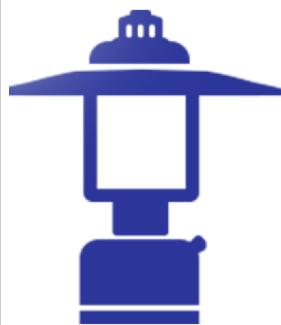


ADAPTACIÓN DE TÉCNICA K-MEANS PARA ANÁLISIS CUALITATIVO DE DATOS

ADAPTATION K-MEANS TECHNIQUE FOR QUALITATIVE DATA ANALYSIS



Fredy Alexander Guasmayan Guasmayan¹*

E-mail: fguasmayan@umariana.edu.co

ORCID: <https://orcid.org/0000-0003-1357-8867>

¹Universidad Mariana. Colombia.

*Autor para correspondencia

Cita sugerida (APA, séptima edición)

Guasmayan Guasmayan, F. A. (2025). Adaptación de técnica k-means para análisis cualitativo de datos. *Revista Conrado*, 21(106), e4404.

RESUMEN

Las técnicas estadísticas para el análisis de datos cuantitativos permiten encontrar resultados numéricos de asociación entre diferentes variables, sin embargo, en diversas investigaciones la información no siempre es cuantitativa sino que además es posible considerar elementos cualitativos que aportan al verdadero significado de los datos con una respuesta con mayor cercanía a la realidad de los fenómenos sociales como lo son las asociaciones que se generan por parte de los estudiantes en una institución educativa en la que se puede relacionar rendimiento académico con variables de tipo socioeconómico y demográfico además de la percepción que se tenga por parte de los individuos respecto a la educación y su importancia para el desarrollo personal de los estudiantes. El artículo trata de datos cuantitativos obtenidos de resultados de evaluaciones numéricas y cualitativas obtenidos por métodos y técnicas cualitativas en los que se generan categorías de análisis emergentes consideradas en diseños desde la teoría fundamentada, que incluyen métodos como la observación participante y entrevisitas personales o grupales que facilitan emitir respuestas y criterios diversos que, después de un análisis por parte del investigador y los argumentos de las respuestas, pueden ser sometidos a una codificación que permita procesar esos datos mediante la estadística.

Palabras clave:

Datos, Conglomerados, Asociación estadística, Tensión.

ABSTRACT

Statistical techniques for the analysis of quantitative data allow finding numerical results of association between different variables, however, in various investigations the information is not always quantitative but it is also possible to consider qualitative elements that contribute to the true meaning of the data with a response that is closer to the reality of social phenomena such as the associations that are generated by students in an educational institution in which academic performance can be related to socio-economic and demographic variables in addition to the perception that individuals have regarding education and its importance. for the personal development of students. The paper deals with quantitative data obtained from the results of numerical and qualitative evaluations obtained by qualitative methods and techniques in which emerging categories of analysis are generated considered in designs based on grounded theory, which include methods such as participant observation and personal or group interviews that facilitate issuing various responses and criteria that, after an analysis by the researcher and the arguments of the responses, can be subjected to a coding that allows processing these data through statistics.

Keywords:

Data, clusters, statistical association, tensions.



Esta obra está bajo una licencia internacional Creative Commons Atribución-NoComercial-SinDerivadas 4.0.

Vol 21 | No.106 | septiembre-octubre | 2025
Publicación continua
e4404



INTRODUCCIÓN

Los métodos de investigación con datos cualitativos muestran gran interés sobre todo en la asociación de variables que presentan relaciones fuertes, para esto, la estadística aborda distintas técnicas para la descripción, la correlación y la inferencia de la información suministrada por una muestra de una población, en este caso es de interés encontrar asociaciones entre variables como el rendimiento académico con aspectos sociodemográficos de estudiantes de una institución educativa pública, para lo cual las estadísticas multivariadas permiten asociar y generar categorías con diferentes grupos de datos por distintas técnicas de correlación por medida de distancia entre el conjunto de datos, sin embargo, para esta investigación se tiene interés en realizar la asociación de las variables mencionadas pero bajo la consideración de la percepción de docentes, estudiantes y padres de familia sobre la calidad de la educación y las expectativas que se tienen sobre la continuación en la formación académica en nivel universitario, para ello, se cuenta con la información de caracterización de los estudiantes que en grado 11 presentaron sus pruebas de evaluación de estado denominadas "Saber 11" aplicadas por el Instituto de Fomento de la Educación Superior ICFES. Sumado a esto, se cuenta con estudios de percepción sobre la motivación de los estudiantes hacia la educación superior.

De manera cuantitativa, la estadística presenta variedad de procesos algorítmicos para categorizar datos y buscar asociación entre ellos y a nivel cualitativo en campos de las ciencias humanas existen también variedad de discusiones empíricas, Merlino y Martínez (2007) afirman y complementan los estudios e investigaciones con mejores interpretaciones que las que se pueden generar de manera independiente en estudios cuantitativos o cualitativos, al integrar estos dos tipos de estudios se aborda tanto la comprensión de discursos y aspectos conductuales de los sujetos de estudio, como la descripción y correlación de los datos con un sentido y direccionado por el criterio del investigador.

Para este estudio, el análisis de datos cualitativos se desarrolla con técnicas de estadísticas multivariadas por conglomerados con un enfoque metodológico que busca explorar y entender patrones complejos en datos no numéricos. Así, los datos cualitativos son inherentemente descriptivos y subjetivos, su análisis mediante herramientas estadísticas avanzadas permite identificar grupos o subgrupos dentro de una población, facilitando la clasificación y la interpretación de fenómenos sociales, culturales o de comportamiento.

El análisis de conglomerados, específicamente, es una técnica de estadística multivariada que agrupa objetos en categorías o conjunto de datos asociados basados en características similares, buscando maximizar la homogeneidad dentro de cada grupo y la heterogeneidad entre ellos. Al aplicar esta técnica a datos cualitativos, como entrevistas, encuestas abiertas o descripciones textuales, se puede identificar patrones subyacentes que no serían fácilmente visibles mediante otros métodos. Este enfoque es particularmente útil en áreas como las ciencias sociales, y particularmente en la educación, donde se requieren metodologías que integren tanto la complejidad de la información cualitativa como la robustez de los análisis estadísticos.

Algunos autores mencionan estudios en este campo, así por ejemplo García Jiménez et al., (1994) y por consiguiente la formulación de propuestas educativas que respeten las diferencias, se enfrenta en la actualidad con problemas que no pueden abordarse desde el mero análisis de las diferencias individuales. La comprensión de cuestiones como el rendimiento académico del estudiante no puede hacerse sin incluir-junto a los elementos tradicionalmente considerados-al centro escolar como unidad de análisis y, por ende, a aspectos tales como el clima y la cultura escolar o el intercambio con el entorno y el medio social. Del mismo modo, la intervención diferenciada requiere hoy propuestas que consideren la acción educativa en su conjunto, incluyendo planteamientos organizativos y pedagógicos: programas formales con estrategias de enseñanza adaptadas; actividades educativas no formales e informales; y esquemas organizativos que mejoren el clima de relaciones en el aula y el centro, y faciliten la implicación de la comunidad. La explicación de estos nuevos y viejos problemas o la mejora de esas situaciones educativas, posiblemente esté reclamando un nuevo marco conceptual y metodológico desde el que entender la problemática de la diferenciación educativa. Así, algunos investigadores están proponiendo aproximaciones conceptuales que se fundamentan en planteamientos sociológicos o antropológicos y aproximaciones metodológicas que se apoyan en planteamientos de investigación que podríamos identificar como cualitativas. Bartolomé Pina (1992) afirman que la transformación de Datos Textuales en Valores Numéricos es posible en la medida que estos indiquen información cualitativa textual con posibilidad de ser transformados en valores numéricos mediante procesos de categorización y codificación, lo que permite aplicar técnicas estadísticas para su análisis.

