

Barreras epistemológicas para la arquitectura de los datos y la significación en el modelo predictivo de la ciencia

Epistemological barriers for the architecture of data and the significance in the predictive model of ciencia

Recibido: agosto 20 de 2020 | Revisado: setiembre 12 de 2020 | Aceptado: octubre 02 de 2020

GEORGE ARGOTA PÉREZ ¹
RITA L. VALENZUELA HERRERA ^{2A}
GLADYS R. HUAMÁN ESPINOZA ^{2B}
ROSA AROSTE ANDÍA ³
EMILY HERNÁNDEZ HUAMANI ³
PERCY GAVILÁN CHÁVEZ ³
KONY L. DURAN LLARO ⁴

RESUMEN

Los datos y la significación representan estructuras secuenciales para la relevancia científica. El propósito del estudio fue identificar barreras epistemológicas en la arquitectura de los datos y la significación del modelo predictivo de la ciencia. El estudio se realizó desde enero hasta julio del 2020 seleccionándose mediante un muestreo probabilístico aleatorio, 100 artículos de Scopus donde se accedió a través, de la plataforma ScienceDirect como herramienta científica de búsqueda. Las estructuras secuenciales se compararon mediante la prueba t-Student considerándose significativos los resultados con un nivel de confianza del 95% y dónde se encontró diferencias entre ellas ($t = -53,88$; $p = 7,09$). Se observó que, el análisis de los datos fue menos relevante en comparación con la importancia que se atribuye a su significación. Se concluyó que, la identificación de las barreras epistemológicas para la arquitectura de los datos y la significación en el modelo predictivo de la ciencia representa una guía a considerarse para la medición de las variables y su interpretación hacia un conocimiento científico.

Palabras clave: ciencia de datos, relevancia estadística, modelos predictivos, puntos críticos, conocimiento científico

ABSTRACT

Data and the significance represent sequential structures for scientific relevance. The purpose of the study was to identify epistemological barriers in the architecture to data and the significance of the predictive model of science. The study was carried from January to July 2020, selecting 100 randomized Scopus articles through a random probabilistic sampling, using the ScienceDirect platform as a scientific search tool. Sequential structures were compared using the t-Student test, the results being considered significant with a confidence level of 95% and where differences were found between them ($t = -53.88$; $p = 7.09$). It was observed that the analysis for the data was less relevant compared to the importance attributed to its significance. It was concluded that the identification of the epistemological barriers for the architecture to data and the significance in the predictive model of science represents a guide to consider for the measurement of the variables and their interpretation towards scientific knowledge.

Key words: data science, statistical relevance, predictive models, critical points, scientific knowledge

- 1 Centro de Investigaciones Avanzadas y Formación Superior en Educación, Salud y Medio Ambiente "AMTAWI". Puno, Perú. george.argota@gmail.com
- 2 Universidad Nacional San Luis Gonzaga de Ica (UNICA). Ica-Perú.
a) Facultad de Farmacia y Bioquímica.
b) Facultad de Odontología.
- 3 Escuela de Estomatología. Universidad Privada San Juan Bautista (UPSJB). Ica-Perú.
- 4 Escuela de Posgrado. Universidad César Vallejo (UCV). Trujillo-Perú.

© Los autores. Este artículo es publicado por la Revista Campus de la Facultad de Ingeniería y Arquitectura de la Universidad de San Martín de Porres. Este artículo se distribuye en los términos de la Licencia Creative Commons Atribución No-comercial – Compartir-Igual 4.0 Internacional (<https://creativecommons.org/licenses/by/4.0/>), que permite el uso no comercial, distribución y reproducción en cualquier medio siempre que la obra original sea debidamente citada. Para uso comercial contactar a: revistacampus@usmp.pe.

<https://doi.org/10.24265/campus.2020.v25n30.08>

Introducción

La afirmación y los conceptos en cualquier disciplina son dos características del debate científico (Chan *et al.*, 2018; Peña, 2019; Kenter *et al.*, 2019) ya que se necesita de argumentos, afirmaciones, juicios, teorías críticas sobre el cómo, podría demostrarse un determinado conocimiento (Couper, 2020). Aunque las conceptualizaciones por lo general, dependen de las escuelas de pensamiento (Rawluk *et al.*, 2019; Kronenburg & Andersson, 2019) uno de los actuales paradigmas para la productividad está en la combinación de la ciencia de los datos (Ceri, 2018) y los pilares del análisis e interpretación de esos datos que proporciona la estadística y que plantea nuevos desafíos en el planteamiento de los problemas metodológicos contrapuestos con los enfoques tradicionales (Secchi, 2018).

