

Eficiencia académica estudiantil: un enfoque de análisis de componentes principales y análisis envolvente de datos

Student academic efficiency: a principal component analysis and data envelopment analysis approach

Rohemi Zuluaga¹
Enrique J. Delahoz-Domínguez²
Hoolyana Ojeda³

DOI: <https://doi.org/10.18041/1909-2458/ingeniare.29.7433>

RESUMEN

En el siguiente artículo se propone el uso de las herramientas de análisis de componentes principales (PCA) y análisis envolvente de datos (DEA) como una metodología para el análisis, la medición y la clasificación de la eficiencia académica de los estudiantes con los resultados de las pruebas Saber PRO en Colombia. La base de datos utilizada pertenece a los resultados de las pruebas Saber PRO del 2018, provenientes del Instituto Colombiano para la Evaluación de la Educación (Icfes). En los resultados se presenta que solo 16,67 % del grupo de estudio es eficiente; con estas unidades se forman quince grupos de referencia. Por otro lado, las unidades no eficientes deben reforzar las competencias de comunicación escrita (CE) y pensamiento científico, matemático y estadístico (PCME).

Palabras clave: Educación; Eficiencia; DEA; PCA.

ABSTRACT

The following paper proposes the use of the Principal Component Analysis (PCA) and Data Envelopment Analysis (DEA) tools as a methodology for the analysis, measurement and classification of students' academic efficiency with the results of the Saber Pro test in Colombia. The database used belongs to the results of the Saber PRO test results of the year 2018 from the Colombian Institute for the Evaluation of Education (ICFES). The results show that only 16.67% of the study group are efficient. Within this group 15 units of reference were formed. On the other hand, the non-efficient units in the other group must reinforce the competences of Written Communication (EC) and Scientific, Mathematical and Statistical Thought (PCME).

Keywords: Education; Efficiency; DEA; PCA.



Como citar este artículo: A. Delgado Tobón y E. Rodríguez Carmona, Valoración de acetinas como reemplazo parcial de aceite nafténico en compuestos de caucho SBR15021, *ingeniare*, vol. 2, n.º 29, dic. 2020.

1. Universidad Tecnológica de Bolívar, Facultad de Ingenierías, programa Ingeniería Industrial. Campus de Ternera, Cartagena, Colombia. Correo: rzuluaga@utb.edu.co

2. Universidad Tecnológica de Bolívar, Facultad de Ingenierías, programa Ingeniería Industrial, Campus de Ternera, Cartagena, Colombia. Correo: edelahoz@utb.edu.co

3. Punto Estratégico, Grupo de investigación Ingenio, Barranquilla, Colombia. Correo: hoolyana0623@gmail.com

1. INTRODUCCIÓN

En Colombia la educación se define como “un proceso permanente y continuo del ser humano, fundamentado en la concepción integral de la persona, su dignidad, sus derechos y deberes” [1].

La Constitución Política colombiana establece que la educación, primero, debe ser ofertada, garantizada y de calidad. Segundo, es un derecho de todo individuo y es un servicio público, por tanto, debe ser regulado y cumplir una función. Tercero, debe promover el desarrollo integral de cada uno de los individuos que la recibe. Finalmente, genera desarrollo económico, social y el capital humano del país.

El sistema educativo colombiano es conformado por la educación inicial, la educación preescolar, la educación básica, la educación media y la educación superior [1].

Se define la educación superior como el conjunto de procesos de aprendizaje que buscan la generación y el desarrollo de habilidades y conocimientos de los educandos de una manera integral [2]. La educación superior se recibe cuando se ha finalizado el ciclo de la educación media.

Globalmente, el reto de los gobiernos para el beneficio de los ciudadanos es lograr la calidad y la cobertura de la educación. No obstante, no solo es un tema que desea el Gobierno, sino que es un objetivo que exige la sociedad. En Colombia, por ejemplo, se presentaron en el 2011 al Ministerio de Educación Nacional (Mineducación) dos propuestas para cambiar el rumbo de la educación: la primera, llamada “Acuerdo por lo Superior 2034: propuesta de política pública para la excelencia de la educación superior en Colombia”, presentada por el Consejo Nacional de Educación Superior. La segunda, la propuesta denominada “Gran Acuerdo Nacional Por la Educación”, presentada por el movimiento Todos por la Educación [3].

La exigencia de la sociedad por la calidad y la cobertura de la educación apunta a una razón, y es que la educación —no solo en el caso de Colombia sino de cualquier país— es una inversión para nuestras sociedades.

Es correcto pensar que la inversión sugiere un contenido económico, pero en este caso queremos señalar el contenido social que representa realizar la inversión en la educación. Esto significa concebir la institución de educación superior (IES) como generadora y desarrolladora de la potencialidad humana necesaria para transformar y desarrollar la sociedad [4].

De acuerdo con todo lo mencionado, la medición de la eficiencia en instituciones de educación superior es esencial para la formulación de planes que busquen optimizar la utilización de recursos. Esto puede lograr el objetivo que tanto persigue la sociedad: la calidad y la cobertura de la educación. La educación se puede observar como cualquier otro sistema productivo al implementar sistemas de medición y control

(por ejemplo, indicadores cuantitativos) sobre los procesos que se realizan a diario en las IES, esto se convierte en una herramienta que permite generar oportunidades de mejora y la toma de decisiones [3].

