

Juan Seguro

3Dresyns by Resyner Technologies S.L.

© 2026 Juan Seguro. Todos los derechos reservados.

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

INF-013 · Inferencia Científica y Razonamiento Experimental · Versión v0.2

Resumen

La inferencia científica no es un único cálculo realizado después de recoger los datos. Es una cadena controlada de compromisos que conecta una afirmación formulada con observaciones, supuestos, explicaciones alternativas, pruebas, incertidumbre y el objetivo al que se pretende aplicar la conclusión. Esta síntesis integra la atribución de contenidos desarrollada en INF-001 a INF-012 en una única secuencia de razonamiento auditable. La secuencia comienza definiendo la afirmación y el objetivo; después pregunta cómo se generaron y seleccionaron los datos, qué estructura causal o mecanística se supone, qué explicaciones rivales siguen siendo compatibles con las observaciones, con qué fuerza el resultado observado discrimina entre ellas, qué conocimiento previo restringe legítimamente la interpretación, qué pruebas podrían revelar error, cómo se separan la incertidumbre estructural y paramétrica, si las conclusiones sobreviven a variaciones analíticas defendibles, si la evidencia convergente es realmente independiente en sus modos de fallo y si la conclusión final puede transportarse más allá de las condiciones estudiadas. El punto final no es la certeza. Es una afirmación cuyo alcance, base evidencial, alternativas no resueltas e incertidumbre residual son suficientemente explícitos para poder ser inspeccionados y cuestionados.

Palabras clave: inferencia científica; razonamiento experimental; inferencia causal; evidencia; identificabilidad; pruebas severas; incertidumbre; robustez; triangulación; generalización; afirmación justificada

1. La inferencia es una cadena, no un resultado

Un resultado numérico puede ser correcto mientras que la conclusión científica asociada esté injustificada. Una estimación precisa puede responder a la pregunta equivocada; una asociación puede no identificar un efecto causal [1,2]; un modelo ajustado puede no ser único; un resultado estable puede depender de decisiones analíticas no informadas; y un resultado internamente válido puede no aplicarse al objetivo pretendido. INF trata por ello la inferencia como una cadena en la que cada eslabón tiene sus propias condiciones de fallo. El orden canónico no es burocrático. Evita que una sofisticación estadística o de modelización posterior intente reparar una ambigüedad previa sobre qué se afirmó realmente.

2. Empezar por la afirmación y el objetivo

INF-001 posee la primera disciplina: las afirmaciones descriptivas, causales, mecanísticas y predictivas son objetos diferentes y requieren evidencias diferentes. La predicción puede funcionar sin identificar un mecanismo causal, y una explicación causal no puede inferirse solo de la precisión predictiva [4]. Un flujo de trabajo defendible formula, por tanto, la afirmación antes de seleccionar la evidencia usada para respaldarla. También declara el objetivo: población, sistema, estado material, intervención, condiciones de contorno, escala y horizonte temporal a los que se pretende aplicar la conclusión. Sin un objetivo no pueden evaluarse ni la relevancia interna ni la generalización posterior.

3. Reconstruir cómo los datos observados llegaron al análisis

Antes de interpretar una asociación, INF-002 e INF-003 exigen una descripción estructural de la generación y selección de los datos. Confusores, mediadores y colisionadores ocupan roles causales diferentes; condicionar por ellos puede eliminar sesgo, bloquear parte del efecto de interés o crear sesgo según la estructura y el estimando [1,2,5]. La selección y los datos faltantes añaden un segundo filtro: las observaciones analizadas pueden ser un subconjunto no aleatorio de los estados que podrían haberse observado. La pregunta pertinente no es solo si faltan datos, sino si la entrada en el conjunto analizado depende de variables conectadas con la exposición, el resultado o sus causas comunes [5,6]. El error de medición puede distorsionar aún más las estimaciones de efecto dentro de esta estructura, y no limitarse a añadir ruido simétrico [7].

4. Hacer explícitas las explicaciones rivales

INF-004 separa observar un patrón de identificar el proceso que lo generó. Cuando varios modelos, mecanismos o estructuras causales predicen las mismas observaciones, la evidencia infradetermina la explicación. La identificabilidad estructural pregunta si el modelo y el diseño declarados determinan de forma única la magnitud o estructura de interés a partir de la distribución observable, incluso con datos ilimitados. La identificabilidad práctica pregunta si los datos finitos realmente disponibles la restringen de forma útil [8]. Las hipótesis auxiliares pertenecen al interior de esta comparación, no fuera de ella: calibración, condiciones de contorno, procesos molestos y formas de modelo pueden rescatar o derrotar una explicación aparente. Una afirmación justificada registra, por tanto, qué se ha excluido, qué sigue siendo viable y qué evidencia discriminaría entre las explicaciones restantes.

