

Cuidado con las inferencias de controles estadísticos (Comentado por Lampadia)

25 de Febrero de 2016

Aporte conceptual de especial interés para economistas

Muy buen artículo de **Ricardo Hausmann**, en el que muestra como algunos métodos de análisis y estudios aplicados por economistas pueden convertirse en una suerte de mecanismos 'automáticos' para llegar a conclusiones no representativas de la realidad. Hausmann se refiere a los ensayos con muestras de control que pueden mostrar inferencias erróneas.

Hace pocos días comentamos en Lampadia un artículo de Timothy Garton Ash, que presentamos como: **"Expectativas Racionales" sobre economistas**. Siendo que las 'expectativas racionales sobre los mercados' y los supuestos de 'mercados perfectos' habían causado entre los economistas, los mismos tipos de errores que ahora comenta Hausmann.

Muchas veces, los seres humanos pensamos que hemos encontrado los polvitos mágicos que nos permiten describir sistemas complejos, ya sean sistemas naturales o de carácter humano. Pero todo tiene su espacio, su tiempo, su contexto y muchas veces lo que nos falta es un poco de humildad.

El problema con las políticas basadas en evidencia

Ricardo Hausmann

Project Syndicate

25 de febrero, 2016

En la actualidad, múltiples organizaciones, desde [entidades gubernamentales](#) hasta [instituciones filantrópicas](#) y de ayuda, requieren que las políticas y los programas "estén basados en evidencia". Es sensato exigir que las políticas se basen en evidencia y que esta sea lo mejor posible, dentro de límites razonables de tiempo y de presupuesto. Sin embargo, la forma en que ahora se implementa este enfoque puede estar causando mucho daño al debilitar nuestra capacidad de aprender y de mejorar lo que hacemos.

Hoy día, el llamado "patrón-oro" de lo que constituye evidencia sólida es el "ensayo de control aleatorizado" o RCT por su sigla en inglés, una idea que se originó en la medicina hace doscientos años, pasó a la agricultura y, durante los últimos veinte años, ha estado causando furor en la economía. Su popularidad se asienta en el hecho de que aborda problemas clave en la inferencia estadística.

Por ejemplo, la gente rica viste ropa fina. ¿Haría ricos a los pobres el que se les distribuyera ropa fina? Este es un caso en que la correlación (entre

la ropa y la riqueza) no implica causalidad.

Los que estudian en la Universidad de Harvard consiguen muy buenos empleos. ¿Es porque en Harvard se enseña bien o porque se escoge bien, a personas inteligentes que hubieran tenido éxito en la vida de todos modos? Este es el problema de sesgo en la selección.

Los RCT abordan estos problemas dividiendo, de manera aleatoria, a los participantes en el ensayo entre quienes van a recibir un “tratamiento” y quienes van a recibir un “placebo”, creando así un grupo de “control”. Mediante la observación de las diferencias que resultan entre los dos grupos con posterioridad a la intervención, se puede evaluar la efectividad del tratamiento. Los RCT se han aplicado en medicamentos, microcréditos, programas de capacitación, herramientas educativas y una miríada de otras intervenciones.

Supongamos que se está considerando introducir *tablets* como medio para mejorar el aprendizaje en la sala de clases. Un RCT exigiría la participación de alrededor de 300 escuelas, de las cuales se asignaría 150 al grupo de control que no va a recibir *tablets*. Antes de que estas se distribuyeran, se realizaría un estudio de referencia inicial con el fin de evaluar cuánto están aprendiendo los niños en la escuela. Luego se distribuiría las *tablets* a las 150 escuelas que están bajo el “tratamiento” y se dejaría pasar un tiempo. Transcurrido el lapso, se llevaría a cabo otro estudio para descubrir si en ese momento existe una diferencia en el aprendizaje entre las escuelas que recibieron y las que no recibieron las *tablets*.

Supongamos que no hay diferencias significativas, como ha sucedido en cuatro RCT que encontraron que la distribución de textos no tenía ningún efecto. Sería equivocado presumir que se llegó a la conclusión de que las *tablets* (o los textos) no mejoran el aprendizaje. Lo que se ha demostrado es que esa *tablet* en particular, con su software particular, empleada en esa estrategia pedagógica particular y en la enseñanza de esos conceptos en particular, no hizo ninguna diferencia.

