



Regresiones:

Rodney Ramcharan

NO CABE duda de que saber leer es importante, y los maestros de primaria se han dado cuenta de que cuanto más calzan los alumnos, con más soltura suelen leer. Ahora bien, ¿significa esto que el ministerio de Educación debería ayudarlos premiando al científico que logre hacerles crecer los pies? Obviamente, el hecho de que el número del calzado y la facilidad de lectura tiendan a aumentar paralelamente no significa que uno lea mejor por tener el pie más grande. A medida que pasa el tiempo, a los niños no solo les van creciendo los pies, sino que también se les va desarrollando el cerebro. Esa evolución natural es lo que explica la simple observación de que el tamaño del pie y la aptitud para la lectura tienden a aumentar juntos; en otras palabras, tienen una correlación positiva. Pero a todas luces no existe una relación: nadie lee con más soltura por calzar un número más grande.

En el campo de la economía las correlaciones abundan, pero no es fácil decidir si la correlación entre dos o más variables representa una relación causal. Los países que comercian más con el resto del mundo gozan de un nivel de ingreso más alto, pero ¿significa esto que el comercio internacional contribuye a la prosperidad? La gente más instruida suele ganar más, pero ¿podemos deducir que la enseñanza se traduce en mejores sueldos? Es importante tener respuestas precisas a estas preguntas: si unos años más de escolaridad producen un ingreso más alto, entonces los gobiernos podrían reducir la pobreza dedicando más fondos a la enseñanza. Si un año extra en el aula añadiera US\$20.000 por

año al nivel salarial, el gasto educativo rendiría mucho más que si el aumento fuera de solo US\$2 por año.

Frente a este tipo de pregunta, los economistas utilizan una herramienta estadística llamada análisis de regresión. Las regresiones sirven para cuantificar la relación entre una variable y las demás variables que supuestamente la explican, y también para comprobar si la relación es estrecha y bien determinada. Hoy por hoy, es fácil y normal hacer miles de regresiones, pero no siempre fue así (véase el cuadro); de hecho, lo difícil es encontrar un estudio académico empírico que no las incluya, y esa abundancia se repite en la esfera de la sociología, la estadística y la psicología.

¿Cómo se hace una regresión?

Para ver cómo funciona una regresión, retomemos el problema del rendimiento de la enseñanza. El gobierno recoge datos sobre el nivel educativo de la población. Pero la gente estudia por distintos motivos: algunos tienen facilidad; otros están más motivados, y hay quienes se dedican a carreras no académicas y terminan ganando muy bien. Como todas estas razones para dedicarse a los estudios pueden influir en el ingreso personal, es difícil saber si la correlación entre la enseñanza y el sueldo representa una relación causal o si responde a otros factores. Es posible que la gente que tiene facilidad de estudio en el aula también la tenga en la oficina y que eso le ayude a subir de sueldo. Es decir, la correlación positiva entre el nivel salarial y el nivel educativo quizá refleje no tanto el efecto de la enseñanza, sino más bien una aptitud innata.

Un modelo teórico quizá muestre mejor cómo y por qué una variable “dependiente” está determinada por una o varias variables “independientes” o “explicativas”. El postulado de que los ingresos de una persona dependen de su escolaridad constituye un ejemplo de un modelo simple con una variable explicativa. La ecuación de regresión correspondiente, supuestamente lineal, sería así:

$$Y = a + bX$$

Del lazo izquierdo tenemos Y , la variable dependiente, que en este caso son los ingresos. A la derecha están a , la constante (o intercepto), y b , el coeficiente (o pendiente) multiplicado por X , la variable independiente (o explicativa), que en este caso es la escolaridad. La regresión expresa en términos algebraicos que “los ingresos dependen solo del nivel educativo y la relación es lineal”, y omite todo otro factor explicativo que pudiera existir.

La magia de la informática

Las primeras conceptualizaciones de la regresión se remontan al siglo XIX, pero el auge no estalló hasta la revolución tecnológica del siglo XX, cuando las computadoras personales se volvieron cosa de todos los días. En los años cincuenta y sesenta, los economistas tenían que calcular las regresiones con calculadoras electromecánicas de escritorio. En 1970 todavía había que esperar hasta 24 horas para recibir el resultado de una regresión hecha en un laboratorio central, después de pasarse horas o días preparando las tarjetas perforadas. Además, bastaba con perforar mal una tarjeta, escribiendo incorrectamente una palabra de control o errando en el valor de un dato, y todo el esfuerzo se iba a la basura.

