

**IMPACTO DE LAS ACTIVIDADES DE TRABAJO AUTÓNOMO SOBRE
LOS RESULTADOS ACADÉMICOS DE ESTUDIANTES DE BAJO NIVEL
SOCIOECONÓMICO: EL CASO DEL MUNICIPIO DE SOACHA***

IMPACT OF AUTONOMOUS WORK ACTIVITIES ON THE
ACADEMICAL RESULTS IN LOW SOCIOECONOMIC LEVEL
STUDENTS: THE CASE OF THE MUNICIPALITY OF SOACHA

Germán Augusto Forero Cantor**
Gisou Díaz Rojo***

* Artículo resultado de trabajo de investigación en la Facultad de Matemáticas de la Universidad de La Habana (Cuba).

** Docente tiempo completo, Facultad de Ciencias Economicas y Administrativas Universidad del Tolima (Colombia). gaforeroc@ut.edu.co

*** Docente tiempo completo, Facultad de Ciencias de la Universidad del Tolima (Colombia). gdiazrj@ut.edu.co

Correspondencia: Universidad del Tolima, Facultad de Ciencias Económicas y Administrativas, Barrio Santa Helena Parte Alta. A.A. 546, Ibagué (Colombia).

RESUMEN

Este artículo implementa algunas técnicas de la estadística multivariada, como el análisis de componentes principales, el escalamiento multidimensional (EMD) y los árboles de clasificación para luego integrarlas con las técnicas propias de la evaluación de impacto y construir un modelo de diferencias que permitiera medir el impacto del programa “Aplicación de actividades de trabajo autónomo para mejorar el rendimiento académico de los estudiantes de la asignatura de Precálculo”. Los resultados evidenciaron un impacto positivo estadísticamente significativo del programa evaluado, así como una influencia de las variables Colegio y Estrato socioeconómico sobre la calificación de los estudiantes, lo que permite pensar lo importante que pueden ser la formación inicial del estudiante y las condiciones socioeconómicas de la familia en estos procesos de aprendizaje.

PALABRAS CLAVE: Rendimiento Académico, Evaluación de impacto, estadística multivariada.

Clasificación JEL: I29.

ABSTRACT

This article implements some techniques of multivariate statistics such as principal components analysis, multidimensional scaling (MDS) and classification trees and integrates them with the techniques of impact valuation to construct a difference model to measure the impact of the program “Application of autonomous activities to improve the academic performance of Pre-calculus students at Corporación Universitaria Minuto de Dios –CMUD- Centro Regional Soacha”. The results evidence a positive and significant influence of the valued program as well as an influence of variables such as School and Socio-economical Stratum on the student scores. This establishes the importance of student’s initial formation and family socioeconomic conditions on the learning process.

KEYWORD: Academic performance, Impact Valuation, Multivariate statistics.

JEL CODES: I29.

1. INTRODUCCIÓN

En el campo de la investigación, particularmente cuando hablamos desde la evaluación de impacto de un programa, el compromiso y responsabilidad de las decisiones tienen repercusión no solo a nivel científico, sino económico y social. La rigurosidad en la formulación de las hipótesis de trabajo y toma de decisiones adquieren, por tanto, una importancia vital. La estadística constituye una parte indisoluble de las investigaciones sociales, desde el diseño hasta la discusión de los resultados.

Las investigaciones relacionadas con la evaluación de impacto son, de manera general, investigaciones sociales aplicadas que hacen uso de herramientas matemáticas y estadísticas no para generar nuevo conocimiento teórico, sino que, por el contrario, buscan utilizar dichas herramientas en función de estimar con precisión el impacto sobre ciertos indicadores sociales a partir de un proyecto para, posteriormente, permitir que los gobernantes o planificadores planteen y organicen estrategias con respecto a la corrección de políticas previas o del proyecto, así como la replicabilidad y, en algunos casos, una reprogramación o discontinuidad del mismo.

En cualquier caso en el que se lleva a cabo un proceso de inversión de recursos (financieros, físicos, etc.), ya sea tanto a nivel privado como público, se hace necesario establecer al cabo de un determinado período si se logró modificar la situación o el problema que motivó dicha inversión. Con frecuencia, a los financiadores y/o ejecutores de un proyecto les preocupa evaluar el impacto que este tuvo sobre la población objetivo; es decir, medir, el(los) cambio(s) que con la ejecución del proyecto se logran alcanzar. Normalmente, realizar ese análisis no es una tarea sencilla, ya que en algunas ocasiones no es posible distinguir cuándo realmente los cambios son producto de la ejecución del proyecto como tal o un resultado de la incidencia de otro tipo de variables que intervienen en el entorno del mismo y no pueden ser involucradas dentro del análisis.

En general, para la evaluación de un proyecto se tienen en cuenta tres dimensiones que son la financiera, la económica y la social. En

la financiera, fundamentalmente, se quiere determinar si el proyecto genera beneficios financieros para quien invirtió recursos en él, por lo que generalmente esta se visualiza desde una perspectiva puramente privada. La económica realiza un análisis desde un entorno que los economistas denominan “macro”, donde el criterio no es exclusivamente financiero sino que se incorporan impactos sobre variables como el empleo y los recursos generados y liberados por el proyecto, entre otros, teniendo en cuenta para ello como pilar fundamental un criterio de eficiencia económica. Finalmente, la dimensión social busca determinar el impacto que los recursos implementados en un proyecto de inversión –generalmente de carácter público– tienen sobre la población objetivo, involucrando para esto no solamente criterios de eficiencia sino de equidad sobre el uso de los recursos.

En ocasiones se le da poca importancia a las técnicas multivariadas como parte de la exploración de datos antes de realizar una evaluación de impacto. Sin embargo, este trabajo es una muestra de lo importante que pueden ser estas técnicas en la evaluación de impacto de un programa de intervención social, permitiendo una mejor comprensión del problema, identificando relaciones entre las variables involucradas y logrando profundizar en los aspectos y factores que componen el problema social estudiado.

