

Estadística y Ciencia de Datos

¿Qué hay de nuevo?

Statistics and Data Science What's New?

Hugo Casanova*

Escuela Venezolana de Planificación - EVP (Venezuela)

casanovade@gmail.com

Resumen

El Cuarto Paradigma de la Ciencia o Ciencia de Datos ha sido explosivo acompañando al desarrollo de las megatendencias de la Industria 4.0. Aquel sería, continuidad en brecha paradigmática, del desarrollo histórico de los precedentes vinculados con las ciencias observacionales, teóricas y experimentales en la construcción del mundo digital a partir del analógico, focalizándose desde una filosofía de la tecnología. Sus postulantes la asocian con la Estadística diferenciándolas por la influencia de la Informática. Esto es comprensible si convergemos la asincronía de la relación Estadística-Informática en contexto del desarrollo de la ciencia y su impacto en el Estado-Nación. La Estadística tiene claramente origen político como descriptora de la sociedad a partir del dato como elemento hermenéutico que fluye por estos tipos de ciencia. Aquí reconstruimos esa historia a partir de la Estadística y su transformación en Ciencia de Datos teniendo de fondo y objeto al Estado-Nación.

Palabras clave: estadística, ciencia de datos, revolución de datos

Código JEL: O33.

Recibido: 30/1/22

Aceptado: 29/04/22

Abstract

The Fourth Paradigm of Science or Data Science has been explosive accompanying the development of the megatrends of Industry 4.0. It would be a continuity in paradigmatic gap, of the historical development of the precedents linked to the observational, theoretical and experimental sciences in the construction of the digital world from the analogical one, focusing from a philosophy of technology. Its proponents associate it with Statistics, differentiating them by the influence of Computer Science. This is understandable if we converge the asynchrony of the Statistics-Informatics relationship in the context of the development of science and its impact on the Nation-State. Statistics clearly has a political origin as a descriptor of society based on data as a hermeneutic element that flows through these types of science. Here we reconstruct this history starting from Statistics and its transformation into Data Science with the Nation-State as its background and object.

Keywords: statistics, data science, data revolution.

JEL Code: O33.

Received: 30/1/22

Accepted: 29/04/22

*. Lic. en Cs Estadísticas, doctorando en Cs Sociales. CEAP-UCV. Docente-Investigador y profesor agregado en Escuela Venezolana de Planificación (EVP). ORCID: <https://orcid.org/0000-0001-7393-5163>.

1. Planteamiento del Problema

La nueva Data Science ha cobrado un gran empuje no solo en las redes sociales sino en las universidades, muchas de ellas se aprestan a crear y ofrecer maestrías y especializaciones en la nueva disciplina y actualmente hay una extensa difusión del tema asociado con la Estadística por el tratamiento de los datos, y como quiera que esta disciplina se ha entendido como ciencia de datos por su origen empírico, habría algo que discernir. Sin embargo, el asunto es más complejo aún pues la *Data Science*, como veremos, es definida como un campo de la eScience, que según los autores de *Fourth Paradigm* se realiza en las *Webs* (1.0 a 6.0). Además, este Cuarto Paradigma de la Ciencia viene a pelo con el desarrollo sin precedentes de la *Industry 4.0* (Cuarta Revolución Industrial). De modo que el dato parece cobrar una relevancia metodológica más allá de ser evidencia científica que argumenta en favor de las hipótesis, pues siendo lo que fluye por Internet, es igualmente elemento realizador de hipótesis y teorías en tecnología, bajo un impacto socioeconómico y político que está en pleno desarrollo.

Curiosamente la Estadística originalmente siendo definida en s. XVIII como una ciencia política de descripción del Estado, explicando su desarrollo desde el dato y en armonía con las nascentes teorías económicas, sufrió un primer giro hacia el dato por la matematización de la ciencia, pero que pese a que las tendencias filosóficas la arrinconaron como ciencia de datos, siguió describiendo al Estado en un efecto de Rey desnudo, pues si algo muestran con elocuencia los organismos internacionales son datos sobre el desempeño de los países. La vinculación parece evidente, de modo que hay que deconstruir esa historia de la Estadística en su impacto socioeconómico y ver su desempeño original asociado con este giro científico de la política y de la economía desde el dato o ciencia empírica. Esto es, el dato es como el

ADN de la ciencia que se convierte en información para transformar el cuerpo social bajo la intencionalidad del Estado y de la Sociedad.

