

# El análisis de conglomerados como herramienta para evaluar el rendimiento académico: una experiencia en la universidad

## Cluster analysis as a tool to evaluate academic performance: an experience at the university

José Gregorio ARÉVALO Ascanio <sup>1</sup>; Santiago PÉREZ-GONZÁLEZ <sup>2</sup>

Recibido: 13/06/2018 • Aprobado: 28/07/2018

### Contenido

1. Introducción
  2. Metodología
  3. Resultados y discusión
  4. Conclusiones
- Referencias bibliográficas

#### RESUMEN:

Este artículo tiene como objetivo presentar la aplicación del análisis de conglomerados como herramienta para evaluar el rendimiento académico de los estudiantes de la Facultad de Ciencias Administrativas y Económicas de la Universidad Francisco de Paula Santander seccional Ocaña. La metodología utilizada se basó en una investigación aplicada, cuantitativa e inductiva con utilización de análisis conglomerados (clúster) para los estudiantes matriculados en el primer y segundo semestre del año 2017, el cual permitió agrupar las variables seleccionadas, tratando de lograr la máxima homogeneidad en cada grupo y su mayor diferencia en distancias. De la función de clúster obtenida y de los estadísticos analizados, se identificaron dos grupos o clústeres claramente definidos y asociados tanto en mayor como en menor grado al rendimiento académico de los estudiantes tomados como unidad de análisis.

**Palabras clave:** Análisis de conglomerados, clasificación de datos e individuos, rendimiento académico, universidad

#### ABSTRACT:

This article aims to present the application of cluster analysis as a tool to evaluate the academic performance of students of the Faculty of Administrative and Economic Sciences of the Universidad Francisco de Paula Santander sectional Ocaña. The methodology used was based on applied, quantitative and inductive research with the use of cluster analysis for students enrolled in the first and second semester of 2017, which allowed grouping the selected variables, trying to achieve maximum homogeneity in each group and its biggest difference in distances. From the cluster function obtained and from the statistics analyzed, two clearly defined groups or clusters were identified and associated to a greater or lesser degree to the academic performance of the students taken as the unit of analysis.

**Keywords:** Cluster analysis, classification of data and individuals, academic performance, university.

## 1. Introducción

El rendimiento académico, por ser multicausal, envuelve una enorme capacidad explicativa de los distintos factores y espacios temporales que intervienen en el proceso de aprendizaje (Garbanzo, 2007). En tal sentido, son diversas las variables que se relacionan con el rendimiento académico, en algunas de ellas median elementos tanto internos como externos a la persona. Sería importante analizar si el rendimiento académico está influenciado por factores como: el tipo de profesor, la metodología utilizada, los índices de capacidad de atención de los estudiantes, como lo demuestran estudios realizados en otros países (Vélez & Roa, 2005) o sí factores como sexo, edad estrato, fecha de culminación del bachillerato o el tiempo dejado de estudiar después de la terminación de estudios de secundaria (vacancia) son también determinantes en el rendimiento académico.

Ahora bien, en el marco de las evaluaciones de desempeño académico realizadas por los comités curriculares de la Facultad, se plantea un aporte desde los programas académicos de Administración de Empresas y Contaduría Pública, desarrollando una propuesta investigativa capaz de dar claridad mediante la aplicación de la técnica de cálculo multivariado del análisis de conglomerado, sobre la clasificación de sujetos u objetos a partir de características similares, como lo sugiere (Torrado & Berlanga, 2013) y al mismo tiempo poder evaluar el rendimiento académico de los estudiantes matriculados en el primer y segundo semestre del año 2017.

El presente artículo se estructuró en tres apartados, como sigue: en primera instancia se efectuó una breve revisión bibliográfica en las bases de datos SCOPUS, SCIENCE DIRECT, EBSCO, SCIELO Y REDALYC, encontrándose un número importante de investigaciones tanto de carácter teórico como empírico basadas en la aplicación de técnicas estadísticas multivariantes en el contexto educativo. En un segundo momento, se muestra que la metodología empleada fue de carácter cuantitativo a través de una investigación aplicada, cuantitativa e inductiva soportada en la técnica de cálculo multivariado del análisis de conglomerado.

Finalmente, en un tercer apartado se presenta el análisis de los resultados obtenidos de la aplicación de un análisis clúster respecto a la distribución de los valores de las variables, análisis con el que se pudo establecer, por un lado, la mayor homogeneidad posible entre cada conglomerado y, por otro lado, que los conglomerados sean muy distintos entre sí dadas sus distancias obtenidas y por último se muestran las conclusiones más importantes.

