

Regionalización de la Pobreza Multidimensional

Ejemplo de aplicación en México 2016 - 2022

García Rodríguez, L . A.

Resumen

La presente investigación, ofrece una metodología para la regionalización de la pobreza multidimensional en México. Este enfoque aborda la pobreza más allá de lo económico, integrando múltiples dimensiones sociales y de derechos, tal como lo estipula la Ley General de Desarrollo Social en su Art36 y lo reconoce la metodología de medición de la pobreza multidimensional del Coneval. Se utilizaron técnicas de estadística multivariante y aprendizaje automático para analizar los datos, superando las limitaciones de las metodologías tradicionales. El objetivo es mejorar la focalización y eficacia de las políticas públicas en México. Los resultados demuestran la aplicabilidad de nuestra metodología en la política y el desarrollo social, ofreciendo una perspectiva más completa y precisa de la pobreza en el país.

Índice

1. Introducción	3
1.1. Contexto y justificación	3
1.2. Problema de investigación	4
1.3. Objetivos de investigación	5
1.4. Preguntas de investigación	5
1.5. Relevancia y contribución	5
2. Revisión de la literatura	5
2.1. Antecedentes	5
2.2. Marco Teórico	7
2.3. Investigaciones previas relacionadas	10
2.4. Brechas del conocimiento actual	12
3. Metodología	12
3.1. Diseño de Investigación	12
3.2. Recopilación de datos	13
3.3. Instrumentos y herramientas	14
3.4. Procedimientos	15
3.4.1. Carga de datos	15
3.4.2. Análisis estadístico descriptivo	18
3.4.3. Análisis Factorial	26
3.4.4. Análisis de Cluster	36
4. Resultados	42
4.1. Regionalizaciones	42
4.2. Estadísticas por cluster	49
5. Discusión	54
6. Conclusiones	54

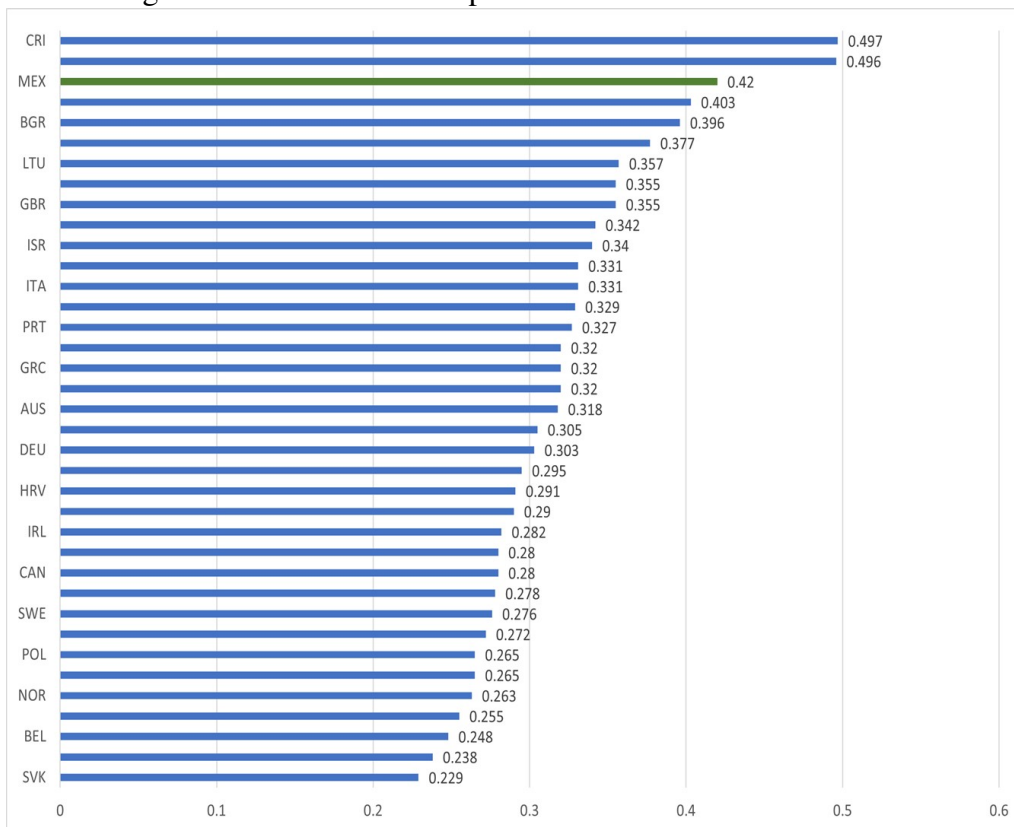
1. Introducción

1.1. Contexto y justificación

La estructura social del sistema económico capitalista, caracterizada por la supremacía de la burguesía sobre el proletariado, han generado una histórica y enorme desigualdad social en nuestro país. Sucesos como la implementación del neoliberalismo, y la firma del TLCAN, han agravado esta desigualdad, generando una desequitativa distribución del ingreso donde los poseedores de los medios de producción han quedado favorecidos, mientras aquellos que recurren a vender su fuerza de trabajo, han sido los más afectados y desfavorecidos.

En 2020 México tuvo un índice de Gini de 0.42, ocupando el tercer lugar de los países de la OCDE con mayor índice de Gini, esta preocupante cifra señala una desigual distribución de la riqueza en nuestro país.(OCDE) [1]

Figura 1: Índice de Gini en países miembros de la OCDE 2020



Elaboración propia con datos de OCDE

El índice de Palma es otra medida de desigualdad de ingresos, este es una ratio entre los ingresos de los más ricos y los más pobres en una sociedad. Específicamente, compara la proporción del ingreso total que posee el decil X, es decir, el 10 % más rico de la población con la proporción que poseen los deciles del I al IV, es decir el 40 % más pobre.

En México, el índice Palma de 2022, calculada con la ENIGH fue de 2.53 (*Gómez. A, 2023*)[2] es decir, que el decil X concentra 2.53 veces lo que tienen los deciles del I al IV juntos. Y esto, aunque no es el nivel más extremo posible, refleja una nación donde una pequeña proporción de la población controla una porción mucho mayor de los ingresos totales en comparación con una sección mucho más grande de la población menos adinerada. Fácilmente puede interpretarse, que el decil X es el mas favorecido, y no es difícil intuir que los más afectados en esta desigual distribución del ingreso, son los primeros deciles. Es decir, los pobres.

La pobreza es un problema público vigente, y es responsabilidad del estado y de los tres poderes de la federación atenderlo, pues el incremento de la pobreza no solo resulta injusto y perjudica la dignidad e integridad de millones de personas, sino que también fomenta una desigualdad, deteriorando la unidad comunitaria y afectando el desarrollo económico.

Como economistas, es nuestro deber ser el ayudar a mejorar la dignidad humana y las condiciones de vida en nuestro país, es nuestro deber ser el contribuir y proponer soluciones a los problemas públicos, como lo es la pobreza. Para poder combatirla, se necesitan de políticas de desarrollo social, y para ello, primero se debe tener una correcta identificación de la pobreza, la cual, no es homogénea en todo el país. De allí nace la necesidad de una metodología de regionalización que pueda clasificar cada una de las entidades federativas del país según los atributos particulares de la pobreza, con un enfoque que abarque su carácter multidimensional y contemple todas sus dimensiones, enfocándose no solo en una cuantificación de la población empobrecida, sino también en las carencias, y la intensidad con la que estas afectan a cada entidad. En ese sentido la presente investigación, se enmarca en la búsqueda teórica y estadística para comprender, cuantificar y mitigar la pobreza multidimensional en México.

1.2. Problema de investigación

El principal problema de investigación que propicia a este estudio es el disentimiento con las propuestas de regionalización de la pobreza en México consultadas en la bibliografía. Dichas regionalizaciones se basan en la cercanía geográfica entre entidades federativas, con un supuesto de contigüidad, lo que, a mi parecer, no logra contribuir a un mejoramiento de la focalización de políticas públicas y de desarrollo social de forma regional.

Este estudio propone trascender estas barreras mediante la formulación de una metodología de regionalización sin considerar la contigüidad, y en su lugar, integrar las propor-

ciones de población que padece cada entidad federativa en las distintas dimensiones de la pobreza, y utilizar esta proporción como el criterio de regionalización.

1.3. Objetivos de investigación

El objetivo principal de este trabajo es la formulación de una metodología de regionalización, que considere la multidimensional de la pobreza y cada una de sus dimensiones como el criterio de regionalización del territorio mexicano. Para ello se requiere una fusión metodologías de estadística multivariante y algoritmos de aprendizaje automático para capturar con mayor fidelidad la multidimensionalidad de la pobreza. La investigación aspira a ser una herramienta para los tomadores de decisión y potenciar así la precisión y eficacia de sus programas y estrategias de intervención.

1.4. Preguntas de investigación

Las preguntas de investigación que guían el estudio, son: ¿Cómo se pueden integrar las diversas dimensiones de la pobreza en un modelo de regionalización?, ¿De qué manera los enfoques estadísticos y el aprendizaje automático pueden mejorar la identificación y comprensión de las áreas más afectadas por la pobreza multidimensional?, ¿Cómo puede esta nueva metodología influir en la formulación y ejecución de políticas públicas más efectivas?

1.5. Relevancia y contribución

La relevancia de este estudio reside en su metodología de regionalización, la cual, es una herramienta analítica, adaptativa y dinámica que permite identificar los distintos niveles de pobreza en el país y su evolución por entidad federativa, para poder mejorar la focalización y efectividad de las políticas públicas y de desarrollo social. La contribución se extiende más allá de lo académico, ofreciendo una herramienta que puede mejorar las estrategias de intervención a nivel nacional, contribuyendo así a la reducción de la pobreza en México.

2. Revisión de la literatura

2.1. Antecedentes

En México, el 20 de enero de 2004, se dio la aprobación por unanimidad (en las Cámaras de Diputados y de Senadores) de la Ley General de Desarrollo Social, cuyo objetivo es: “garantizar el pleno ejercicio de los derechos sociales consagrados en la Constitución

Política de los Estados Unidos Mexicanos, asegurando que toda la población se beneficie del desarrollo social” (*Coneval, 2023*).[3]

Esta misma ley dio lugar a el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL), “como la instancia con autonomía técnica y de gestión con las atribuciones de medir la pobreza y evaluar la política de desarrollo social” (*Coneval*).[4]

En el *decreto por el que se regula el Consejo Nacional de Evaluación de la Política de Desarrollo Social*[5], se estipula en su artículo tercero que este tiene el objeto de:

- I Normar y coordinar la evaluación de la Política Nacional de Desarrollo Social y las políticas, programas y acciones que ejecuten las dependencias públicas, sin perjuicio de las atribuciones que en materia de control y evaluación tienen las secretarías de Hacienda y Crédito Público y de la Función Pública, y;
- II Establecer los lineamientos y criterios para la definición, identificación y medición de la pobreza, garantizando la transparencia, objetividad y rigor técnico en dicha actividad.

La multidimensionalidad de la pobreza se expresa en el artículo 36 de la *Ley General de Desarrollo Social*[6], que dicta: Los lineamientos y criterios que establezca el Consejo Nacional de Evaluación de la Política de Desarrollo Social para la definición, identificación y medición de la pobreza son de aplicación obligatoria para las entidades y dependencias públicas que participen en la ejecución de los programas de desarrollo social, y deberá utilizar la información que genere el Instituto Nacional de Estadística, Geografía e Informática, independientemente de otros datos que se estime conveniente, al menos sobre los siguientes indicadores:

- I Ingreso corriente per cápita;
- II Rezago educativo promedio en el hogar;
- III Acceso a los servicios de salud;
- IV Acceso a la seguridad social;
- V Calidad y espacios de la vivienda digna y decorosa;
- VI Acceso a los servicios básicos en la vivienda digna y decorosa
- VII Acceso a la alimentación nutritiva y de calidad;
- VIII Grado de cohesión social, y;
- IX Grado de Accesibilidad a carretera pavimentada.

2.2. Marco Teórico

Para cumplir con sus atribuciones establecidas en la Ley General de Desarrollo Social y conscientes de la discusión internacional en la conceptualización de la pobreza multidimensional, el CONEVAL, adopto desde 2009 un enfoque multidimensional en la metodología de medición de pobreza, cuyos lineamientos se muestran en la tercera edición de la Metodología para la medición multidimensional de la pobreza en México. (*Coneval,2019*)[7]

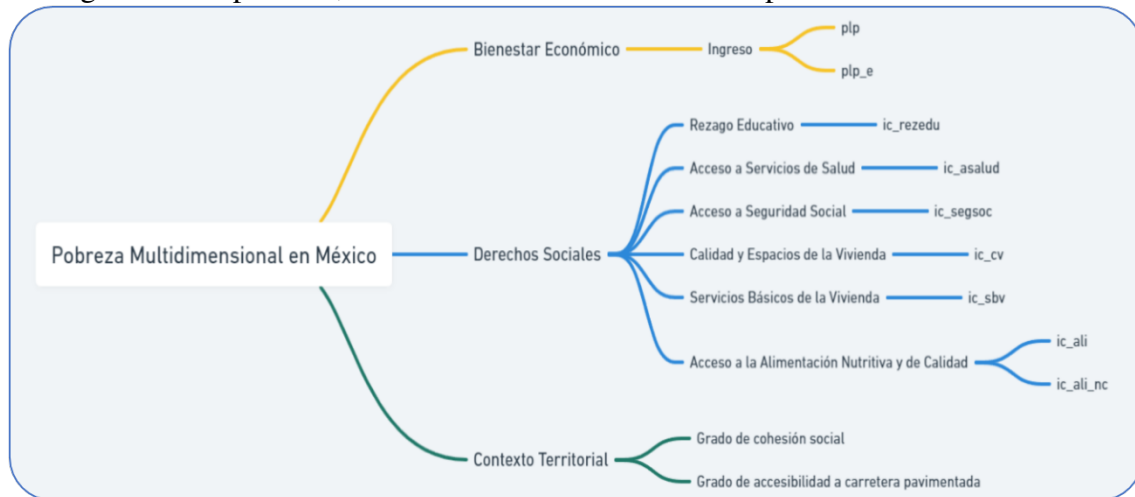
Esta metodología está basada en dos perspectivas que analizan este carácter de la pobreza, que son la perspectiva de bienestar y la de derechos. La primera se basa en la premisa de que “el ingreso es fundamental para la adquisición de una variedad de bienes y servicios que son indispensables para la satisfacción de las necesidades esenciales” (CONEVAL,2023) , mientras que la segunda considera el cumplimiento de los derechos sociales establecidos como derechos (garantías) constitucionales, y se concibe a la pobreza como una negación de estos derechos. (*CONEVAL,2023*)[7]

Las dimensiones de la pobreza estipuladas en el artículo 36, son retomadas por el CONEVAL, que hace un análisis de los resultados de la ENIGH en contraste con fuentes

externas, y también se realiza un proceso de validación interna, para posteriormente para publicar los resultados y estimaciones de la pobreza multidimensional a nivel nacional, estatal y por grupos poblacionales. Además, este realiza una validación de dichas estimaciones con un agente externo. (CONEVAL, 2023) [3]

Las perspectivas del enfoque multidimensional y las dimensiones pueden expresarse de la siguiente forma:

Figura 2: Perspectivas, dimensiones e indicadores de la pobreza multidimensional



Elaboración propia con información de la presentación de las estimaciones de la medición de pobreza 2022, recuperada de:
<https://www.youtube.com/watch?v=2mkkdaxBTPY>

En este diagrama de árbol, el primer nivel de nodos representa la perspectiva de la pobreza; el segundo nivel, las dimensiones de la pobreza; y el tercer nivel, los indicadores que genera el CONEVAL con la información de la ENIGH para representar estas dimensiones.

De forma concreta, la definición de pobreza del CONEVAL es la siguiente:

“Una persona se encuentra en situación de pobreza multidimensional cuando no tiene garantizado el ejercicio de al menos uno de sus derechos para el desarrollo social, y sus ingresos son insuficientes para adquirir los bienes y servicios que requiere para satisfacer sus necesidades.” (CONEVAL, 2019)[7]

Y de manera específica, el CONEVAL identifica a un individuo, en alguna de las siguientes clasificaciones:

Pobreza extrema: Aquellas personas cuyo ingreso es menor a la línea de pobreza extrema por ingresos y que presentan 3 o más carencias.

Pobreza moderada: Aquellas personas cuyo ingreso es menor a la línea de pobreza por ingresos y que presentan 1 o 2 carencias.

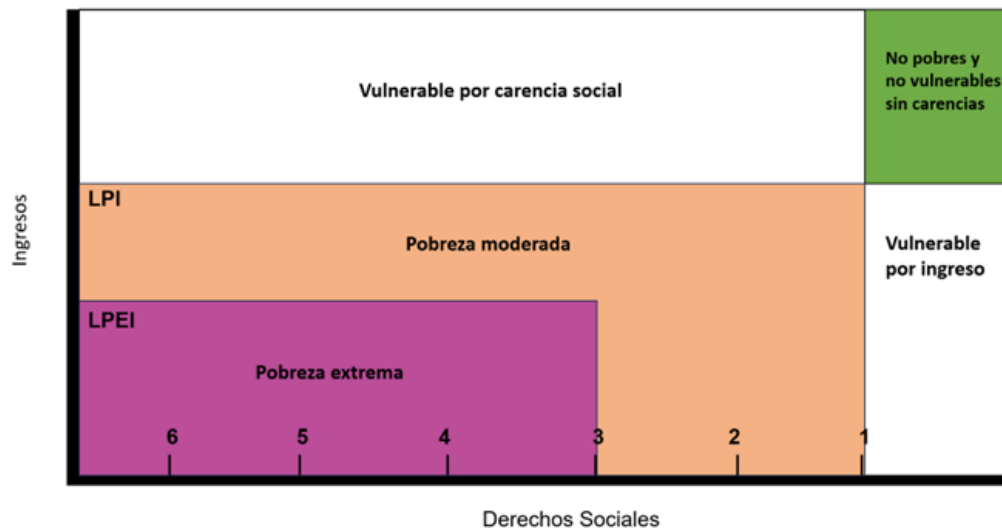
Vulnerable por carencia social: Aquellas personas cuyo ingreso es mayor a la línea de pobreza por ingresos y que presentan 1 o más carencias.

