

Comparación de conglomerados creados según indicadores de progreso social del año 2019 para la caracterización de los cantones de Costa Rica

Jose Pablo Mora¹, Andrés Rojas¹, Estefanía Villalobos¹

josepablo_99@outlook.com, andresrc.ucr@gmail.com, evillalobos97@hotmail.com

RESUMEN

La pobreza es una condición estudiada por diferentes enfoques; esta puede ser medida mediante indicadores sociales y permite realizar comparaciones dado al nivel de desagregación de la información. En Costa Rica, se realiza un seguimiento de la pobreza por zonas, regiones de planificación y otras áreas de interés social para identificar posibles brechas e impulsar acciones correctivas. Debido a esto, se planteó el objetivo de agrupar los cantones de Costa Rica según variables relacionadas al Índice de Progreso Social del 2019, con el fin de conocer la condición social y ambiental de los cantones por agrupamiento. A partir de 15 indicadores, se realiza un análisis de componentes principales para definir las cinco variables de agrupamiento utilizadas en el estudio de conglomerados. Seguidamente, utilizando métodos jerárquicos aglomerativos y el método de k-medias, se conforma un grupo con 28 cantones y otro con 53. Se observa que el primer grupo, conformado mayoritariamente por cantones rurales y de otras regiones diferentes de la Central, presentan puntajes menores en los componentes principales asociados a variables de bienestar social.

PALABRAS CLAVES: Componentes principales, agrupamiento, distancias

INTRODUCCIÓN

Con el paso del tiempo, la conceptualización de la pobreza ha tenido diferentes percepciones, sobre todo al determinar las metodologías y clasificaciones de quienes se consideran en condición de pobreza o no. Actualmente, dadas estas modificaciones, se han agregado diferentes criterios multidimensionales para abordar la medición de la pobreza (Sánchez y Barrera, 2020). Según Céspedes y Jiménez (2006), no existe una definición única de pobreza, sin embargo, se puede dividir en distintos criterios; dos de estos criterios son la pobreza según la subsistencia y según las necesidades básicas. El primer criterio hace referencia a la falta de bienes y servicios materiales que cubren las necesidades biológicas que son indispensables para vivir, como lo son la alimentación, el vestido y también el alojamiento. Por otro lado, el segundo criterio se relaciona con la posibilidad de tener acceso a los servicios que ofrece el entorno como la educación, salud, transporte e infraestructura.

Ahora bien, la falta de acceso a estos bienes y servicios puede ser mayor en ciertas áreas o zonas geográficas. Las zonas se pueden clasificar según el grado de urbanización que se presente, siendo urbanas las áreas que cuenten con cuerdas definidas, calles, aceras, alumbrado público, recolección de basura y actividades económicas; y aquellas áreas que se clasifiquen como rurales cuentan con viviendas muy alejadas una de otra y también se desarrollan

¹ Estudiantes de Estadística de la Universidad de Costa Rica

actividades agropecuarias, pecuarias, silvícola y turísticas. O también, se pueden clasificar por regiones de planificación, que se refiere a la división del territorio nacional, en Costa Rica se divide en seis regiones: Región Central, Región Chorotega, Región Pacífico Central, Región Brunca, Región Huetar Caribe y Región Huetar Norte (INEC, 2016).

Principalmente las políticas públicas se han enfocado en el avance de zonas más industrializadas y con mayor concentración de personas dejando de lado el desarrollo de las zonas rurales, presentando mayor desigualdad y dificultando el acceso a ciertos recursos. De esta manera, históricamente la pobreza rural siempre ha sido mayor que la pobreza urbana (Sáenz, Barboza, y Rodríguez, 2020). En Costa Rica, la pobreza enfocada en la subsistencia y necesidades básicas ha tenido un impacto en el 4.4% de la población durante el 2019 (INEC, 2019). Porras (2018) menciona que, para la clasificación de área urbano-rural, las personas que residen en zonas rurales son las que se ven más afectadas por la pobreza; y según las regiones de planificación, las regiones Huetar Caribe y Huetar Norte son las que presentan mayor cantidad de personas en situación de pobreza.

De esta manera, la pobreza al ser una realidad a la que se ha enfrentado la humanidad desde sus inicios, llevar a cabo estudios que midan la pobreza son de suma importancia para lograr disminuir la desigualdad y las brechas sociales presentes en la población, ya que a partir de los resultados se puede determinar la magnitud de pobreza y cuál ha sido la tendencia a través del tiempo. De este modo, se puede identificar las zonas más perjudicadas, con el fin de orientar y mejorar programas destinados a la lucha contra la pobreza y así asegurar una mejor calidad de vida a las personas (Boltvinik y Damián, 2020).

Este estudio tiene como objetivo agrupar los cantones de Costa Rica según variables relacionadas al Índice de Progreso Social (IPS) del 2019, con el fin de conocer la condición social y ambiental de los cantones por agrupamiento.

METODOLOGÍA

El análisis se llevó a cabo consultando los datos en línea desarrollados por el Centro Latinoamericano para la Competitividad y Desarrollo Sostenible (CLACDS) de INCAE Business School y la organización Social Progress Imperative (2019), que tiene como objetivo medir el desempeño social y ambiental de los 81 cantones de Costa Rica para el año 2019. También, se consultó el Manual de Clasificación Geográfica con Fines Estadísticos de Costa Rica del Instituto Nacional de Estadística y Censos (2016) para crear variables de clasificación según grado de urbanización y por región de planificación.