La evaluación de impacto es un tipo de evaluación “completa”, en el sentido de que no solo involucra una evaluación cualitativa del proyecto después de haber sido aplicado, sino que desarrolla diferentes estrategias metodológicas, tanto cualitativas como cuantitativas, que permiten comprobar el impacto de un proyecto de inversión y donde se hace necesario contar con mediciones antes y después del desarrollo del proyecto de variables involucradas en el problema (Valdés, 2008).

De esta manera, el objetivo general de este trabajo fue evaluar el impacto del programa “Aplicación de actividades de trabajo autónomo de Precálculo dirigida a los estudiantes de la CUMD-Centro Regional Soacha”. Se plantearon tres objetivos específicos: el primero: Estudiar y aplicar diferentes técnicas de análisis multivariado para descubrir posibles relaciones entre las variables estudiadas; el segundo: Analizar las técnicas de evaluación de impacto y aplicar la más adecuada en el

problema de investigación, y el tercero: Integrar ambos métodos en la evaluación del impacto del programa aplicado en esta institución educativa.

2. REVISIÓN DE LITERATURA

En general, los proyectos de inversión pública o de intervención, fundamentalmente, buscan mitigar problemas de carácter socioeconómico, como bajos niveles de educación, mala nutrición, mortalidad infantil, etc. —sólo para mencionar algunos—, los cuales son característicos de países y/o regiones pobres alrededor del mundo. Es por esto que la evaluación de impacto ha sido aplicada mayoritariamente a programas que se implementan en países de América Latina, África y Asia con niveles de pobreza elevados.

La evaluación de impacto tiene como objetivo determinar de forma general si un programa o proyecto produjo los efectos deseados en las personas, instituciones o regiones donde se implementó, y si esos efectos son resultado de la intervención del programa, permitiendo además examinar consecuencias no previstas en los beneficiarios (Baker, 2000). Por ello, la evaluación debe identificar si existen o no relaciones de causa-efecto entre el programa y los resultados obtenidos y esperados, ya que pueden existir otros factores que pueden estar presentes durante el período de intervención del programa, que podrían estar correlacionados con los resultados y que no son tenidos en cuenta al momento de la evaluación. Fundamentalmente, el problema de evaluación de impacto consiste en establecer la diferencia entre la variable de resultado previamente establecida del individuo que participó en el programa y la variable de resultado de ese individuo si no hubiera participado en el programa. A esta diferencia se le llama efecto del tratamiento o programa.

La idea de la evaluación de impacto de programas no es nueva, ya que desde los años setenta del siglo XX se ha venido llevando a cabo. Esta surge durante un período en que el desarrollo social y cultural de algunos países estaba relacionado con un gran número de programas dirigidos a mejorar el bienestar social y económico de los individuos. Dada entonces la proliferación de programas, en forma

paralela se generó la necesidad de conocer los resultados –impactos– de la aplicación de estos, a través de mecanismos de evaluación que deberían permitir detectar de manera temprana sus aciertos y fallas, con el objetivo de poder mejorar las estrategias de los programas y estructurar mejor los programas futuros (Añorga & Valcárcel, 1999).

La literatura alrededor del tema de la Evaluación de Impacto se ha enriquecido de manera importante en los últimos años. En este sentido, y como resultado de un análisis pormenorizado, y de lo que podría denominarse una “Evaluación de la Evaluación de Impacto”, Vara-Horna (2010) hace un análisis de artículos e informes internacionales especializados publicados en el período 1976-2006 que giran alrededor de la evaluación de programas y evaluación de impacto. Entre los principales hallazgos del estudio se encuentra un incremento significativo hacia fines de la década de los noventa y principios de la presente década. La mayoría de estos estudios (77.3%) se publicaron durante el período 2000-2006, mientras que alrededor del 15% durante la década de los noventa y sólo un 7.4% desde 1976 hasta 1989. Las principales áreas de estudio e interés de las evaluaciones de impacto para el análisis fueron los programas educativos (26.3%), de salud (21.5%), de pobreza extrema (14.6%) y de empleo (11.4%). En menor porcentaje se han centrado en la evaluación de programas alimentarios, infraestructura y electrificación, agricultura y ganadería, entre otros.

En términos metodológicos, Castro (2008) presenta una síntesis de diferentes métodos utilizados en la priorización de inversiones públicas en los sectores de salud y educación, durante todo el ciclo del proyecto –desde la identificación hasta la evaluación–. Menciona, además, las metodologías de evaluación de impacto haciendo énfasis en los estimadores comúnmente utilizados en diseños no experimentales: estimador antes y después, el estimador de sección cruzada, el estimador de diferencia en diferencia, el estimador Matching, entre otros; todo con el propósito de abordar los dos problemas fundamentales en este tipo de evaluaciones: primero, elegir el grupo de control adecuado para llevar a cabo la comparación de la situación sin y con proyecto de los beneficiarios y, segundo, garantizar que los beneficios recibidos

por la población objetivo (beneficiarios) pueden ser atribuibles en su totalidad al proyecto.

Luego de haber consultado la bibliografía relacionada con este tema, se puede decir que con frecuencia se presentan dificultades al tratar de establecer las variables que influyen en la determinación del impacto de un programa, por tanto es importante mencionar que no existen modelos ideales de planeación ni de prácticas de evaluación, y cada estudio de evaluación de impacto debe adaptarse a los objetivos y condiciones de cada programa.

En este sentido, Valdés (2008) expone que la evaluación de impacto es un proceso integral que evalúa no solo el logro de los objetivos, sino que debe tener en cuenta los diferentes factores que influyen en el cómo serían el contexto, el proceso, los resultados y el producto. Por tal motivo, se plantea la necesidad de acudir a la implementación y uso de diferentes instrumentos y etapas según el entorno donde se desarrolle.

Las técnicas estadísticas utilizadas en la evaluación de impacto son diversas, ya que estas dependen del área de aplicación donde se realice el estudio, así como de los objetivos del programa. Por ejemplo, en los estudios relacionados con la evaluación de impacto ambiental, las técnicas utilizadas se dividen en dos grupos (Plazas et al., 2009); el primero está conformado por los métodos tradicionales de evaluación de proyectos (análisis costo-beneficio), con el cual se realizan mediciones en términos monetarios; el segundo hace referencia a métodos cuantitativos, los cuales intentan aplicar escalas valorativas para los diferentes impactos, medidos originalmente en sus respectivas unidades físicas.