La fijación de los economistas

Ahora bien, si creemos que el mundo es mucho más complejo y que hay una variedad de factores que podrían explicar la influencia de la instrucción sobre los ingresos, haríamos una regresión de múltiples variables que se vería así:

$$Y = a + b_1X_1 + b_2X_2 + \dots$$

Ahora tenemos varias variables X (capacidad, inteligencia, edad, escolaridad, estado civil, nivel educativo de los padres) para explicar la variable Y (ingresos). Los coeficientes b no hacen más que medir el impacto de cada una de estas variables en los ingresos, presumiendo que las demás variables se mantienen constantes.

¿La inteligencia es oro?

Intentemos hacer una regresión partiendo de la teoría que el salario por hora (la variable dependiente) depende del nivel educativo (la variable explicativa). Supongamos que hay otra variable explicativa posible —la aptitud, medida con un test de cociente intelectual (CI)— que no tiene ningún efecto sobre el salario salvo el que pueda ejercer a través de la enseñanza. Incorporamos todos los datos sobre los ingresos y los niveles educativos y la regresión nos da:

$$Y = 5,40 + 1,06 \text{ EDU}$$

El coeficiente b muestra que por cada año extra de escolaridad, el salario por hora aumenta US\$1,06. Y para los que no han recibido instrucción ($\text{EDU} = 0$), la constante indica que el promedio salarial es de US\$5,40 por hora.

Ahora bien, ¿qué pasa si incluimos el cociente intelectual en la ecuación; es decir, si suponemos que los ingresos dependen tanto del nivel educativo como del cociente intelectual? Ingresamos los datos correspondientes a los tests de CI y vemos que:

$$Y = 5,40 + 0,83\text{EDU} + 0,001\text{IQ}$$

O sea que las personas con un puntaje más alto en el test de CI también ganan más por hora. Además, aunque el impacto del nivel educativo sigue siendo positivo, es un 27% menor si se tienen en cuenta los resultados de los tests de CI (el 27% representa la diferencia entre los coeficientes: $100(1,06 - 0,83)/0,83$). Esto significa que originalmente sobreestimamos el efecto de la escolaridad sobre los ingresos porque omitimos la influencia del CI, que está correlacionado con la escolaridad.

Posibles obstáculos

A pesar de sus ventajas, las regresiones no están exentas de problemas y suelen estar mal utilizadas, sobre todo por las siguientes cuatro razones.

Omisión de variables. Hay que tener un buen modelo teórico que sugiera qué variables podrían explicar la variable dependiente. Si se trata de una simple regresión de dos variables, es necesario pensar en los demás factores que podrían entrar en juego. En el ejemplo anterior, la correlación entre escolaridad e ingresos podría reflejar algún otro factor, además del CI. Es decir, las personas que integran la muestra podrían distinguirse por algo que no se ha observado y que explica los sueldos que terminan ganando, algo que quizá tenga que ver con las alternativas a su disposición. Los miembros de una familia acaudalada suelen gozar de un mejor acceso a la escolaridad, y a la vez la riqueza familiar puede dar lugar a contactos laborales que les reditúan sueldos más altos. Por lo tanto, la situación económica de la familia puede ser otra variable a tener en cuenta.

Causalidad inversa. Muchos modelos teóricos predicen una causalidad bidireccional; en otras palabras, una variable dependiente puede provocar cambios en una o más variables explicativas. Por ejemplo, un mejor nivel salarial permite invertir más en la propia educación, lo cual a su vez permite ganar más. Este es un factor que complica la estimación de una regresión y exige técnicas especiales.

Errores de medición. Existe la posibilidad de medir mal los factores. Por ejemplo, la aptitud es difícil de medir y los tests de CI presentan problemas bien conocidos. Por lo tanto, una regresión que incluya el CI quizá no tenga debidamente en cuenta la aptitud, lo cual podría desvirtuar las correlaciones entre la instrucción y los ingresos.

Excesiva estrechez. Un coeficiente de regresión muestra de qué manera los cambios que sufre una variable alteran a otra variable, pero solo si esos cambios son pequeños. Es decir, un pequeño cambio del nivel educativo probablemente influya en los ingresos, pero eso no permite hacer generalizaciones sobre el efecto de cambios profundos. Por ejemplo, si la totalidad de una población va a la universidad al mismo tiempo, es poco probable que un flamante graduado gane mucho más, ya que la oferta total de trabajadores con título universitario habrá aumentado drásticamente. ■

Rodney Ramcharan es Economista en el Departamento de Estudios del FMI.