### 1.1. Investigaciones sobre la aplicación de técnicas estadísticas multivariantes en el contexto educativo

Con el objetivo de explorar los estudios sobre el uso de técnicas estadísticas multivariantes en el contexto educativo, se realizó una revisión sistemática en las bases de datos (SCOPUS, SCIENCE DIRECT, EBSCO, SCIELO y REDALYC), a partir de categorías como: análisis clúster y rendimiento académico. Finalmente, se encontraron ocho documentos (empíricos) y uno teórico, que se relacionaban de forma directa con las categorías planteadas, estos son:

Inicialmente (Lamos & Giraldo, 2011), concluyen en su estudio sobre el análisis del desempeño académico de los estudiantes de cálculo I en la UNAB que los resultados obtenidos sirven como referente para las diversas Instituciones de Educación Superior para implementar estrategias que mejoren los indicadores existentes sobre la mortalidad estudiantil. En la misma línea (Mendoza & Herrera, 2013), pretendieron determinar qué factores diferencian el grupo de estudiantes que culminan con éxito un periodo académico (o una asignatura), del grupo de estudiantes que lamentablemente fracasan en el alcance de logros y finalmente no aprueban el periodo académico.

De modo similar el trabajo de (González, 2004), presenta que en sus resultados trataron de delimitar por medio de la realización de un análisis discriminante las diferencias entre un rendimiento medio, alto y bajo entre 807 estudiantes de diferentes titulaciones, concluyeron que los alumnos clasificados como el grupo de rendimiento bajo, consideran que elemento que define una buena formación es la integración laboral de los titulados, lo cual no es posible sin la posesión de un diploma que acredite esa formación.

En su tesis doctoral (Garijo, 2016), concluye que las variables explicativas que determinan la función discriminante y que diferencian los grupos de rendimiento analizados, pueden trabajarse dentro del sistema educativo con el fin de intentar mejorarlas, además añade que las otras variables están relacionadas con aspectos familiares y son difíciles de modificar.

Por su parte (Martínez, Urrutia, & Ponce, 2013), realizaron un estudio de tipo analítico, retrospectivo y transversal, cuyo objetivo fue construir un

modelo predictivo a partir de un contexto multivariado, de algunos condicionantes que determinan la obtención del grado en la maestría y el doctorado. En tal investigación, se obtuvo información de 17 variables, en las cuales se reveló la importancia de variables que, de manera sinérgica contribuyen a explicar el perfil del estudiante de posgrado con éxito académico.

Análogamente (Santander, Rubén, & González, 2014), en su estudio sobre caracterización del rendimiento académico de los estudiantes del Plan de Estudios de Enfermería de la Universidad de Ciencias Médicas de La Habana, en el cual mediante un análisis factorial, de conglomerados y discriminante se detectó que la trayectoria del rendimiento se relaciona con los resultados que se obtienen en el Informe final de Investigación, en los cuales se aplican integralmente los conocimientos y las habilidades adquiridos para el trabajo de investigación.

De forma paralela (Fontalvo, Carrillo, & Cantillo, 2015), presentaron un estudio que tuvo como propósito efectuar un análisis multivariado para determinar los factores más relevantes de deserción estudiantil en los alumnos del programa de Ingeniería Industrial de la Universidad Autónoma del Caribe, en él encontraron como resultado dos grupos o clústeres claramente definidos, el primero agrupa 8 variables (jornada, tipo de colegio, género, rango de edad, ciudad de residencia, tipo de ingreso, departamento de residencia); estos factores fueron asociados con factores personales del estudiante. Lo anterior indica que están fuertemente ligados entre sí. El segundo grupo está compuesto de 12 variables: factor de elección de la universidad, factores académicos, vida universitaria, ayudas de la institución, factores económicos, ambiente educativo, ambiente familiar, financiación, promedio ponderado, orden académico, entre otros. Estos factores fueron asociados con el servicio educativo.

