Historia

Sección a cargo de

Luis Español González

Un antecedente histórico de regresión lineal: la estimación mediana propuesta por Boscovich

por

María Dolores Pérez Hidalgo, Jesús Basulto Santos y José Antonio Camúñez Ruiz

Introducción

El análisis de regresión es una de las técnicas estadísticas más utilizadas en la actualidad cuando se intenta relacionar una variable objetivo con una o varias variables explicativas. Así, por ejemplo, si Y es la variable objetivo, o variable a modelizar, o variable explicada (variable efecto en una relación del tipo causa-efecto) y X es una variable explicativa (ilustramos con el caso más simple de una sola variable explicativa), entonces los clásicos introducen la regresión lineal como la estimación de los parámetros de la ecuación $y_i = \alpha + \beta x_i + u_i$, $i = 1, \dots, n$, donde x_i e y_i son los valores observados de las dos variables que intervienen en el análisis (X e Y, respectivamente), para las n observaciones o los n elementos muestrales; u_i es una variable aleatoria latente, no observada, que garantiza la igualdad en la ecuación anterior y recibe el nombre de perturbación aleatoria; por último, α y β son dos parámetros desconocidos que establecen la linealidad de la relación y cuyas estimaciones, $\hat{\alpha}$ y $\hat{\beta}$, se desean alcanzar a partir de los datos muestrales observados. Lógicamente, se busca la mejor estimación posible. O sea, las estimaciones en las que haya la máxima garantía de proximidad a los verdaderos valores de los parámetros que, no olvidemos, son desconocidos. Si tuviésemos los valores estimados de esos parámetros, para cada valor muestral podríamos construir un valor «ajustado» de la variable explicada, de la siguiente forma: $\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$, $i = 1, \dots, n$. De tal manera que, asociado a cada elemento muestral u observación i, tenemos dos valores relacionados con la variable explicada: el observado y_i y el ajustado $\hat{y}_i$. La diferencia entre ambos es el error o residuo asociado a la observación:

$$e_i = y_i - \hat{y}_i = y_i - \hat{\alpha} - \hat{\beta} x_i.$$

Pues bien, la estimación $\hat{\alpha}$ y $\hat{\beta}$ de los parámetros se consigue minimizando la suma de cuadrados de residuos, o sea, resolviendo este problema de optimización:

$$\text{mín} \sum_{i=1}^{n} (y_i - \hat{\alpha} - \hat{\beta} x_i)^2.$$

El método, conocido como estimación por mínimos cuadrados, es atribuido a Legendre [7] y data de 1805. En la segunda mitad del siglo XVIII, astrónomos, geodestas y otros científicos prácticos se vieron obligados a perfeccionar sus propios métodos para resolver sistemas de ecuaciones en los que había más observaciones que incógnitas. Desde los matemáticos del Renacimiento hasta 1750, la única solución posible a este problema fue procurar que hubiera tantas ecuaciones como incógnitas. En 1750, Tobias Mayer sugirió que las incógnitas deberían determinarse estableciendo ciertos sistemas de ecuaciones homogéneas. Una aportación casi definitiva fue el método de mínimos cuadrados de Legendre, aunque su argumento fue completamente algebraico, sin contenido estadístico. Posteriormente, en la sección no estadística de su *Theoria Motus Corporum Coelestium* de 1809 [4], Gauss reprodujo esencialmente las deliberaciones de Legendre, pero usando la suma de las cuartas potencias, de las sextas potencias, etc., como posibles alternativas a la suma de los errores al cuadrado. De estas propuestas, la más simple es la suma de los errores al cuadrado, que Gauss dice que había usado en cálculos prácticos desde 1795, para gran disgusto de Legendre, que publicó una respuesta venenosa en 1820 (véase [13]).

La solución al problema de optimización antes planteado es una media. Los valores estimados $\hat{\alpha}$ y $\hat{\beta}$ proporcionan la mejor estimación posible de esta media o esperanza, $E(Y|x_i) = \alpha + \beta x_i$, bajo la condición de que las perturbaciones aleatorias u_i cumplan unas determinadas hipótesis estocásticas.

Al inicio de la descripción de las técnicas relacionadas con el método de mínimos cuadrados, Mosteller y Tukey, en su influyente texto [8], señalan:

> Lo que la curva de regresión hace es dar un gran resumen de los promedios de las distribuciones correspondientes al conjunto de las X. Podríamos ir más lejos y calcular varias curvas de regresión diferentes correspondientes a los diferentes puntos porcentuales de las distribuciones y así obtener una imagen más completa del conjunto. Normalmente esto no se hace y, por lo tanto, la regresión a menudo da una imagen bastante incompleta. Así como la media da una imagen incompleta de una sola distribución, la curva de regresión da una imagen incompleta para un conjunto de distribuciones.

¿Por qué la estimación por mínimos cuadrados del modelo de regresión lineal impregna las estadísticas aplicadas? ¿Qué la convierte en una herramienta exitosa? Como establece Koenker [6], tres posibles respuestas se sugieren. No se debe descartar el hecho obvio de que el método genera estimadores lineales, cuyo manejo computacional es extremadamente atractivo. Seguramente este fue el impulso inicial para su éxito. En segundo lugar, si los residuos de las observaciones se distribuyen normalmente, esto es, de modo gaussiano, se sabe que los métodos de los mínimos cuadrados gozan de cierta optimalidad. Pero, como lo fue para el propio Gauss, esta respuesta a menudo parece ser una racionalización *ex post* destinada a reemplazar a la primera respuesta. Más convincente es la observación relativamente reciente de que los métodos de mínimos cuadrados proporcionan un enfoque general para estimar funciones de medias condicionadas.