Es por todo lo anterior que en este artículo se hace uso de las herramientas análisis de componentes principales (PCA, por sus siglas en inglés) y análisis envolvente de datos (DEA, por sus siglas en inglés) con el objetivo de proponer una metodología para el análisis, la medición y la clasificación de la eficiencia académica de los estudiantes. En este propósito se utilizan los resultados de las pruebas Saber PRO. Para que los resultados sean certeros se debe realizar una adecuada identificación de las unidades del modelo (entradas y salidas), con lo cual es posible asegurar una buena clasificación entre unidades eficientes e ineficientes, además de realizar buenas inferencias.

2. REVISIÓN DE LITERATURA

Para aplicar correctamente la herramienta DEA, primero, se realiza un análisis exploratorio de datos utilizando el análisis de componentes principales (PCA, por sus siglas en inglés).

El PCA es un método que propone Harold Hotelling a principios del siglo XX, basado en los trabajos de Karl Pearson y las investigaciones sobre ajustes ortogonales por mínimos cuadrados [5] desde hace mucho tiempo se han desarrollado las técnicas multivariadas, pero sólo con la evolución de los computadores y diversos paquetes de software que procesan amplios conjuntos de datos ha llegado a ser notoria la potencia de la estadística multivariada. Se ha tomado como pretexto el problema del desplazamiento de personas del campo a la ciudad para consolidar el concepto y la aplicación de la técnica de componentes principales (ACP. Teniendo en cuenta las diversas definiciones que son utilizadas en la literatura, el PCA se puede definir como una técnica de análisis exploratorio de datos que, mediante el uso del análisis multivariado, el álgebra y la estadística, busca, a partir de una matriz de datos inicial, generar una nueva matriz de datos final con información resumida y estructurada [6]. El método consiste en seleccionar la información contenida en un conjunto de n variables de interés en m nuevas variables independientes; mediante la combinación lineal de las variables originales se otorga la posibilidad de resumir la información, es decir, absorber la mayor variabilidad de la información (variables) en componentes principales (combinación lineal de variables) que reducen la dimensión de la matriz. Sin embargo, a fin de que el método se aplique correctamente las variables introducidas no deben tener correlación entre ellas (escalar variables) y deben tener su variabilidad inicial completa (no tener modificaciones adicionales diferentes al primer inciso) [7].

Por otra parte, una duda de muchos investigadores es la selección de los componentes principales, para lo cual existen dos criterios de selección [8] from the analysis of meteorological fields to the evaluation of air quality monitoring networks (AQMN). El primer criterio, llamado “Kaiser”, considera conservar los componentes principales con autovalores mayores a uno, aunque muchas veces este criterio puede ser muy restrictivo [9]. El segundo criterio se denomina “porcentaje de varianza acumulada (PVA)” y

busca retener las componentes principales cuyos porcentajes de varianza acumulada superen un valor determinado. En este último criterio el porcentaje dependerá del tipo de problema y el conocimiento que tenga el investigador sobre el tema. También algunos autores recomiendan establecer un valor de varianza acumulada en 90 % [10].

La segunda herramienta usada en este artículo es el análisis envolvente de datos (DEA). La herramienta DEA es una metodología no paramétrica basada en modelos de programación lineal propuesta en 1978 por Charnes, Cooper y Rhodes con el fin de estudiar la eficiencia relativa de una serie de unidades de decisión en la que existen múltiples entradas y salidas [11].

El problema que seleccionó Rhodes para usar DEA fue titulado como “Data envelopment analysis and approaches for measuring the efficiency of decision-education” [12]. Rhodes, con este problema, buscaba determinar la eficiencia relativa de varias instituciones de educación mediante el uso de diversas variables de entrada y salida. Esta primera respuesta se presenta como la primera formulación matemática de DEA, en su variante CCR [13].

DEA surge como una técnica para evaluar la eficiencia de un conjunto observado, denominado como unidades de toma de decisión (DMU corresponde a *decision making unit*). El modelo DEA relaciona, para cada DMU, una serie de entradas y salidas [14]. La premisa principal en la que se desarrolla DEA es que la eficiencia de una DMU se define como la habilidad en que esta DMU, a partir de una serie de entradas, puede transformarlas en salidas deseadas. Cabe resaltar que todas las entradas y salidas del modelo deben estar en una escala y unidad comparables [15] en este trabajo pretendemos dotar a los directivos de una herramienta que les permita establecer cuándo el tamaño de los equipos no es el adecuado, en qué casos no utilizan los recursos de una forma eficiente y cuáles son las recomendaciones sobre utilización de los recursos para mejorar la eficiencia de la organización. Los análisis se realizan utilizando dos modelos. En el primer modelo se analiza la eficiencia del conjunto de los equipos, mientras que en el segundo se separan los equipos rurales de los urbanos para que las organizaciones comparadas sean lo más homogéneas posibles y la información sea de más utilidad. Data Envelopment Analysis (DEA).