También se puede resumir información mediante tablas de Contingencia con el fin de estudiar la relación entre

dos o más variables categóricas, permitiendo calcular medidas de asociación como el chi-cuadrado, coeficiente de contingencia, y V de Cramer, con lo que se procede a realización de Análisis Factorial y Escalamiento Multidimensional, técnicas que facilitan la representación gráfica contenida en matrices de datos, facilitando la identificación de estructuras y relaciones entre variables como es el caso particular de análisis de conglomerados, método que agrupa objetos o individuos con características similares, utilizando criterios de proximidad o lejanía, y representaciones gráficas como árboles invertidos para visualizar las jerarquías de clases formadas.

Los autores Merlino y Martínez (2007) mencionan que las ciencias sociales presentan un interés creciente en la integración metodológica que por lo general son por medio de triangulación de métodos, a lo cual presentan una investigación que incluye el aspecto empírico desde un supuesto central de que un objetivo que no puede resolverse mediante el enfoque cuantitativo, requiere de un aporte cualitativo que en este caso se realiza en función de la experiencia que aporta el investigador inmerso en el conocimiento y vivencia de los sujetos investigados en acuerdo con la teoría fundamentada (Globales, 2004) en la cual se busca la generación de categorías mediante procesos de codificación abierta, generación de subcategorías y finalmente categorías de análisis de la información cualitativa.

En lo específico del artículo, se considera el uso de la técnica de agrupamiento de datos por medio del algoritmo denominado K-means, el cual es una técnica de aprendizaje no supervisado utilizada para la agrupación de datos. Su objetivo es dividir un conjunto de datos en K grupos o clústeres basándose en la similitud entre los puntos. Funciona asignando aleatoriamente K centroides y ajustándolos iterativamente para minimizar la distancia entre los puntos de datos y su centroide más cercano. El proceso continúa hasta que los centroides dejan de cambiar significativamente (Ossama et al., 2011). En este caso consiste en tomar una base de datos numéricos transformados a partir de información recolectada tanto cuantitativa como cualitativa y asociarlos por medio de la medida de distancia euclíadiana. La propuesta planteada utiliza un coeficiente que pondera la distancia entre los datos y el centroide, permitiendo asociar dichos datos en función además de la ponderación desarrollada de manera cualitativa y a criterio del investigador que conoce la población de estudio.

MATERIALES Y MÉTODOS

Bajo el paradigma mixto se realizó una revisión bibliográfica inicial sobre el manejo de datos cuantitativos y

cualitativos con técnicas estadísticas, en lo que resulta común el análisis de la información por categorías de datos multivariados con técnicas de clúster o conglomerados, así la metodología de investigación mixta combina enfoques cuantitativos y cualitativos con la finalidad de obtener una mejor comprensión y completa de un fenómeno, que en este caso se desarrolla en el campo educativo, la metodología permite aprovechar la precisión y generalización de los datos numéricos, propios de los estudios cuantitativos, junto con la profundidad y el contexto proporcionados por los enfoques cualitativos, ofreciendo una visión holística del problema de investigación, además se considera una implementación secuencial y simultánea, que en este caso fue posible recolectar la información cualitativa de diferentes grupos de estudiantes y sus docentes durante el proceso de preparación de los mismos estudiantes para las evaluaciones de estado.

Técnicas e instrumentos

En la investigación se realiza la recolección de la información cuantitativa por medio de los resultados de evaluaciones externas suministradas por el ICFES para los estudiantes de grado 11 de una Institución Educativa pública, la información respecto a aspectos sociodemográficos fueron consultados directamente por medio de encuestas a los estudiantes y la información cualitativa sobre motivación y apreciación de la educación se realiza mediante entrevistas a una muestra de estudiantes y sus docentes en un grupo focal.

Búsqueda bibliográfica

Los resultados de investigaciones previas sobre la categorización de datos cuantitativos, que incorporan características cualitativas en la asociación de variables de estudio, han generado un gran interés en la comunidad científica. Este enfoque permite identificar factores que influyen en el rendimiento académico de los estudiantes y sus resultados en pruebas externas, considerando también características sociodemográficas. En el ámbito educativo, la integración de aspectos cualitativos es fundamental, ya que aporta conceptos valiosos desde las ciencias humanas, enriqueciendo la comprensión y el análisis de los procesos de aprendizaje. Al complementar con la fortaleza del análisis estadístico la información cuantitativa permite dar mayor robustez y validez a los datos, reafirmando que el enfoque mixto se acerca de mejor manera a la realidad de los resultados de las investigaciones. Al combinar métodos cuantitativos (como encuestas y pruebas estandarizadas) con métodos cualitativos (como entrevistas y observaciones), se obtiene una comprensión profunda del entorno escolar. Este enfoque facilita la identificación de tendencias numéricas mientras

explora experiencias y percepciones de estudiantes y docentes. En este contexto, las perspectivas epistemológicas desempeñan un papel fundamental tanto teórico como conceptual en el planteamiento de preguntas complejas desde perspectivas oncológicas tales como el realismo, el construcción social y el pragmatismo (Medina Romero et al., 2023).

Algunos ejemplos de estudios que mencionan este campo de investigación se mencionan a continuación resaltando su aporte para el presente estudio.

Los autores García Jiménez et al. (1994) y por consiguiente la formulación de propuestas educativas que respeten las diferencias, se enfrenta en la actualidad con problemas que no pueden abordarse desde el mero análisis de las diferencias individuales. La comprensión de cuestiones como el rendimiento académico del estudiante no puede hacerse sin incluir-junto a los elementos tradicionalmente considerados-al centro escolar como unidad de análisis y, por ende, a aspectos tales como el clima y la cultura escolar o el intercambio con el entorno y el medio social. Del mismo modo, la intervención diferenciada requiere hoy propuestas que consideren la acción educativa en su conjunto, incluyendo planteamientos organizativos y pedagógicos: programas formales con estrategias de enseñanza adaptadas; actividades educativas no formales e informales; y esquemas organizativos que mejoren el clima de relaciones en el aula y el centro, y faciliten la implicación de la comunidad. La explicación de estos nuevos y viejos problemas o la mejora de esas situaciones educativas, posiblemente esté reclamando un nuevo marco conceptual y metodológico desde el que entender la problemática de la diferenciación educativa. Así, algunos investigadores están proponiendo aproximaciones conceptuales que se fundamentan en planteamientos sociológicos o antropológicos y aproximaciones metodológicas que se apoyan en planteamientos de investigación que podríamos identificar como cualitativas. Bartolomé Pina (1992) mencionan en su ponencia algunos aspectos concluyentes respecto del análisis de la información cualitativa por medio de técnicas estadísticas, así algunos aportes consisten en que la información cualitativa se puede transformar a partir de los datos cualitativos textuales en valores numéricos mediante procesos de categorización y codificación, lo que permite aplicar técnicas estadísticas para su análisis, haciendo énfasis en la importancia del analista de los datos para la transformación mencionada. Una forma sencilla de transformación de los datos es por medio de un resumen categórico descrito por medio de Tablas denominadas de Contingencia, estas tablas se utilizan para estudiar la relación entre dos o más variables categóricas, permitiendo calcular medidas