Y, sin embargo, como Mosteller y Tukey sugieren, la media rara vez es un fin satisfactorio en sí mismo, incluso para el análisis estadístico de una sola muestra. Medidas de dispersión, asimetría, curtosis, diagramas de cajas, histogramas y estimación de densidad más sofisticada se emplean frecuentemente para obtener mayor información. ¿Puede algo similar hacerse en la regresión? Un punto de partida natural para esto sería complementar las medias condicionales estimadas por mínimos cuadrados con estimaciones de varios cuantiles condicionados y, en particular, con la estimación de la mediana condicionada. Tendríamos así una información mucho más completa sobre el comportamiento de la variable explicada cuando se condiciona a los valores de la variable explicativa. Boscovich lo hizo medio siglo antes de la aparición del método de mínimos cuadrados con respecto a la estimación de la mediana condicionada, y es lo que ha dado pie a lo que hoy se conoce como regresión cuantil.

Pretendemos en este trabajo describir cómo Boscovich se convirtió en uno de los pioneros del análisis de regresión, con un método antecedente al de mínimos cuadrados, método que hoy día tiene una vigencia extraordinaria bajo la denominación ya nombrada de regresión cuantil. Describiremos cómo este autor llevó a cabo la creación de su método y la aplicación del mismo a los datos sobre los que estaba interesado. Para redactar este trabajo hemos usado la obra original de Boscovich y Maire en su segunda edición [2, pp. 482–484, 501–510], de 1770, y referencias secundarias.[1]

1. Algunos datos biográficos sobre Boscovich

Rogerius Ioseph Boscovich nació el 18 de mayo de 1711 en Dubrovnik, hoy día en Croacia y entonces una ciudad independiente rodeada por el Imperio Otomano y la República de Venecia.[2]

Asistió al Colegio de los Jesuitas en Dubrovnik hasta 1725. Su talento hizo que recibiera una invitación, a la edad de 14 años, para ir a Roma a convertirse en sacerdote jesuita y estudiar en el *Collegium Romanum*, la escuela insignia de la formación jesuita. Allí sus estudios incluyeron un curso de filosofía de tres años, en el que se impartía lógica durante el primer año, física aristotélica con matemáticas euclidianas y astronomía durante el segundo, y la metafísica y ética de Aristóteles en el tercero.

[1]Los comentarios de Todhunter [15, vol. 1, pp. 309, 319, 346–375] de 1873), el capítulo 1 del libro de Koenker [6], el capítulo 6 de Hald [5], el artículo de Sheynin [11], el de Stigler [12] y el importante capítulo 1 del libro de este mismo autor [14].

[2]Hay cierta ambigüedad sobre cómo escribir el nombre de Boscovich. En inglés se le conoce generalmente como Roger Joseph Boscovich, en italiano como Ruggiero Giuseppe Boscovich, mientras que en su nativo croata se le llama Ruder Josip Boškovic (pronunciado como «Boshkovich»).

En 1735 Boscovich empezó a estudiar a Newton. Cuando Orazio Borgondio accedió a rector del *Collegium* en 1740, dejó la cátedra de matemáticas a su protegido Boscovich. Durante los siguientes 20 años, Boscovich se convirtió en uno de los científicos jesuitas más prominentes de su época, dejando huella en diversas disciplinas como la matemática, la astronomía, la óptica, la estática, la hidráulica, la geodesia y la filosofía natural. En 1744 fue ordenado sacerdote jesuita. Desde 1757 a 1763 viajó por toda Europa como diplomático del estado papal. A su vuelta a Italia, fue profesor de matemáticas y astronomía, primero en la Universidad de Pavía y después en Milán. Tras la supresión de la orden jesuita en 1773, sus colegas científicos franceses le convencieron para que se instalara en París, donde permaneció durante nueve años como Director de Óptica de la Armada Francesa. Pasó sus últimos años en Milán.

2. Sobre la esfericidad de la Tierra

Encontramos las primeras dudas sobre la esfericidad perfecta de la Tierra en las experimentaciones de Richer en Cayena (Guayana Francesa), cerca del ecuador, realizadas en 1672 y publicadas en 1679 en las *Memorias de la Academia de París* [10]. Aquí Richer observó que para que un péndulo oscile en la Guayana con el mismo periodo que otro en París, su longitud había de ser ligeramente más corta, lo que interpretó como que la Tierra no era perfectamente esférica, sino que debía tener un achatamiento por los polos, puesto que un péndulo en el ecuador era menos afectado por la gravedad que lo sería el mismo péndulo oscilando en París. Ya en 1687, Newton, en los *Principia* [9], escribe que era aceptable la hipótesis según la cual el giro de la Tierra ha provocado un ensanchamiento en la zona del ecuador y un aplastamiento en los polos, de forma que la Tierra es un esferoide achatado o esferoide oblato. Llevado a la práctica, eso significa que el radio de la Tierra en el ecuador es superior al mismo en los polos. Ese exceso lo calculó Newton en la fracción aproximada de 1/230 del radio en el ecuador, y lo llamó la elipticidad de la Tierra.

