.07	1.28	98.98
comp 16	0.05	1.02	100.00

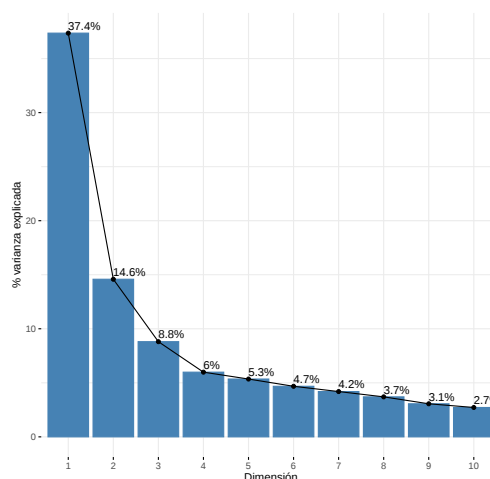


Figura 6-1.: Porcentaje de varianza

6.2.2. Ponderación de los grupos de variables

El AFM trata de equilibrar cada grupo de variables, donde a cada variable se le asigna un peso. Este peso es igual para todas las variables del mismo grupo Escofier & Pagès (1992). Los ACP separados de los tres grupos de variables han dado como primer valor propio 3.16 para las siete variables o ítems de probabilidad de contagio, 2.73 para las cuatro variables de severidad y 3.23 para las cinco variables de susceptibilidad. En el AFM, los pesos de las variables de cada uno de los tres grupos son 0.32, 0.37 y 0.31 respectivamente. Observándose que el grupo severidad presenta el mayor peso y los grupos restantes, tienen un peso similar.

Tabla 6-2.: Descomposición de la inercia de las primeras dimensiones

	Dim.1	Dim.2	Dim.3
Inercia total	1.95	0.76	0.46
Probabilidad de contagio	0.43	0.59	0.07
Severidad	0.76	0.10	0.21
Susceptibilidad	0.77	0.07	0.18

La Tabla 6-2 demuestra las relaciones que tiene cada grupo con la primera, segunda y tercera dimensión. La primera dimensión se relaciona con los grupos Severidad y Susceptibilidad, puesto que son los grupos más próximos a 1 (máximo) y así, esta dimensión genera una

dirección de inercia importante para cada grupo mencionado. La segunda dimensión está relacionada con el grupo Probabilidad de contagio; los grupos de susceptibilidad y severidad no contribuyen a esta dirección.

6.2.3. Tipología de los grupos

La Figura 6-2 se muestra la correlación de los grupos con las dos primeras dimensiones, donde la Susceptibilidad y Severidad son las que mejor están representadas en la primera dimensión. Además, se nota una relación entre los dos grupos mencionados. El grupo de Probabilidad de contagio en la segunda dimensión cuenta con las coordenadas más altas.

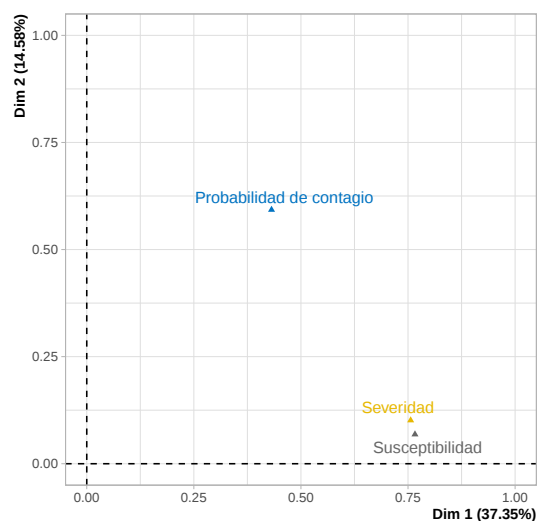


Figura 6-2.: Representación de los grupos

En la Tabla 6-3 se observan los coeficientes de correlación entre las proyecciones de los grupos y de la nube global del AFM, afirmando que la primera dimensión demostró la existencia de una dirección de dispersión casi análoga en las nubes de los grupos, puesto que los grupos presentan correlaciones altas. Por otro lado, el grupo que más se correlaciona con la segunda dimensión es la Probabilidad de contagio, concluyendo que la segunda dimensión solo es común a un grupo.

Tabla 6-3.: Correlaciones entre las proyecciones del AFM y de los grupos

Grupo	Dim.1	Dim.2	Dim.3
Probabilidad de contagio	0.66	0.78	0.44
Severidad	0.87	0.33	0.58
Susceptibilidad	0.88	0.28	0.47

Según Abdi (2007), el coeficiente RV es similar al coeficiente de correlación, oscila en un rango de 0 a 1 porque trabaja con matrices semidefinidas positivas, donde presentar un

valor de cero es que no existe relación. Por otro lado, el coeficiente Lg cuantifica el número de factores de inercia considerables de cada grupo y no presenta un límite máximo.

En la Tabla 6-4 se observa que todos los coeficientes RV son menores a 0.44, indicando una alta relación entre las variables de los grupos, por tanto los grupos son diferentes. El valor más alto se halló entre Severidad y Susceptibilidad, obteniendo un valor de 0.44.

Tabla 6-4.: Coeficientes RV

	Probabilidad de contagio	Severidad	Susceptibilidad	AFM
Probabilidad de contagio	1.00	0.12	0.12	0.62
Severidad	0.12	1.00	0.44	0.73
Susceptibilidad	0.12	0.44	1.00	0.73
MFA	0.62	0.73	0.73	1.00

6.2.4. Representación de las variables

Las interpretaciones del círculo de correlación se realizan de igual forma que en ACP o en ACM. La primera dimensión presenta una inercia de 37.4 % y la segunda dimensión, 14.6 %. Por tanto, se grafican las dimensiones 1-2 con el objetivo de tener una visión del comportamiento de las variables estudiadas.

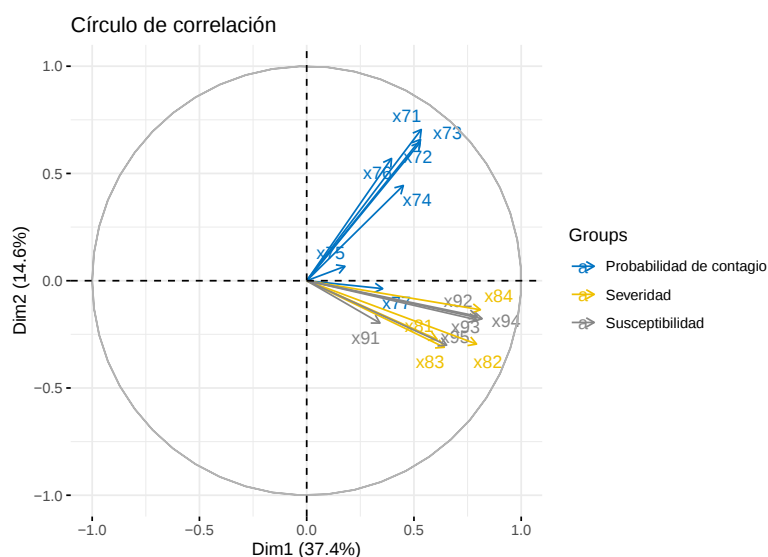


Figura 6-3.: Representación del Círculo de correlación de las variables del AFM

En la Figura 6-3 (Dimensión 1-2) se aprecian los tres grupos activos. Entre las variables activas se observa una correlación positiva entre las variables y las dimensiones. Se presenta

el factor tamaño, el cual es útil y/o necesario para construir el Índice de Percepción de Riesgo Global (IPRG). En el primer cuadrante se ubican la mayoría de las variables del grupo Probabilidad de contagio a excepción de la variable x_{77} sobre la enfermedad reiterada. En el cuarto cuadrante se encuentran todas las variables de los grupos Severidad y Susceptibilidad.

Las contribuciones de las variables a las dos primeras componentes principales pueden ser observadas en la Figura 6-4, donde los grupos de variables que más contribuyen a la primera componente principal son: Las cuatro variables de Severidad y cuatro de las cinco variables de Susceptibilidad (Por su condición de salud, probabilidad de afectarse de COVID-19, por su edad, probabilidad de afectarse por COVID-19, por su salud actual, probabilidad de afectarse por COVID-19 en comparación con personas de su edad y gravedad para usted de contraer la enfermedad COVID-19).