A fin de establecer el modelo usado en la metodología DEA se debe seguir los siguientes pasos [12]: 1) se debe estudiar nuestro problema y conocer su naturaleza; 2) identificar correctamente las entradas y salidas del modelo; 3) seleccionar el tipo de modelo DEA a implementar (CCR que alude a la eficiencia técnica o BCC que mide a la eficiencia global), el retorno de escala (CRS-Constante o VRS-Variable) y la orientación del modelo (entrada o salida). Este último paso depende del tipo de modelo seleccionado.

Una vez establecido el modelo se procede a resolver el modelo como un problema de programación lineal, en el cual la función objetivo debe maximizar el cociente de las salidas entre sus entradas. Además, como restricciones se tiene que la eficiencia no debe ser mayor a uno [16]. De acuerdo con lo anterior y con el fin de ilustrar los modelos CCR y BCC, se presentan la Figura 1 correspondiente.

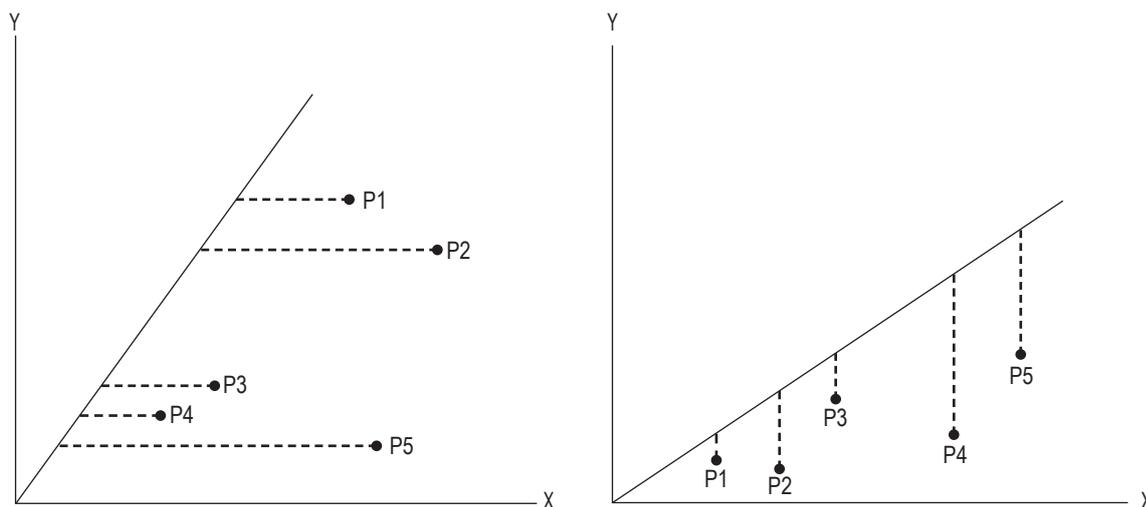


Figura 1. Comparación de los modelos CCR y BCC

Fuente: elaboración propia.

En la Figura 1 se muestra la comparación entre los modelos CCR y BCC. En este ejemplo la solución que ofrecen los dos modelos es la misma debido a que coinciden en la zona de frontera. La principal diferencia es que el modelo CCR busca calcular la eficiencia global, mientras el modelo BBC hace uso de la eficiencia técnica.

3. METODOLOGÍA

El conjunto de datos usado en este artículo pertenece a los resultados de las pruebas saber PRO del 2018. Estos datos los comparte al público investigador el Gobierno de Colombia en la página del Instituto Colombiano para la Evaluación de la Educación (Icfes).

El conjunto de datos está compuesto por 106 columnas por observación (estudiante) que contienen la información personal y el resultado de las pruebas por estudiante. Para realizar este artículo se seleccionaron nueve variables que corresponden al resultado de las pruebas específicas, al resultado de las pruebas genéricas y a la modalidad de financiación de la carrera universitaria. Por otro lado, se seleccionan los resultados del 2018 de la carrera Ingeniería Industrial de la Universidad Tecnológica de Bolívar: en total son 114 observaciones [17].

El primer conjunto de variables que corresponde a las pruebas específicas de Ingeniería Industrial cuenta con tres áreas: formulación de proyectos de ingeniería (FPI), pensamiento matemático y estadístico (PCME), y diseño de procesos productivos y logísticos (DPL). El segundo conjunto de variables, que corresponde a las pruebas genéricas, cuenta con cinco áreas: comunicación escrita (CE), Inglés (IN), lectura crítica (LC), razonamiento cuantitativo (RC) y competencias ciudadanas (CC). La última variable

es la modalidad de financiación que tiene tres niveles: becado tiene un valor de cero (0), crédito tiene un valor asignado de uno (1) y cuenta propia que tiene un valor de dos (2).

A fin de realizar la medición de la eficiencia se sigue la metodología planteada en la Figura 2. Primero se realiza un análisis exploratorio de datos usando la herramienta PCA y luego se aplica la metodología DEA en la medición de la eficiencia de los resultados de las pruebas. Por otra parte, para la aplicación de las técnicas mostradas en la metodología se usa el *software* libre llamado R versión 3.6.1.

