



## SEIO y Ciencia de Datos

(Aprobado por el Consejo Ejecutivo y los Consejos Académicos de la SEIO 25/11/2024)

### 1) Origen de la Ciencia de Datos

Los orígenes se remontan a 1962, cuando John W. Tukey discutía **el futuro de la Estadística Matemática como ciencia empírica** en su artículo *"The Future of Data Analysis"*<sup>1</sup>. En 1974, Peter Naur usó el término **"Ciencia de Datos"** como sustituto de la Ciencia Computacional en su libro *Concise Survey of Computer Methods*<sup>2</sup>.

A partir de 1977, el término se fue integrando en congresos de ámbito estadístico y computacional. Ese año, se estableció la **International Association for Statistical Computing (IASC)** como una sección del International Statistical Institute (ISI): *"La misión del IASC es vincular la metodología estadística tradicional, la tecnología informática moderna y el conocimiento de los expertos en la materia, para convertir los datos en información y conocimiento."*<sup>3</sup>

William S. Cleveland fue el primero en tratar la Ciencia de Datos como una disciplina independiente en 2001: ***Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics***<sup>4</sup>.

Jim Gray (ganador de un premio Turing) planteó en 2007 un "cuarto paradigma" de la ciencia, **la ciencia intensiva en datos** (luego rebautizada como impulsada por datos), dado que "todo lo relacionado con la ciencia está cambiando debido al impacto de la tecnología de la información y el diluvio de datos".<sup>5</sup>

Actualmente, entre las sociedades científicas tradicionalmente dedicadas a la Estadística hay una corriente para incorporar la Ciencia de Datos. Por ejemplo, la Royal Statistical Society (RSS) de Reino Unido, lanzó en octubre de 2023, en colaboración con la American Statistical Association, el proyecto **"Real World Data Science"** para ofrecer un espacio donde estudiantes, profesorado, profesionales y líderes en Ciencia de Datos puedan compartir conocimiento, aprender y encontrar inspiración en aplicaciones de datos en el mundo real. La RSS ya viene incluyendo el término en sus últimos congresos anuales ([RSS International Conference 2024- the UK's annual showcase for statistics & data science](#)).

### 2) Definición de la Ciencia de Datos

Una de las primeras definiciones de Ciencia de Datos es la de Hayashi que la define como *un concepto para unificar la Estadística, el Análisis de Datos, la Informática, y sus métodos relacionados para comprender y analizar fenómenos reales con datos*<sup>6</sup>.

De entre las muchas definiciones, una de la más completas define *la Ciencia de Datos como un campo interdisciplinar que utiliza la Estadística, la Ciencia Computacional, la Programación y conocimientos especializados para recopilar, procesar y analizar datos con el fin de adquirir conocimientos o resolver un problema. La Ciencia de Datos también incluye compartir los conocimientos adquiridos a través de la visualización científica y otros medios para la comunicación. La Ciencia de Datos suele emplear métodos como el Aprendizaje Automático, la Inteligencia Artificial, el Procesamiento del Lenguaje Natural, algoritmos y otras herramientas analíticas para procesar y comprender los datos*<sup>7</sup>.

<sup>1</sup> Tukey, J.W. (1962). [The Future of Data Analysis](#). *The Annals of Mathematical Statistics* 33, 1-67.

<sup>2</sup> Naur, P. (1974). *Concise Survey of Computer Methods*. Studentlitteratur, Lund, Sweden. Petrocelli Books.

<sup>3</sup> Press, G. [«A Very Short History Of Data Science»](#). *Forbes*.

<sup>4</sup> Cleveland, W.S. (2001). [Data science: An action plan for expanding the technical areas of the field of statistics](#). *International Statistical Review*, 69, 21-26

<sup>5</sup> Gray, J. (2009). [Jim Gray on Science: A Transformed Scientific Method](#). En T. Hey, S. Tansley, & K. Tolle, *Fourth Paradigm. Data-Intensive Scientific Discovery*. Microsoft Corporation

<sup>6</sup> Hayashi, C. (1998). [What is Data Science ? Fundamental Concepts and a Heuristic Example](#). In Hayashi, C. et al. (eds.). *Data Science, Classification, and Related Methods*. Springer Japan. [doi:10.1007/978-4-431-65950-1\\_3](#)

<sup>7</sup> [National Library of Medicine](#)



### 3) Análisis DAFO SEIO y Ciencia de Datos

Se presenta a continuación un análisis DAFO (Debilidades, Amenazas, Fortalezas y Oportunidades) para explorar de forma estructurada la relación entre las disciplinas propias de la SEIO y la Ciencia de Datos.