Se tomaron en cuenta 15 variables continuas para el estudio, correspondiente a los indicadores de Nutrición y Cuidados Médicos Básicos, Agua y Saneamiento, Vivienda, Seguridad Personal, Acceso a Conocimientos Básicos, Acceso a Información y Comunicaciones, Salud y Bienestar, Calidad Medioambiental, Derechos Personales, Libertad Personal y de Elección, Inclusión, Acceso a Educación Superior, Necesidades Humanas Básicas, Fundamentos del Bienestar y Oportunidades, que miden características sociales y ambientales para los diferentes cantones. Al tratarse de indicadores, la escala es la misma para las 15 variables, sin embargo, se estandarizaron al encontrar varianzas muy distintas entre estas.

Seguidamente, se realizó un análisis de valores extremos donde se encontró el cantón de Nandayure, no obstante, al tratarse de un estudio de agrupamiento de los cantones de Costa Rica, se decidió conservarlo. También, al encontrar altas correlaciones entre las variables, se desarrolló un análisis de componentes principales. Se determinó cantidad de componentes principales al evaluar 1) que la variabilidad explicada acumulada por los componentes principales llegue al menos a un 70% y 2) incluir aquellos componentes que cumplan con la condición $\lambda_j \geq 1$, por lo que, se eligieron los primeros cinco cuya varianza explicada alcanza un 77.091% del total.

Utilizando los componentes principales mencionados como variables de agrupamiento, se utilizaron las distancias Euclídea y Manhattan entre individuos y las distancias de vecino más cercano, vecino más lejano, salto promedio y la distancia de Ward entre grupos para comparar las 12 combinaciones de distancias que ajustan de mejor manera los grupos, por medio de métodos jerárquicos aglomerativos. Los grupos se representaron de manera visual con dendrogramas que posteriormente se verificaron y compararon con ayuda del coeficiente de correlación entre las distancias *cophenetic* y la matriz de distancias original, para revisar la verdadera similitud con las observaciones.

Luego de elegir la partición del dendrograma, se utilizó el método de k-medias para considerar las diferentes agrupaciones de los datos a partir de las medias o centroides de los grupos seleccionados. En otras palabras, para el desarrollo de agrupamiento de los cantones, se utilizó la combinación entre métodos jerárquicos aglomerativos y k-medias. Asimismo, se evaluaron los criterios de la suma de cuadrados entre grupos (SCDG) y el método de la silueta promedio para elegir la cantidad de grupos (k) y posteriormente, se empleó el ancho de silueta promedio para validar internamente la composición de los agrupamientos.

Finalmente, se caracterizaron los grupos resultantes con los componentes principales elaborados anteriormente. También, se utilizaron las variables (no incluidas en el análisis de conglomerados) de: **grado de urbanización**, que corresponde al promedio aplicando redondeo al más próximo de la escala de urbanización, para las categorías urbano, predominantemente urbano, predominantemente rural y rural de los distritos de cada cantón, y **región de planificación**, la cual divide los cantones en las regiones Central, Chorotega, Pacífico Central, Brunca, Huetar Caribe y Huetar Norte, para la caracterización de los grupos.

El análisis se realizó con el software R (R Core Team, 2022), junto con las librerías biotools (da Silva, 2021), corrgram (Wright, 2021), dendextend (Galili, 2015), dplyr (Wickham François, Henry y Müller, 2022), factoextra (Kassambara y Mundt 2020), fpc (Hennig, 2020), gg dendro (de Vries y Ripley, 2022), ggplot2 (Wickham, 2016), NbClust (Charrad, Ghazzali, Boiteau, y Niknafs, 2014), plotly (Sievert, 2020) y psych (Revelle, 2022).

RESULTADOS

Para el análisis fue posible recopilar información de los 81 cantones de Costa Rica. Entre las variables se encuentran diferentes indicadores de progreso social, los cuales se pueden clasificar en 4 grupos principales: en seguridad se encuentra un 26.67% de las variables, en salud

un 26.67%, en educación un 26.67% y otros indicadores complementarios sobre bienestar un 20.00% (Tabla 3, sección Anexos).

Previo al análisis de conglomerados, se estudió la presencia de valores extremos en el conjunto de datos y se determinó que existe un único posible valor extremo, correspondiente al cantón de Nandayure (Ver Figura 4, sección Anexos). A pesar de esto, se tomó la decisión de mantener la observación ya que este cantón no se encuentra tan alejado del criterio de decisión de los valores extremos y su aporte es importante para abarcar la totalidad de cantones del país.

Luego, al calcular la matriz de correlaciones, se observaron correlaciones muy altas entre las variables tomadas en cuenta para el estudio de conglomerados, por lo que se estimó el determinante de la matriz de correlaciones y se obtuvo un valor cercano a cero (Ver Figura 5, sección Anexos). Tomando esto en cuenta, se optó por realizar un análisis de componentes principales con el fin de reducir la dimensionalidad del conjunto de datos y arreglar el problema de correlaciones entre las variables.

Antes de desarrollar el análisis de componentes principales, se tomó la decisión de utilizar la matriz de correlaciones para la construcción de estos, ya que se observaron varianzas muy distintas entre las variables. Para que ambos criterios de selección de los componentes principales se cumplan, se definió una cantidad total de 5 componentes principales (Ver la Tabla 4, sección Anexos) explicando un 77.09% de la variabilidad total y de esta manera, consolidar las variables de conglomeración para el análisis.

