



ESCENARIOS

Boletín de Prospectiva

Número 25

Julio, 2024

OFICINA DE PLANEACIÓN

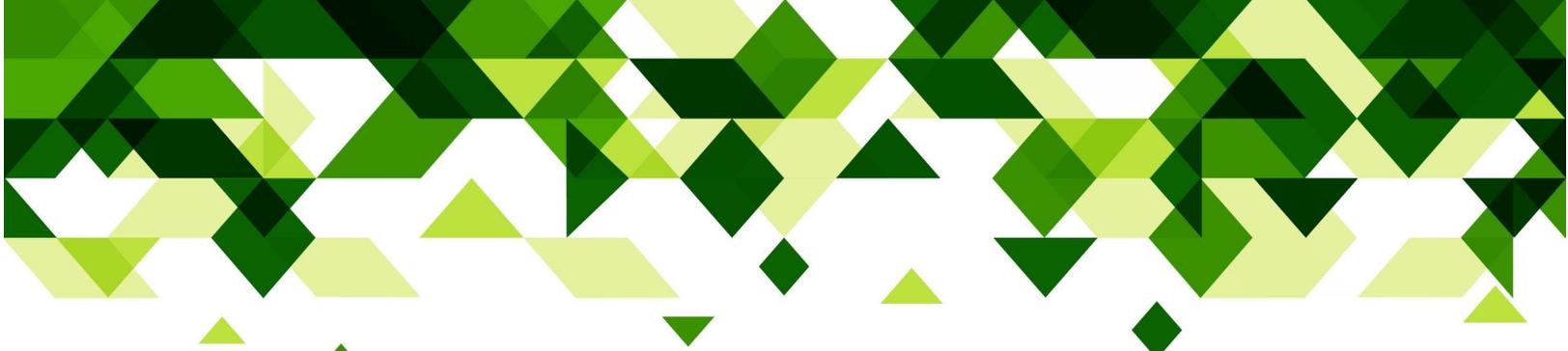
DETERMINANTES DE LA DEMANDA DE EDUCACIÓN SUPERIOR: UN ANÁLISIS ESTADÍSTICO

Realizar un análisis estadístico multivariado de los determinantes de la demanda de educación superior puede ser un proceso iterativo (práctica y error) que recorre diversos caminos en busca de identificar patrones y tendencias. En esta ocasión, se presentan los resultados de un ejercicio estadístico preliminar utilizando el análisis de componentes principales (ACP).

|| NÚMERO 25-2024



UNIVERSIDAD
La Gran Colombia



El supuesto utilitarista plantea que los estudiantes actúan como sujetos racionales, utilizando la educación como un medio para acceder a mejores ingresos y una posición social más prestigiosa. Aunque el supuesto utilitarista tiene sus limitaciones en términos de medición y no se ignora la importancia de las circunstancias específicas y culturales que rodean al actor, el universo de expectativas sigue siendo un elemento crucial en estudios de este tipo.

Estas expectativas son, en ocasiones, determinadas por condiciones del entorno, especialmente aquellas que repercuten en la calidad de vida de las personas. En este sentido, el análisis de variables globales desde una perspectiva macroeconómica permite aproximarse a comprender los determinantes generales de la demanda de educación superior en el país. Asimismo, el uso de herramientas estadísticas es fundamental para reducir la dimensionalidad de los datos e identificar posibles asociaciones y tendencias relevantes.

En los números [23](#) y [24](#) de **ESCENARIOS**, se realizó una aproximación general a los determinantes de la demanda en educación superior utilizando inteligencia artificial generativa y revisión de literatura. El primero

analizó elementos como la cobertura y la calidad, brindando un contexto para la identificación de posibles factores de interés en un diálogo con Chat GPT. El segundo permitió dar solidez a algunas de las ideas planteadas en el número anterior por la inteligencia artificial y contrastar factores analizados por otros autores en el pasado.