Vulnerable por ingreso: Aquellas personas cuyo ingreso es menor a la línea de pobreza por ingresos y que no presentan carencias.

No pobres y no vulnerables sin carencias: Aquellas personas cuyo ingreso es mayor a la línea de pobreza por ingresos y que no presentan carencias.

Lo anterior puede visualizarse de manera gráfica en la siguiente recreación de la dispositiva de la presentación del resumen ejecutivo de la medición de la pobreza en 2022.

Figura 3: Identificación de la pobreza multidimensional



Recreación de la diapositiva de la conferencia de prensa "Presentación de estimaciones de la Medición de Pobreza 2022" recuperada de:
<https://www.youtube.com/watch?v=2mkkdaxBTP>[8]

La metodología establecida por el CONEVAL ofrece una estimación puntual de la pobreza. Se trata de un enfoque que analiza la pobreza, no solo desde la perspectiva de los ingresos, sino también considerando las carencias sociales con la perspectiva de derechos sociales.

2.3. Investigaciones previas relacionadas

Regionalización de la pobreza y pobreza extrema en México 2008-2018[9]

Este estudio de Islas Ochoa y Torres Busqueño, publicado en "Tiempo Económico" de la Universidad Autónoma Metropolitana, da un análisis detallado de las tendencias de la pobreza en México, utilizando una metodología de regionalización. Los autores emplean datos del CONEVAL para examinar las tasas de pobreza y pobreza extrema a nivel nacional y regional, señalando cómo estas han evolucionado entre 2008 y 2018. La regionalización se basa en la división geoeconómica propuesta por Ángel Bassols Batalla en 1967. Esta metodología divide al país en siete regiones definidas: Centro, Golfo de México, Norte, Pacífico Norte, Pacífico Central, Sureste y Península de Yucatán.

La investigación resalta la heterogeneidad en la intensidad y características de la pobreza entre las regiones, que sugiere la necesidad de políticas diferenciadas y focalizadas para combatir eficazmente estos problemas.

Las conclusiones del estudio señalan que a pesar de los esfuerzos nacionales, la pobreza sigue siendo una cuestión arraigada y regionalmente diversa en México. Se subraya la importancia de desarrollar estrategias de política pública que no solo se enfoquen en reducir las cifras generales de pobreza, sino que también reconozcan y aborden las particularidades regionales.

La pobreza en México y sus regiones: un análisis de impacto del programa Oportunidades en el periodo 2002-2006 [10]

La investigación de Joaquín Bracamontes-Nevarez y Mario Camberos-Castro, se centra en medir la pobreza en México utilizando un enfoque monetario y el método de líneas de pobreza. Para la regionalización se consideran las cuatro regiones migratorias establecidas por el Consejo Nacional de Población. El Conapo divide el territorio del país en cuatro grandes regiones migratorias y agrupa a las entidades federativas por la cercanía geográfica y su tradición en cuanto a intensidad migratoria. Dentro de la metodología se emplean técnicas de micro simulación estática para evaluar el impacto de las transferencias monetarias del programa Oportunidades en la reducción de la pobreza, particularmente en la región Sur-Sureste. Los autores identificaron resultados poco exitosos del programa en la disminución de la pobreza, lo que cuestiona la eficiencia de Oportunidades debido a recursos insuficientes y errores de focalización.

El estudio utiliza datos de la Encuesta Nacional de Ingresos y Gastos de los Hogares (2002 y 2006) para evaluar la pobreza desde varias perspectivas. Inicialmente, adopta el enfoque monetario, centrándose en los ingresos y gastos para medir la pobreza, y luego aborda el enfoque multidimensional.

El estudio aplica el método de líneas de pobreza desarrollado por Amartya Sen, que com-

para el ingreso corriente de los hogares con una línea de pobreza especificada para identificar a los hogares pobres. Además, detalla cómo se utilizan los índices de pobreza global y estandarizado para medir la proporción y la intensidad de la pobreza en diferentes entidades y regiones. La micro simulación estática, una técnica crucial en su análisis, que permite evaluar los efectos de las políticas fiscales y sociales en la calidad de vida, calculando el ingreso disponible para cada hogar en una muestra representativa y simulando cambios en la política.

Los autores concluyen que las transferencias monetarias del programa Oportunidades han tenido un impacto limitado en la reducción de la pobreza, especialmente en el Sur-Sureste de México. Sugieren que los recursos insuficientes y los errores de focalización son factores clave detrás de este impacto reducido. Además, resaltan la importancia de considerar la diversidad regional en el diseño de políticas públicas para combatir la pobreza de manera más efectiva.

Una propuesta para el análisis regional de la pobreza en México [11]

En este estudio de Marco Antonio Pérez Méndez se analiza la pobreza en México desde una perspectiva regional. El autor propone una regionalización homogénea de la Tasa de Pobreza por entidad federativa, empleando el algoritmo Max-P para identificar el óptimo entre pobreza y desigualdad. Este enfoque busca identificar la pobreza y cómo esta se manifiesta en términos regionales debido al crecimiento económico y la desigualdad.

La metodología de regionalización se basa en la identificación de sendas regionales diferenciadas que explican la relación triangular entre pobreza, crecimiento y desigualdad de ingresos. El algoritmo Max-P permite identificar regiones homogéneas en términos de estas variables, optimizando la relación entre pobreza y desigualdad y asegurando que cada región satisfice un umbral mínimo predefinido por un atributo espacialmente extensivo.

Pérez Méndez realiza un análisis de las tasas de pobreza a nivel nacional y regional, considerando la evolución económica del país y las variaciones en las distribuciones de ingreso. Resalta cómo, a pesar de ciertas reducciones en las tasas de pobreza antes de la crisis económica de 2008, las tendencias se invirtieron durante y después de la crisis, llevando a una década perdida en términos de reducción de la pobreza y distribución del ingreso.

El estudio concluye que las transferencias monetarias y otras políticas implementadas hasta la fecha han tenido un impacto limitado en la reducción de la pobreza, especialmente en el sur-sureste de México, donde las tasas de pobreza son más altas. La investigación resalta la importancia de adoptar un enfoque regionalizado en el diseño de políticas públicas para combatir la pobreza de manera más efectiva, considerando las diferencias significativas en las condiciones económicas, sociales y culturales entre las regiones.

2.4. Brechas del conocimiento actual

Tras la revisión de la literatura, se identifica una brecha significativa en la comprensión y metodología de regionalización que afecta la efectividad de las políticas públicas y el desarrollo social. A pesar de la existencia de diversos enfoques y estudios sobre la pobreza multidimensional, persisten limitaciones fundamentales en cómo se conceptualiza y aborda esta problemática a nivel regional. Las brechas identificadas son las siguientes:

- 1 **Inadecuación de la regionalización basada en contigüidad geográfica:** Las metodologías de regionalización actuales, comúnmente se basan solo en la contigüidad geográfica, y no reflejan adecuadamente la complejidad y heterogeneidad de la pobreza en México. Esta forma de regionalización implica que la pobreza en diferentes áreas puede no estar siendo identificada ni abordada con la precisión necesaria para formular y aplicar políticas específicas y efectivas.
- 2 **Desafíos en la aplicación y efectividad de políticas públicas:** La falta de una metodología de regionalización precisa y matizada puede estar llevando a una focalización y asignación inadecuada de recursos, y por ende, a una menor eficacia en la reducción de la pobreza y mejora del bienestar.

Contribución:

Frente a estas brechas, esta investigación propone el desarrollo de una metodología de regionalización innovadora que no considere la contigüidad geográfica, sino que integre las proporciones de población afectadas por las diferentes dimensiones de la pobreza como criterio de regionalización. Esta metodología busca incorporar estadística multivariante y algoritmos de aprendizaje automático para ofrecer una regionalización más profunda y detallada de la pobreza multidimensional. La investigación aspira a proporcionar una herramienta analítica, adaptativa y dinámica que mejore la focalización y efectividad de las políticas públicas, contribuyendo significativamente a la lucha contra la pobreza en México.

3. Metodología

3.1. Diseño de Investigación

La presente investigación utiliza un diseño cuantitativo, longitudinal descriptivo y correlacional. Es correlacional, porque analiza las relaciones entre diversas variables reflejos de la pobreza, buscando entender cómo estas se relacionan y cómo el tenerlas a todas en cuenta puede dar lugar a una regionalización multidimensional de la pobreza. Es descriptiva, ya que con esta información busca detallar la pobreza y sus carencias de forma regional en el país, para mejorar en la formulación de políticas públicas.

Metodológicamente es un enfoque mixto que combina técnicas de estadística multivariante y de aprendizaje automático para procesar y analizar datos de este fenómeno multidimensional. Este enfoque permite abordar la complejidad y la naturaleza de la pobreza, no solo desde un punto de vista económico sino también considerando diversas dimensiones sociales y de derechos. Al integrar múltiples variables y técnicas analíticas, la metodología ofrece una visión más profunda y detallada de la pobreza y sus múltiples dimensiones.

3.2. Recopilación de datos

La metodología de análisis requiere de la estipulación de un año base, este será aquel año con el que se comparara el dinamismo de la pobreza multidimensional en años posteriores, y por su puesto este debe actualizarse constantemente, óptimamente, una vez cada seis años; pues puede ser también una forma de evaluar los resultados en el combate a la pobreza de cada sexenio.

En este ejemplo de aplicación se utiliza como año base 2016, y se hacen 3 comparaciones con esta base, que serán los años 2018, 2020 y 2022.

Las bases de datos empleadas para este análisis se obtuvieron de los de los programas de cálculo y bases de datos que el CONEVAL pone a disposición de la población cada 2 años. [12]. Y corresponden específicamente a los archivos *pobreza_16.csv*, *pobreza_18.csv*, *pobreza_20.csv*, y *pobreza_22.csv*, los cuales, para un acceso mas sencillo, fueron cargados en el *repositorio de Github de la investigación*. [13]

Las variables que fueron seleccionadas para el estudio son las que reflejan la multidimensionalidad de la pobreza según lo estipulado en el Artículo 36 de la LGDS (Ley General de Desarrollo Social). Estas incluyen:

plp_e Población por debajo de la línea de pobreza extrema por ingresos

plp Población por debajo de la línea de pobreza por ingresos

ic_rezedu Indicador de carencia por rezago educativo

ic_asalud Indicador de carencia por acceso a los servicios de salud

ic_segsoc Indicador de carencia por acceso a la seguridad social

ic_cv Indicador de carencia por calidad y espacios de la vivienda digna y decorosa

ic_sbv Indicador de carencia por acceso a servicios básicos en la vivienda

ic_ali Indicador de carencia por acceso a la alimentación

ic_ali_nc Indicador de carencia por acceso a la alimentación nutritiva y de calidad

Estas reflejan tanto la perspectiva de bienestar económico como la de derechos sociales.

Además se incluyó la variable `ent`, que nos indica la entidad federativa a la que corresponde cada observación.

3.3. Instrumentos y herramientas

Para la limpieza, transformación, manipulación, y modelado de los datos se usaron los software R y Python. Y se usaron las siguientes librerías:

R:

1. `ggplot2`
2. `GGally`
3. `reshape2`
4. `dplyr`
5. `moments`

Python:

1. `numpy`
2. `pandas`
3. `sklearn`
5. `geopandas`
6. `plotly`
7. `matplotlib`
8. `pandas.plotting`
9. `factor_analyzer`

Además se construyeron funciones en el mismo lenguaje para facilitar el análisis. Estas, al igual que todo el procedimiento de extracción, limpieza, transformación y análisis de datos están en el notebook de la investigación, disponible en el repositorio de la investigación [14].

3.4. Procedimientos

3.4.1. Carga de datos

Inicialmente, cargaron las base de datos originales desde el *repositorio de la investigación* [13] a un dataframe correspondiente a cada año (2016, 2018, 2020, 2022). Posteriormente se revisaron las respectivas dimensiones de cada base, es decir, el número de variables y observaciones para cada año. Dicha información se expresa en el siguiente cuadro:

Cuadro 1: Observaciones por año y por variable

Año	Variables	Observaciones
2016	10	257658
2018	10	269065
2020	10	315619
2022	10	269065

También se revisaron las observaciones por entidad federativa en cada año.

Cuadro 2: Observaciones por año y por entidad federativa

Estados	Obs_16	Obs_18	Obs_20	Obs_22
Aguascalientes	9302	9017	10040	9017
Baja California	12435	11413	13917	11413
Baja California Sur	7526	7756	8875	7756
Campeche	6385	6799	7781	6799
Coahuila	12023	11582	13692	11582
Colima	9196	9265	10611	9265
Chiapas	7383	7207	8275	7207
Chihuahua	10582	12517	15009	12517
CDMX	5710	7462	9035	7462
Durango	7462	8440	10113	8440
Guanajuato	9664	10539	11894	10539
Guerrero	6183	7261	9253	7261
Hidalgo	6944	6721	7819	6721
Jalisco	7621	7760	9976	7760
EDOMEX	11050	10982	13643	10982
Michoacán	7080	6720	7483	6720
Morelos	7516	7225	8737	7225
Nayarit	6607	6052	6979	6052
Nuevo León	10548	10446	11991	10446
Oaxaca	7067	8139	9516	8139
Puebla	8023	7136	8142	7136
Querátaro	8299	11678	13835	11678
Quintana Roo	6185	6065	7533	6065
San Luis PotosÃ	6338	8021	9342	8021
Sinaloa	10755	10229	11883	10229
Sonora	7844	7216	8135	7216
Tabasco	6456	6479	7312	6479
Tamaulipas	6358	6383	7754	6383
Tlaxcala	6800	7195	8490	7195
Veracruz	6309	8231	9366	8231
Yucatán	8997	9164	10305	9164
Zacatecas	6999	7956	8858	7956

Para identificar el tipo de variables con las que se está trabajando, es importante realizar una inspección de los valores únicos de los datos. Tras esta inspección se identificó que todas las variables de las cuatro bases, son dicotómicas, es decir, que toman valores 0 o 1. 0 en ausencia de algún atributo o carencia, y 1 en presencia del mismo.

También se inspeccionó la existencia de datos faltantes (NA), en el conjunto de datos. Los datos faltantes encontrados se resumen en la siguiente tabla:

Cuadro 3: Valores Na por base y por variable

Variable	2016	2018	2020	2022
ent	0	0	0	0
plp_e	0	0	0	0
plp	0	0	0	0
ic_rezedu	0	0	0	0
ic_asalud	0	0	0	0
ic_segsoc	0	0	0	0
ic_cv	0	0	0	0
ic_sbv	0	0	0	0
ic_ali	0	0	0	0
ic_ali_nc	0	0	0	0

Para este análisis es necesario trabajar con variables cuantitativas, por lo que se requiere de una transformación de datos en proporciones.

Por ejemplo, al contar las observaciones totales en la Base16 y la variable plp, que son iguales a 257658, y a su vez contáramos las observaciones de la misma base y variable, cuyo valor fuese igual a 1, que son igual a 126502 ; podríamos obtener una proporción de población que presenta plp, lo que se podría interpretar como la proporción de la población por debajo de la línea de pobreza por ingresos, que sería igual a $126502/257658 = 0.4909 = 49\%$

Se procede a obtener estas proporciones pero segmentadas por entidad federativa en cada una de las variables de interés, y en cada una de las bases. Los resultados de dicha transformación de variables dicotómicas discretas a continuas se encuentran en las hojas: "Base16_prop", "Base18_prop", "Base20_prop", y "Base22_prop" del archivo Cuadros_RegPoMex, disponible en el *repositorio de la investigación* [15]

Una vez terminada la transformación de variables dicotómicas a continuas, a través de los indicadores de proporción, se procede a realizar un análisis estadístico de las bases.

3.4.2. Análisis estadístico descriptivo

Se realizó un análisis estadístico para todas las bases en cada año de estudio, donde se calcularon los siguientes estadísticos:

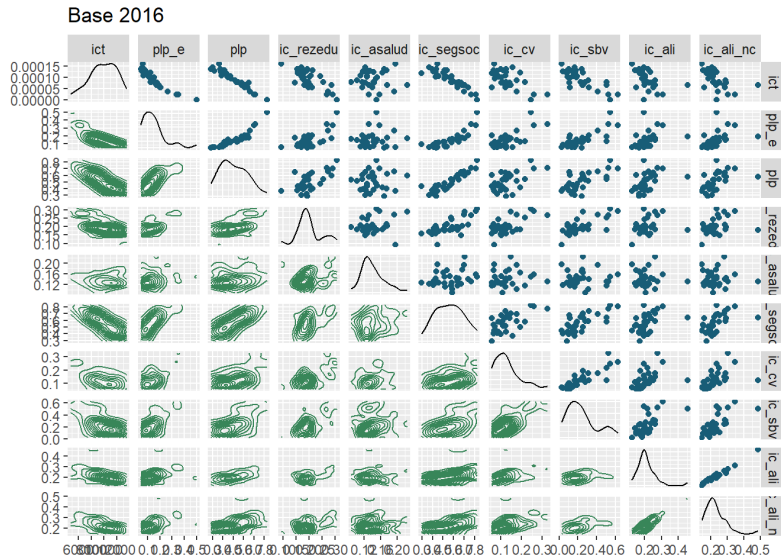
- Media
- Mediana
- Desv. Est.
- Varianza
- Rango
- IQR
- Rango
- MAD
- Q1
- Q3

Los resúmenes estadísticos de cada base se encuentran disponibles en las hojas Resumen16",Resumen18",Resumen20", yResumen22"del archivo Cuadros_RegPoMex, disponible en el *repositorio de la investigación*. [15]

A continuación se muestran gráficos de pares de las densidades univariantes, bivariantes y las dispersiones de cada año de estudio, así como su respectiva matriz de correlaciones.