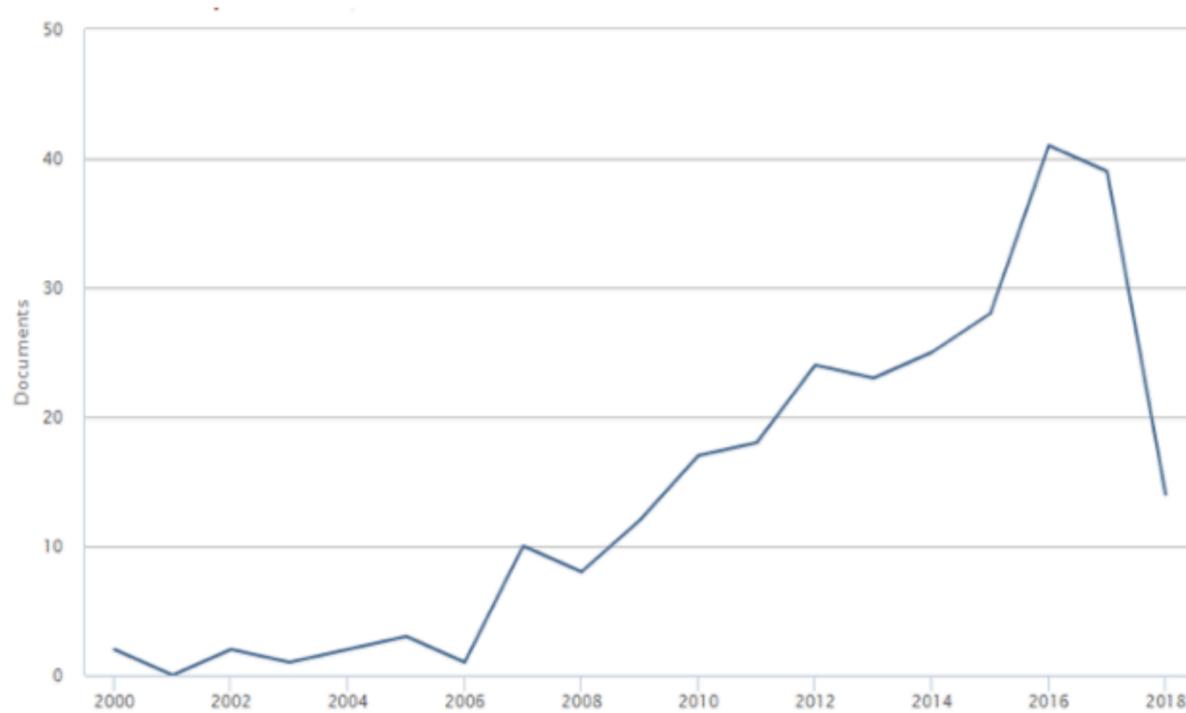
Del mismo modo (Gargallo, Pérez, & Serra, 2007), en su investigación sobre las "Actitudes ante el aprendizaje y rendimiento académico en los estudiantes universitarios", después de realizar un análisis de clúster pudieron delimitar cuatro grupos de alumnos, uno de ellos con perfil actitudinal débil, dos de ellos con un perfil actitudinal ligeramente más adecuado y uno de ellos con un buen perfil actitudinal.

Finalmente, (Leiva & Torres, 2010), presentan una investigación teórica sobre el análisis de conglomerados en la que hacen una comparación de diversos métodos de partición del análisis de conglomerados, usualmente conocidos como métodos no jerárquicos. Además presentan un estudio de simulación y aplicación con el fin de exponer y comparar los algoritmos de partición, según distintos esquemas de partición

A continuación, se aprecia el análisis de las investigaciones que sobre el tema se han realizado por parte de la comunidad científica, en donde se reportan 270 artículos usando solo el descriptor "rendimiento académico" en la base de datos SCOPUS (desde inicios del año 2000 hasta junio de 2018).

En cuanto al número de publicaciones por año se obtuvo la información que se observa en la Figura 1.

**Figura 1**  
Publicaciones de rendimiento académico de 2000 a 2018



Fuente: Los autores a partir de la base de datos SCOPUS

En la imagen anterior, se deja ver un avance muy lento en los primeros 10 años (2000 a 2009), es realmente a partir del año 2010 en donde comienza a presentarse un incremento en el número de divulgaciones, de manera puntual en el año 2016 es donde se alcanza el pico más alto de publicaciones entre los investigadores.

## 2. Metodología

La metodología se basó en una investigación aplicada, cuantitativa e inductiva, soportada en la técnica de cálculo multivariado del análisis de conglomerados, para contrastar si en dos periodos distintos se puede evidenciar la aplicación del análisis clúster y evaluar el rendimiento académico de los estudiantes tomados como unidad de análisis durante los dos semestres académicos del año 2017, para lo cual se estudiaron las variables: promedio académico de cada semestre, sexo, edad, estrato, fecha de culminación del bachillerato, fecha de ingreso y vacancia (entendida esta última como el espacio de tiempo comprendido entre la finalización del bachillerato y el ingreso a la universidad).

La población para esta investigación la constituyeron la totalidad de los estudiantes de la Facultad de Ciencias Administrativas y Económicas de la Universidad Francisco de Paula Santander seccional Ocaña (UFPSO), específicamente de los programas académicos de Administración de Empresas y Contaduría Pública.