Aparte de la variabilidad en las oscilaciones del péndulo según la latitud del lugar, un segundo procedimiento para comprobar la existencia de esa elipticidad era la medición de la longitud de un arco del meridiano, para un grado, en diferentes latitudes ampliamente separadas, y comparar dichas longitudes. Así, la Academia Francesa organizó expediciones a Perú, Laponia y el Cabo de Buena Esperanza, entre 1735 y 1754, con el objetivo de hacer esas mediciones y compararlas con medidas similares en las proximidades de París. Estas expediciones fueron agotadoras, costosas y lentas, requiriendo gran destreza en mediciones astronómicas y geodésicas.

Toda esta actividad despertó el interés en otros países. En 1750, el Papa Benedicto XIV comisionó a Boscovich y al jesuita inglés Christopher Maire, rector del Colegio Jesuita Inglés en Roma, para medir un arco del meridiano y construir un nuevo mapa de los estados papales. Su informe, escrito en latín, fue publicado en 1755. Fueron dos años de aventuras para Boscovich y Maire, realizando mediciones, y tres años para producir su estimación de la forma de la Tierra. La obra fue publicada en cinco volúmenes bajo el título *De Litteraria Expeditione* [2]. En 1770 el libro, con algunas adiciones, fue traducido al francés y publicado bajo el nombre de *Voyage Astronomique et Géographique* [3].

En el texto de Boscovich aparece por primera vez un método robusto de ajustar una línea recta a través de datos observados. Como ya se ha dicho, fue un trabajo precursor de lo que hoy conocemos como regresión cuantil. Podemos considerar entonces dicho ajuste una de las aportaciones de Boscovich al desarrollo de la estadística, y lo hizo medio siglo antes de que Legendre publicase en 1805 su método de ajuste de rectas mediante mínimos cuadrados. Fue, por tanto, uno de los antecedentes al tan popular y exitoso método de mínimos cuadrados.

3. La primera regresión: una regresión cuantil

Como ya se ha dicho, Newton concluye en sus *Principia*[3] que la Tierra es un esferoide achatado en los polos y que la proporción entre el diámetro en el ecuador y en los polos era de 230 a 229, por lo que la elipticidad de la Tierra, o proporción de exceso de un diámetro sobre otro, con respecto al mayor, es de 1/230. Sin embargo, esta conclusión sobre la forma o la figura de la Tierra no fue indiscutible. Giovanni Domenico Cassini, director del Real Observatorio de París, pensaba de hecho que la Tierra era un esferoide prolato, aplanado en el ecuador, no en los polos. Se hicieron varios intentos en el siglo siguiente para determinar si la Tierra estaba oblata o prolata y para medir la desviación de su forma respecto de la esférica.

Los dos principales métodos para determinar la figura de la Tierra fueron experimentos de péndulo y medidas de arco. Nos ocupamos aquí del método empleado por Boscovich: la determinación de la figura de la Tierra con medidas de arco. La idea era medir la longitud lineal de un grado de latitud en dos (o más) latitudes ampliamente separadas. Si se encuentra que un grado cerca del ecuador es más corto que uno más cercano al polo, entonces la forma de la Tierra es oblata y la diferencia entre las dos mediciones se puede usar para calcular la elipticidad.

La latitud de un punto en la superficie de la Tierra no es el ángulo formado entre las dos líneas que salen desde el centro de la Tierra, una dirigida al punto dado y la otra a la intersección del plano ecuatorial y el plano del meridiano de dicho punto. La latitud es el ángulo entre una línea al cenit desde el punto dado y el plano ecuatorial o, alternativamente, el complementario del ángulo entre las líneas desde el punto dado hasta el cenit y hasta la estrella polar. Por dicha razón, un grado más corto cerca del ecuador indica un abultamiento en el mismo. La figura 1 muestra arcos de 10° de latitud para una Tierra exageradamente oblata.

La relación entre longitud de arco y latitud puede ser derivada de la geometría de secciones cónicas, siendo dada la relación exacta por una integral elíptica. Sin embargo, para arcos pequeños la relación entre la longitud del arco y la latitud era aproximadamente lineal para nuestro autor. Del tipo $y = \alpha + \beta x$, donde y denota la longitud de un arco de un grado del meridiano y $x = \operatorname{sen}^2 L$, donde L es la latitud del punto medio del arco (ya que los arcos medidos son de diferentes grados, y es

[3]Pág. 407, formando parte de la Proposición XIX, Problema III.

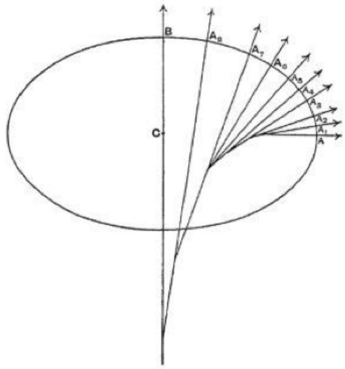

Figura 1: Una vista lateral de una Tierra exageradamente oblata, que ilustra el alargamiento de grados de arco hacia el polo. El cuadrante meridiano AB se divide en nueve segmentos, cada uno de 10° de latitud [1, p. 277].

considerada como la longitud del arco correspondiente medida en toises[4] dividida por la diferencia entre las latitudes de los puntos finales). Por tanto, es una longitud por unidad de arco.