Según la matriz de vectores propios, el primer componente principal, que explica un 31.54% de la varianza total, se asocia con las variables de Vivienda, Acceso a Información y Comunicación, Libertad personal y de Elección. El segundo componente principal se relaciona con las variables de Calidad Medioambiental y Fundamentos del Bienestar, explicando un 17.15% de la variabilidad total. El tercer componente principal se vincula a las variables de Seguridad Personal, Calidad Medioambiental, Inclusión, Necesidades Humanas Básicas y Fundamentos del Bienestar y explica 11.94% de la variabilidad. Las variables de Nutrición y Cuidados Médicos Básicos, Acceso a Conocimientos Básicos, Derechos Personales e Inclusión se asocian al cuarto componente principal, que explica el 9.72% de la variabilidad total. El quinto componente principal se une a las variables de Nutrición y Cuidados Médicos Básicos y Acceso a Conocimientos Básicos explicando un 6.74% de la variabilidad total (Ver Tabla 4 y Tabla 5, sección Anexos).

Ahora, respecto al análisis de conglomerados, para contemplar el uso de los métodos de distancias entre individuos más adecuados, se tomó en cuenta que las variables de agrupamientos son continuas y linealmente independientes, es decir, no están correlacionadas. Por lo que se utilizaron las distancias entre individuos Euclídea y Manhattan.

También, para cada una de las distancias entre individuos, se evaluaron todas las posibles combinaciones de métodos de agrupación jerárquica entre grupos, que hace referencia a: vecino más cercano, vecino más lejano, salto promedio y distancia de Ward. En la Tabla 1, se muestran las diferentes combinaciones de las distancias junto con sus correlaciones cophenetic

y una categorización de la conformación de grupos con las visualizaciones obtenidas de los dendrogramas (buena o mala visualización).

Tabla 1

Correlación de cophenetic – para evaluar la estructura de los dendrogramas

Distancias entre individuo	Distancias entre grupo	Correlación cophenetic	¿Se visualiza una conformación de grupos?
Euclídea	Vecino más cercano	0,532	😞
	Vecino más lejano	0,500	😎
	Salto promedio	0,621	😎
	Ward	0,531	😎
Manhattan	Vecino más cercano	0,504	😞
	Vecino más lejano	0,459	😎
	Salto promedio	0,649	😞
	Ward	0,525	😎

Leyenda

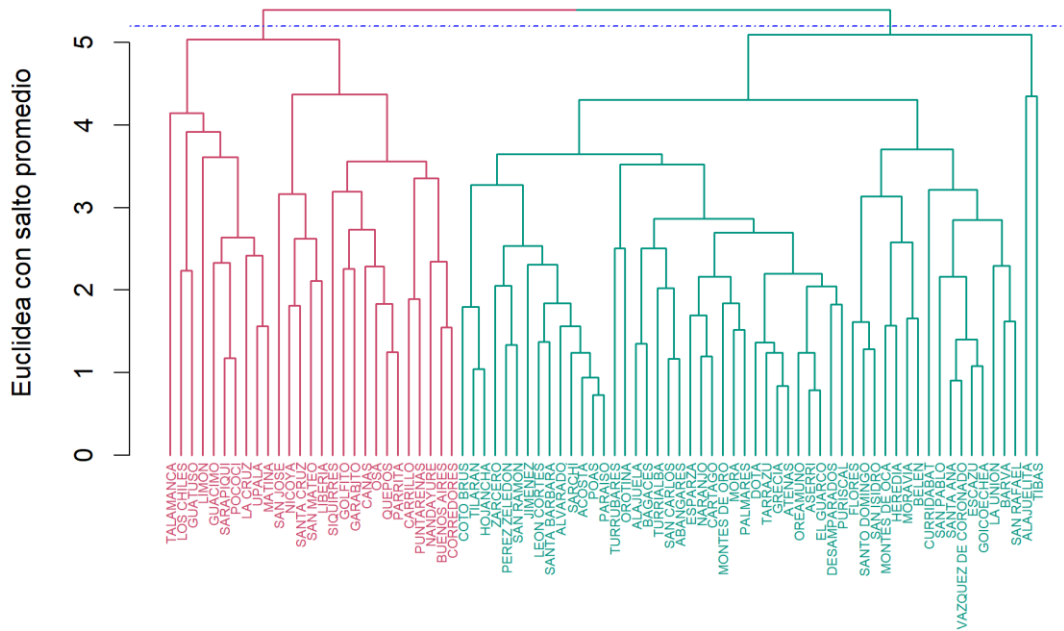
😞 : Mala visualización

😎 : Buena visualización

A partir de estos resultados, se observó que entre las combinaciones que están mejor definidas, según las correlaciones de cophenetic, se encuentran: euclídea con salto promedio, euclídea con Ward, Manhattan con salto promedio y Manhattan con Ward. A pesar de que la conformación de Manhattan con salto promedio presentó el mayor valor en la correlación cophenetic (0.649), la visualización de los grupos a partir del dendrograma no es muy clara; por esta razón, se eligió el análisis con el agrupamiento jerárquico aglomerativo utilizando la distancia Euclídea entre individuos y salto promedio entre grupos (correlación cophenetic 0.621). La Figura 1, presenta la división de los 2 grupos conformados con el método jerárquico. La distribución quedó definida en 28 cantones en el grupo 1 y 53 cantones en el grupo 2.

