

## El gasto en educación en los hogares colombianos

Agosto 2022

### Abstract

El objetivo de este estudio es predecir el gasto de los hogares colombianos en educación en función a algunas de sus características socioeconómicas. A partir de la información que proporciona la Encuesta de Calidad de Vida 2021 (ECV) del DANE, se diseñó un modelo predictivo con dos finalidades. En primer lugar, entender qué características, de los hogares y sus integrantes, pueden tener mayor efecto en ese nivel de gasto. En segundo lugar, establecer cuál es el mejor modelo para estimar el gasto de los hogares en educación, con las limitantes de información que tienen las encuestas nacionales. Los resultados muestran que, el ingreso de la unidad de gasto, el contar con servicio de luz, un alto logro educativo del jefe del hogar y el número de integrantes en edad escolar tienen un impacto significativo en el aumento del gasto. Por el otro lado, el ejercicio demostró que para la muestra tomada y bajo las características analizadas, el árbol de regresión dio los mejores resultados en cuanto a la minimización del error cuadrático medio.

### Introducción

La formación académica y el rol que juega la educación en el desarrollo socioeconómico de un país no es menor. A pesar del aumento sostenido del gasto público en educación durante los últimos años en Colombia (3,5% del PIB en el 2000 a 4,7% en 2018), los niveles de cobertura aún son bajos, un fenómeno que afecta principalmente a los más vulnerables, la brecha en calidad y acceso entre hogares de bajos y altos ingresos y entre zonas urbanas y rurales es significativa, entre otras cosas. Por ejemplo, la tasa de asistencia escolar en niños entre los 0 y 5 años en 2021 fue de 46% para niños de hogares pobres y del 59% para no vulnerables. Bajo ese panorama, hoy, más que nunca, es importante conocer y analizar los determinantes socioeconómicos que podrían explicar el gasto o los recursos que destina un hogar a educación.

En Colombia, el énfasis sobre este tema en particular se ha puesto en el gasto público, algo que se puede medir en el tiempo. El aumento del presupuesto en educación en los últimos treinta años es significativo y ha tenido impactos importantes en indicadores de acceso, sobre todo en los niveles de primaria y secundaria. Sin embargo, no se ha ahondado lo suficiente en cómo los hogares invierten o consumen sus ingresos cuando se trata de rubros destinados a la educación. Por ejemplo, si ha habido cambios en las preferencias de consumo, cuánto pueden pesar bienes educativos en la canasta de un hogar de bajos ingresos y uno de altos, entre otras cosas. Las limitantes en términos de información son una barrera importante para este tipo de análisis. Por esa razón este estudio se pregunta ¿cuál es el mejor modelo predictivo para estimar el gasto de los hogares en educación?, teniendo en cuenta las restricciones de información existentes.

A propósito de esto, se tomó como referencia los estudios de Angus Deaton (2003), que profundizan en la importancia de las encuestas de hogares para la medición de ingresos y consumo de los hogares,

como base para la generación de mejores mediciones en términos de pobreza, índices de precios, bienes y servicios y diferencias en costos de vida según región y condición socioeconómica.

En ese sentido, también se tomó en cuenta el documento de Chi y Qian (2016), que estudia los cambios del gasto en educación de los hogares en China y sus factores asociados. El estudio concluyó que el gasto en educación aumenta desproporcionadamente cuando aumenta el ingreso del hogar. Por otro lado, Bayar y Ilhan (2016) examinan las determinantes del gasto y el impacto de los diversos grupos de ingresos en África y encuentran que, los hogares que perciben mayores ingresos y que tienen mejor capital humano, invierten más en educación. Tilak (2002) evaluó la elasticidad del gasto de los hogares de la India en educación y encontró que, independientemente de los estratos socioeconómicos, los individuos destinan una gran cantidad de sus ingresos en educación, en especial en el grado de primaria.