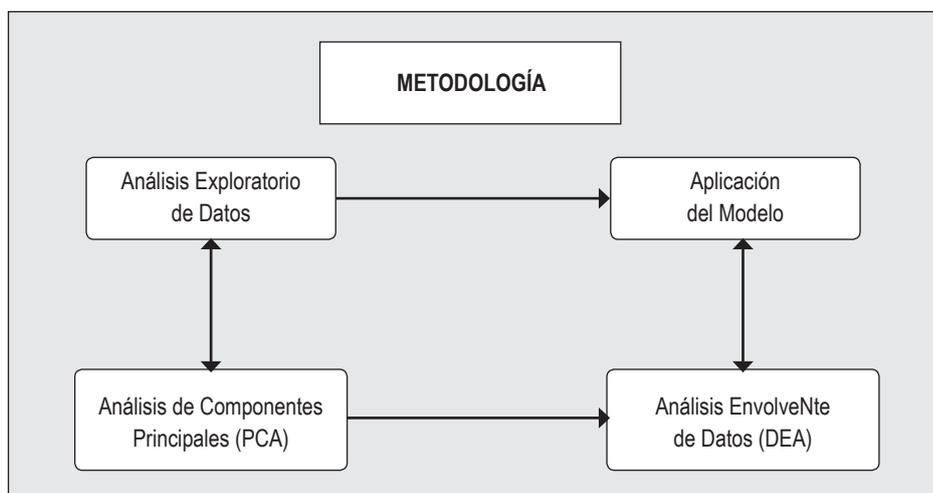


Figura 2. Metodología implementada

Fuente: elaboración propia.

4. RESULTADOS

Como se mencionó, primero se realiza el análisis exploratorio de los datos con la herramienta PCA. En la Tabla 1 se muestran los resultados obtenidos al aplicar el análisis de componentes principales.

Tabla 1. Componentes principales

Componente	Autovalor	Porcentaje de varianza	Porcentaje de varianza acumulado
1	3,80	47,53	47,53
2	1,05	13,23	60,77
3	0,69	8,69	69,47
4	0,62	7,77	77,24
5	0,58	7,35	84,60
6	0,47	5,90	90,50
7	0,41	5,22	95,73
9	0,34	4,26	100,00

Fuente: elaboración propia.

La primera columna corresponde a las componentes (se tiene tantas componentes como variables), la segunda columna corresponde a los autovalores de cada componente, la tercera el porcentaje de la varianza de cada componente y la última corresponde al porcentaje de la varianza acumulada; la variable de modalidad de pago se tiene en cuenta para clasificar las observaciones, es decir, de carácter informativo.

Con el objetivo de complementar la Tabla 1 se presenta la gráfica del método Elbow (véase la Figura 3). Esta información se utiliza para seleccionar las componentes principales que explican la mayor variabilidad de la información.

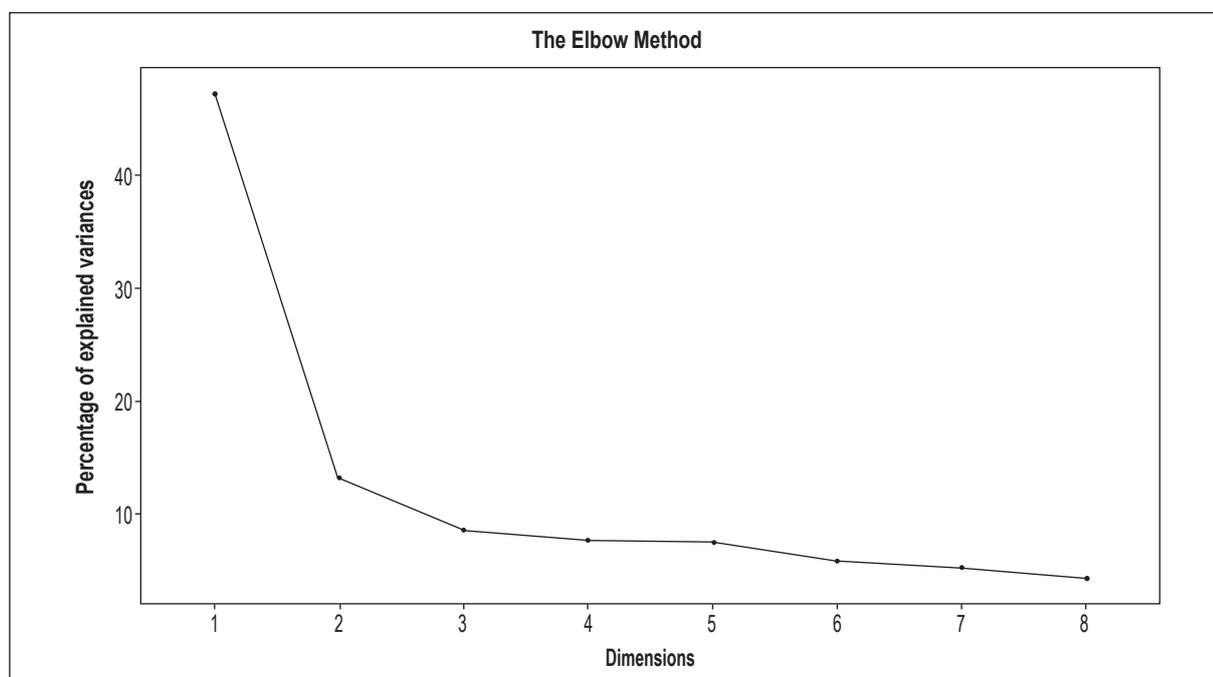


Figura 3. Método Elbow

Fuente: elaboración propia.