Figura 1

Dendrograma de agrupamiento jerárquico según distancia euclídea y salto promedio entre grupo



Se aplicó el método de k-medias definiendo la cantidad de grupos (k=2) y la conformación de los grupos a partir del método jerárquico aglomerativo. Como resultado, los cantones de San José, San Mateo y Liberia pasaron del grupo 1 al grupo 2 y los cantones de San Carlos, Turrialba, Abangares y Coto Brus se movieron del grupo 2 al grupo 1, por lo que, el grupo 1 se conformó por 28 cantones y el grupo 2 por 53 cantones. Mediante los criterios de la suma de cuadrados entre grupos (SCDG) y el método de la silueta promedio, se determinó que el número de grupos adecuado es k = 2 (Ver Figura 6, sección Anexos). Por lo tanto, los grupos se conformaron como se muestra en la Tabla 2, presentando el grupo 1 una suma de cuadrados entre grupos de 236.80 y el grupo 2 de 415.57.

Tabla 2

Agrupamiento de cantones según grupo

Grupo 1		Grupo 2			
San Carlos	Buenos Aires	San José	Montes de Oca	Orotina	San Rafael
Upala	Osa	Escazú	Turrubares	Zarcero	San Isidro
Los Chiles	Quepos	Desamparados	Dota	Sarchí	Belén
Guatuso	Golfito	Puriscal	Curridabat	Cartago	Flores
Turrialba	Coto Brus	Tarrazú	Pérez Zeledón	Paraíso	San Pablo
Sarapiquí	Parrita	Aserri	León Cortés	La Unión	Liberia
Nicoya	Corredores	Mora	Alajuela	Jiménez	Bagaces
Santa Cruz	Garabito	Goicoechea	San Ramón	Alvarado	Tilarán
Carrillo	Limón	Santa Ana	Grecia	Oreamuno	Hojancha
Cañas	Pococí	Alajuelita	San Mateo	El Guarco	Esparza

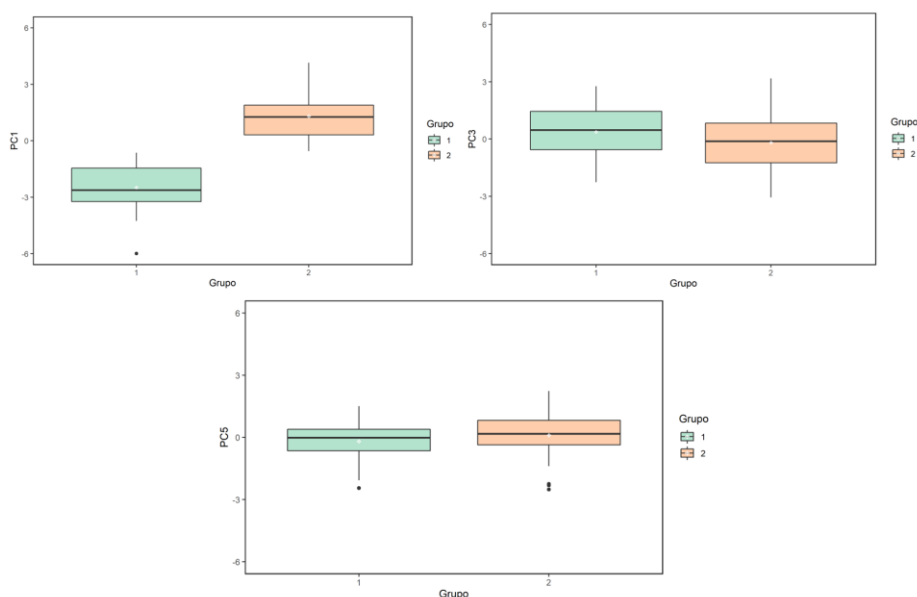
Abangares	Siquirres	Vázquez de Coronado	Atenas	Heredia	Montes de Oro
Nandayure	Talamanca	Acosta	Naranjo	Barva	
La Cruz	Matina	Tibás	Palmares	Santo Domingo	
Puntarenas	Guácimo	Moravia	Poás	Santa Bárbara	

Para validar el agrupamiento, se llevó a cabo el ancho de silueta, dando como resultado un promedio de 0.26, lo cual indica que algunas observaciones se ubican en un punto intermedio entre los conglomerados (Ver Figura 7, sección Anexos). Por otro lado, no se poseen observaciones con valores negativos, es decir, no apuntan a una asignación incorrecta de los cantones a los grupos.

Luego, se realizó una representación con gráficos de cajas para determinar cuáles componentes principales tuvieron un mayor peso para la consolidación de los grupos. En la Figura 2, se observó que, para el primer componente principal, cuyas variables asociadas eran Vivienda, Acceso a Información y Comunicación, Libertad personal y de Elección, las medias de los puntajes de los grupos son muy diferentes, siendo -2.48 para el primer grupo y 1.31 para el segundo. En el tercer componente principal, ligado a las variables Seguridad Personal, Calidad Medioambiental, Inclusión, Necesidades Humanas Básicas y Fundamentos del Bienestar, los puntajes promedio en el grupo 1 (0.37) son mayores que en el grupo 2 (-0.19). También, para el quinto componente principal, asociado a Nutrición y Cuidados Médicos Básicos y Acceso a Conocimientos Básicos, se aprecia una leve diferencia entre las medias de los puntajes del grupo 1 (-0.19) y grupo 2 (0.10) como también, una baja variabilidad entre los cantones que pertenecen a estos grupos. Cabe destacar, que para los demás componentes principales no se contemplaron grandes diferencias entre los grupos (Ver Figura 8, sección de Anexos).