El objetivo de este número de **ESCENARIOS** es **explorar algunos de los determinantes de la demanda de educación superior mediante la aplicación del análisis de componentes principales (ACP)**. Para ello, se realizó una aproximación conceptual a la técnica y, posteriormente, se presentaron los resultados obtenidos en el ACP a partir de algunas de las variables más importantes y cuya información fue de fácil acceso. Finalmente, se presentan algunas consideraciones finales.

APROXIMACIÓN CONCEPTUAL AL ANÁLISIS DE COMPONENTES PRINCIPALES

Una de las técnicas que se pueden usar para estructurar un conjunto de datos multivariados, sin que el resultado sea un modelo estadístico, es el análisis de componentes principales (ACP). En términos

generales, el ACP permite reducir la dimensionalidad y realizar un análisis exploratorio de los datos sin que previamente estos tengan que cumplir con todos los supuestos que exige otro tipo de técnica, como la regresión. Asimismo, se usa en casos donde las variables son numéricas¹.

De acuerdo con Laura Pla, los objetivos más relevantes del ACP son:

Generar nuevas variables que puedan expresar la información contenida en el conjunto original de datos.

Reducir la dimensionalidad del problema que se está estudiando, como paso previo para futuros análisis.

Eliminar, cuando sea posible, algunas de las variables originales si ellas aportan poca información. (1986, p. 15)

Las nuevas variables generadas, que son una integración de los datos analizados, se denominan **componentes principales**.

En este análisis, ante la sospecha de colinealidad y dependencia entre las variables resultado de un análisis de regresión realizado previamente, se privilegió el uso del

ACP para estudiar los datos y tomar decisiones sobre la inclusión, cambio o eliminación de las variables en futuros modelos. Las ventajas del uso de esta técnica multivariada son los siguientes:

Los componentes principales no están correlacionados y tienden a ser independientes.

Cada componente principal sintetiza la máxima variabilidad residual contenida en los datos (Pla, 1986, p. 17)

A continuación, se describen y presentan los datos usados, así como sus fuentes.

DATOS Y FUENTES

Para el estudio de los determinantes de la demanda de educación superior en Colombia, **se tomó como variable dependiente el número de matrículas de primer curso** registradas en el Sistema Nacional de Información de la Educación Superior SNIES. Como variables independientes se tuvieron en cuenta:

¹ Para variables categóricas existe el análisis de correspondencias (AC), el cual fue aplicado en [Escenarios 15](#).

- El **Índice de Costos de la Educación Superior** (ICES). Este dato es consolidado semestralmente por el DANE. Para efectos del análisis, se promedió el resultado del primer y segundo semestre de cada año. Su unidad de medida fue ‘porcentaje’.
- El **Coefficiente de Gini** (GINI), el cual mide la desigualdad de ingresos en los países. Su unidad de medida fue ‘porcentaje’ y se tomó del Banco Mundial.
- El **número de personas entre 17 y 21 años** en Colombia por año se tomó de las proyecciones y retroproyecciones del DANE de acuerdo con el Censo Nacional de Población y Vivienda 2018.
- El **Índice de Friz** (FRIZ) que representa la proporción de la población en el grupo 0 a 19 años en relación con la de 30 a 49 años. Si este índice es mayor a 160 la población se considera jóvenes, si se halla entre 60 y 160, madura, y si es menor de 60, vieja. Este índice fue calculado con los datos de las proyecciones y retroproyecciones del DANE.
- El **número de programas activos** de pregrado y posgrado en el país por año, el cual fue consultado al Ministerio de Educación Nacional.

En la tabla 1 se presentan los datos.