De acuerdo a la segunda dimensión, las variables del grupo que más contribuye a esta dimensión, pertenecen a la Probabilidad de contagio, las cuales son: Probabilidad de afectarse en los próximos 2 meses de COVID-19, probabilidad de familiares/amigos sean diagnosticados con COVID-19 en los próximos 2 meses, sea diagnosticado con COVID-19 en 2 meses en comparación con personas de su edad, probabilidad de la mayoría de la población se contagie de COVID-19 y probablemente me enfermaré con el coronavirus / COVID-19.

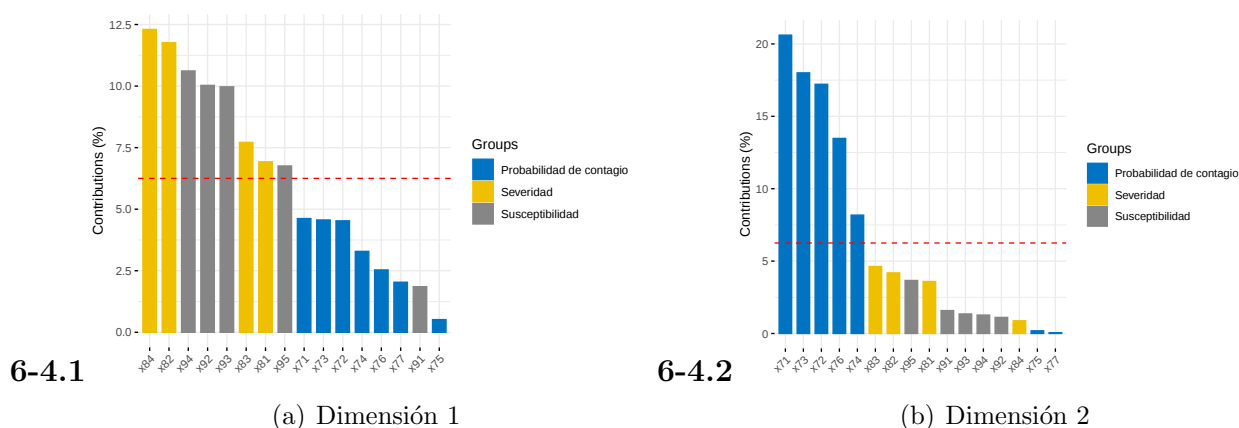


Figura 6-4.: Contribución de las variables cuantitativas

En la Tabla 6-5 se presentan las coordenadas y contribuciones de las variables en las dos primeras dimensiones. Se observa que la variable Probabilidad de sufrir una enfermedad grave por COVID-19 comparado con personas de su edad es la que más contribuye al primer eje factorial en el grupo de Percepción de riesgo: Severidad. De igual manera, las variables Probabilidad de enfermarse gravemente si usted fuera diagnosticado con COVID-19 y; por

su salud actual, probabilidad de afectarse por COVID-19 en comparación con personas de su edad.

De acuerdo a la segunda dimensión, las variables que más contribuyen hacen parte del grupo de Probabilidad de contagio, las cuales son: Probabilidad de afectarse en los próximos 2 meses de COVID-19, probabilidad de familiares/amigos sean diagnosticados con COVID-19 en los próximos 2 meses y sea diagnosticado con COVID-19 en 2 meses en comparación con personas de su edad.

Tabla 6-5.: Coordenadas y contribuciones de las variables

	Coordenadas		Contribuciones	
	Dim.1	Dim.2	Dim.1	Dim.2
x71	0.53	0.70	4.62	20.61
x72	0.53	0.64	4.53	17.21
x73	0.53	0.66	4.56	18.00
x74	0.45	0.44	3.29	8.18
x75	0.18	0.07	0.52	0.19
x76	0.40	0.57	2.53	13.48
x77	0.35	-0.04	2.03	0.06
x81	0.61	-0.27	6.92	3.59
x82	0.79	-0.29	11.76	4.19
x83	0.64	-0.31	7.71	4.63
x84	0.81	-0.14	12.30	0.88
x91	0.34	-0.20	1.85	1.58
x92	0.80	-0.16	10.03	1.11
x93	0.79	-0.18	9.97	1.35
x94	0.82	-0.18	10.62	1.28
x95	0.65	-0.30	6.75	3.66

6.2.5. Representación de los individuos

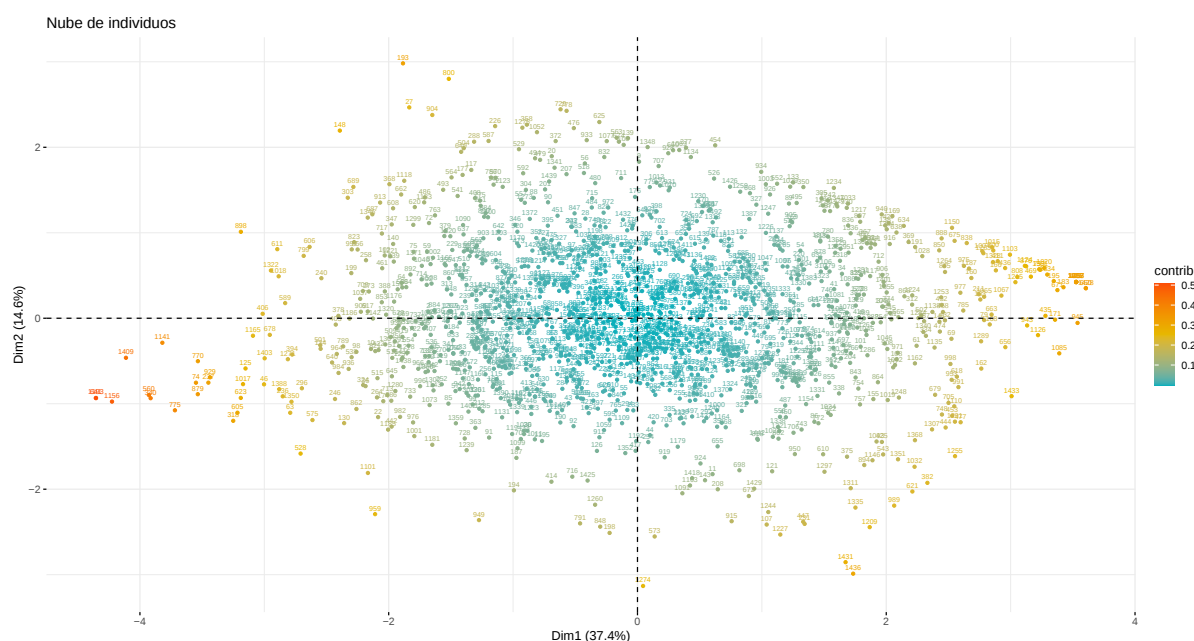


Figura 6-5.: Nube de individuos del AFM

La primera dimensión se asocia con la Severidad y Susceptibilidad, mientras que la segunda dimensión se asocia con la probabilidad de contagio. Se presenta la nube de los individuos en el primer plano factorial mediante la Figura 6-5, donde los individuos 641, 1193, 1156, 1409, 560, 340 y 1141 cuentan con las contribuciones más altas en la construcción de la primera dimensión (superiores a 50 %), estos individuos se sienten confiados y/o seguros, es decir que presentan un bajo perfil de riesgo y sus indicadores frente a la enfermedad COVID-19 fueron buenos, mientras que los individuos 945, 1428, 650, 183 (extrema derecha), presentan un perfil de percepción de riesgo alto o mayor.

En la segunda dimensión, los individuos que aportan en gran medida a la conformación de esta dimensión son 1274, 1436, 193, 1431, 800, 573, 1227 y 198, con unas contribuciones superiores al 55 % y son individuos con una percepción media, es decir, se perciben con riesgo de COVID-19 y de igual forma cuida su salud.