2. Antecedentes

Existen revistas especializadas en el tema como el *Journal de Data Science*, revista oficial del centro de estadística aplicada, escuela de estadística, Universidad Renmin de China; la *Revista internacional de ciencia de datos y análisis* de la Universidad de Tecnología de Sídney, Australia y la *Revista de Educación en Estadística y Ciencia de Datos* de la Asociación Estadounidense de Estadística (ASA) todas de reciente creación; el Instituto Internacional de Estadística (ISI) estableció la Asociación Internacional de Computación Estadística como una de sus secciones. Para Jiménez (2020) este fenómeno posiblemente se deba al florecimiento del mundo Startup, formado por empresas que promueven un rápido crecimiento usando TIC's (Google, Facebook, Twitter, entre otras) fenómeno que puede observarse en EEUU o en algunos países de Europa y Asia.

En Twitter¹ alguien bromeaba con la frase “A data scientist is a statistician who lives in San Francisco” o en Silicon Valley y, por otra parte, en términos operativos se vincula al científico de datos con el estadístico a través de la informática como lo hace Wills (2012) “A data scientist is someone who is better at statistics than any software engineer and better at software engineering than any statisticians”; esto es, son científicos de datos los que de entre todos los estadísticos dominan más la informática y los que entre los informáticos saben más de estadística, de modo que la Informática haría un doble efecto, de deslinde en términos profesionales en tanto que alguien conozca más de una o de otra y de efecto unificador disciplinalmente, la Ciencia de Datos (CD) vendría de dos

1. Disponible en: <https://twitter.com/cdixon/status/428914681911070720>

corrientes, una más longeva, la Estadística (E)² y otra actual, la Informática (I)³ en lo que tiene visos de continuidad. Según Jeff Wu (1949-) la Estadística debe llamarse ahora Ciencia de Datos y los estadísticos, científicos de datos, si es así ¿qué justifica este salto nominal? Esto se viene mostrando como conjunción de herramientas, explosión tecnológica; creemos que hay algo sustantivo, más que cambio nominal pues, en otra ocasión, a la Estadística se la ha definido como la *tecnología del método científico* (INEGI. Instituto Nacional de Estadística y Geografía, 2014, pág. 1) o para Rao (1995, pág. 365) *Información*, salto de dato a información, por su parte, el nombre Ciencia de Datos proviene del estadístico Jeff Wu (1949-) y del informático Willians Cleveland (1943-) existiendo otros antecedentes directos en lo que luce como un desarrollo de la Informática que reclama a la Estadística. Esto es, propiamente la nueva ciencia no tendría nada propio que no sea fruto de un largo desarrollo de otras dos disciplinas, la Estadística y la Informática madurando y desprendiéndose con vida propia, pero quedaría la pregunta ¿es realmente una nueva ciencia o una ramificación de otra?, cosa no tan sencilla de responder pues las ciencias sufren mutaciones por transversalización y conjunción de temas entre ellas. Es el caso de la Estadística y la Informática que recientemente se han conjugado.

En este artículo exploraremos la relación entre Estadística a Informática desde la fundación de la primera como ciencia política en el siglo XVII, el largo debate sobre esta contra el de metodología de la ciencia en el s. XVIII, que produjo una sospechosa dualidad como

2. Que en forma general agrupa a la probabilidad, la Teoría del Riesgo, la Inferencia Inductiva, Análisis Multivariante, Data Mining, Análisis de Datos, entre otras.
3. Como Informática, grosso modo, incluimos por brevedad, Big Data, Dara Mining, Análisis de Datos, entre otras herramientas computacionales o algorítmicas. En términos de la Estadística sería mecanización de sus algoritmos dándoles eficacia

métodos cualitativos contra cuantitativos y que acarreo su atasco metodológico con la caída de la Concepción Heredada a finales del s. XX, su oculto relanzamiento desde el s. XIX con técnicas que requirieron paulatina mecanización de sus procesos, hasta la maduración de la Informática que, con la Big Data e Internet, hicieron de la Estadística, Ciencia de Datos, esta vez reclamada desde la Informática y sin perder su costado metodológico de ciencia empírica. En términos de su objeto, la Estadística, describiendo extensivamente al Estado, incluye luego a la sociedad describiendo y prescribiendo al ser humano y ahora a la sociedad comercial moderna globalizada en las redes e Internet.

3. Nota Metodológica

La IV Revolución Industrial (*Industry 4.0*), asociada con el desarrollo tecnológico actual tiene una infraestructura que si bien toca al sistema-mundo real de la economía, la política y la sociedad, parece asentarse en un sistema-mundo que, por no llamarlo virtual, se coloca como una capa sobre el mundo real sustrayendo de este las comunicaciones, la información y en general todo aquello que parecía depender del mundo tetradimensional (3-D del espacio y 1-D del tiempo) atado a mecanismos que llamaremos por rapidez, Cara a Cara, para integrarlo con otros que, minimizando este espacio-tiempo, hace que se encuentren las personas Mente a Mente (preterido el cuerpo). El impacto de *Industry 4.0* es creciente “integrando los sistemas de fabricación físicos y virtuales” (Schwab, 2016, pág. s/n) desarrollo desigual respecto de las revoluciones 2.0 y 3.0 caracterizadas grosso modo por cadenas de montaje basadas en redes eléctricas y en controladores programables. En esta ocasión el cambio de producción se basa en tecnologías online y en la fábrica inteligente.