Asimismo, para el desarrollo de esta investigación se utilizó como fuentes primarias los reportes de los rendimientos académicos generados por la División de Sistemas de la institución de educación superior objeto de estudio, lo cual permitió la aplicación la técnica de clúster para analizar el rendimiento académico en los periodos seleccionados.

Es pertinente señalar que para realizar el estudio de análisis de conglomerado, se utilizó el software SPSS 18, aplicación con la cual se establecieron las funciones de clúster y se estudiaron los diferentes estadísticos. Con el uso de esta técnica de cálculo multivariado se pudo aglomerar elementos en grupos homogéneos en función de las similitudes o similaridades entre ellos (Peña, 2012).

Como criterio de significancia estadística para el análisis de conglomerados se utilizaron algunas de las medidas de similitud habituales, como: la distancia euclidiana y distancia Euclídea al cuadrado; además se eligieron como métodos de conglomeración: el método de ward y la agrupación de centroides. Del mismo modo, se utilizó el dendograma, para representar gráficamente el resultado del proceso de agrupamiento en forma de árbol, el cual sirvió para mostrar cómo se va formando la clasificación jerárquica de las variables, y la homogeneidad dentro de los grupos resultantes y de igual manera visualizar las distancias entre ellos.

## 3. Resultados y discusión

### 3.1 Análisis de conglomerados

Uno de los objetivos fundamentales de este trabajo fue aplicar el análisis clúster como herramienta para evaluar el rendimiento académico de los estudiantes de la Facultad de Ciencias Administrativas y Económicas de la Universidad Francisco de Paula Santander seccional Ocaña. Los resultados expuestos a continuación, apuntan en la dirección de analizar lo que ocurre con los grupos concretos de estudiantes.

Por lo anterior, se lleva a cabo un análisis de clúster para determinar qué grupos de alumnos se constituían, usando medidas de similitud como la distancia euclidiana, la distancia Euclídea al cuadrado junto con los métodos de conglomeración: el método de Ward y la agrupación de centroides.

Se realizaron cuatro combinaciones, que permitieron identificar el grupo con mayor homogeneidad, en función de la similitud del rendimiento académico y las demás variables involucradas, siendo la segunda combinación la que parece más adecuada, en razón a que los dos grupos identificados son lo suficientemente distantes entre ellos, pero con homogeneidad dentro de cada grupo. Las combinaciones realizadas se resumen en el cuadro 1.

**Cuadro 1**  
Métodos de agrupación y medidas de similitud

COMBINACIONES	MÉTODO (de conglomeración)	DISTANCIA (medida de similitud)	No. DE CONGLOMERADOS
PRIMERA COMBINACIÓN	WARD	EUCLIDEA	3
SEGUNDA COMBINACIÓN	WARD	EUCLIDEA CUADRADO	3
TERCERA COMBINACIÓN	CENTROIDE	EUCLIDEA	3
CUARTA COMBINACIÓN	CENTROIDE	EUCLIDEA CUADRADO	3

Fuente. Elaboración propia

Partiendo de la Matriz de distancias y el Dendograma de conglomerados arrojados por el software estadístico SPSS-18 para cada una de las combinaciones, y comparando la mayor homogeneidad dentro de cada conglomerado, así como las mayores distancias entre ellos, se seleccionó la segunda combinación (Método Ward con medida de similitud Euclídea cuadrado) como aquella que logra la máxima homogeneidad en cada grupo y la mayor distancia entre los mismos. A pesar de parametrizar el sistema para la conformación de tres conglomerados, el sistema agrupó las variables en dos conglomerados, como se ilustra en el dendograma (figura 2) de combinaciones, nivel de distancias entre conglomerados (diagrama de codo, figura 3) y el cuadro de Historial de conglomeración (cuadro 2)

**Cuadro 2**  
Historial de conglomeración.

**Historial de conglomeración**

Etapa	Conglomerado que se combina		Coeficientes	Etapa en la que el conglomerado aparece por primera vez
	Conglomerado 1	Conglomerado 2		Conglomerado 1
	1	1		4
2	1	3	11077,444	1
3	2	7	653684,765	0
4	5	6	2710503,707	0
5	1	2	5200010,439	2
6	1	5	1,423E7	5

**Historial de conglomeración**

Etapa	Etapa en la que el conglomerado aparece por primera vez	Próxima etapa
	Conglomerado 2	
1	0	2
2	0	5
3	0	5
4	0	6
5	3	6
6	4	0