#### - Debilidades:

- las dificultades en la aplicación y explicación de las metodologías estadísticas formales en algunos contextos de aplicación pueden transmitir una imagen que limita el alcance de las aportaciones de la comunidad estadística;
- los modelos estadísticos se formulan en ocasiones bajo hipótesis difíciles de cumplir o de comprobar para garantizar la adecuación de la inferencia, lo que reduce su capacidad de respuesta práctica;
- si bien los avances tanto en Estadística Matemática como Estadística Computacional han sido importantes en los últimos años, no se ha trabajado sobre la imagen de marca de la disciplina, que sigue muy ligada a las Matemáticas, sin alcanzar, para la sociedad en general, una visión práctica en la resolución de problemas con grandes volúmenes de datos;
- la formación básica en Estadística fuera de los ámbitos más específicos, en muchas ocasiones se limita a ejemplos pequeños o a la formalización matemática, sin llegar a transmitir su potencial para problemas reales de gran tamaño;

#### - Amenazas:

- no se identifica a las y los profesionales de la Estadística y de la Investigación Operativa como artífices de los avances en Ciencia de Datos;
- la Estadística es percibida como la Ciencia de Datos para pocos datos;
- términos ya conocidos en Estadística son renombrados y problemas ya superados en Estadística vuelven a aparecer (sesgos...), volviendo a hacer el camino que ya ha recorrido la Estadística en muchos casos, y corriendo el riesgo de quedarnos fuera de la discusión;
- en el ámbito empresarial, las y los científicos de datos tienden a extender sus desarrollos a la toma de decisiones, muchas veces sin conocimientos específicos de optimización o decisión, disminuyendo parte del potencial de la Ciencia de Datos o reinventando métodos ya existentes en Investigación Operativa, e invisibilizando esta disciplina;
- la popularidad de la Ciencia de Datos puede provocar que sociedades de otras disciplinas científico-técnicas reclamen como suyo este ámbito.

#### - Fortalezas:

- La Estadística, como disciplina académica/científica, es reconocida por la sociedad, resultando próxima a la ciudadanía, en parte por la labor de la Estadística Oficial;
- la SEIO es una sociedad reconocida y respetada como autoridad científica en Estadística, Probabilidad e Investigación Operativa;
- la Estadística es la primera Ciencia de Datos que ha habido, y tiene ya un largo camino recorrido en el trabajo con datos;
- la mayoría de los problemas abordados en Ciencia de Datos forman parte de la actividad y el conocimiento de buena parte de la comunidad de la SEIO, donde ya se han identificado dificultades y propuesto soluciones, incluso para retos futuros desde la perspectiva teórica;



- la comunidad de la SEIO cuenta con profesionales de la Ciencia de Datos en su sentido más amplio, como puede verse en la actividad de sus grupos de trabajo;
- en la SEIO recibimos con brazos abiertos nuevos enfoques para la Ciencia de Datos, y muy en especial desde la optimización.

- **Oportunidades:**

- la SEIO necesita una actualización de sus estatutos (periodos de dedicación de equipos editoriales, lenguaje inclusivo, incorporación de tesorero/a en el consejo ejecutivo, plazos y tipos de asambleas, forma de votación...), que se pretende abordar en el primer semestre de 2025, y podría aprovecharse para incorporar la Ciencia de Datos de forma explícita;
- la SEIO aunque tiene una buena implantación geográfica a nivel estatal, tiene margen de mejora para su implantación interdisciplinar, y, hoy en día, el mayor alcance se puede conseguir integrando de forma explícita la disciplina de Ciencia de Datos y atrayendo a la comunidad dedicada a ella;
- la incorporación explícita de la Ciencia de Datos en la SEIO puede atraer a una comunidad orientada a esta disciplina que ahora no se identifica con los objetivos de la SEIO, fomentando el diálogo, el intercambio de conocimiento y la colaboración en beneficio mutuo;
- cada vez hay más datos e interés en extraer información de ellos, y la SEIO, en el marco de sus competencias como agente dinamizador de la investigación en Estadística e Investigación Operativa, puede aprovechar el auge de la Ciencia de Datos para acercar los retos y necesidades de esta disciplina a su comunidad;
- cada vez hay más interés en aprovechar la información obtenida mediante la Ciencia de Datos en la toma de decisiones posteriores ("*data-driven decision making*") lo que amplía las oportunidades profesionales e investigadoras para la Investigación Operativa;
- la optimización se está haciendo cada vez más necesaria para la Ciencia de Datos, con un crecimiento exponencial del interés por sus aplicaciones en Ciencia de Datos, y la SEIO es la sociedad científica que cuenta con las personas expertas en optimización.

#### 4) **Sondeo y posible modificación de estatutos de la SEIO**

Los Consejos Ejecutivo y Académicos de la SEIO lanzan un sondeo entre los miembros de la Sociedad para recabar su opinión ante una posible iniciativa para incorporar la Ciencia de Datos entre las disciplinas que comprende la SEIO, aprovechando la modificación de estatutos que se pretende abordar en 2025. Este sondeo presenta alternativas para el nombre, acrónimo y estructura en secciones de la SEIO, incluyendo también la opción de no modificar ninguno de estos aspectos.

Los Consejos de la SEIO tienen una propuesta, pero del resultado del sondeo dependerá la propuesta final con la que se desarrollen los estatutos que se someterán a aprobación en una Asamblea General específica durante el próximo Congreso Nacional de la SEIO (Lleida, 10 a 13 de junio de 2025). Dicha propuesta es:

- Incorporar explícitamente la Ciencia de Datos como disciplina de la SEIO
- Nombre de la Sociedad: Sociedad de Estadística, Investigación Operativa y Ciencia de Datos
- Acrónimo: SEIO
- Secciones: Estadística y Ciencia de Datos, Investigación Operativa

Para que la información obtenida del sondeo sirva para tomar decisiones es necesaria la máxima participación. **¡¡¡Por favor participa!!!**