Ahora bien, si las revoluciones industriales anteriores tuvieron de base desarrollos científicos asociados con la mecánica, la electricidad y

la digitalización, a estos avances científicos no se los llamó *revolución*, esto es, no hay diacronía entre las revoluciones industriales y las científicas⁴ que son vistas más en el esquema de Kuhn como cambios de *Paradigmas* que obviamente afectan a la economía pero que se asocian más con formas de pensar la realidad. Sin embargo, asociado a *Industry 4.0* aparece, desde la ciencia, un Cuarto Paradigma que tiene poco que ver con las clásicas revoluciones científicas de Kuhn. Hey, et all. (2009) exponen en su libro *Fourth Paradigm* todo este nuevo desarrollo vinculado con un Tsunami de Datos, llamada *Data Science* o *Ciencia Intensiva de Datos* y curiosamente mientras que los paradigmas de la ciencia se dan desde visiones generales, cambios de formas de pensar los fenómenos, conceptualismo, idealismo, la revolución de datos viene del empirismo debido al proceso metodológico que propone. Gray (2009, pág. xviii) al hablar de este Cuarto Paradigma lo define en un proceso histórico con tres paradigmas⁵ precedentes, exponemos el esquema tal como lo escribe en el artículo del libro.

1. Hace mil años: la ciencia era empírica (describía fenómenos)
2. Últimos cientos de años: rama teórica (uso de modelos, generalizaciones)
3. Últimas décadas: una rama computacional (simulando fenómenos complejos)
4. Hoy: exploración de datos (eScience) (unifica teoría, experimento y simulación.
 - 4.1. Datos capturados por instrumentos o generados por simulador
 - 4.2. Procesado por software
 - 4.3. Almacén de información/conocimiento en computadora
 - 4.4. El científico analiza la base de datos/archivos utilizando la gestión de datos y las estadísticas.

3. Separación que tiene que ver con la dualidad teoría-práctica.

4. No vamos a discutir el término Paradigma de lo cual hay mucha literatura y es comprendido como una forma de pensamiento en ciencia normal.

El esquema es bastante elocuente y ciertamente tiene un aire histórico que puede defenderse en un sentido que veremos, lo que no puede defenderse es su continuidad como automatización de la investigación o uso de estos recursos para mejorar el desempeño científico, pues, si el paso del primer paradigma al segundo se justifica por vía de extender la descripción de los fenómenos observados a generalizaciones empíricas, que luego se sostuvieron como leyes; y el paso del segundo al tercero como desarrollo del cómputo matemático mediante la lógica y la automatización de la modelación, el paso del segundo y tercero al cuarto es más difícil en la misma línea argumentativa, pues acá hablamos de aprendizaje de máquinas, modificación genética, entre otras formas tecnológicas que son más que aplicaciones de la ciencia, Schwab (2016) lo va a expresar claramente en el punto 1.2 de su libro “La premisa de este libro es que la tecnología y la digitalización lo revolucionarán todo, lo cual validará el trillado refrán «esta vez será diferente»” [comillas angulares del autor], la pregunta deviene obvia, no solo por la diferencia pues los hechos están ahí, sino otra ¿podemos seguir pensando la tecnología como la combinación de ciencia y técnica? ¿No será la propia ciencia, técnica de transformación de todo? Esa parece ser la real diferencia. Las mega-tendencias físicas que describe el autor, por citar solo una, la componen los vehículos autónomos, la impresión 3D, la robótica avanzada y los nuevos materiales. La explicación tendría que venir de una filosofía de la tecnología que englobe estos paradigmas del *Fourth Paradigm* y de la *Industry 4.0*, pues hablar de positivismo, relativismo, etc., no va a ser satisfactorio.

Ahora bien, para Gray (2009) esta fusión de paradigmas que implica el *Fourth Paradigm* lo ubica como eScience y al preguntarse por el qué, responde que es el “lugar donde las tecnologías de información se encuentran con los científicos”, las webs 1.0 a 6.0. Por otra parte, Fox y Hendler (2009, pág. 147) van a decir que

la Ciencia Intensiva de Datos (o Data Science) es uno de los componentes de la eScience dado que lo que fluye por las Webs son datos en sus múltiples formas y como hasta ahora la Estadística ha sido conocida como ciencia de datos veremos por qué el predicado se ha puesto como sujeto usando una episteme tecnológica que explica la ciencia en su decurso hacia tecnología.