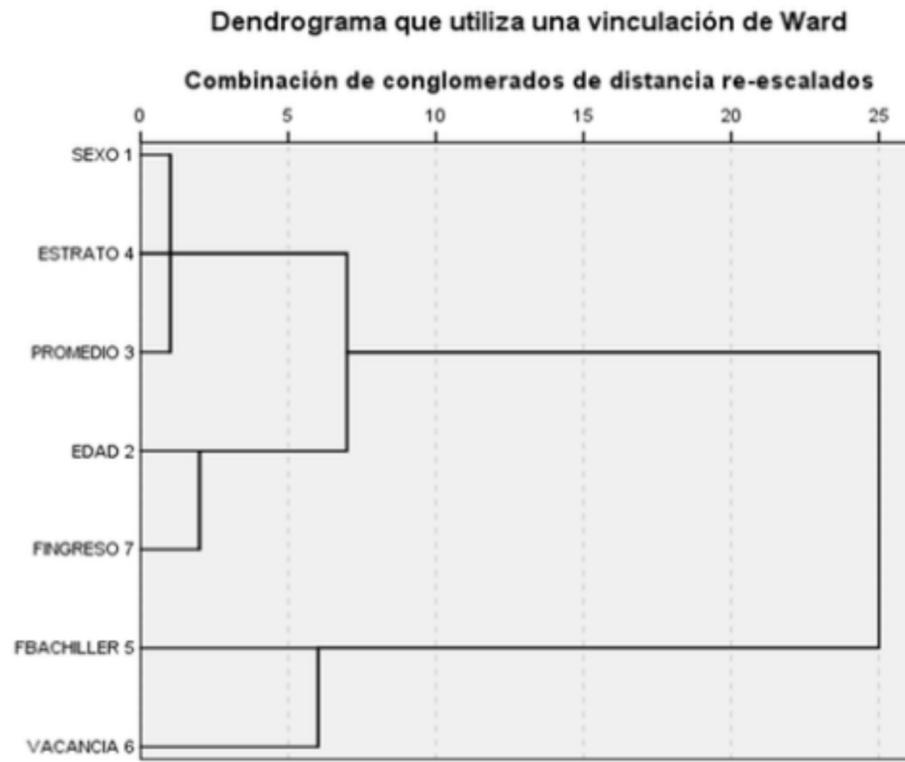
**Conglomerado de pertenencia**

Caso	3 conglomerados
SEXO	1
EDAD	2
PROMEDIO	1
ESTRATO	1
FBACHILLER	3
VACANCIA	3
FINGRESO	2

Fuente. Elaboración propia, a partir de SPSS.

----

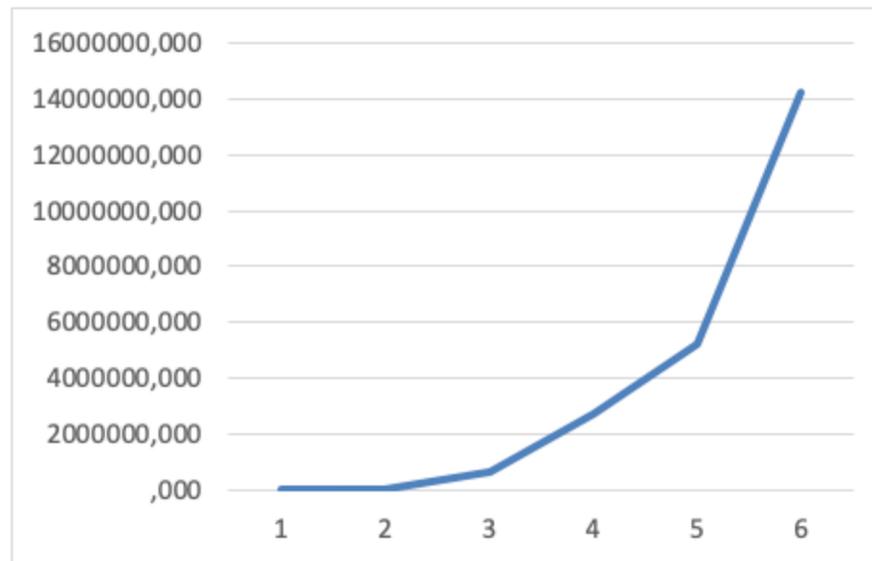
**Figura 2**  
Dendograma de combinación de conglomerados



Fuente. Elaboración propia, a partir de SPSS.

-----

**Figura 2b**  
Diagrama de codo



Fuente. Elaboración propia, a partir de SPSS.