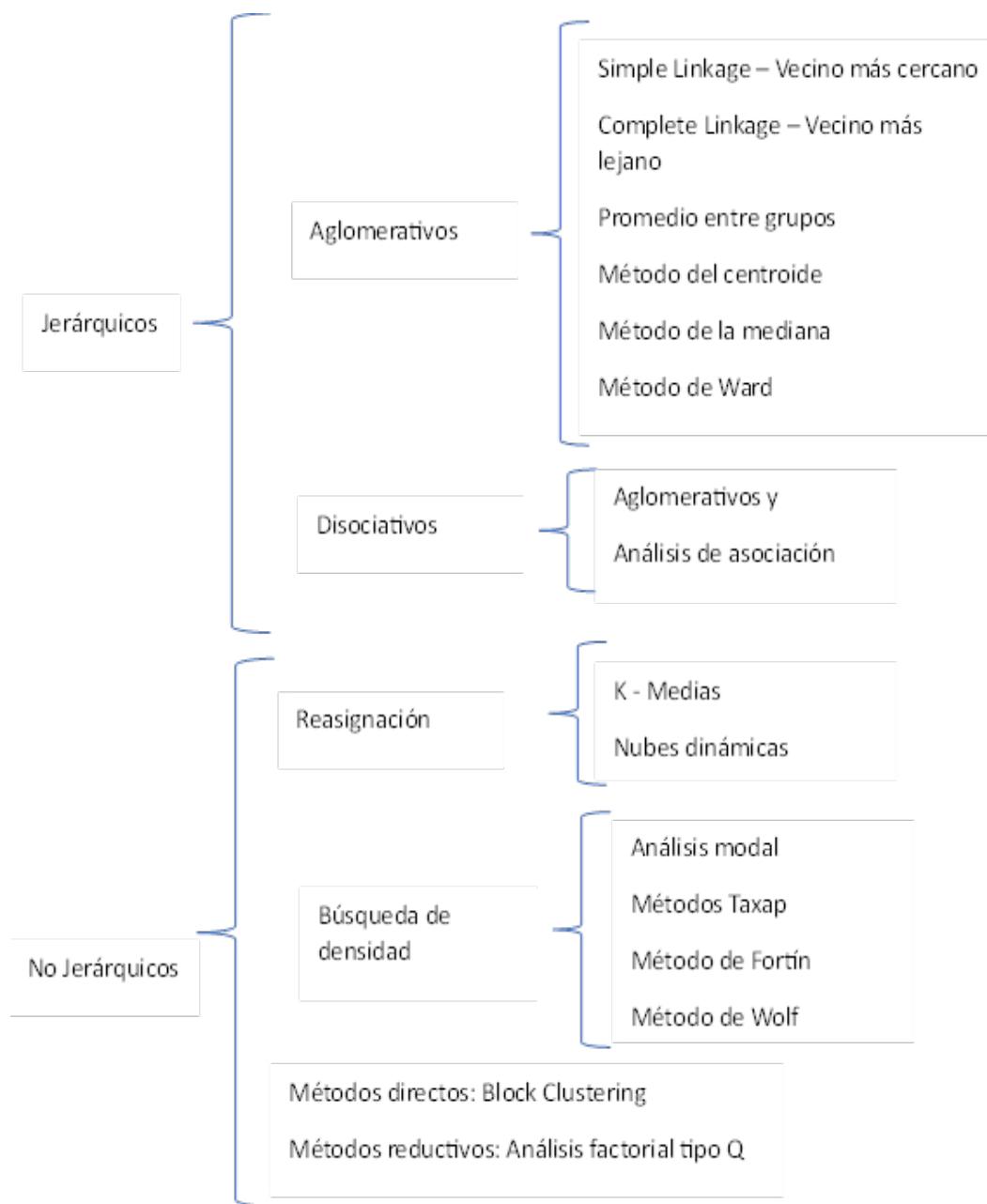
de asociación como el chi-cuadrado, coeficiente de contingencia, y V de Cramer; elementos estadísticos que dan información relevante para la correlación de información y conclusión de parámetros de una muestra y/o población, resaltando que son datos provenientes de características o categorías generadas a partir de información cualitativa.

Por otra parte, desde las estadísticas para variables categóricas los mismos autores mencionan la técnica del análisis secuencial aplicada para comprender cómo se producen los datos en secuencias temporales, para lo cual se utiliza estadísticos descriptivos y comparaciones con modelos teóricos. Aunque en este estudio no se consideran variables de tipo temporal, es importante recalcar que la información suministrada por estudiantes y docentes sí tiene posibilidad de cambios de percepción en el tiempo y más, cuando se trata de valoraciones que provienen de información cualitativa. No obstante, las técnicas estadísticas para datos categóricos por lo general son abordados desde dos métodos tradicionales basados en el análisis factorial y escalamiento multidimensional, técnica que permite representar gráficamente la información contenida en matrices de datos, facilitando la identificación de estructuras y relaciones entre variables; por otra parte, la técnica de Análisis de Conglomerados agrupa objetos o individuos con características similares, utiliza criterios de proximidad o lejanía, y representaciones gráficas como árboles invertidos para visualizar las jerarquías de clases formadas.

En un gran número de investigaciones con datos categóricos se menciona el análisis de los datos con la técnica de agrupación en categorías y sus relaciones o conglomerados, De la Fuente-Fernández (2011) mencionan al análisis de conglomerados o (cluster analysis) como una técnica estadística multivariante utilizada para agrupar elementos en función de sus características, destacando sus métodos, etapas y medidas de asociación. Así, desde lo teórico se indican algunos procesos o métodos que se basan ya sea en medir distancias de vecindad o alejamiento, que por lo general buscan la correlación entre todos los datos para generar categorías basadas por lo general en el centro de los grupos de datos y variación de la media entre los grupos, para ello se construyen técnicas algorítmicas que optimizan para encontrar valores mínimos de distancia y la generación de gráficos denominados dendogramas en los que se identifican dichas cercanías en los grupos de datos que indican alguna semejanza entre ellos.

La Figura 1 indica algunos de los principales métodos algorítmicos para el análisis y generación de asociación de datos multivariados.

Fig. 1: Métodos de análisis de conglomerados o cluster



Fuente: (De la Fuente - Fernández, 2011)

Todos los métodos de asociación por conglomerados agrupan elementos o variables para maximizar la homogeneidad dentro de cada grupo y la diferencia entre los grupos, es una técnica descriptiva y exploratoria, no inferencial, que se basa en criterios geométricos. Existen dos categorías principales de algoritmos de formación de conglomerados así, los Algoritmos de partición que dividen el conjunto de observaciones en k conglomerados, donde k es definido por el usuario y los Algoritmos jerárquicos que crean una jerarquía de divisiones del conjunto de elementos en conglomerados, que pueden ser aglomerativos (comienzan con cada observación como un conglomerado y los unen sucesivamente) o disociativos (comienzan con un gran conglomerado y lo dividen sucesivamente). Algunas etapas generales del análisis de conglomerados inician con la elección de las variables que pueden ser cualitativas (ordinales, nominales) o cuantitativas (discretas, continuas), posteriormente se hace elección de la medida de asociación como

una distancia (euclídea) o una similaridad (coeficiente de correlación), luego se realiza una elección de la técnica cluster la cual por lo general se basa en métodos jerárquicos (aglomerativos o disociativos) y no jerárquicos (como K-medias) para finalmente validar los resultados con las distancias entre conglomerados por medio de técnicas de optimización tales como el vecino más cercano, el vecino más lejano, el promedio de grupo, o el centroide.

En lo particular del análisis de datos cualitativos y cuantitativos Villanueva-Díaz et al. (2023) presentan un estudio que trata del análisis la calidad multidimensional de los datos en la apreciación que tiene una población sobre el consumo de carne de conejo desde la perspectiva del consumidor, enfocándose en factores nutricionales y sanitarios que influyen en su consumo, el análisis de datos cualitativos y cuantitativos sobre la calidad multidimensional en los datos revelan que los consumidores valoran principalmente las dimensiones nutricionales y sanitarias, considerándola nutritiva y saludable, sin embargo, la percepción de la calidad tecnológica es moderada, con un interés creciente en productos transformados, en contraste, las dimensiones de uso y tradicionalidad son vistas como limitantes, ya que la carne de conejo no está bien posicionada en la cultura gastronómica, también se menciona que la falta de acceso y la baja frecuencia de consumo en preparaciones tradicionales reflejan una barrera significativa para su aceptación generalizada. A esto se menciona que el aporte principal es el uso de distintos datos tanto cualitativos como cuantitativos es posible desde la concepción estadística junto con la aportación del investigador. Los datos cuantitativos proporcionan una base objetiva y medible sobre aspectos como el valor nutricional, la seguridad sanitaria y la calidad tecnológica de la carne, permitiendo evaluar su composición y beneficios de manera precisa. Por otro lado, los datos cualitativos aportan una comprensión profunda de las percepciones, preferencias y experiencias de los consumidores, revelando cómo estos factores influyen en sus decisiones de compra y consumo. Al combinar ambos enfoques, se obtiene una visión integral que no solo cuantifica las características de la carne de conejo, sino que también contextualiza su aceptación y uso en la cultura gastronómica, identificando barreras y oportunidades para su mayor adopción.