Algunos de estos métodos se centran en la identificación y síntesis de los impactos y otras involucran la explicación de las bases de cálculo de cada impacto generado. Para esto recurren a modelos matemáticos que utilizan ecuaciones que representan las relaciones que existen entre las variables, las cuales corresponden a los atributos o características de los impactos. Para llegar a un modelo adecuado, algunos autores proponen el uso de técnicas multivariadas, como el análisis de componentes principales, análisis de *cluster* y el análisis discriminante.

Por otra parte, en los estudios de evaluación de impacto de programas sociales se destacan: el modelo de diferencias, el modelo de diferencias en diferencias, los métodos de emparejamiento, variables instrumentales, regresión discontinua, funciones de control y los modelos estructurales.

Desde un punto de vista particular, se encuentran estudios de evaluación de impacto enfocados específicamente en programas educativos o de formación cívica-ciudadana, que aunque están dentro de los llamados programas sociales su análisis e interpretación tienen en cuenta los métodos cualitativos y los modelos teóricos, enmarcándose en lo que se podría denominar una metodología mixta. Lo usual en el análisis estadístico de este tipo de modelos “mixtos” es desarrollar un análisis de factores exploratorios (se calculan medidas descriptivas) en dos momentos que son antes y después de la intervención, y en los dos grupos (control y tratamiento), los cuales se comparan a través de algunos de los métodos confirmatorios con el fin de determinar posibles diferencias entre los grupos. Adicionalmente, para profundizar en algunos resultados es común encontrar que se realicen algunos análisis de tipo multivariado para poder así comparar los cambios en la composición de los grupos previamente identificados (Valdés, 2008). La pertinencia de cada uno de estos métodos depende de los datos, de las características del problema y de los objetivos del programa que se está evaluando.

3. METODOLOGÍA

3.1. ANÁLISIS MULTIVARIADO COMO TÉCNICA DE EVALUACIÓN DE IMPACTO

Las técnicas multivariadas son cada vez más utilizadas tanto en estudios científicos como en las áreas sociales. Los métodos de análisis exploratorio y confirmatorio de datos, que generalmente se utilizan de forma combinada, requieren de un conocimiento previo del problema y de los factores con que está relacionado. Se puede definir el análisis multivariado como aquella rama de la estadística interesada en el

estudio de las relaciones entre series de variables (independientes o no) y de los individuos que la sustentan.

Según Linares (1990), las técnicas de Análisis Multivariado se dividen en dos: unas que tienen en cuenta las relaciones de interdependencia entre las variables, y otras que se basan en las relaciones de dependencia. Las técnicas basadas en relaciones de dependencia establecen una distinción entre las variables a explicar y las variables explicativas u observadas. Las variables a explicar son las llamadas dependientes, y las variables explicativas, independientes. Tales técnicas tienen por objeto establecer la relación entre las variables como base para realizar una predicción. Las técnicas basadas en relaciones de independencia no establecen ninguna diferenciación entre las variables, recibiendo todas ellas el mismo tratamiento. El objetivo que se persigue al utilizar estas técnicas es el de organizar los datos de forma que sean más manejables para el investigador y ofrezcan una mayor comprensión global.

Las principales técnicas de análisis multivariado son:

i) El *análisis de Componentes Principales* (ACP), desarrollado por Pearson y por Hotelling, se utiliza con propósitos descriptivos y de reducción de dimensionalidad (Johnson y Wichern, 2002). El ACP es un análisis interno, concerniente a las varianzas y covarianzas de los elementos de un vector aleatorio, por lo que no involucra relaciones externas con cualquier otro vector. Debe tenerse en cuenta que al transformar y reducir el número de variables originales, parte de la información contenida en la matriz de varianzas y covarianzas de las variables se sacrifica, pero en el ACP esta pérdida de información es mínima (Linares, 1990).

ii) El *análisis de cluster* pretende buscar grupos dentro de un conjunto de datos, donde en general se tienen datos y se busca dividirlos en cierto número de grupos de modo que dos datos dentro del mismo grupo sean similares entre sí y qué datos de diferentes grupos sean diferentes.

iii) El *escalamiento multidimensional* es un conjunto de técnicas y métodos de análisis cuya finalidad es obtener la estructura subyacente de los datos y la representación geométrica de los mismos, a través

de la construcción de una configuración de puntos en un espacio de baja dimensión, cuando se conoce una determinada información sobre proximidades entre objetos. Las pocas restricciones que se imponen a los datos es lo que otorga al escalamiento multidimensional una gran aplicabilidad en diversas áreas, especialmente en el campo de la medición de factores humanos (Conchillo et ál., 1993).

iv) Los *árboles de clasificación* constituyen una herramienta flexible y eficaz que permite explorar estructuras de datos complejas, y son útiles cuando necesitamos clasificar o hacer predicciones de resultados (Sistachs, Joya & Díaz, 2011). Este procedimiento crea un modelo de clasificación basado en árboles, y clasifica casos en grupos o pronostica valores de una variable dependiente basada en valores de variables independientes, proporcionando además herramientas de validación para análisis de clasificación exploratorios y confirmatorios.

3.2 OTRAS TÉCNICAS DE EVALUACIÓN DE IMPACTO

En términos modernos, la bibliografía que gira entorno a las medidas de evaluación de impacto ha tenido amplios desarrollos a partir de las ciencias económicas y sociales. Aunque son bastante utilizadas, la bibliografía relacionada con estas es bastante escasa, y se destacan básicamente los textos de Bernal y Peña (2011) y Aedo (2005).

El problema de la evaluación de impacto consiste en establecer la diferencia entre la variable de resultado del individuo participante en el programa en presencia del programa y la variable resultado de ese individuo en ausencia del programa. Esta diferencia es lo que se conoce como *efecto del tratamiento o programa*. Es evidente que el problema que se enfrenta se deriva de que no se pueden observar ambos resultados para el mismo individuo al mismo tiempo. Por tanto, el resultado del individuo participante, si el programa no existiera, es hipotético y no se observa. En la literatura de evaluación de impacto este resultado hipotético se denomina *resultado contrafactual*.