Figura 2

Comparación de los grupos formados según PC1, PC3 y PC5

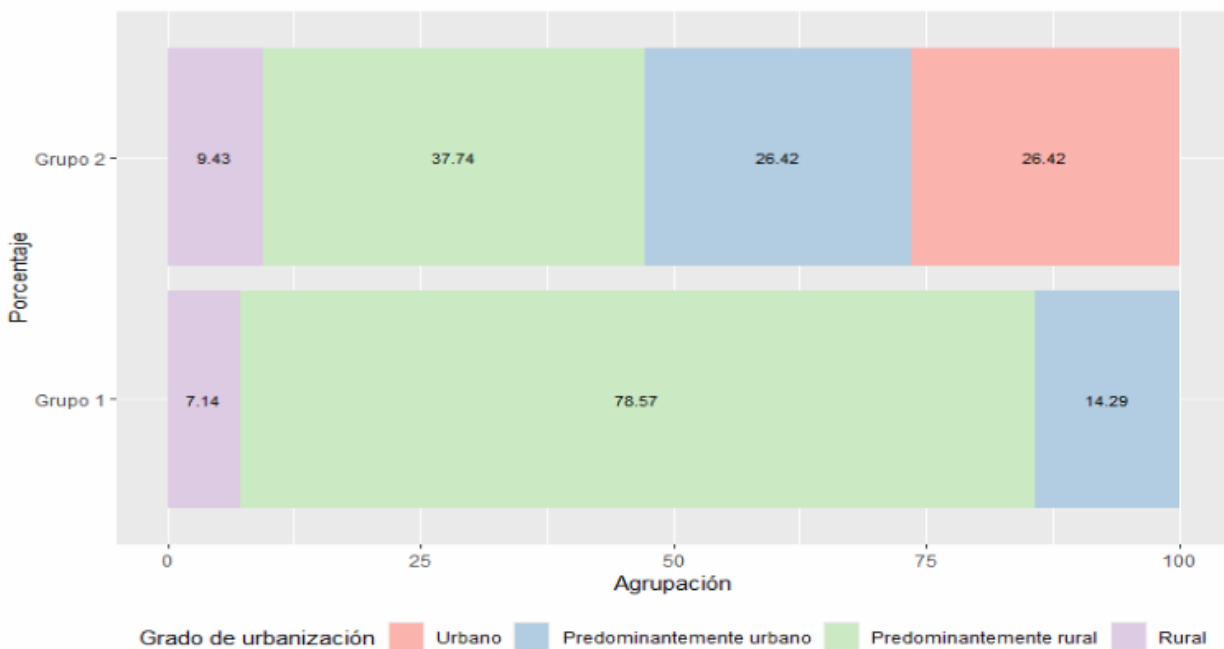


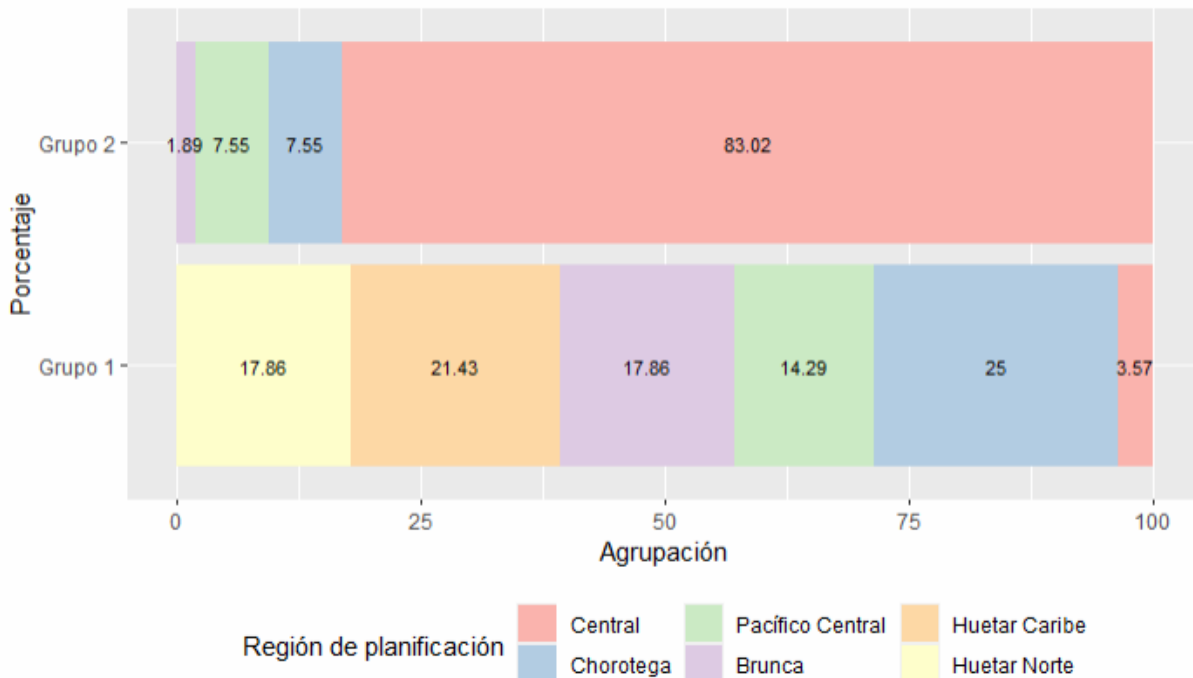
Por último, se caracterizaron los grupos según grado de urbanización, lo cual mostró que el grupo 1 está conformado mayoritariamente por cantones predominantemente rurales. Se encontró un 7.14% cantones rurales y 78.57% predominantemente rurales. Para el grupo 2, se observó que todos los cantones urbanos se agruparon en este, sin embargo, también hubo un 9.93% de cantones rurales y 37.74% predominantemente rurales. En términos generales, el grupo 1 está constituido por cantones rurales y el grupo 2 está formado por todos los cantones urbanos como también cantones rurales.