6.2.6. Proyección de ejes parciales de cada grupo

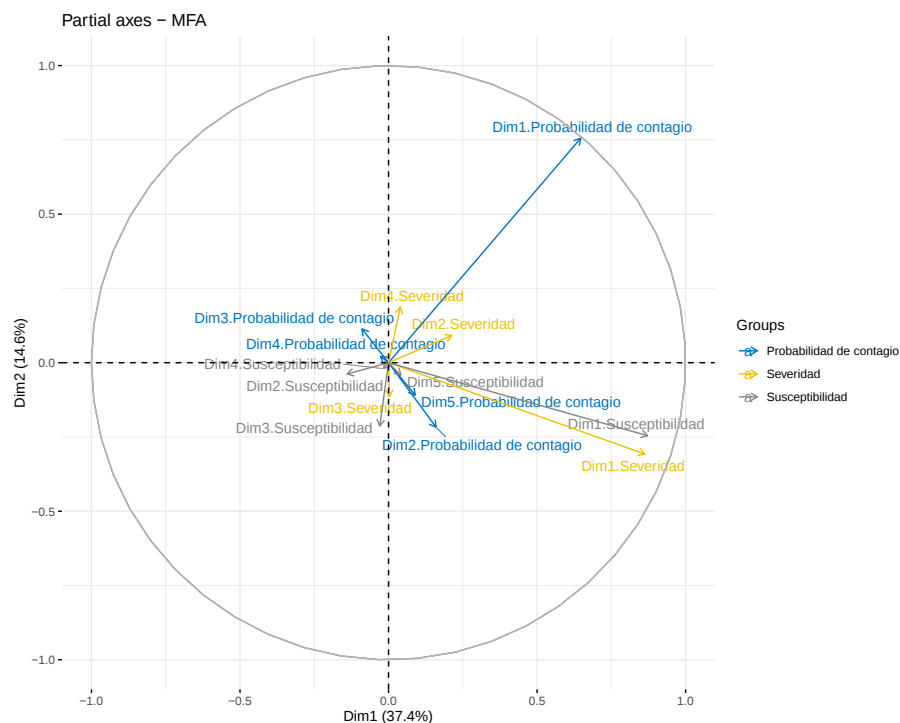


Figura 6-6.: Contribución de los grupos a las dimensiones

Mediante la Figura 6-6 se muestra que la primera dimensión se correlaciona con la primera dimensión de los grupos de las variables. Para la segunda dimensión se revela muy poca la variabilidad presente en la estructura de los grupos, sin embargo el grupo que más se resalta en esta dimensión es Probabilidad de contagio. Las dimensiones tres, cuatro y cinco

no cuentan con una relación relevante en este caso. Por tanto, con la primera dimensión se construirá el índice a continuación.

6.3. Formulación de un índice sintético

6.3.1. Índices usando métodos multivariados de la estadística

Índices por dimensión

Se construyó un índice¹ de cada individuo i para cada grupo de variables cuantitativas que conformaron el AFM, a través de las siguientes expresiones, seguido de la metodología planteada en la sección 5.2:

$$IPPC_i = 0,202x_{i71} + 0,192x_{i72} + 0,193x_{i73} + 0,150x_{i74} + 0,044x_{i75} + 0,158x_{i76} + 0,061x_{i77}$$

$$IPSE_i = 0,245x_{i81} + 0,251x_{i82} + 0,256x_{i83} + 0,248x_{i84}$$

$$IPSU_i = 0,134x_{i91} + 0,231x_{i92} + 0,224x_{i93} + 0,228x_{i94} + 0,183x_{i95}$$

Se tomó la primera componente principal de su respectivo grupo en el análisis separado, posteriormente se estimaron los pesos relativos de las variables de cada grupo, donde su respectiva suma da uno y se transforman los valores del vector propio asociado a la primer componente principal, por medio de la división de cada valor, sobre la suma de todos ellos y multiplicado por 100; se transformó en una escala de 0 a 100 para facilitar las interpretaciones como se denotó en el capítulo 5.

De acuerdo a la Tabla **6-6**, el IPSU presentó la mediana más baja en comparación con los demás índices, seguido de IPSE; esto se presentó tanto en las ciudades de Palmira, Cali, como en conjunto, demostrando que los individuos tienden a sentirse o percibirse más propensos y con una mayor probabilidad de presentar un evento adverso frente a la enfermedad COVID-19 en el periodo analizado.

Índice de Percepción de Riesgo Global (IPRG)

Después de haber realizado el AFM, se procedió a construir el Índice de Percepción de Riesgo Global (IPRG) puesto que en la Figura **6-3** todas las variables tienen una misma dirección, lo cual indica la existencia del factor tamaño. De acuerdo al vector propio de la primera componente principal, se obtuvieron los pesos relativos de las variables empleadas en el análisis anterior, donde el IPRG se formó así,

1

¹ Índice de Percepción de Probabilidad de Contagio (IPPC), Índice de Percepción de Severidad (IPSE), Índice de Percepción de Susceptibilidad (IPSU).

$$IPRG_i = 0,058x_{i71} + 0,057x_{i72} + 0,057x_{i73} + 0,049x_{i74} + 0,019x_{i75} + 0,042x_{i76} + 0,038x_{i77} + \\ 0,066x_{i81} + 0,086x_{i82} + 0,069x_{i83} + 0,088x_{i84} + 0,037x_{i91} + 0,086x_{i92} + 0,086x_{i93} + \\ 0,089x_{i94} + 0,071x_{i95}.$$

Con la metodología planteada en la sección 5.2, el índice del AFM como Índice de Percepción de Riesgo Global (IPRG) transformado mantuvo las siguientes medidas de tendencia central para los habitantes y/o trabajadores de las ciudades de Cali y Palmira observadas en la Tabla 6-6.

6.3.2. Índice usando información de expertos

Índice por dimensión

De acuerdo a las ponderaciones asignadas por el especialista (Tabla 5-4), se construyeron los índices para cada una de las dimensiones que conforman la Percepción de Riesgo Global, mediante las siguientes expresiones

$$IPPC_i = 0,164x_{i71} + 0,145x_{i72} + 0,164x_{i73} + 0,127x_{i74} + 0,109x_{i75} + 0,145x_{i76} + 0,145x_{i77} \\ IPSE_i = 0,241x_{i81} + 0,276x_{i82} + 0,276x_{i83} + 0,207x_{i84} \\ IPSU_i = 0,222x_{i91} + 0,222x_{i92} + 0,194x_{i93} + 0,167x_{i94} + 0,194x_{i95}$$

Índice de Percepción de Riesgo Global (IPRG)

Se procedió a construir el Índice de Percepción de Riesgo Global (IPRG) empleando las ponderaciones asignadas por la investigadora tanto para las i -ésimas dimensiones y sus respectivos j -ésimos factores, mediante la siguiente expresión

$$IPRG_i = 0,061x_{i71} + 0,055x_{i72} + 0,061x_{i73} + 0,048x_{i74} + 0,041x_{i75} + 0,055x_{i76} + 0,055x_{i77} + \\ 0,080x_{i81} + 0,092x_{i82} + 0,092x_{i83} + 0,069x_{i84} + 0,065x_{i91} + 0,065x_{i92} + 0,057x_{i93} + \\ 0,049x_{i94} + 0,057x_{i95}$$

La distribución de los puntajes de los individuos de acuerdo al IPRG, observados a través del histograma (Ver Figuras A-1 y A-2) en ambas metodologías, es asimétrico; por lo tanto, la medida de tendencia central es la mediana. A partir del *PECM*, el IPRG con el menor valor del error cuadrático medio fue el formado mediante el AFM tanto para las ciudades por separado como para las dos ciudades juntas. Por lo tanto, se seleccionó finalmente el índice construido con AFM como IPRG, el cual presentó las medidas de tendencia central para los individuos observadas en la Tabla 6-6. Para el índice de Severidad, se presentó la misma situación anterior; sin embargo, para IPPC y IPSU se presentó el menor error cuadrático medio con la metodología heurística.

Tabla 6-6.: Medidas de los Índices por dimensión

Metodología	Ciudad	Índice	Q1	Mediana	Media	Q3	Desv.Est	PECM
AFM	Cali	IPPC	49.02	60.59	61.03	73.73	19.25	0.72
		IPSE	37.60	50.26	52.99	66.90	21.99	0.83
		IPSU	21.30	32.61	33.82	45.43	17.80	0.11
		IPRG	39.76	50.89	52.05	63.37	17.59	0.01
	Palmira	IPPC	51.83	64.84	64.67	77.22	19.26	3.33
		IPSE	45.23	58.40	57.86	71.08	21.75	0.53
		IPSU	22.68	35.42	36.23	48.33	17.59	0.79
		IPRG	44.15	57.09	57.03	69.84	18.04	0.28
	Cali y Palmira	IPPC	48.23	60.96	61.38	74.47	19.63	1.30
		IPSE	41.66	54.28	55.17	70.83	22.01	0.03
		IPSU	21.91	34.02	34.30	45.38	17.50	0.15
		IPRG	41.24	53.47	54.10	66.20	17.83	0.03
Heurística	Cali	IPPC	49.10	60.53	61.01	74.21	18.75	0.10
		IPSE	37.35	50.58	52.76	66.67	22.01	152.06
		IPSU	17.26	29.06	30.93	42.53	17.11	0.35
		IPRG	42.48	52.90	53.75	64.29	16.43	8.15
	Palmira	IPPC	50.98	64.30	63.98	76.94	19.47	32.63
		IPSE	44.39	58.05	57.79	71.85	21.81	10.52
		IPSU	27.41	43.75	43.88	57.69	21.04	750.59
		IPRG	45.18	57.71	57.78	70.11	17.59	0.19
	Cali y Palmira	IPPC	49.98	62.57	62.52	75.69	18.68	0.64
		IPSE	40.80	54.02	55.12	70.41	22.08	39.34
		IPSU	22.31	33.46	34.43	45.41	17.35	0.05
		IPRG	43.79	54.97	55.55	66.98	17.07	0.12

En relación con los puntos de corte (Tabla 6-7), el nivel de IPRG bajo quedó definido por ciudadanos que presentan un índice de percepción de riesgo de entre 0 y 44.26, el nivel medio por ciudadanos con un IPRG mayor a 44.26 pero menor a 66.05 y el nivel alto por ciudadanos cuya percepción de riesgo va de 66.05 a 100. En el caso de los 1443 encuestados, se presentó un individuo con un IPRG de 0, el cual indica que presenta una percepción de total seguridad o improbabilidad de adquirir la enfermedad COVID-19.