Como se mencionó en el marco teórico, a fin de seleccionar el número de componentes principales se utilizan las técnicas Kaiser y Varianza acumulada; para efectos de este análisis adicionamos la gráfica Elbow que consiste en elegir el número de componentes cuando la gráfica toma la forma de un codo.

Con el criterio Kaiser las componentes seleccionadas son las componentes 1 y 2, debido a que sus autovalores tienen un valor mayor a uno. El criterio de varianza acumulada se fija con un valor de 70 % [9], entonces, para este criterio las componentes son 1, 2, 3 y 4. En el método Elbow las componentes seleccionadas son las componentes 1, 2 y 3. Para seleccionar el número de componentes principales en este artículo se selecciona el criterio de varianza acumulada y se eligen las componentes que explican, al menos, el 70 % de la información (en este caso son las componentes 1, 2, 3 y 4, las cuales explican el 77,24 % de la información).

Para entender mejor los resultados que brinda el PCA usaremos el gráfico Biplot (Figura 4) que es la representación simultánea entre la nube de puntos formados por las observaciones, las variables estudiadas y el plano formado por las dos primeras dimensiones.

Las dimensiones que conforman el plano del gráfico Biplot se muestran en la Tabla 2. Esta tabla muestra la representación del resultado de la prueba sobre alguna dimensión valorada por un puntaje entre -1 y 1, donde -1 es una representación negativa, 0 es pobremente representado y 1 es una representación positiva. Por otro lado, se gráfica las dos primeras dimensiones debido a que tienen los valores de representación más altos.

En este artículo se consideran como nombre de las dimensiones 1 y 2, Prom_Global_Prueba y Compe_Comun, respectivamente.

Tabla 2. Dimensiones

Prueba	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7	Dim.8
RC	0,730	-0,295	-0,112	0,229	0,442	-0,005	0,118	-0,325
LC	0,711	0,142	-0,387	-0,359	0,171	0,372	0,026	0,168
CC	0,743	0,145	-0,140	-0,327	-0,281	-0,371	0,254	-0,137
IN	0,654	0,310	-0,379	0,502	-0,146	-0,126	-0,158	0,138
CE	0,358	0,795	0,424	0,056	0,212	0,040	0,105	-0,012
DPL	0,736	-0,320	0,263	-0,108	0,256	-0,296	-0,141	0,319
FPI	0,787	-0,027	0,215	-0,124	-0,234	0,145	-0,433	-0,232
PCME	0,704	-0,316	0,275	0,227	-0,312	0,266	0,308	0,118

Fuente: elaboración propia.

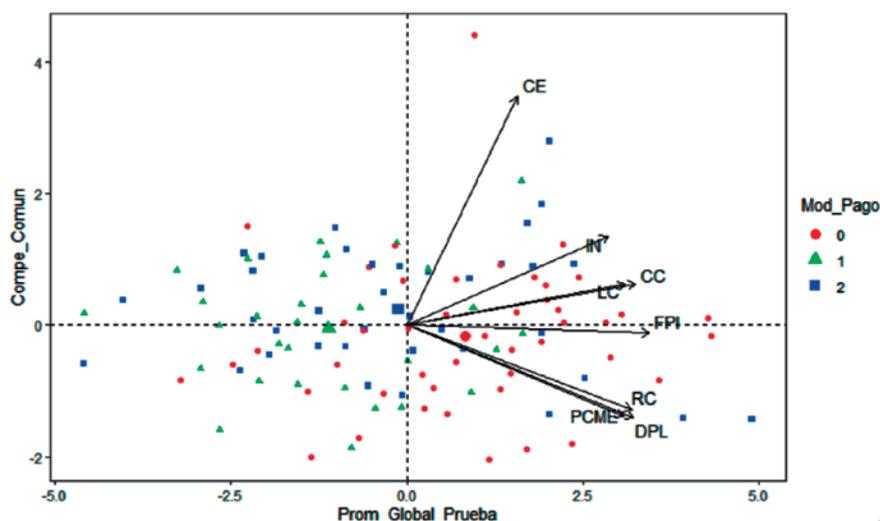


Figura 4. PCA

Fuente: elaboración propia.

Cabe resaltar que esto es un análisis de carácter comparativo entre las observaciones, no evaluativo. Las observaciones pueden ser buenas o malas pero este análisis determina únicamente las diferencias entre las observaciones. Por otra parte, en el gráfico Biplot no se busca proximidad entre observación y variables, lo importante son las direcciones.

En el primer cuadrante tenemos que , luego se caracteriza por tener buen puntaje en CE, IN, CC, LC, FPI, RC, PCME y DPL; además, , por tanto, se caracteriza por tener buen puntaje en CE, IN, CC y LC.

De la información de las variables y el gráfico Biplot se puede decir que los mejores puntajes los tienen los estudiantes becados y quienes pagan por su cuenta; sin embargo, los becados tienen una mayor concentración del lado derecho del gráfico que es la dirección con la que crecen las variables. En la Tabla 3 se presenta el conteo de los niveles de la variable modalidad de pago por cuadrante, y son los cuadrantes 1 y 4 aquellos que tienen la dirección positiva de las variables.