Madrid Casado (2019) haciendo una reseña del libro de Martín Jiménez (Filosofía de la Técnica y de la Tecnología, 2018) expresa que este hallando ocho modelos de explicación de lo tecnológico, la Filosofía de la Técnica de G. Bacca (1901-1992) sería un precedente del octavo sosteniendo “que las técnicas son el camino a las ciencias y el germen del ser humano y su mundo”. Aretxaga (1998) anteriormente había expuesto que este filósofo de la técnica⁶ sintetiza las tres tradiciones generales sobre el tema, esto es, las filosofías de la tecnología desde las perspectivas de la ingeniería, de las ciencias sociales y de la Teoría Crítica, síntesis que se manifiesta en su apuesta global por un ser humano recreador de lo natural al querer transustanciar todo en artificial “como la tendencia ascendente superior del universo capaz de fundar el dominio artificial-técnico y extenderlo progresivamente en detrimento del natural consiguiendo imponerse a él al recrearlo en mundo *para sí*” [cursivas del autor] (Aretxaga Burgos, 1998, pág. 68).

Para Bacca (1984) es posible construir una historia epistémica de la ciencia en capas superpuestas de ciencia natural- observacional, experimental y sintética (síntesis de ciencia y técnica) o ciencias empíricas las dos primeras y trascendentales las dos últimas, como un Plan que va desde lo sensible-entendible, hasta

lo estrictamente construido, la ciencia iría de mayor a menor positividad, en acto creador de artefactos sin anularse mutuamente, como no lo hacen, obviamente con su distancia tecno-cognitiva, los remedios caseros y las tabletas de elevado tratamiento tecnológico o la tracción a sangre, la combustión interna o el motor solar, habiendo sí un cambio de eficacia; grosso modo, el calor como fenómeno sensible estaría en ciencia natural, a escala humana, manifestándose con toda su crudeza y sus usos, por ejemplo, para cocción; ahora, a partir de la teoría del calórico se haría ciencia observacional (existiría un fluido o energía que atraviesa los cuerpos), pasando a experimental con la construcción del calorímetro y toda la teoría moderna de las unidades físicas, etc.; con las tecnologías del láser, dispositivos de efecto fotoeléctrico automáticos, entre otros, se haría ciencia sintética adquiriendo la categoría de *cosa en sí*, el costado antropológico lo halla al manifestar que esas capas implican una paulatina cesión de racionalidad hacia la “cosa”, va a decir Bacca (1984) que la ciencia es supra e infra estructurada, objetivadora y entitativa usando la metáfora de la estructura del átomo en una metodología que propone al dato como elemento que fluye entre las capas como saltos epistémicos. De modo que, sin contradecir la estructura de la ciencia basada en filosofías conceptualistas propone que el conocimiento producido se trasciende en transformar el mundo de natural en artificial, por lo tanto, Gray (2009) y todos los que escribieron *Fourth Paradigm*, además de Schwab (2016) parecen estar en esa línea de pensamiento filosófico. Veamos ahora el desempeño de la Estadística bajo esta concepción constructiva y en la narrativa daremos algunas características de cada nivel de ciencia. Primero la relación entre Estadística e Informática y luego su desempeño en las diferentes capas de la ciencia.

6. Bacca no distingue entre técnica y tecnología, cosa que no discutiremos pues nuestra referencia es sobre lo que se conoce actualmente como tecnología o conjugación entre ciencia y técnica

4. Relación entre Estadística e Informática

Pese al debilitamiento del positivismo lógico a partir de la segunda mitad del siglo XX que pareció dejar de lado a *la frecuencia de datos* como unidad de conocimiento, la Informática las toma como primer procesamiento de las etiquetas de los predicados en su construcción de bases de datos, cosa que la Estadística ya hacía en sus inicios en la antigüedad. Moscoloni (2011, pág. 16) haciendo un recuento histórico de los métodos estadísticos desde la antigüedad dice que estos habrían evolucionado, respecto de su objetivo, en tres grandes grupos, 1. Enumerar, inventariar, administrar, 2. Medir, describir, comparar, resumir y 3. Predecir y dominar la incertidumbre. Estos tres momentos no podrían ser excluyentes, el primero, dice, se refería a los censos de los Estados, incluyendo al registro civil y a las oficinas públicas, de modo que este sería un largo proceso hasta entrada la modernidad, s. XVII (coincidiendo con la formación del Estado Nación, la expansión del capital y desarrollo de la ciencia); el segundo, lo refiere a la Estadística Alemana y a la Aritmética Política en el s. XVI-II (desarrollo de los Estados europeos, de sus colonias) y la tercera a partir de la probabilidad s. XIX (matematización de la Estadística y desarrollo de la tecnología). Por su parte Efron (1988) al hacer una síntesis de la historia de la Estadística moderna coloca en los vértices de un triángulo a las Aplicaciones, la Matemática y la Computación siendo el lado de los vértices Aplicaciones-Matemática la etapa teórica, el lado de los vértices Matemática-Computación la etapa metodológica y allí se detiene, pues del lado de los vértices Computación-Aplicaciones no dice nada, coloca signos de interrogación pues dice que no sabe hacia donde se dirige. Habría que agregar el desarrollo del Análisis de Datos y de la Big Data.