Por otra parte, Torres y Buitrago (2020) indican un documento en el que se analiza la desigualdad en la elección de carrera en Colombia, enfocándose en factores sociodemográficos como género, área de origen, estrato y edad. En este sentido el análisis de datos cualitativos y cuantitativos en este estudio reveló una fuerte asociación

entre variables sociodemográficas como el género, la zona de procedencia y el estrato socioeconómico con la elección de carrera en la Universidad Nacional de Colombia. Utilizando técnicas estadísticas como el análisis de Chi-cuadrado y el Análisis de Correspondencias Múltiples, se evidenció que los hombres tienden a elegir carreras en ingeniería y ciencias exactas, mientras que las mujeres se inclinan por áreas de la salud. Además, se observó una diferencia significativa en la elección de carrera entre estudiantes de zonas urbanas y rurales, así como entre aquellos de distintos estratos socioeconómicos. Estos hallazgos subrayan la influencia de factores culturales y sociodemográficos en la configuración de las decisiones académicas y profesionales, sugiriendo la necesidad de políticas educativas que promuevan una mayor equidad de género y origen en el acceso a la educación superior.

Los autores Merlino y Martínez (2007) afirman que la integración de métodos cuantitativos y cualitativos con técnicas de *clusters* y triangulación aportan tanto a la formalidad estadística como la flexibilidad en la comprensión e inferencia de los datos de una realidad con percepciones de mayor cercanía a la naturaleza humana que no solo se caracteriza por la medición de variables sino también por aspectos que percibe una población y su investigador siendo posible el análisis de un fenómeno social con la comprensión cultural, psicológica y económica implícitos en el discurso de la población y los entrevistados. Se resalta de los autores la metodología basada en la teoría fundamentada para el establecimiento de categorías y generación de conceptos que se desprenden de la relación semántica de los conceptos investigados con los datos que coinciden con aspectos metodológicos teóricos mencionados en investigación cualitativa, enfocándose en la búsqueda de significados (Taylor y Bogdan, 1994). En este sentido la investigación cualitativa cobra importancia en función de metodologías que permiten desarrollar conceptos desde los datos cualitativos y en específico desde el manejo de datos que se pueden categorizar, Balderas Gutiérrez (2016) menciona que la importancia de la investigación cualitativa radica en la neutralidad, flexibilidad, objetividad y legitimidad en la explicación causal de los fenómenos y el establecimiento de leyes y teorías.

“La realidad no es una sola, la que destaca en importancia porque es donde el hombre piensa y actúa es en la realidad de la vida cotidiana; Berger y Luckmann (1994) la llaman la realidad por excelencia o suprema realidad, es ésta la que para nuestros fines intentaremos acercarnos a estudiar.” (Balderas Gutiérrez, 2016. p 8).

En este sentido Álvarez Rodríguez et al. (2022) mencionan en su estudio de transformación digital en las instituciones educativas por medio de estrategias de análisis de información por métodos multivariados actuales enfocados en aprendizaje de máquina o machine *learning* donde el manejo de datos en las estadísticas multivariadas involucra el tratamiento de información cuantitativa y cualitativa a través de diversas técnicas para la extracción de patrones, predicciones y toma de decisiones para lo cual los datos pueden ser numéricos (edad, salario, número de hijos) o categóricos (estado civil, nivel educativo, empleo propio). La combinación de estos tipos de datos permite modelos más robustos en la predicción y análisis y su procesamiento se realiza depurando datos eliminando valores nulos y transformando variables categóricas en numéricas, posterior a esto se menciona que es importante aplicar normalización con métodos de optimización de *MinMax scaler* en modelos de predicción. En el caso de machine *learning* aplicado a instituciones educativas, se usan técnicas como regresión logística, árboles de decisión, *clustering* y redes neuronales artificiales, para ello se utilizan técnicas de correlación y análisis de conglomerados para identificar relaciones entre variables con algoritmos de clasificación que en lo particular de la educación permiten predecir factores de riesgo en estudiantes, como la deserción escolar y el rendimiento académico considerando en el análisis institucional la toma de decisiones basadas en datos históricos mediante la predicción y segmentación de datos.

Los autores Delgado Orrillo et al. (2023) complementan respecto al análisis de datos multivariados con técnicas de predicción con patrones de aprendizaje en estudiantes de posgrado en salud pública, basado en información cuantitativa y cualitativa desde la perspectiva de las estadísticas multivariadas en la que se identifican cuatro patrones principales a través de análisis de conglomerados Dirigido a los significados por medio del aprendizaje basado en la comprensión profunda y autorregulación, Dirigido a la aplicación orientado al uso del conocimiento y aplicación práctica, Dirigido a la reproducción basado en la memorización con cierta flexibilidad en las estrategias y estrategias No dirigidas con escasa regulación y orientación ambivalente.

En lo particular del método mixto de análisis de información para interpretar las relaciones entre identidad profesional, formación pedagógica y trabajo docente, combinando la riqueza interpretativa del análisis cualitativo con la precisión del análisis cuantitativo en un contexto sociopolítico en la estructuración de los patrones de aprendizaje, la articulación entre teoría y práctica, la vocación

docente y la orientación a la enseñanza participativa se identificaron como elementos clave en la modelización de los datos en la evaluación de patrones de aprendizaje que permite un análisis más preciso y contextualizado de cómo los estudiantes procesan la información y regulan su aprendizaje promoviendo enfoques más críticos y reflexivos.

Según Pardo et al. (2007), presentan un estudio que combina diferentes métodos multivariados numéricos basados en estadísticas factoriales para datos cuantitativos y categóricos. Este uso de métodos factoriales combinados con análisis de conglomerados en estadísticas multivariadas proporciona una estrategia robusta para la exploración y clasificación de datos cuantitativos y cualitativos, la estrategia combinada de análisis factorial mediante componentes principales propios del análisis de datos cuantitativos y conglomerados permite previamente reducir la dimensionalidad y extraer las características principales de los datos y posteriormente con conglomerados y algoritmos jerárquicos (método de Ward) y no jerárquicos (K-medias) se clasifica los datos en grupos homogéneos, la eficiencia y ventajas de la combinación de métodos permite mejor interpretación en el preprocesamiento facilitando la identificación de estructuras en los datos antes de la clasificación, reducción del ruido eliminando dimensiones menos relevantes, mejorando la calidad de los conglomerados. El ajuste final con K-medias garantiza que las asignaciones de los individuos sean precisas.

Según Eraso et al. (2012) toman un caso particular en el manejo de información y su análisis desde la recolección de información cualitativa en instrumentos de entrevistas, para ello usan la técnica del Análisis Cualitativo Multinivel (ACM) como un enfoque que combina técnicas cualitativas y cuantitativas para analizar entrevistas en profundidad. Este modelo permite estructurar el análisis de datos de manera más rigurosa y contextualizada, facilitando la interpretación del discurso en distintos niveles, para ello transforman los datos en variables categorizadas y así permitir un análisis estructurado con el principio de parsimonia que busca simplificar la información sin perder riqueza interpretativa con la creación de tipologías que identifican patrones en el discurso, facilitando su contextualización y con el uso de técnicas estadísticas básicas como métricas de la distancia entre sujetos y pruebas estadísticas (Chi-cuadrado, Kruskal-Wallis) para validar los grupos identificados indicando la utilidad en estudios de sociología, educación y ciencias políticas, donde la interpretación del discurso es clave dada su flexibilidad ajustable a los objetivos y tipo de datos.

En complemento Leiva et al. (2010) indican diferentes métodos algorítmicos de partición más comunes en el análisis de conglomerados, presentada en un estudio comparativo en el análisis de conglomerados el cual destaca que cada técnica tiene ventajas y limitaciones dependiendo de la estructura y naturaleza de los datos. Los algoritmos más comunes incluyen K-medias, K-medoides, y modelos basados en densidad y jerarquía. K-medias es ampliamente utilizado por su eficiencia en conjuntos de datos grandes, pero su desempeño se ve afectado por la presencia de valores atípicos y la necesidad de definir previamente el número de grupos. K-medoides, al ser menos sensible a valores extremos, mejora la robustez del análisis, aunque es computacionalmente más costoso. Los algoritmos basados en densidad, de otra parte, permiten detectar conglomerados de formas arbitrarias y manejar ruido en los datos, pero dependen de parámetros sensibles para su efectividad. Por otro lado, los métodos jerárquicos ofrecen una representación más interpretable, aunque su complejidad limita su aplicabilidad en grandes volúmenes de datos. Respecto a esto los autores recomiendan que la elección del algoritmo debe considerar el tamaño del conjunto de datos, la distribución de los puntos y la necesidad de interpretar los resultados, combinando a menudo varias metodologías para optimizar la segmentación.