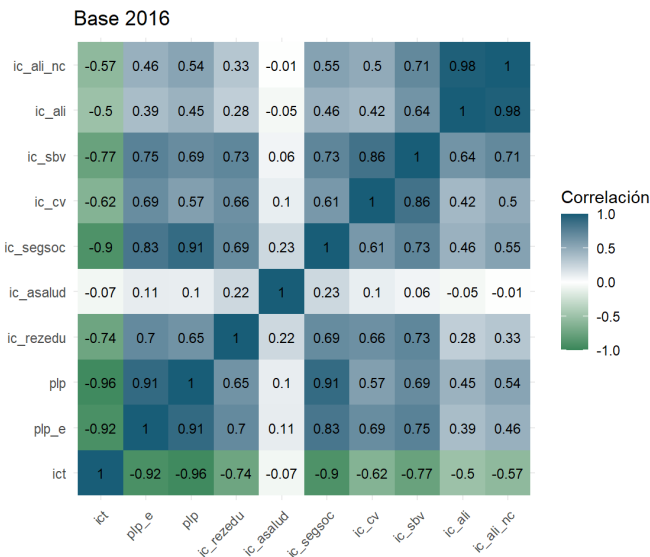
2016

Figura 4: Matriz de densidades y dispersiones para 2016



Elaboración propia con datos de CONEVAL

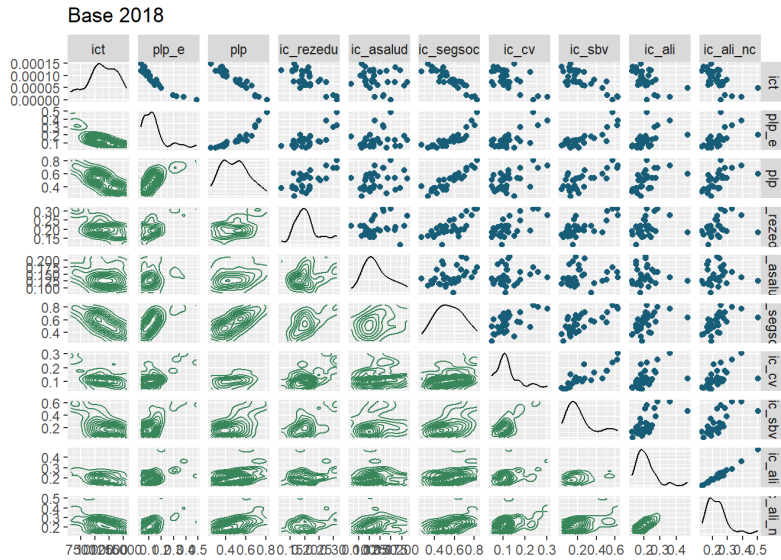
Figura 5: Matriz de correlaciones para 2016



Elaboración propia con datos de CONEVAL

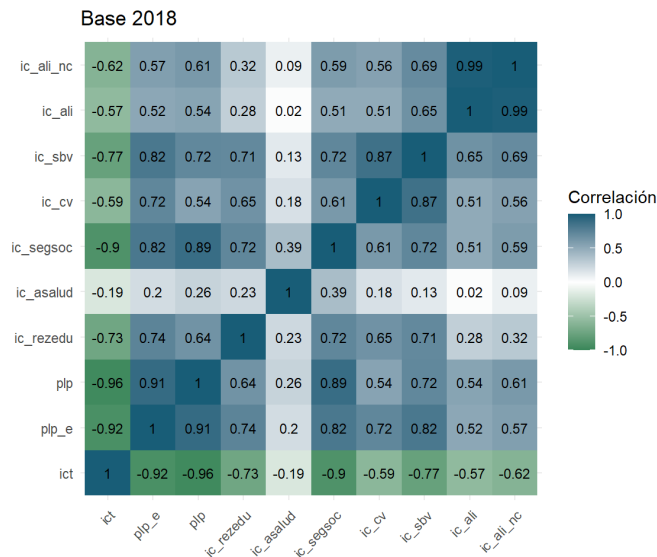
2018

Figura 6: Matriz de densidades y dispersiones para 2018



Elaboración propia con datos de CONEVAL

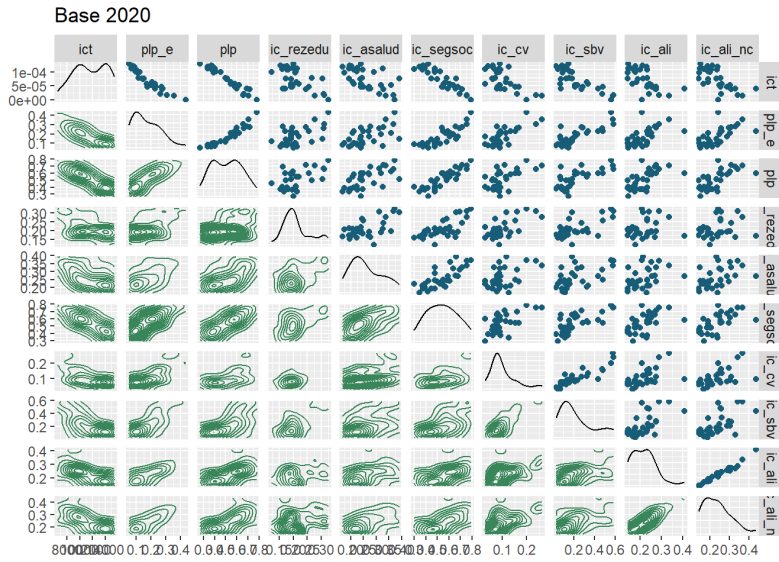
Figura 7: Matriz de correlaciones para 2018



Elaboración propia con datos de CONEVAL

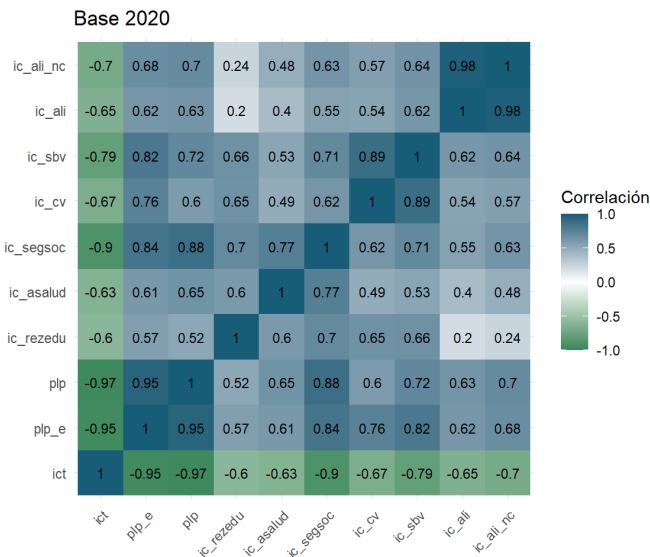
2020

Figura 8: Matriz de densidades y dispersiones para 2020



Elaboración propia con datos de CONEVAL

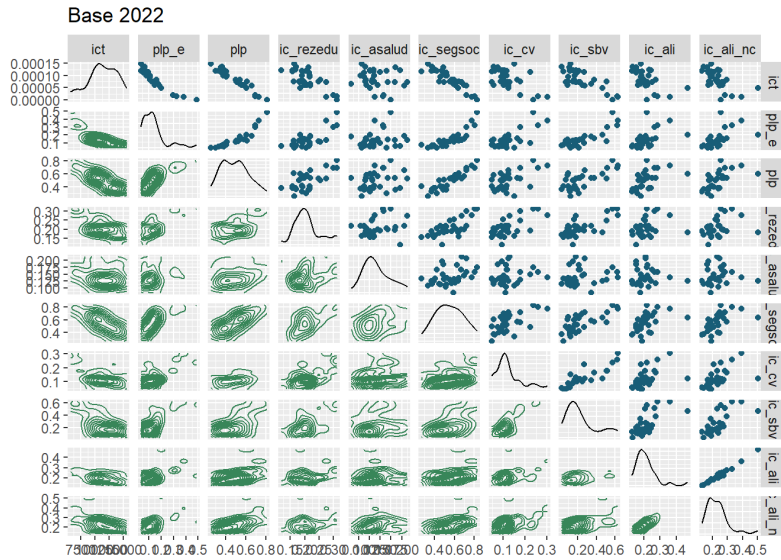
Figura 9: Matriz de correlaciones para 2020



Elaboración propia con datos de CONEVAL

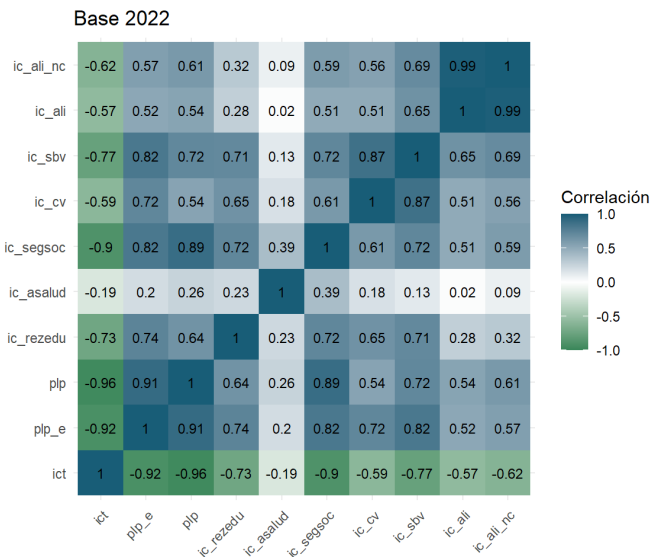
2022

Figura 10: Matriz de densidades y dispersiones para 2022



Elaboración propia con datos de CONEVAL

Figura 11: Matriz de correlaciones para 2022



Elaboración propia con datos de CONEVAL

Analizando las estadísticas, se identifica que de forma general, los indicadores de la pobreza multidimensional tuvieron una tendencia a la baja de 2016 a 2018, lo que se interpreta como una disminución de la pobreza multidimensional en dichos años. Desafortunadamente y a causa de la terrible situación de pandemia que sufrió nuestro país al igual que el resto del mundo en 2020, esta rompió con la tendencia, pues la mayoría de indicadores no mostraron variaciones significativas, a excepción de el *ic_asalud*, que evidentemente por la situación pandémica sufrió una alza que hizo pasar el promedio nacional de esta carencia de 14 % en 2018 a 26 % en 2020, un aumento del 12 % lo cual es muy significativo.

En los gráficos anteriores, identificamos que en este periodo de 2016 a 2022, la estructura de los datos se han mantenido prácticamente igual, sin mucha variación en sus distribuciones, y asociaciones lineales. A continuación, nos enfocarnos a describir un poco de esta estructura de datos.

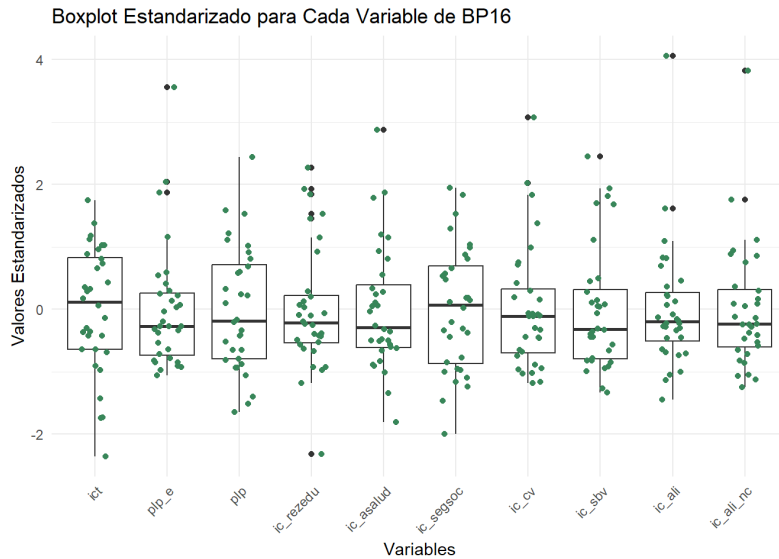
Se identifico que variables como *plp_e*, *ic_rezedu*, *ic_asalud*, *ic_cv*, *ic_sbv*, *ic_ali* y *ic_ali_nc* tienen distribuciones apuntaladas y sesgadas, cargadas a la izquierda, lo que indica que la mayoría de estados tienen proporciones menores al 30 % en estos indicadores. Encontramos distribuciones con un comportamiento distinto, como el de *ic_segsoc* que es mucho más normal, lo que apunta que a nivel nacional, la mayoría de los estados tiene una proporción de esta carencia en alrededor del 50 %, con una desviación estándar del 14 %. Otra distribución particular es la de *plp*, la cual es una distribución bimodal, que tiene dos medias, que en promedio se centran en un valor de 48-49 % con una desviación estándar del 13 %, cabe señalar también que la correlación entre ambas en los cuatro años es muy cercana al 0.90, lo indica que están altamente coleccionadas. Al igual que esta variable, existen más correlaciones importantes en el tiempo, como los es el caso de *ic_segsoc* y *ic_cv* y *ic_sbv*, que rondan entre el 0.90 y 0.80 respectivamente, lo que podría indicar de que aquellos que carecen de seguridad social, también carecen de servicios de vivienda y calidad de la misma. *ic_rezedu*, también presenta una correlación cercana al 0.70 con *ic_sbv*, lo que también induce a la idea de que una parte de quienes carecen de servicios básicos de la vivienda, también tienen rezago educativo. Esta última también presenta una correlación estable en el tiempo con *ic_ali* en alrededor de 0.60, lo que lleva a la misma idea de que quienes padecen de servicios básicos de vivienda, también tienen carencia por alimentación, solo que en menor intensidad.

En los gráficos de pares, bajo la diagonal principal, se encuentran las densidades variantes entre pares de variables, estas ayudan a conocer la distribución de los datos en un espacio bidimensional y también a identificar outliers o datos atípicos en el mismo espacio. Se observa que la mayoría de los pares de variables se distribuyen como una normal bivariante, salvo a algunos casos, que son las distribuciones que ven más deformes y no siguen una forma elipsoidal. Además identificamos grupos de atípicos como pequeñas islas separadas del conjunto de datos más denso, y esto se nota especialmente en las variables

relacionadas con *plp_e*.

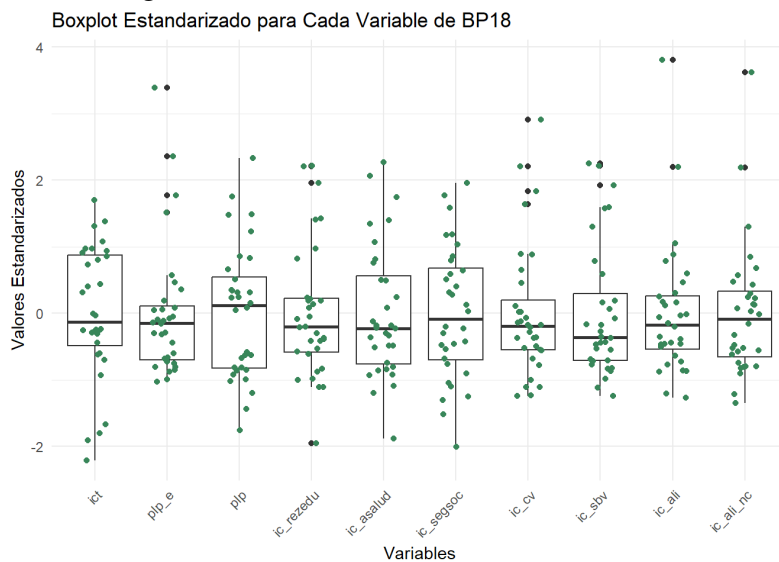
Para identificar datos atípicos, volvemos al espacio univariante, así que estandarizamos y empleamos gráficos de cajas y bigotes, también conocidos como boxplots.