Respecto de la Informática, Da Costa (1998, págs. 220-221) ubica un largo período

de Prehistoria desde las cavernas hasta Charles Babbage (s. XIX), una *Antigüedad* hasta el primer cuarto del s. XX con el nacimiento de IBM; una etapa de *Modernidad* hasta 1944 (con la Mark I electromecánica y la ENIAC primera digital) y una etapa *Contemporánea* hasta la actualidad (habría que actualizar los períodos hasta la computación cuántica que se inicia desde 2020 o un poco antes). Ahora bien, para comparar estas dos líneas de desarrollo tomaremos en cuenta que Da Costa (1998, pág. 218) parte su narrativa de la historia de la Informática definiendo *información* como el agregado de, *herramientas*, formas de expresión, soporte y datos; de las cuales las herramientas, desde la antigüedad hasta hoy, son la pintura, la escritura, el ordenador; la *forma de expresión*, los dibujos, los símbolos, caracteres, el lenguaje; el soporte, las paredes, las tablillas de arcilla, el papel, el soporte magnético, entre otros y los *datos* las formas simbólicas de lo que se transmite o se graba.

El problema para establecer una relación epistemológica entre ambas (no mero soporte instrumental) estriba en que la Estadística, como ha sido narrada, tiene su asiento en la Ciencia, mientras que la Informática, en la Tecnología, de modo que luce como una *aplicación* de esta hacia aquella por el desarrollo algorítmico; esto es problemático y confunde, no quedando claro que la Informática sea únicamente Tecnología pues, aunque su medio de soporte actual sean las máquinas, hemos visto que puede hablarse de ella desde la antigüedad al referirse a *Información* y sus modos de expresión desde instrumentos rudimentarios como pinturas rupestres hasta los ordenadores actuales, de modo que no habría tal *aplicación* (sino transversalidad) si consideramos que el costado de procesamiento de datos le corresponde a la Informática. Por otra parte, la Estadística no podría catalogarse como *ciencia pura* pues siendo fundada como ciencia de descripción del Estado Nación en el siglo XVIII, tiene un elevado componente

sociopolítico. Ocurre que, al desarrollarse como matematización, debido a la influencia del positivismo en el s. XIX, se creó un largo vacío en el s. XVIII que se llenó con la actividad censal de los países, pero lo que sucedía epistemológicamente era una discusión sobre ella cuyo objeto era sacarle el efecto político para convertirla en ciencia pura y provocar un *efecto avestruz*, convertirla en ciencia de procesamiento de datos reduciendo el censo a mero instrumento de recolección.

De modo que en la actualidad tenemos una Informática pegada a las máquinas y una Estadística pegada a la Matemática comunicadas como soporte de la primera a la segunda o mecanización de procesos matemáticos, cuando en realidad fue un binomio indistinguible desde tiempos arcaicos. El costado epistemológico que se da a la relación es el mismo que tuvo la Estadística, obtener conocimiento del procesamiento de datos por medios inductivos, cosa innecesaria y confundente pues este es el objetivo de todas las disciplinas científicas, todas procesan sus datos para obtener conocimiento. Lo que vemos es que así es como la Estadística originalmente se creó para describir al Estado y luego se asumió como ciencia social, la nueva Ciencia de Datos amplía el ámbito de acción al describir y descubrir nuevas relaciones en el mundo económico y empresarial que se mueve en Internet y las redes sociales. Así la expresión de Gil (2013) tendría sentido, la Ciencia de Datos sería “un plan para ampliar las principales áreas de trabajo técnico del campo de la estadística”, corregiríamos que no solo técnico, sino que ahora tiene un mayor dominio epistémico.

5. El paralelismo en las definiciones de Estadística y Ciencia de Datos

Si usamos las definiciones de Estadística proporcionadas a Google por Oxford Languages hallaremos dos acepciones, la primera se refiere a una “Ciencia que utiliza conjuntos

de datos numéricos para obtener, a partir de ellos, inferencias basadas en el cálculo de probabilidades” y la segunda como el “Estudio que reúne, clasifica y recuenta todos los hechos que tienen una determinada característica en común, para llegar a conclusiones a partir de los datos numéricos extraídos”, son definiciones bastante amplias pero que aportan dos aspectos cruciales en la historia de E, el costado lógico-probabilístico y el descriptivo. Ahora bien, si buscamos la de Ciencia de Datos (CD) vemos que “is an interdisciplinary field about processes and systems to extract knowledge or insights from large volumes of data in various forms either structured or unstructured.” (Liu, 2015) es decir, la define como un campo interdisciplinario sobre distintos procesos y sistemas que sirve para extraer conocimiento de un gran volumen de datos. Este costado interdisciplinario lo especifica Vázquez (2020) quien señala que para aplicar CD se requiere desarrollo en cuatro áreas, Programación, Estadística, Comunicación y Conocimiento de Dominio, posiblemente por esto diga que la Estadística sea inescapable para aplicarla.