Para esa relación lineal, si $L = 0$ entonces $x = 0$ e $y = \alpha$. Por tanto, podemos interpretar el parámetro α como la longitud de un arco correspondiente a un grado en el ecuador. Si $L = \frac{\pi}{2}$ tendremos $x = 1$ y, según la relación lineal propuesta, $y = \alpha + \beta$. Por tanto, interpretamos β como el excedente de arco correspondiente a un grado en el polo con respecto a su valor en el ecuador.

Si a es el radio de la Tierra en el ecuador y b lo es en un polo, se define la elipticidad de la Tierra como el cociente $\eta = \frac{a-b}{a}$. Pues bien, Boscovich pensaba que una buena aproximación a la misma era $\frac{\beta}{3\alpha}$, según la interpretación dada a ambos parámetros, por lo que era necesario «calcular» α y β para así poder «calcular» la elipticidad. Realmente, los investigadores de la época hablaban en términos de la inversa de la elipticidad, o sea, $\frac{1}{\eta}$ (por ejemplo, $\frac{1}{230}$). Para Boscovich, $\frac{1}{\eta} = \frac{3\alpha}{\beta}$. Posteriormente, Laplace usó una aproximación ligeramente mejor: $\frac{1}{\eta} = \frac{3\alpha}{\beta} + \frac{5}{3}$, pero la diferencia entre las dos aproximaciones es pequeña. Otros investigadores usaron $\frac{1}{\eta} = \frac{3\alpha}{\beta} + \frac{3}{2}$, o incluso $\frac{1}{\eta} = \frac{3\alpha}{\beta} + 2$ [14].

PRIMERA ACTUACIÓN DE BOSCOVICH. Cuando Boscovich abordó por primera vez este problema en 1755 (en un capítulo de su trabajo conjunto con Maire, pero en el que él se hizo único responsable del capítulo), alcanzó un éxito limitado. El autor era consciente, como otros antes que él, de que para obtener una determinación precisa de la figura de la Tierra sería necesario comparar mediciones ampliamente separadas en la latitud, ya que pequeños errores cometidos en mediciones del arco en lugares próximos podrían suponer errores exagerados en el cálculo final. Por lo tanto, Boscovich concentró su atención en solo cinco mediciones que se realizaron en ubicaciones

[4] 1 toise es igual a 1949 m.

bien separadas y que probablemente fueran bastante precisas. Dichas mediciones dieron lugar a las cinco parejas de valores (para x y para y) del cuadro 1. En su texto no encontramos una descripción analítica del manejo de estos datos; aquí, como en otros lugares, siguió la tradición newtoniana de dar descripciones geométricas en lugar de analíticas.[5] Sin embargo, será más fácil relacionar los diferentes esfuerzos de Boscovich con el trabajo posterior si adoptamos una formulación analítica desde el principio.

Boscovich ordena las observaciones $(x_1, y_1), \ldots, (x_n, y_n)$ de acuerdo con el tamaño de x. Ya que $x_1 = 0$, tenemos $y_1 = a$ y señala que puede calcular un valor de β usando cada par de observaciones de la siguiente forma:

$$\beta = \frac{y_j - y_i}{x_j - x_i}, \qquad i, j = 1, \ldots, n;\ i < j.$$

De las quince observaciones realizadas, once son tomadas en Francia, lo que significa que están demasiado cercanas para dar fiabilidad a los valores de β a causa de la influencia de los errores de medida. Por ello solo se queda con una de estas observaciones y en consecuencia sus datos constan de las cinco observaciones mostradas en el cuadro 1.

Lugar	L = Latitud	$x = \operatorname{sen}^2 L$	y = longitud de arco
(a) Quito	$0°0'$	0.0000	56751
(b) Cabo de Buena Esperanza	$33°18'$	0.2987	57037
(c) Roma	$42°59'$	0.4648	56976
(d) París	$49°23'$	0.5762	57074
(e) Laponia	$66°19'$	0.8386	57422

Cuadro 1: Datos de Boscovich sobre latitudes y longitudes de arco de meridiano de $1°$ en toises. El autor denotaba θ la latitud y, en 1755, ponía $x = \operatorname{sen}^2 \theta \times 10^4$ [2].

[5] Isaac Todhunter se encuentra entre aquellos que se han sentido frustrados porque Boscovich mantuvo los engorrosos aparatos geométricos de Newton en lugar de utilizar las fórmulas analíticas más elegantemente concisas de Clairaut y Euler. Al comentar una de las pruebas de Boscovich, Todhunter escribe: «Boscovich propone usar solo la Geometría: pero la Geometría consiste principalmente en denotar la longitud de cada línea recta con dos letras mayúsculas en lugar de una sola letra pequeña: esta extraña noción de Geometría ha sobrevivido hasta nuestros propios tiempos en la Universidad de Cambridge» [15, vol. 1, p. 309]. Todhunter aprovechó más tarde otra oportunidad para vincular sarcásticamente a Boscovich con su propia Universidad de Cambridge: «el autor había prescrito la condición de suministrar investigaciones geométricas, por lo que el Cálculo Diferencial no se introdujo. Debemos considerar el tratado más bien como el trabajo de un profesor con fines de instrucción que como el de un investigador para el avance de la ciencia, entonces podríamos premiar con alabanza que la tarea propuesta se lleva a cabo de manera justa. Hubiera sido más deseable estudiar la obra de Clairaut que limitarse a los métodos geométricos de Boscovich: pero la experiencia de nuestra propia universidad nos muestra que es posible encontrar de vez en cuando métodos utilizados para enseñar con algunos años de atraso con respecto a los utilizados para la investigación» [15, vol. 1, p. 319].