Para la región de planificación, se encontró que el grupo 1 está constituida por todos los cantones de la región Huetar Caribe y Huetar Norte, también por 5 de los 6 cantones de la región Brunca, 4 cantones de la región Pacifico Central, 7 de la región Chorotega y solamente 1 cantón de la región Central, correspondiente a Turrialba. Por otro lado, el grupo 2 está conformado por todos los cantones de la región Central, menos Turrialba, correspondiente a 44 cantones. También hay 4 cantones de la región Chorotega y 4 de la región Pacifico Central y solo un cantón de la región Brunca, correspondiente a Pérez Zeledón.

Figura 3

Caracterización de grupos según grado urbanización y región de planificación





CONCLUSIONES

A partir de los 81 cantones de Costa Rica, con ayuda de las variables del Índice de Progreso para el año 2019, se logró realizar el agrupamiento de las observaciones en dos conglomerados, a partir de la combinación entre métodos jerárquicos aglomerativos y k-medias. Además, se caracterizaron según el nivel de urbanización y región de planificación.

El primer grupo, conformado mayoritariamente por cantones predominantemente rurales y más alejados a la región Central, posee promedios en los puntajes del primer componente principal más bajos que el promedio de los puntajes del segundo grupo, cuya composición predominante son de cantones urbanizados y de la región Central, lo que quiere decir que, en términos de las variables de Vivienda, Acceso a Información y Comunicación, Libertad personal y de Elección, el grupo 1 se encuentra en desventaja respecto al grupo 2. Cabe destacar que, estas agrupaciones concuerdan con el artículo de Porras (2018) en el que establece que los hogares ubicados en las zonas rurales y alejadas del valle central del país son las más afectadas por la pobreza, como también los hogares en las regiones de Huetar Caribe y Huetar Norte. Sin embargo, en el grupo 2, se encuentran los cantones de Pérez Zeledón, San Mateo, Orotina, Liberia, Bagaces, Tilarán, Hojancha, Esparza y Montes de Oro, mostrando que no todos los cantones fuera de la región Central poseen puntajes tan bajos en estos indicadores.

Por otro lado, los puntajes promedio del grupo 1 son superiores que el grupo 2 en el tercer componente principal, el cual está asociado positivamente a variables de Calidad Medioambiental, Inclusión y Fundamentos del Bienestar y negativamente en Seguridad Personal y Necesidades Humanas Básicas. En otras palabras, el grupo conformado en su totalidad por cantones rurales predominantemente rurales y predominantemente urbanos, presentan puntajes más altos en el componente que está ligado a indicadores medioambientales y

bienestar, sin embargo, también quiere decir que tienen una desventaja en seguridad personal y necesidades humanas respecto al grupo 2.

Finalmente, el comportamiento de los grupos en el quinto componente principal, asociado positivamente a Nutrición y Cuidados Médicos Básicos y negativamente a Acceso a Conocimientos Básicos, es muy similar con una ventaja leve para el grupo 2 sobre el grupo 1. Quiere decir que, en términos de nutrición, salud y conocimientos básicos, no hay grandes diferencias entre los cantones en los diferentes grupos.

BIBLIOGRAFÍA

- Boltvinik, J., & Damián, A. (2020). Medición de la pobreza de México: análisis crítico comparativo de los diferentes métodos aplicados. Recomendaciones de buenas prácticas para la medición de la pobreza en México y América Latina. Recuperado de https://repositorio.cepal.org/bitstream/handle/11362/45555/S2000335_es.pdf?sequence=1&isAllowed=y
- Céspedes, V., & Jiménez, R. (2006). Pobreza en Costa Rica. Academia de Centroamérica. Recuperado de <https://www.academiaca.or.cr/wp-content/uploads/2017/06/Librojornada3.pdf>
- Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software*, 61(6), 1-36. URL <http://www.jstatsoft.org/v61/i06/>
- da Silva, A. R. (2021) biotools: Tools for Biometry and Applied Statistics in Agricultural Science. R package version 4.2. <https://cran.r-project.org/package=biotools>
- de Vries, A. & Ripley, B. (2022). gg dendro: Create Dendrograms and Tree Diagrams Using 'ggplot2'. R package version 0.1.23. <https://CRAN.R-project.org/package=ggdendro>
- Galili, T. (2015). dendextend: an R package for visualizing, adjusting, and comparing trees of hierarchical clustering. *Bioinformatics*. DOI: 10.1093/bioinformatics/btv428
- Hennig, C. (2020). fpc: Flexible Procedures for Clustering. R package version 2.2-9. <https://CRAN.R-project.org/package=fpc>
- INCAE Business School. (2019). *INCAE Business School*. Recuperado de Índice de Progreso Social Cantonal 2019: <https://www.incae.edu/es/clacds/proyectos/indice-de-progreso-social-cantonal-2019.html>
- INEC (2016). Manual de Clasificación Geográfica con Fines Estadísticos de Costa Rica. San José Costa Rica. Recuperado de <https://www.inec.cr/sites/default/files/documentos-biblioteca-virtual/meinstitucionalmcgfecr.pdf>
- INEC (2019). Encuesta Nacional de Hogares (ENAH0) 2019. Resultados generales. San José, Costa Rica. Recuperado de <https://www.inec.cr/sites/default/files/documentos-biblioteca-virtual/reenaho2019.pdf>