Tabla 6-7.: Valores mínimos y máximos por nivel de Índice.

Índice	Nivel de percepción	Mínimo	Máximo	Cali n (%)	Palmira n (%)	Cali y Palmira n (%)
IPPC (Heurística)	Bajo	0.00	<49.60	199 (25.1)	143 (22.0)	342 (23.7)
	Medio	≥ 49.60	<72.69	378 (47.6)	291 (44.8)	669 (46.4)
	Alto	≥ 72.69	100.0	217 (27.3)	215 (33.1)	432 (29.9)
IPSE	Bajo	0.00	<41.53	220 (27.7)	133 (20.5)	353 (24.5)
	Medio	≥ 41.53	<70.50	385 (48.5)	331 (51.0)	716 (49.6)
	Alto	≥ 70.50	100.0	189 (23.8)	185 (28.5)	374 (25.9)
IPSU (Heurística)	Bajo	0.00	<27.48	321 (40.4)	211 (32.5)	532 (36.9)
	Medio	≥ 27.48	<49.38	341 (42.9)	290 (44.7)	631 (43.7)
	Alto	≥ 49.38	100.0	132 (16.6)	148 (22.8)	280 (19.4)
IPRG	Bajo	0.00	<44.26	254 (32.0)	179 (27.6)	433 (30.0)
	Medio	≥ 44.26	<66.05	373 (47.0)	273 (42.1)	646 (44.8)
	Alto	≥ 66.05	100.0	167 (21.0)	197 (30.3)	364 (25.2)

La distribución de los participantes de las ciudades en conjunto (Cali y Palmira) presentó las frecuencias más altas del nivel del IPRG en medio y bajo, con 44.8 % (n=646) y 30.0 % (n=433) respectivamente; esta misma tendencia se encontró en la ciudad de Cali. Por otro lado, en la ciudad de Palmira, la percepción de riesgo se presentó en mayor medida entre niveles medio y alto, demostrando así, que en los individuos de Palmira se sienten un poco más vulnerables y/o inseguros frente a la enfermedad COVID-19, que los individuos de la ciudad de Cali sin presentar diferencias significativas.

6.4. Asociaciones entre variables

De acuerdo a los puntos de corte establecidos, se realizó la clasificación de los participantes de las ciudades Cali y Palmira en conjunto, con las respectivas características sociodemográficas y el IPRG por niveles. Con la prueba chi-cuadrado y un nivel de significancia de 0.05 se demostró que el grupo etario, estrato, nivel de educación, ingresos y ocupación están relacionados con IPRG.

Tabla 6-8.: Índice de Percepción de Riesgo Global en Cali y Palmira

Características		Bajo n=433 (%)	Medio n=646 (%)	Alto n=364 (%)	p-valor
Grupo etario	De 18 a 29 años	171 (39.9)	175 (40.8)	83 (19.3)	0.000
	De 30 a 59 años	247 (26.8)	429 (46.5)	246 (26.7)	
	Más de 60 años	15 (16.3)	42 (45.6)	35 (38.0)	
Género	Hombre	163 (34.4)	204 (43.0)	107 (22.6)	0.030
	Mujer	269 (27.9)	439 (45.5)	257 (26.6)	
Raza	Blanco/Mestizo	394 (30.1)	593 (45.3)	322 (24.6)	0.209
	Otro	39 (29.1)	53 (39.6)	42 (31.3)	
Área de residencia	Rural	25 (22.3)	52 (46.4)	35 (31.2)	0.122
	Urbana	408 (30.6)	594 (44.6)	329 (24.7)	
Estrato	1 y 2	101 (25.0)	176 (43.6)	127 (31.4)	0.002
	2 y 3	219 (31.3)	309 (44.1)	172 (24.6)	
	3 y 4	113 (33.3)	161 (47.5)	65 (19.2)	
Nivel educativo	Esp/Maestría/Doctorado	144 (32.1)	218 (48.7)	86 (19.2)	0.001
	Técnico/Bachillerato o menos	124 (27.9)	180 (40.5)	140 (31.5)	
	Universitario	165 (29.9)	248 (45.0)	138 (25.0)	
Ingresos	> 3 SMMLV	174 (32.5)	253 (47.2)	109 (20.3)	0.000
	< 1 SMMLV	141 (35.1)	166 (41.3)	95 (23.6)	
	1 SMMLV a 3 SMMLV	118 (23.4)	227 (45.0)	160 (31.7)	
Ocupación	Jubilado/Desempleado	54 (24.5)	93 (42.3)	73 (33.2)	0.000
	Independiente	89 (33.6)	113 (42.6)	63 (23.8)	
	Estudiante	89 (43.6)	95 (46.6)	20 (9.8)	
	Empleado	201 (26.6)	345 (45.7)	208 (27.6)	

En la Tabla 6-8 se observan las mayores puntuaciones en los niveles de medio y alto, revelando la presencia de riesgo de contraer la enfermedad en gran parte de las características sociodemográficas de la población estudiada. La población con la percepción de riesgo más baja son los jóvenes de 18 a 29 años y los estudiantes.

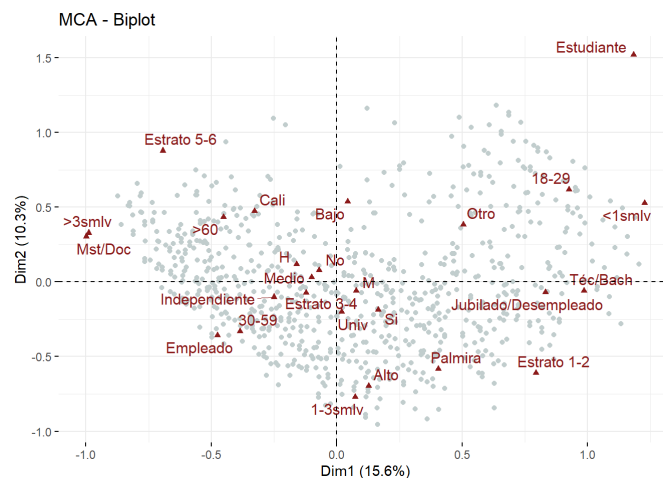


Figura 6-7.: Plano factorial simultáneo. M=Mujer, H=Hombre.

Se realizó el ACM con las variables significativas de la Tabla 6-8 para complementar el análisis y de acuerdo a la Figura 6-7 se alcanzó a recoger un 25.94 % donde se puede observar una relación entre las categorías de las variables sociodemográficas y la clasificación del IPRG, demostrando una diferencia de percepción de riesgo en los municipios de Palmira y Cali. Las categorías que más contribuyen a la primera componente son ingresos menores a 1 SMLV, ingresos mayores a 3 SMLV, nivel de estudios Maestría/ Doctorado y Técnico/Bachillerato, seguidos de edad entre 18 y 29 años y estrato 1 y 2. Para la segunda componente las categorías que más contribuyen son ocupación estudiante, ingresos entre 1 y 3 SMLV, estrato 5 y 6, edades entre 18-29 años seguidas de los municipios Palmira y Cali.