Figura 12: Identificación de outliers en 2016

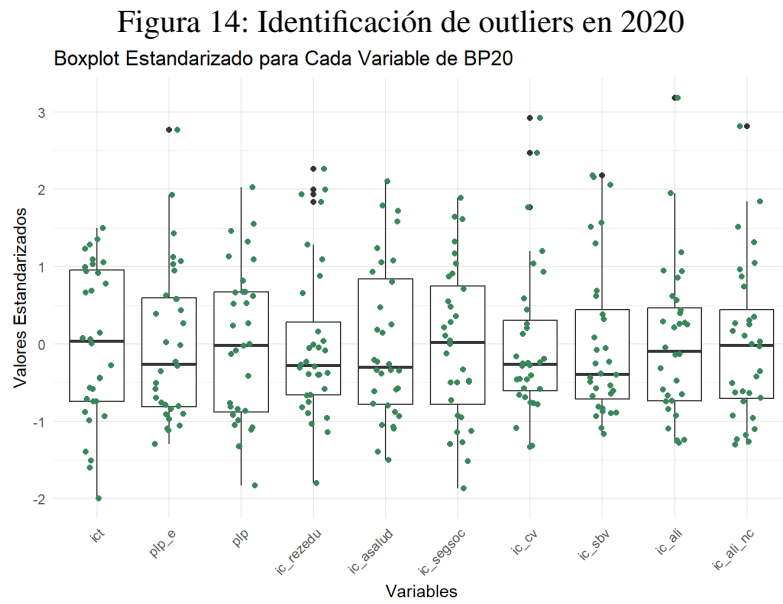


Elaboración propia

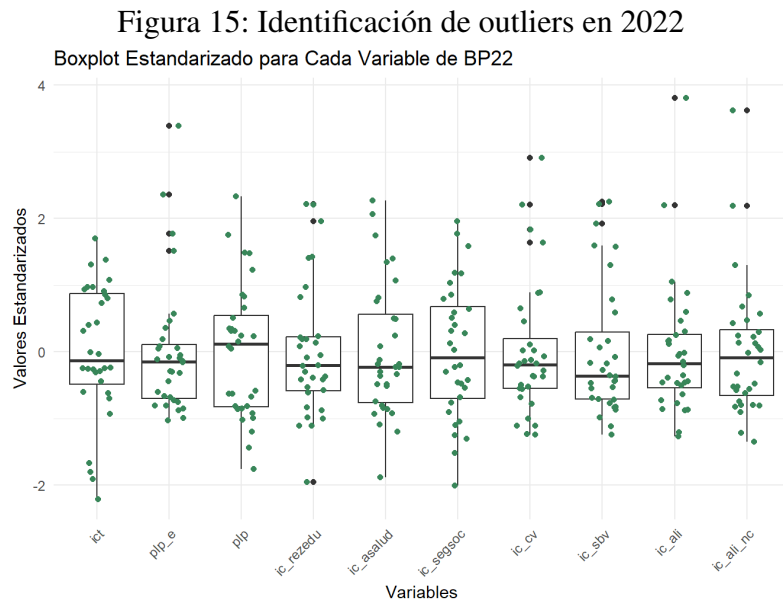
Figura 13: Identificación de outliers en 2018



Elaboración propia



Elaboración propia



Elaboración propia

Se identificó que en todas las variables se presentan datos atípicos, salvo plp, ic_segroc, e ic_asalud, que son las variables mas centradas y normalmente distribuidas de forma univariante.

Una vez realizado el análisis estadístico descriptivo, procedemos a la parte inferencial. Para la metodología de regionalización, se implementan dos modelos de análisis de datos. El primero es un análisis Factorial de componentes principales, mientras es segundo es una análisis de cluster con las puntuaciones factoriales del análisis factorial.

3.4.3. Análisis Factorial

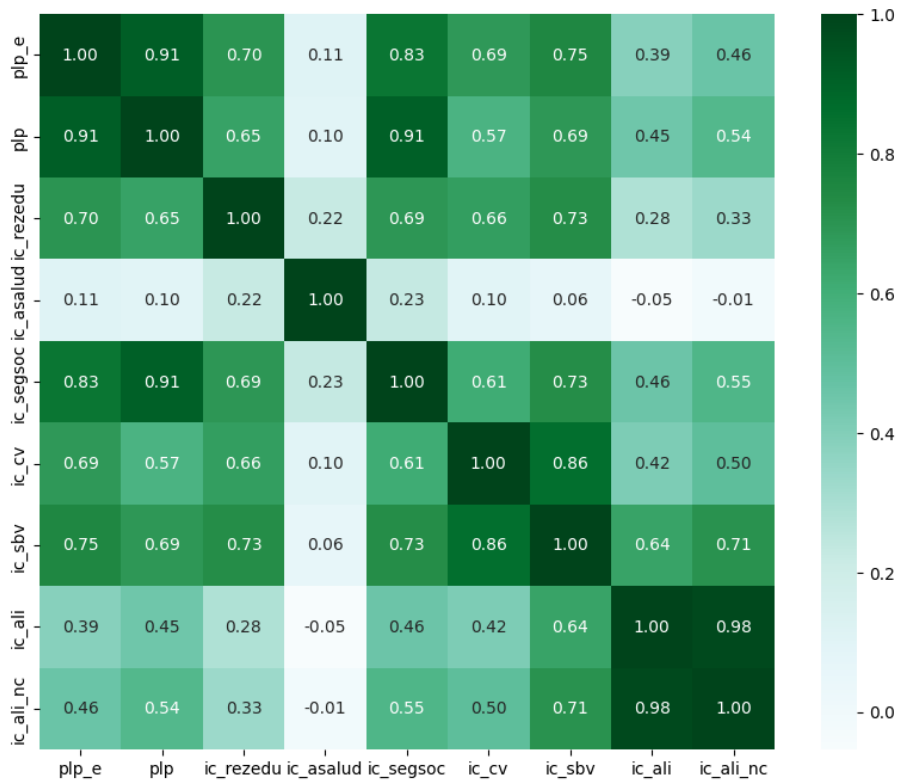
El Análisis Factorial, es un método de estadística multivariante, cuyo propósito es definir la estructura subyacente en una matriz de datos. Aborda el problema del análisis de las correlaciones entre un gran número de variables con la definición de una serie de dimensiones subyacentes comunes, conocidas como Factores. El análisis factorial permite identificar las dimensiones separadas de una estructura y posteriormente determinar el grado en el que se justifica cada variable por cada dimensión. Una vez determinadas estas dimensiones subyacentes, y la explicación de cada variable, se logran dos objetivos: el resumen y la reducción de datos y dimensiones. Es decir, el análisis factorial permite obtener dimensiones subyacentes de una matriz de datos, que cuando son entendidas e interpretadas, permiten describir los datos con un número menor de conceptos que las variables individuales originales. Y se puede obtener esta reducción de datos, al calcular las puntuaciones factoriales para cada dimensión subyacente y sustituir estas por las variables originales. (*Hair, J. F., 1999*)[16]

Para obtener las puntuaciones factoriales, se necesita de la matriz de cargas, esta matriz muestra la relación entre cada variable observada y los factores subyacentes o no observados identificados durante el análisis. Cada elemento de esta matriz representa la carga factorial, que es el coeficiente que indica cuánto de la varianza en una variable observada es explicada por un factor.

Es decir, la matriz de cargas dice que tan fuertemente asociada está cada variable observada con cada factor. Lo que permite interpretar las puntuaciones factoriales, al saber el peso de todas las variables observadas en los factores latentes no observados.

Existen distintos métodos para estimar la matriz de cargas, en este caso, se implementa el análisis Factorial de componentes principales, que está basado en la descomposición espectral de la matriz de correlaciones, a través de los valores y vectores propios de dicha matriz, para ello, lo primero es encontrar la matriz de correlaciones de los datos estandarizados de 2016, la cual se muestra a continuación:

Figura 16: Matriz de correlaciones 2016



Elaboración propia

Posteriormente, se realiza la siguiente operación:

$$R_x - \lambda I$$

Donde R_x es la matriz de correlaciones y λ es una variable que representa los valores propios. Esta operación consiste en restar λ a cada elemento de la diagonal de R_x . Posteriormente, se debe obtener el determinante de esta matriz, que en realidad será el polinomio característico de la matriz R_x , que es el siguiente:

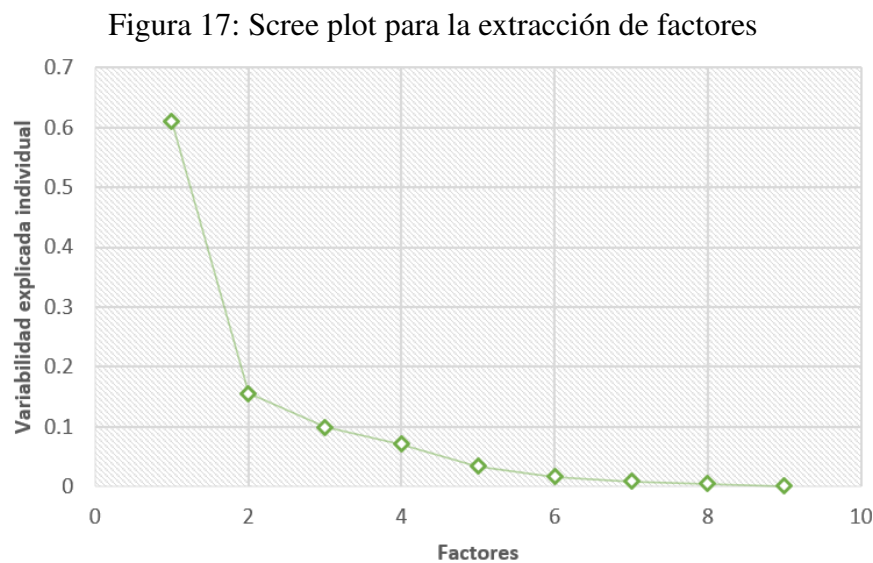
$$-\lambda^9 + 9.00\lambda^8 - 23.80\lambda^7 + 27.51\lambda^6 - 15.55\lambda^5 + 4.36\lambda^4 - 0.593\lambda^3 + 0.036\lambda^2 - 0.00089\lambda + 4.73 \times 10^{-6}$$

Una vez obtenido este polinomio, se procede a la obtención de los valores propios, estos son las raíces del polinomio característico, y son a su vez, los valores de λ que satisfacen la ecuación.

Los valores propios de R_x se muestran a continuación:

1. $\lambda_1 = 5.486335404485111$
2. $\lambda_2 = 1.4032014687383696$
3. $\lambda_3 = 0.8954299307150807$
4. $\lambda_4 = 0.6361611470143665$
5. $\lambda_5 = 0.30275391707446164$
6. $\lambda_6 = 0.145330079263165$
7. $\lambda_7 = 0.08233887649522643$
8. $\lambda_8 = 0.04122295812963896$
9. $\lambda_9 = 0.007226218084579581$

Para determinar el número de factores a estimar, se emplea el criterio de caída de la variabilidad explicada acumulada, lo que puede visualizarse en un scree plot que nos muestra el porcentaje de variabilidad explicada por cada valor propio. El criterio consiste en extraer x factores, donde x es el punto donde la curva de variabilidad explicada acumulada del scree plot comienza a rectificarse.



Elaboración propia

En este caso, el gráfico indica que la rectificación sucede a los tres valores propios, es decir, se sugiere la extracción de 3 factores, que acumularán un 86 % de la variabilidad

total. Por lo tanto, se deben extraer los vectores propios asociados a esos 3 primeros valores propios, los cuales son los siguientes:

$$\lambda_1 = 5.486335404485111 \quad v_1 = \begin{bmatrix} -0.3769359 \\ 0.20673322 \\ 0.2188745 \\ -0.22640449 \\ 0.22793296 \\ -0.636818 \\ -0.04653166 \\ -0.49484166 \\ -0.10719889 \end{bmatrix}$$

$$\lambda_2 = 1.4032014687383696 \quad v_2 = \begin{bmatrix} -0.37465822 \\ 0.14129258 \\ 0.15317393 \\ -0.49207846 \\ 0.09822353 \\ -0.06427165 \\ 0.13770563 \\ 0.72425528 \\ 0.12952568 \end{bmatrix}$$

$$\lambda_3 = 0.8954299307150807 \quad v_3 = \begin{bmatrix} -0.3317368 \\ 0.31244079 \\ 0.08407707 \\ 0.28171538 \\ -0.80906029 \\ -0.08007141 \\ 0.19957543 \\ 0.01806412 \\ -0.06893684 \end{bmatrix}$$

Se procede a la estimación de la matriz de cargas, para ello se plantea que:

$$LM16 = V_3^{Rx} (D_3^{Rx})^{\frac{1}{2}}$$

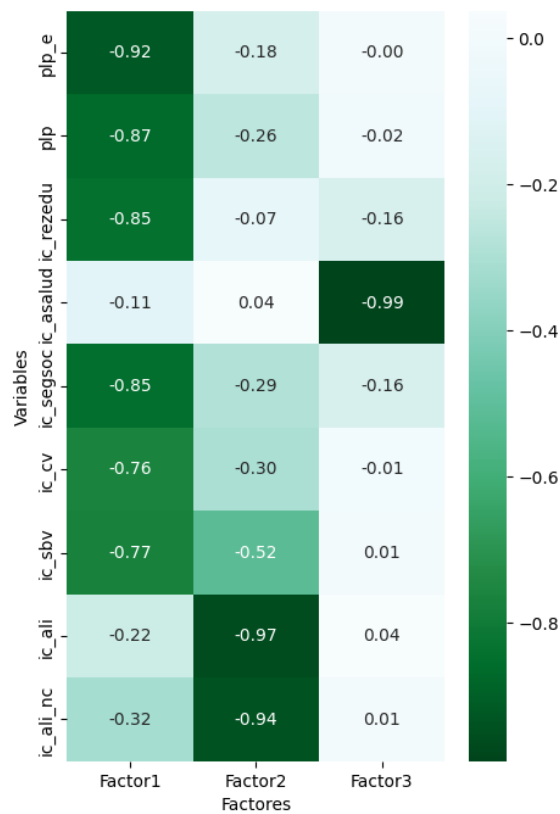
donde:

- $LM16$: es la matriz de cargas para 2016;
- V_3^{Rx} : es la matriz con los 3 vectores propios de R_x ; y
- D_3^{Rx} : es la matriz diagonal de los 3 valores propios más grandes de R_x .

Una vez concluido este proceso, se aplicará un criterio de rotación a los datos que beneficie la interpretación de los mismos, en este caso se emplea varimax, que es una rotación que respeta la ortogonalidad de los factores.

La matriz de cargas para 2016, se muestra a continuación:

Figura 18: Matriz de cargas factoriales para 2016 (LM16)



La coloración de la matriz proporciona una representación visual de la fuerza y de las asociaciones y los pesos de los factores: los tonos más oscuros indican las cargas más fuertes, mientras que los más claros, son los más débiles.

Se identifican asociaciones muy importantes de *plp_e*, *plp*, *ic_rezedu*, *ic_segsoc*, *ic_cv*, *ic_sbv* en el primer factor; *ic_ali* y *ic_ali_nc* en el segundo; y de *ic_asalud* en el tercero.

Las cargas factoriales tienen mayormente signo negativo, lo que implica que al calcular las puntuaciones factoriales (que son el producto punto de los vectores fila de la matriz original estandarizada por el vector columna de la matriz de cargas respectiva a cada factor) estas se interpretarán de la siguiente forma:

A mayor valor en las variables originales, más negativa será la puntuación de estas en los tres factores, indicando así, que los estados con puntuaciones más pequeñas (más negativas) serán los que mayor presencia de pobreza multidimensional presentan. Mientras que los de mayores puntuaciones (menos negativas) serán los menos afectados. Es decir, mientras más negativa sea una puntuación factorial, más pobre multidimensionalmente será el respectivo estado en el respectivo factor.

Gracias a el análisis factorial, las puntuaciones factoriales pueden interpretarse como un rango de la pobreza multidimensional, ya que al ser el producto punto de los vectores fila de la matriz original y los vectores columna de la matriz de cargas; las puntuaciones factoriales son combinaciones lineales de todas las variables reflejos de la multidimensionalidad de la pobreza estipulada en la LGDS y su Art.36.

Utilizando la matriz de cargas de 2016, se calculan las puntuaciones factoriales para 2018, 2020, y 2022, esto con la finalidad, de que las puntuaciones sean comparables entre sí.