Tabla 1. Datos usados para el análisis de determinantes de la demanda en educación superior

Año	Primer curso	ICES	GINI	17 A 21	FRIZ	Número de programas
2010	471 308	89,235	54,6	4 031 864	138,562813	9 716
2011	647 731	92,54	53,5	4 083 849	136,411185	10 256
2012	695 488	96,17	52,6	4 125 876	134,248215	10 423
2013	760 316	99,73	52,6	4 156 312	132,121119	10 699
2014	782 781	103,515	52,6	4 175 772	130,024329	11 000
2015	791 736	108,19	51	4 185 606	127,943687	11 448
2016	952 988	116,135	50,6	4 190 194	125,890961	11 891
2017	912 468	122,205	49,7	4 189 178	123,858465	11 878
2018	813 217	129,145	50,4	4 201 009	121,855438	12 302
2019	824 840	133,595	51,3	4 228 704	119,746023	16 893
2020	855 706	140,295	53,5	4 226 577	116,075514	16 981
2021	911 131	143,57	55,1	4 198 289	113,172644	14 785
2022	901 664	155,525	54,8	4 159 592	110,249169	15 043

Fuente: elaboración propia a partir de datos del Ministerio de Educación Nacional, Departamento Administrativo Nacional de Estadística y el Banco Mundial.

RESULTADOS DEL ANÁLISIS DE COMPONENTES PRINCIPALES

La matriz de correlaciones bivariadas (tabla 2) muestra que existe una fuerte correlación positiva entre la variable primer curso, el Índice de Costos de la Educación Superior y la población entre 17 y 21 años. Esto podría indicar que las tres variables cambian en el mismo sentido. Asimismo, la variable dependiente tiene una fuerte correlación negativa con el Índice de Friz y una moderada correlación de esta naturaleza con el índice de Gini. En este sentido, estos datos indican que estas últimas cambian de manera inversa a la variable dependiente.

Tabla 2. Correlaciones

	PC	ICES	GINI	17A21	FRIZ
PC	-	,748	-,333	,853	-,782
ICES	,748	-	,105	,691	-,994
GINI	-,333	,105	-	-,439	-,113
17A21	,853	,691	-,439	-	-,723
FRIZ	-,782	-,994	-,113	-,723	-
NP	,580	,876	,136	,704	-,866

Determinante: 1,156E-5

Fuente: elaboración propia en SPSS

El ACP arrojó que, con un plano factorial de dos dimensiones, es posible explicar el 91,94 % de la variabilidad de los datos (tabla 3). En estas dos dimensiones, que corresponden cada uno de los componentes

principales, se explican las variables en más del 80 %. La variable primer curso se explica en un 86,1 %, el ICES en un 95,7 %, el GINI en un 94,8 %, la población de 17 a 21 años en un 91,9 %, el Índice Friz en un 97,3 % y el número de programas en un 85,8 %. Esto muestra que las variables analizadas están altamente correlacionadas.

Tabla 3. Varianza total explicada

Componente	Total	Autovalores iniciales	
		% de varianza	% acumulado
1	4,144	69,068	69,068
2	1,372	22,874	91,942
3	,287	4,788	96,731
4	,148	2,471	99,202
5	,047	,781	99,983
6	,001	,017	100

Fuente: elaboración propia en SPSS

Los componentes principales generados (tabla 4) son el resultado de las matrices de componentes y de los coeficientes de puntuación de componentes, los cuales están representados en la figura 1. Las figuras generadas a partir de estos componentes permiten interpretar y comprender mejor los datos utilizados. Estos resultados no solo permiten obtener dos nuevas variables que integran las anteriores, sino que el nivel de explicación de cada variable en el análisis permite tomar decisiones sobre qué variables excluir.