Mediante un análisis cluster con clasificación jerárquica de las variables sociodemográficas expuestas en la Tabla 6-8 y el IPRG con los niveles construidos en la Tabla 6-7; se obtuvieron los siguientes tres clusters teniendo en cuenta los valores test (Anexo A):

- **Cluster 1 (Negro):** Individuos con las puntuaciones del IPRG más altas y pertenecientes a la clase alta de la ciudad de Cali (ingresos mayores o iguales a 3 SMLV, nivel educativo Maestría/Doctorado, estrato 5 y 6, edades mayores a los 30 años y del municipio de Cali).
- **Cluster 2 (Rojo):** Individuos con puntuaciones en el IPRG altas y pertenecientes a la clase media-alta de la ciudad de Palmira (ingresos entre 1 y 3 SMLV, ocupación empleados, nivel educativo universitario, estrato 3 y 4, edades entre 30 y 59 años y pertenecientes al municipio de Palmira).
- **Cluster 3 (Verde):** Individuos con puntuaciones en el IPRG bajas-medias y pertenecientes a la clase baja de Cali y Palmira (ingresos menores a 1 SMLV, ocupación estudiante, edades entre 18 y 29 años, educación Técnico/Bachillerato completo/Bachillerato incompleto, estratos 1 y 2).

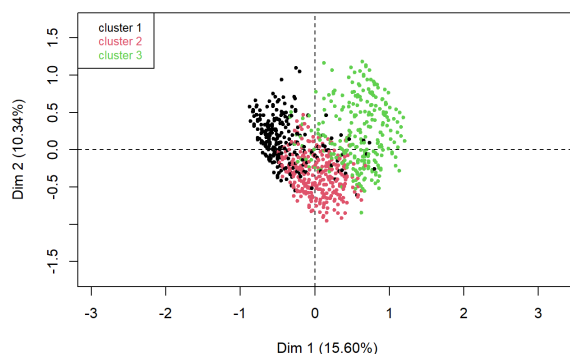


Figura 6-8.: Identificación de tipos de individuos (clústers) según sus características sociodemográficas y los niveles del IPRG.

7. Conclusiones y recomendaciones

7.1. Conclusiones

Evaluar la variable latente percepción de riesgo frente a la enfermedad COVID-19 es una tarea compleja para los profesionales y/o investigadores de la salud, puesto que no es una variable medible directamente, por lo tanto se emplea la estimación estadística para lograr una aproximación de lo que sería la variable latente, dado que no se cuenta con formas distribucionales para la expresión de la misma.

Frente a la mencionada situación problema, este trabajo de grado permitió cuantificar la percepción de riesgo global en individuos de las ciudades de Cali y Palmira en uno de los picos más altos de dicha enfermedad (Enero-Abril del 2021), mediante dos metodologías de construcción de índices, una empleando la estadística multivariada con el Análisis Factorial Múltiple y otra, de manera heurística; siendo adaptativo para ser replicado por otros municipios y/o departamentos.

También fue posible construir índices de Probabilidad de contagio (IPPC), Severidad (IPSE) y Susceptibilidad (IPSU), las cuales son dimensiones que conforman la percepción de riesgo.

Todos los resultados obedecen de un análisis empírico utilizando metodologías no paramétricas, donde se halló el índice con mejor desempeño en cada una de las situaciones de acuerdo al Pseudo Error Cuadrático Medio (PECM), donde el índice de Percepción de Riesgo Global seleccionado fue con la metodología del AFM, el cual es más estable y robusta porque presenta una menor varianza, puesto que construye pesos a partir de las respuestas de los individuos; al igual que el índice de Percepción de Severidad; mientras que los índices de Probabilidad de Contagio y Susceptibilidad, presentaron mejor desempeño con la metodología heurística. Posteriormente, para construir el índice de Percepción de Riesgo Global se establecieron puntos de corte, pruebas chi-cuadradas y clasificación de individuos, los cuales serán útiles para realizar intervenciones por parte de las entidades de salud y gubernamentales de los municipios estudiados.

Se logró un comportamiento similar del IPRG tanto en las ciudades por separado, como en conjunto. Por medio del Análisis cluster, se clasificaron los individuos de acuerdo al IPRG

y las variables sociodemográficas, donde se evidenció una coherencia entre las agrupaciones del índice y las características de los individuos; los ciudadanos de clase baja (menor nivel educativo y menor estrato socioeconómico), presentaron un nivel bajo de percepción de riesgo frente a la enfermedad COVID-19 en ambas ciudades; a diferencia de los ciudadanos de clase alta (estratos socioeconómicos altos y nivel de estudio superior), presentan un nivel de percepción de riesgo frente a la COVID-19 alto.

Se debe tener en cuenta a nivel estadístico que, haber construido el IPRG por medio de la metodología del AFM presenta una gran cantidad de ventajas sobre la metodología heurística, tales como: Reducir la información de gran cantidad de variables en pocos índices, genera una comprensión del problema más clara, disminuye el ruido de los datos originales; los factores nuevos pueden ser utilizados como un suministro en próximos procedimientos, sin redundancia de información.

Operativamente es más sencillo construir índices desde los métodos heurísticos si se cuenta con el apoyo del profesional en salud todo el tiempo, de lo contrario, los métodos multivariados son la mejor alternativa para el análisis de los datos.

7.2. Recomendaciones

Durante la realización del presente trabajo de grado se pudo encontrar algunas inconsistencias en los datos, por lo tanto se recomienda realizar un formulario más corto, con preguntas más concretas sin la opción de otro con respuesta abierta; sería bueno hacer un estudio de validación de las propiedades psicométricas del cuestionario por medio de un Análisis Factorial Confirmatorio, con el fin de ver si se reduce la cantidad de preguntas.

Para medir la percepción de riesgo de los individuos frente a la enfermedad COVID-19, se recomienda emplear las mismas 16 preguntas utilizadas en este proyecto para obtener el IPRG y así, generar comparaciones de la población estudiada en el tiempo.

En caso de repetirse el estudio, podría realizarse un muestreo por conglomerados o un muestreo estratificado. También se recomienda para futuras investigaciones construir índices a través de otras metodologías como PLS o SEM para pretender obtener un menor error cuadrático medio. También se recomienda realizar una metodología asumiendo métodos paramétricos y una forma distribucional con el objetivo de realizar clusters paramétricos.