Las puntuaciones factoriales para cada año, se presentan a continuación:

Figura 19: Puntuaciones factoriales 2016

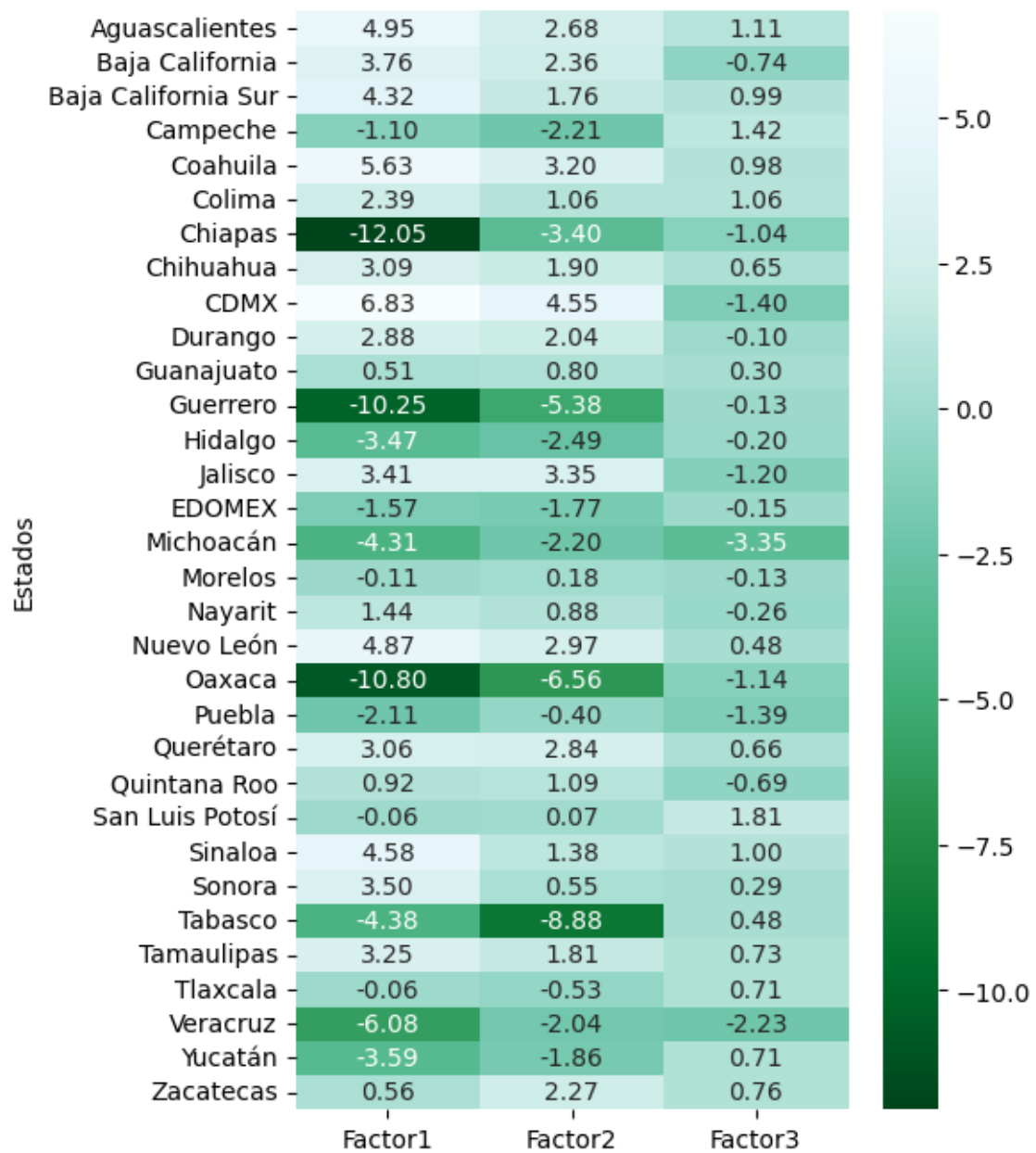


Figura 20: Puntuaciones factoriales 2018

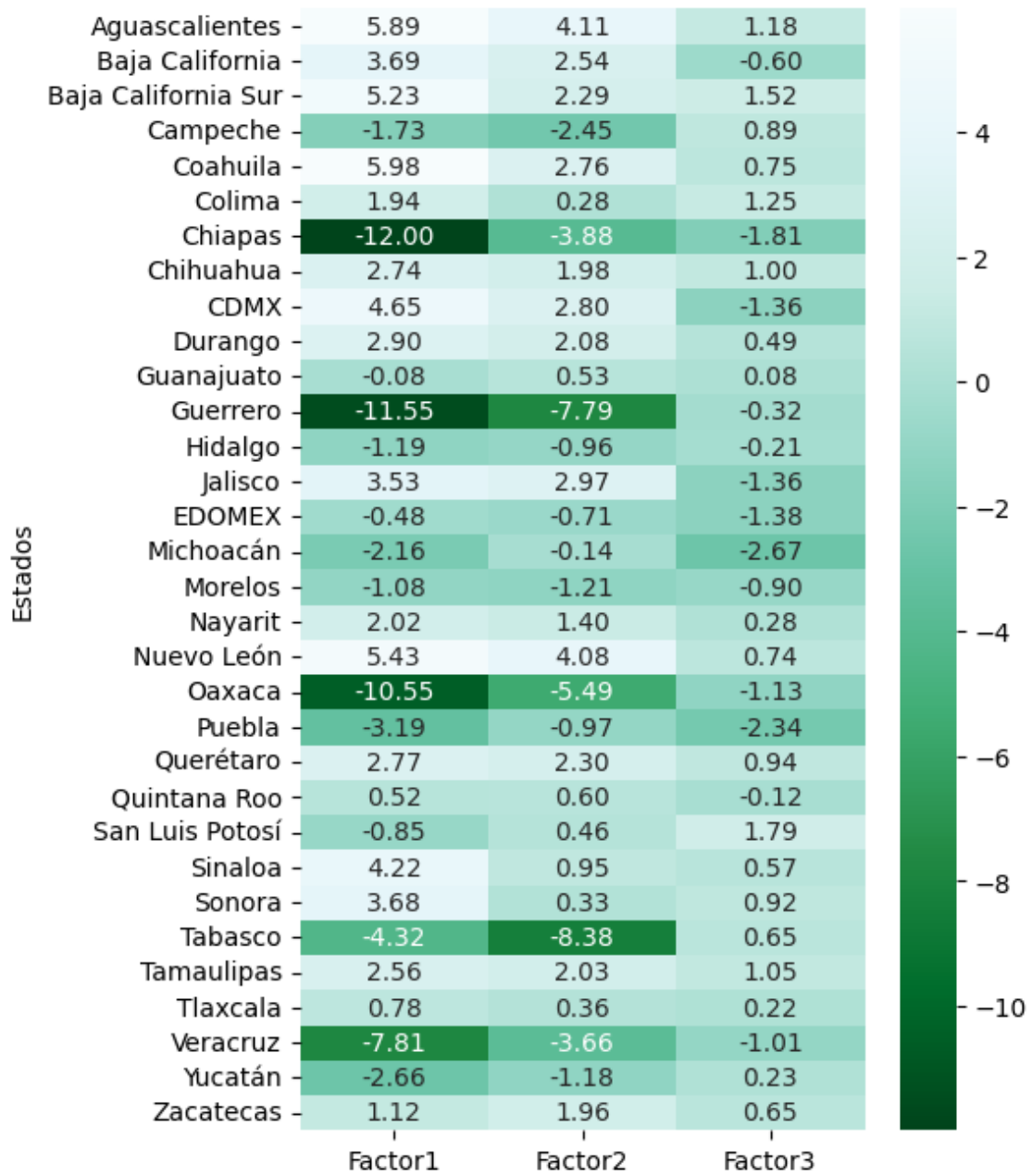


Figura 21: Puntuaciones factoriales 2020

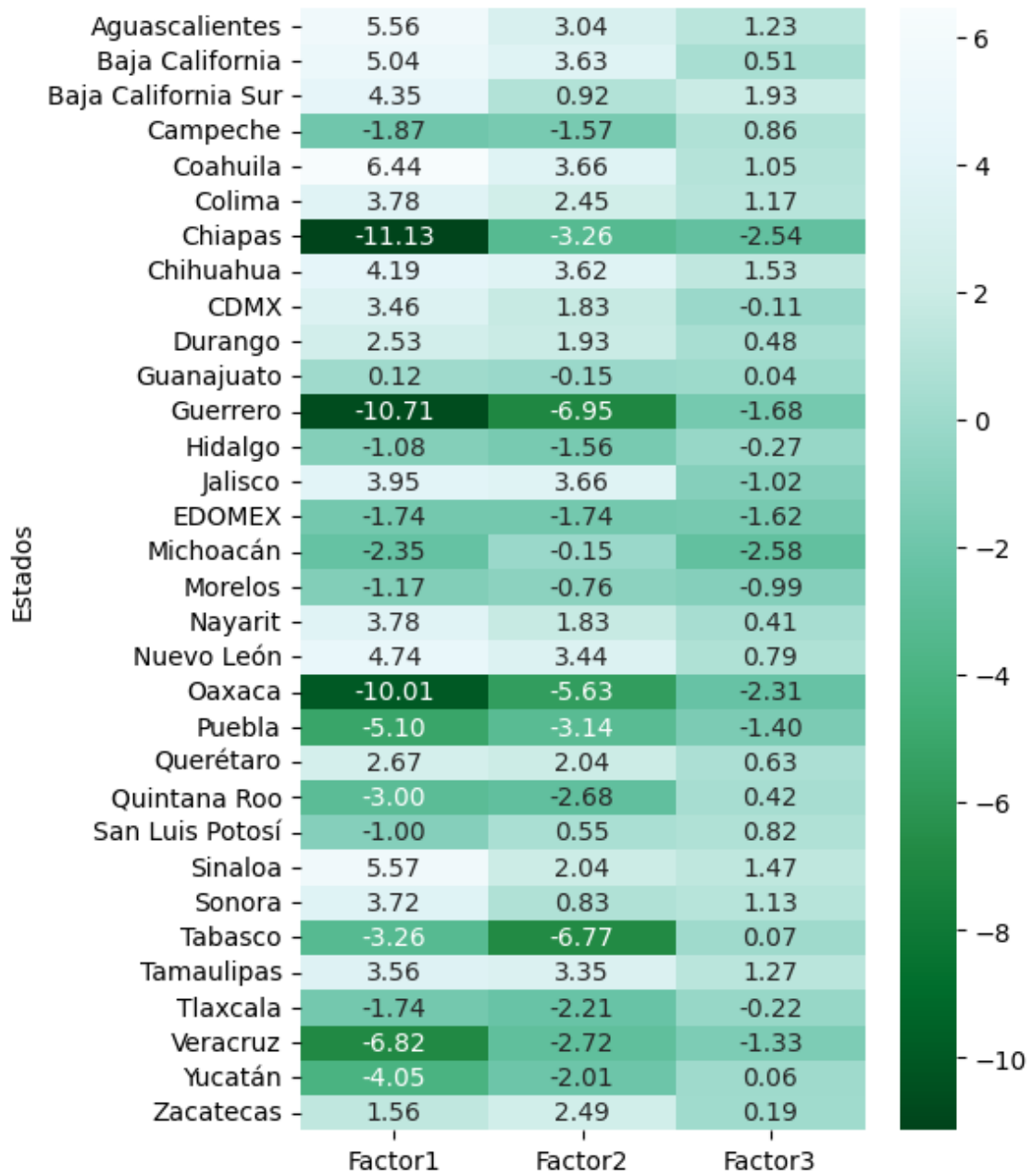
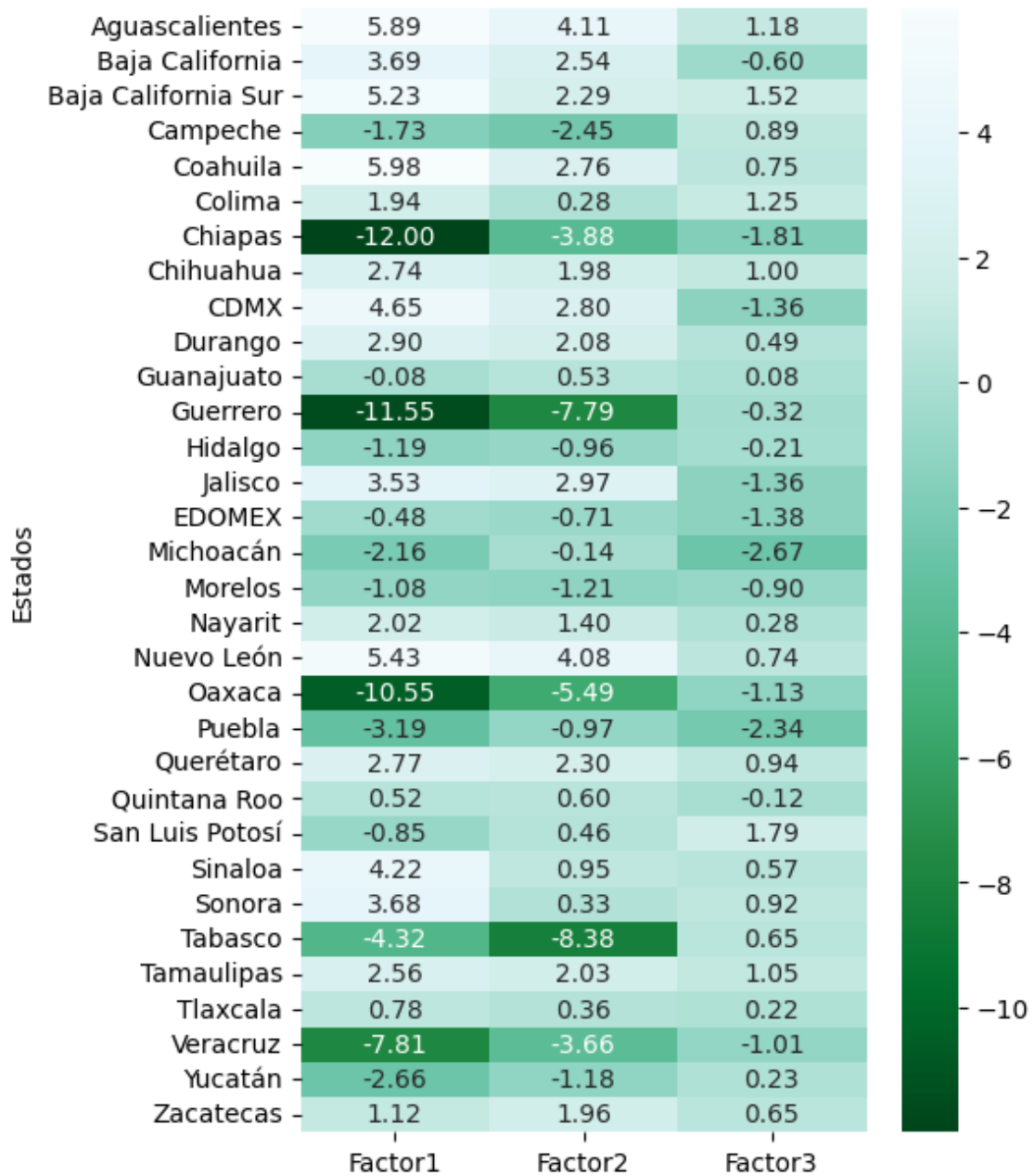


Figura 22: Puntuaciones factoriales 2022



3.4.4. Análisis de Cluster

El análisis de cluster es un conjunto de técnicas multivariantes cuyo principal propósito es el de agrupar objetos, basándose en las características que estos poseen. El análisis de cluster clasifica objetos de tal forma que cada objeto es muy parecido a los que existen en el mismo conglomerado respecto a algún criterio de selección determinado. Los conglomerados de objetos resultantes de este análisis deben ser homogéneos internamente (dentro del conglomerado) y heterogéneos externamente (con los otros conglomerados).[16]

Para este análisis, se utiliza un algoritmo de cluster de aprendizaje de máquina no supervisado, llamado K-Means.

En este algoritmo, se eligen K puntos al azar como los centros iniciales de los grupos, a estos se les denomina centroides. Posteriormente, cada punto del conjunto de datos se asigna al grupo cuyo centroide es el más cercano. Esto utilizando alguna medida de distancia, que en este caso será la distancia euclídea.

La fórmula para la distancia euclídea entre dos puntos x y y en un espacio de dimensiones N es:

$$d(x, y) = \sqrt{\sum_{i=1}^N (x_i - y_i)^2}$$

Donde x_i y y_i son las coordenadas de los puntos x y y , respectivamente.

Una vez todos los puntos han sido asignados a un grupo, se recalculan los centroides de cada grupo. El nuevo centroide de un grupo j tiene los puntos P_1, P_2, \dots, P_n , el nuevo centroide C_j se calcula como:

$$C_j = \frac{1}{n} \sum_{i=1}^n P_i$$

Aquí, n es el número de puntos en el grupo j .

Estos pasos se repiten hasta que se cumpla un criterio de detención. Este criterio puede ser un número fijo de iteraciones, que en este caso es de 100.

El objetivo del K-Means es minimizar la suma de las distancias cuadradas de cada punto a su centroide, lo cual es conocido como la inercia o la varianza dentro del grupo. Matemáticamente, la inercia se define como:

$$\text{Inercia} = \sum_{j=1}^K \sum_{i \in S_j} (x_i - C_j)^2$$

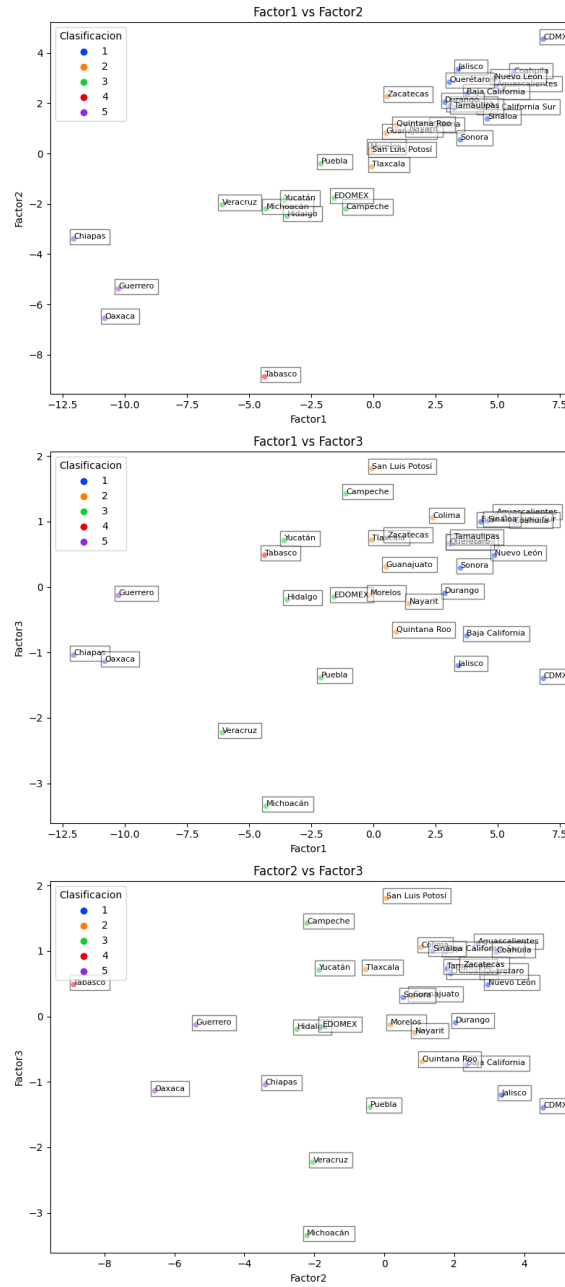
Donde S_j es el conjunto de puntos en el grupo j y C_j es el centroide de ese grupo.