Tabla 3. Conteo de observaciones por cuadrante

Tipo de pago	Total			
	Cuadrante 1	Cuadrante 2	Cuadrante 3	Cuadrante 4
0	15	5	9	17
1	3	12	11	4
2	9	12	9	8

Fuente: elaboración propia.

El conteo confirma lo observado en la gráfica Biplot: los estudiantes becados (0) son los que poseen mejores resultados en las pruebas.

Realizado el análisis y entendido los datos, se procede a realizar la aplicación de la herramienta DEA. Para la aplicación del modelo DEA se usa el modelo CCR orientado a las entradas. De acuerdo con esto tenemos lo siguiente:

$$\max_{u, v} h_o = \frac{\sum_{r=1}^s u_r y_{ro}}{\sum_{i=1}^m v_i x_{io}} \quad (1)$$

Sujeto a:

$$\frac{\sum_{r=1}^s u_r y_{rj}}{\sum_{i=1}^m v_i x_{ij}} \leq 1, j = 1, \dots, n \quad (2)$$

Donde,

n : cantidad de DMU del conjunto de datos.

x_{ij} : cantidad de entrada i consumido por la j -ésima DMU.

- x_{io} : cantidad de entrada i consumida por la DMU o .
 y_{rj} : cantidad observada de salidas r producida por la j -ésima DMU.
 y_{ro} : cantidad observada de la salida r producida por la DMU o .
 u_r : ponderación asignada por el modelo de la salida virtual r .
 v_i : ponderación asignada por el modelo de la entrada virtual i .

La ecuación (1) corresponde a la función objetivo del modelo de optimización de la herramienta DEA y la ecuación (2) a la restricción de la eficiencia.

Las entradas del modelo son los resultados de las pruebas genéricas y las salidas corresponden a los resultados de las pruebas específicas. De acuerdo con lo anterior se obtienen los resultados que se expresan la Figura 5.

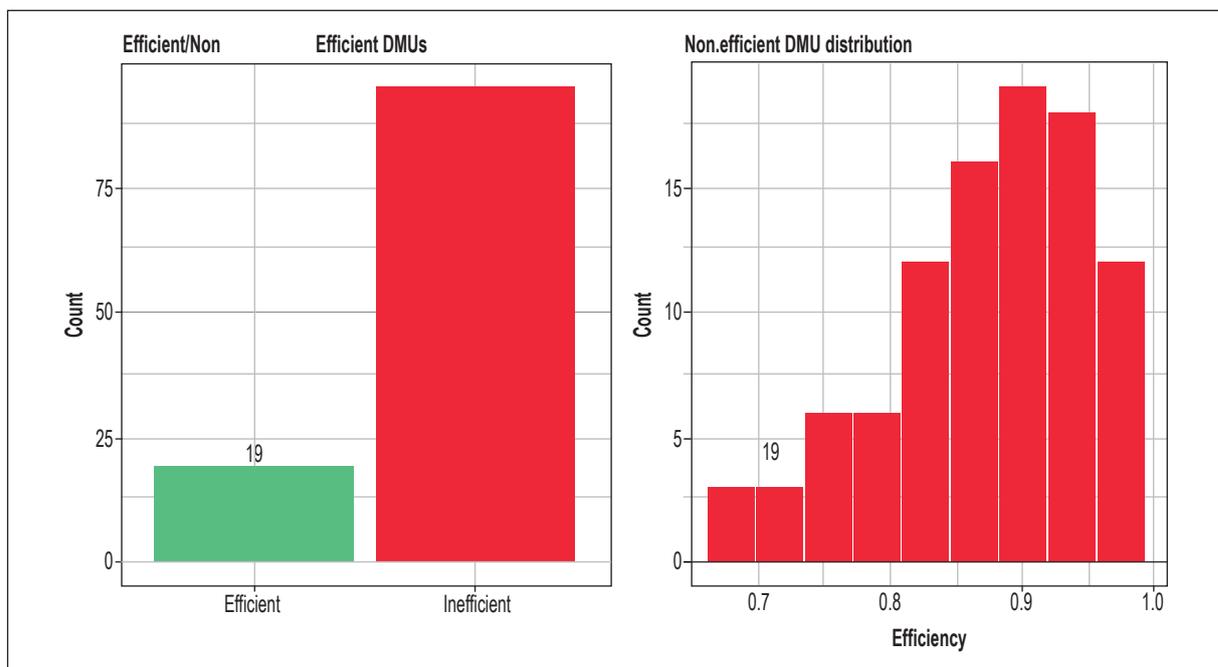


Figura 5. Unidades eficientes e ineficientes

Fuente: elaboración propia.

Se observa en los gráficos de la Figura 5 que de 114 *decision making units* (DMU) solo 19 son eficientes y 95 DMU son ineficientes. A modo de ejemplo se presentan algunas DMU eficientes y otras ineficientes.

Tabla 4. DMU eficientes y DMU ineficientes

DMU	Eficiencia	Tipo Pago	RC	LC	CC	IN	CE	DPL	FPI	PCME
U30	1	0	160	185	97	177	172	150	183	155
U34	1	0	124	129	116	189	111	138	140	119
U44	1	0	162	137	87	130	122	152	162	155
U101	0,99	2	185	161	174	186	130	185	191	160
U10	0,98	0	203	196	182	141	126	182	163	155
U68	0,98	2	181	160	155	161	179	126	194	117

Fuente: elaboración propia.