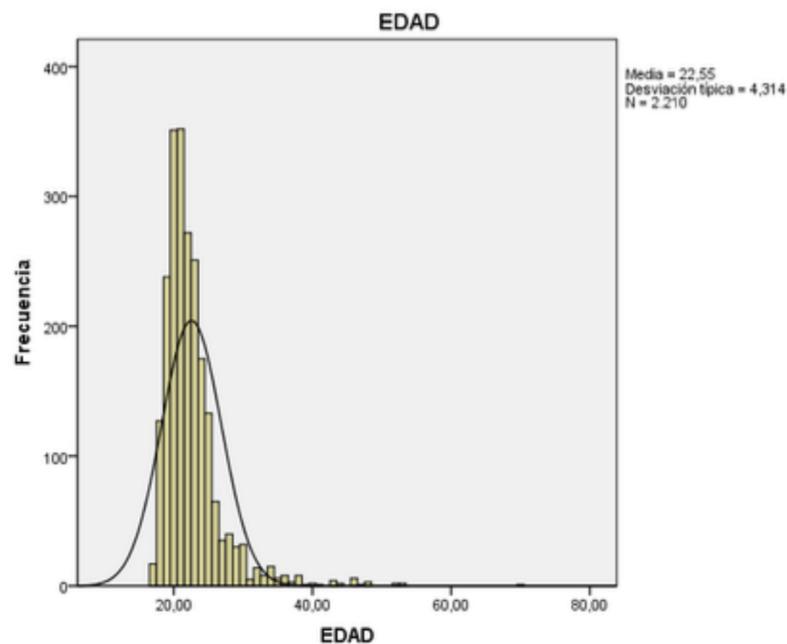
Al Interpretar el dendrograma resultante del análisis, el aspecto visual del gráfico parece indicar que hay 2 grupos, uno que encierra las variables sexo, estrato, promedio, edad, fecha de ingreso y otro con las variables bachiller y vacancia. Vale destacar que las variables sexo, estrato y promedio muestran una alta homogeneidad en la participación en el análisis dentro de uno de los grupos, así como la edad y el ingreso muestran alta similitud dentro de ese mismo grupo.

Estos dos grupos claramente diferenciados, podrían representar mejor la estructura de los datos, teniendo en cuenta la forma en que se van homogenizando los clúster y la medida de similitud o distancia entre ellos, este nivel de distancia entre conglomerados se refleja en el diagrama de codo.

### 3.2. Análisis estadístico de variables

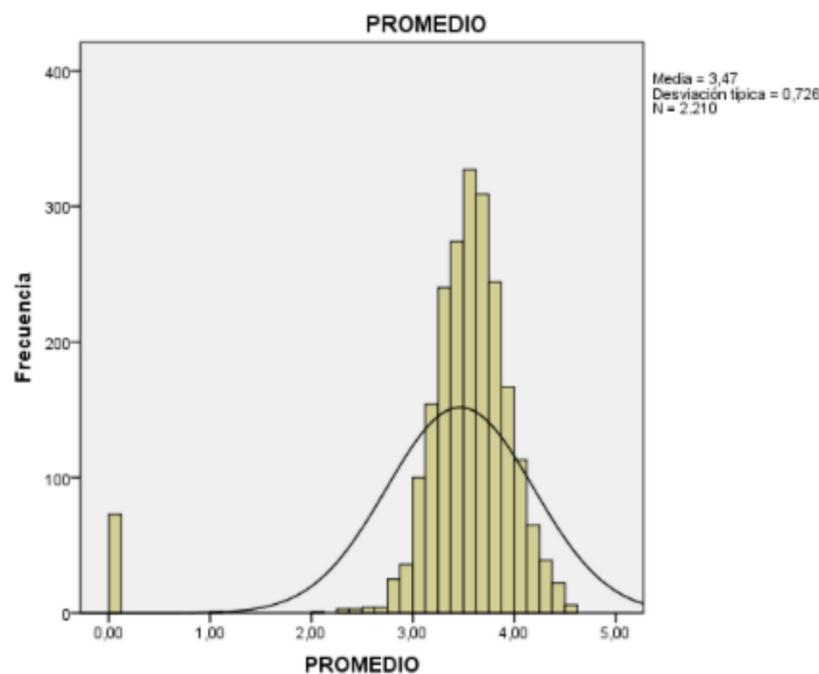
Con el objeto de resaltar la participación de cada una de las variables en el análisis realizado, se destaca a continuación el comportamiento de algunas de las variables relevantes por su impacto en el comportamiento del rendimiento académico en términos de las frecuencias en que ocurren. Para tal efecto, a continuación se describen las variables Edad, Promedio académico y Vacancia (entendida como el tiempo dejado de estudiar para ingresar a la universidad), en las figuras 3, 4 y 5 se pretende inferir o estimar las características de la población objeto de estudio, que sirven de ayuda para acciones de mejoramiento en la gestión académica.

