

Ciencia Latina
Internacional

Ciencia Latina Revista Científica Multidisciplinar, Ciudad de México, México.
ISSN 2707-2207 / ISSN 2707-2215 (en línea), noviembre-diciembre 2024,
Volumen 8, Número 6.

https://doi.org/10.37811/cl_rcm.v8i6

**APLICACIONES DE LOS MÉTODOS DE
ANÁLISIS DE CLÚSTER Y CORRESPONDENCIA
EN EL ESTUDIO DE RESULTADOS DE EXAMEN
DE ADMISIÓN DE LA UNIVERSIDAD NACIONAL
DE SAN ANTONIO ABAD DEL CUSCO, 2022**

APPLICATIONS OF CLUSTER AND CORRESPONDENCE
ANALYSIS METHODS IN THE UNIVERSIDAD NACIONAL DE
SAN ANTONIO ABAD DEL CUSCO ADMISSION EXAM
RESULTS STUDY, 2022

Luz Marina Catunta Guillén

Universidad Nacional de San Antonio Abad – Cusco

Santiago Soncco Tumpi

Universidad Nacional de San Antonio Abad – Cusco

Abelardo Huamán Guzmán

Universidad Nacional de San Antonio Abad – Cusco

DOI: https://doi.org/10.37811/cl_rcm.v8i6.14633

Aplicaciones De Los Métodos De Análisis De Clúster y Correspondencia En El Estudio De Resultados De Examen De Admisión De La Universidad Nacional De San Antonio Abad Del Cusco, 2022

Luz Marina Catunta Guillén¹luz.catunta@unsaac.edu.pe<https://orcid.org/0009-0009-1203-6493>Universidad Nacional de San Antonio Abad –
Cusco
Perú**Santiago Soncco Tumpi**santiago.soncco@unsaac.edu.pe<https://orcid.org/0000-0002-4380-7555>Universidad Nacional de San Antonio Abad –
Cusco
Perú**Abelardo Huamán Guzmán**Abelardo.huaman@unsaac.edu.pe<https://orcid.org/0000-0002-7813-9956>Universidad Nacional de San Antonio Abad –
Cusco
Perú

RESUMEN

El objetivo de la presente investigación fue, analizar las APLICACIONES DE LOS MÉTODOS DE ANÁLISIS DE CLÚSTER Y CORRESPONDENCIA EN EL ESTUDIO DE RESULTADOS DE EXAMEN DE ADMISION DE LA UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO, 2022, se aplicaron técnicas estadísticas del análisis de conglomerados como el algoritmo bietapico, PAM (partition around medoids) y clara (Clustering Large Applications) para poder cumplir con el objetivo trazado, la investigación tuvo un enfoque cuantitativo, de alcance descriptivo, mientras que el diseño de investigación es no experimental, de tipo transversal; así mismo la población de estudio son los estudiantes que, ya egresaron de la educación básica regular quienes postularon en los diferentes semestres de la universidad San Antonio Abad del Cusco.

Se han utilizado 3 algoritmos para poder conglomerar a los individuos, como el algoritmos de PAM (Partitioning Around Medoids) y CLARA (Clustering Large Applications) y el algoritmo bietapico, se calculó una matriz de distancias con la metodología de distancias mixtas de gower, en tanto se decidió utilizar el algoritmo bietapico por tener mejor medida de silueta de cohesión y separación; se determinó que 4 clústeres eran adecuados para describir los perfiles de los postulantes donde se observó que el clúster 1, tiene prevalencia de alumnos que no ingresaron cuando postularon, su edad promedio es 18.49 años y su nota promedio fue de 7.44, su procedencia en su mayoría es del Cusco, de sexo femenino, procedencia de colegio nacional y mayormente postulan al grupo D. En el clúster 2, tiene prevalencia de alumnos que si lograron una vacante en la UNSAAC, con edad promedio de 20.7 años y su nota promedio fue de 12.58, su procedencia mayormente es del Cusco, de sexo masculino, procedencia de colegio nacional; en el clúster 3, tiene prevalencia de alumnos que no ingresaron cuando postularon, su

¹ Autor principal

Correspondencia: luz.catunta@unsaac.edu.pe

edad promedio es 18.49 años y su nota promedio fue de 7.67, su procedencia en su mayoría es del Cusco, de sexo masculino, procedencia de colegio particular y mayormente postulan al grupo A; en el clúster 4, tiene prevalencia de alumnos que no ingresaron cuando postularon, su edad promedio es 18.50 años y su nota promedio fue de 7.75, su procedencia en su mayoría es del Cusco, de sexo femenino, procedencia de colegio particular y mayormente postulan al grupo A.

Palabras clave: perfil, postulantes, Algoritmo bietápico, PAM, CLARA



Applications Of Cluster And Correspondence Analysis Methods In The Universidad Nacional De San Antonio Abad Del Cusco Admission Exam Results Study, 2022

ABSTRACT

The objective of this research was to analyze the APPLICATIONS OF THE CLUSTER AND CORRESPONDENCE ANALYSIS METHODS IN THE UNSAAC ADMISSION EXAM RESULTS STUDY, 2022, were applied statistical techniques of cluster analysis such as the two-stage algorithm, pam (partition around medoids) and clear (Clustering Large Applications) in order to meet the stated objective, the research had a quantitative approach, descriptive in scope, while the research design is non-experimental, transversal type; Likewise, the study population is the students who have already graduated from regular basic education who applied in the different semesters of the San Antonio Abad University of Cusco.