Ahora se procede a calcular el conjunto de referencia. Algunas DMU son más importantes en el momento de evaluar DMU ineficientes. Esto se evidencia en la Tabla 5; por ejemplo, las DMU de esta tabla, a excepción de la DMU U73, se usan como referencias varias veces, mientras que la U73 no se usa como referencia de alguna otra DMU, aunque es declarada como eficiente. Esto es un criterio para la clasificación de DMU eficientes y brinda una mejor posición a aquellas DMU que aparecen más veces en la evaluación de las demás.

Como se mencionó, el número de veces que una DMU eficiente aparece como referencia de una DMU ineficiente puede servir como herramienta de clasificación. En la Tabla 5 se presenta la clasificación por grupo de las DMU eficientes, en la que el grupo 1 es el que tiene las DMU más eficientes y, a medida que sube el número del grupo, aumenta y será menos eficiente que las primeras.

Tabla 5. Clasificación de las DMU

DMU	Número de veces que aparece como referencia	Grupo de clasificación
U44	63	1
U88	45	2
U2	40	3
U86	36	4
U107	34	5
U79	34	5
U60	31	6
U93	31	6
U28	19	7
U67	14	8
U75	13	9
U76	11	10
U51	9	11

DMU	Número de veces que aparece como referencia	Grupo de clasificación
U98	9	11
U87	8	12
U30	6	13
U34	6	13
U65	5	14
U73	0	15

Fuente: elaboración propia.

Luego del análisis de los resultados del modelo DEA se procede a observar el comportamiento de las DMU ineficientes. A manera de ejemplo se presenta en la Tabla 6 los resultados de las pruebas de las DMU y en la Tabla 7 los objetivos que deben alcanzar para ser eficientes.

Tabla 6. Puntaje de las DMU

DMU	Eficiencia	Tipo Pago	RC	LC	CC	IN	CE	DPL	FPI	PCME
U3	0,97	2	158	122	92	161	129	154	136	101
U5	0,86	0	186	186	180	201	183	178	185	139

Fuente: elaboración propia.

Tabla 7. Holgura de las variables

DMU	S-RC	S-LC	S-CC	S-IN	S-CE	S-DPL	S-FPI	S-PCME
U3	0	0	0	23,83	0	0	15,04	36,17
U5	0	0	0	0	0	0	0	14,38

Fuente: elaboración propia.

La DMU U3 tiene una eficiencia del 97 %, aproximadamente. De acuerdo con la tabla de holgura (Slack-S), debería enfocarse en estudiar FPI y PCME, y de acuerdo con la interpretación de las holguras, es lo que puede disminuir (entradas) o aumentar (salidas) para llegar a la eficiencia. En este contexto, las DMU U3 debería “disminuir” su puntaje de inglés en 23,83 puntos, lo cual indica que tiene un puntaje superior a lo que necesita para ser eficiente (es decir, tiene buen puntaje en las entradas o competencias específicas). No obstante, sus competencias específicas FPI y PCME tienen un puntaje bajo y deben aumentar 15,04 y 36,17 puntos, respectivamente, para llegar a ser una DMU eficiente. En el caso de la DMU 5 tiene una eficiencia de 86 % y las holguras de las entradas están en cero, lo cual indica que tiene un puntaje en las competencias genéricas bueno, sin embargo, la competencia específica PCME debe reforzarse para subir 14,38 puntos y ser una DMU eficiente.

5. CONCLUSIÓN

En este artículo se presenta la herramienta DEA como una herramienta para la medición de la eficiencia del resultado de una prueba académica, así como con miras a la clasificación de observaciones (estudiantes) eficientes [18]. También se presenta la técnica PCA como herramienta para el análisis exploratorio de datos y determinar las componentes principales que reducen la dimensión del problema y agilizan el modelo. Estas dos herramientas generan una metodología robusta y sencilla que permite analizar, clasificar e interpretar los resultados de cualquier prueba académica, en este caso la prueba de estados Saber PRO del 2018. De acuerdo con los resultados obtenidos en este estudio se concluyen los siguientes puntos.

1. La utilización del DEA como herramienta para la medición de la eficiencia de los estudiantes de la universidad ha permitido establecer una hipótesis según la cual la mayor cantidad de estudiantes con puntajes eficientes son becados, le siguen las personas que pagan por cuenta propia y, finalmente, las personas con crédito.
2. Con respecto a las holguras en las entradas, se debe a que el estudiante tiene buenas bases en las competencias genéricas, sin embargo, aplicarlas en su contexto profesional se les dificulta (lo que se evidencia en la Tabla 7), por lo que hay estudiantes que necesitan aumentar solo los puntajes en sus competencias específicas.
3. De acuerdo con el PCA, los estudiantes deben enfocar sus esfuerzos en fortalecer las competencias de comunicación escrita (CE) y pensamiento científico, matemático y estadístico (PCME).