### Data

De acuerdo con la literatura analizada, la base de datos que se empleó fue la Encuesta Nacional de Calidad de Vida del 2021 (ECV 2021). Para la construcción de las variables del modelo se fusionaron los módulos de educación, trabajo, características del hogar, servicios del hogar y datos de la vivienda. Posteriormente, se filtró la base en función a los hogares que reportaron gastos en alguno de los rubros relacionados con educación, así se obtuvo una muestra final de 17.301 hogares con reporte efectivo.

#### *Variables del modelo y construcción:*

- Gasto en educación (dependiente): variable numérica en logaritmo. Se sumaron los gastos del hogar reportados en pensión educativa, útiles, alimentos, uniformes, transporte, crédito educativo.
- Miembros del hogar: Número de integrantes por hogar.
- Jefe del hogar: Variable dummy que toma el valor de “1” si el individuo es jefe del hogar y “0” si cumple cualquier otro rol dentro del núcleo familiar.
- Sexo del jefe del hogar: Variable dummy que toma el valor de “1” si el individuo es hombre y “0” si es mujer.
- Nivel educativo jefe del hogar: Variable multinomial que muestra el máximo nivel educativo alcanzado por el jefe del hogar.
- Área: Variable dummy que toma el valor de “1” si el hogar se encuentra en una zona urbana y “0” en zona rural.
- Estado civil jefe del hogar: Variable dummy que toma el valor de “1” si el individuo es casado o vive en convivencia y “0” en caso contrario.
- Edad del jefe del hogar: Variable numérica que expresa los años cumplidos del jefe del hogar.
- Ocupados en el hogar: Variable numérica que indica el número de ocupados por hogar.
- Servicio de electricidad: Variable dummy que toma el valor de “1” si el hogar posee servicio de luz y “0” en caso contrario.
- Número de personas en edad escolar: Variable numérica que cuantifica el número de personas en el hogar en edad escolar.

- Ingreso total del hogar: Variable numérica que cuantifica los ingresos mensuales totales del hogar.

Una primera exploración estadística de la base muestra cosas que vale la pena resaltar. Tal como se puede ver en la tabla 1, el gasto promedio en educación en hogares con jefatura femenina es menor en un 12% que el de los hogares con jefatura masculina. En el primer caso (mujeres) el gasto está por debajo del promedio de la muestra (\$200.970), mientras que en el segundo (hombres) está por encima. De igual forma, la tabla 2, se puede observar que el gasto en hogares urbanos es un 50% más alto que en hogares rurales.

Tabla 1.

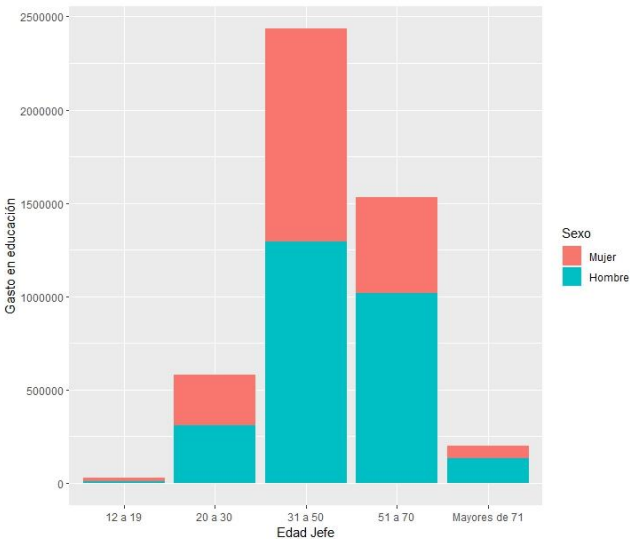
Sexo jefe hogar	Promedio de gasto en educación mes
Mujer	\$ 186.477,92
Hombre	\$ 212.787,71

Tabla 2.

Área	Promedio de gasto en educación mes
Rural	\$ 135.798,87
Urbano	\$ 247.912,87

Ahora, si se analiza el gasto en educación en relación con la edad del jefe del hogar (Gráfico 1), se puede ver que los gastos en educación son más altos en los hogares que tienen jefes entre los 31 y los 50 años. Eso puede deberse, entre otras cosas, a que son unidades de gasto con un mayor número de integrantes en edad escolar.

