

Regresión de borde

Por: Luis Pérez G.*
Germán Cabarcas J.*

RESUMEN

Se introduce el criterio de multicolinealidad agregada y el efecto que sobre ella ejerce la estandarización. Dentro de la familia de los estimadores sesgados para un modelo de regresión, se presentan los estimadores de borde de una manera detallada en los diversos aspectos analítico, gráfico, geométrico, de uso práctico y de existencia.

1.0 INTRODUCCION

Considere el modelo de regresión lineal estandarizado:

$$\underline{Y} = \underline{X} \underline{b} + \underline{\varepsilon}$$

donde \underline{Y} es un vector $n \times 1$ de observaciones; \underline{X} es una matriz $n \times p$ de rango p cuyas columnas han sido ya estandarizadas; \underline{b} es un vector $p \times 1$ compuesto por parámetros desconocidos b_1, b_2, \dots, b_p ; y $\underline{\varepsilon}$ es un vector de variables aleatorias independientes y normales con media cero y matriz de covarianzas $\sigma^2 I_n$. Un vector \underline{X}_j se estandariza mediante la resta a cada uno de sus elementos del valor promedio, y luego de esta centralización se divide cada uno de sus valores por la desviación estándar multiplicado por $\sqrt{n-1}$; observe que si \underline{X}_j está estandarizado se cumple que $\underline{1} \cdot \underline{X}_j = 0$, y, $\underline{X}_j' \underline{X}_j = 1$.

En la adaptación de los modelos lineales a situaciones prácticas abundan las dificultades si el deseo es construir un modelo efectivamente adecuado; una dificultad común es el problema de la multicolinealidad para el cual no se ha encontrado hasta la fecha un remedio absolutamente contundente. La multicolinealidad es un problema intrínseco en los datos y se detecta por la existencia de dependencia lineal entre dos o más vectores columna

de la matriz \underline{X} . Curiosamente, se presentan diferencias respecto al concepto de dependencia lineal que se utiliza en Algebra Lineal con el usado en Estadística. Si X_1, X_2, \dots, X_p son las variables columna de \underline{X} , por Algebra Lineal se dice que son linealmente dependientes si y sólo si existen reales a_1, a_2, \dots, a_p (diferentes de cero) tales que:

$$a_1 X_1 + a_2 X_2 + \dots + a_p X_p = 0. \quad (1)$$

En Estadística sin embargo existe tal dependencia cuando la igualdad (1) es cierta o aproximadamente cierta. Bajo este criterio se presenta la multicolinealidad como una enfermedad medible en escalas donde el caso (1) es el caso extremo; si el coeficiente de correlación entre las variables es alto la dependencia será alta, y si es bajo la multicolinealidad será asimismo baja.

Las causas de multicolinealidad son múltiples; entre ellas pueden citarse las siguientes:

1. Sobredefinición del modelo; esto es, relativamente más variables que observaciones.
2. Defectos en la recolección de la información en el sentido de muestrear sólo un subespacio total de variables independientes.
3. Restricciones físicas en el modelo o en la población donde pueden presentarse efectos entre las variables consideradas. Como ejemplo note que la temperatura afecta intrínsecamente la presión y viceversa.

Los efectos de la multicolinealidad sobre el modelo son altamente perjudiciales. El modelo se torna inestable y la estimación de los parámetros b_1, \dots, b_p por el método de los mínimos cuadrados es completamente inadecuada pues a pesar de ser insesgados su varianza es considerablemente alta. Estos efectos se ven matemáticamente al recordar que el estimador de los mínimos cuadrados $\hat{\underline{b}}$ es:

$$\hat{\underline{b}} = (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{Y}$$

*Profesores Facultad de Ciencias, Universidad Nacional, Medellín.

con matriz de covarianzas:

$$V(\hat{\underline{b}}) = \sigma^2 (X^T X)^{-1},$$

y con varianzas individuales:

$$V(\hat{b}_i) = \sigma^2 (1 - R_i^2)^{-1},$$

donde R_i es el coeficiente de determinación parcial entre la variable i y las demás variables independientes. En el caso de multicolinealidad los R_i son cercanos a 1 y esto hace las varianzas considerablemente altas y la matriz $(X^T X)$ tiende a tener características de singularidad.