En cuanto a las oportunidades para futuros trabajos e investigaciones, DEA se puede usar en la clasificación de grupos eficientes y como indicador de gestión educativa. Esto puede ser útil en la predicción del puntaje de alguna prueba que tenga como entrada los grupos de clasificación existentes. Además, finalmente, la comprobación de la hipótesis planteada en el inciso 1 de la conclusión.

REFERENCIAS

- [1] Mineducación, “Art 1, Ley 115 de 1994”, 1994. https://leyes.co/se_expide_la_ley_general_de_educacion/1o.htm (consultado sep. 11, 2019).
- [2] Congreso de Colombia, “Ley 30 de 1992”, 1992. http://www.legal.unal.edu.co/rlunal/home/doc.jsp?d_i=34632 (consultado sep. 11, 2019).
- [3] O. Buitrago, A. Espitia, y L. Molano, “Análisis envolvente de datos para la medición de la eficiencia en instituciones de educación superior: una revisión del estado del arte”, *Rev. Científica Gen. José María Córdova*, vol. 15, n.º 19, pp. 147-173, jun. 2017, doi: 10.21830/19006586.84.
- [4] J. Vela Valdés, “Educación superior: inversión para el futuro*”, *Educ. Médica Super.*, vol. 14, n.º 2, pp. 171-183, ago. 2000.

- [5] Á. León, H. Llinás y J. Tilano, "Análisis multivariado aplicando componentes principales al caso de los desplazados", *Rev. Científica Ing. Desarro.*, vol. 23, n.º 23, pp. 119-142-142, jul. 2011.
- [6] D. Peña, "Análisis de datos multivariantes", p. 515, 2002.
- [7] C. Lozares y P. López, "El análisis de componentes principales: aplicación al análisis de datos secundarios", *Pap. Rev. Sociol.*, vol. 37, n.º 0, pp. 31-63, jul. 1991, doi: 10.5565/rev/papers/v37n0.1595.
- [8] J. M. Polanco Martínez, "El papel del análisis por componentes principales en la evaluación de redes de control de la calidad del aire", *Comun. En Estad.*, vol. 9, n.º 2, p. 271, sep. 2016, doi: 10.15332/s2027-3355.2016.0002.06.
- [9] E. R. Pérez y L. A. Medrano, "Análisis factorial exploratorio: Bases conceptuales y metodológicas", *Rev. Argent. Cienc. Comport. RACC*, vol. 2, n.º 1, pp. 58-66, 2010.
- [10] J. C. M. Pires, M. C. Pereira, M. C. M. Alvim-Ferraz, y F. G. Martins, "Identification of redundant air quality measurements through the use of principal component analysis", *Atmos. Environ.*, vol. 43, n.º 25, pp. 3837-3842, ago. 2009, doi: 10.1016/j.atmosenv.2009.05.013.
- [11] J. Pino, F. Solís, M. Delgado, y R. el C. Barea, "Evaluación de la eficiencia de grupos de investigación mediante análisis envolvente de datos (DEA)", 2010, doi: 10.3145/epi.2010.mar.06
- [12] J. Perdomo y J. Mendieta, "Factores que afectan la eficiencia técnica y asignativa en el sector cafetero colombiano: una aplicación con análisis envolvente de datos*", *Rev. Desarro. Soc.*, jun. 2017, doi: 10.13043/dys.60.1.
- [13] J. Ferrera, F. Pedraja-Chaparro, y J. Salinas-Jimenez, "Measuring Efficiency in Education: An Analysis of Different Approaches for Incorporating Non-discretionary Inputs", *Appl. Econ.*, vol. 40, pp. 1323-1339, feb. 2008, doi: 10.1080/00036840600771346.
- [14] J. Ligarda y M. Ñaccha, "La eficiencia de las organizaciones de salud a través del análisis envolvente de datos: Microrredes de la Dirección de Salud IV Lima Este 2003", *An. Fac. Med.*, vol. 67, n.º 2, pp. 142-151, jun. 2006.
- [15] S. G. Legaz, "El análisis envolvente de datos como sistema de evaluación de la eficiencia técnica de las organizaciones del sector público: aplicación en los equipos de atención primaria", *Rev. Esp. Financ. Contab.*, vol. 27, n.º 97, pp. 979-1004, 1998.
- [16] T. Fontalvo, E. De La Hoz, E. De La Hoz, T. Fontalvo, E. De La Hoz, y E. De La Hoz, "Método Análisis Envolvente de Datos y Redes Neuronales en la Evaluación y Predicción de la Eficiencia Técnica de Pequeñas Empresas Exportadoras", *Inf. Tecnológica*, vol. 29, n.º 6, pp. 267-276, dic. 2018, doi: 10.4067/S0718-07642018000600267
- [17] E. Delahoz-Dominguez, R. Zuluaga y T. Fontalvo-Herrera, "Dataset of academic performance evolution for engineering students", *Data Brief*, vol. 30, p. 105537, jun. 2020, doi: 10.1016/j.dib.2020.105537.
- [18] J. A. Calderón Velasco, G. A. Amarillo Cárdenas, L. A. Silva Bahamon, y C. G. Donoso Albarracín, "Biometría dactilar: una nueva alternativa de controlar efectivamente la asistencia a clases", *Inv. Inn. Ing.*, vol. 6, n.º 1, pp. 27-39, ene. 2018.