

# Avance y Perspectiva

Revista de divulgación del CINVESTAV

## ¿Existe una crisis de reproducibilidad en los estudios científicos?

Karina Galache · Monday, December 13th, 2021

Categorías: Ciencias Naturales y de la Salud, Zona Abierta

Es frecuente que algunos medios de comunicación publiquen notas que hacen referencia a artículos científicos, indicando que ciertos productos son dañinos para la salud y que después de un tiempo salga otra nota contradiciendo a la primera, afirmando que el consumo de tales productos tiene efectos saludables. Por ejemplo, la página <http://kill-or-cure.herokuapp.com/> lleva un registro de todas las notas publicadas por el diario británico Daily Mail referentes a lo que previene o causa cáncer. Ahí se puede ver que hay notas que parecen contradecirse; por ejemplo, en el caso del vino, hay múltiples notas afirmando que causa y previene el cáncer. Esto hace preguntarnos: ¿dónde está el error? ¿Se escribieron las notas sin entender el artículo científico base? ¿O los resultados publicados son erróneos? Parte del enigma es que en ocasiones las notas periodísticas se escriben descuidando aspectos clave de la investigación que dio lugar a los artículos base. Otra parte del problema, que es más difícil de resolver, proviene de los modelos y los métodos estadísticos que se usan para reportar los hallazgos.

Ésta y otras situaciones llevaron a los científicos a cuestionar la validez de los estudios publicados, y, sobre todo, su reproducibilidad en tiempo y forma razonable. Por reproducibilidad, no nos referimos a los p-valores, sino a la capacidad de encontrar resultados congruentes (reproducibles) a través de diversos estudios con las mismas condiciones y, por ende, sus conclusiones puedan ser extendidas a la población bajo estudio. Una de las discusiones más influyentes de estos aspectos [1], aseguraba que la mayoría de los estudios publicados en las ciencias experimentales tenían resultados incorrectos. Dicho escrito fue muy polémico por sostener la tesis de que la mayoría de la investigación científica no servía; con el tiempo, sus conclusiones fueron desmentidas, pues algunos de los supuestos que se empleaban en el análisis eran incorrectos. Las versiones corregidas de ese análisis indican, por ejemplo, que entre 8.6% y 17.6% de los artículos científicos en psicología tiene resultados erróneos. ¿Se debe ese porcentaje a una investigación mal desarrollada? Algunos artículos científicos pueden ser producto de investigación defectuosa, pero la principal razón en el contexto de esos análisis es el empleo de métodos y modelos estadísticos inadecuados (ver también [2,3]). Esto nos lleva a preguntarnos, ¿cuál es la tasa de falsos descubrimientos? Y, ¿a qué se puede adjudicar dicha tasa?

Antes de avanzar en la discusión, es importante aclarar algunas ideas muy generales que han ido

minando el camino para el uso correcto de las técnicas estadísticas en todas las ramas de las ciencias y humanidades. En principio, aunque no necesariamente son correctos dado que los modelos estadísticos no implican relaciones causales, tampoco son totalmente inapropiados. Por otra parte, muchas veces son aplicados sin tomar en cuenta los supuestos que dan lugar al modelo estadístico y sólo se ajustan mecánicamente. La peor parte corresponde a la forma de tratar la incertidumbre y sobre todo cómo se emplean, interpretan o comparan las medidas generadas para ello, sin tomar en cuenta la contribución científica que está detrás. Un claro ejemplo son los abusos de los p-valores que se interpretan como “nivel de significancia”, en lugar de simplemente considerarlos como evidencia a favor de una hipótesis. Esto ha implicado el rechazo de múltiples contribuciones por no pasar las reglas impuestas por las revistas, comunidades científicas y académicas y desperdiciar así un gran cúmulo de nuevos conocimientos (ver, e.g., [4-7]).

La pobreza que ha habido en el proceso de hacer reproducibles los resultados obtenidos, ha generado una crisis cada vez mayor en la efectividad de las aseveraciones científicas como las descritas anteriormente. Mediante las nuevas formas de tratar este problema desde la perspectiva estadística, así como considerando los principios de las teorías que sustentan los estudios, es factible encontrar soluciones que permitan corregir este sesgo, producido en mucho por la presión de obtener resultados que justifiquen los fondos invertidos mundialmente en investigación. Una excelente discusión sobre el problema de reproducibilidad de resultados, a través de la repetición de los experimentos, puede encontrarse en la página:

<https://www.nature.com/articles/d41586-021-02486-7>.