Pero la pregunta que realmente queríamos responder es ¿cómo se deberían utilizar las *tablets* para maximizar el aprendizaje? En este ámbito, el espacio para el diseño es realmente enorme, y los RCT no permiten testear más de dos o tres diseños al mismo tiempo –y a paso de tortuga–. ¿Es posible hacerlo de una manera mejor?

Consideremos el siguiente experimento mental: En la *tablet* se incluye algún mecanismo que informe al maestro en tiempo real qué tan bien sus alumnos están absorbiendo el material que se les está enseñando. A todos los maestros se les da libertad para experimentar con distintos tipos de software, de estrategias y de modos de emplear la nueva herramienta. El rápido circuito de retroalimentación que se producirá, hará que los maestros ajusten sus estrategias para maximizar el desempeño de sus alumnos.

Con el correr del tiempo, se observará que algunos maestros han dado con estrategias que son altamente efectivas. Entonces, se comparte lo que ellos han logrado con otros maestros.

Notemos lo radicalmente diferente que es este método. En lugar de testear la validez de un diseño haciendo que 150 de 300 escuelas implementen exactamente el mismo programa, este método explora el espacio de diseño al hacer que cada maestro busque resultados. En lugar de tener un estudio de referencia inicial y otro final, provee retroalimentación sobre el desempeño de manera constante. En lugar de que el aprendizaje lo haga un econometrista de modo centralizado y luego informe a todos sobre los resultados del experimento, son los maestros quienes hacen el aprendizaje de manera descentralizada e informan al centro sobre sus descubrimientos.

Es claro que los maestros van a confundir la correlación con la causalidad mientras ajusten sus estrategias, pero estos errores se irán corrigiendo prontamente a medida que sus presunciones equivocadas no produzcan resultados mejores. Igualmente, es posible que ocurra un sesgo en la selección (a algunos lugares les puede ir mejor que a otros porque difieren en otros aspectos); pero si diferentes contextos requieren estrategias diferentes, el sistema las va a encontrar tarde o temprano. Esta estrategia se asemeja más a la implementación social de un [algoritmo de aprendizaje automático](#) que a un ensayo clínico.

En economía, los RCT han estado causando gran furor, especialmente en el campo del desarrollo internacional, a pesar de las críticas formuladas por el ganador del Premio Nobel [Angus Deaton](#), así como por [Lant Pritchett](#) y [Dani Rodrik](#), quienes han atacado las exageradas aserciones de los defensores de los RCT. Una grave deficiencia es su validez externa. Las lecciones no viajan bien: Si a través de un RCT se descubre que darles micronutrientes a los niños de Guatemala mejora su capacidad de aprendizaje, ¿se les debería dar micronutrientes a los niños noruegos?

El problema más importante que yo tengo con los RCT es que nos hacen pensar sobre las intervenciones, las políticas y las organizaciones de manera errónea. A diferencia de los dos o tres diseños que se testean lentamente a través de un RCT (como colocar *tablets* o [rotafolios](#) en las escuelas), la mayor parte de las intervenciones sociales tiene millones de posibilidades de diseño y los resultados dependen de combinaciones complejas entre ellas. Esto conduce a lo que el científico de la complejidad Stuart Kauffman denomina un "[paisaje adaptativo accidentado](#)".

En todos los casos, obtener la combinación de parámetros correcta, es de importancia crítica. Ello exige que las organizaciones implementen estrategias de evolución basadas en ensayar las cosas y aprender rápidamente sobre sus resultados a través de circuitos de retroalimentación acelerados, según lo [han sugerido](#) Matt Andrews, Lant Pritchett y Michael Woolcock del [Center for International Development](#) de la Universidad de Harvard.

Bien puede que los RCT sean apropiados para los ensayos clínicos de medicamentos. Sin embargo, en una extremadamente amplia gama de ámbitos relativos a políticas, el impacto del movimiento RCT ha sido equivalente a poner a auditores a cargo de un departamento de investigación y desarrollo. Esta es la forma equivocada de diseñar cosas que funcionan bien. Solamente creando organizaciones que aprenden a aprender, del mismo modo en que la "producción optimizada" o [lean manufacturing](#) lo ha hecho para el sector

industrial , podremos acelerar el progreso.

Traducción de Ana María Velasco