Generalmente, para la evaluación de impacto es útil contar con estimaciones de la variable de resultado antes y después de la aplicación de un programa; sin embargo, si la aleatorización fue realizada correctamente, no se requiere contar con estimaciones de la variable

de resultado en la situación antes del tratamiento. Esto se debe a que una asignación al azar en los grupos de tratamiento y control garantiza que ambos grupos presenten, en promedio, el mismo valor de la variable resultado y, por ende, toda la diferencia, en caso de existir, radicaría en los valores finales de la variable de resultado después de aplicado el programa. De esta manera, una comparación entre las medias de la variable de resultado entre el grupo de tratamiento y control después de la aplicación del programa permitiría obtener una medición del impacto del programa.

Dentro de las técnicas de evaluación de impacto más utilizadas se destacan:

i) El *Modelo de diferencias*, en el cual se parte del supuesto que los datos son aleatorios e independientes y que, por tanto, los grupos a comparar tienen características iniciales similares (Bernal & Peña, 2011). Como cada uno de los individuos elegibles para participar en el programa puede realizar dos acciones alternativas, participar o no participar, por lo que tiene dos resultados posibles respecto a la variable de interés (variable resultado). Se define el indicador del tratamiento como D_i . En el caso de que el tratamiento es binario, entonces $D_i = 1$ si el individuo i participa en el programa y 0 de lo contrario. El impacto del programa se puede medir como la diferencia de medias en la variable de resultado entre los grupos de tratamiento y control después de la intervención, a través de un modelo de regresión lineal simple con variables indicadoras.

ii) El *Modelo de diferencias en diferencias*, en el que bajo un diseño cuasi experimental es probable que los grupos control y tratamiento presenten diferencias aun antes de la aplicación del programa, cuestión que se debe tener en cuenta para estimar el efecto del programa sobre la variable de resultado. El modelo de diferencias en diferencias controla de alguna manera las posibles diferencias preexistentes entre los dos grupos en estudio. Según Bernal y Peña (2011), este modelo incorpora el cambio esperado en Y entre el período posterior y el período anterior a la implementación del programa en el grupo de tratamiento menos la diferencia esperada en Y en el grupo control durante el mismo período. Este estimador requiere de la existencia

de observaciones de los mismos individuos antes y después de la implementación del programa.

iii) El *Método de variables instrumentales*, que se utiliza en caso de que el diseño del estudio realizado sea no experimental, y los datos no provengan de una asignación aleatoria al programa, podemos aplicar algunos métodos para solucionar el sesgo de autoselección.

iv) El *Método de regresión discontinua* (RD), que se emplea generalmente cuando el programa a implementar utiliza una discontinuidad en la regla (variable continua observada Z) de asignación del tratamiento al grupo de beneficiarios; esta regla será utilizada como una asignación aleatoria en un vecindario para estimar el impacto del programa. La idea fundamental es que los individuos que están justo arriba del corte utilizado para asignar el tratamiento, son muy parecidos a los individuos que han sido clasificados justo por debajo del corte. Por tal motivo, para la evaluación del impacto se compararán los resultados obtenidos entre ambos grupos de individuos tratados y no tratados.

Este método tiene como supuestos que el umbral es arbitrario y que los determinantes de las variables X son parecidos a la izquierda y a la derecha del umbral. Por tal motivo, la comparación de estos dos grupos resulta un estimador válido del impacto del programa alrededor del punto de corte. El parámetro estimado por RD es de carácter local, porque se centra en los individuos de una vecindad arbitrariamente pequeña a la izquierda y a la derecha del umbral elegido (Imbens & Lemieux, 2008).

4. EVIDENCIA EMPÍRICA

4.1. GENERALIDADES DEL PROYECTO

El proyecto “Aplicación de actividades de trabajo autónomo de Precálculo dirigida a los estudiantes de la CUMD-Centro Regional Soacha” surge como respuesta a las dificultades que presentan los estudiantes en el proceso de aprendizaje de las ciencias básicas, en particular de las matemáticas. En la búsqueda de alternativas para llevar a cabo con éxito el proceso enseñanza-aprendizaje, se propuso

desarrollar un conjunto de actividades para el trabajo autónomo de los estudiantes, acompañadas por los docentes, para estimular y desarrollar el pensamiento matemático, la generalización y aplicación de conceptos matemáticos y, por tanto, el desempeño en la asignatura de Precálculo, la cual es la base para la carrera en estudio y su futura práctica profesional. Este proyecto empezó a efectuarse en el primer semestre de 2009.

1. El proyecto se planificó para realizar en tres etapas principales:
2. Diseño de las actividades de trabajo autónomo.
3. Desarrollo de las actividades de trabajo autónomo.

Evaluación de los resultados obtenidos a través de las actividades de trabajo autónomo.

A continuación, se muestran en la tabla 1 la relación de las actividades de trabajo autónomo propuestas por el proyecto.

Tabla 1. Títulos de las actividades de trabajo autónomo propuestas

Actividad	Título
1	Conjuntos numéricos y operaciones básicas con números enteros
2	Números racionales
3	Exponentes, radicales y logaritmos
4	Introducción al álgebra
5	Productos, cocientes notables y factorización
6	Fracciones algebraicas
7	Ecuaciones de primer grado
8	Problemas que se resuelven por ecuaciones de primer y segundo grado
9	Desigualdades lineales y cuadráticas
10	Funciones
11	Función lineal y ecuación de la recta

El proyecto se desarrolló teniendo en cuenta las características y condiciones de los estudiantes que ingresan a los programas que ofrece la Universidad (especialmente aquellos de la jornada nocturna), relacionadas con el tiempo disponible (que es poco), falta de hábitos de estudio autónomo, poco desarrollo de competencias, habilidades y conocimientos adquiridos en la educación media y las dificultades de tipo social, aspectos que usualmente los lleva al fracaso académico.