3 algorithms have been used to cluster individuals, such as the PAM (Partitioning Around Medoids) and CLARA (Clustering Large Applications) algorithms and the two-stage algorithm, a distance matrix was calculated with Gower's mixed distance methodology, in Therefore, it was decided to use the two-stage algorithm because it had a better measure of cohesion and separation silhouette; It was determined that 4 clusters were suitable to describe the profiles of the applicants where it was observed that cluster 1 has a prevalence of students who did not enter when they applied, their average age is 18.49 years and their average grade was 7.44, their majority origin She is from Cusco, female, comes from a national school and mostly applies to group D. In cluster 2, there is a prevalence of students who did achieve a vacancy at UNSAAC, with an average age of 20.7 years and their average grade was 12.58. , his origin is mostly from Cusco, male, from a national school; In cluster 3, there is a prevalence of students who did not enter when they applied, their average age is 18.49 years and their average grade was 7.67, their origin is mostly from Cusco, they are male, they come from a private school and they mostly apply to the group A; In cluster 4, there is a prevalence of students who did not enter when they applied, their average age is 18.50 years and their average grade was 7.75, their origin is mostly from Cusco, they are female, they come from a private school and they mostly apply to the group A.

Keywords: profile, applicants, two-stage algorithm, PAM, CLARA.

Artículo recibido 09 noviembre 2023

Aceptado para publicación: 12 diciembre 2023



INTRODUCCION

El Análisis de agrupamiento indica Elguera (2018) es conocido como Análisis de conglomerados, es una técnica estadística multivariada cuyo propósito es agrupar un conjunto de objetos, tratando de lograr la máxima homogeneidad en cada grupo y la mayor diferencia entre los grupos.

El término "Análisis de Clúster" refiere Chávez (2020) abarca una diversidad de métodos muy útiles que se emplean para generar clasificaciones significativas. Este enfoque polifacético ofrece una gama amplia de herramientas que pueden ser aprovechadas con flexibilidad, permitiendo la creación de agrupaciones discernibles. Al utilizar el Análisis de Clúster, de acuerdo con Levy & Varela (2005) se abre un abanico de posibilidades para estructurar y organizar datos de manera efectiva, facilitando la identificación de patrones y relaciones relevantes. Este recurso multifacético se erige como un aliado valioso en la tarea de dar sentido y orden a conjuntos de información diversos. De acuerdo con lo manifestado por Fontalvo y Hoz (2020) para conformar grupos, se manejan medidas de distancias y criterios de eslabonamiento para la clústerización de las observaciones, evaluándose su nivel de disimilaridad o similaridad. En análisis de conglomerados y clusters de acuerdo con Sánchez y Velasco (2012) representa una técnica estadística multivariante, cuyo propósito es dividir un conjunto de objetos en grupos con vistas a que los perfiles de los objetos dentro de un mismo grupo sean análogos entre sí (cohesión interna) y los de los objetos de clúster diferentes sean desiguales (aislamiento externo). Cuando se busca la solución cluster el método sugerido es el algoritmo K medias (k-means) de acuerdo con Martínez y Marín (2017) que se caracteriza porque en cada etapa se comprueba si la configuración clúster que se tiene, hasta ese momento, puede ser mejorada reasignando observaciones de un grupo a otro. K-medias es el algoritmo de clustering por excelencia y también se le denomina algoritmo de Lloyd (S/F), pues fue propuesto por primera vez por Stuart Lloyd, en 1957

En el fascinante mundo del Análisis de Clúster o conglomerado, según lo precisa la Universidad Autónoma de Madrid (2017) destaca la singularidad de contar con escasa o nula información previa sobre la estructura de las categorías. Este aspecto lo diferencia notablemente de los métodos multivariantes de asignación y discriminación, donde la comprensión previa es esencial. En esencia, el Análisis de Clúster de acuerdo a lo manifestado por Fleiss y Zubin (1969) se convierte en una herramienta intrigante para explorar y revelar la complejidad inherente en conjuntos de datos, sin tener



prejuicios previos sobre la disposición de las categorías. Más concretamente manifiesta Cuadras (2019) el objetivo es ordenar las observaciones en grupos tales que el grado de asociación natural es alto entre los miembros del mismo grupo y bajo entre miembros de grupos diferentes.

Aunque la estructura de las categorías suele ser un misterio inicialmente según lo manifestado por Everitt & Torsten (2011) que, es común contar con algunas ideas sobre las características deseables e indeseables al definir un esquema de clasificación específico. A pesar de la incertidumbre inicial sobre la disposición de las categorías, a menudo se poseen ciertas percepciones acerca de los atributos que se consideran ideales o inaceptables al establecer un determinado sistema de clasificación. Este enfoque flexible dentro de la dinámica de los sistemas según Forrester (1961) permite adaptarse a la complejidad del análisis, permitiendo la incorporación de criterios claros y relevantes en el proceso de clasificación.