### Tasa de falsos descubrimientos

Para encontrar posibles soluciones, conviene entender primero por qué se origina este problema y, segundo, cómo tratar de mitigar el efecto. Para ello es importante definir lo que se llama la tasa de falsos descubrimientos (TDF), ver [5], la cual se puede describir como la proporción de descubrimientos erróneos en un conjunto de pruebas de hipótesis. Por ejemplo, consideremos un estudio en el que se intenta determinar los genes asociados con una enfermedad para una población determinada; por simple azar, incluso bajo los controles más estrictos, siempre existe la posibilidad de que las pruebas estadísticas por sí solas indiquen erróneamente que algunos genes tienen asociación con la enfermedad. La proporción de genes equívocamente asociados sería, en este caso, la tasa de falsos descubrimientos del estudio. Aquí es importante entender que esto ocurre incluso en los estudios más controlados y aunque se usen correctamente las mejores pruebas de hipótesis disponibles, el causante de estos errores es la aleatoriedad inherente en el estudio, que es producida, en parte, por la enorme cantidad de factores que intervienen en el proceso y refleja la heterogeneidad de los individuos de la muestra. No hay forma de eliminar esta aleatoriedad. Entonces, ¿cómo podemos confiar en los resultados de los estudios científicos si siempre hay errores que no se pueden eliminar? Primero, los hallazgos requieren ser validados varias veces para poder decir que en efecto son plausibles; un solo estudio, por más extenso que sea, es insuficiente para proveer evidencia irrefutable. Esto sirve como filtro para estos errores. El segundo punto es que, si bien la tasa de falsos descubrimientos no se puede eliminar, sí se puede controlar. Es decir, planteando un buen estudio y considerando las técnicas estadísticas apropiadas, es posible acotar este error. Obviamente, entre menor sea la cota, se requerirán mayores controles y recursos en el estudio. Sin embargo, no es necesario controlar la tasa de falsos descubrimientos hasta valores ínfimos, pues, como ya dijimos, un estudio por sí solo es insuficiente para establecer asociaciones

irrefutables.

Entre las técnicas estadísticas que permiten controlar la TFD, se encuentra la generación de copias sintéticas de las variables bajo estudio que reproduzcan su comportamiento conjunto, ver [8,9]. La idea de estas técnicas, a grandes rasgos, es producir copias de las variables originales, y en muchas ocasiones, estas copias pueden tener incluso información más relevante con respecto al fenómeno bajo estudio que las variables originales. Esto resulta equivalente a pensar como si el experimento global se repitiese. Por otra parte, las muestras se diseñan de tal forma que puedan controlar la tasa de falsos descubrimientos fijando una cota, dando lugar a que los estudios sean reproducibles. Este aspecto fundamental es ignorado en la mayoría de los estudios.

Los problemas actuales en muchas áreas científicas, tecnológicas y sociales implican un volumen muy alto de información en diversos sentidos, estructurados o no estructurados que no fueron en principio recolectados para responder a preguntas o hipótesis científicas, sino simplemente, están disponibles, lo que hace más complejo el problema de replicabilidad de los resultados. Tal vez donde se tienen mayores avances corresponde a problemas donde existe una alta dimensionalidad, aún bajo estructuras de datos en forma de matrices o cubos, por mencionar una estructura. Las ideas antes mencionadas sobre el control de la TDF, han sido extendidas a estructuras cada vez más complejas, permitiendo mayor flexibilidad en el tratamiento de las copias sintéticas, así como selección de variables (ver [9]). Muchos de estos avances han sido producidos en áreas de trabajo multidisciplinarias, no solamente por estadísticos, sino también por otros científicos y sus respectivos laboratorios trabajando en equipo. Otro de los retos que se persigue es la inclusión de otras medidas de dependencia más allá de correlaciones, esto es, dejando atrás las asociaciones lineales y los modelos aditivos. Esta nueva forma de pensar en el uso correcto de la estadística y la validez de los resultados científicos implica un cambio de cultura que comienza con los académicos, las revistas científicas y los científicos en general.

¿Es un desafío que vamos a seguir soslayando?

---

## Referencias

1. Ioannidis, J.P.A. (2005). [Why Most Published Research Findings Are False](#). PLOS Medicine. 2 (8).
2. He, F.J. y MacGregor, G.A. (2018). Role of salt intake in prevention of cardiovascular disease: controversies and challenges. *Nat Rev Cardiol* 15, pp. 371–377.
3. Jager, L.R. y Leek, J.T. (2014). [An estimate of the science-wise false discovery rate and application to the top medical literature](#). Oxford Academic. pp. 1–12.
4. Wasserstein R.L. y Lazar N.A. (2016). [The ASA’s statement on p-values: context, process, and purpose](#). *The American Statistician*. 70 (2): pp. 129–133.
5. Colquhoun D. (2014). [An investigation of the false discovery rate and the misinterpretation of p-values](#). *Royal Society Open Science*. 1 (3): 140216. arXiv:5296.
6. Senn, S. (2001). OPINION: Two cheers for P-values?. *Journal of Epidemiology and Biostatistics* 6 (2), 193–204.
7. Wasserstein, R.L., Schirm, A.L. y Lazar, N.A. (2019). [Moving to a World Beyond “p<?0.05”](#). *The American Statistician*. 73 (1): pp. 1–19.

8. Barber, R..F. y Candès, E.J. (2015). Controlling the false discovery rate via knockoffs. *The Annals of Statistics*. 43(5): pp. 2055-2085.
9. Sesia, M., Sabatti, C., y Candès, E.J. (2019) Gene hunting with hidden Markov model knockoffs. *106 (1)*: pp. 1–18.

[Foto de portada de Markus Winkler](#)

This entry was posted on Monday, December 13th, 2021 at 11:20 am and is filed under [Ciencias Naturales y de la Salud](#), [Zona Abierta](#)

You can follow any responses to this entry through the [Comments \(RSS\)](#) feed. Both comments and pings are currently closed.