Igualmente habría paralelismo metódico, Merayo (Data Science), expone que en CD habría unos cuatro o cinco pasos (similares con los de Estadística), captura de datos, gestión y procesamiento, exploración, análisis y comunicación. Por otra parte, Colopy (2021) señala que CD usa tres procesos de inferencia, el razonamiento deductivo, el inductivo y el abductivo, asociando este último con la inferencia sobre la mejor explicación (verosimilitud) y a la probabilidad. Creemos que el autor se anticipa en su juicio pues estos procesos de razonamiento formaron parte de la discusión lógico-metodológica que se dio en el s. XIX a propósito de la formación de la metodología y la Estadística con el método inductivo, estaba en ese proceso a propósito de la lógica inferencial, lo que creemos es que el autor busca soporte metodológico para CD, pero la salida a la creación de una dualidad irresoluble entre

Estadística y CD es colocarlas en un solo proceso en el cual la Informática juega su papel. Igualmente pasa con Leo Breiman (citado por Gil (2013)) según el cual existirían dos culturas en el uso de modelos estadísticos, la primera supone el uso de modelos estocásticos en Estadística y la segunda modelos algorítmicos en CD, en nuestra perspectiva, resulta inútil como demarcación entre ambas pues vemos que hay elementos histórico-metodológicos más gruesos.

El desarrollo de la Informática es realmente reciente, pero explosivo, incorporando lo metodológico en lo tecnológico produciéndose una mezcla que hoy es difícil de separar, la Brecha Tecnológica no es sino la imposibilidad de acceder a lo metodológico oculto en lo tecnológico, esto es, solo cuando la Informática madura podrá reclamar espacios metodológicos comenzando con registros de bases datos. La Estadística ya reclamaba automatización temprana, en tiempos de Pearson (1857-1936) y Spearman (1863-1945) los nuevos procesos estadísticos multivariantes necesitaban automatización y recursividad pues eran operaciones inductivas de lograr el mejor ajuste de datos a los modelos necesitando una algorítmica, cuando esto se produce en el s. XX, se echan los fundamentos para la unificación de la Estadística y la Informática a través del dato, pero más allá, de la renovación de la Estadística en procesos algorítmicos.

6. Estadística en Ciencia de Datos ¿A dónde queremos llegar?

En términos de Ciencia Observacional, la que ocurre a escala humana (construcciones de la realidad con sentidos naturales y plan de observación) la estadística implica, por ejemplo, métodos de clasificación, que van desde la separación de objetos de modo espontáneo (que haría un niño, experimento de Piaget) para ordenar su mundo, subiendo a las clasifica-

ciones de juegos sociales, las de estratificación socioeconómica, la construcción de taxones biológicos según algunas características, hasta las clasificaciones estadísticas por métodos de taxonomía numérica, las de los métodos multivariantes o clúster. Las clasificaciones en Ciencia Experimental implican las que se hacen por instrumentación o pruebas científicas como las clasificaciones según las ratios de colesterol, tipos de sangre, enfermedades, entre otras muchas; obsérvese que en estos casos hay un mayor grado de transferencia de racionalidad al instrumento (diríamos mayor artificialidad) o correlativamente disminución de subjetividad por objetividad pues son los instrumentos los que hacen la medición. Finalmente, mayor incremento de racionalidad a máquinas para reconocimiento de patrones o *Deep Learning* en, por ejemplo, robots que separen humanos de no humanos, caso de los vehículos automatizados que pueden procesar obstáculos sin ayuda humana, aquí estaríamos en presencia de Ciencia Sintética.

Otro ejemplo más complejo es ir de la probabilidad subjetiva, sentida-entendida por la experiencia, pasando por la que se calcula con las frecuencias de ocurrencia a la que se realiza en Inteligencia Artificial y los computadores cuánticos “Una nueva herramienta desarrollada en el MIT democratiza el cuarto paradigma de la ciencia, la ciencia de datos: demuestra que los programas probabilísticos se pueden sintetizar automáticamente, en lugar de ser escritos por personas” (T21, 2019). Así, desde una episteme tecnológica se evitarían las inútiles metáforas como la que usa Toca Rey (2015) quien, no hallando la forma de pasar de la Estadística Clásica a la Big Data, dice que

4. En algunos casos, estas dos últimas pruebas robustas no permiten discriminar entre un SAR y un SEM. Son posibles varias posibilidades. El primero consiste en estimar un modelo que contiene ambos términos espaciales (SARAR)

esta última tiene como abuela a la primera, curiosamente no dice quién sería la madre.