Parece ser muy diferenciable que la ciencia de datos no es estadística *per se*, pues en las habilidades del modelado e inferencias no se enfatiza. En la primera, existe el almacén y acceso a los datos mediante algoritmos de comprensión sobre el cómo, implementarse un método de análisis elegido por cuanto, no se establece el desarrollo de teorías estadísticas (Klein *et al.*, 2014; Dunson, 2018). Algunas preguntas epistemológicas se pueden formular para asumir que algunos modelos de predicción científicos, constituyen explicaciones de la realidad pues de lo contrario, detalles significativos no se considerarían y por tanto, influirían en las inferencias.

Entre otros puede mencionarse a la falta de equilibrio en los diseños aleatorios (Little, 2011), la relación entre la observación que se establece con el

objeto de referencia seleccionado (Cox, 2018), la heterogeneidad en el análisis de poblaciones grandes (Bühlmann & Meinshausen, 2016), el reconocimiento de las réplicas y la precisión durante la agregación y estratificación (Ashley, 2016; Bühlmann & van der Geer, 2018).

El propósito del estudio fue identificar barreras epistemológicas en la arquitectura de los datos y la significación del modelo predictivo de la ciencia.

Materiales y Métodos

El estudio se realizó desde enero hasta julio del 2020 y se seleccionó mediante un muestreo probabilístico aleatorio, 100 artículos de Scopus a los cuales se accedió a través, de la plataforma ScienceDirect como herramienta científica de búsqueda (Figura 1).

Para la selección de los artículos científicos se consideró las 12 características principales de la ciencia como son: ser fáctica, contrastación empírica, objetividad, especificidad, aplicabilidad, comunicación, verificación, metódica, predictividad, utilidad, sistematicidad, legalidad. Se analizó dos estructuras metodológicas del modelo predictivo de la ciencia: los datos y la significación de la teoría desde una concepción en el pensamiento filosófico de Popper (1962, 1976), Kuhn (1962, 1982), Lakatos (1987) y Feyerabend (1958, 1989).

Se reflexionó en una simplificación estructural del probable modelo de la ciencia con basamento al programa de progreso de la propia ciencia haciéndose un análisis de las barreras epistemológicas las cuales consistieron en lo siguiente: Las

constantes preguntas sobre la generación y validación del método científico y la

verdad o falsedad de la teoría según la aprehensión del conocimiento científico.



Figura 1. Selección de artículos científicos / año / títulos / revistas

Se analizó en el programa estadístico profesional Statgraphics Centurion XVIII las barreras epistemológicas entre la arquitectura de los datos y la significación comparándose mediante la prueba t-Student. Los resultados se consideraron significativos con un nivel

de confianza del 95%.

Resultados

Se muestra las estructuras del modelo de predicción de la ciencia para el análisis de las barreras epistemológicas (Figura 2).

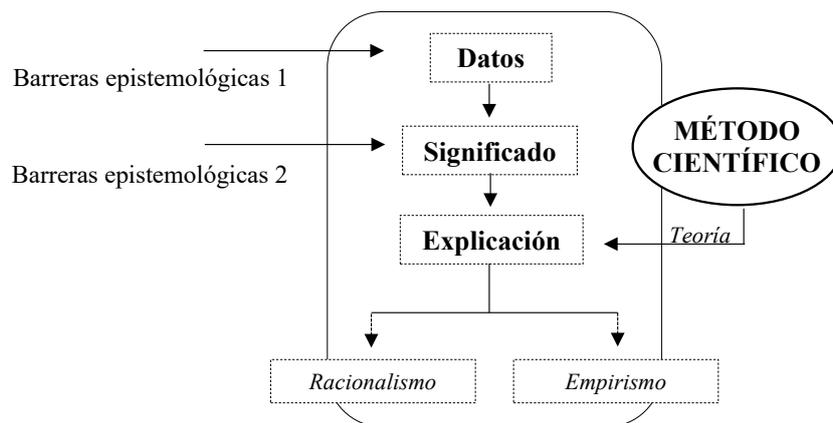


Figura 2. Modelo de predicción de la ciencia / análisis de las barreras epistemológicas

Se muestran las barreras epistemológicas según las dos estructuras

metodológicas del modelo de predicción de la ciencia (Tabla 1).

Tabla 1
Barreras epistemológicas / estructuras metodológicas

Estructuras metodológicas	Barreras epistemológicas
Datos	<ol style="list-style-type: none"> 1. ¿Se origina de una observación? 2. ¿Es orientativo de la variable? 3. ¿Pertenece a una adecuada clasificación y agrupación? 4. ¿Permitirá la comparación? 5. ¿Se reconoce para la evidencia científica?
Significado	<ol style="list-style-type: none"> 1. ¿Se permite la interpretación? 2. ¿Cómo influye en la información para generar conocimiento? 3. ¿La contrastación que se indique dará distinción al conocimiento científico? 4. ¿Se considera resaltante la interpretación ante las características de los datos agrupados? 5. ¿Con el tratamiento que se propone, los objetivos de la investigación se cumplirán?