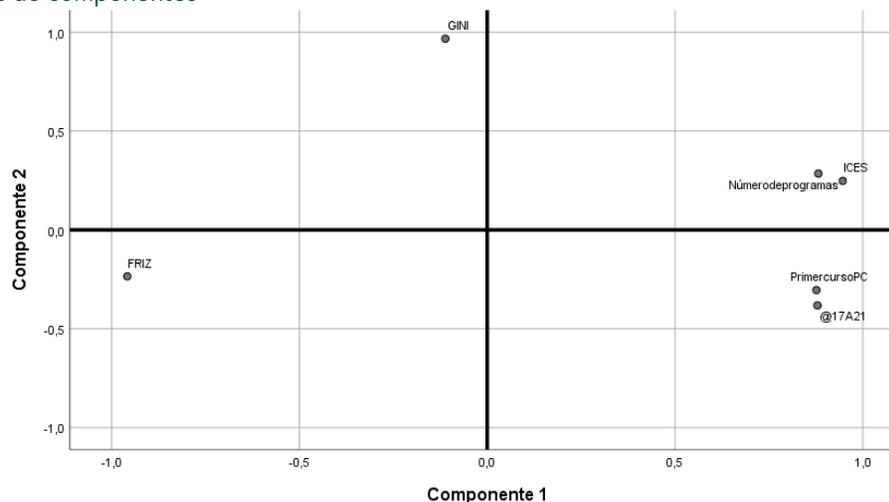
Además del análisis numérico, el análisis de componentes principales, al igual que otras técnicas de reducción de la dimensionalidad, permite un análisis visual. Si se observa el cuadrante positivo en el eje horizontal de la Figura 1, se puede ver cómo el ICES, el número de programas, la población de 17 a 21 años y los estudiantes de primer curso son variables que aumentan en conjunto. Este comportamiento fue más claro entre 2016 y 2022, según lo observado en la Figura 1. Por otro lado, el Índice de Gini y el Índice de Friz tienen un aporte negativo. El efecto de estas variables fue más evidente, de acuerdo con la Figura 1, en el periodo de 2010 a 2015.

Tabla 4. Componentes principales generados

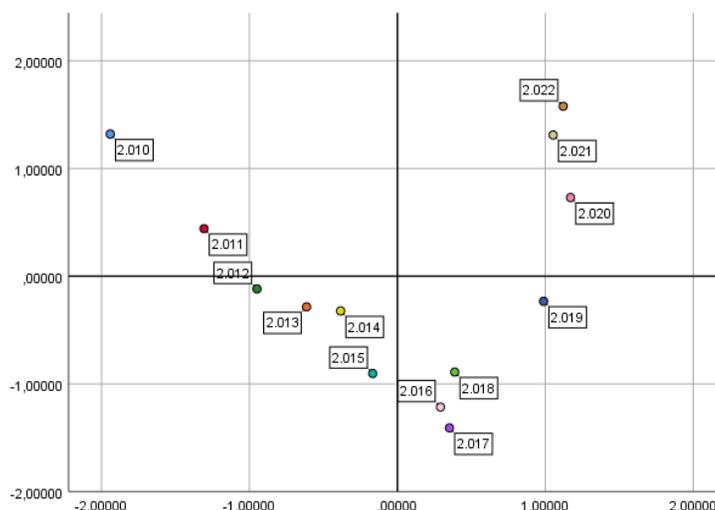
Año	Componente 1	Componente 2
2010	-1,94115	1,31924
2011	-1,3073	0,44124
2012	-0,94946	-0,1183
2013	-0,61383	-0,28532
2014	-0,3842	-0,32285
2015	-0,16697	-0,90385
2016	0,2916	-1,2152
2017	0,35202	-1,40892
2018	0,3875	-0,89054
2019	0,98826	-0,23342
2020	1,1704	0,72966
2021	1,05261	1,31072
2022	1,12052	1,57755