Son múltiples los remedios propuestos para resolver el problema de multicolinealidad sin obtener siempre resultados completamente satisfactorios. Un intento moderno es el de buscar estimadores sesgados que sean garantía de una varianza pequeña; esto es, modernamente hay tendencia a preferir un sesgo y una varianza pequeña sobre un estimador de sesgo 0 y alta varianza.

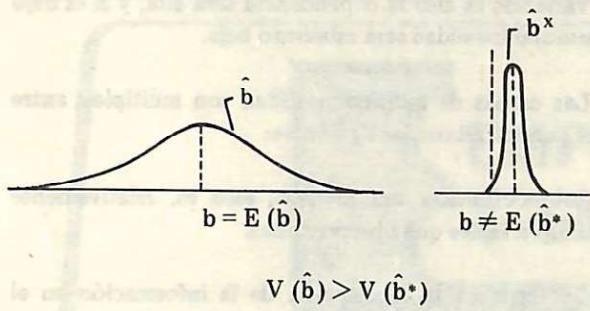


Figura 1

Dentro de la familia de estimadores sesgados aparecen con gran aplicabilidad práctica los estimadores de borde alrededor de los cuales girará todo el artículo. Antes de iniciar el estudio detallado, es de gran conveniencia práctica introducir el criterio de multicolinealidad agregada y diferenciarlo del de multicolinealidad propiamente dicha o esencial. La multicolinealidad agregada surge del efecto de adicionar al modelo variables que son a su vez funciones de las variables iniciales, y del efecto del uso de escalas de valores arbitrarios; adicionar variables cuadráticos e interacciones es traer multicolinealidad agregada al modelo. Tal tipo de multicolinealidad se reduce casi en su totalidad mediante la centralización y/o estandarización de las variables pues se disminuye el factor de

inflación de la varianza y se remueve totalmente la correlación entre el término constante y todos los otros términos lineales. Se deduce entonces que la multicolinealidad agregada es un asunto de más fácil atención, y que es esencial la que necesita tratamiento más complejo y para la que se proponen los estimadores de borde.

2.0 CONCEPTOS BASICOS

Los estimadores de borde propuestos por Hoerl y Kennard (1970) para controlar la varianza alta y la inestabilidad del modelo de regresión son del tipo:

$$\hat{\underline{b}}^* = (X^T X + k I_p)^{-1} X^T Y$$

donde k es una constante positiva para determinar. A diferencia de $\hat{\underline{b}}$, $\hat{\underline{b}}^*$ no es un estimador insesgado como se puede ver en su esperanza y varianza transcritas a continuación:

$$E(\hat{\underline{b}}^*) = (X^T X + k I)^{-1} X^T X \underline{b}$$

y,

$$V(\hat{\underline{b}}^*) = (X^T X + k I)^{-1} X^T X (X^T X + k I)^{-1} \sigma^2.$$

Su error en media cuadrática $MSE(k)$ está dado por:

$$E[(\hat{\underline{b}}^* - \underline{b})^T (\hat{\underline{b}}^* - \underline{b})] = \sigma^2 \sum \lambda_i (\lambda_i + k)^{-2} + k^2 \underline{b}^T (X^T X + k I)^2 \underline{b}$$

con $\lambda_1, \lambda_2, \dots, \lambda_p$ como valores propios de $X^T X$, y de los cuales algunos de ellos tienden a cero si hay síntomas de multicolinealidad. Conviene notar que el primer término de la derecha es la varianza total y el segundo es el sesgo al cuadrado; el primero se denominará por $f(k)$ y el segundo por $g(k)$.

La idea central de la regresión de borde es escoger un valor k tal que la reducción en varianza compense el sesgo en el cual se incurre.