En 1755 Boscovich usa erróneamente 0.2987 como valor de $\operatorname{sen}^2 33^\circ 18' = 0.3014$; en 1770, [3, p. 482] da el valor 0.3015. Este error no tuvo influencia esencial en el resultado de Boscovich.

Para estimar los parámetros, Boscovich primero usa el método de puntos seleccionados, eligiendo la primera y última observación para conseguir un contraste tan grande como sea posible. Esto conduce a la recta

$$\hat{y} = 56751 + 800x,$$

o sea, $\alpha = 56751$, el valor de y para $x = 0$, y $\beta = \frac{57422-56751}{0.8386} \simeq 800$. Para los otros tres valores observados, los tres intermedios, los residuos (diferencia entre el valor observado y el ajustado según la ecuación) que resultan, -46, 144, 138, los considera excesivos.

Entonces busca otra solución y propone considerar todas las posibles combinaciones de las ecuaciones de condición con $\alpha = y_1$, ya que $x_1 = 0$. Por tanto, solo tiene que calcular los diez valores de β, que resultan variar desde -350 a 1327. Finalmente, calcula las elipticidades $\frac{\beta}{3\alpha}$, que varían desde $-\frac{1}{486}$ hasta $\frac{1}{128}$. Promediando los diez valores de β, encuentra 222 y la correspondiente elipticidad $\frac{222}{56751} = \frac{1}{255}$. La gran variación de las elipticidades le desanima tanto que concluye: «es evidente que las determinaciones de estos grados no pueden conciliarse con la elipse de Newton ni con ninguna otra elipse más o menos oblata».

Estos resultados y los que siguen se encuentran en la traducción francesa del informe de Boscovich y Maire (véase [3, pp. 484, 494, 497]). Por error de imprenta, la elíptica es aquí dada como $\frac{1}{155}$ en lugar de $\frac{1}{255}$. En una discusión más detallada de las posibles causas de la discrepancia entre el valor promedio de Boscovich y el $\frac{1}{230}$ encontrado por Newton (entre ellas la no homogeneidad de la densidad de la Tierra), Boscovich usa el valor correcto $\frac{1}{255}$. Que el valor $\frac{1}{155}$ es un error de imprenta se sigue también del hecho de que Boscovich en su análisis posterior descarte los dos valores más pequeños de las β, lo que le lleva a una media para β de 286 y a una elipticidad de $\frac{1}{198}$, que es un valor menor (en lugar de mayor) que $\frac{1}{155}$.

Boscovich no puede decidir qué promedio usar; después de una larga discusión vagamente concluye: «estoy convencido de que, de acuerdo con las observaciones hechas hasta ahora, es muy probable que la Tierra sea aplastada en los polos».

Usando el valor medio de los diez valores de β, la ecuación de la recta se convierte en

$$\hat{y} = 56751 + 655x$$

y los residuos son 87, -81, -60, 113. Boscovich no hace el cálculo de estos valores, y no parece darse cuenta de que su método utiliza efectivamente estos datos. No lo hizo, pero podía haber resuelto todas las diez combinaciones de las ecuaciones de condición para ambos α y β, y entonces habría conseguido la ecuación (usando el valor correcto de x_2)

$$\hat{y} = 56729 + 666x,$$

y los residuos 22, 108, -59, -38, 135, que indican un más pobre ajuste que el de arriba. La elipticidad ahora es igual a $\frac{1}{256}$.

Nuevo método. En 1757 Boscovich publicó un resumen del informe de 1755, en el que volvió al problema de reconciliar ecuaciones inconsistentes y formuló un nuevo método de resolución, ahora conocido como el método de mínimas desviaciones absolutas. Declaró que el método era nuevo y lo aplicó a los datos anteriores en una nota contenida en un tratado latino sobre filosofía natural de Stay de 1760.[6]

La formulación por Boscovich del principio es puramente verbal: dado un cierto número de observaciones, encontrar las correcciones que deben ser hechas en cada una (esto es, encontrar los valores ajustados y calcular los residuos o correcciones) con las siguientes condiciones observadas:

1. La suma de las correcciones positivas será igual a la suma de las correcciones negativas (independencia del signo): $\sum_{i=1}^{n}(y_i - \alpha - \beta x_i) = 0$.
2. La suma de todas las correcciones, ya sean positivas o negativas, será la menor posible: será minimizada la suma $\sum_{i=1}^{n} |y_i - \alpha - \beta x_i|$.

Su motivación para la primera condición es la simetría de la distribución de los errores. La segunda se requiere para aproximar la observación lo más cerca posible, esto es, asume que la bondad del ajuste es medida por la suma de las desviaciones absolutas, como era costumbre en aquel momento. Por ejemplo, Galileo (1632) comparó dos hipótesis sobre la posición de una estrella comparando la correspondiente suma de los errores absolutos [5, §10.3].

Boscovich es el primero en formular un criterio para ajustar una recta a los datos basado en la minimización de una función de los residuos.

