

Impacto de la ciencia de datos en la investigación

Pedro José Albarracín Vargas

Con la preocupación del drástico aumento de HARKing (Hypothenisizing After the Results are Known), en investigaciones científicas en las dos últimas décadas [2], la comunidad científica busca alternativas que eviten publicaciones de informes falsos y repetitivos, que sean resultados creíbles, concretos y transparentes. Para garantizar la disminución de fracasos en investigación y también, la mejora continua en la metodología aplicada, el diseño de estudio, calidad de presentación y la accesibilidad de la información, la ciencia de datos permite que estos factores sean llevados a la investigación como resultados confiables de manera que sea reproducible, abierto para el cualquier público.

Además del HARKing, los investigadores con el fin de mejorar las posibilidades de facilitar la publicación de sus resultados, obtener objetivos forzados por la manipulación de datos hasta que el resultado pase el umbral del error estadístico, el p-hacking, es otro reto que debe considerarse que deteriora la investigación. En la formación metodológica, en ausencia de desarrollar recursos educativos accesibles, suelen cometerse errores de interpretación de resultados, para esto debe hacerse revisión antes y después de la publicación por parte de un grupo calificado por parte de una revista o editorial. En el proceso de revisión, se torna lento y tiende a ser de acceso limitado para la persona interesada [1]. Luego, es necesaria la colaboración distribuida a una comunidad más amplia que facilite el trabajo de los pares y que sea más rápida la retroalimentación del investigador.

La ciencia de datos promueve la ciencia libre y así, los recursos para imponer los métodos de evaluación en mejorar la calidad de informes, la transparencia de los resultados (gracias a pruebas estadísticas que las respalda, datos adquiridos, análisis de datos e interpretación de resultados), refleja la transparencia de la ciencia de datos en ser reproducible. Debido a lo anterior, las revistas y financiadores requieren prácticas abiertas, datos abiertos, para ello ofrecen publicaciones gratuitas a sus investigadores con acceso abierto, la participación de infraestructuras que incluyen la presentación de informes

sin ánimo de lucro, que promueve la transparencia a sus investigadores. La colaboración activa de una comunidad diversa por medio de la ciencia abierta, es buen camino para el avance de divulgación de la información, más que la confianza, la credibilidad de los resultados, de modo que también se ofrezca intercambio de datos para que exista un beneficio contributivo. Es la ciencia de datos la que impulsa a una mejor formación de la mano de obra científica en métodos y alfabetización estadística, la forma para construir el ideal científico en progreso de la investigación.

Referencias

- [1] Alawi A Alsheikh-Ali y col. “Public availability of published research data in high-impact journals”. En: *PloS one* 6.9 (2011), e24357.
- [2] Mark Rubin. “The Costs of HARKing”. En: *The British Journal for the Philosophy of Science* 73 (ene. de 2022). DOI: 10.1093/bjps/axz050.