4.2. DATOS

En el transcurso de los años 2009 y 2010 se incluyeron en el estudio 872 estudiantes de cuatro programas tecnológicos de la Universidad del Minuto de Dios del Centro Regional Soacha. Se conformaron dos grupos para el análisis de impacto del programa: *grupo tratamiento* (estudiantes que recibieron el programa de intervención, en este caso, el refuerzo en matemáticas a través de los talleres de trabajo autónomo propuestos por el departamento de Ciencias Básicas con más horas de clases) y *grupo control* (estudiantes que no recibieron el programa de refuerzos en matemáticas).

Los datos se obtuvieron a partir de bases de datos extraídas del sistema general de información que maneja UNIMINUTO. Se conformaron bases en Excel para cada año según el programa de estudio y luego toda la información se recopiló en una base general de datos, y el análisis estadístico se realizó con el programa SPSS v18. Las variables que se midieron se muestran en la tabla 2. Las variables explicativas se transformaron en dicotómicas.

Tabla 2. Variables recogidas en el estudio

Variable	Categorías
Variable dependiente: Nota	0.00 a 5.00
Grupo	1: Programa, 2: Control
Edad	1: menor o igual a 25 años, 2: más de 25 años
Género	1: masculino, 2: femenino
Estrato socioeconómico	1: estratos bajos (1-2), 2: estratos medios (3-4)
Estado civil	1: soltero, 2: no soltero
Hijos	1: tiene, 2: no tiene
Colegio	1: público, 2: privado
Trabajo	1: trabaja, 2: no trabaja

Entre las técnicas relacionadas con la metodología de evaluación de impacto se utilizó un modelo de diferencias, el cual se desarrolló además incluyendo los resultados que se generaron en el análisis multivariado, lo que permitió así la construcción de un modelo de diferencias que recoge mayor información del problema de investigación y una mejor comprensión de la realidad estudiada por la universidad.

4.3. MODELO DE DIFERENCIAS UTILIZADO PARA MEDIR EL IMPACTO

El modelo de diferencias parte del supuesto que los datos son aleatorios e independientes y que, por tanto, los grupos a comparar tienen características iniciales similares.

Como cada uno de los individuos elegibles para participar en el programa puede realizar dos acciones alternativas, participar o no participar, por lo que tiene dos resultados posibles respecto a la variable de interés (variable resultado), se define el indicador del tratamiento como D_i . En el caso de que el tratamiento sea binario, entonces $D_i = 1$ si el individuo i participa en el programa y 0 de lo contrario.

El impacto del programa se puede medir como la diferencia de medias en la variable de resultado entre los grupos de tratamiento y

control después de la intervención, a través del modelo de *regresión lineal simple con variables indicadoras*:

$$Y_i = \beta_0 + \beta_1 D_i + u_i \quad (4.1)$$

Donde Y_i es la variable de resultado, D_i es el indicador de tratamiento y u_i es el error $u_i \sim N(0, \sigma^2)$.

El modelo ajustado de regresión lineal simple es, entonces:

$$\tilde{y} = \hat{\beta}_0 + \hat{\beta}_1 D \quad (4.2)$$

y al estimador en (4.2), $\hat{\beta}_1$, se llama *estimador de diferencias*, dado que es igual a la diferencia en medias entre los grupos de tratamiento y control.

Cuando se tiene más información sobre los individuos participantes en el estudio, la cual está contenida en otras variables que pueden estar afectando el resultado, como en nuestra investigación, se realiza un modelo de **regresión lineal múltiple**:

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_{1i} + \dots + \beta_{k+1} X_{ki} + u_i \quad (4.3)$$

donde X_{1i} a X_{ki} son características individuales que preceden al tratamiento y que no se ven afectadas por este. Los parámetros β_j , $j=0, 1, \dots, K$ se llaman coeficientes de regresión, y el parámetro β_j representa el cambio esperado en la respuesta y por cambio unitario en x_j cuando todas las demás variables regresoras x_i ($i \neq j$) se mantienen constantes. Dentro de ciertos márgenes de las variables regresoras, el modelo de regresión lineal es una aproximación adecuada a la función verdadera desconocida. En este caso $\hat{\beta}_j$, a se le denomina *estimador de diferencias con regresores adicionales*.

Con respecto a este tema Cuadras (2010) propone que por medio del Análisis de Componentes Principales es posible efectuar una regresión lineal múltiple de Y sobre X_1, \dots, X_k , considerando las primeras componentes principales extraídas Y_1, Y_2, \dots, Y_p como variables explicativas, y realizar una regresión de Y sobre Y_1, Y_2, \dots, Y_p .

evitando así efectos de colinealidad, aunque las últimas componentes principales también pueden influir. Exactamente esta propuesta fue la que se tuvo en cuenta en este trabajo para integrar los resultados del análisis multivariado con el modelo de diferencias utilizado en la evaluación de impacto.

4.4. ANÁLISIS DE RESULTADOS

Como en todo estudio exploratorio de datos, se hizo un análisis descriptivo de las variables recogidas, y aunque en nuestro trabajo no fue un objetivo profundizar en esto, este se ejecutó y ofrecemos un resumen de los resultados más interesantes. El 55,2 % de los estudiantes evaluados son de género Femenino y el 44,8% Masculino, el 64,8% tienen estado civil Soltero, y el 35,2% están distribuidos entre Casados, Unión Libre, Divorciado, Separado, Viudo, o Religioso. En cuanto a la Edad, el 67,2 % tiene edad menor o igual a 25 años, y el 32,8% mayor que 25 años. Si analizamos el Estrato socioeconómico, tenemos que el 14% de los estudiantes provienen de Estrato 1, el 66,5% procede de Estrato 2, el 18,6% viene de Estrato 3 y el 0,9% proviene de Estrato 4. También se pudo observar que el 45,1% de los estudiantes tiene hijos, mientras que el 54,9% no tiene, que el 89,3% de los estudiantes trabaja, contrastado con un 10,7% que no trabaja, y que el 88,9% de los estudiantes vienen de colegios públicos frente a un 11,1% que viene de colegios privados (ver tabla 2).

Los grupos comparados (control y programa) fueron homogéneos para todas las variables descriptivas recogidas, lo que permitió atribuir las diferencias encontradas al programa de intervención.