En la prueba estadística t-Student entre la estructura metodológica del dato y la significación, el intervalo de confianza se extendió entre $-46,27$ y $-41,73$ donde la $t = -53,88$ y el valor de $p = 7,09$.

Discusión

Ante la diferencia estadísticamente significativa ($p < 0,05$) entre la estructura secuencial de los datos y la significación se observó que el análisis de los datos fue menos relevante en comparación con la importancia que se atribuye a su significación. En el modelo de predicción de la ciencia se establece el análisis de las barreras epistemológicas según los datos y el significado que espera lograrse en cualquier proceso de investigación científica pero, una de las principales dificultades está en la mediación de las variables y su posible reconocimiento para presentar datos fidedignos.

Canziani & Tullar (2017) señalan el pensamiento crítico del objeto de observación para que los datos representen hechos reales. El dato que se considera es producto del acto de aprehensión con el

objeto de investigación (Teckchandani & Khanin, 2014; Baltag *et al.*, 2019; Karini & Kamandi, 2019). Ante las barreras epistemológicas que se mostraron para la arquitectura de los datos parece que, la construcción de un probable conocimiento nuevo en el proceso de la investigación científica será limitado (Rekalde, Vizcarra & Macazaga, 2014; Red'ko, 2016; Jaime & Ladino, 2018).

El significado que se le proporcione a los datos, no necesariamente debe orientar a la evidencia científica, pues del mismo modo, existe alta probabilidad de no ofrecerse novedad en la investigación científica con el implícito riesgo del término "estadísticamente significativo" (Marshall & Hughes, 2020). Tergiversar el significado de los datos va más allá de lo ético, pues se compromete el propio futuro de los métodos en la ciencia para hallazgos con ciertos intervalos de confianza (Harris, Chivers & Drew, 2019).

Dentro de las limitaciones del estudio se menciona que otras estructuras metodológicas no se consideraron para la formulación de barreras epistemológicas y

que denoten potenciales contribuciones al desarrollo de modelos predictivos de la ciencia.

Se concluye que, la identificación de las barreras epistemológicas para la

arquitectura de los datos y la significación en el modelo predictivo de la ciencia representa una guía a considerarse para la medición de las variables y su interpretación hacia un conocimiento científico.

Referencias

- Ashley, E.A. (2016). Towards precision medicine. *Nature Reviews Genetics*; 17, 507-522. Doi: 10.1038/nrg.2016.86
- Baltag, A., Gierasimczuk, N., Özgün, A., Vargas Sandoval, A.L. & Smets, S. (2019). Una lógica dinámica para la teoría del aprendizaje. *Journal of Logical and Algebraic Methods in Programming*; 109, 100485. Doi: 10.1016/j.jlamp.2019.100485
- Bühlmann, P. & Meinshausen, N. (2016). Maging: maximin aggregation for inhomogeneous large-scale data. *Proceedings of the IEEE*; 104, 126-135. Doi: 10.1109/JPROC.2015.2494161
- Bühlmann, P. & van der Geer, S. (2018). Statistics for big data: a perspective. *Statist. Probab. Lett*; 136, 37-41. Doi: 10.1016/j.spl.2018.02.016
- CANZIANI, B. & TULLAR, W.L. (2017). DEVELOPING CRITICAL THINKING THROUGH STUDENT CONSULTING PROJECTS. *JOURNAL OF EDUCATION FOR BUSINESS*; 92, 271-279. DOI: 10.1080/08832323.2017.1345849
- Ceri, S. (2018). On the role of statistics in the era of big data: A computer science perspective. *Statistics and Probability Letters*; 136, 68-72. Doi: 10.1016/j.spl.2018.02.019
- Chan, K.M., Gould, R.K. & Pascual, U. (2018). Editorial overview: Relational values: what are they, and what's the fuss about? *Current Opinion in Environmental Sustainability*; 35, 1-7. Doi: 10.1016/j.cosust.2018.11.003
- Couper, P.R. (2020). Epistemology. *International Encyclopedia of Human Geography*. Second edition, 275–284. Doi: 10.1016/b978-0-08-102295-5.10640-7
- Cox, D.R., Kartsonaki, C. & Keogh, R.H. (2018). Big data: Some statistical issues. *Statistics Probability Letters*; 136, 111-115. Doi: 10.1016/j.spl.2018.02.015
- Dunson, D. (2018). Statistics in the big data era: Failures of the machine. *Statistics Probability Letters*; 136, 4-9. Doi: 10.1016/j.spl.2018.02.028
- Feyerabend, P. (1958). Review of mathematical foundations of quantum mechanics. By John von Neumann. *The British Journal for the Philosophy of Science*; 8(32),

