26 de febrero de 2026

Discurso de agradecimiento en la ceremonia de entrega de los Premios SEIO–Fundación BBVA

Daniel García Rasines y G. Alastair Young

Sra. Presidenta Saliente de la SEIO, Sra. Directora de RRII de la Fundación BBVA, autoridades, distinguidos compañeros de la SEIO, señoras y señores:

Vivimos en un mundo lleno de datos. Se recogen enormes cantidades en casi todas las áreas: salud, ciencias básicas, economía, finanzas, y muchas más. Contar con esa información debería ayudarnos a entender mejor el mundo y a tomar decisiones más justas y eficaces. En ese sentido, la Estadística es una herramienta central, pues convierte los datos en evidencia, y la evidencia en decisiones.

En los últimos años, sin embargo, se ha hecho visible una crisis, especialmente en las ciencias. Demasiados resultados de investigación no se replican en experimentos posteriores. Cuando esto ocurre, la sociedad paga un coste: se pierden tiempo y recursos, y se debilita la confianza en la ciencia. Y en ámbitos como la salud, la economía o las políticas públicas, ese coste puede ser muy alto.

Entre los factores que contribuyen a este problema está el uso inadecuado de herramientas estadísticas. En particular, cuando los datos son complejos o muy grandes, es habitual explorarlos primero para identificar qué aspectos merece la pena estudiar. Esa exploración es razonable y útil, pues permite focalizar el esfuerzo y gastar los recursos en los elementos más significativos del problema. El problema es que muchos métodos clásicos asumen que las preguntas se fijan antes de mirar los datos y, cuando no es así, la inferencia deja de ser válida. A este reto lo llamamos inferencia selectiva.

26 de febrero de 2026

Aquí, aparece una tensión natural entre dos fases: la fase de selección, donde queremos elegir bien qué preguntas conviene plantear, y la de inferencia, donde queremos dar respuestas a estas preguntas que sean válidas y a la vez precisas. Una solución sencilla sería dividir los datos en dos conjuntos, uno para seleccionar y otro para inferir. Sin embargo, en muchas ocasiones esto no es viable; por ejemplo, cuando hay dependencias complejas o cuando dividir reduce demasiado la información de la muestra.

En este trabajo desarrollamos una idea sencilla y general para desacoplar las fases de selección e inferencia. Esto lo conseguimos mediante la introducción de aleatoriedad controlada, de tal manera que en la fase de selección se usa una versión contaminada de los datos. Luego, la inferencia se realiza con una función aleatorizada e independiente. Es decir, se usan los mismos datos en ambas fases, pero "difuminados" de una manera que permite aplicar los métodos clásicos sin perder validez.

Esto es importante porque hace que las conclusiones sean más fiables, y por tanto mejora la evidencia para guiar decisiones en salud, regulación o evaluación de políticas, así como en grandes retos como el clima, la desigualdad o los riesgos financieros. Una Estadística moderna y rigurosa es clave para que la disponibilidad masiva de datos sea una oportunidad real, y no una fuente de errores.

Para concluir, queremos agradecer a la Fundación BBVA, a la Sociedad Española de Estadística e Investigación Operativa, y al jurado por este reconocimiento. Es un honor que este trabajo haya sido distinguido con este premio. Muchas gracias.