A. Anexo: Resultados

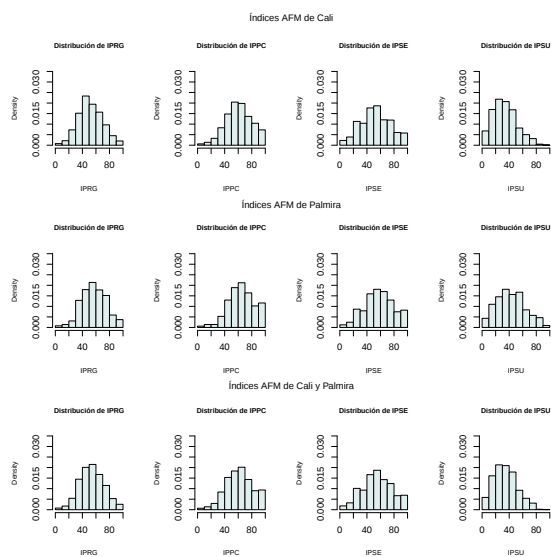


Figura A-1.: Distribuciones de los índices AFM

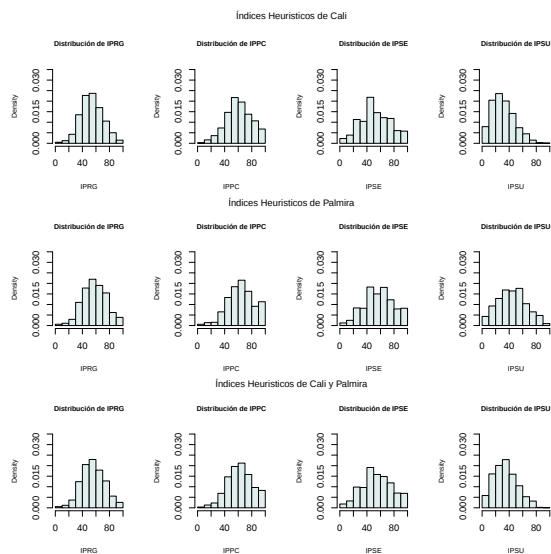


Figura A-2.: Distribuciones de los índices heurísticos

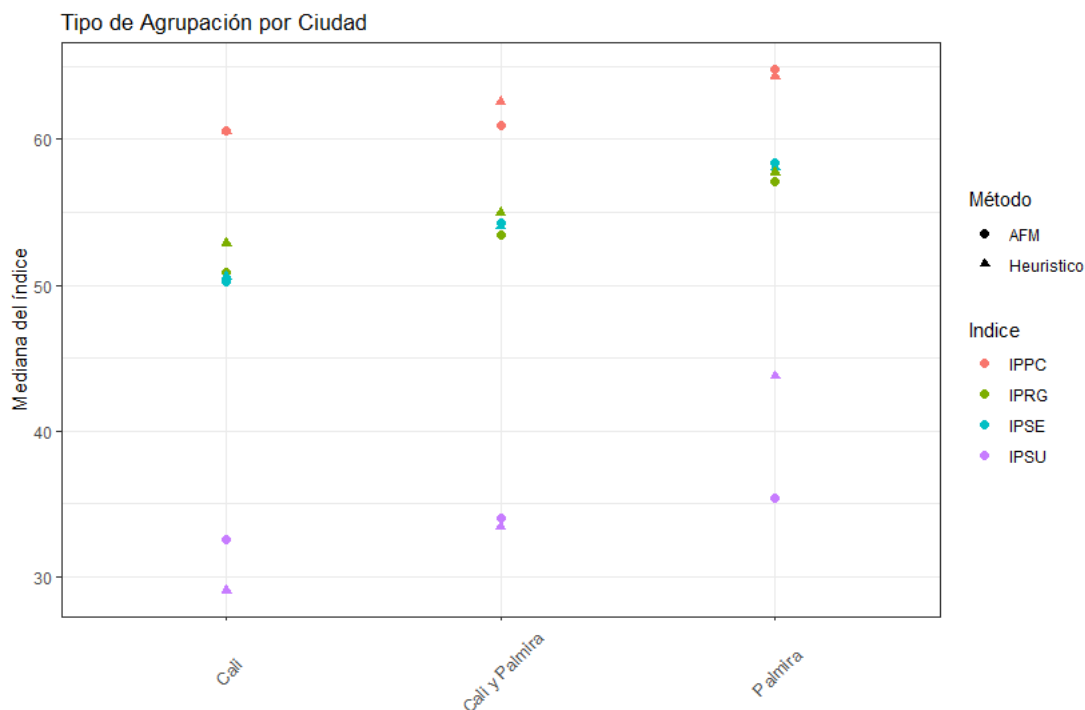


Figura A-3.: Mediana de cada índice construido en las respectivas ciudades

Tabla A-1.: Descripción de cluster 1

Característica	Cla/Mod	Mod/Cla	Global	valor p	v.test
Ingresos=>3smlv	83.582090	86.990291	37.144837	2.763024e-202	30.346871
Educacion=Mst/Doc	85.491071	74.368932	31.046431	2.156318e-160	26.986024
Estrato=Estrato 5-6	70.501475	46.407767	23.492723	2.895141e-51	15.061640
Edad_g=>60	96.739130	17.281553	6.375606	3.086511e-38	12.929039
Municipio=Cali	48.055207	74.368932	55.232155	2.032146e-28	11.056802
Edad_g=30-59	43.817787	78.446602	63.894664	2.224735e-18	8.745275
Ocu=Empleado	44.488712	65.048544	52.182952	2.469413e-13	7.320558
Sexo=H	44.514768	40.970874	32.848233	1.206921e-06	4.854498
ViveConEnferCronicas=No	38.180020	74.951456	70.062370	2.385014e-03	3.037561
Ocu=Independiente	42.264151	21.747573	18.364518	1.436419e-02	2.448024
ViveConEnferCronicas=Si	29.861111	25.048544	29.937630	2.385014e-03	-3.037561
Ocu=Jubilado/Desempleado	26.244344	11.262136	15.315315	1.225439e-03	-3.232892
IPRG=Alto	28.021978	19.805825	25.225225	3.625769e-04	-3.565924
Sexo=M	31.502591	59.029126	66.874567	2.864759e-06	-4.680285
Educacion=Univ	19.782214	21.165049	38.184338	3.976028e-24	-10.132200
Municipio=Palmira	20.433437	25.631068	44.767845	2.032146e-28	-11.056802
Ocu=Estudiante	4.901961	1.941748	14.137214	7.974452e-29	-11.140418
Ingresos=1-3smlv	11.089109	10.873786	34.996535	1.827993e-51	-15.092006
Estrato=Estrato 1-2	7.425743	5.825243	27.997228	6.910301e-52	-15.156049
Edad_g=18-29	5.128205	4.271845	29.729730	2.840273e-67	-17.329041
Educacion=Téc/Bach	5.180180	4.466019	30.769231	4.565094e-70	-17.695212
Ingresos=<1smlv	2.736318	2.135922	27.858628	7.124286e-75	-18.308155

Tabla A-3.: Descripción de cluster 3

Característica	Cla/Mod	Mod/Cla	Global	valor p	v.test
Ingresos=<1smly	90.547264	86.0520095	27.858628	1.422635e-227	32.206688
Ocu=Estudiante	91.666667	44.2080378	14.137214	2.671355e-93	20.489498
Edad_g=18-29	64.568765	65.4846336	29.729730	2.374451e-78	18.739138
Educacion=Téc/Bach	61.036036	64.0661939	30.769231	4.935667e-67	17.297229
Ocu=Jubilado/Desempleado	69.683258	36.4066194	15.315315	2.812254e-42	13.625822
Estrato=Estrato 1-2	48.267327	46.0992908	27.997228	6.604890e-22	9.619660
Municipio=Palmira	35.758514	54.6099291	44.767845	1.388987e-06	4.826578
IPRG=Bajo	37.875289	38.7706856	30.006930	3.981743e-06	4.612333
Sexo=M	32.435233	73.9952719	66.874567	1.854597e-04	3.738041
IPRG=Alto	25.274725	21.7494090	25.225225	4.911288e-02	-1.967610
IPRG=Medio	25.851393	39.4799054	44.767845	9.227392e-03	-2.603512
Ocu=Independiente	20.377358	12.7659574	18.364518	2.958028e-04	-3.618948
Sexo=H	22.784810	25.5319149	32.848233	1.173631e-04	-3.851570
Estrato=Estrato 5-6	20.058997	16.0756501	23.492723	1.198242e-05	-4.377907
Educacion=Univ	22.504537	29.3144208	38.184338	6.588344e-06	-4.506572
Municipio=Cali	24.090339	45.3900709	55.232155	1.388987e-06	-4.826578
Estrato=Estrato 3-4	22.857143	37.8250591	48.510049	1.584836e-07	-5.242421
Edad_g=>60	3.260870	0.7092199	6.375606	5.245939e-11	-6.563783
Educacion=Mst/Doc	6.250000	6.6193853	31.046431	4.948088e-45	-14.081314
Ingresos=1-3smly	7.524752	8.9834515	34.996535	1.725981e-46	-14.316501
Edad_g=30-59	15.509761	33.8061466	63.894664	5.426837e-52	-15.171917
Ingresos=>3smly	3.917910	4.9645390	37.144837	8.326091e-72	-17.919373
Ocu=Empleado	3.718459	6.6193853	52.182952	2.752197e-124	-23.711374

Tabla A-2.: Descripción de cluster 2

Característica	Cla/Mod	Mod/Cla	Global	valor p	v.test
Ingresos=1-3smly	81.386139	81.386139	34.996535	1.821093e-169	27.748412
Ocu=Empleado	51.792829	77.227723	52.182952	4.386362e-46	14.251519
Educacion=Univ	57.713249	62.970297	38.184338	1.182407e-45	14.182108
Municipio=Palmira	43.808050	56.039604	44.767845	2.860689e-10	6.306146
Edad_g=30-59	40.672451	74.257426	63.894664	1.167717e-09	6.084615
Estrato=Estrato 3-4	42.000000	58.217822	48.510049	6.178149e-08	5.413567
IPRG=Alto	46.703297	33.663366	25.225225	9.257423e-08	5.340728
Estrato=Estrato 1-2	44.306931	35.445545	27.997228	4.790900e-06	4.573742
ViveConEnferCronicas=Si	40.509259	34.653465	29.937630	4.368762e-03	2.850230
Edad_g=18-29	30.303030	25.742574	29.729730	1.465903e-02	-2.440695
ViveConEnferCronicas=No	32.640950	65.346535	70.062370	4.368762e-03	-2.850230
IPRG=Bajo	23.556582	20.198020	30.006930	1.272217e-09	-6.070867
Municipio=Cali	27.854454	43.960396	55.232155	2.860689e-10	-6.306146
Edad_g=>60	0.000000	0.000000	6.375606	1.183326e-18	-8.816273
Ocu=Estudiante	3.431373	1.386139	14.137214	2.056435e-31	-11.659334
Ocu=Jubilado/Desempleado	4.072398	1.782178	15.315315	1.636841e-32	-11.872905
Estrato=Estrato 5-6	9.439528	6.336634	23.492723	5.364841e-34	-12.155476
Ingresos=>3smly	12.500000	13.267327	37.144837	5.082929e-47	-14.401232
Ingresos=<1smly	6.716418	5.346535	27.858628	1.396016e-52	-15.260765
Educacion=Mst/Doc	8.258929	7.326733	31.046431	2.985654e-53	-15.361079

B. Anexo: Cuestionarios



COVID-19: ¿CÓMO PERCIBEN SU RIESGO LOS CALEÑOS Y LAS CALEÑAS?