Para lo cual se recurrió a la metodología sistémica en opinión de Aracil (1995) El proceso de modelado consiste en el conjunto de operaciones mediante el cual, tras el oportuno estudio y análisis, se construye el modelo del aspecto de la realidad que resulta problemático

La presente investigación se ejecutó mediante la separación y cohesión con el método PAM

```
> library(cluster)
> gower_dist <- daisy(data_nueva,metric = "gower")
> # pam
> library(fpc)
> a=pamk(gower_dist,criterion="asw")
> a$crit
```

```
[1] 0.0000000 0.1731278 0.1715935 0.2169045 0.1992636 0.2015927 0.1
```

```
883585
```

```
[8] 0.1804238 0.1973837 0.1919144
```

```
> a$nc
```

```
[1] 4
```

a) Índice de separación y cohesión con el método CLARA

```
> # clara
> b=pamk(gower_dist,criterion="asw",usepam=FALSE)
```



> b\$crit

[1] 0.0000000 0.2993381 0.2444897 0.3370344 0.3422099 0.3442859 0.3

038037

[8] 0.3516095 0.3186699 0.2580091

> b\$nc

[1] 8

Información de Campo

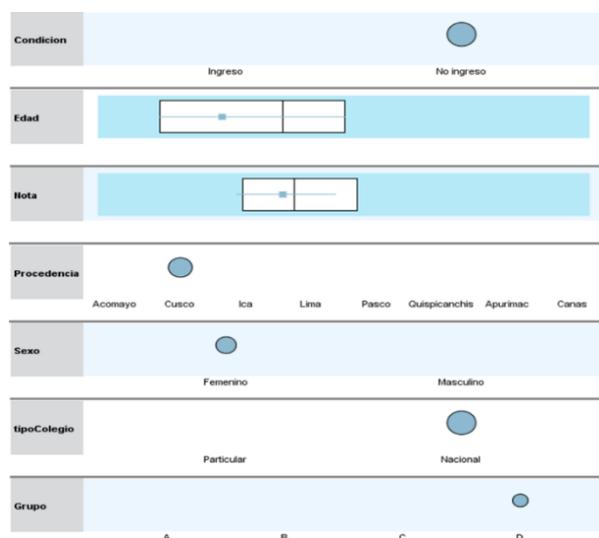
La muestra considerada para poder evaluar en la investigación estuvo conformada por 24,810 postulantes a la casa de estudios de la universidad Nacional de San Antonio Abad del Cusco la cual tiene la siguiente distribución de datos, según el cuadro siguiente por periodo de postulación.

Tabla 1. Postulantes a la UNSAAC por Semestre

Frecuencia	Porcentaje
2021-1	3,974
2021-2	4,215
2022-1	6,316
2022-2	5,895
2023-1	4,410
Total	24,810
	100,0

Análisis de Carga

Figura 1. Conglomerado 1



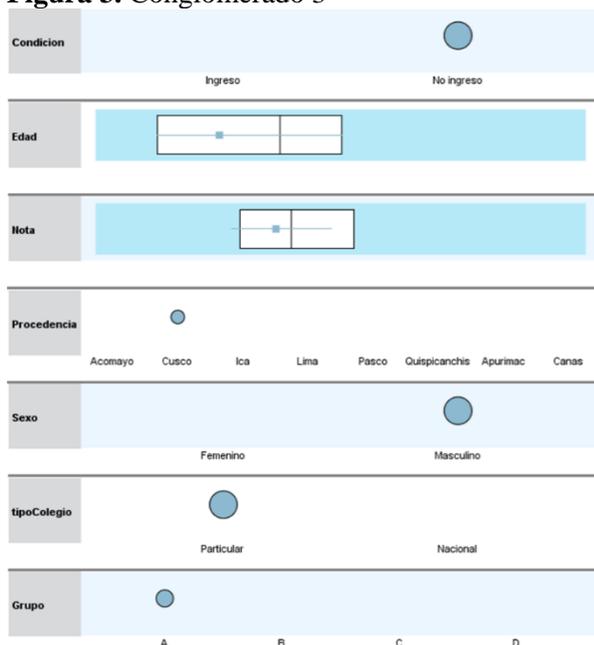
Se observa que en este conglomerado o clúster 1, tiene prevalencia de alumnos que no ingresaron cuando postularon, su edad promedio es 18.49 años y su nota promedio fue de 7.44, su procedencia en su mayoría es del Cusco, de sexo femenino, procedencia de colegio nacional y mayormente postulan al grupo D.

Figura 2. Conglomerado 2



Se observa que en este conglomerado o clúster 2, tiene prevalencia de alumnos que ingresaron cuando postularon, su edad promedio es 20.7 años y su nota promedio fue de 12.58, su procedencia en su mayoría es del Cusco, de sexo masculino, procedencia de colegio nacional y mayormente postulan al grupo A, obviamente que hay ingresantes a todos los grupos, pero hay más ingresantes al grupo A, donde se sitúa la mayor cantidad de escuelas profesionales en la universidad.

Figura 3. Conglomerado 3



Se observa que en este conglomerado o clúster 3, tiene prevalencia de alumnos que no ingresaron cuando postularon, su edad promedio es 18.49 años y su nota promedio fue de 7.67, su procedencia en su mayoría es del Cusco, de sexo masculino, procedencia de colegio particular y mayormente postulan al grupo A, donde se sitúa la mayor cantidad de escuelas profesionales en la universidad.