**Figura 3**  
Diagrama variable EDAD



La edad promedio de los estudiantes participantes del análisis es de 22.5 años, donde se identifica en la gráfica una curtosis grande o gran apuntamiento, pero con una leve asimetría a la izquierda, lo que permite suponer que la mayoría de los datos que se acercan a la media, están alrededor de los 20 años de edad.

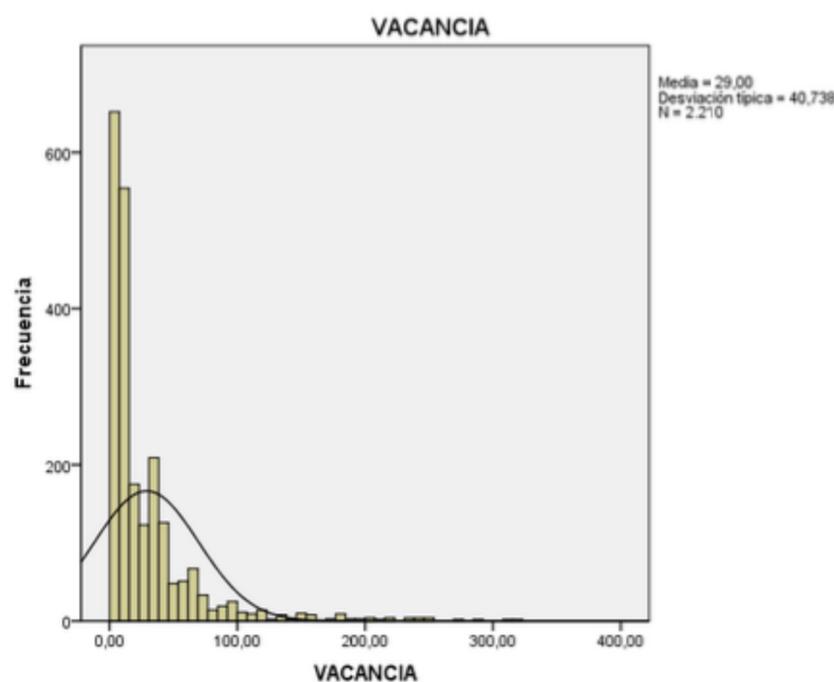
**Figura 4**  
Diagrama variable PROMEDIO ACADÉMICO



Fuente. Elaboración propia, a partir de SPSS

El rendimiento académico muestra una media de calificación del 3.47. En este caso, los datos muestran una curtosis no tan escarpada o apuntada y adicionalmente se distribuyen en forma simétrica alrededor de esta media.

**Figura 5**  
Diagrama variable VACANCIA



Fuente. Elaboración propia, a partir de SPSS.

La variable Vacancia, entendida como el tiempo dejado de estudiar para ingresar a la universidad, muestra un promedio o media de 29 meses (2,4 años), pero la frecuencia de participación de los datos en este análisis deja ver una alta asimetría en la gráfica, hecho que invita a monitorear de manera prioritaria esta variable en la institución. Adicionalmente, por el valor tan alto de la desviación, el comportamiento de esta variable no es representativo frente a la media encontrada.

## 4. Conclusiones

Se insiste en la importancia, que para el contexto educativo se reconozca la utilidad de técnicas estadísticas multivariantes como el análisis de conglomerados, en razón a que con ésta técnica puede asegurarse la identificación correcta y segura de la aplicación de métodos estadísticos para la aglomeración de diferentes observaciones y comportamientos en grupos homogéneos de variables que aporten a la solución de problemas en el ámbito educativo.

El estudio presenta como resultado dos grupos o clústeres claramente definidos, el primero agrupa 5 variables (sexo, estrato, promedio, edad y fecha de ingreso); el segundo grupo está compuesto de 2 variables: fecha de culminación del bachillerato y vacancia. Las variables de ambos clúster están asociadas tanto en mayor como en menor grado al rendimiento académico de los estudiantes.

Algunos de los tópicos que adquieren especial relevancia al momento de estudiar las variables que influyen en el rendimiento académico de los estudiantes son precisamente la edad y la vacancia, entendida esta última como el espacio de tiempo comprendido entre la finalización del bachillerato y el ingreso a la universidad. En este sentido, resulta muy útil encontrar las correlaciones significativas entre dichas variables, para de esta forma hacer una primera aproximación a la posible relación existente entre ellas.