En este mismo sentido Pérez Londo et al. (2023) plantean una comparación de técnicas estadísticas multivariadas aplicadas a datos cualitativos, que permiten evaluar su eficiencia en términos de tiempo de ejecución y uso de memoria computacional. Este estudio comparativo en estadísticas multivariadas aplicadas a datos cualitativos evalúa su eficiencia computacional e indican que los métodos basados en Análisis de Correspondencias Múltiples (ACM) muestran un desempeño moderado en términos de tiempo de ejecución, pero requieren un uso de memoria considerable cuando los conjuntos de datos son grandes. Así mismo, los algoritmos de Análisis de Componentes Principales (ACP) adaptado a datos cualitativos son más eficientes en memoria, aunque pueden perder interpretabilidad en comparación con ACM. Por otra parte, respecto a la precisión y estabilidad, los modelos jerárquicos aplicados a datos cualitativos generan estructuras de agrupamiento más interpretables, aunque su tiempo de ejecución aumenta exponencialmente con el tamaño del conjunto de datos. Los métodos de partición, como K-medoides, ofrecen una segmentación más robusta con menor uso de memoria, pero pueden verse afectados por la inicialización de los centroides.

Respecto a la escalabilidad, los modelos basados en redes neuronales y machine learning permiten un análisis más detallado de datos cualitativos, pero su implementación requiere más recursos computacionales, a esto los métodos como DBSCAN son útiles para detectar patrones en datos no estructurados, aunque dependen de parámetros sensibles como el radio de agrupamiento. Finalmente de acuerdo con la aplicabilidad en distintos contextos en el análisis de encuestas y entrevistas, ACM (Correspondencias múltiples) y análisis factorial destacan por su capacidad para revelar relaciones latentes entre categorías, en estudios de segmentación de mercados o educación, los métodos de *clustering* como K-medoides y DBSCAN se muestran más eficientes al manejar datos heterogéneos a lo que se concluye que la selección de la técnica estadística multivariada más adecuada para datos cualitativos depende del balance entre precisión, interpretabilidad y eficiencia computacional, siendo los métodos como ACM los que ofrecen mayor interpretabilidad, mientras que técnicas de clustering son más eficientes en el manejo de grandes volúmenes de datos.

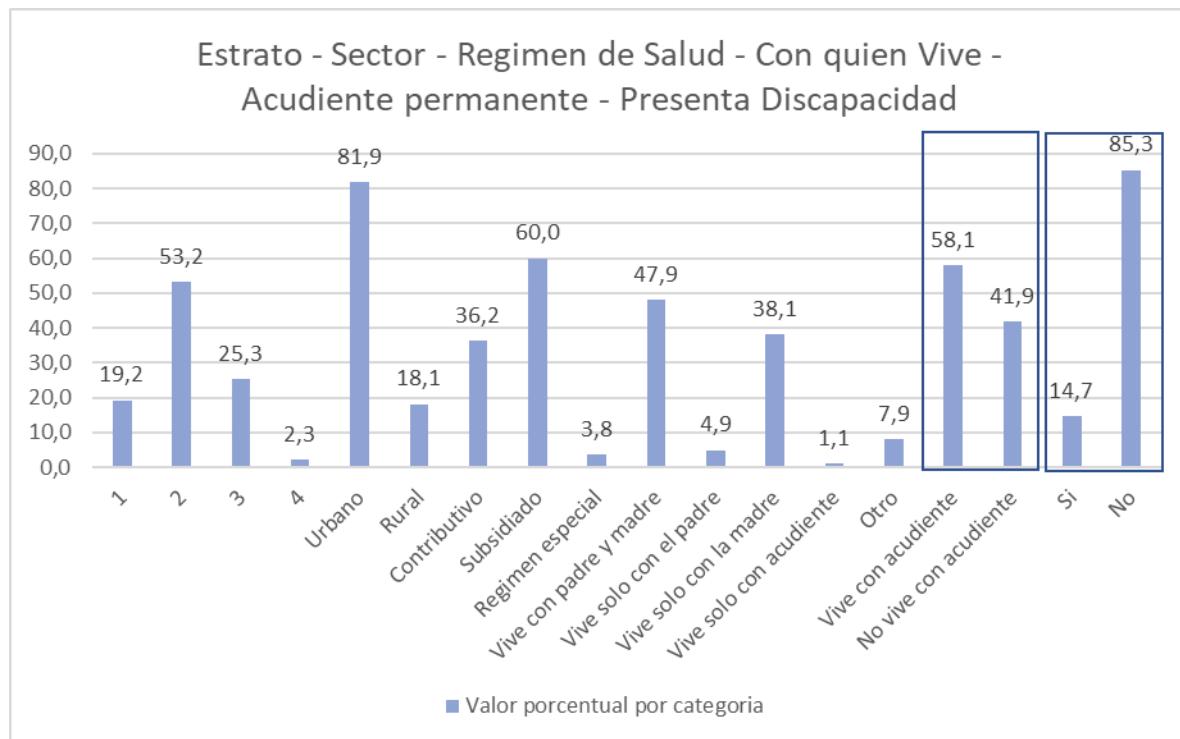
RESULTADOS-DISCUSIÓN

En lo particular del análisis de datos, se realiza en primer lugar la recolección de información, para ello se considera una muestra de 265 estudiantes de educación media de quienes se tiene registro completo sobre su caracterización sociodemográfica, a quienes se solicita diligenciar encuesta que indaga sobre género, estrato socioeconómico, actividades de sus padres, nivel de escolaridad de sus padres, número de hermanos, nivel salarial de sus padres y/o acudientes, lugar donde viven, entre otras variables. Sumado a esto para la muestra, se considera las valoraciones externas mediante exámenes de competencias practicadas por el ICFES (Instituto Colombiano de Fomento para la Educación Superior) denominadas Pruebas Saber 11, valoraciones que se miden en una escala de 0 a 500 puntos en cada área que la prueba evalúa.

La información cualitativa que complementa el estudio, se basa en una serie de entrevistas y grupos focales para los docentes y algunos de sus estudiantes de la muestra (265), quienes reflejan apreciación y percepción respecto a las relaciones que puedan existir en los resultados de pruebas externas con la caracterización socioeconómica.

Dentro de la descripción de los datos de caracterización, son de interés para la investigación el estrato socioeconómico, el sector donde vive, el régimen de salud, con quienes vive el estudiante, si tiene o no acudiente permanente y si presenta o no algún tipo de discapacidad; en este orden la Figura 1 resume estos valores en datos porcentuales:

Fig. 2: Valores porcentuales para variables socioeconómicas en la muestra de estudiantes



Fuente: Elaboración de autor

En la Figura 2 se identifican varios datos de referencia, en primer lugar, se considera el estrato socioeconómico numerado del 1 al 4, siendo la muestra que se ubica en estrato 2 con un 53,2% en su mayoría y 2,3% en estrato 4, esto indica que la mayoría de estudiantes corresponden a estratos socioeconómicos bajos. De igual manera la mayoría de estudiantes 81,9% corresponden al sector rural. Respecto al régimen de salud el 36,2% corresponde a contributivo, es decir que los padres o acudientes laboran de manera formal sumado al 3,8% que corresponde a regímenes de salud especiales como son los hijos de profesores y militares, aunque se observa que la mayoría (el 60%) corresponde a régimen de salud subsidiado. Siguiendo el gráfico hacia la derecha se encuentran los valores porcentuales de la variable "Con quien vive" esto indica que la mayoría con un 47,9% corresponden a hogares tradicionales, aunque un 38% vive únicamente con la madre; también es importante notar que el 58,1% de los estudiantes tienen acudiente permanente diferente a sus padres y finalmente existe un 14,7% de estudiantes que presentan discapacidad.