Figura 4. Conglomerado 4

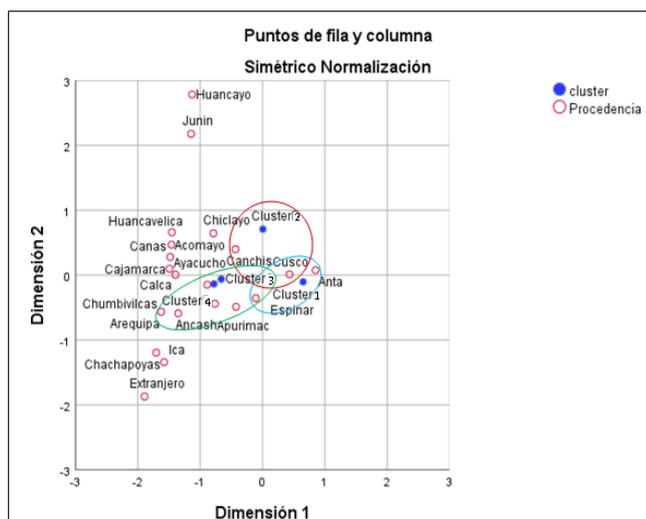


Se observa que en este conglomerado o clúster 4, tiene prevalencia de alumnos que no ingresaron cuando postularon, su edad promedio era de 18.50 años y su nota promedio fue de 7.75, su procedencia en su mayoría es del Cusco, de sexo femenino, procedencia de colegio particular y mayormente postulan al grupo A, donde se sitúa la mayor cantidad de escuelas profesionales en la universidad

Clúster	Masa	Puntuación en dimensión		Inercia	Contribución				
		1	2		Del punto en la inercia de dimensión		De la dimensión en la inercia del punto		
					1	2	1	2	Total
Clúster 1	,459	,651	-,104	,080	,472	,069	,996	,004	1,000
Clúster 2	,125	,007	,708	,005	,000	,868	,000	,996	,996
Clúster 3	,208	-,662	-,062	,039	,222	,011	,960	,001	,962
Clúster 4	,208	-,778	-,136	,053	,306	,053	,973	,005	,978
Total activo	1,000			,177	1,000	1,000			

En la tabla 2 anterior se puede evidenciar que para la dimensión 1 el clúster 1, es el que más influye, mientras que para la segunda dimensión el clúster 2, es el que más influye, así mismo se tienen las coordenadas en el biplot como por ejemplo para el clúster 1, su ubicación en el biplot es (0,459; 0,651) para el clúster 2, su ubicación en el biplot es (0,125; 0,007) y así sucesivamente para los demás puntos.





En la figura se observa que al superponer los puntos del clúster y la procedencia de los estudiantes se observa coincidencia del clúster 1, más cercano a las provincias de Cusco, Anta y Espinar, así mismo el clúster 2, está más próximo a Canchis y Cusco, así mismo el clúster 3 y 4 se encuentra más próximo a Cusco, Ancash, Apurímac, Espinar, Calca y Chumbivilcas.

OBJETIVO DEL ESTUDIO

El objetivo de la presente investigación fue la de analizar las APLICACIONES DE LOS MÉTODOS DE ANÁLISIS DE CLÚSTER Y CORRESPONDENCIA EN EL ESTUDIO DE RESULTADOS DE EXAMEN DE ADMISION DE LA UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO, 2022

METODOLOGIA

La metodología utilizada para la presente investigación corresponde a un diseño no experimental según lo manifiesta Sampieri (2006) el diseño no experimental es un tipo de investigación que se caracteriza por la **observación de fenómenos que ya ocurrieron**, sin la intervención del investigador. Esto implica que el investigador no puede controlar las variables independientes, solo puede observarlas y analizar sus relaciones con las variables dependientes.

Se aplicó en la presente investigación la transversalidad según lo manifiesta Hernández et al (2014) es un tipo de estudio que se caracteriza por la recolección de datos de una muestra en un único punto en el tiempo. Arias (2006) precisa que, este enfoque, también conocido como **transeccional**, permite analizar las variables o la relación entre ellas en un momento específico el diseño transversal como un tipo de

investigación no experimental que se centra en la observación de fenómenos en un momento determinado. Este enfoque se diferencia de los diseños longitudinales, que recopilan datos de la misma muestra en diferentes momentos en el tiempo.

Con el propósito de identificar los grupos de estudiantes se utilizó el algoritmo clustering que en opinión de Rai y Sing (S/F) además de Raulji (2014) y Velmurugam y Santhman consiste en un método exploratorio multivariado iterativo no supervisado, el cual describe el comportamiento de los objetos en grupos en fase exploratoria de investigación. Los resultados logrados en el análisis en opinión de Wang y Zang (2007) se refiere al modo en el cual el analista no asigna clases previamente, el mismo como lo manifiesta Faveró y Belfiore (S/F) y Scoltock (1982) fue utilizado en diversas áreas desde los años sesenta.

De acuerdo con lo manifestado por Aggarwal (2014), Everit y Hothorn (2011) y Ketchen y Shoock (1996) el algoritmo clasifica los objetos, asignándolos en grupos internamente homogéneos, y también heterogéneos entre los mismos.