Al realizar el ACP se buscaba principalmente describir la relación entre las variables y reducir la dimensión del problema con la creación de nuevas variables. Este método se realizó con todas las variables que describen las características de los estudiantes. Se seleccionaron, entonces, dos componentes que explican el 42% del problema, que aunque no es mucho, permite hacer una interpretación lógica de las componentes seleccionadas.

Según la tabla 3 y el gráfico 1, la primera componente (fac_1) contiene mayormente información de las variables Edad, Hijos y

Estado Civil, es decir, que los estudiantes con edad menor o igual a 25 años, solteros y que no tienen hijos es lo que caracteriza a la primera componente principal, por lo que se puede pensar en nombrarla componente *familiar*. La segunda componente (*fac_2*) contiene información de las variables Estrato socioeconómico y Colegio, es decir, los estudiantes que provienen de estrato bajos y de enseñanza pública, y esto caracteriza a la segunda componente principal, que hemos decidido denominar componente *procedencia*. Esto puede corroborarse en la tabla de las componentes, lo que permite hacer una mejor interpretación del problema y ver la formación de estas agrupaciones.

En el análisis de *cluster* se utilizaron todas las variables descriptivas de los estudiantes y estas se agruparon utilizando la distancia euclídea para variables binarias en unión con el método de Ward.

Como se puede observar en el gráfico 3, se forman dos grupos: el primero, que encabeza la unión correspondiente a Trabajo, Colegio, Estrato socioeconómico y Género, esta última variable se agrupa al final, por lo que, además, podemos pensarla de forma independiente. En un segundo grupo aparece la unión de las variables Estado Civil, Hijos y Edad. Estos resultados confirman la relación entre estas variables que se mostraba en el análisis de componentes principales.

Tabla 3. Matriz de componentes extraídos

	Componente	
	1	2
Genero	-,263	-,242
edad_2c	,619	-,129
Estado C	,806	,029
hijos_2c	,808	-,029
Colegio	,051	,746
Trabajo	,054	,214
estrato_2c	-,048	,697

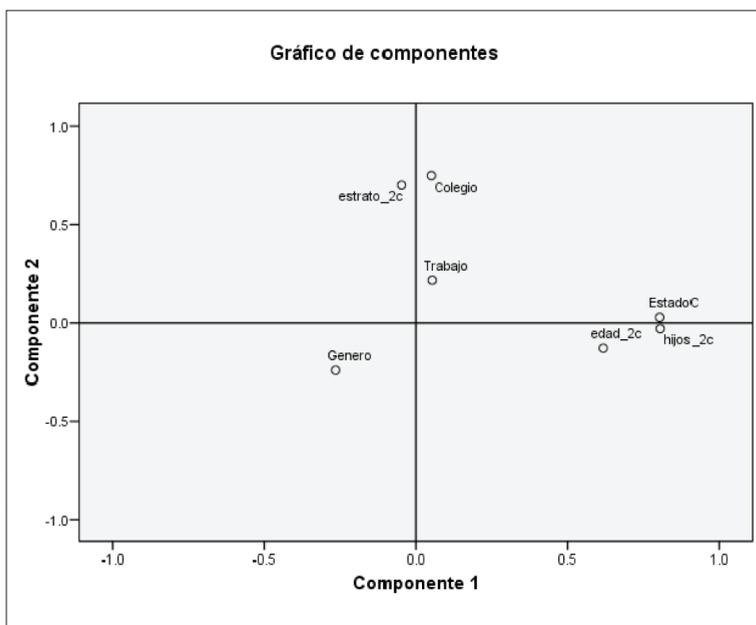


Gráfico 1. Componentes principales

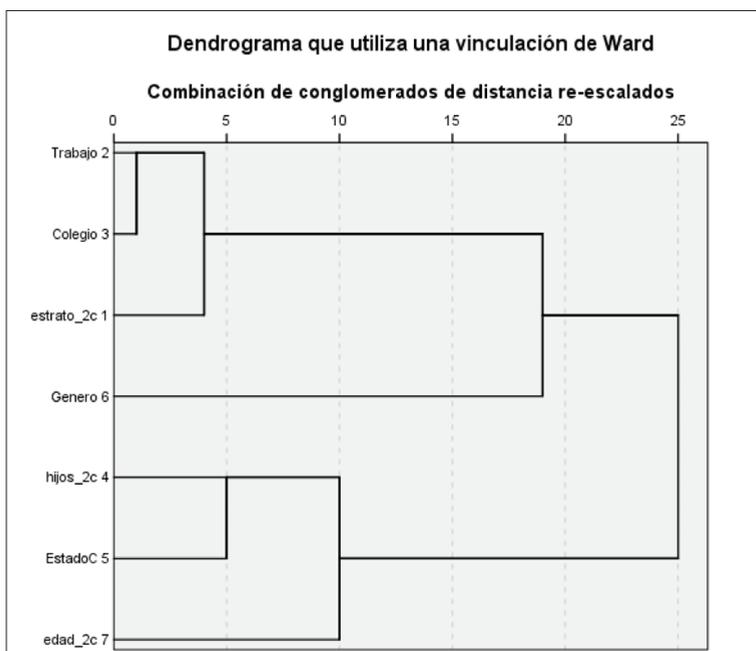


Gráfico 2. Análisis de *cluster*

En el gráfico 3 se muestra el resultado del análisis de escalamiento multidimensional. De cierto modo, se corroboran los resultados obtenidos para este caso con el uso de la técnica de análisis de *cluster* y análisis de componentes principales. Se identifican claramente dos grupos de variables: 1. Edad, Estado civil, Hijos; 2. Colegio, Estrato, Trabajo. Además, se puede apreciar que la variable Género no se agrupa con ninguna otra variable, lo que hace pensar en la posibilidad de que esta no esté relacionada o influya en las variables antes mencionadas. Además, se obtuvo un índice de ESTRESS de 0,051 (según la escala de Kruskal: bueno), lo que indica que la configuración es adecuada para describir las disimilaridades entre las variables.

