
REGRESIÓN CUANTÍLICA APLICADA EN EDUCACIÓN EN COLOMBIA

Alfonso Javier Peñuela

alfonso.penuela@usa.edu.co

Universidad Sergio Arboleda (Colombia)

Asunto: Proyecto de Investigación

Temática: Estadística descriptiva e Inferencial

RESUMEN

Esta comunicación muestra una sinopsis del artículo “Valued-added in higher education: Ordinary least squares and quantile regression for a colombian case” (Bogoya, Bogoya, Peñuela, 2017). Se mencionan los datos usados, con una breve estadística descriptiva y luego se encamina a hablar de las metodologías de regresión usadas, a saber, regresión por mínimos cuadrados ordinarios y regresión cuantílica, en seguida se muestran los resultados obtenidos bajo las dos metodologías y se hace una rápida comparación para culminar con algunas conclusiones.

PALABRAS CLAVE

Saber11, SaberPro, Mínimos cuadrados, Regresión cuantílica.

INTRODUCCIÓN

Este trabajo recopila los resultados nacionales de las pruebas Saber Pro 2009 para los programas de administración de empresas con más de 20 estudiantes que presentaron dicho examen, a cada uno de estos estudiantes se le efectuó la respectiva correspondencia de su resultado en la prueba Saber 11 así como el género y estrato socioeconómico, obteniendo una base de datos con 10.783 registros.

En estos datos se toma como variable dependiente los resultados de la prueba Saber Pro y los resultados de la prueba Saber 11, el Estrato y el Género como variables dependientes. Se hace una revisión básica de manera descriptiva de las variables y se construyen dos modelos, uno de regresión lineal y otro de regresión cuantílica. Teniendo en cuenta que la regresión por mínimos cuadrados modela el impacto lineal que tienen las covariables en la media condicional de la variable independiente, esto implica subestimaciones y sobreestimaciones de dichos impactos. La regresión cuantílica a partir de estimar los parámetros de sumas ponderadas superiores e inferiores de las desviaciones

absolutas muestra los impactos, pasando por los diferentes cuantiles. De manera gráfica se presentan los dos resultados mostrando las diferencias y similitudes en las dos técnicas.

MARCO DE REFERENCIA

En Colombia se presentan dos exámenes de estado obligatorios denominados Saber 11 y Saber Pro, el primero de ellos se presenta finalizando el bachillerato (la educación media) y el otro finalizando los estudios universitarios (después de haber cursado el 70% de los créditos de la carrera). Estos exámenes miden conocimientos en diferentes áreas del saber, y son en cierto modo una medida de la calidad de la educación nacional. El primero de estos exámenes es requisito obligatorio para acceder a la educación superior, el segundo aún no es pre requisito para inserción en el mundo laboral ni en estudios de posgrado, pero si para graduarse de profesional. Se toman los resultados del examen Saber Pro para la carrera Administración de empresas y a cada individuo se le buscó su respectivo resultado en el examen Saber 11 para tenerlo como proxy de dotaciones iniciales académicas y su estrato como proxy de nivel socioeconómico, así como el género que ha mostrado ser en ocasiones un elemento diferenciador.

Con estos datos, el trabajo consiste en explicar el resultado de las pruebas Saber Pro por medio de las otras tres variables. Esta explicación se hace por medio de la regresión lineal básica bajo un modelo de estimación por mínimos cuadrados ordinarios, estudiando los supuestos, luego de este proceso se identifican los impactos en la media. Seguidamente, los datos se toman como referencia para usar un modelo donde se puedan ver los impactos en los diferentes cuantiles en el examen Saber Pro, terminando con los resultados numéricos y gráficos de los dos modelos, y se procede a hacer su comparación.

DESARROLLO DEL TEMA

Aspectos metodológicos

La información fue recolectada a partir de cruzar algunas bases de datos publicadas por el Instituto Colombiano para el Fomento de la Educación Superior [ICFES]. Este trabajo surgió a partir de las ideas iniciales sobre valor agregado de la educación superior en Colombia (Bogoya & Bogoya, 2013) donde el problema se aborda desde una perspectiva meramente lineal buscando una medida de valor agregado sin hacer ninguna comparación con otro posible modelo. En este trabajo el enfoque es más comparativo de la estimación clásica con la propuesta econométrica de regresión cuantílica (Koenker & Bassett, 1978).

Desarrollo de la propuesta

Se inicia con el planteamiento del problema como motivación y para despertar el interés del público. Una vez explicado el problema se procede a describir los datos con los que se realizó el trabajo, se presentan algunos asuntos de la estadística descriptiva sin

entrar en demasiados detalles. En la Tabla 1 se presentan algunos valores descriptivos importantes de la población.

| | | | | |
|----------------|-----------|-----|-----------|-----|
| Hombres 39% | Estrato 1 | 7% | Estrato 4 | 16% |
| | Estrato 2 | 26% | Estrato 5 | 10% |
| Mujeres 61% | Estrato 3 | 35% | Estrato 6 | 6% |

Tabla 1. Datos descriptivos de la población
Fuente. Creación propia

Se procede a dar una breve explicación de las técnicas econométricas en el contexto de Mínimos Cuadrados Ordinarios y Regresión Cuantílica.

Sean:

Y : Saber Pro es un vector de 10.783×1 .

X : La matriz de diseño conformada por las variables explicativas y el vector de unos, el tamaño de esta matriz es de 10.783×4 .

β, E , son el vector de coeficientes y el vector de términos de perturbación respectivamente.