Dentro de las opciones que se utilizaron en opinión de Boehmke y Greenwell (2014) y Janseen y Wan (2020) recomiendan utilizar el algoritmo clustering es el k-means que consiste en la técnica de distribuir los objetos mediante el sistema de particiones en un número k de clusters definidos previamente tal como lo menciona MacQueen (S/F) y Hartigan y Wong (1979).

Sin embargo, manifiesta Adams et al (2019) y Hair et al (2018) y Loperfido (2020) la utilización de este enfoque presente inconvenientes tales como la presencia de elementos con outliers; de acuerdo a lo manifestados por Acock (2014) Irizarry (2016) y Malhora (2018) los cuales pueden presentar efecto extremo en el análisis y ocasionar un agrupamiento inadecuado.

DEFINICION DE LA POBLACION

En la presente investigación se desarrolló con todos postulantes que se presentaron al examen de admisión ordinario desde el 2021-1 hasta el 2023-1 haciendo un total de 24,810 postulantes. Lo cual representa una muestra censal según lo define López & Fachelli (1998) la muestra censal es aquella porción que representa a toda la población, podríamos precisar también que se trata de un estudio.

INSTRUMENTO

El instrumento utilizado en la presente investigación fue el método Multivariado análisis de

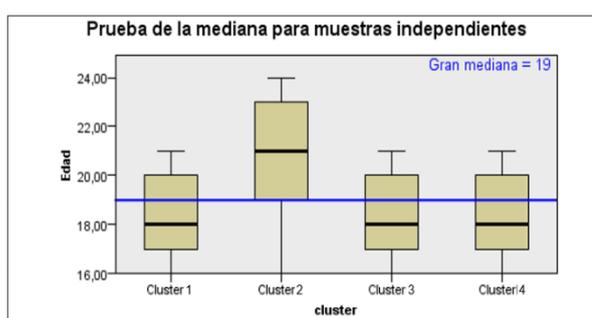


Correspondencia y Análisis de Agrupamiento o Clusterización que como manifiestan Sangaro y Matamoros (2020), son aquellas muestras que analizan múltiples características medidas en un individuo que por estar interrelacionadas no tiene sentido medir su efecto de manera aislada. Las mismas son más frecuentemente empleadas en los estudios observacionales acerca de la etiología y el pronóstico de una enfermedad, ya que permiten explicar, predecir, controlar variables confusoras y obtener mayor consistencia en la inferencia estadística

RESULTADOS

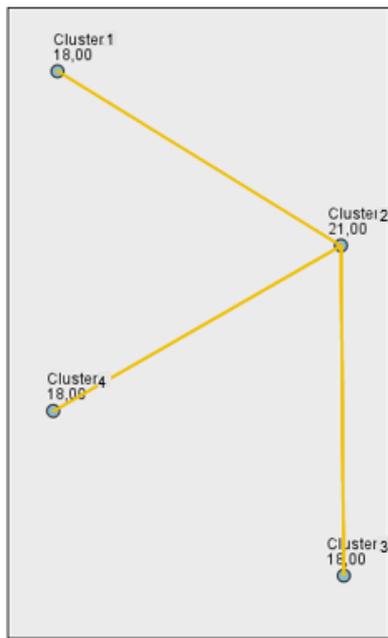
1.1. Comparación de la edad según clúster de pertenencia

N total	24.810
Mediana	19,000
Estadístico de contraste	1.626,785
Grados de libertad	3
Sig. asintótica (prueba bilateral)	,000



De la prueba estadística de Kruskal-Wallis (sig o pvalor = $0.00 < 0.05$) se evidenció que si existen diferencias de la edad en los diferentes clústeres, es así que observando el gráfico de cajas se evidencia que el clúster 2, en cuanto a edad se refiere presenta mayor nota promedio en comparación al resto, esto puede deberse por que por información anterior se vio que este clúster aglomeraba mayormente a los ingresantes a la universidad, lo explicado se puede observar en los gráficos posteriores y pruebas de comparaciones por pares.

Comparaciones entre parejas de cluster



Cada nodo muestra la mediana de la muestra de cluster.

Muestra 1-Muestra 2	Estadístico de contraste	Sig.	Sig. ajust.
Cluster.1.Cluster 3	,090	,764	1,000
Cluster.1.Cluster 4	,155	,694	1,000
Cluster.1.Cluster 2	1.358,052	,000	,000
Cluster 3.Cluster 4	,398	,528	1,000
Cluster 3.Cluster 2	1.199,094	,000	,000
Cluster 4.Cluster 2	1.148,803	,000	,000

DISCUSION

Huapaya et al (2011) indican que al analizar el nivel del conocimiento de los estudiantes encontró tres perfiles: individual, colectivo y colaborativo, en nuestro caso la investigación nos brindó 4 perfiles de alumnos postulantes los cuales tienen diferencias debido a sus características.