De la fórmula dada para $\hat{\underline{b}}^*$ se deduce que coincide con $\hat{\underline{b}}$ si $k = 0$, y que $\hat{\underline{b}}$ es una transformación lineal de $\hat{\underline{b}}$ ya que:

$$\hat{\underline{b}}^* = Z \hat{\underline{b}}$$

$$\begin{aligned} \text{donde } Z &= (X^T X + k I)^{-1} X^T X = I - k (X^T X + k I)^{-1} \\ &= I + k (X^T X)^{-1}. \end{aligned}$$

La inclusión del k trae novedades en los valores de los parámetros. En la Figura 2 se presenta una situación práctica (2) donde por medio de un modelo lineal se relacionan las importaciones de Francia con las variables consumo interno, producción nacional bruta y formación de inventario. Allí se ve que para cada k existe un valor de \hat{b}^* que puede incluso cambiar de signo al valor del parámetro original lo cual cambiaría completamente la interpretación práctica y el efecto propiamente dicho. Usando los mínimos cuadrados ($k = 0$) la producción nacional (b_1) tiene un efecto negativo sobre las importaciones; usando $k = .2$ se tiene un estimador de borde con efecto positivo sobre las mismas!

3.0 COMPARACION DE VARIANZAS

Se dijo que los estimadores de borde se definen para un k real positivo tal que el estimador obtenido establece el sistema y ofrezca una varianza menor que la de los mínimos cuadrados y justifique el sesgo.

Como σ^2 es un parámetro desconocido, en toda comparación de varianzas debe darse un término de referencia que sirva para tal fin. El error medio cuadrático (MSE), definido respectivamente para el estimador de los mínimos cuadrados y para el estimador de borde como:

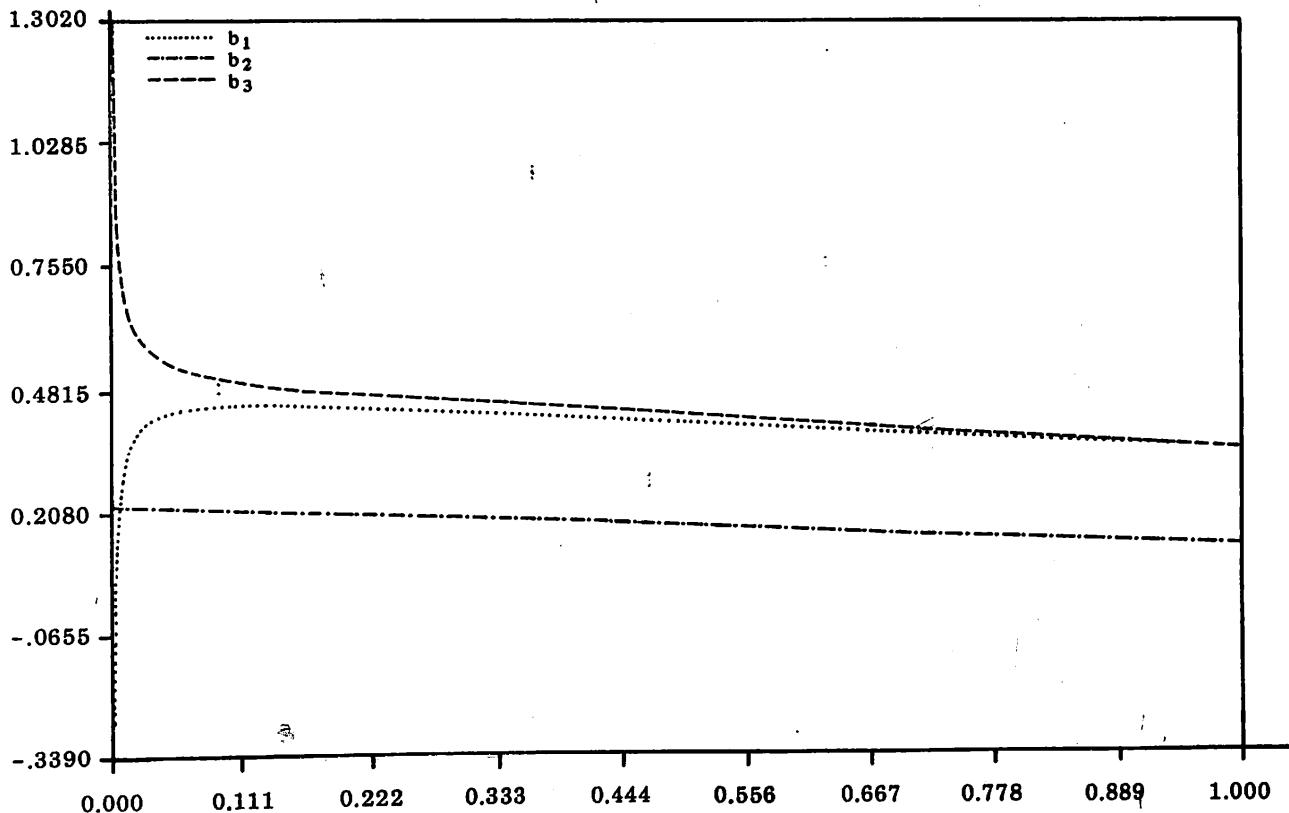


Figura 2

$$MSE(0) = E[(\hat{b} - b)^T (\hat{b} - b)]$$

$$= \sigma^2 \text{traza}(X^T X)^{-1} = \text{constante}$$

y,

$$MSE(k) = E[(\hat{b}^* - b)^T (\hat{b}^* - b)]$$

$$= \sigma^2 \sum \lambda_i (\lambda_i + k)^{-2} + k^2 b^T (X^T X + kI)^{-2} b$$

ha sido universalmente usado para estimar y comparar varianzas y desde este punto de vista se mostrará cómo el MSE de \hat{b} es menor que el de \hat{b}^* .

Una impresión gráfica es siempre importante y es así como la Figura 3 da la relación deseada. En ella se observan cuatro curvas. La primera representa el MSE de los mínimos cuadrados el cual es constante o independiente de k . Las dos siguientes se refieren a las funciones $f(k)$ y $g(k)$ componentes del MSE del estimador de borde. Como:

$$f(k) = \sigma^2 \sum \lambda_i (\lambda_i + k)^{-2},$$

es $f(k)$ función continua y monótona decreciente de k ,

su derivada respecto a k tiende a $-\infty$ a medida que k va a 0^+ y si $X^T X$ es mal condicionada o lo que es lo mismo si λ_p tiende a 0. Para:

$$g(k) = k^2 \underline{b}^T (X^T X + k I)^{-2} \underline{b}$$

se observa una función continua y monótona creciente de k . La cuarta curva es la punteada y es la correspondiente al MSE del estimador de borde dado por la suma de f y g . En la línea punteada de la figura para diversos valores de k , $MSE(k) < MSE(0)$; y existe además un mínimo de alguna manera tentador!

Aunque muchos autores restringen a k en $(0,1)$, esto no es absolutamente necesario. El parámetro k podría tomar cualquier valor positivo pero, sin embargo, no es de interés considerarlo a partir del momento en el cual $MSE(k) > MSE(0)$ —como se nota en la gráfica— ni cuando su valor es grande debido a que muy posiblemente el sesgo no compensaría la disminución en varianza.

4.0 EXISTENCIA Y CALCULO DE k .

Aunque en la figura está clara la bondad del estimador de borde, se hace necesario mostrar que efectivamente existe un valor k que logre alcanzar un mínimo similar al de la línea punteada y que sea posible su cálculo.

Hoerl y Kennard (1970) probaron que:

“Siempre existe un $k > 0$ tal que:

$$MSE(k) < MSE(0) = \sigma^2 \sum (\lambda_j)^{-1}.$$

A pesar de ello, como ocurre con frecuencia en matemáticas, se logró encontrar la existencia del k pero no el valor exacto ni la manera precisa de calcularlo. Las dificultades para este cálculo son obvias pues k es a su vez función complicada de parámetros desconocidos. El asunto del cálculo de k tiene abiertas muchas discusiones teóricas y son diversos los métodos existentes de cálculo aproximados para el estimador de borde y que por razones de espacio y de objetivos se omitirán aquí. Los interesados pueden referirse a Cabarcas y Pérez (1984), quienes recogen críticamente los métodos existentes y proponen además algunas posibilidades nuevas de trabajo basadas en el número condicional de una matriz.