De la primera condición se sigue que $\bar{y} = \alpha + \beta\bar{x}$, lo que significa que la línea ajustada pasa por el centro de gravedad de los puntos observados. Usando este resultado para eliminar α de la segunda condición, obtenemos la suma

$$S(\beta) = \sum_{i=1}^{n} |y_i - \bar{y} - \beta(x_i - \bar{x})|,$$

que será minimizada con respecto a β.

Hoy día, si queremos efectuar una regresión cuantil, o sea, estimar a través de un modelo lineal del tipo $y_i = \alpha + \beta x_i + u_i$ (usamos una variable explicativa tal como Boscovich lo hizo en su trabajo) un determinado cuantil de la variable objetivo Y, cuyos valores observados son los y_i, la estimación de un cuantil τ (con $0 \leq \tau \leq 1$) de dicha variable objetivo condicionada a los valores de la variable explicativa consiste en resolver

$$\min_{\beta}\Big(\sum_{y_i \geq \alpha+\beta x_i}^{n} \tau|y_i - \alpha - \beta x_i| + \sum_{y_i < \alpha+\beta x_i}^{n} (1-\tau)|y_i - \alpha - \beta x_i|\Big),$$

donde los valores obtenidos para α, β que minimicen esa suma ponderada son las estimaciones cuantiles del modelo. Obviamente, si $\tau = 0.5$ obtenemos la estimación

[6] Una traducción al francés de esta nota se encuentra en [3, pp. 501–506]. Boscovich cita las *Actes de l'Institut de Bologne*. En sus *Suppléments de la Philosophie*, tomo 2, p. 420, aparece Benoit Stay.

de la mediana de Y condicionada a los valores de las variables explicativas. Por tanto, lo que Boscovich propone como método de estimación es lo que hoy conocemos como regresión cuantil para el caso particular de estimación de la mediana de la variable objetivo.

Ahora bien, hay que enfatizar que esta formulación analítica no se encuentra en el trabajo de Boscovich. Su establecimiento de las condiciones fue solo dado en la versión verbal que hemos señalado, y en ningún lugar de su trabajo dio otra cosa que no fuera una descripción geométrica o mecánica de su solución al problema. Pero a pesar de que esta restricción a un enfoque geométrico era costosa porque se limita la generalización del enfoque, hubo un importante beneficio de compensación: el enfoque geométrico de Boscovich sugirió una solución al problema que habría sido mucho menos evidente con una formulación analítica.

A pesar de ello, Boscovich comienza señalando que su primera condición puede ser expresada analíticamente (en el sentido de que los arcos corregidos son expresables como una ecuación de primer grado que involucra a los senos versos) y que, si esto fuese dado, la segunda condición sería expresable como una suma en la que se implican coeficientes desconocidos. Sin embargo, como se ha dicho, Boscovich no presenta esta expresión analítica, sino que trabaja en la tradición newtoniana que requiere una formulación geométrica.

La discusión de Boscovich estaba acompañada en la traducción de 1770 de un diagrama reproducido aquí como figura 2. Podemos considerar AF como una unidad de longitud, A como el origen y A, B, C, D y E como representando los cinco valores de $\operatorname{sen}^2 \theta$ situados en el intervalo de longitud unidad. Las longitudes Aa, Bb, Cc, Dd y Ee pueden ser consideradas representantes de las longitudes (en toises por grado) de los correspondientes arcos de medida. El problema era encontrar una línea recta $A'H$ tal que las correcciones aA', bO, cK, dL y eM satisfagan las condiciones. Boscovich comienza observando que la primera condición, interpretada como la descripción de una condición de equilibrio mecánico, implica que la línea debe pasar por el centro de gravedad G de los puntos. Esto reduce el problema a encontrar que la línea que pasa por G sea tal que la segunda condición sea satisfecha.

Para encontrar la solución, Boscovich imaginó la línea recta SGT que rotase en sentido horario, con centro de giro en G. A medida que la línea gira, la suma de las correcciones (tomadas sin tener en cuenta el signo) disminuirá hasta que se alcance un mínimo y después empezarían a crecer. Dado que las correcciones para los arcos individuales cambian a medida que la línea gira, es suficiente señalar el orden en el que los puntos son encontrados e ir efectuando la suma de las correcciones, AS, BS, SC, SD y SE, que corresponden a esos puntos, mediante suma acumulada incrementando en cada caso un sumando. Cuando dicha suma supere la mitad de la suma total $AS + BS + SC + SD + SE$, la recta obtenida verificará la condición de mínimo. Simplificó aún más la solución al observar que la línea giratoria encontraría los cinco puntos en el orden inverso a las pendientes de las cinco líneas desde G a dichos puntos: $\frac{aX}{AS}$, $\frac{bo}{BS}$, etcétera.

El notable diagrama de Boscovich muestra al mismo tiempo los datos, las pendientes $\beta_i = \frac{y_i - \bar{y}}{x_i - \bar{x}}$ y la línea ajustada, y demuestra cómo encontrar $\hat{\beta}$, el valor de β que minimiza $S(\beta)$.