Las variables consideradas para el análisis estadístico (Edad, Promedio y Vacancia), dejan ver una distribución simétrica alrededor de las medias obtenidas en los rangos establecidos, con excepción de la vacancia, que muestra una alta desviación entre el tiempo dejado de estudiar, es decir el comportamiento de esta variable no es representativo.

Se espera que este manuscrito, contribuya a la comprensión de la técnica estadística de análisis multivariado y que del mismo modo su aplicación al contexto educativo, sirva de referente para que se use no solo para analizar el rendimiento académico sino también se aplique por parte de los estudiantes e investigadores en los múltiples eventos de estudio propios del contexto educativo.

## Referencias bibliográficas

Fontalvo, W., Carrillo, M., & Cantillo, S. (2015). Análisis multivariado para determinar los factores más relevantes de deserción estudiantil presentes en el programa de Ingeniería Industrial de una Universidad del Caribe colombiano. *Prospect*, 86-98.

Garbanzo, G. (2007). Factores asociados al rendimiento académico en estudiantes universitarios, una reflexión desde la calidad de la educación superior pública. *Revista Educación*, 43-63.

Gargallo, B., Pérez, C., & Serra, B. (2007). Actitudes ante el aprendizaje y rendimiento académico en los estudiantes universitarios. *Revista Iberoamericana de Educación*, 1-25.

Garijo, R. (2016). *Análisis de los factores determinantes del bajo rendimiento académico en educación secundaria obligatoria en la provincia de Albacete*. Tesis Doctoral, Escuela de Doctorado UNED.

González, I. (2004). Realización de un análisis discriminante explicativo del rendimiento académico en la universidad. *Revista de Investigación Educativa*, 43-59.

Lamos, H., & Giraldo, J. (2011). Un modelo conceptual para el análisis del desempeño académico de los estudiantes de cálculo I en la UNAB. *Revista Educación en Ingeniería*, 115-125 .

Leiva, S., & Torres, F. (2010). Una revisión de los algoritmos de partición más comunes en el análisis de conglomerados: un estudio comparativo. *Revista Colombiana de Estadística*, 321-339.

Martínez, A., Urrutia, M., & Ponce, R. (2013). Perfil del estudiante de posgrado con éxito académico en la UNAM. *tarbiya* 32, 133-145.

Mendoza, A., & Herrera, R. (2013). *Propuesta para la predicción del rendimiento académico de los estudiantes de la Universidad del Atlántico, basado en la aplicación del análisis discriminante*. Cartagena: Innovación en investigación y educación en ingeniería: Factores claves para la competitividad global.

Peña, D. (07 de enero de 2012). *www.researchgate.net*. Recuperado el 05 de Junio de 2018, de [https://www.researchgate.net/profile/Daniel\\_Pena4/publication/40944325\\_Analisis\\_de\\_Datos\\_Multivariantes/links/549154880cf214269f27ffae/Analisis-de-Datos-Multivariantes.pdf](https://www.researchgate.net/profile/Daniel_Pena4/publication/40944325_Analisis_de_Datos_Multivariantes/links/549154880cf214269f27ffae/Analisis-de-Datos-Multivariantes.pdf)

Santander, A., Rubén, M., & González, C. (2014). Caracterización del rendimiento académico de los estudiantes de enfermería en la disciplina informática en salud. *Revista Cubana de Informática Médica*, 120-139.

Torrado, M., & Berlanga, V. (2013). Análisis Discriminante mediante SPSS. *REIRE*, 150-166.

Vélez, A., & Roa, C. (2005). Factores asociados al rendimiento académico en estudiantes de medicina. *Educación Médica*, 74-82.

---

1. Administrador de Empresas, Magíster en Administración de empresas, Candidato a Doctor en Administración. Docente- investigador del grupo GIDSE. Universidad Francisco de Paula Santander Ocaña (Colombia): Correo: [jgarevaloa@ufpso.edu.co](mailto:jgarevaloa@ufpso.edu.co)

2. Ingeniero Industrial, Magíster en Sistemas Integrados de Gestión QHSE, Candidato a Doctor en Administración. Docente- investigador de la Universidad de Boyacá, sede Sogamoso (Colombia): Correo: [sanperez@uniboyaca.edu.co](mailto:sanperez@uniboyaca.edu.co)

---

Revista ESPACIOS. ISSN 0798 1015  
Vol. 39 (Nº 46) Año 2018

[Índice]

[En caso de encontrar algún error en este website favor enviar email a [webmaster](mailto:webmaster)]