Ahora haremos un recorrido de la Estadística como ciencia observacional, experimental y sintética en un aire histórico epistémico y en línea de nuestra idea de llegar a un estadio tecnológico, para lo cual, vamos a delimitar conceptualmente estos tipos de ciencia; básicamente tienen que ver, primero, con una creciente intervención del observador sobre los eventos o fenómenos como control de variables o dominio sobre el fenómeno y segundo, con una mayor transferencia del lenguaje científico a los dispositivos (por ejemplo caso de manipulación genética o de robótica), en el caso de la Estadística será como apoyo en unos casos y como control experimental en otros. Pueden observarse estadios intermedios entre ciencia observacional y experimental como el cuasi-experimental, pero no nos detendremos en esto, finalmente contextualizaremos estas ciencias en un ámbito histórico de desarrollo del Estado-Nación. Esto es necesario porque se trata de un enfoque desde el dato en tanto que elemento de construcción del mundo pasa sí y no desde un discurso general de marcos teóricos.

7. Estadística en Ciencia Observacional. El dato interpretado

La creación de la Estadística coincide con la fundación de los estados modernos, paz de Wesfalia (1646 y 1648), realidad que será explosiva en todos los órdenes, pero la trascendencia para el tema es que implica una cerradura sistémica que propicia el crecimiento interno, búsqueda de relaciones entre los estados y un crecimiento territorial complejo por los límites propios impuesto por geopolítica. En la Modernidad todo será reinterpretado. Foucault (2006, pág. 28 sig) tratando el problema del espacio y las ciudades, expresa que luego de su encierro en murallas, las ciudades deben abrirse por problemas de circulación. El crecimen-

to del comercio, de la demografía, entre otros, hará que el confinamiento en murallas (Edad Media) sea inútil, pero igualmente la comunicación entre ciudades europeas e internamente deberán crearse costumbres que guíen el comportamiento de todos (nacimiento de la policía), la evitación de peligros, el control de enfermedades, entre otros aspectos, marca el inicio de muchas actividades nuevas, pero al mismo tiempo la necesidad de información para saber y decidir.

Por lo tanto, con las ciudades, los estados y la nueva sociedad surgen nuevas ciencias, las político sociales y económicas, el término *Estadística* se le atribuye a G. Achenwall (1719-1772) quien perteneciente al historicismo alemán, viene de un movimiento que ya tenía importancia como descripción de estos estados recién creados, la cerradura territorial era exigencia del desarrollo social y político pues nace la sociedad industrial, el capitalismo y una población que se especializa y estratifica mucho más. Originalmente esta disciplina nace como una ciencia política del Estado-Nación de carácter inductivo que tuvo por objeto cuantificar la riqueza del Estado tanto individual como de las empresas, por lo tanto, era de perfil económico, para Loyo (1971, pág. 158) Achenwall en 1749 declararía que “La Estadística es la ciencia del Estado que se ocupa de determinar la riqueza individual” y dado que la concepción actual se refiere más a modelos matemáticos y conteos de frecuencias para la ciencia, evitando la concepción política original, debemos expresar que los propios fundadores introducen estos elementos metodológicos, según el mismo autor, Achenwall va a expresar en 1752 que:

Puesto que la palabra ciencia del Estado implica política, dejo a los filósofos su ciencia, en cuanto a sus bases generales se refiere, y a la estadística que se ocupa exclusivamente de experiencias prácticas, le dejaré solamente el nombre

de método del Estado o simplemente método de administración u organización estatal de la riqueza” (Loyo, 1971, pág. 159).⁷

Ciencia inductiva significaba histórico-inductiva, de ahí que su discípulo Schlözer (1735-1809) declararía que la Historia es estadística en movimiento y la Estadística historia puntual, frase ya famosa que muestra el programa historicista que tenía esa tendencia (de hecho, hoy existen métodos longitudinales y transversales). Este mismo ilustrado creó poco tiempo después el término *etnografía* para referirse a “una ciencia que trataba sobre pueblo y naciones” (Buccellato, 2021, pág. 32) y aquí nos detendremos brevemente.