- Kassambara, A., & Mundt, F. (2020). factextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.7.
- Porras, Á. J. (2018). Construcción de un índice de pobreza a partir del equipamiento de los hogares en Costa Rica. *Revista Posgrado y Sociedad*, 16(1), 17-36. Recuperado de <https://revistas.uned.ac.cr/index.php/posgrado/article/view/2096/2432>
- 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/>.
- Revelle, W. (2022) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, <https://CRAN.R-project.org/package=psych> Version = 2.2.3,.
- Sáenz, F., Barboza, L., & Rodríguez, A. (2020). Las políticas de desarrollo rural en Costa Rica: avances y desafíos desde las perspectivas del territorio. Recuperado de <https://repositorio.una.ac.cr/bitstream/handle/11056/17686/3017-Texto%20del%20art%3%adculo-9617-1-10-20200629.pdf?sequence=1&isAllowed=y>
- Sánchez, A., Ruiz, J., & Barrera, M. (2020). La transformación del concepto de pobreza: un desafío para las ciencias sociales. *Intersticios sociales*, (19), 39-65. Epub 24 de febrero de 2020. Recuperado de http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2007-49642020000100039&lng=es&tlng=es
- Sievert, C. (2020). Interactive Web-Based Data Visualization with R, plotly, and shiny. Chapman and Hall/CRC Florida.
- Wickham. H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.
- Wright, K. (2021). corrgram: Plot a Correlogram. R package version 1.14. <https://CRAN.R-project.org/package=corrgram>
- Wickham, H., François, R., Henry, L., & Müller, K (2022). dplyr: A Grammar of Data Manipulation. R package version 1.0.8. <https://CRAN.R-project.org/package=dplyr>

ANEXOS

Tabla 3

Grupos principales de clasificación de las variables

Grupo	Variable	Porcentaje
Seguridad	Seguridad Personal	
	Libertad Personal y de Elección	26.67
	Inclusión	

	Derechos Personales	
	Nutrición y Cuidados Médicos Básicos	
Salud	Agua y Saneamiento	26.67
	Salud y Bienestar	
	Necesidades Humanas Básicas	
Educación	Acceso a Conocimientos Básicos	
	Acceso a Educación Superior	26.67
	Oportunidades	
	Acceso a Información y Comunicaciones	
Indicadores complementarios de bienestar	Calidad Medioambiental	
	Fundamentos del Bienestar	20.00
	Vivienda	

Figura 4

Identificación de valores extremos

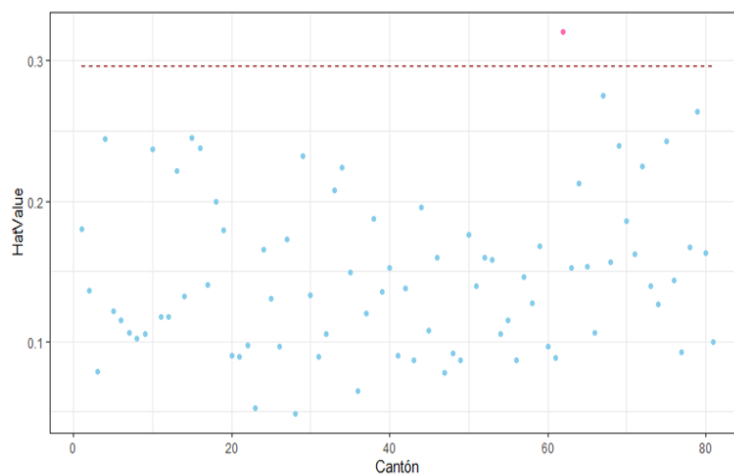


Figura 5

Correlación entre variables originales

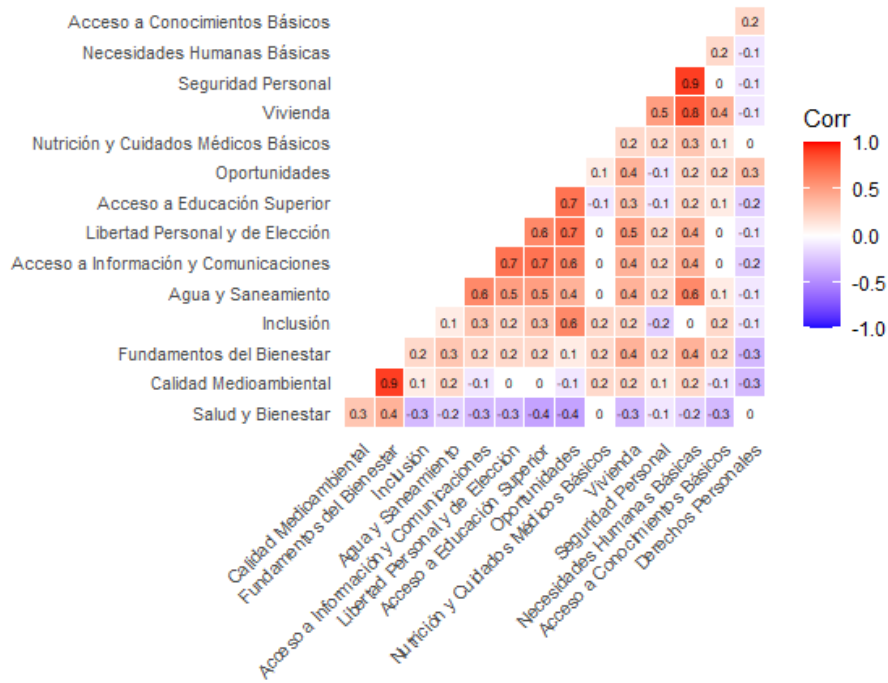


Tabla 4

Valores propios y varianza explicada de los componentes principales

Componente	λ_j	% varianza	% varianza acumulado
1	4.73	31.54	31.54
2	2.57	17.15	48.69
3	1.79	11.94	60.63
4	1.46	9.72	70.35
5	1.01	6.74	77.09
6	0.98	6.52	83.61
7	0.63	4.22	87.83
8	0.52	3.47	91.30
9	0.48	3.17	94.47
10	0.33	2.22	96.70
11	0.27	1.80	98.49