Zuniga et al (2022) conglomeró las carreras profesionales de medicina de distintas universidades usando un enfoque basado en dendrogramas identificó agrupamientos. Después de analizar 11 variables, 5 asociadas a los estudiantes y otras 6 asociadas a la institución o universidad. Como resultado, se obtuvieron dos dimensiones de clasificación: 1) un proxy de la calidad de las instituciones y de sus estudiantes, y 2) un proxy del costo anual y del perfil socioeconómico de los estudiantes. A un primer nivel de disimilitud aparecen dos grandes grupos de carreras de medicina: 1) perfil tradicional, regional,

con mejores indicadores de calidad institucional en promedio, y 2) perfil de instituciones privadas jóvenes, con casa central en Santiago de Chile y con estudiantes de mayor nivel socioeconómico. Se concluye que es posible caracterizar las carreras de medicina; de la misma manera que nosotros utilizamos una técnica de clustering (bietápico) para poder desarrollar nuestro propósito, si bien es cierto nosotros no usamos los dendogramas por la cantidad de sujetos a evaluar, pero el criterio de agrupar es el mismo y los resultados difieren puesto que a nosotros nos resultó 4 grupos de perfiles de estudiantes. De acuerdo con lo manifestado por Cerda (2011) quien nos manifiesta que, el razonamiento lógico inductivo o inteligencia lógica, presenta un carácter psico-evolutivo y su promoción temprana puede erigirse también, como una ayuda para el aprendizaje y comprensión de las matemáticas y por ende constituir una herramienta para hacer frente al fracaso escolar dado que presenta relaciones significativas con el rendimiento en matemáticas en la población escolar chilena

En Arora et al (2016) se aplicaron las técnicas más populares K- Means y K-Medoids sus resultados de la comparación muestran que el tiempo empleado en la selección de los valores iniciales y la complejidad espacial de la superposición del clúster es mucho mejor en K-Medoids que en K-Means. Además, K-Medoids es mejor en términos de tiempo de ejecución, no sensible a valores atípicos y reduce el ruido en comparación con KMeans, ya que minimiza la suma de las diferencias de los objetos de datos, estos resultados se asemejan a los de Arbin, Suhailayani, et al (2015); Nosotros no compartimos la misma conclusión debido que al comparar los algoritmos bietapico, PAM y CLARA logramos un mejor índice de cohesión y separación con el algoritmo mietapico esto debido a la gran cantidad de datos que nosotros utilizamos y el hecho de tener tanto variables cualitativas como cuantitativas hizo que este algoritmo tenga un mejor performance Chávez (2020) cuando caracterizo el perfil del ingresante de una universidad pública aplicando algoritmos clustering k-prototypes y k-medoides”, lograron identificar 3 tipos de alumnos: Ingresante previsto, Ingresante en proceso y el Ingresante en inicio; cada uno con características peculiares, las cuales permitirán a los responsables de las políticas educativas y en especial a los profesores consejeros saber el tipo de alumno que tienen a su cargo desde que ingresa a la universidad y empezar con ello políticas educativas como el emprendimiento del acompañamiento especializado, sistemático e integral; buscando la realización del paradigma del aprendizaje que la universidad se ha propuesto en su Modelo Educativo; la diferencia



con nuestro estudio es que nosotros analizamos a todos los postulantes mientras que Chávez analizo solo a los ingresantes, nosotros obtuvimos 4 perfiles de postulantes y el método usado fue el bietápico y no el K-medoides.

CONCLUSIONES

Se han utilizado 3 algoritmos para poder conglomerar a los individuos, con el algoritmo (bietápico) dio un índice de silueta y cohesión en 0.37 donde el número ideal de clusters fue 4; así mismo en los algoritmos de PAM (Partitioning Around Medoids) y CLARA (Clustering Large Applications) en el programa R, se calculó una matriz de distancias con la metodología de distancias mixtas de gower y con ello poder usar los otros algoritmos obteniéndose índices de 0.2169 con 4 clústeres en PAM y en el algoritmo CLARA se obtuvo 0.3516 con 8 clústeres, En tanto se decidió utilizar el algoritmo bietápico por tener mejor medida de silueta de cohesión y separación.

Se observa que en el conglomerado o clúster A, tiene prevalencia de alumnos que no ingresaron cuando postularon, su edad promedio es 18.49 años y su nota promedio fue de 7.44, su procedencia en su mayoría es del Cusco, de sexo femenino, procedencia de colegio nacional y mayormente postulan al grupo D. en el clúster B, tiene prevalencia de alumnos que ingresaron cuando postularon, su edad promedio es 20.7 años y su nota promedio fue de 12.58, su procedencia en su mayoría es del Cusco, de sexo masculino, procedencia de colegio nacional y mayormente postulan al grupo A, obviamente que hay ingresantes a todos los grupos, pero hay más ingresantes al grupo A, donde se sitúa la mayor cantidad de escuelas profesionales en la universidad; en el clúster C, tiene prevalencia de alumnos que no ingresaron cuando postularon, su edad promedio es 18.49 años y su nota promedio fue de 7.67, su procedencia en su mayoría es del Cusco, de sexo masculino, procedencia de colegio particular y mayormente postulan al grupo A; en el conglomerado D, tiene prevalencia de alumnos que no ingresaron cuando postularon, su edad promedio es 18.50 años y su nota promedio fue de 7.75, su procedencia en su mayoría es del Cusco, de sexo femenino, procedencia de colegio particular y mayormente postulan al grupo A, donde se sitúa la mayor cantidad de escuelas profesionales en la universidad.

RECOMENDACIONES

Investigar la efectividad en la formación de estudiantes de educación básica regular que pretenden ingresar a una Universidad Nacional como lo es la Universidad Nacional San Antonio Abad del Cusco,



debido a las bajas notas registradas.