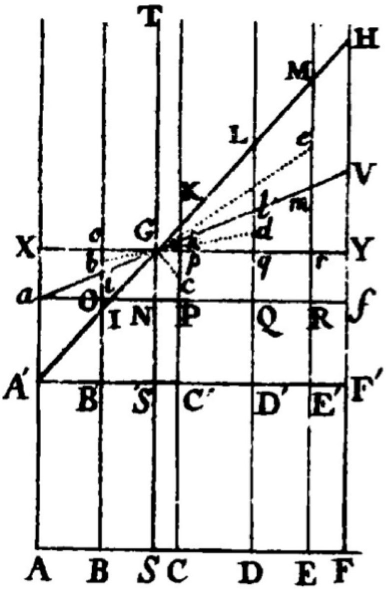

Figura 2: Gráfico de Boscovich de los datos y la línea ajustada. La línea discontinua que pasa por el centro es la línea móvil de Boscovich [3, p. 502]. El eje horizontal AF da $\operatorname{sen}^2\theta$, donde θ es la latitud del punto medio del arco: el eje vertical AX da la longitud de arco en toises por grado. Los cinco arcos están indicados por a, b, c, d y e; G es el centro de gravedad.

Visto de manera analítica, para estudiar cómo $S(\beta)$ depende de β, Boscovich introduce la línea antes citada que pasa por el centro de gravedad y con pendiente β (la línea discontinua en la gráfica), que se mueve en el sentido de las agujas del reloj desde la posición vertical hasta la horizontal y hacia abajo hasta la vertical; esto es, β decrece de más a menos infinito. Al mismo tiempo, $S(\beta)$ decrece continuamente desde el infinito a un mínimo, y crece después al infinito. La línea móvil pasa por las observaciones en el orden e, a, d, b, c. Boscovich es consecuente con dicho orden y efectúa una reordenación de las observaciones introduciendo las desviaciones del centro de gravedad $X_i = x_i - \bar{x}$, $Y_i = y_i - \bar{y}$, por lo que $\beta_i = \frac{Y_i}{X_i}$, como se indica por las líneas de puntos, y así tenemos

$$S(\beta) = \sum_{i=1}^{n} |Y_i - \beta X_i| = \sum_{i=1}^{n} |X_i|\,|\beta_i - \beta|,$$

donde el subíndice i ahora se refiere a la nueva ordenación de las observaciones del cuadro 2.

El razonamiento de Boscovich está basado en el hecho de que el valor absoluto de los residuos, $|e_i| = |X_i|\,|\beta_i - \beta|$, es proporcional a $|X_i|$, por lo que $|e_i|$ decrece linealmente con β (la pendiente de la línea en movimiento) cuando β decrece desde infinito hasta β_i y crece con β decreciendo desde β_i hasta menos infinito. Por tanto,

Lugar (i)	X_i	Y_i	β_i	$\Sigma\|X_i\|$	$S(\beta_i)$
(e)	0.40294	369.4	917	0.40294	416
(a)	−0.43566	−301.6	692	0.83860	340
(d)	0.14054	21.4	152	0.97914	627
(b)	−0.13696	−15.6	114	1.11610	658
(c)	0.02914	−73.6	−2526	1.14524	3527

Cuadro 2: Reordenamiento de Boscovich de las observaciones de acuerdo con las pendientes decrecientes. El autor usa $\frac{1}{\beta}$ en lugar de β [3, p. 506]. Los valores de $S(\beta)$ han sido calculados posteriormente.

$S(\beta)$ es una función lineal a trozos de β, con una pendiente que depende de la posición de β en relación a la de los β_i, siendo la pendiente de $S(\beta)$ la diferencia entre las correspondientes sumas de los $|X_i|$. Para $\beta > \beta_1$, la pendiente es igual a $|X_1|+|X_2|+\cdots+|X_n|$. Para $\beta_2 < \beta < \beta_1$, la pendiente es igual $|X_2|+\cdots+|X_n|-|X_1|$, y así sucesivamente vamos construyendo la pendiente para diferentes valores de β. El mínimo de $S(\beta)$ es así obtenido para, digamos, $\beta = \beta_k$, donde k está determinada por la desigualdad

$$\sum_{i=k+1}^{n} |X_i| - \sum_{i=1}^{k} |X_i| \leq 0 < \sum_{i=k}^{n} |X_i| - \sum_{i=1}^{k-1} |X_i|,$$

o, equivalentemente,

$$\sum_{i=1}^{k-1} |X_i| < \frac{1}{2} \sum_{i=1}^{n} |X_i| \leq \sum_{i=1}^{k} |X_i|,$$

que es la forma usada por Boscovich.[7]

Para los números que aparecen en el cuadro 2, Boscovich encuentra que

$$|X_1| = 0.40294 < 0.57262 < 0.83860 = |X_1| + |X_2|,$$

por lo que $\hat{\beta} = \beta_2 = 629$.

La ecuación para la línea de mejor ajuste es $\hat{y} = \bar{y} + \hat{\beta}(x - \bar{x})$, $\hat{\beta} = \beta_k$. De la definición de β_k se sigue que el residuo $e_k = 0$, por lo que la ecuación puede ser escrita como

$$\hat{y} = y_k + \beta_k(x - \bar{x}).$$

Usando los valores numéricos, en 1770 Boscovich encuentra [3, p. 506]

$$\hat{y} = 56751 + 692.3x$$

y da los residuos 0, 79.2, −93.8, −75.9, 90.5, siendo la suma de los valores absolutos igual a 340. La elipticidad es igual a $\frac{1}{246}$, que está bastante próxima al valor $\frac{1}{255}$ previamente encontrado y arriba citado.

[7] Hoy β_k es llamada la mediana ponderada de los β_i.