Gráfico 1. Gasto en educación por rango edad y sexo del jefe del hogar



Modelo

Para el diseño del modelo final se eliminaron las variables que no fueron estadísticamente significativas para la explicación de la dependiente (Gasto). Una vez hecho ese primer análisis, a continuación se expresa la forma funcional:

$$\ln(\text{Gasto}) = \beta_0 + \beta_1 \text{Escolares} + \beta_2 \text{Ocupados del hogar} + \beta_3 \text{Logro educativo jefe} + \beta_4 \text{Servicio eléctrico} + \beta_5 \text{Ingreso del hogar} + e_i$$

Resultados

Una vez planteado el modelo, los resultados de la regresión lineal mostraron que el ingreso de la unidad de gasto, el contar con servicio de luz, un alto logro educativo del jefe del hogar y el número de integrantes en edad escolar tienen un impacto significativo en el aumento del gasto en educación por hogar. Ahora, dado que el objetivo era predecir y por ende tratar de reducir la varianza del modelo, se emplearon distintos métodos de regularización. En ese sentido, se dividió la base de datos en una base de entrenamiento (70%) y testeo (30%) y se emplearon los métodos de: Predicción Lineal, Lasso, Ridge, Elastic Net y árboles de regresión. Los resultados del ejercicio muestran que el mejor modelo en términos de minimización del error cuadrático medio es el de árboles de regresión (Tabla 3).

Tabla 3.

Modelo	RMSE
MCO	1,3600
Lasso	1,3619
Ridge	1,3615
Elastic Net	1,3615
Árbol	1,3498

Una vez elegido el modelo se pasó a predecir en la base de testeo. Se diseñó una metodología de clasificación para evaluar la capacidad predictiva del modelo. Se calculó la diferencia entre el valor original del gasto en educación y el valor predicho en la base de testeo (*Dif*) y se creó un rango para ubicar las predicciones correctas y las incorrectas, todos los valores por fuera del rango se clasificaron como 1 y los restantes como 0.

$$Si\ Dif \geq \$150.000 \mid Si\ Dif \leq -\$150.000 \Rightarrow 1, 0$$

Los resultados muestran que, el 85% de los valores predichos fueron correctos o estaban dentro del umbral estipulado por el ejercicio. Por otro lado, solo el 15% de las predicciones se encontraron por fuera del rango. De modo que se podría decir, de manera preliminar y bajo las condiciones planteadas, la predicción se acerca a su valor original. Como muestra la tabla 4, el valor medio de la muestra original y el de los valores predichos son muy cercanos.

Tabla 4.

Valor medio muestra original	Valor medio predicción
\$ 70.000,00	\$ 65.560,00

## Conclusiones

- De las variables elegidas se emplearon, en función a su significancia, las siguientes: el número de integrantes del hogar en edad escolar, los ocupados en el hogar, el logro educativo del jefe del hogar, contar con servicio de energía y el ingreso del hogar.
- Por otra parte, de los métodos de regularización elegidos, el que mejor minimizó el error cuadrático medio fue el árbol de regresión.
- Una vez elegido el mejor modelo, realizada la predicción y establecido un criterio de clasificación de los valores predichos, los resultados muestran que hubo una correcta clasificación del 85% de las observaciones de la muestra.

## Repositorio

Repositorio GiHub donde se alojan los archivos:

[https://github.com/CamilaCiurlo/Educac\\_Gasto](https://github.com/CamilaCiurlo/Educac_Gasto)

## Referencias

Bayar, A., y Ilhan, B. (2016, 05). *Determinants of household education expenditures: Do poor spend less on education?* Topics in Middle Eastern and North African Economies, 18.

Chi, W., y Qian, X. (2016). *Human capital investment in children: An empirical study of household child education expenditure in China, 2007 and 2011*. China Economic Review, 37, 52-65.

Deaton, A. (2003, 02). *Household surveys, consumption, and the measurement of poverty*. Economic Systems Research, 15, 135-159.

Tilak, J. (2002, 07). *Determinants of household expenditure on education in India*. International Journal of Development Planning Literature, 17, 165-82.