En este caso, los criterios que se emplean para el cluster, serán las puntuaciones factoriales en los tres factores obtenidos del análisis factorial. Esto para cada año de estudio. Y los grupos a formar serán 5, esto es un criterio a priori y por elección propia.

A continuación se muestran los resultados del cluster y sus proyecciones. El espacio debería idealmente proyectarse en 3 dimensiones, sin embargo dada la imposibilidad de graficar de esta forma en el documento, se harán 3 proyecciones bidimensionales que proyecten dicho espacio tridimensional.

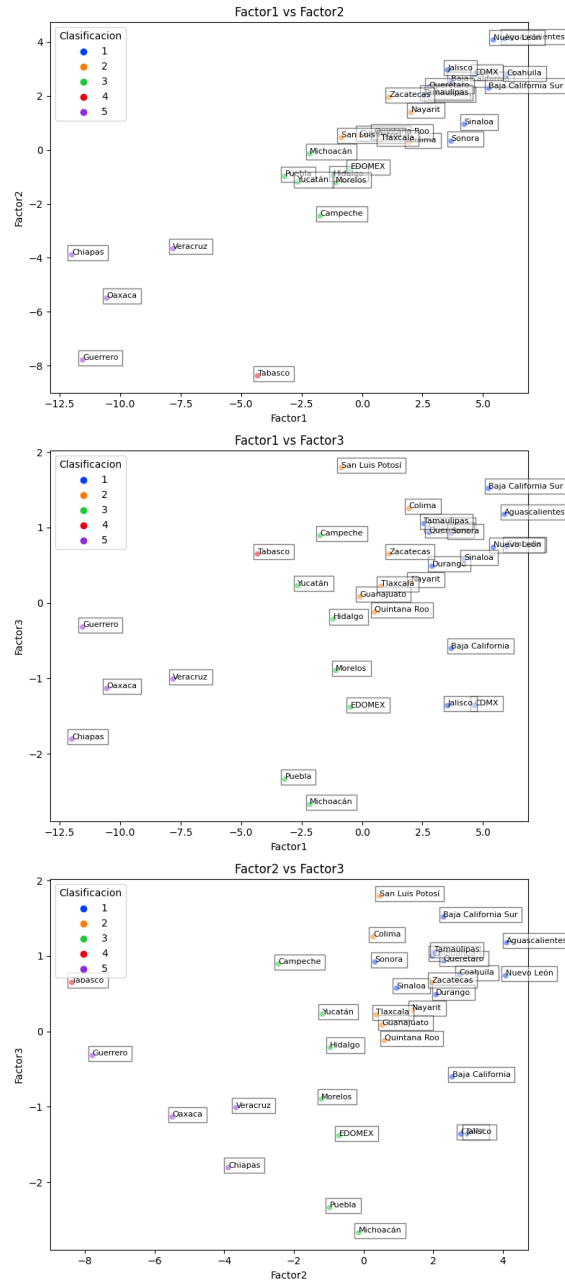
Resultados y proyecciones del cluster 2016

Figura 23: Proyecciones 2016



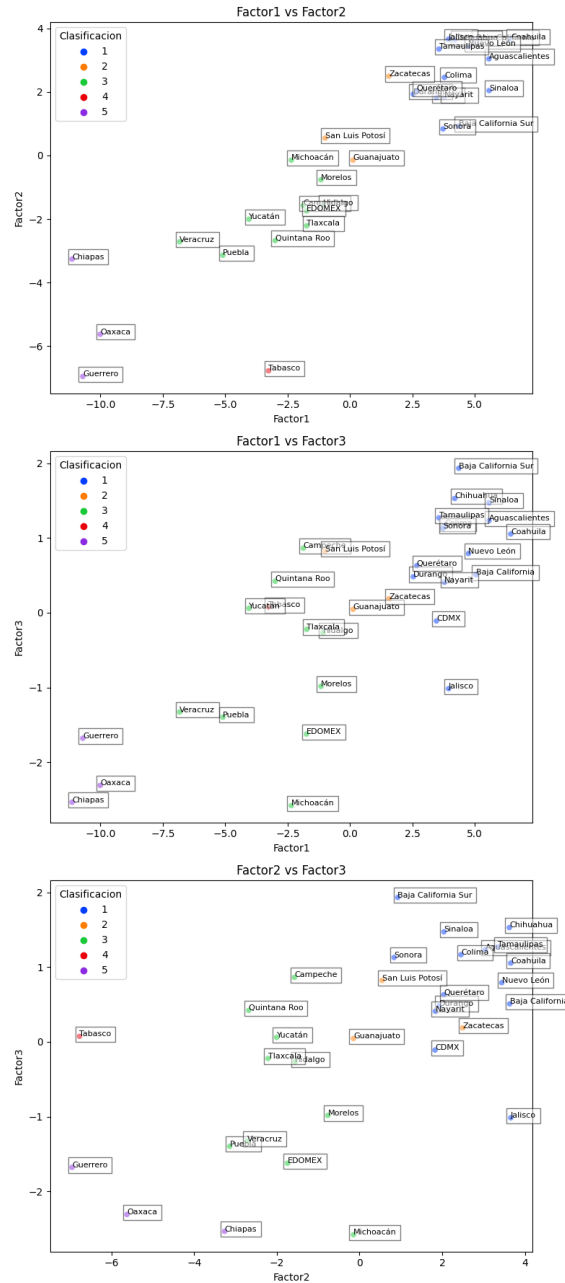
Resultados y proyecciones del cluster 2018

Figura 24: Proyecciones 2018



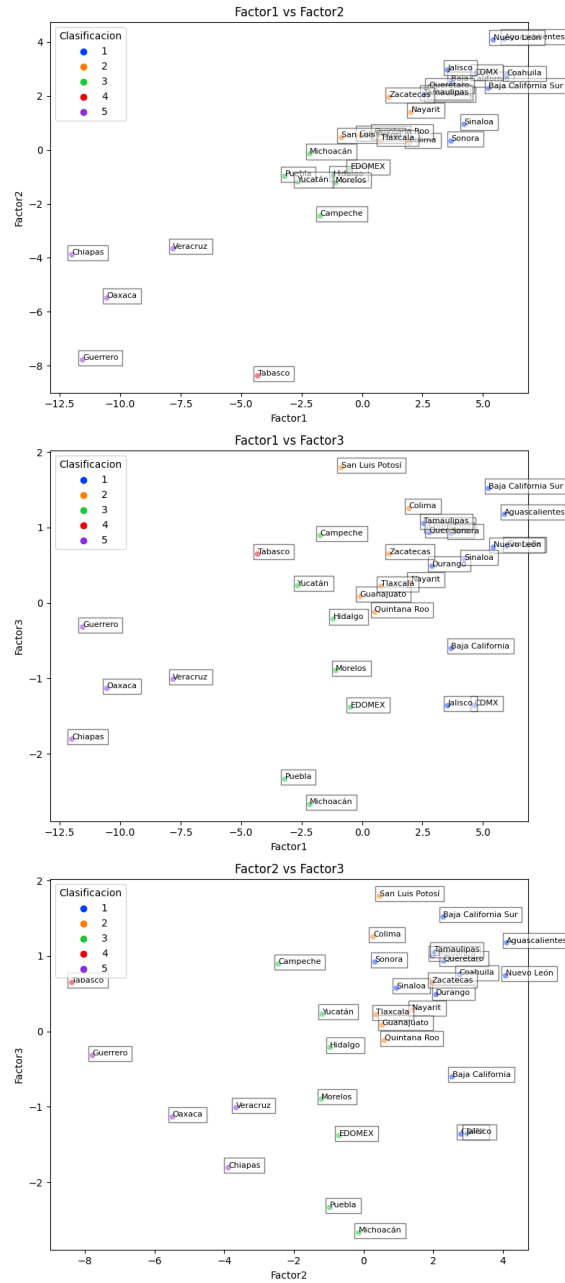
Resultados y proyecciones del cluster 2020

Figura 25: Proyecciones 2020



Resultados y proyecciones del cluster 2022

Figura 26: Proyecciones 2022



Como puede visualizarse en los gráficos anteriores, se clasificaron las entidades federativas de la república en 5 grupos, cada uno identificado con un color y un número. Donde 1 es el grupo de los estados menos pobres multidimensionalmente y 5 es el de los estados con mayor pobreza multidimensional. En los gráficos y las dispersiones de los factores puede observarse como la mayoría de los grupos son homogéneos internamente, entre sus miembros, y son heterogéneos con los demás grupos.

Sin embargo también existen casos como los de los grupos 4, y 5. Donde el grupo 4 está únicamente conformado por el estado de Tabasco, cuyas proyecciones factoriales fueron tan heterogéneas al resto de estados que el algoritmo le asignó un cluster propio y único. El caso del grupo 5, también es bastante particular, porque gráficamente se ve como difiere de los demás grupos e incluso es medianamente heterogéneo entre los miembros del cluster. Y además, es el cluster con mayor pobreza multidimensional.

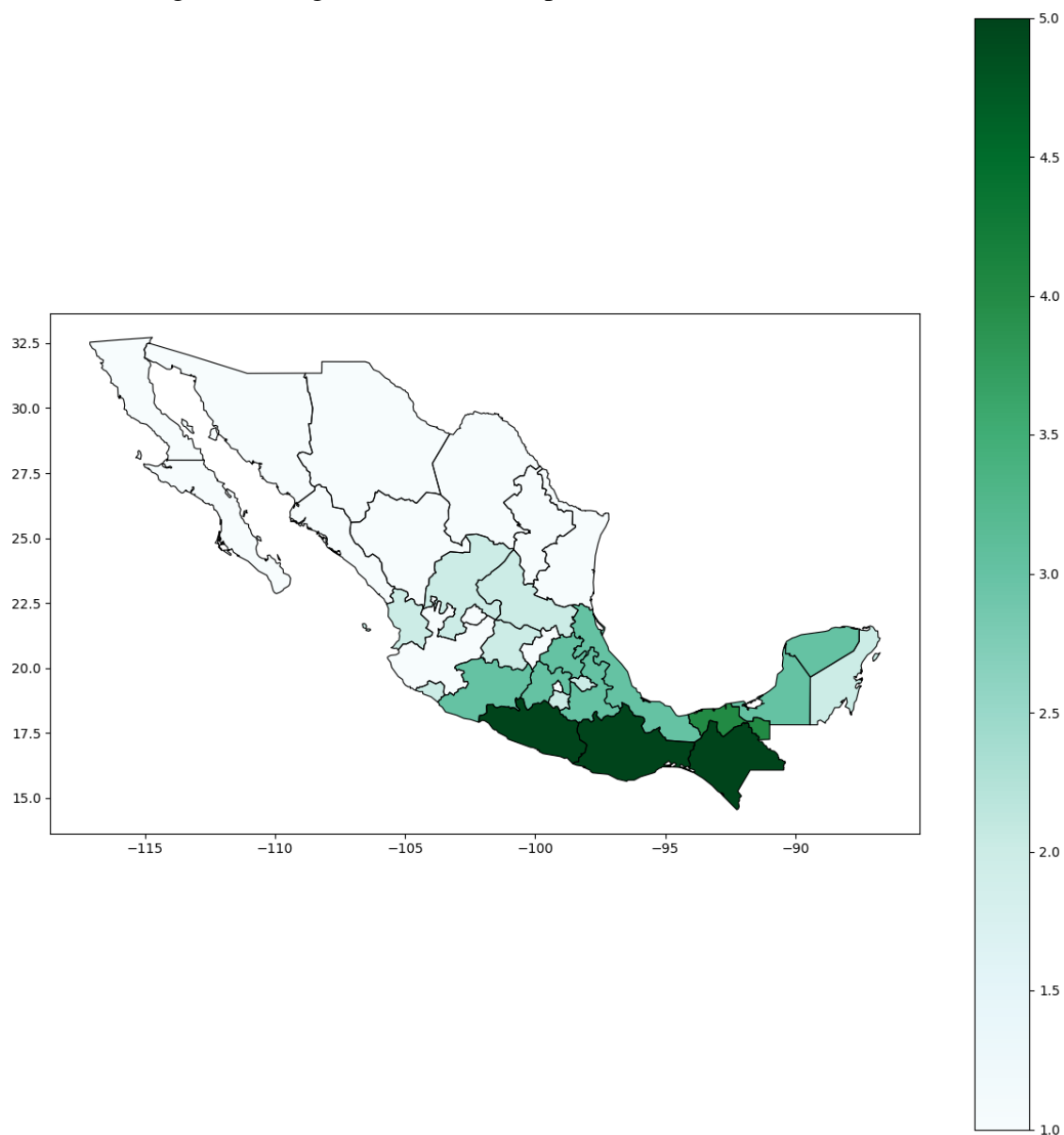
4. Resultados

4.1. Regionalizaciones

Concluidos los procesos de análisis de datos que incluyeron un modelo de análisis factorial con componentes principales y otro de cluster con k-means, se obtuvieron las regionalizaciones de la pobreza multidimensional para cada año estudiado. Los resultados de dicha regionalización se muestran a continuación.

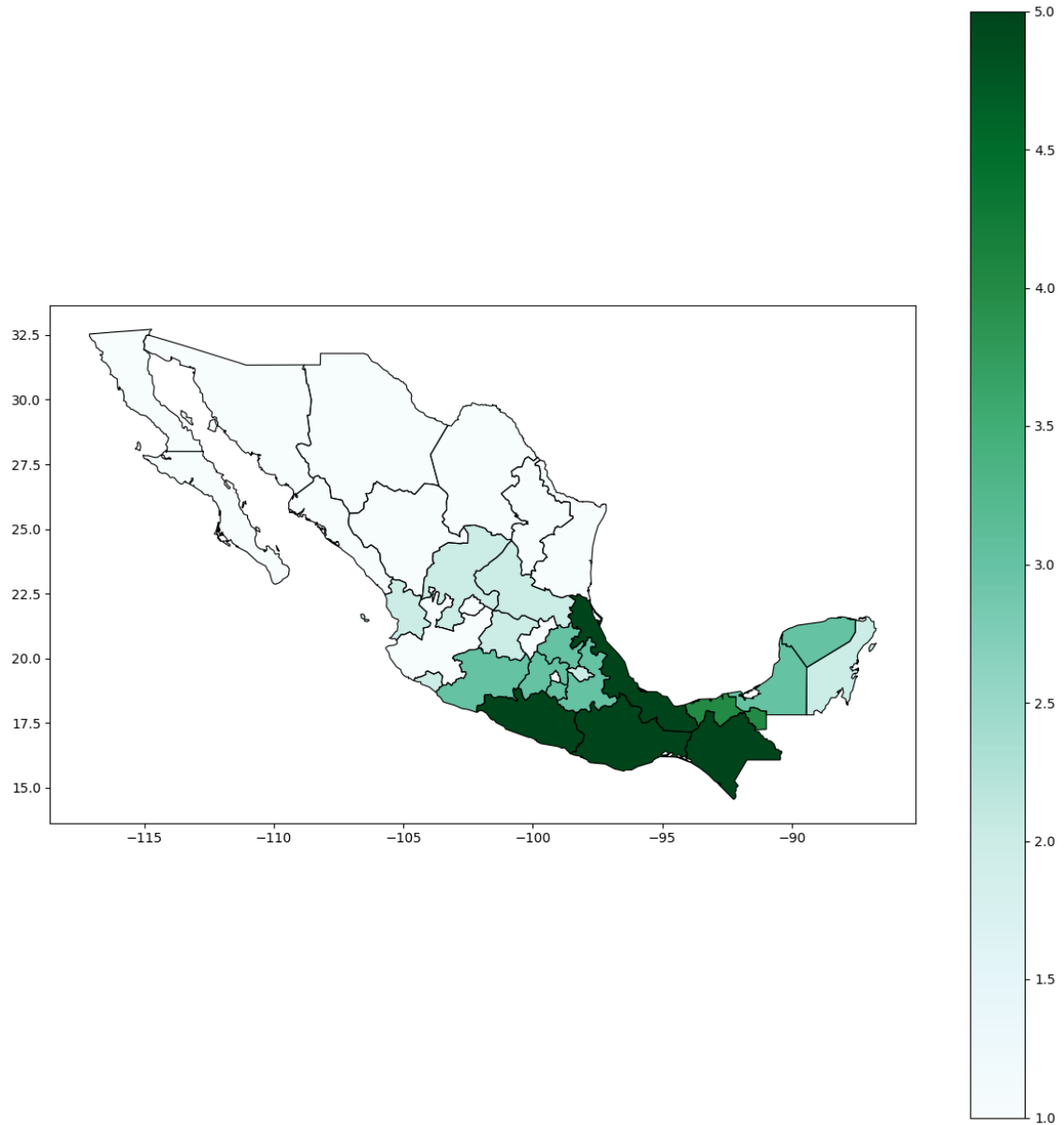
Regionalización de la pobreza multidimensional 2016