5. La fuerza de la evidencia es comparativa

INF-005 e INF-006 distinguen la existencia de un resultado de cuánto debería modificar una creencia o preferencia entre explicaciones. Un valor p no es la probabilidad de que una hipótesis sea verdadera ni una medida directa del tamaño o importancia práctica de un efecto [9]. Las razones de verosimilitud y los factores de Bayes hacen explícito el carácter

comparativo de la evidencia al preguntar lo bien que se predicen las observaciones bajo explicaciones competidoras [10,11]. El conocimiento previo puede restringir legítimamente la inferencia cuando es explícito, pertinente y está abierto a ser cuestionado; se convierte en una fuente oculta de sesgo cuando los supuestos entran informalmente después de conocerse el resultado. Las tasas base importan porque el mismo incremento evidencial puede conducir a conclusiones posteriores distintas cuando las plausibilidades iniciales difieren [11,12].

6. Preferir pruebas que podrían haber revelado que la explicación era errónea

INF-007 convierte las explicaciones rivales en palanca experimental. Un experimento útil no es simplemente el que produce otra medición; produce resultados que difieren entre las explicaciones comparadas. La inferencia fuerte y las pruebas severas comparten esta disciplina: diseñar la prueba alrededor de errores o alternativas que el procedimiento tenga una posibilidad real de revelar [3,13]. No rechazar una explicación tras una prueba insensible constituye evidencia débil. A la inversa, una intervención que modifica el discriminador predicho y el resultado aguas abajo puede reforzar la atribución causal, siempre que la intervención no altere simultáneamente vías no controladas. La falsación operativa es, por tanto, local y específica de la prueba, no una declaración universal de que una teoría ha quedado permanentemente refutada.

7. Separar la incertidumbre en valores de la incertidumbre en la estructura

INF-008 distingue la incertidumbre paramétrica dentro de un modelo aceptado de la incertidumbre estructural sobre el propio modelo. Intervalos de confianza o posteriores estrechos no pueden expresar la incertidumbre creada por un mecanismo omitido, una forma funcional errónea, una condición de contorno incorrecta o una estructura no identificable. La calibración del modelo también puede confundir incertidumbre paramétrica con discrepancia de modelo si esta última no está representada [14]. El análisis de sensibilidad solo es informativo cuando varía los supuestos que importan para la conclusión; no puede sustituir la consideración de estructuras alternativas que nunca se incluyeron [15].

8. Preguntar si la conclusión sobrevive a una variación analítica defendible

INF-009 e INF-010 abordan dos problemas diferentes. El análisis de robustez varía deliberadamente supuestos plausibles, formas de modelo o especificaciones y pregunta qué conclusiones sobreviven. Los enfoques de curva de especificaciones y multiverso hacen visible esta dependencia cuando existen muchos análisis defendibles [16,17]. La multiplicidad analítica es un riesgo distinto, creado cuando exclusiones de datos, reglas de parada, resultados, covariables, transformaciones o modelos se seleccionan después de ver los resultados. La flexibilidad no declarada puede aumentar enormemente la probabilidad de presentar un resultado favorable como si estuviera determinado de forma única por los datos [18]. El preregistro puede restringir algunas formas de esta flexibilidad [19], pero no vuelve informativo un mal diseño ni elimina la necesidad de análisis de sensibilidad.

9. Integrar evidencia por independencia de debilidades, no contando estudios

INF-011 trata la triangulación como convergencia estructurada. Varias mediciones son más informativas cuando sus sesgos, supuestos y modos de fallo difieren. El acuerdo entre métodos que comparten el mismo sesgo no reconocido puede crear falsa tranquilidad, mientras que el desacuerdo entre métodos puede revelar qué supuesto requiere examen. La triangulación pregunta, por tanto, si rutas independientes restringen la misma afirmación desde direcciones distintas, no cuántas publicaciones informan una respuesta similar [20,21]. La integración de evidencias sigue siendo específica de la afirmación: las mediciones que respaldan predicción, mecanismo y causalidad no son intercambiables solo porque se refieran al mismo sistema.