Estimado (a) participante,

El estudio "Percepción de riesgo COVID-19 y sus factores relacionados en Cali, 2020-2021" es conducido por la Pontificia Universidad Javeriana Cali. El objetivo es identificar el nivel de percepción de riesgo que la población tiene frente a las posibilidades sufrir de esta enfermedad y los factores que puedan influenciar dicha percepción. Estamos invitando a participar entre 1.000 y 6.000 personas mayores de edad que vivan o trabajen en la ciudad de Cali. Le tomará aproximadamente 15 minutos responder esta encuesta, cuyas preguntas están relacionadas con sus características sociales y económicas y la salud. Se realizarán 3 mediciones de la percepción de la población durante 6 meses. Usted puede participar una o más veces, cada vez que reciba la encuesta. Esta información podría ayudar a los gobernantes y a la ciudadanía a implementar intervenciones específicas para mantener la epidemia con bajos niveles de transmisión en este periodo pos-cuarentena.

Más información:

1. Su participación es voluntaria y anónima puede cancelar o dejar de diligenciar la encuesta en cualquier momento sin necesidad de dar explicaciones
2. Su participación en la investigación no lo expone a ningún riesgo significativo. Las preguntas no son sobre temas sensibles. Sus respuestas pueden ayudar para mejorar la respuesta a la epidemia en beneficio de la población
3. Los datos serán tratados con total confidencialidad
4. Sus respuestas serán usadas exclusivamente con propósitos científicos.
5. Los datos anonimizados se almacenarán en bases de datos custodiadas por investigadores en la universidad. No es posible sacar conclusiones de cada una de las personas, los resultados serán reportados para los grupos poblacionales. Los datos podrán utilizarse para investigación y docencia en el futuro.
6. El estudio no es comercial y es financiado por la Pontificia Universidad Javeriana Cali. Usted no recibirá pago por su participación.

¿Cómo puede contactarnos?

Responsables:

Sheila Gómez. sagomez@javerianacali.edu.co (<mailto:sagomez@javerianacali.edu.co>). 3186996218

Diana Caicedo, diana.caicedob@javerianacali.edu.co
(<mailto:diana.caicedob@javerianacali.edu.co>). 3127217092

Miembros del comité de ética: etica.investigación@javerianacali.edu.co
(<mailto:etica.investigación@javerianacali.edu.co>).

🔗 AL FINAL ENCONTRARÁ INFORMACIÓN DE INTERÉS SOBRE LA ENFERMEDAD COVID-19



COVID-19: ¿CÓMO PERCIBEN SU RIESGO LOS PALMIRANOS?

Estimado (a) participante,

El estudio "Percepción de riesgo COVID-19 y sus factores relacionados en Cali y en Palmira, 2020-2021" es conducido por la Pontificia Universidad Javeriana Cali. El objetivo es identificar el nivel de percepción de riesgo que la población tiene frente a las posibilidades sufrir de esta enfermedad y los factores que puedan influenciar dicha percepción. Estamos invitando a participar entre 1.000 y 6.000 personas mayores de edad que vivan o trabajen en la ciudad de Palmira. Le tomará aproximadamente 15 minutos responder esta encuesta, cuyas preguntas están relacionadas con sus características sociales y económicas y la salud. Se realizarán 3 mediciones de la percepción de la población durante 6 meses. Usted puede participar una o más veces, cada vez que reciba la encuesta. Esta información podría ayudar a los gobernantes y a la ciudadanía a implementar intervenciones específicas para mantener la epidemia con bajos niveles de transmisión en este periodo pos-cuarentena.

Más información:

1. Su participación es voluntaria y anónima puede cancelar o dejar de diligenciar la encuesta en cualquier momento sin necesidad de dar explicaciones
2. Su participación en la investigación no lo expone a ningún riesgo significativo. Las preguntas no son sobre temas sensibles. Sus respuestas pueden ayudar para mejorar la respuesta a la epidemia en beneficio de la población
3. Los datos serán tratados con total confidencialidad
4. Sus respuestas serán usadas exclusivamente con propósitos científicos.
5. Los datos anonimizados se almacenarán en bases de datos custodiadas por investigadores en la universidad. No es posible sacar conclusiones de cada una de las personas, los resultados serán reportados para los grupos poblacionales. Los datos podrán utilizarse para investigación y docencia en el futuro.
6. El estudio no es comercial y es financiado por la Pontificia Universidad Javeriana Cali. Usted no recibirá pago por su participación.

¿Cómo puede contactarnos?

Responsables:

Sheila Gómez. sagomez@javerianacali.edu.co (mailto:sagomez@javerianacali.edu.co). 3186996218

Diana Caicedo, diana.caicedob@javerianacali.edu.co
(mailto:diana.caicedob@javerianacali.edu.co). 3127217092

Miembros del comité de ética: etica.investigación@javerianacali.edu.co
(mailto:tica.investigación@javerianacali.edu.co).

📞 AL FINAL ENCONTRARÁ INFORMACIÓN DE INTERÉS SOBRE LA ENFERMEDAD COVID-19

¡COMPARTELA CON TUS CONOCIDOS!

C. Anexo: Certificado de ponencia



III Congreso Internacional de Estadística Aplicada

Certifica que

LAURA LUCIA DOMINGUEZ BARRIOS

Participo como ponente en el Tercer Congreso Internacional de Estadística Aplicada, llevado a cabo del 14 al 16 de julio de 2022.

Ponencia: Uso de variables latentes para evaluar la percepción de riesgo de contagio por Covid-19 en Cali y Palmira

Comité organizador



D. Anexo: Código R

```
library(bootstrap);library(factoextra);library(FactoMineR); library(ade4)

CaliyPalmira #Carga completa de la base de datos
Datos=CaliyPalmira[,c(19:34)]
CaliyPalmira.FMA<-MFA(Datos, #16 Variables para percepción
                      group=c(7,4, 5), type=c('s', 's','s'),
                      name.group=c("Probabilidad de contagio","Severidad",
                                   "Susceptibilidad"), graph=FALSE)

#Índice de Percepción de Riesgo Global (AFM)
res.mfa=CaliyPalmira.FMA
Coord <-res.mfa$global.pca$var$coord[,1]
vl_p<-res.mfa$eig[1] #Valor Propio
vt_p<-Coord/sqrt(vl_p) #Vector Propio
Pesos<-(vt_p/sum(vt_p)) # Pesos relativos
sum(Pesos) #Verificación

IndiceOriginal<-as.matrix(Datos)%*%Pesos
IndiceMinimo<-min(IndiceOriginal); IndiceMaximo<-max(IndiceOriginal)

Indice<-round((((IndiceOriginal-IndiceMinimo)/(IndiceMaximo-IndiceMinimo))*100,2)
C8<-cbind(Indice,Datos)

set.seed(1234)
kmeans <- kmeans(C8$Indice, 3, iter.max = 1000, nstart = 10)
C8$cluster <- kmeans$cluster;summary(C8)
kmeans$centers; tapply(Indice, kmeans$cluster, summary)

## ECM
NfinalMedia<-c() #cada repeticion bootstrap
for (i in 1:1000) {
  res.mfa_severidad=CaliyPalmira.FMA
  #Datos
```

```

Coord <-res.mfa$global.pca$var$coord[,c(1)]
vl_p<-res.mfa$eig[c(1)]
vt_p<-Coord/sqrt(vl_p)
Pesos<-(vt_p/sum(vt_p))
data.frame(round(Pesos,3))
IndiceOriginal<-as.matrix(Datos)%*%Pesos
IndiceMinimo<-min(Ind)
IndiceMaximo<-max(Ind)
Indice<-round(((Ind-IndiceMinimo)/(IndiceMaximo-IndiceMinimo))*100,2)
NfinalMedia[i]=median(sample(Indice,replace = TRUE))
}
hist(NfinalMedia)
I_mediana=round(median(NfinalMedia,3))
EE<-(1/(1000-1))*sum(NfinalMedia-median(C8$Indice))^2+(median(NfinalMedia)-median(C8$Indice))^2
EE

#ACM
dataACM<-DatosExploratorio
res.mca<-MCA(dataACM)
res.mca$call$Xtot ## Tabla disjuntiva completa
res.mca$eig
res.mca$var$contrib ## Contribuciones
fviz_mca_biplot(res.mca, repel=TRUE,cex=0.6,geom.ind = c("point"),
geom.var = c("point", "text"),col.ind = "#C1CDCD", col.var="#8B1A1A")

cluster<-HCPC(res.mca,nb.clust=-1,consol=FALSE,graph=FALSE)
plot.HCPC(cluster,choice='map',draw.tree=FALSE,title = '',
axes = c(1,2),ylim=c(-1.7,1.7),ind.names = FALSE)
summary(cluster)
cluster$desc.var #valores test