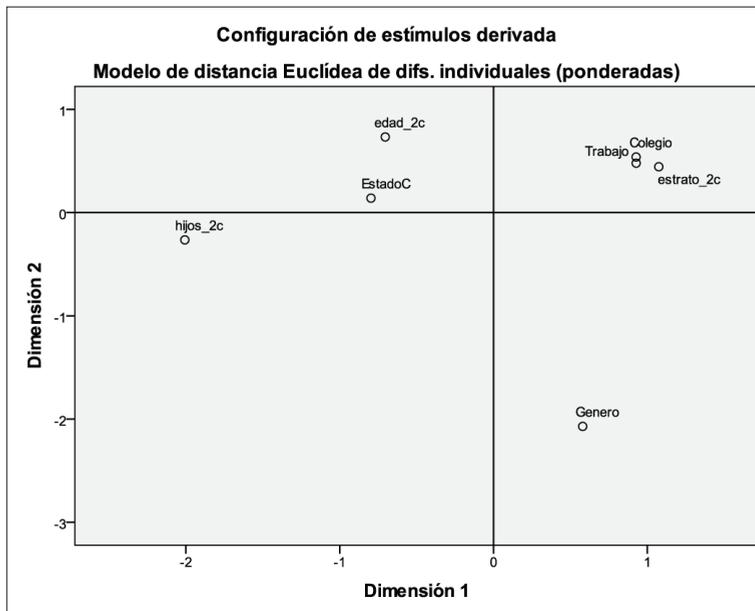


Gráfico 3. Escalamiento multidimensional

El gráfico 4 muestra el árbol de clasificación obtenido por el método CART. Los resultados encontrados permitieron caracterizar los siguientes grupos de estudiantes según reglas de decisión:

- a) El primer perfil con alta probabilidad (0,85) de aprobar la asignatura corresponde al grupo de estudiantes que proviene de colegios privados.
- b) Si el estudiante proviene de un colegio público y no es soltero, entonces es clasificado como que aprobó la asignatura (probabilidad de 0,64).
- c) Si el estudiante proviene de un colegio público, es soltero y es femenino, es clasificado como que aprobó la asignatura (probabilidad de 0,61).
- d) Si el estudiante proviene de un colegio público, es soltero y es masculino, es clasificado como que aprobó la asignatura (probabilidad de 0,49).

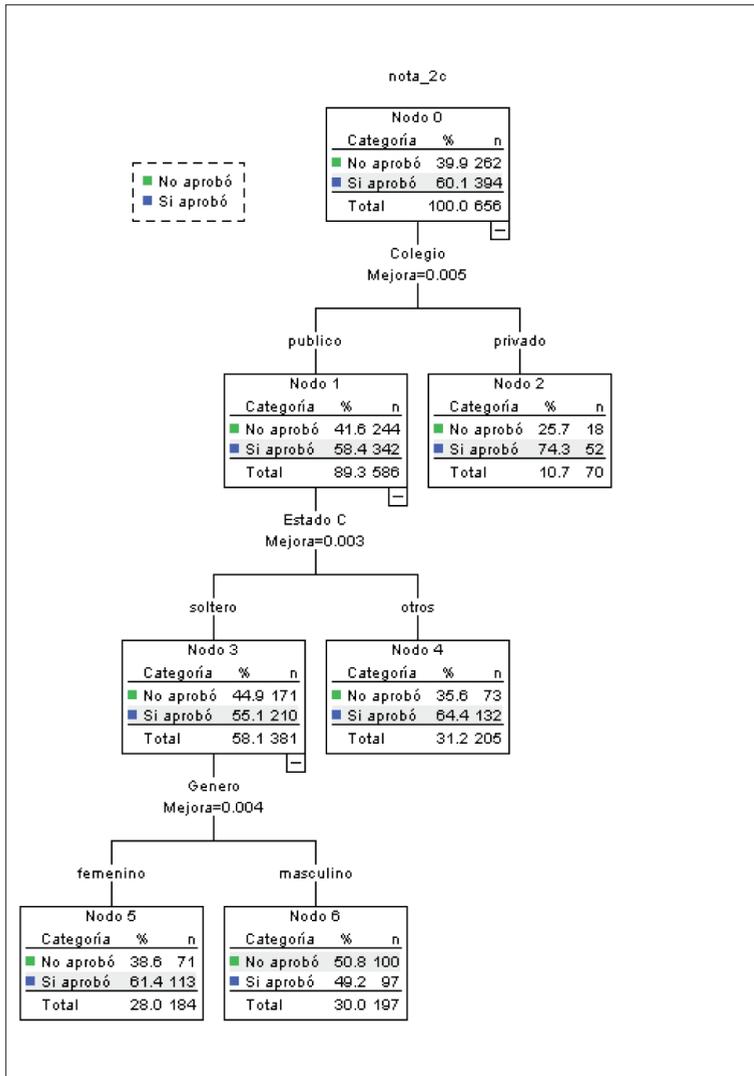


Gráfico 4. Árbol de clasificación

Como resumen de los resultados, podemos señalar que el análisis multivariado permitió caracterizar los estudiantes de las diferentes carreras que fueron intervenidas por el programa “Aplicación de acti-

vidades de trabajo autónomo de Precálculo”, descubrir relaciones entre las variables de caracterización social, reduciendo así la dimensión del mismo, permitiendo una mejor comprensión del contexto social y económico en que se desenvuelven los estudiantes que ingresan a la universidad. Las nuevas variables que resultaron de este análisis fueron utilizadas para obtener el modelo de regresión y de esta manera determinar el impacto del programa evaluado.

El modelo planteado para determinar el impacto del proyecto fue un modelo de diferencias (regresión lineal simple con variable indicadora), que tiene una estructura de la forma:

$$Nota_i = \beta_0 + \beta_1 Grupo_i + u_i \quad (4.1)$$

Donde la variable “Nota” hace referencia a la calificación obtenida por el estudiante i y la variable “Grupo” es una variable dicotómica que determina si el estudiante tuvo participación o no en el programa establecido para el mejoramiento de las calificaciones. Al aplicar este modelo obtenemos los resultados que se muestran en la tabla 4, en la que se observa que para quienes participaron en el programa el impacto es de 0.369 (estimación de β_1). El p-value encontrado para este caso muestra que el coeficiente de la variable Grupo es estadísticamente significativo en el modelo.