A la información socioeconómica se agregan los resultados en evaluación externa en competencias de lectura crítica, matemáticas, ciencias sociales - ciudadanas, ciencias naturales y finalmente competencia en inglés. Además del valor del puntaje global. La escala de estos datos está de 0 a 100 en las 5 competencias y de 0 a 500 en el puntaje global. Finalmente, a los datos se agrega información cualitativa sobre intención de estudios superiores, para lo cual se indaga a los estudiantes sobre la percepción que tienen para acceder a estudios universitarios, para ello se entrevista a estudiantes y profesores con interrogantes como "¿Qué cree que hace que se repreuben las materias?", "¿Qué le motiva a continuar con sus estudios?", entre otras y con ello se establecen categorías de análisis por medio de la metodología de la teoría fundamentada. La categoría principal es si están o no motivados a continuar sus estudios.

Dado que en este estudio interesa en mayor medida el análisis multivariado de la información con componentes cualitativos, se enfatiza en la técnica propuesta para generar la asociación de las variables por medio de un dendograma generado por la técnica de k-medias.

En este sentido, se involucra un concepto planteado a la comunidad científica basado en la definición de "Tensión" como elemento que genera la necesidad de integrar conceptos desde sus fronteras, sin romper el contacto ni colonizarlo, para lo cual se requiere de un diálogo disciplinar que construye conocimiento y desplaza fronteras (Guasmayan

Guasmayán, 2023). Esto conduce a proponer que la medición de distancias para la asociación de variables no pueda ser únicamente por medio de cálculos euclídeos y algorítmicos, sino que al tratarse de involucrar información cualitativa en los datos, esto genera “Tensión” entre los mismos datos, aportando así, un componente ponderador de las distancias entre la media de los centroides de los grupos de datos, es decir que el cálculo matemático de grado de asociación de los conglomerados por técnica de k-medias se debe realizar considerando un coeficiente ponderador que indique la importancia de agrupar una variable con otra, a este elemento matemáticamente se define como una tensión.

Como adición al algoritmo de k-medias, se propone un coeficiente ponderador que en este particular se denomina tensión, el cual se establece como una transformación lineal $T(x)=ax+b$, sin embargo, para suavizar su ponderación en la medida de distancia se realiza por medio de una función exponencial como se indica en la ecuación 1. (Alpala et al., 2015)

$$T(e^k) = \frac{1 - e^k}{1 - e} \quad (1)$$

La ecuación considera una ponderación de forma exponencial de acuerdo con el concepto que tengan los estudiantes sobre su intención y apreciación para el ingreso a estudios superiores, ponderación que cobra importancia de manera exponencial en el criterio de asociación. Para ello matemáticamente se incluye la tensión como un coeficiente que afecta directamente a la distancia euclídea entre los datos al generar el agrupamiento al ejecutar el algoritmo de k-medias hacia los centroides de cada grupo (Leiva et al., 2010).

El resultado principal se indica en el dendograma generado desde un algoritmo de k-medias corrido en Python con algunos de los datos seleccionados para los estudiantes y sus variables.

Fig. 3: Variables ingresadas al algoritmo K-medias

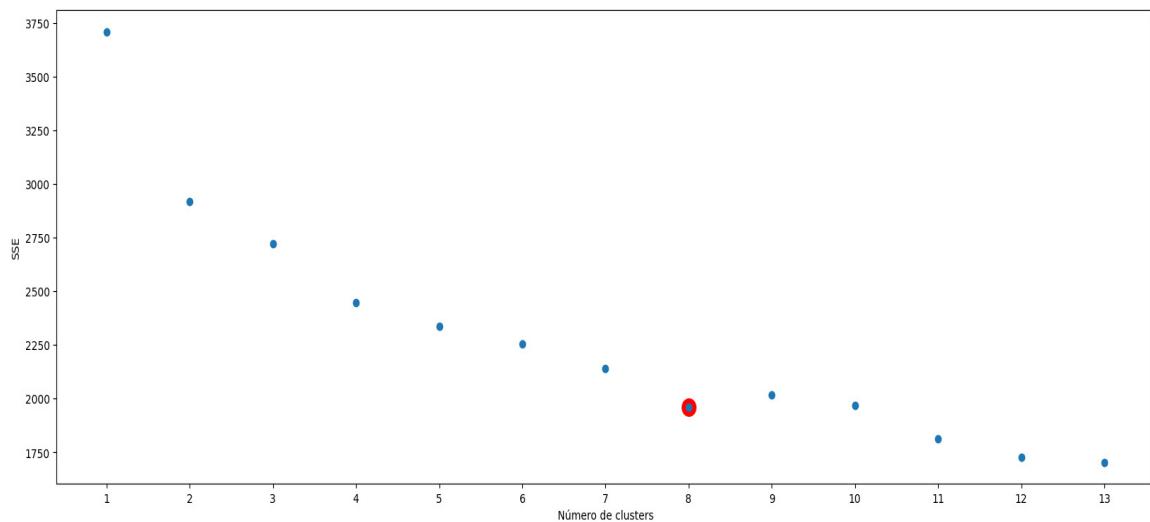
SEC	ENFASIS	L.CRI	MAT	SYC	CNAT	ING	PJE	ESTRATO	Sector	Salud	V_Padre	V_Madre	Discap	Tension
15	ACA	63	68	47	62	55	298	2	1	1	0	1	0	0.8204581737194935
14	ACA	57	56	56	52	41	271	2	1	2	0	1	0	0.07972387243569501
2	ACA	57	55	54	60	48	279	2	1	2	1	1	0	0.8127643069361902
14	ACA	59	45	47	49	57	253	1	2	1	0	1	1	0.061721219417957496
2	ACA	42	31	39	45	49	200	2	1	1	1	1	0	0.8694418022930479

Fuente: Elaboración de autor

En la Figura 3 se observan las variables consideradas en la ejecución del algoritmo, estas son la categoría de los estudiantes o “Énfasis”, en esta variable se identifica el tipo de bachillerato que tiene la Institución Educativa así: Académica, Electricidad, Sistemas, Robótica, Comercio, Salud ocupacional, Primera infancia y agropecuaria. Luego aparecen los resultados de pruebas externas para Lectura crítica, matemáticas, Ciencias naturales, inglés y el puntaje general; luego variables socioeconómicas como Estrato socioeconómico de sus viviendas, el sector de clasificación como rural o urbano, el tipo de empresa prestadora de salud como pública, privada o de régimen especial; la categoría de si vive con su padre, con su madre, si presenta alguna discapacidad y el coeficiente de tensión.

Al ejecutar el algoritmo se indica cómo evoluciona el error en el método de optimización:

Fig. 4: Evolución de error

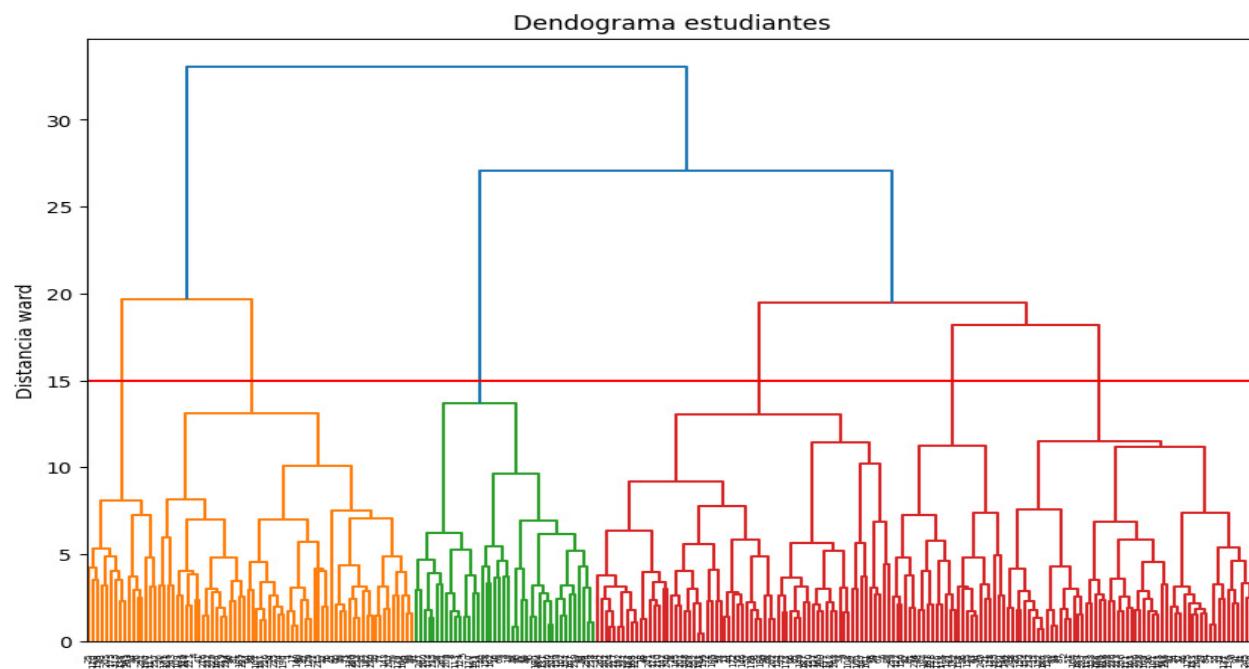


codigo

Fuente: Elaboración de autor

En la Figura 4, se indica la evolución del error y por el método de “acodo” se muestra que una clasificación favorable es con 8 grupos, sin embargo, después de varias corridas se logra un resultado favorable con 5 grupos.

Fig. 5: Dendograma de clasificación



Fuente: Elaboración de autor

En la clasificación se ubican tres agrupamientos principales de los datos vistos en tres colores.

El agrupamiento se observa en la matriz que se genera con el método como lo indica la imagen de la Figura 6.

Fig. 6: Agrupamiento de datos por medio de distancia euclídea y método Ward

```
# Creamos el modelo
cluster = AgglomerativeClustering(n_clusters=5, metric='euclidean', linkage='ward')

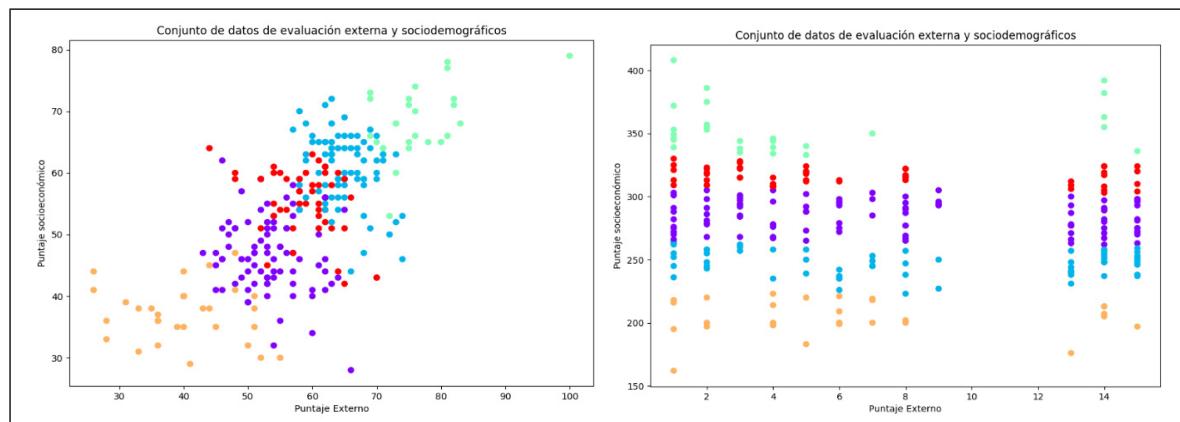
# Lo ajustamos con los datos
cluster.fit_predict(sdata) # fit_predict hace lo mismo que fit pero devuelve el vector

array([1, 0, 4, 0, 3, 3, 2, 0, 0, 2, 2, 4, 0, 1, 4, 0, 4, 0, 1, 1, 3, 1,
       3, 1, 1, 1, 3, 0, 2, 4, 2, 1, 1, 0, 0, 0, 1, 1, 0, 0, 3, 0, 0,
       1, 0, 4, 2, 1, 1, 4, 4, 1, 2, 0, 0, 0, 1, 0, 2, 0, 0, 3, 2,
       1, 2, 2, 4, 0, 4, 2, 0, 0, 1, 3, 0, 2, 0, 4, 3, 2, 0, 4, 4, 0, 4,
       3, 2, 1, 4, 1, 4, 0, 1, 2, 3, 1, 0, 1, 0, 1, 0, 1, 1, 0, 0, 1, 3,
       4, 0, 0, 4, 1, 4, 1, 0, 4, 0, 4, 1, 1, 0, 0, 4, 3, 1, 0, 4, 0, 1,
       1, 4, 0, 0, 4, 3, 4, 0, 0, 0, 4, 0, 0, 1, 4, 2, 1, 2, 0, 1, 4, 0,
       0, 0, 1, 4, 1, 4, 2, 1, 1, 4, 4, 1, 3, 1, 1, 0, 3, 3, 2, 1, 0, 0,
       1, 1, 2, 1, 2, 4, 0, 1, 1, 3, 4, 2, 4, 1, 3, 0, 4, 1, 1, 1, 0, 0,
       0, 1, 1, 1, 2, 3, 4, 0, 4, 0, 0, 1, 3, 1, 0, 3, 0, 3, 4, 1, 3,
       1, 3, 1, 0, 3, 0, 0, 3, 0, 3, 0, 4, 3, 4, 2, 0, 1, 1, 0, 4, 1, 1,
       4, 3, 1, 1, 4, 1, 3, 0, 0, 3, 1, 0, 1, 4, 0, 0, 1, 1, 0, 3, 1, 1,
       1])
```

Fuente: Elaboración de autor

El arreglo de la Figura 6 muestra una clasificación de los datos en los clústeres 0,1,2,3 y 4.

Fig. 7: Agrupamiento de datos por medio de distancia euclídea y método Ward



Fuente: Elaboración de autor

Finalmente la Figura 7 indica la clasificación de los datos con dos coeficientes de tensión, la clasificación inicial (Gráfica a la izquierda) con un coeficiente de 1 presenta sectores agrupados con mayor aleatoriedad, sin embargo al considerar la tensión (Gráfica a la derecha) como un coeficiente definido con la experiencia del investigador se presenta un agrupamiento con mayor dispersión entre los datos de cada grupo pero con mayor claridad en los grupos, es decir que al conocer muy bien la población de manera cualitativa es posible sesgar los datos a conveniencia con el fin de clasificar a una población de acuerdo a sus necesidades de formación, sus características sociodemográficas y sus valoraciones académicas.

CONCLUSIONES

El estudio propuesto, es un acercamiento al análisis de datos mixtos, tanto cuantitativos como cualitativos con el fin de relacionar variables del ámbito educativo con técnicas clásicas de estadísticas multivariadas como lo es el análisis por conglomerados con la adición de un factor ponderante. De acuerdo con los resultados preliminares, es posible presentar a la comunidad científica un contexto que abra caminos y posibilidades de investigación en estos ámbitos que permiten acercarse de mejor manera a la realidad de las ciencias humanas desde la combinación de estrategias de asociación de información cuantitativa con un aporte cualitativo. Esta combinación de perspectivas permite una caracterización completa de los estudiantes, incorporando no solo sus resultados académicos y características socioeconómicas, sino también su percepción sobre la educación y motivaciones hacia la formación universitaria, permitiendo una segmentación precisa de los estudiantes en función de sus características académicas y sociodemográficas.

La propuesta de la consideración de un coeficiente de tensión como un ponderador basado en la percepción cualitativa de los estudiantes respecto a su intención de acceder a la educación superior permite mejorar la clasificación, destacando la importancia de considerar aspectos subjetivos en la asociación de datos, logrando identificar diferentes perfiles de estudiantes con base en su rendimiento académico, contexto socioeconómico y motivaciones personales, lo que abre posibilidades para diseñar estrategias educativas más adaptadas a las necesidades individuales y de grupos, esto gracias a la implementación de técnicas mixtas en la investigación educativa que enriquecen el análisis y que también ofrecen una base sólida a partir de estrategias validadas desde lo estadístico, con el fin de proponer una toma de decisiones en políticas educativas y programas de acompañamiento estudiantil de manera efectiva. Finalmente, se recomienda continuar explorando la integración de modelos cualitativos y cuantitativos en futuras investigaciones para mejorar la precisión en la identificación de patrones y tendencias en el ámbito educativo.