10. La generalización es otra inferencia

INF-012 cierra la cadena separando la validez en las condiciones estudiadas de la aplicabilidad a un objetivo. La generalización y la transportabilidad requieren supuestos explícitos sobre modificación de efecto, soporte y qué relaciones permanecen estables entre estudio y objetivo [22,23]. La reponderación o modelización solo pueden tender un puente entre distribuciones cuando se miden las variables necesarias y el objetivo se encuentra dentro de un soporte adecuado. La extrapolación hacia estados no representados depende cada vez más de supuestos estructurales. En sistemas físicos y de ingeniería se aplica la misma lógica a escala, geometría, historia de proceso, estado del equipo y entorno: la replicación procedimental por sí sola no establece que el mismo efecto o mecanismo gobierne el nuevo objetivo.

11. Registro mínimo para una afirmación científica justificada

- Indique el tipo de afirmación y el objetivo al que se aplica.
- Describa cómo se generaron, midieron, seleccionaron y excluyeron las observaciones.
- Haga explícita la estructura causal o mecanística necesaria para la afirmación.
- Enumere explicaciones rivales creíbles e identifique qué sigue sin ser identificable.
- Indique qué aporta el resultado observado respecto de esas alternativas y del conocimiento previo pertinente.
- Identifique una prueba discriminante o severa y especifique qué error es capaz de revelar.
- Separe la incertidumbre paramétrica de la incertidumbre estructural y de la discrepancia del modelo.
- Compruebe si la conclusión sobrevive a supuestos y especificaciones analíticas defendibles.
- Integre evidencia adicional considerando la independencia de sesgos y modos de fallo.
- Justifique cualquier transferencia desde las condiciones estudiadas hasta el objetivo y marque explícitamente la extrapolación.
- Informe la conclusión más fuerte respaldada por la evidencia, junto con las alternativas no resueltas y la incertidumbre residual.

12. Regla de parada: justificado no significa definitivo

Una conclusión científica está justificada en la medida en que no se formula con más fuerza de la que permite la discriminación de la evidencia frente a las alternativas pertinentes dentro del alcance y los supuestos declarados, y en la medida en que las principales vías por las que podría ser artefactual se han hecho explícitas y se han probado mediante procedimientos capaces de revelarlas. Cuando los métodos disponibles carecen de esa capacidad, la afirmación se estrecha en consecuencia en lugar de declararse justificada. Este estándar no exige certeza ni implica permanencia. Nueva evidencia puede alterar la estructura causal, revelar selección, mostrar un mecanismo no modelizado previamente, cambiar el balance evidencial o demostrar que el resultado no se transporta. La regla de parada correcta es, por tanto, evidencial y no retórica: detenerse en la afirmación más fuerte que la evidencia presente puede sostener, no en la historia más completa que pueda escribirse.

13. Síntesis

La arquitectura INF convierte el razonamiento experimental en una secuencia inspeccionable. INF-001 define la afirmación; INF-002 e INF-003 establecen las condiciones estructurales y de selección; INF-004 decide qué explicaciones rivales permanecen activas; INF-005 e INF-006 disciplinan la actualización evidencial; INF-007 exige pruebas capaces de discriminar; INF-008 separa tipos de incertidumbre; INF-009 e INF-010 exponen la dependencia de decisiones analíticas; INF-011 integra evidencia complementaria; e INF-012 define las condiciones de transferencia a un objetivo. La serie termina, por tanto, donde un artículo científico o una decisión técnica deberían comenzar su conclusión: con una afirmación cuyo significado, evidencia, supuestos, alternativas, incertidumbre y dominio de validez son visibles. El valor del marco no es que elimine el juicio, sino que hace trazables los fundamentos científicos de ese juicio.