Consideremos el método de mínimas desviaciones absolutas bajo el supuesto de que β es conocido. Poniendo $\alpha_i = y_i - \beta x_i$, la menor suma de las desviaciones absolutas se obtiene minimizando $\sum |\alpha_i - \alpha|$, por lo que a se convierte en la mediana de las α_i. Dado que $\beta = 692$, la media de los diez valores β es independiente de α. La menor suma de las desviaciones absolutas se obtiene eligiendo α como la mediana de los cinco valores α_i, lo que lleva a $\alpha = 56.751$, esto es, al valor usado por Boscovich para la construcción mediante medias. Por tanto, si combinamos el valor medio de los β para cada par de observaciones con la mediana de los α_i, tendremos otro método simple y «objetivo» de estimación. Si Boscovich hubiese comparado el ajuste obtenido usando sus dos líneas con diferentes pendientes, pero con el mismo α, habría descubierto que son prácticamente igual de buenas, puesto que las dos sumas de desviaciones absolutas son 340 y 341, respectivamente.

Boscovich no hace mención sobre la posibilidad de minimizar la suma de las desviaciones absolutas, sin restricción alguna, sobre α y β.

En 1770, Boscovich tuvo a su disposición cuatro observaciones más que en 1760. Analizó las nueve observaciones con el mismo método que usó para las cinco y presentó una tabla análoga al cuadro 2. Entonces siguió un camino peligroso rechazando primero una y luego dos observaciones más, debido a los grandes residuos. De este modo, encontró elípticas con un mejor acuerdo con sus ideas teóricas (véase [3, pp. 482, 506–510]).

4. Conclusión

Con este trabajo mostramos a Boscovich como un pionero del análisis de regresión. Una estimación tan particular como la de la mediana condicionada fue previa a la más habitual de media condicionada. En los tiempos actuales, la regresión propuesta por este autor está siendo exitosa, pues cada vez encontramos más trabajos empíricos en los que se usa la regresión cuantil como complementaria, unas veces, y como sustitutiva, otras, de la regresión por mínimos cuadrados.

Referencias

[1] A. Berry, *A Short History of Astronomy*, John Murray, London, 1898; reimpresión: Dover, New York, 1961.

[2] R. J. Boscovich y C. Marie, *De litteraria expeditione per pontificiam ditionem ad dimitiendos duos Meridiani gradus*, Palladis, Roma, 1755.

[3] R. J. Boscovich y C. Marie, *Voyage astronomique et géographique, dans l'état de l'église*, N. M. Tilliard, Paris, 1770. Incluye una traducción al francés de [2].

[4] C. F. Gauss, *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*, Perthes et Besser, Hamburg, 1809. Reproducido en *Werke* **7**, 1–280, Perthes, Gotha, 1871. Traducción inglesa: *Theory of the motion of the heavenly bodies moving about the sun in conic sections*, Little, Brown and co., Boston, 1857; reimpresión: Dover, New York, 1963.

[5] A. HALD, *A history of Mathematical Statistics from 1750 to 1930*, John Wiley & Sons, New York, 1998.

[6] R. KOENKER, *Quantile Regression*, Cambridge University Press, Cambridge, 2005.

[7] A. M. LEGENDRE, *Nouvelles méthodes pour la détermination des orbites des comètes*, Firmin Didot, Paris, 1805. Páginas 72–75 del apéndice reimpresas en [14, p. 56]; traducción al inglés de estas páginas en D. E. Smith, *A Source Book of Mathematics*, 576–579, McGraw-Hill Book Company, New York, 1929.

[8] F. MOSTELLER Y J. TUKEY, *Data analysis and regression: A second course in statistics*, Addison-Wesley, Reading, 1977.

[9] I. NEWTON, *Philosophiae naturalis principia mathematica*, Royal Society, London, 1687. Traducción inglesa: *The mathematical principles of natural philosophy*, Motte, London, 1729; traducción revisada con notas de F. Cajori, University of California Press, Berkeley, 1934.

[10] J. RICHER, *Observations astronomiques et physiques faites en l'isle de Caïenne*, Imprenta real, Paris, 1679.

[11] O. SHEYNIN, R. J. Boscovich's work on probability, *Arch. History Exact Sci.* **9** (1973), 306–324.

[12] S. M. STIGLER, Boscovich, Simpson and a 1760 manuscript note on fitting a linear relation, *Biometrika* **71** (1984), 615–620.

[13] S. M. STIGLER, Gauss and the invention of least squares, *Ann. Statist.* **9** (1981), 465–474.

[14] S. M. STIGLER, *The history of statistics. The measurement of uncertainty before 1900*, Harvard University Press, Cambridge, 1986.

[15] I. TODHUNTER, *A history of the mathematical theories of attraction and the figure of the Earth from the time of Newton to that of Laplace*, 2 vols., Macmillan, London, 1873; reimpresión: Dover, New York, 1962.

MARÍA DOLORES PÉREZ HIDALGO, DPTO. DE ECONOMÍA APLICADA I, UNIVERSIDAD DE SEVILLA
Correo electrónico: mdperez@us.es

JESÚS BASULTO SANTOS, DPTO. DE ECONOMÍA APLICADA I, UNIVERSIDAD DE SEVILLA
Correo electrónico: Basulto@us.es

JOSÉ ANTONIO CAMÚÑEZ RUIZ, DPTO. DE ECONOMÍA APLICADA I, UNIVERSIDAD DE SEVILLA
Correo electrónico: camunez@us.es