```

Bibliografía

- Abascal, E. & Landaluce, M. I. (2002), 'Análisis factorial múltiple como técnica de estudio de la estabilidad de los resultados de un análisis de componentes principales', *Qüestió* **26**(1-2).
- Abdi, H. (2007), 'Rv coefficient and congruence coefficient', *Encyclopedia of measurement and statistics* **849**(853), 92.
- Andersen, K., Rambaut, A., Lipkin, W., Holmes, E. & Garry, R. (2020), 'The proximal origin of sars-cov-2', *Nat Med* **26**, 450–452.
- Becerra, M. (2010), 'Comparación del análisis factorial múltiple (afm) y del análisis en componentes principales para datos cualitativos (prinqual), en la construcción de índices', *Universidad de Nacional de Colombia* .
- Caicedo, D., Gómez, S., Estrada, V., Bermúdez, E., Pérez, M., Aristizábal, J., Holguín, J., Rodríguez, M., Gutiérrez, K., Velásquez, A. & Benítez, N. (2021a), 'Percepción de riesgo de covid-19 en cali', *Pontificia Universidad Javeriana* .
URL: <https://www.javerianacali.edu.co/noticias/los-lugares-donde-menos-se-cumplen-las-medidas-de-bioseguridad-resultados-de-encuesta>
- Caicedo, D., Gómez, S., Estrada, V., Bermúdez, E., Pérez, M., Aristizábal, J., Holguín, J., Rodríguez, M., Gutiérrez, K., Velásquez, A. & Benítez, N. (2021b), 'Percepción de riesgo de covid-19 en palmira', *Pontificia Universidad Javeriana* .
URL: <https://www.javerianacali.edu.co/noticias/los-lugares-donde-menos-se-cumplen-las-medidas-de-bioseguridad-resultados-de-encuesta>
- Davis, L., Melmed, A. & Krop, R. (2003), *Individual preparedness and response to chemical, radiological, nuclear, and biological terrorist attacks*, Rand Corporation, pp. 147–152.
- Dryhurst, S., Schneider, C., Kerr, J., Freeman, A., Recchia, G., Van der Bles, A., Spiegelhalter, D. & Van der Linden, S. (2020), 'Risk perceptions of covid-19 around the world', *Journal of Risk Research* **23**(7-8), 994–1006.
- Duan, T., Jiang, H., Deng, X., Zhang, Q. & Wang, F. (2020), 'Government intervention, risk perception, and the adoption of protective action recommendations: evidence from the covid-19 prevention and control experience of china', *International journal of environmental research and public health* **17**(10), 3387.

- Escofier, B. & Pagès, J. (1992), *Análisis Factoriales Simples y Múltiples: Objetivos, Métodos e Interpretación*, Universidad del País Vasco, España, pp. 117–170.
- Fetjah, D., Ihssane, B., Idardare, Z., Ainhout, L. & Bouqbis, L. (2022), ‘Limiting the hurtful oxidative stress and seasonal physiological adaptations in seashore paspalum through the use of banana waste biochar and compost mixtures’, *Journal of Ecological Engineering* **23**(1), 216–227.
- Fujii, R., Suzuki, K. & Niimi, J. (2021), ‘Public perceptions, individual characteristics, and preventive behaviors for COVID-19 in six countries: a cross-sectional study’, *Environmental Health and Preventive Medicine* **26**(1).
- González Rojas, V. (2014), ‘Análisis conjunto de múltiples tablas de datos mixtos mediante PLS’, *Universitat Politècnica de Catalunya* .
- Hernández Sampieri, R., Fernández, C. & Baptista, P. (2014), *Metodología de la investigación*, 6 edn, McGRAW-HILL, pp. 194–267.
- Johns Hopkins University (2022), *COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)*.
URL: <https://coronavirus.jhu.edu/map.html>
- Kassambara, A. & Mundt, F. (2020), *factoextra: Extract and Visualize the Results of Multivariate Data Analyses*. R package version 1.0.7.
URL: <https://CRAN.R-project.org/package=factoextra>
- Lê, S., Josse, J. & Husson, F. (2008), ‘FactoMineR: A package for multivariate analysis’, *Journal of Statistical Software* **25**(1), 1–18.
- Lebart, L., Morineau, A. & Piron, M. (1995), *Statistique exploratoire multidimensionnelle*, Vol. 3, Dunod Paris.
- Matar-Khalil, S., Ortiz Barrero, M. & González-Campos, J. (2021), ‘Diseño y validación de un cuestionario para evaluar la percepción de riesgo de contagio de covid-19 en población colombiana’, *Rev. Perú. medicina Exp. salud publica* **38**(4), 512–520.
URL: <https://doi.org/10.17843/rpmesp.2021.384.9298>.
- Mendivelso, F. & Rodríguez, M. (2018), ‘Prueba chi-cuadrado de independencia aplicada a tablas 2xn’, **21**, 92 – 95.
- Ochoa Muñoz, A. (2018), ‘Análisis de Correspondencias Múltiples en presencia de datos faltantes: el principio de datos disponibles del algoritmo NIPALS (ACMpdd)’, *Universidad del Valle* .

- OMS (2020a), *Brote de enfermedad por coronavirus (COVID-19): orientaciones para el público*.
URL: <https://www.who.int/es/emergencies/diseases/novel-coronavirus-2019/advice-for-public>
- OMS (2020b), *Managing the COVID-19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation*.
- OPS (2020), *Coronavirus*.
URL: <https://www.paho.org/es/temas/coronavirus>
- Organización Panamericana de la Salud (2020), *COVID-19 Consejos para informar. Guía para periodistas*.
- Pardo, C. & Cabarcas, G. (2001), Métodos estadísticos multivariados en investigación social, in ‘Simposio de estadística’, Vol. 13.
- Pardo, C. E. (2020), *Estadística descriptiva multivariada*, Universidad Nacional de Colombia.
- Pardo, C. E. & Del Campo, P. C. (2007), ‘Combinación de métodos factoriales y de análisis de conglomerados en R: el paquete FactoClass’, *Revista Colombiana de Estadística* **30**, 231 – 245.
- R Core Team (2022), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: <https://www.R-project.org/>
- Rohrmann, B. (2008), Risk perception, risk attitude, risk communication, risk management: A conceptual appraisal, in ‘15th International Emergency Management Society (TIEMS) Annual Conference’.
- Secretaría de Salud del Valle del Cauca (2022), *Boletín Covid-19 de fecha 01/01/2022*.
URL: <https://www.valledelcauca.gov.co/loader.php?lServicio=Tools2lTipo=viewpdfid=58696>
- Silva, L. (1997), *Cultura estadística e investigación científica en el campo de la salud: Una mirada crítica*, Editorial Diaz de Santos., Madrid.
- Solanas, A. & Olivera, V. (1992), ‘Bootstrap: fundamentos e introducción a sus aplicaciones’, *Anuario de psicología a The UB Journal of Psychology*.
- Tovar, J., Vargas, K., Díaz, J. & Ochoa, A. (2021), ‘Pleasure for independent reading: a proposal for its measurement’, *PUBLICACIONES* **51**(2), 289–325.
- Trejos, J., Castillo, W. & Varela, J. (2014), *Análisis multivariado de datos: Métodos y aplicaciones*, Editorial UCR, San José, CR, pp. 23–83.

- Varela, M., Ochoa, A. & Tovar, J. (2018), ‘Measuring eating habits and physical activity in children: Synthesis of information using indexes and clusters’, *Revista Mexicana de Trastornos Alimentarios/Mexican Journal of Eating Disorders* **9**, 264–276.
- Visbal-Cadavid, D., Martínez-Gómez, M. & Escorcia-Caballero, R. (2020), ‘Exploring university performance through multiple factor analysis: A case study’, *Sustainability (Switzerland)* **12**(3).
- Xian, S., Xia, H., Yin, Y., Zhai, Z. & Shang, Y. (2016), ‘Principal component clustering approach to teaching quality discriminant analysis’, *Cogent Education* **3**(1), 1194553.