Figura 27: Regionalización de la pobreza multidimensional 2016



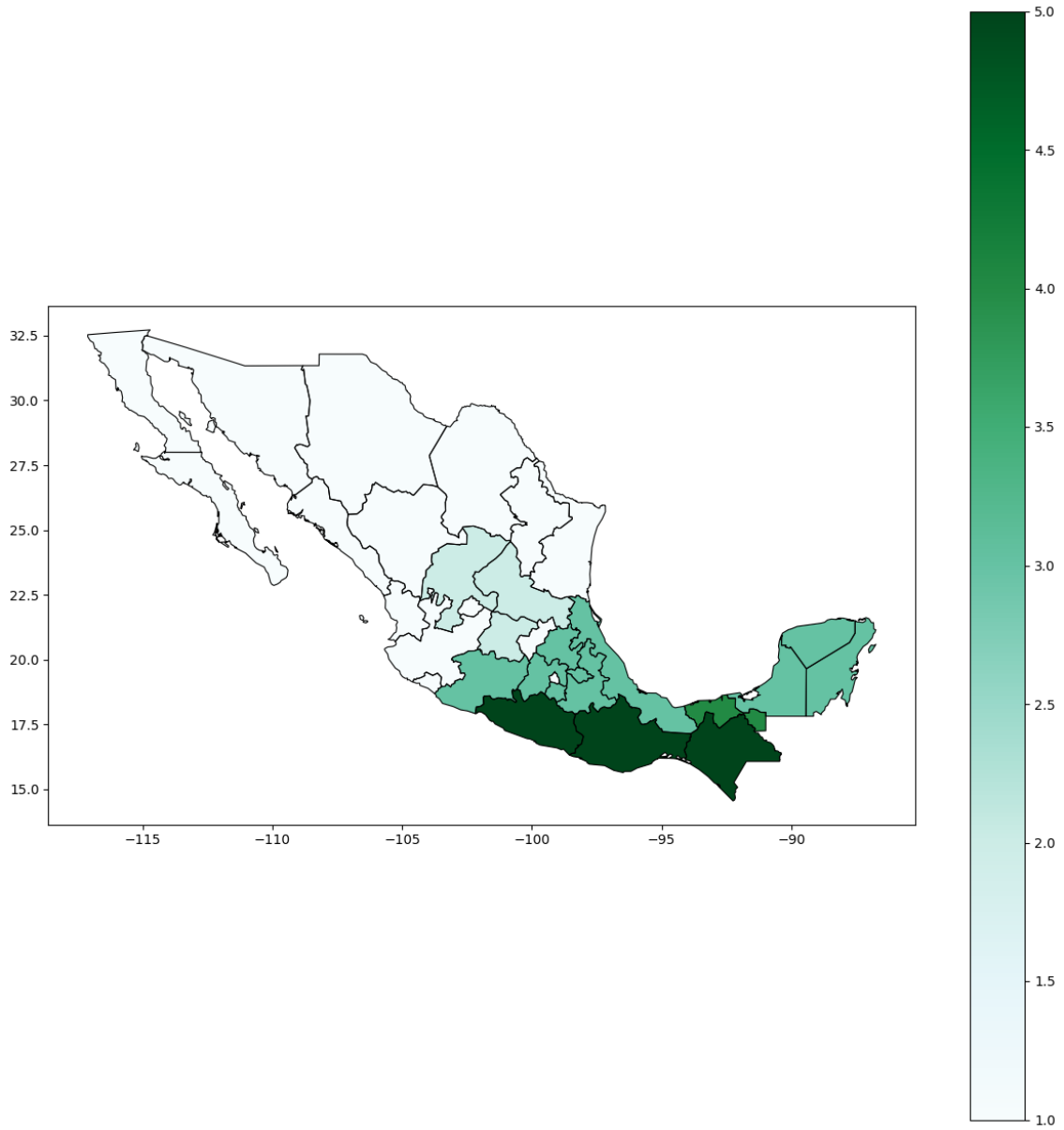
Regionalización de la pobreza multidimensional 2018

Figura 28: Regionalización de la pobreza multidimensional 2018



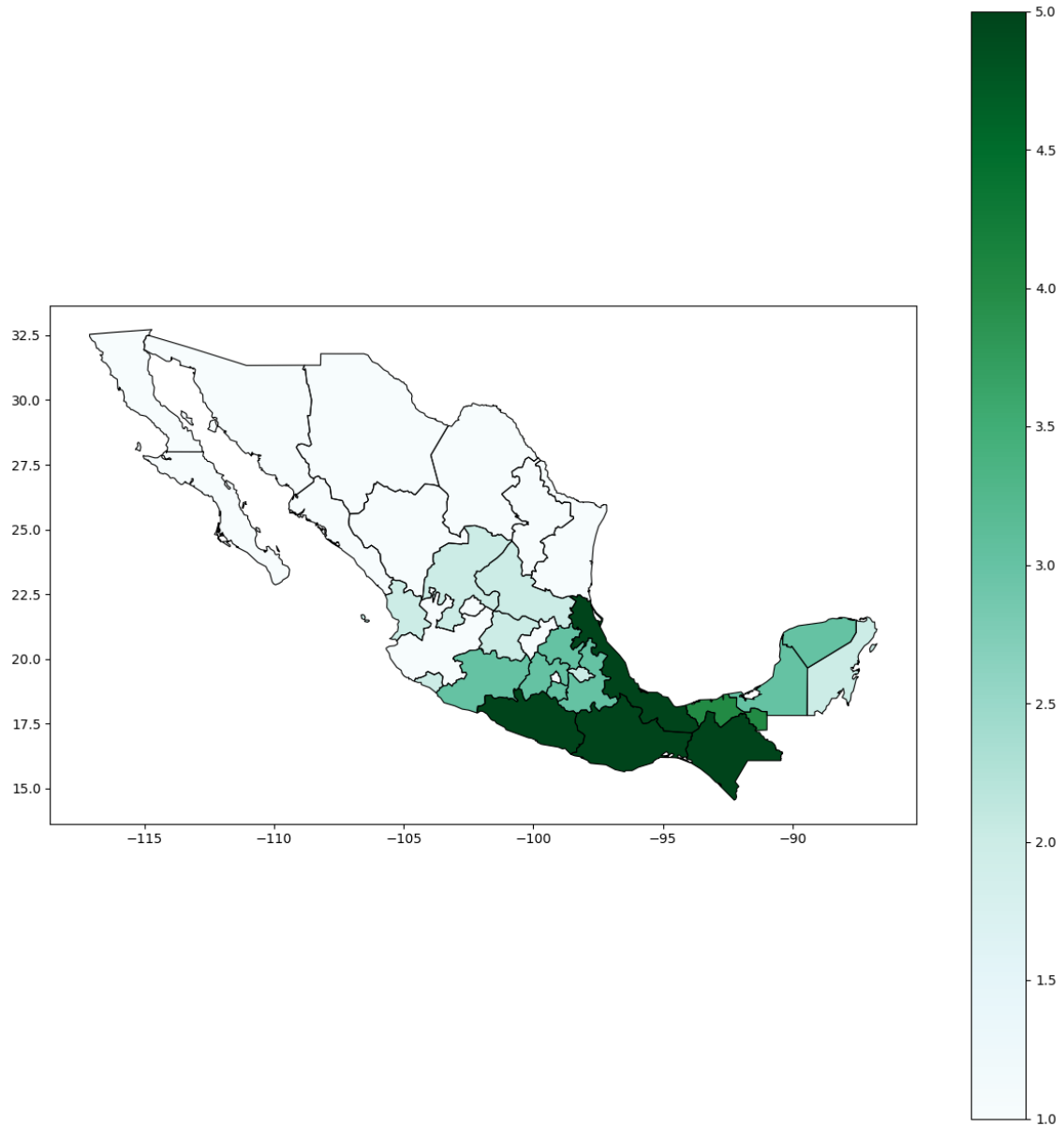
Regionalización de la pobreza multidimensional 2020

Figura 29: Regionalización de la pobreza multidimensional 2020



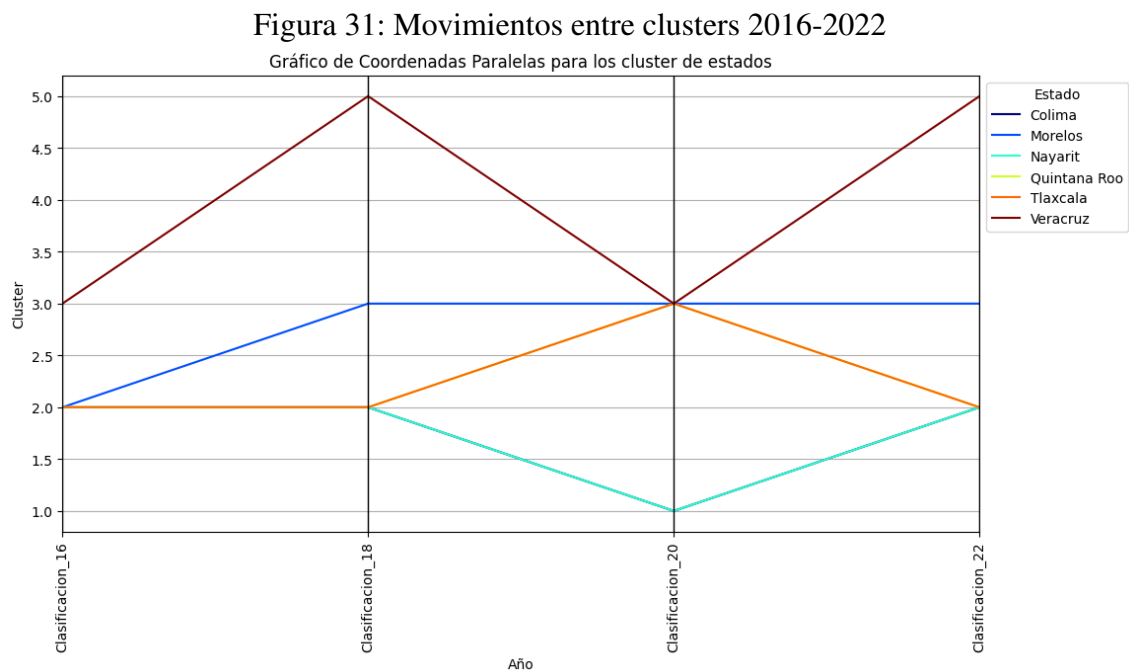
Regionalización de la pobreza multidimensional 2022

Figura 30: Regionalización de la pobreza multidimensional 2022



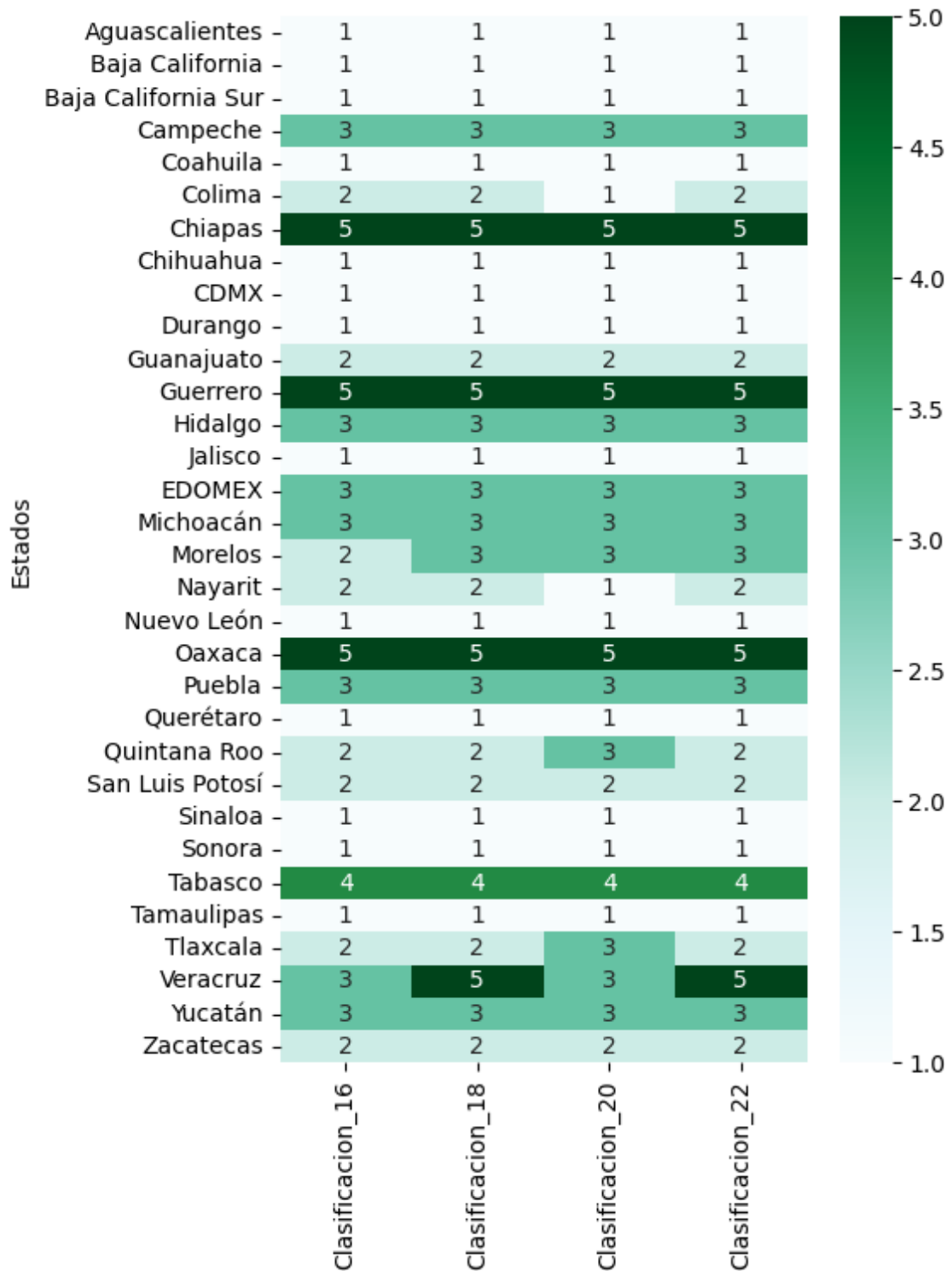
Tras las visualizaciones, se observa como aunque hay contigüidad geográfica, esta no es perfecta, lo que demuestra la importancia de una regionalización enfocada en la pobreza multidimensional y como esta puede mejorar la focalización de políticas públicas y de desarrollo social, mejor que una regionalización geográfica a priori.

Analizando los mapas, se observa que la mayoría de los estados se mantuvieron en su cluster inicial en el intervalo de los 6 años estudiados, siendo Colima, Morelos, Nayarit, Quintana Roo, Tlaxcala y Veracruz, los únicos que se movieron.



Las clasificaciones de los estados en clusters y su evolución en el tiempo, se resumen en el siguiente gráfico.

Figura 32: Clasificación de los estados en clusters 2016-2022



4.2. Estadísticas por cluster

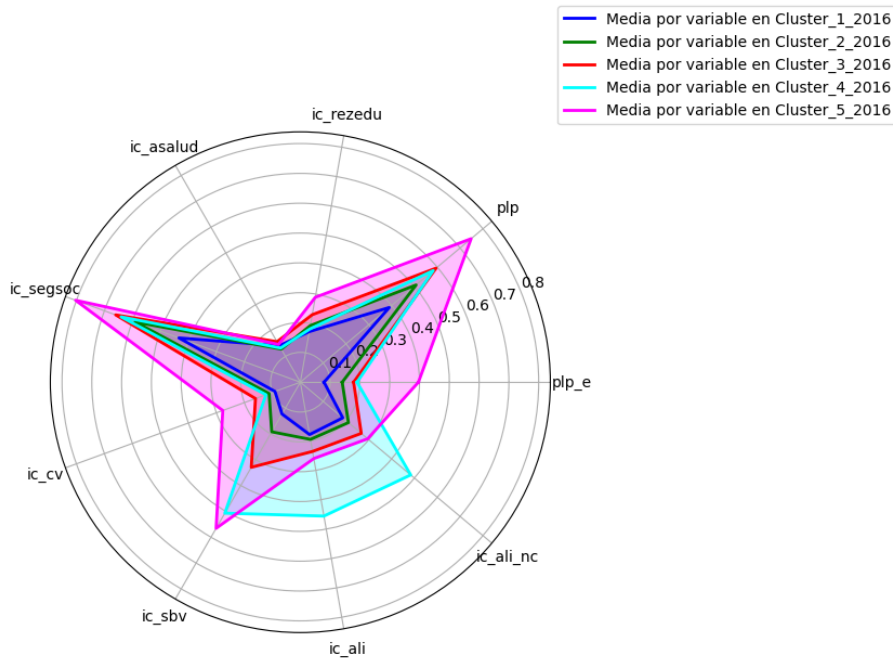
Una vez clasificado cada estado en un cluster, se pueden obtener estadísticas descriptivas más acertadas y menos sesgadas, debido a la homogeneidad entre los grupos formados. Un ejemplo de ello, se muestra en los siguientes gráficos, donde se obtiene la media por variable de cada cluster. Esto, proporciona una visualización de como la pobreza multidimensional se manifiesta de distinta forma en cada cluster/región, donde las carencias e intensidades entre los grupos formados difiere significativamente, ayudando a identificar los problemas de pobreza en cada una de las dimensiones, y como estas afectan a cada grupo, y como han evolucionado en el tiempo; proporcionando más y mejor información, para una focalización precisa y efectiva de las políticas publicas y de desarrollo social.