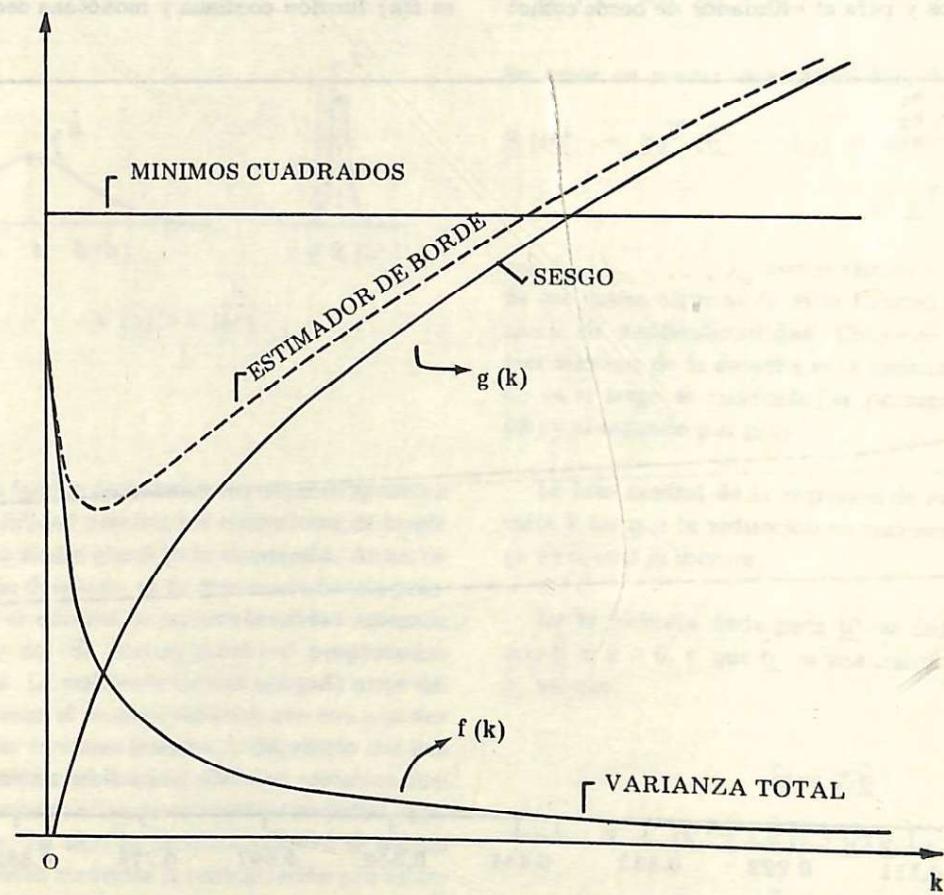


Figura 3

5.0 INTERPRETACION GEOMETRICA DE LOS ESTIMADORES.

En un modelo de **regresión** lineal, la suma de los errores al cuadrado se **utiliza** para hallar el estimador de $\underline{b} = (b_1, b_2, \dots, b_p)$. En el espacio de los parámetros o espacio geométrico generado por b_1, b_2, \dots, b_p , la función de la suma de los cuadrados —que es función sólo de \underline{b} — puede ser representada por los contornos de varias superficies elipsoidales donde el centro de la familia de esos elipsoides corresponde al estimador de mínimos cuadrados como se nota en la Figura 4 para el caso bidimensional.