En los colegios fortalecer el tema psicológico específicamente para abordar la ansiedad frente a las evaluaciones y la ansiedad social en materias como matemáticas y lenguaje, con el objetivo de mejorar su desempeño de los estudiantes a la hora de postular a la universidad.

A la universidad se recomienda tener en cuenta los perfiles encontrados en la presente investigación y así poder tener mayor amplitud a la hora de convocar a exámenes de admisión.

Implementar un programa de orientación específico para estudiantes de provincias que brinde información detallada sobre el proceso de postulación, becas disponibles y recursos de apoyo académico y emocional. Establecer una ruta para acceder a fondos de becas destinado exclusivamente a estudiantes de provincias, con criterios que tengan en cuenta las circunstancias económicas y las necesidades específicas de esta población; por otro lado, desarrollar campañas de concientización que destaquen los beneficios de estudiar en una universidad nacional, haciendo hincapié en las oportunidades académicas, la red de contactos y el prestigio asociado.

REFERENCIAS BIBLIOGRAFICAS

Acock A. A gentle introduction to Stata. 4th ed. College Station: Stata Press. 2014

Adams J, Hayunga D, Mansi S, Reeb D, Verardi V. Identifying and treating outliers in finance. *Financial Management*. 2019; 48(2): 345–384. Disponible en: <https://cutt.ly/9hawFSN>.

Arbin, N., Suhailayani, N., Zafirah, N., & Othman, Z. (2015). Comparative Analysis between K-Means and K-Medoids for Statistical Clustering. 3rd International Conference on Artificial Intelligence, Modelling & Simulation (AIMS), (págs. 117-121). doi:10.1109/AIMS.2015.82.

Aggarwal C. An introduction to cluster analysis. In C. Aggarwal, C. Reddy (Eds.). *Data clustering: Algorithms and applications* (pp. 1-28). New York: CRC Press. 2014.

Arias (2006) Fideas Arias libro El proyecto de investigación Introducción a la metodología científica Editorial Episteme

Arora, P., Deepali, & Varshney, S. (2016). Analysis of K-Means and K-Medoids Algorithm for Big Data. *Physics Procedia*, 78, 507-512.

Boehmke B, Greenwell B. K-means Clustering. In *Hands-On Machine Learning with R* (pp. 399–416). 1st ed. New York: CRC Press. 2014. Disponible en: <https://cutt.ly/KhaqBcJ>.



- Cerda, G., Pérez, C., Aguilar, M., & Aragón, E. (2018). Algunos factores asociados al desempeño académico en matemáticas y sus proyecciones en la formación docente. *Educ Pesqui*, São Paulo, 1-19. doi:<https://doi.org/10.1590/S1678-4634201706155233>
- Chávez, L. (2020). Caracterización del perfil del ingresante de una Universidad Pública aplicando algoritmos clustering K-Prototypes y K-Medoids. Lima, Perú: Universidad Nacional Agraria La Molina.
- Cuadras, C. (2019). Nuevos métodos de análisis multivariante. Barcelona, España: CMC Editions.
- Cuenca, R. (2015). La educación universitaria en el Perú: democracia, expansión y desigualdades. Lima: IEP Instituto de Estudios Peruanos.
- Elguera, R. (2018). “Segmentación de clientes de un casino utilizando el algoritmo partición alrededor de medoides (PAM) con datos Mixtos. Lima, Perú: Universidad Nacional Agraria La Molina.
- Everitt, B. (2011). *Cluster analysis 5th Editions*. Londres: Wiley.
- Everitt, B., & Torsten Hothorn. (2011). *An Introduction to Applied Multivariate Analysis with R*. Londres: Springer.
- Everitt B, Hothorn, T. Cluster analysis. In B. Everitt, T. Hothorn, *An Introduction to Applied Multivariate Analysis with R* (pp. 163–200). 1st ed. New York: CRC Press. 2011.
- Fávero L, Belfiore P. Análise de agrupamentos. In *Manual de análise de dados: Estatística e modelagem multivariada com Excel, SPSS e Stata* (pp. 309–378). 1st ed. São Paulo: GEN. 2017.
- Fleiss y Zubin J (1969) On the methods and theory of clustering. *Multivariate Behavioral Research* 4,235-250
- Forrester, J. W. (1961). *Industrial Dynamics*. The M.I.T. Press
- Fontalvo-Herrera, T. J., y Hoz-Granadillo, E. D. L. (2020). Conglomerate method-discriminant analysis-data envelopment analysis to classify and evaluate business efficient. *Entramado*, 16(2), 46-55.
- Hair J, Black W, Babin B, Anderson R. *Multivariate data analysis*. 8th ed. Ireland: Cengage Learning EMEA. 2018
- Hartigan J, Wong M. Algorithm AS 136: A K-means clustering algorithm. *Journal of the Royal Statistical Society*. 1979; 28(1): 100-108. Disponible en: <https://bit.ly/30jLpV1>.