Fuente: elaboración propia en SPSS

Figura 1. Gráficos de componentes



(a) Componentes por variables



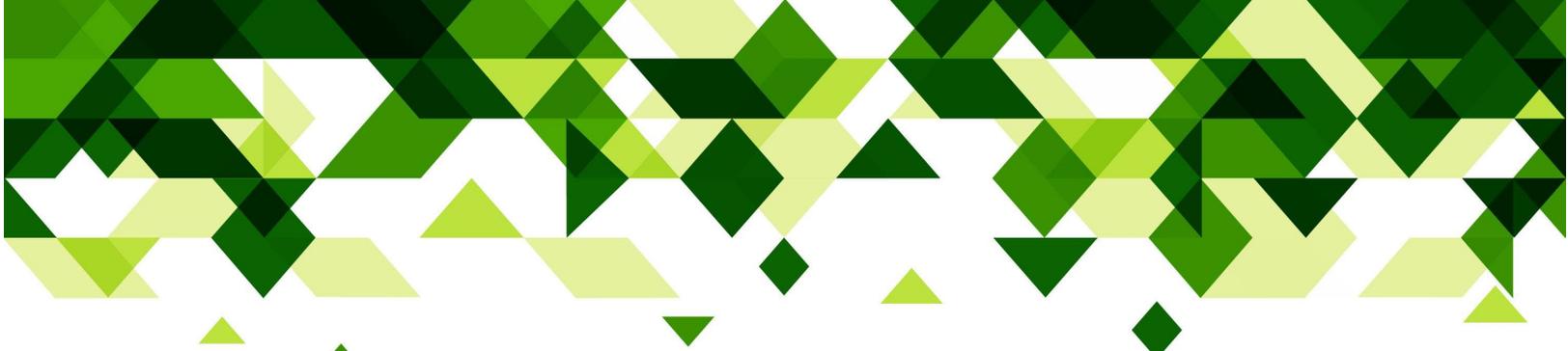
(b) Componentes por casos
Fuente: elaboración propia con SPSS

CONSIDERACIONES FINALES

Los resultados de cualquier análisis estadístico deben ser interpretados en su contexto. El ACP confirma que **existe un efecto positivo** de la población de 17 a 21 años y del tamaño de la oferta de programas **sobre la matrícula de primer curso**. Además, se cuenta con evidencia sólida para afirmar que el Índice de Gini tiene un efecto negativo en esta variable. Sin embargo, el análisis cuestiona la relevancia del Índice de Costos de la Educación Superior y del Índice Friz. Aunque estas variables tienen una alta correlación con la variable de interés y son

explicadas en un porcentaje muy alto por los componentes generados, los resultados no ofrecen un sentido claro. Esto invita a considerar su eliminación y buscar otras variables proxy para los costos de la educación superior y el cambio demográfico.

El ACP fue útil y permitió conocer mejor los datos analizados. Se destaca que este ejercicio estadístico evidencia la necesidad de un pensamiento iterativo en el desarrollo de modelos más robustos con un sentido explicativo. El análisis de regresión previo permitió identificar problemas de colinealidad y dependencia entre las



variables. El ACP nos mostró cuáles de las variables podrían ser excluidas del análisis y reemplazadas por otras. En general, todo tipo

de técnicas univariadas y multivariadas permiten tomar decisiones sobre la información.