- 343-347. Doi.org/10.1093/bjps/VIII.32.343
- Feyerabend, P. (1989). Realism and the historicity of knowledge. *The Journal of Philosophy*; 86(8), 393-406. Doi.org/10.2307/2026649
- Harris, S., Chivers, P. & Drew, M. (2019). The science of statistics: reporting statistical results with confidence. *Journal of Science and Medicine in Sport*; 22, Pp 5. Doi: 10.1016/j.jsams.2019.08.040
- Jaime, M.G.M. & Ladino, L.D. (2018). El método científico como alternativa didáctica de educación en valores para escuelas de ingeniería. *Formación Universitaria*; 11, 3-10. Doi: 10.4067/S0718-50062018000500003
- Karini, H. & Kamandi, A. (2019). A learning-based ontology alignment approach using inductive logic programming. *Expert Systems with Applications*; 125, 421-424. Doi: 10.1016/j.eswa.2019.02.014
- Kenter, J.O., Raymond, C.M., van Riper, C.J. & et al. (2019). Amar el desorden: navegar por la diversidad y el conflicto en los valores sociales para la sostenibilidad. *Sustainability Science*; 14, 1439-1461. Doi: 10.1007/s11625-019-00726-4
- Kleiner A., Talwalkar A., Sarkar P. & Jordan M.I. (2014). A scalable bootstrap for massive data. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*; 76, 795-816. Doi: 10.1111/RSSB.12050
- Kronenberg, J. & Andersson, E. (2019). Integrando valores sociales con otras dimensiones de valor: uso paralelo vs. combinación vs. integración total. *Science*; 14, 1283-1295. Doi: 10.1007/s11625-019-00688-7
- Kuhn, T. (1962) La Estructura de las Revoluciones Científicas. Fondo de Cultura Económica. México. 319 Pp.154
- Kuhn, T. (1982). Objetividad, juicios de valor y elección de teoría. In: Kuhn, T., (Ed.). La tensión esencial. Estudios selectos sobre la tradición y el cambio en el ámbito de la ciencia. México: FCE. 344-364.
- Lakatos, I. (1987). Historia de las ciencias y sus reconstrucciones racionales. *Comparación crítica de metodologías: La historia como prueba de sus reconstrucciones racionales*. 1ª Reimpresión. Editorial Tecnos S.A., Madrid. Pp 43-72. ISBN: 84-309-0538-3
- Little, R.J. (2011). Calibrated Bayes, for Statistics in General, and Missing Data in Particular with Discussion and Rejoinder. *Statistical Science*; 26, 162-186. Doi: 10.1214/10-STS318
- Marshall, A.P. & Hughes, I. (2020). Statistics: The grammar of science. *Australian Critical Care*; 33, 113-115. Doi: 10.1016/j.aucc.2020.02.001
- Peña, G.D.M. (2019). French historical epistemology: discourse, concepts and rules of rationality. *Studies in History and Philosophy of Science*;

- 79, 68-76. Doi: 10.1016/j.shpsa.2019.01.006
- Pooper, K.R. (1962). La lógica de la investigación científica. Madrid: Tecnos. 31-32.
- Pooper, K.R. (1976). A note on verisimilitude. *The British Journal for the Philosophy of Science*; 27(2), 147-159. www.jstor.org/stable/686164
- Rawluk, A., Ford, R., Anderson, N. & *et al.* (2019). Exploración de múltiples dimensiones de valores y valoración: un marco conceptual para mapear y traducir valores para la investigación y la práctica socioecológica. *Sustainability Science*; 14, 1187-1200. Doi: 10.1007/s11625-018-0639-1
- Red'ko, V.G. (2016). Epistemological foundations of investigation of cognitive evolution. *Biologically Inspired Cognitive Architectures*; 18, 105-115. Doi: 10.1016/j.bica.2016.10.001
- Rekalde, I., Vizcarra, M.T. & Macazaga, A.M. (2014). La observación como estrategia de investigación para construir contextos de aprendizaje y fomentar procesos participativos. *Educación XXI*; 17, 199-220. Doi: 10.5944/educxx1.17.1.1074
- Secchi, P. (2018). On the role of statistics in the era of big data: A call for a debate. *Statistics and Probability Letters*; 136, 10-14. Doi: 10.1016/j.spl.2018.02.041
- Teckchandani, A. & Khanin, D. (2014). The instructor's role in the student consulting process: Working with the student team. *Small Business Institute Journal*; 10, 11-24. Doi: 10.5465/AMBPP.2014.13187abstract