Tabla 4. Coeficientes del modelo de diferencias

Modelo	Coeficientes no estandarizados		Coeficientes estandarizados	t		Sig.		Intervalo de confianza para B al 95%	
	B	Error típ.	Beta	Límite inferior	Límite superior	B	Error típ.	B	Error típ.
(Constante)	3,031	,038		79,463	,000	2,956	3,106		
Grupo	,369	,053	,230	6,976	,000	,265	,473		

Con el objetivo de enriquecer y comparar los resultados obtenidos con el modelo de diferencias simple, realizamos este mismo modelo de diferencias pero utilizando las nuevas variables obtenidas del análisis de componentes principales.

La variable *fac_1* (que agrupa las variables Hijos, Edad, Estado civil) y la variable *fac_2* (que agrupa las variables Estrato y Colegio), esta integración de los resultados de los métodos señalados en los capítulos 2 y 3, es uno de los principales aportes de este trabajo: integrar los resultados del análisis multivariado con los modelos de regresión utilizados para medir el impacto del programa y las variables relacionadas con este.

Al realizar la regresión con las nuevas variables *fac1_1* y *fac2_2*, se puede observar que el coeficiente de la variable *fac1_1* no tiene una influencia significativa sobre los resultados obtenidos por los estudiantes, lo que puede interpretarse como que las variables con las que se estableció *fac1_1* (Hijos, Edad, Estado Civil) no tienen incidencia sobre los calificaciones finales de los estudiantes.

Por otro lado, el coeficiente de la variable *fac2_2* (Estrato y Colegio) sí presenta una influencia estadísticamente significativa, lo que implica que las variables Estrato socioeconómico y Colegio inciden en las calificaciones alcanzadas por los estudiantes, dando un papel fundamental a la formación inicial del estudiante, las condiciones para estudiar y el apoyo económico familiar (ver tabla 5).

Tabla 5. Coeficientes del nuevo modelo de diferencias

Modelo	Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
	B	Error típ.	Beta		
(Constante)	3,034	,038		79,978	,000
<i>fact_1</i>	-,042	,026	-,053	-1,604	,109
<i>fact_2</i>	-,082	,026	-,102	-3,106	,002
Grupo	,364	,053	,227	6,918	,000

CONCLUSIONES

Con el uso de las técnicas multivariadas estudiadas (análisis de componentes principales, análisis de *cluster*, escalamiento multidimensional y árboles de clasificación) fue posible la caracterización del problema, y se formaron esencialmente tres grupos entre las variables: uno compuesto por Estado Civil - Hijos- Edad; otro, por Estrato socioeconómico - Colegio, y la variable Género quedaría sola en un tercer grupo.

A través del método de árboles de clasificación (CART) se establecieron reglas de decisión que permiten clasificar a los estudiantes según las características en estudio, lo que facilita extraer inferencias del impacto del programa.

Las variables Estrato socioeconómico y Colegio (a través de la nueva variable *fact_2* del ACP) inciden en las calificaciones alcanzadas por los estudiantes, otorgando así un papel importante a la formación inicial y previa del alumno, a las condiciones para estudiar y al apoyo económico familiar.

El modelo de diferencias permitió establecer un impacto positivo estadísticamente significativo en la calificación de los estudiantes que participaron en el programa “Aplicación de actividades de trabajo autónomo de Precálculo dirigida a los estudiantes de la CUMD- Soacha”, que resultó más realista al incluir los resultados del análisis multivariado.

BIBLIOGRAFÍA

- Aedo, C. (2005). Evaluación de Impacto. *Serie Manuales*, 47. Nueva York: Naciones Unidas, División de Desarrollo Económico, CEPAL.
- Añorga, J. & Valcárcel, N. (1999). *Diseño teórico-práctico del Modelo de Evaluación de Impacto. Informe parcial del Proyecto de Investigación sobre la evaluación de Impacto del Postgrado*. Cuba.
- Baker, J. L. (2000). *Evaluación de impacto de los proyectos de desarrollo en la pobreza. Manual para profesionales*. Washington, D.C.: Banco Mundial.
- Bernal, R. & Peña, X. (2011). *Guía práctica para la evaluación de impacto*, Bogotá: Ediciones Uniandes.

- Castro, R. (2008). Evaluación ex ante y ex post de proyectos de inversión pública en educación y salud. Metodologías y estudios de caso. *Documentos CEDE 2008-12*. Universidad de los Andes, Bogotá.
- Conchillo, J. A. & Ruiz, T. G. (1993). *Escalamiento Multidimensional: Una metodología de análisis en el campo de los factores humanos*. Madrid: Facultad de Psicología, Universidad Complutense de Madrid.
- Cuadras, C. M. (2010). *Nuevos Métodos de Análisis Multivariante*. Barcelona: CMC Editions.
- Imbens, G. & Lemieux, T. (2008). Regression Discontinuity Designs: A guide to practice. *Journal of Econometrics*, 142 (2), 15-635.
- Johnson, R. & Wichern, D. (2002). *Applied Multivariate Statistical Analysis*. Upper Saddle River, Nueva Jersey: Pearson Education.
- Linares, G. (1990). *Análisis de Datos*. La Habana (Cuba): ENPES, Ministerio de Educación Superior.
- Plazas, J. A., Lema, A. & León, J. D. (2009). Una propuesta estadística para la evaluación del impacto ambiental de Proyectos de desarrollo. *Revista Facultad Nacional de Agronomía*, 62 (1), 4937-4955.
- Sistachs, V., Joya, G. & Díaz, M. A. (2011). *Early Pigmentary Retinosis Diagnostic Based on Classification Trees*. En Advances in Computational Intelligence, 11th International Work-Conference on Artificial Neural Networks, LNCS 6691, p. 329.
- Valdés, M. (2008). *La evaluación de impacto de proyectos sociales: Definiciones y conceptos*. Recuperado el 5 de marzo de 2011 de: http://www.mapunet.org/documentos/mapuches/Evaluacion_impacto_de_proyectos_sociales.pdf
- Vara-Horna, A. (2010). Tendencias metodológicas de la evaluación de impacto de los Programas sociales: 1976-2006. *Innovación & Emprendimiento. Revista Latinoamericana de Ciencias Empresariales*, 1 (1), 69-88.