A continuación se presentan las medias variable y por cluster, junto con su representación gráfica para cada año

Cuadro 4: Media por variable y cluster en 2016

	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
plp_e	0.07947509	0.14048007	0.17875066	0.19021066	0.39627791
plp	0.38963875	0.50719272	0.59469242	0.58224907	0.74829802
ic_rezedu	0.17230016	0.1926473	0.22981487	0.18091698	0.28906881
ic_asalud	0.14114386	0.1281905	0.15673738	0.13228005	0.14615611
ic_segsoc	0.43382281	0.59049192	0.65977655	0.64048947	0.80301166
ic_cv	0.09066984	0.11165835	0.16033824	0.12283147	0.27712537
ic_sbv	0.12316023	0.19204787	0.32942488	0.50697026	0.56556436
ic_ali	0.17822436	0.19470523	0.23566531	0.45570012	0.25985869
ic_ali_nc	0.18588253	0.2106485	0.26689947	0.48358116	0.29498204

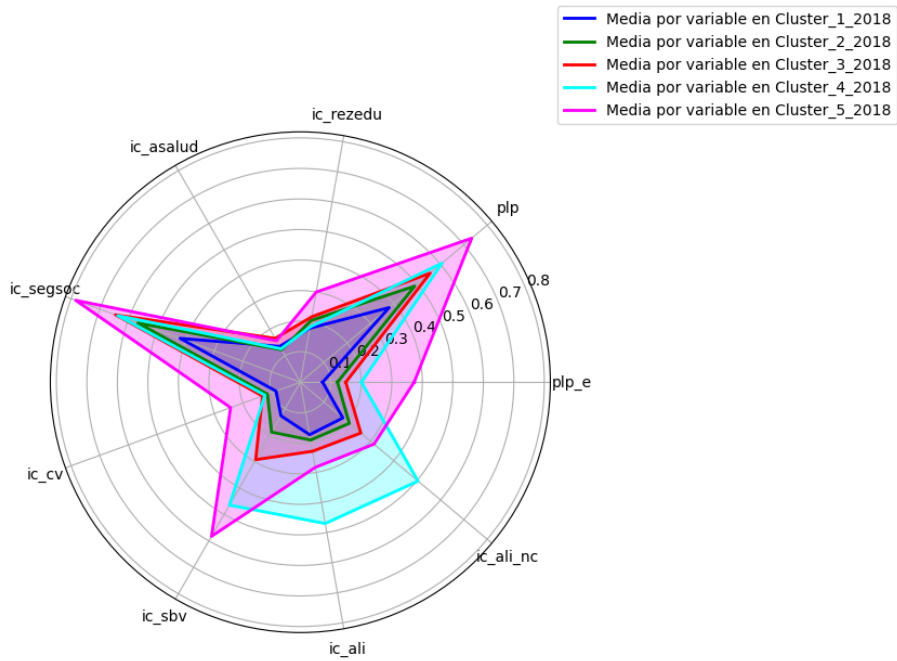
Figura 33: 'Pobreza multidimensional por cluster 2016



Cuadro 5: Media por variable y cluster en 2018

	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
plp_e	0.07947509	0.14310016	0.15838469	0.19021066	0.36338374
plp	0.38963875	0.50316437	0.576369	0.58224907	0.72713729
ic_rezedu	0.17230016	0.19312128	0.21793795	0.18091698	0.28491858
ic_asalud	0.14114386	0.12608983	0.14952604	0.13228005	0.15796072
ic_segsoc	0.43382281	0.58167525	0.65325036	0.64048947	0.77673172
ic_cv	0.09066984	0.10981892	0.15274562	0.12283147	0.2522647
ic_sbv	0.12316023	0.18669608	0.30293168	0.50697026	0.52791396
ic_ali	0.17822436	0.19387661	0.23235914	0.45570012	0.25080621
ic_ali_nc	0.18588253	0.20930345	0.26400978	0.48358116	0.28130944

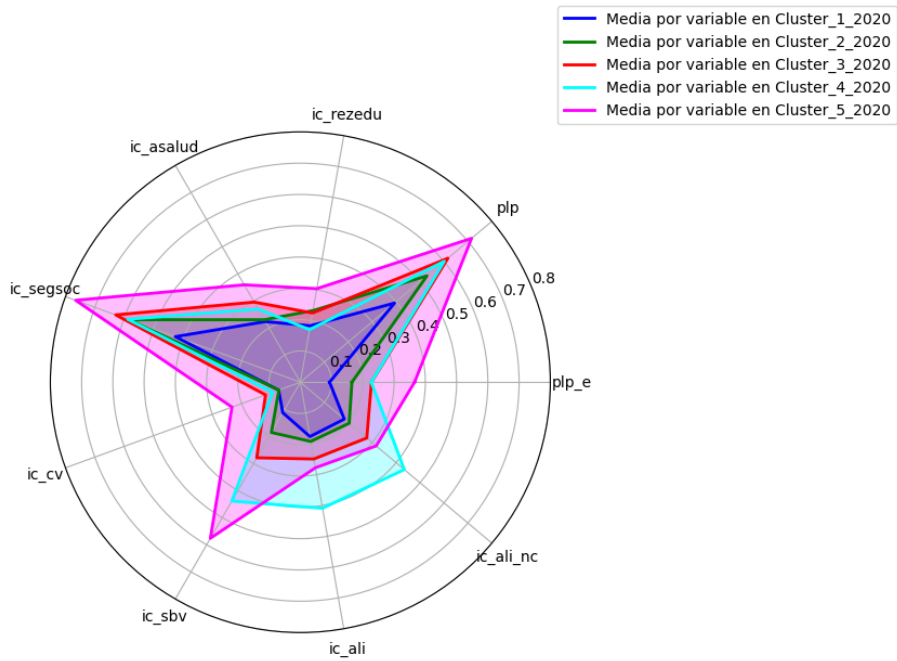
Figura 34: Pobreza multidimensional por cluster en 2018



Cuadro 6: Media por variable y cluster en 2020

	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
plp_e	0.08339191	0.16073132	0.16751987	0.19021066	0.39627791
plp	0.39338999	0.53962118	0.57659792	0.58224907	0.74829802
ic_rezedu	0.17502638	0.21297309	0.21254696	0.18091698	0.28906881
ic_asalud	0.13987916	0.11068118	0.15273249	0.13228005	0.14615611
ic_segsoc	0.44846167	0.6026985	0.64470471	0.64048947	0.80301166
ic_cv	0.09347373	0.08366414	0.1541244	0.12283147	0.27712537
ic_sbv	0.13046019	0.19555155	0.28998826	0.50697026	0.56556436
ic_ali	0.18146086	0.18477725	0.2247971	0.45570012	0.25985869
ic_ali_nc	0.18934005	0.19549905	0.25433593	0.48358116	0.29498204

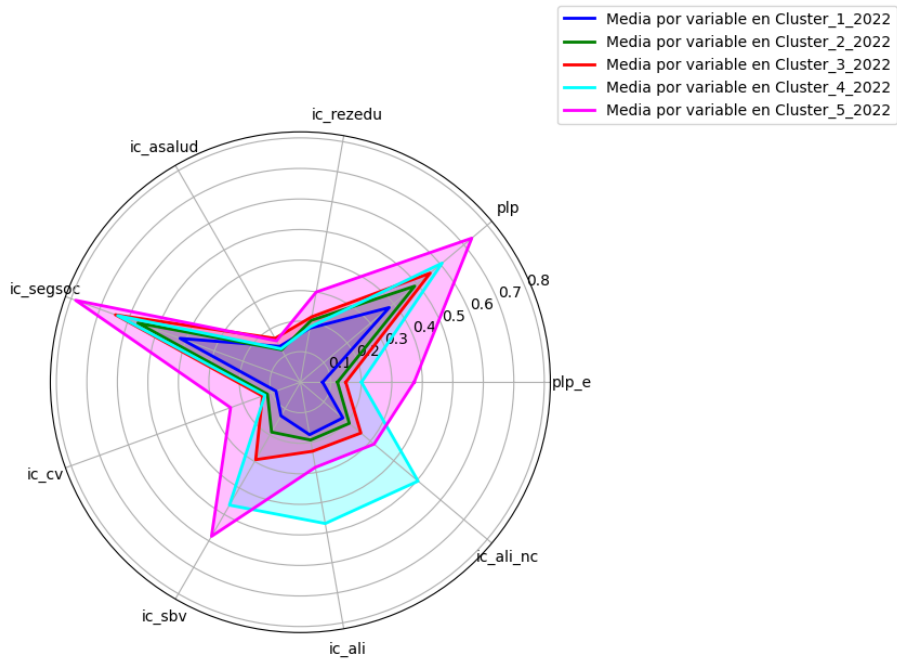
Figura 35: Pobreza multidimensional por cluster en 2020



Cuadro 7: Media por variable y cluster en 2022

	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
plp_e	0.07947509	0.14310016	0.15838469	0.19021066	0.36338374
plp	0.38963875	0.50316437	0.576369	0.58224907	0.72713729
ic_rezedu	0.17230016	0.19312128	0.21793795	0.18091698	0.28491858
ic_asalud	0.14114386	0.12608983	0.14952604	0.13228005	0.15796072
ic_segsoc	0.43382281	0.58167525	0.65325036	0.64048947	0.77673172
ic_cv	0.09066984	0.10981892	0.15274562	0.12283147	0.2522647
ic_sbv	0.12316023	0.18669608	0.30293168	0.50697026	0.52791396
ic_ali	0.17822436	0.19387661	0.23235914	0.45570012	0.25080621
ic_ali_nc	0.18588253	0.20930345	0.26400978	0.48358116	0.28130944

Figura 36: Pobreza multidimensional por cluster 2022



5. Discusión

Dentro de los resultados se puede observar la movilidad social de Colima, Morelos, Nayarit, Quintana Roo, Tlaxcala y Veracruz. Los años donde se aprecian cambios en la mayoría de estos estados fueron 2020 y 2022, siendo Morelos la excepción moviéndose en 2018 para luego mantenerse constante.

Según los cambios en la región, los estados se dividen en dos grupos:

1. Aquellos que se catalogan en una región menos pobre para 2020, en relación al 2018, para luego volver a la región donde estaban para el 2022: Colima, Nayarit y Veracruz.
2. Los que se movieron en sentido opuesto, es decir, se catalogan en una región más pobre para 2020, en relación al 2018, para luego volver a la región donde estaban para el 2022: Quintana y Tlaxcala.

La presencia de una movilidad social positiva o negativa puede deberse a distintos factores, tanto internos como externos. Analizando el caso de ambos grupos mencionados anteriormente se plantean las posibles causas:

En el año 2020, debido a la pandemia, los estados más afectados fueron los que se dedicaban al turismo, siendo uno de ellos Quintana Roo. En cambio, Colima, Nayarit y Veracruz, con su condición de puertos, continuaron con sus actividades comerciales aún en el 2020.

Durante el año 2022 Quintana Roo pudo retomar sus actividades turísticas y este estado, al igual que el resto del sureste, se vio favorecido por la inversión pública en la región.

A diferencia de los estudios antes mencionados sobre regionalización de la pobreza, áreas geográficamente separadas no eran comúnmente asociadas; por ejemplo: Colima, Nayarit y Veracruz.

Finalmente, encontramos similitudes con Ros (2014) respecto a que la creación de política pública (beneficiar a un sector en específico en comparación del resto) y como esta puede favorecer el desarrollo del sector en cuestión.

6. Conclusiones

En esta investigación, se ha propuesto una forma de integrar las diversas dimensiones de la pobreza en un modelo de regionalización multidimensional. Dicho modelo consiste en la obtención de proporciones por entidad federativa de cada uno de los indicadores que usa el CONEVAL estipulados en su metodología de medición de la pobreza multidimensional, para posteriormente resumir y reducir las dimensiones de dichos datos a través de un análisis factorial, con la cual se obtienen puntuaciones factoriales para un posterior

análisis de cluster, el cual permite clasificar entidades federativas según la similitud entre su grado de pobreza multidimensional, obteniendo así una regionalización no basada en la contigüidad geográfica, sino en la similitud e intensidad de la pobreza multidimensional que cada una padece.

Para dichos procesos, la estadística multivariante y los algoritmos de aprendizaje automático o machine learning son fundamentales, pues estos hacen posible el correcto tratamiento de la multidimensionalidad y son capaces de capturar y procesar información que sin algoritmos computacionales sería muy difícil de desentrañar.

La metodología de regionalización propuesta, tiene potencial para mejorar la formulación, evaluación y ejecución de la política pública y de desarrollo social, ya que permite identificar correctamente las carencias de cada entidad federativa y agruparlas en similitudes para ofrecer una mejor focalización de la pobreza. Además, esta metodología es perfectamente micro escalable, y puede aplicarse a estados, alcaldías, localidades e incluso municipios siempre y cuando se cuente con la información y los datos necesarios. Además, no es únicamente en temas de pobreza esta aplicación, pues la metodología en realidad puede adaptarse a medir otros fenómenos de carácter multidimensional como los pueden ser la economía en su conjunto y sus interacciones y la seguridad pública

En resumen, la presente investigación ha destacado la importancia de la metodología de la medición multidimensional de la pobreza, al mismo tiempo que ha propuesto la regionalización de la pobreza multidimensional en México, como una metodología que combina estadística multivariante y métodos computacionales, y que puede ser complementaria a la actual del CONEVAL. Se identificaron clústeres de entidades federativas con características similares en términos de pobreza multidimensional, lo que es una herramienta útil para diseñar y ejecutar políticas públicas y sociales más efectivas y focalizadas.

Esta metodología desde luego necesita de datos actualizados y de alta calidad, pero puede mejorar la eficiencia y efectividad de los programas dedicados al combate a la pobreza en el país. Esta aportación metodológica puede significar una comprensión más integral, y una guía para el micro-escalamiento de la técnica, lo que puede resultar en una intervención gubernamental y de políticas públicas que junto con la metodología actual, pueden hacer estas más eficientes, para combatir esta problemática compleja y multidimensional.

Los datos, transformaciones y cuadros resumen se encuentran disponibles en el repositorio de git de la investigación, en:

https://github.com/GARMA350/Regionalizacion_de_la_pobreza_en_Mexico_2016_2022

Referencias

- [1] OCDE. Desigualdad en los ingresos, income inequality. <https://www.oecd.org/espanol/estadisticas/diferencia-ingresos.htm>, 2020.
- [2] A. E. G. Gómez. Enigh 2022, ¿cómo vamos en desigualdad? <https://mexicocomovamos.mx/animal-politico/2023/08/enigh-2022-como-vamos-en-desigualdad/>, 2023.
- [3] CONEVAL. Medición de pobreza 2022: Resumen ejecutivo. https://www.coneval.org.mx/Medicion/MP/Documents/MMP_2022/Pobreza_multidimensional_2022.pdf, 2023.
- [4] CONEVAL. ¿quiénes somos?, creación del coneval. <https://www.coneval.org.mx/quienessomos/Paginas/Creacion-del-Coneval.aspx>.
- [5] Diario Oficial de la Federación. Decreto por el que se regula el consejo nacional de evaluación de la política de desarrollo social. <https://www.coneval.org.mx/rw/resource/coneval/normateca/2343.pdf>, 8 2005.
- [6] Ley general de desarrollo social. <https://www.coneval.org.mx/quienessomos/Documents/Ley-General-de-Desarrollo-Social.pdf>, 2004.
- [7] CONEVAL. Metodología para la medición multidimensional de la pobreza en México. <https://www.coneval.org.mx/InformesPublicaciones/InformesPublicaciones/Documents/Metodologia-medicion-multidimensional-3er-edicion.pdf>, 2019.
- [8] CONEVALVIDEO. Presentación de estimaciones de la medición de pobreza 2022. <https://www.youtube.com/watch?v=2mkkdaxBTPY>, 8 2023.
- [9] L. A. Islas Ochoa and K. Y. Torres Busqueño. Regionalización de la pobreza y pobreza extrema en México 2008-2018. *Tiempo Económico, Universidad Autónoma Metropolitana*, 2019.
- [10] J. Bracamontes-Nevarez and M. Camberos-Castro. La pobreza en México y sus regiones: un análisis de impacto del programa Oportunidades en el periodo 2002-2006. https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1405-74252011000100006, 2011.
- [11] M. A. Pérez Méndez. Una propuesta para el análisis regional de la pobreza en México. https://rde.inegi.org.mx/RDE36/RDE36_art4, 2022.

- [12] CONEVAL. Programas de cálculo y bases de datos 2016, 2018, 2020 y 2022. https://www.coneval.org.mx/Medicion/MP/Paginas/Programas_BD_2022.aspx.
- [13] Luis Armando Garcia Rodríguez. Repositorio de la investigación regionalización multidimensional de la pobreza". https://github.com/GARMA350/Regionalizacion_de_la_pobreza_en_Mexico_2016_2022/tree/main/Bases_Pobreza.
- [14] Luis Armando Garcia Rodríguez. Repositorio de la investigación regionalización multidimensional de la pobreza". https://github.com/GARMA350/Regionalizacion_de_la_pobreza_en_Mexico_2016_2022/tree/main/Codigo.
- [15] Luis Armando Garcia Rodríguez. Repositorio de la investigación regionalización multidimensional de la pobreza". https://github.com/GARMA350/Regionalizacion_de_la_pobreza_en_Mexico_2016_2022/blob/main/Cuadros_RegPoMex.xlsx.
- [16] J. F. Hair. *Análisis multivariante*. PRENTICE HALL, 1999.
- [h] [h]