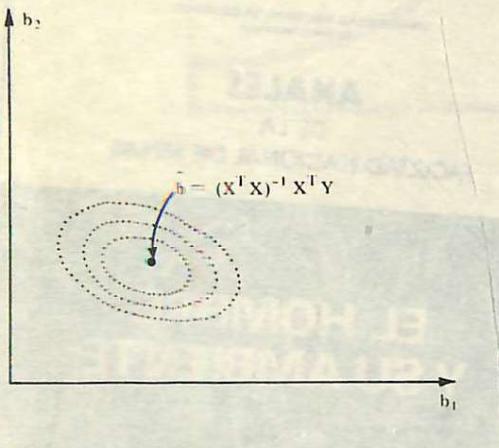


Figura 4

Con el fin de lograr la interpretación geométrica del estimador de borde es necesario notar que éste es equivalente al estimador de mínimos cuadrados sujeto a una restricción esférica sobre los parámetros. En símbolos, si se trata de minimizar

$$(X^T X) \underline{b} = X^T Y$$

Sujeta a:

$$\underline{b}^T \underline{b} \leq c^2 \text{ con } c^2 \text{ conocido;}$$

usando el **método** de multiplicadores de Lagrange y hallando la derivada respecto a \underline{b} , este problema de minimización es equivalente a resolver:

$$(X^T X + k I) \underline{b} = Z^T Y$$

lo cual es la **solución** de borde y sujeta a $\underline{b}^T \underline{b} = c^2$.

Puede entonces verse la regresión de borde como el uso del **método** de los mínimos cuadrados restringido a

la esfera con origen en 0 y radio \hat{b}^* . En la Figura 5, \hat{b} es el estimador de los mínimos cuadrados, \hat{b} es el estimador de borde el cual corresponde al punto donde la esfera de radio \hat{b} corta una elipse del espacio de los parámetros. Además como:

$$S(\hat{b}) = (Y - X \hat{b})^T (Y - X \hat{b})$$

y,

$$\hat{b} = (X^T X)^{-1} X^T Y = (X^T X)^{-1} \underline{h} \text{ con } \underline{h} \text{ gradiente de } S(\hat{b})$$

tal gradiente \underline{h} es perpendicular al espacio de $S(\hat{b})$ a través del origen. Observe que el estimador de borde está siempre entre \hat{b} y \underline{h} ; y el ángulo β entre el vector \hat{b}^* y el gradiente \underline{h} se vuelve cada vez menor a medida que se aumenta k .

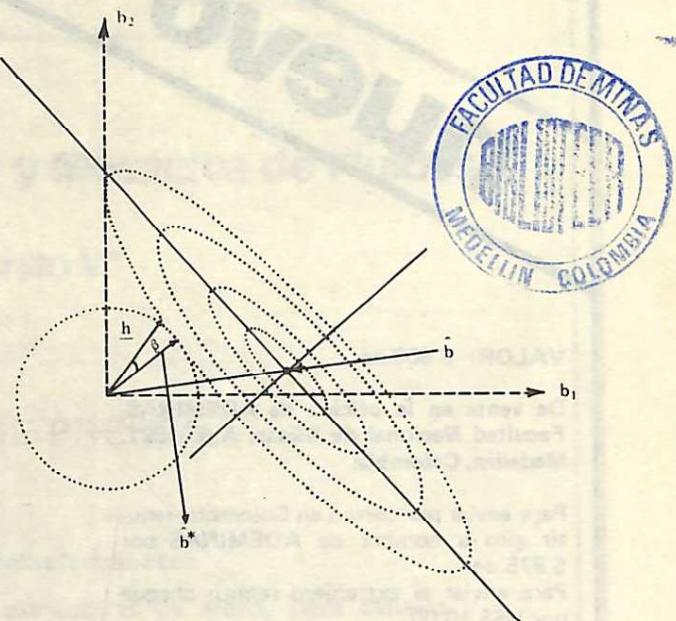


Figura 5

6.0 NOTA FINAL

La regresión de borde es una alternativa real para la modelación matemática. Aunque todavía su uso no está generalizado, muchas veces por desconocimiento o por no entrar en complejidades, constituye innegablemente este método científico una herramienta novedosa que cada vez encuentra más simpatizantes debido a las posibilidades teóricas y prácticas que ofrece.

La regresión de borde aireó el ambiente teórico de los modelos lineales en la década del 70 y muy seguramente la década del 80 será una posibilidad para extender su uso en la práctica ayudado por los métodos computacionales más modernos.