Para el primer modelo, en su sentido básico toma la forma:

$$Y = X'\beta + E$$

Esta metodología busca minimizar una función objetivo, constituida por la suma de los errores cuadráticos de los residuos, y luego del proceso de derivación y verificados los supuestos clásicos, se obtienen las estimaciones por medio de la operación matricial:

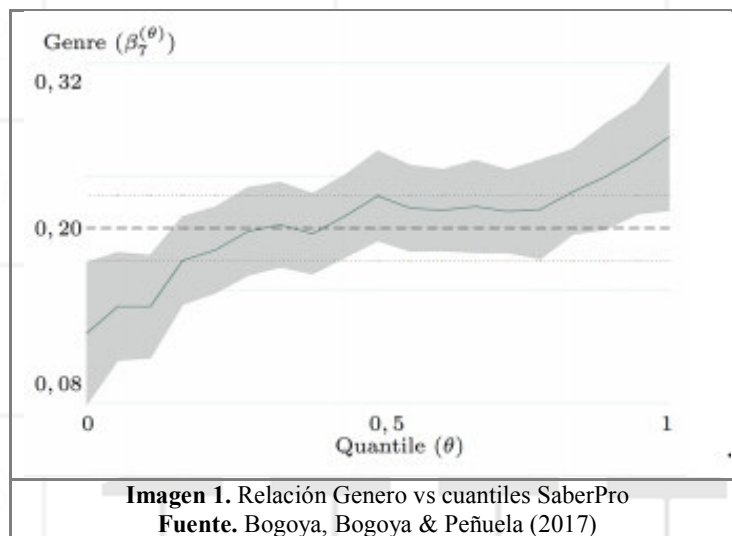
$$\tilde{\beta} = (X'X)^{-1}X'Y.$$

Estos valores miden el impacto lineal de las covariables en la media esperada de la variable dependiente (resultados examen Saber Pro). Si bien es cierto la metodología explicada puede cumplir los supuestos requeridos propios de ella en su forma estadística, finalmente es una estimación de la media que puede llevar a subestimaciones y sobre estimaciones en algunos segmentos de la población. La regresión cuantílica (Koenker & Bassett, 1978) no busca minimizar la suma cuadrática de los residuos sino la suma ponderada de las desviaciones absolutas a través de los diferentes cuantiles.

$$\min(\beta) \left[\sum_{Y_t \geq X_t\beta}^n \theta |Y_t - X_t\beta| + \sum_{X_t\beta \geq Y_t}^n (1 - \theta) |Y_t - X_t\beta| \right]$$

Dicha función objetivo al no ser diferenciable no puede encontrar sus soluciones por medio de la derivación y hallar una expresión cerrada como en el caso de MCO. Para solucionar dicho problema se hace uso de diferentes métodos numéricos (Konker & Hallock, 2001), siendo que el software estadístico ya los tiene implementados (Cameron & Trivedi, 2009).

La Imagen 1 relaciona la estimación del género a través de los cuantiles de los resultados del examen Saber pro. Se puede observar la sobreestimación de MCO para los cuantiles bajos y la subestimación para cuantiles altos.



La Imagen 2 muestra resultados de mínimos cuadrados que se compararán con el método de regresión cuantílica.

| Saber Pro | β_t | σ_t | 95 % conf. int. |
|-----------|-----------|------------|-----------------|
| Saber 11 | 0,65 | 0,008 | [0,64; 0,67] |
| Stratum 2 | 0,15 | 0,030 | [0,095; 0,21] |
| Stratum 3 | 0,20 | 0,030 | [0,14; 0,26] |
| Stratum 4 | 0,24 | 0,033 | [0,18; 0,31] |
| Stratum 5 | 0,26 | 0,036 | [0,19; 0,33] |
| Stratum 6 | 0,13 | 0,041 | [0,05; 0,21] |
| Gender | 0,15 | 0,015 | [0,13; 0,18] |

Imagen 2. Resultados de mínimos cuadrados
Fuente. Bogoya, Bogoya & Peñuela (2017)

CONCLUSIONES

Las conclusiones más relevantes tienen que ver con:

1. Las variables resultan ser significativas en la explicación de los resultados de saber pro.
2. El impacto de saber 11 en los cuantiles más bajos de saber pro es sobre estimado en el modelo clásico de regresión lineal. Con lo cual se podrían tomar decisiones a partir de una importancia más ponderante a Saber 11, con las personas que ocupan los lugares más bajos en sus resultados.
3. El género tiene implicaciones directas sobreestimando el impacto en las personas de menores resultados y subestimando el impacto en los resultados superiores. El hecho de ser hombre hace concluir de acuerdo a los modelos unos resultados mayores comparados con ser mujer.
4. El estrato socioeconómico uno presenta una especie de “lastre” de tipo académico comparado con todos los otros. Cuando se controla por otro estrato no hay diferencias tan marcadas.

REFERENCIAS

- Bogoya, J. D., Bogoya, J. M. & Peñuela A. J. (2017). Value-added in higher education: ordinary least squares and quantile regression for a Colombian case. *Ingeniería e Investigación*, 37(3), 30-36.
- Bogoya, J. D. & Bogoya, J. M. (2013). *An academic valueadded mathematical model for higher education in Colombia*. *Ingeniería e Investigación*, 33(2), 76-81.
- Koenker, R. & Bassett, G. (1978). *Regression quantiles*. *Econometría*, 46(1), 33-50.
- Koenker, R. & Hallock, K. (2001). *Quantile regression*. *Journal of Econometrics Perspectives*, 15(4), 143-156.
- Cameron, A. C. & Trivedi, P. K. (2009). *Microeconometrics using Stata*. Stata Press, Texas 2009.