12	0.23	1.51	1
13	0.00	0.00	1
14	0.00	0.00	1
15	0.00	0.00	1

Tabla 5

Matriz de vectores propios de componentes principales elegidos

	PC1	PC2	PC3	PC4	PC5
Nutrición y Cuidados Médicos Básicos	0.0822	-0.1563	-0.0482	0.3812	0.6584
Agua y Saneamiento	0.3367	-0.0291	-0.0015	-0.2035	-0.0668
Vivienda	0.3473	-0.1602	-0.2234	0.1548	-0.1547
Seguridad Personal	0.1863	-0.3351	-0.4610	-0.0260	0.1030
Acceso a Conocimientos Básicos	0.1240	0.0533	-0.0828	0.5321	-0.5998
Acceso a Información y Comunicaciones	0.3516	0.1337	0.0433	-0.2909	0.1090
Salud y Bienestar	-0.2020	-0.3333	0.2426	-0.0452	0.1121
Calidad Medioambiental	0.0963	-0.4533	0.3970	0.0734	-0.0400
Derechos Personales	-0.0828	0.2251	-0.2021	0.4098	0.0820
Libertad Personal y de Elección	0.3514	0.1291	0.0503	-0.1717	0.1199
Inclusión	0.1824	0.1704	0.3538	0.3456	0.2145
Acceso a Educación Superior	0.3269	0.2369	0.1838	-0.1857	-0.1265
Necesidades Humanas Básicas	0.3430	-0.2778	-0.3536	0.0139	0.0559

Fundamentos del Bienestar	0.2036	-0.4049	0.3904	0.1285	-0.1793
Oportunidades	0.3253	0.3255	0.1740	0.2132	0.1527

Figura 6

Definición de la cantidad de grupos

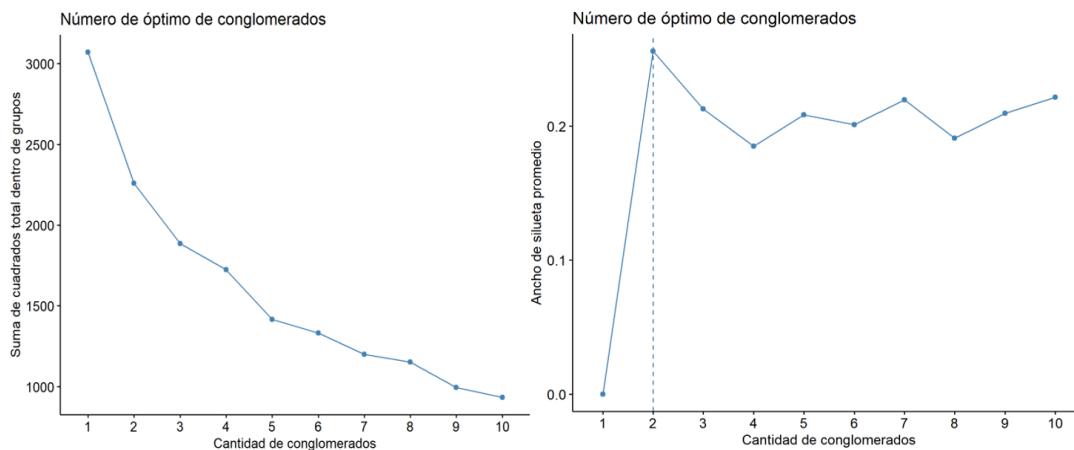


Figura 7

Ancho de silueta promedio por grupo

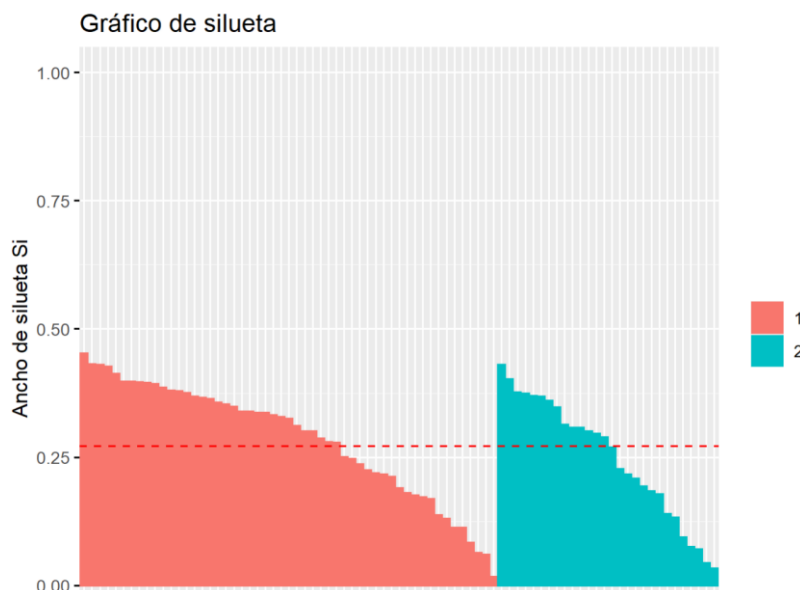


Figura 8

Comparación de los grupos formados según variables construidas