REFERENCIAS

1. CABARCAS, G. y PEREZ, L. Cálculo de los estimadores de borde. Congreso Nacional de Estadística. Universidad de Medellín. 1984.
2. CHATTERJEE, S. and PRICE, B. Regression analysis by example. Wiley. 1977.
3. DRAPER, N. and SMITH, H. Applied regression analysis. Second edition. Wiley. 1981.
4. GOLDSTEIN, M. and SMITH, A. Ridge type estimators for regression analysis. Journal of Royal Statistical Society. B-36. Págs. 284-91. 1974.
5. HOERL, A. and KENNARD, R. Ridge regression: Biased estimation for non orthogonal problems. Technometrics. Volume 12. Págs. 55-67. 1970.
6. MARQUARDT, D. and SNEE, R. Ridge regression in practice. American Statistician. 29. 3-19. 1975.
7. MASON, R., GUNST, R. and WEBSTER, J. Regression analysis and problems of multicollinearity. Communications in Statistics. 4 (3). 277-92. 1975.

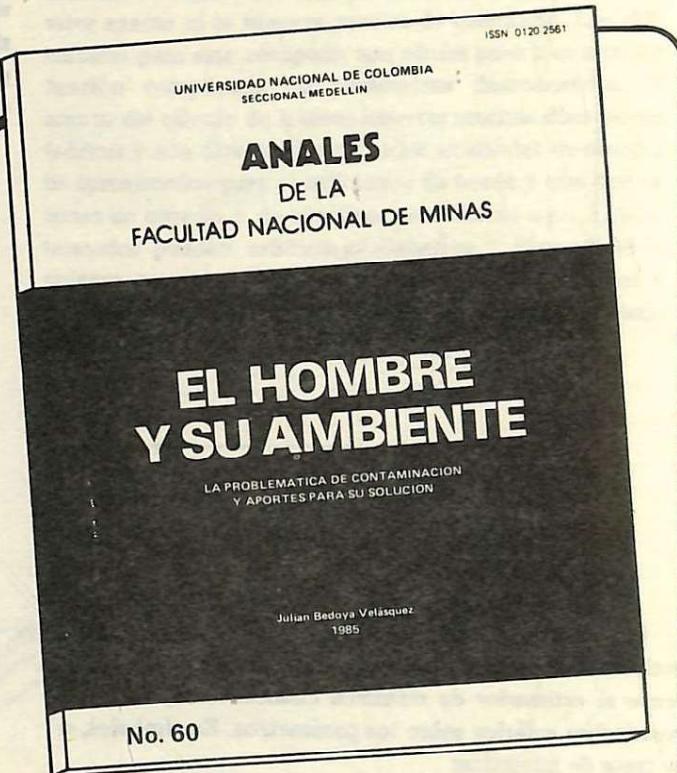
nuevo

VALOR: \$ 800.00

De venta en la oficina de ADEMINAS,
Facultad Nacional de Minas, A.A. 1027.
Medellín, Colombia.

Para enviar por correo en Colombia, remitir giro a nombre de ADEMINAS por \$ 875.00

Para enviar al extranjero remitir cheque por US\$ 10.00



OTROS NUMEROS DE ANALES DE VENTA EN LA FACULTAD:

No.	NOMBRE:	VALOR
40	Análisis Químico y Análisis Petrográfico de las Rocas de Antioquia. Antonio Durán	\$ 100.00
44	La Flotación. Robert Wokitel	\$ 100.00
46	Pisos para carreteras colombianas. Gabriel Hernández Salazar	\$ 150.00
48	Grado Colectivo de 1939.	\$ 150.00
54	Ecuaciones diferenciales. Jorge Mejía Ramírez	\$ 100.00
56	Cálculo vectorial. Luis de Greiff Bravo	\$ 100.00
59	Mineralogía determinativa (Ensayos Químicos). Gabriel Trujillo Uribe, 1984	\$ 450.00

Para envíos por correo en Colombia, adicionar \$ 50.00 por cada ejemplar solicitado.