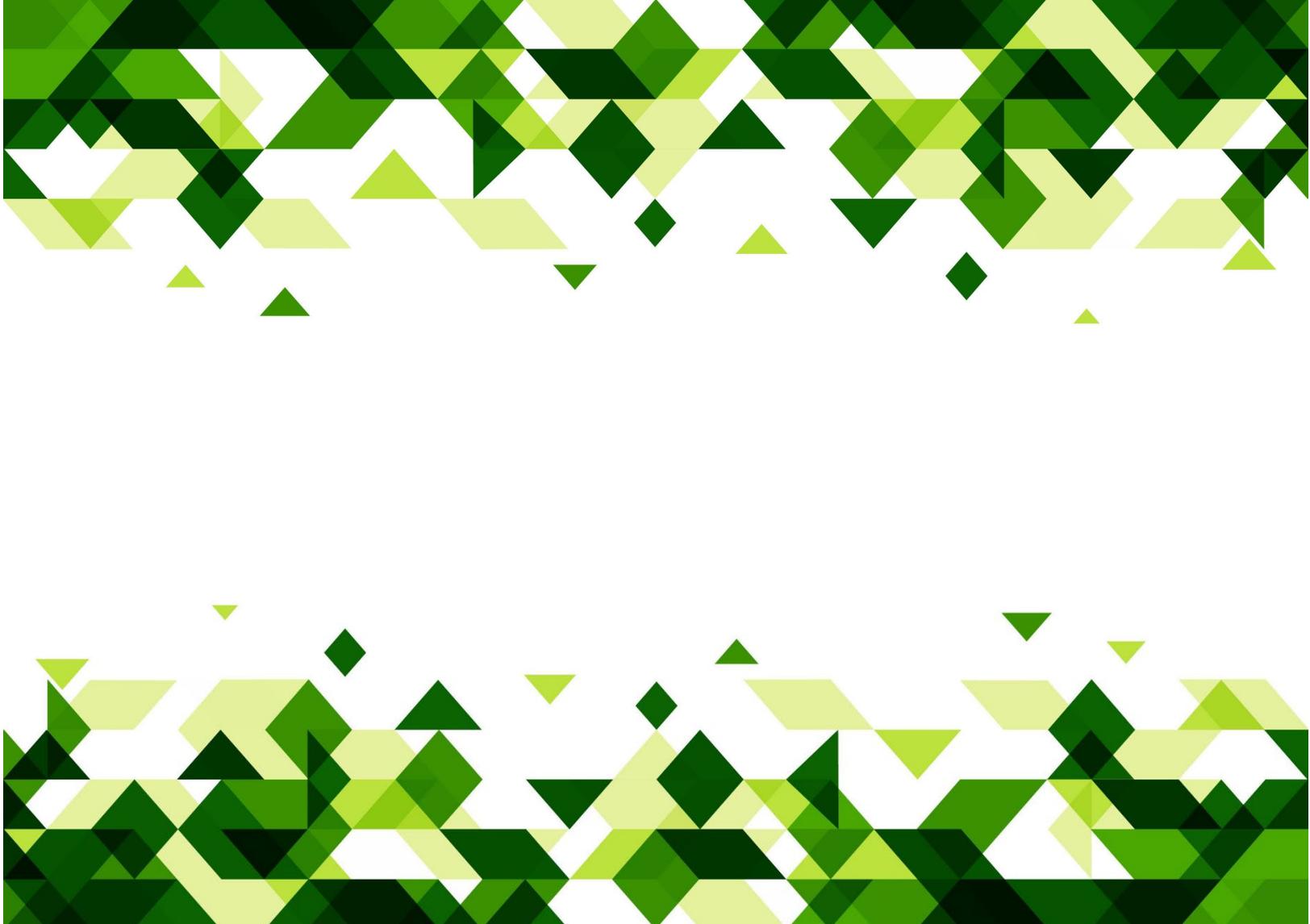
REFERENCIAS

Gil, F. (1995). El estudiante como actor racional: objeciones a la teoría del capital humano. *Revista de Educación*, (306). 315 - 327. <https://dialnet.unirioja.es/servlet/articulo?codigo=19062>

Pla, L. (1986). *Análisis multivariado: método de componentes principales*. Organización de los Estados Americanos (OEA).

BASES DE DATOS CONSULTADAS

- Banco Mundial
- Departamento Administrativo Nacional de Estadística de Colombia
- Ministerio de Educación Nacional de Colombia



Realización:

ROBINSON GÓMEZ ZAPATA
Asesor de Prospectiva

Revisión:

DEIXA MORENO CASTRO
Coordinadora de Investigación Científica
Dirección de Investigaciones

ESCENARIOS

Es el boletín de prospectiva de la Oficina de Planeación de la Universidad La Gran Colombia, con el que se busca generar interacción sobre estos temas en la comunidad universitaria local y nacional.

Deje sus comentarios, sugerencias, inquietudes y solicitudes de temas a
robinson.gomez@ugc.edu.co



UNIVERSIDAD
La Gran Colombia