- Hernández Sampieri, R., Fernández Collado, C., y Baptista Lucio, P. (2014). Metodología de la investigación. Gobierno de Colombia. <http://observatorio.epacartagena.gov.co/>
- Huapaya, C., Lizarralde, F., & Arona, G. (2011). Propuesta para Construir Perfiles Cognitivos en la Evaluación del Estudiante. XIII Workshop de Investigadores en Ciencias de la Computación (págs. 920-924). Argentina: Red de Universidades con Carreras en Informática (RedUNCI).
- Irizarry R, Love M. Data analysis for the life sciences with R. 1st ed. United Kingdom: Chapman and Hall/CRC. 2016.
- Janssen A, Wan P. K-means clustering of extremes. *Electronic Journal of Statistics*. 2020; 14(1): 1211–1233. Disponible en: <https://cutt.ly/ihaupE6>
- Ketchen D, Shook C. The application of cluster analysis in strategic management research: An analysis and critique. *Strategic Management Journal*. 1996; 17(6): 441–458. Disponible en: <https://cutt.ly/Whaq1Kh>.
- Levy Mangin, J. & Varela Mallou, Jesús (2005). Análisis multivariable para las ciencias sociales. Madrid-España. Editorial Pearson Educación
- Lloyd, Stuart P. (1982). "Least squares quantization in PCM" (PDF). *IEEE Transactions on Information Theory* **28** (2): 129–137. CiteSeerX 10.1.1.131.1338. doi:10.1109/TIT.1982.1056489. S2CID 10833328. Retrieved 2009-04-15.
- López & Fachelli (1998) Pedro López Roldán y Sandra Fachelli Metodología de la Investigación social cuantitativa Universidad Autónoma de Barcelona
- Loperfido N. Kurtosis-based projection pursuit for outlier detection in financial time series. *The European Journal of Finance*. 2020; 26(2–3); 142–164. Disponible en: <https://cutt.ly/dhaq0Oc>.
- Malhotra N. Marketing research: An applied orientation. 7th ed. New York: Pearson. 2018
- Martínez-Pérez, Martín Agustín. (2017). Agrupación de patrones en series de tiempo usando una red neuronal autoorganizativa (SOM) para el análisis de concentraciones de SO₂. *Revista de Ciencias Ambientales y Recursos Naturales*. 3(7): 16-24
- MacQueen J. Some methods for classification and análisis of multivariate observations. *Proceedings of the Berkeley symposium on mathematical statistics and probability*. 1967; 1: 281–297. Disponible en: <https://cutt.ly/YhaubYD>.



- Rai P, Singh S. A Survey of Clustering Techniques. International Journal of Computer Applications. 2010; 7(12): 1-5. Disponible en: <https://cutt.ly/OhauJpX>
- Raulji G. A Review on Fuzzy C-Mean Clustering Algorithm. International Journal of Modern Trends in Engineering and Research. 2014; 2(2): 751-754. Disponible en: <https://bit.ly/2FSxewM>.
- Sampieri (2006) Roberto Hernández Sampieri, Carlos Fernández Collado, Pilar Baptista Lucio
Editores: McGraw Hill España **Año de publicación:** 2014 **Edición:**6ª **País:** España **Idioma:** español **ISBN:** 978-1-4562-2396-
- Sagaró del Campo, Nelsa María, & Zamora Matamoros, Larisa. (2020). Técnicas estadísticas multivariadas para el estudio de la causalidad en Medicina. *Revista de Ciencias Médicas de Pinar del Río*, 24(2), 287-300. Epub 01 de marzo de 2020. Recuperado en 04 de noviembre de 2024, de http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S156131942020000200287&lng=es&tlng=es.
- Sánchez, P. P. I., Maldonado, C. J., & Velasco, A. P. (2012). Caracterización de las Spin-Off universitarias como mecanismo de transferencia de tecnología a través de un análisis clúster. *Revista Europea de Dirección y Economía de la empresa*, 21(3), 240-254.
- Scoltock J. A survey of the literature of cluster analysis. *The Computer Journal*. 1982; 25(1), 130–134. Disponible en: <https://cutt.ly/Ghaq8Rg>
- Universidad Autónoma de Madrid (2017) Análisis Conglomerado Cluster Facultad Ciencias Económicas Empresariales Universidad Autónoma de Madrid obtenido de chrome-extension://efaidnbnmnibpcjpcglclefindmkaj/https://www.estadistica.net/Master-Econometria/Análisis_Cluster.pdf
- Velmurugan T, Santhanam T. A comparative analysis between K-medoids and fuzzy C-means clustering algorithms for statistically distributed data points. *Journal of Theoretical and Applied Information Technolog*. 2011; 27: 19-29. Disponible en: <https://bit.ly/3867V6o>
- Joaquin Zuniga-Soria¹, Sergio Zuniga-Jara² y Karla Soria-Barreto^{2*} (1) Facultad de Medicina, Univ. Católica del Norte, Coquimbo-Chile (correo-e: joaquin.zuniga@alumnos.ucn.cl) (2) Escuela de Ciencias Empresariales, Univ. Católica del Norte, Coquimbo-Chile (correo-e: sz@ucn.cl; ksoria@ucn.cl) * Autor a quien debe ser dirigida la correspondencia Recibido May. 5, 2022;



Aceptado Jun. 20, 2022; Versión final Ago. 3, 2022, Publicado Dic. 2022

Wang W, Zhang Y. On fuzzy cluster validity indices. *Fuzzy Sets and Systems*. 2007; 158(19): 2095-2117. Disponible en: <https://cutt.ly/DhaifXB>