Referencias

- [1] Pearl, J. *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge University Press, 2009. DOI: 10.1017/CBO9780511803161.
- [2] Hernán, M.A.; Robins, J.M. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC, 2020. DOI: 10.1201/9781315374932.
- [3] Mayo, D.G. *Statistical Inference as Severe Testing: How to Get Beyond the Statistics Wars*. Cambridge University Press, 2018. DOI: 10.1017/9781107286184.
- [4] Shmueli, G. To Explain or to Predict? *Statistical Science* 25(3), 289-310 (2010). DOI: 10.1214/10-STS330.
- [5] 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.
- [6] Little, R.J.A.; Rubin, D.B. *Statistical Analysis with Missing Data*. 3rd ed. Wiley, 2019. DOI: 10.1002/9781119482260.
- [7] Carroll, R.J.; Ruppert, D.; Stefanski, L.A.; Crainiceanu, C.M. *Measurement Error in Nonlinear Models: A Modern Perspective*. 2nd ed. Chapman & Hall/CRC, 2006. DOI: 10.1201/9781420010138.
- [8] Raue, A.; Kreutz, C.; Maiwald, T.; Bachmann, J.; Schilling, M.; Klingmüller, U.; Timmer, J. Structural and Practical Identifiability Analysis of Partially Observed Dynamical Models by Exploiting the Profile Likelihood. *Bioinformatics* 25(15), 1923-1929 (2009). DOI: 10.1093/bioinformatics/btp358.
- [9] Greenland, S.; Senn, S.J.; Rothman, K.J.; Carlin, J.B.; Poole, C.; Goodman, S.N.; Altman, D.G. Statistical Tests, P Values, Confidence Intervals, and Power: A Guide to Misinterpretations. *European Journal of Epidemiology* 31, 337-350 (2016). DOI: 10.1007/s10654-016-0149-3.
- [10] Royall, R. *Statistical Evidence: A Likelihood Paradigm*. London: Chapman & Hall, 1997. DOI: 10.1201/9780203738665.
- [11] Kass, R.E.; Raftery, A.E. Bayes Factors. *Journal of the American Statistical Association* 90(430), 773-795 (1995). DOI: 10.1080/01621459.1995.10476572.
- [12] Greenland, S. Bayesian Perspectives for Epidemiological Research: I. Foundations and Basic Methods. *International Journal of Epidemiology* 35(3), 765-775 (2006). DOI: 10.1093/ije/dyi312.
- [13] Platt, J.R. Strong Inference. *Science* 146(3642), 347-353 (1964). DOI: 10.1126/science.146.3642.347.
- [14] Kennedy, M.C.; O'Hagan, A. Bayesian Calibration of Computer Models. *Journal of the Royal Statistical Society: Series B* 63(3), 425-464 (2001). DOI: 10.1111/1467-9868.00294.
- [15] Saltelli, A.; Ratto, M.; Andres, T.; Campolongo, F.; Cariboni, J.; Gatelli, D.; Saisana, M.; Tarantola, S. *Global Sensitivity Analysis: The Primer*. Wiley, 2008. DOI: 10.1002/9780470725184.
- [16] Simonsohn, U.; Simmons, J.P.; Nelson, L.D. Specification curve analysis. *Nature Human Behaviour* 4, 1208-1214 (2020). DOI: 10.1038/s41562-020-0912-z.
- [17] Steegen, S.; Tuerlinckx, F.; Gelman, A.; Vanpaemel, W. Increasing Transparency Through a Multiverse Analysis. *Perspectives on Psychological Science* 11(5), 702-712 (2016). DOI: 10.1177/1745691616658637.
- [18] Simmons, J.P.; Nelson, L.D.; Simonsohn, U. False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science* 22(11), 1359-1366 (2011). DOI: 10.1177/0956797611417632.
- [19] Nosek, B.A.; Ebersole, C.R.; DeHaven, A.C.; Mellor, D.T. The preregistration revolution. *Proceedings of the National Academy of Sciences* 115(11), 2600-2606 (2018). DOI: 10.1073/pnas.1708274114.

[20] Lawlor, D.A.; Tilling, K.; Davey Smith, G. Triangulation in aetiological epidemiology. *International Journal of Epidemiology* 45(6), 1866-1886 (2016). DOI: 10.1093/ije/dyw314.

[21] Munafò, M.R.; Davey Smith, G. Robust research needs many lines of evidence. *Nature* 553, 399-401 (2018). DOI: 10.1038/d41586-018-01023-3.

[22] Dahabreh, I.J.; Robertson, S.E.; Tchetgen Tchetgen, E.J.; Stuart, E.A.; Hernán, M.A. Generalizing causal inferences from individuals in randomized trials to all trial-eligible individuals. *Biometrics* 75(2), 685-694 (2019). DOI: 10.1111/biom.13009.

[23] Dahabreh, I.J.; Robertson, S.E.; Steingrimsson, J.A.; Stuart, E.A.; Hernán, M.A. Extending inferences from a randomized trial to a new target population. *Statistics in Medicine* 39(14), 1999-2014 (2020). DOI: 10.1002/sim.8426.

Declaración de datos y materiales

No se generó ningún conjunto de datos experimental nuevo para este artículo. El artículo es una síntesis metodológica basada en la bibliografía citada y en la arquitectura de atribución de contenidos congelada de INF-001 a INF-012.