El estudio propuesto, es un acercamiento inicial al análisis de datos tanto cuantitativos como cualitativos con el fin de relacionar variables del ámbito educativo con técnicas clásicas de estadísticas multivariadas como lo es el análisis por conglomerados. De acuerdo con los resultados preliminares, es posible presentar a la comunidad científica un contexto que abra caminos y posibilidades de investigación en estos ámbitos que permiten acercarse de mejor manera a la realidad de las ciencias humanas desde la combinación de estrategias de asociación

de información cuantitativa con un aporte cualitativo. Esta combinación de perspectivas permite una caracterización completa de los estudiantes, incorporando no solo sus resultados académicos y características socioeconómicas, sino también su percepción sobre la educación y motivaciones hacia la formación universitaria, permitiendo una segmentación precisa de los estudiantes en función de sus características académicas y sociodemográficas.

La propuesta de la consideración de un coeficiente de tensión como un ponderador basado en la percepción cualitativa de los estudiantes respecto a su intención de acceder a la educación superior permite mejorar la clasificación, destacando la importancia de considerar aspectos subjetivos en la asociación de datos, logrando identificar diferentes perfiles de estudiantes con base en su rendimiento académico, contexto socioeconómico y motivaciones personales, lo que abre posibilidades para diseñar estrategias educativas más adaptadas a las necesidades individuales y de grupos, esto gracias a la implementación de técnicas mixtas en la investigación educativa que enriquecen el análisis y que también ofrecen una base sólida a partir de estrategias validadas desde lo estadístico, con el fin de proponer una toma de decisiones en políticas educativas y programas de acompañamiento estudiantil de manera efectiva. Finalmente, se recomienda continuar explorando la integración de modelos cualitativos y cuantitativos en futuras investigaciones para mejorar la precisión en la identificación de patrones y tendencias en el ámbito educativo.

REFERENCIAS BIBLIOGRÁFICAS

- Alpala, L. O., Peluffo-Ordóñez, D. H., y Guasmayan, F. (2015). Deforming objects via exponential homotopy: A first approach. *2015 20th Symposium on Signal Processing, Images and Computer Vision (STSIVA)* (Issue 2, pp. 1-6). <https://doi.org/10.1109/STSIVA.2015.7330401>
- Álvarez Rodríguez, F. J., Sánchez Guerrero, L., y García Gaona, A. R. (2022). *Transformación digital de las instituciones educativas* (1^a ed.). Alfa-Omega.
- Balderas Gutiérrez, I. (2016). Qualitative research. *INNOVA Research Journal*, 1(2), 1-9. <http://www.sagepublications.com>
- Berger, P. y Luckmann, Th. (1994). *La construcción social de la realidad*. Madrid: Amorrortu-Murguía.
- De la Fuente-Fernández, S. (2011). *Análisis conglomerados*. España: Facultad de Ciencias Económicas y Empresariales. Universidad Autónoma de Madrid (UAM). <https://www.fuenterebollo.com/Economicas/ECONOMETRIA/SEGMENTACION/CONGLOMERADOS/conglomerados.pdf>

- Delgado Orrillo, Y. A., Ciraso-Calí, A., Quesada-Pallarès, C., Matos Nunes, T. C., de Oliveira Figueiredo, G., García-Orriols, J., Rodrigues Guilam, M. C., y Martínez-Fernández, J. R. (2023). Patrones de aprendizaje de estudiantes de posgrado en salud pública: Relaciones con la identidad, la formación y el trabajo de profesores en Brasil. *Educación Médica*, 24(4). <https://doi.org/10.1016/j.edumed.2023.100829>
- Eraso, B., Brunet Icart, A., y Gosálbez, P. I. (2012). El diseño del análisis cualitativo multinivel: Una aplicación práctica para el análisis de entrevistas. *EM-PIRIA. Revista de Metodología de Las Ciencias Sociales*, 26, 15–44. http://www.redalyc.org/articulo_oa?id=297124737002
- García Jiménez, E., Flores, J. G., y Rodríguez Gómez, G. (1994). Análisis de datos cualitativos en la investigación sobre la diferenciación educativa. *Revista de investigación educativa*, 23, 179–213. <https://idus.us.es/items/f05161f8-80d2-4d67-b0e5-b41d26192170>
- Globales, D. (2004). Aplicación de la teoría fundamentada (*Grounded Theory*) al estudio del proceso de creación de empresas. En *Decisiones basadas en el conocimiento y en el papel social de la empresa: XX Congreso anual de AEDEM*, Vol. 2, 2007 (Comunicaciones), pág. 44
- Guasmayan Guasmayan, F. A. (2023). Diseño mecatrónico como acercamiento a la transdisciplinariedad en ingeniería mecatrónica. *Revista Mutis*, 13(2), 1–15. <https://doi.org/10.21789/22561498.1988>
- Leiva-Valdebenito, S. A., y Torres-Avilés, F. J. (2010). Una revisión de los algoritmos de partición más comunes en el análisis de conglomerados: Un estudio comparativo. *Revista Colombiana de Estadística*, 33, 321–339.
- Medina Romero, M. Á., Hurtado Tiza, D. R., Muñoz Murillo, J. P., Ochoa Cervantez, D. O., y Izundegui Ordóñez, G. (2023). *Método mixto de investigación: Cuantitativo y cualitativo*. Instituto Universitario de Investigación en Ciencia y Tecnología. <https://doi.org/10.35622/inudi.b.105>
- Merlino, A. y Martínez, A. (2007). Integración de métodos cualitativos y cuantitativos: Construyendo e interpretando clusters a partir de la teoría fundamentada. *FQS Forum: Qualitative Social Research*, 8, 1–21. <https://www.ssoar.info/ssoar/handle/document/8903>
- Ossama, O., Mokhtar, H. M. O., y El-Sharkawi, M. E. (2011). An extended k-means technique for clustering moving objects. *Egyptian Informatics Journal*, 12(1), 45–51. <https://doi.org/10.1016/j.eij.2011.02.007>
- Pardo, C. E., César, P., y Campo, D. (2007). Combinación de métodos factoriales y de análisis de conglomerados en R: El paquete FactoClass. *Revista Colombiana de Estadística*, 30(2), 231–245. <https://dialnet.unirioja.es/servlet/articulo?codigo=10152875>
- Pérez-Londo, N. A., Lema-Londo, D. S., Batallas-Carrillo, N. A., y Pazmiño-Maji, R. A. (2023). Comparación de técnicas estadísticas multivariadas usadas en datos cualitativos. *Revista Científica Universidad de Las Ciencias*, 9, 1725–1741. <https://doi.org/10.238557/dc.v9i3.3522>
- Taylor, S. J. y Bogdan, R. (2000). Introducción a los métodos cualitativos de investigación. *Ediciones Paidós*. <https://asodea.wordpress.com/wp-content/uploads/2009/09/taylor-s-j-bogdan-r-metodologia-cualitativa.pdf>
- Torres, N. I. y Buitrago, D. (2020). ¿Hay desigualdad en la elección de carrera en Colombia?: Un análisis de algunos factores sociodemográficos. *HSE Social and Education History*, 9(2), 128–153. <https://doi.org/10.17583/HSE.2020.4303>
- Villanueva-Díaz, A., Espinosa-Ayala, E., Hernández-García, P. A., Márquez-Molina, O., Hidalgo-Milpa, M., y Mireles-Arriaga, A. I. (2023). Calidad multidimensional de la carne de conejo, atributos cuantitativos y cualitativos desde la perspectiva del consumidor. *Estudios Sociales. Revista de Alimentación Contemporánea y Desarrollo Regional*, 33, 1–24. <https://doi.org/10.24836/es.v33i61.1287>