

Juan Seguro

3Dresyns by Resyner Technologies S.L.

© 2026 Juan Seguro. Todos los derechos reservados.

Correspondencia: info@3Dresyns.com · www.3Dresyns.com

INF-003 · Inferencia Científica y Razonamiento Experimental · Versión v0.2

Resumen

La inferencia se realiza a partir de las observaciones que llegan al conjunto de datos analizado, no de las observaciones que podrían haber existido bajo un proceso ideal de muestreo y medición. La entrada en el conjunto de datos puede depender de la participación, la supervivencia, la detección, el seguimiento, la integridad del espécimen, las reglas de control de calidad, los requisitos de casos completos o las decisiones del analista. Cuando la selección depende de variables relacionadas con la exposición, el resultado o sus causas, el conjunto analizado puede diferir sistemáticamente de la población objetivo. Condicionar por un colisionador es un mecanismo estructural capaz de generar esa distorsión, pero el sesgo de selección es más amplio que la terminología de colisionadores. Los datos faltantes añaden un problema relacionado: la validez del análisis de casos completos, la ponderación, la imputación o el análisis de sensibilidad depende del proceso de ausencia y del estimando objetivo. Este artículo separa selección, ausencia, supervivencia y selección inducida por colisionador, y proporciona una ruta disciplinada desde la población objetivo hasta la muestra analizada. No trata la no detección y la censura como intercambiables con los datos faltantes; esos problemas diagnósticos de medición permanecen en HEB-007.

Palabras clave: sesgo de selección; datos faltantes; sesgo de supervivencia; sesgo de colisionador; análisis de casos completos; participación; seguimiento; población objetivo

1. El conjunto de datos analizado es un estado seleccionado

Todo conjunto de datos representa una secuencia de eventos de inclusión. Algunas unidades nunca se muestrean; algunas rechazan participar; algunas se pierden durante el seguimiento; algunos especímenes no superan el control de calidad; algunos resultados no se miden; algunas observaciones se eliminan porque faltan covariables requeridas. La pregunta científica, por tanto, no es solo qué muestran los valores observados, sino por qué esas observaciones concretas llegaron a estar disponibles para el análisis.

2. La selección es un proceso causal

Un enfoque estructural trata la selección como parte del sistema causal y no como una propiedad genérica del tamaño muestral. Hernán, Hernández-Díaz y Robins mostraron que el sesgo de selección puede representarse mediante condicionamiento por efectos comunes u otras estructuras que crean asociaciones no causales [1]. Cole y colaboradores ilustraron cómo condicionar por un colisionador puede inducir una asociación entre causas que, de otro modo, serían independientes [2]. La consecuencia práctica es que la selección no puede diagnosticarse solo a partir de la similitud o disimilitud marginal de la muestra retenida.

3. La ausencia de datos no es un único mecanismo

Los datos faltantes pueden surgir por mecanismos diferentes y requieren supuestos distintos para una inferencia válida. Little y Rubin distinguen situaciones en las que la ausencia puede ignorarse bajo el modelo de análisis de aquellas en las que el proceso de ausencia debe modelizarse o explorarse mediante análisis de sensibilidad [3]. Un análisis de casos completos puede ser insesgado en algunas estructuras y estar gravemente sesgado en otras. La imputación puede recuperar información bajo modelos apropiados, pero no puede fabricar información sobre un mecanismo no observado sin supuestos.

4. Supervivencia y condicionamiento por estados posteriores a la exposición

El sesgo de supervivencia aparece cuando la inclusión exige seguir siendo observable después de un proceso que, a su vez, está relacionado con la exposición, el resultado o causas comunes. El término se utiliza a menudo de forma laxa; INF lo usa solo cuando la supervivencia o persistencia constituye un evento de selección relevante para la inferencia objetivo. Condicionar por un estado posterior a la exposición también puede inducir sesgo de selección aunque no exista supervivencia biológica literal. Algunos ejemplos son analizar solo dispositivos que superaron una etapa de cribado, pacientes que permanecieron en seguimiento o experimentos que produjeron una señal medible.

5. Selección sobre un colisionador

INF-002 define estructuralmente un colisionador como un efecto común en una vía especificada. INF-003 posee la consecuencia inferencial cuando la inclusión en el conjunto de datos condiciona por esa variable o por uno de sus descendientes. La asociación resultante no constituye evidencia de una relación causal únicamente porque aparezca tras la estratificación o selección. El estatus de colisionador depende del grafo, por lo que el mismo nombre de variable no puede etiquetarse como “un colisionador” al margen de la estructura causal [5].

6. Distinguir datos faltantes de no detección y censura

Un valor faltante puede indicar que no se obtuvo una medición, mientras que una no detección puede ser un resultado observado de un método cuya respuesta quedó por debajo de un umbral de capacidad, y la censura puede codificar información parcial sobre un valor. Estos estados contienen información diferente. HEB-007 posee el tratamiento diagnóstico de la no detección y la censura, incluida la forma de distinguirlas de un valor que no se midió. INF-003 utiliza esas distinciones únicamente para evitar aplicar procedimientos de datos faltantes al estado de observación equivocado.

7. Diagramas de selección expresados en palabras

- Defina la población objetivo o el conjunto completo de unidades elegibles.
- Enumere cada evento requerido para que una unidad entre en el conjunto de datos analizado.
- Para cada evento de selección, identifique las variables que influyen sobre él.
- Pregunte si condicionar por ese evento abre una vía no causal o cambia la población objetivo.
- Separe el muestreo previo a la exposición de la selección posterior a la exposición.
- Indique qué variables faltan y qué supuestos hacen válido el método elegido para tratarlas.
- Utilice análisis de sensibilidad cuando no puedan descartarse mecanismos alternativos plausibles de selección o ausencia.

8. Qué pueden y qué no pueden hacer la ponderación y la imputación

La ponderación por probabilidad inversa, los métodos de verosimilitud y la imputación múltiple pueden reducir el sesgo o recuperar eficiencia bajo modelos especificados y condiciones de positividad/soporte. No convierten la selección en ignorable por declaración. El análisis sigue condicionado a la adecuación del modelo de participación, ausencia o resultado [4]. Cuando determinantes no observados de la selección afectan materialmente a la relación objetivo, un análisis de sensibilidad puede ser científicamente más honesto que una única estimación corregida.

9. Reportar el estado seleccionado

Un informe defendible debe indicar la población objetivo, el conjunto elegible, las unidades reclutadas o generadas, las exclusiones, los datos faltantes por variable, los motivos de exclusión cuando se conozcan, el momento de la selección respecto de la exposición y el resultado, y los supuestos utilizados por la ponderación o la imputación. Esta información forma parte de la cadena de evidencia porque el objetivo de la inferencia puede cambiar a medida que se pierden unidades.

Referencias

- [1] Hernán, M.A.; Hernández-Díaz, S.; Robins, J.M. A Structural Approach to Selection Bias. *Epidemiology* 15(5), 615-625 (2004). DOI: 10.1097/01.ede.0000135174.63482.43.
- [2] Cole, S.R.; Platt, R.W.; Schisterman, E.F.; Chu, H.; Westreich, D.; Richardson, D.; Poole, C. Illustrating Bias Due to Conditioning on a Collider. *International Journal of Epidemiology* 39(2), 417-420 (2010). DOI: 10.1093/ije/dyp334.
- [3] Little, R.J.A.; Rubin, D.B. *Statistical Analysis with Missing Data*. 3rd ed. Wiley, 2019. DOI: 10.1002/9781119482260.
- [4] Hernán, M.A.; Robins, J.M. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC, 2020. DOI: 10.1201/9781315374932.
- [5] Pearl, J. *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge University Press, 2009. DOI: 10.1017/CBO9780511803161.

Declaración de datos y materiales

No se generó ningún conjunto de datos experimental nuevo para este artículo. El artículo es un análisis metodológico basado en la bibliografía citada y en la arquitectura de atribución de contenidos congelada de la serie INF.