Ambos ilustrados de la Universidad de Göttingen y el segundo discípulo del primero (de Achenwall) estarían mostrando lo que sería un programa de investigación (al estilo de Lakatos) respecto de la creación del Estado-Nación, cosa que sucedería en otros países como Francia; la etnografía alemana tendría sentido en tanto que descripción de la gran diversidad de territorios a ser anexados en el proceso de unificación de Alemania. En segundo lugar, la ciencia en general era tanto inductiva como deductiva, este último costado partía de principios generales como los de Adam Smith (1723-1790) que creía en la modelación, pero solo la inductiva ve la realidad cuantificando la riqueza y potencialidades de los países. Igualmente sucedería con la demografía y la epidemiología que quedarían bajo Estadística por razones de método y ciencia de estado-nación. Hoy la dualidad metódica se diluye como otras tantas

dualidades creadas por el positivismo y su lógica. Ambas ciencias Estadística y Etnografía siguen vigentes en torno a la descripción de los estados.

Así los gobiernos comenzarán a fundar oficinas de estadística y desde la sociedad civil muchos investigadores comenzarán a registrar y narrar todo como si se redescubriese la vida social, y la ciencia junto con los gobiernos se darán a la tarea de registrar los casos de las enfermedades, los nacimientos, las defunciones, los matrimonios (nacería el registro civil), la planificación de las edificaciones, de las aguas servidas, de las blancas (nacería la planificación), los censos adquirirán otras perspectivas, no solo las de tributación de las colonias. Moreau de Jonnes (1857, pág. 21) es explícito en este tema, en su libro *la Estadística en los s. XVII y XVIII* se dividía en trece partes, Territorio, Población, Agricultura, Industria, Comercio interior, exterior, Navegación, Colonias, Admón. Pública, Económica y de Justicia, Fuerzas Militares e Instrucción Pública; en el siglo XX, bajo la creciente influencia del positivismo lo hará en Estadística Descriptiva e Inferencial y, curiosamente, desde ahora será en Estadística Metodológica e Ingeniería de Datos o Estadística y Ciencia de Datos, también llama la atención que, el libro de Moreau de Jonnes, siendo académico, no muestre en aquella Francia cálculos de ningún tipo, que sea meramente descriptivo de las divisiones territoriales, la cantidad de población, producción, demografía, la técnica censal, entre otros temas que hoy formarían parte de un informe de censo de los países, igualmente llama la atención que, en otro libro posterior de 1873, escrito en España por Mariano Carreras y José Manuel Piernas (1873), no aparezca el tema de probabilidades, solo algunas menciones, mostrando el carácter etnográfico y descriptivo de la Estadística.

Dos operaciones destacan entre todas, los censos y los catastros. Los primeros, aun

7. La discusión entre Estadística Alemana y Aritmética Política inglesa según la cual la primera sería cualitativa y la segunda cuantitativa iniciándose este dualismo metodológico y además trasladando el origen de la Estadística a la isla como precursora de la Contabilidad Nacional es tendenciosa, ambas eran cuantitativas, como lo fue la Estadística fundada en Francia. El enfoque cualitativo nació con la Etnografía Alemana, de modo que la dualidad era innecesaria

cuando vienen desde la antigüedad, en la modernidad tienen que ver, además de los países de Europa, con las colonias bajo objetivos de construcción de territorio. Desde el siglo XVI “Visionarios, directores de contabilidad y generales proyectaron censos en muchos tiempos y lugares” (Hacking, 1995, pág. 39), existe documentación de la “monumental actividad estadística de la Corona española desde los primeros momentos” INEGI (2014, pág. 33) desarrolladas desde 1540 pero que vienen del Antiguo Régimen en Europa, los registros parroquiales y las estadísticas vitales se le dejaron a la iglesia, de modo que esta actividad no era centralizada ni organizada mundialmente. Los censos continuos se hicieron en Canadá en la década de 1660. En Francia, Colbert (ministro de finanzas francés) ordenó la realización de censos en todo el país, las islas azucareras del Caribe informaban sobre su producción y en New York se realizó un censo en 1698. “habría que hacer historia de cada fenómeno estadístico nacional y colonial; cada uno tiene características propias” (Hacking, 1995, págs. 39-40). La actividad censal global se haría a partir de las teorías de A. Quetelet (1796-1874) quien introduce en los censos mediciones humanas realizándose el primer congreso internacional de Estadística en 1853 donde comenzaría la estandarización de algunas de las características antropométricas en una mirada más experimental, pero igualmente estadística y etnográfica.

La modernidad implicó una explosión de datos, pero que no solo protagonizada por los gobiernos sino por los Ilustrados (científicos, aristócratas y capitalistas), movimiento en ciencia que reorganiza la sociedad, la va a hacer más sistémica. A modo de ilustración, Goethe (1891, págs. 23-24) describe en sus viajes por Italia que a la Feria de Botzen iban muchos mercaderes a reembolsar dinero, recibir pedidos y abrir nuevos créditos y que teniendo él deseos de observar de cerca estos intercambios expresó que “Además, confío que en nuestros

tiempos de estadística todo se encuentra impreso y puede aprenderse en los libros” [cursivas mías], en otro episodio paradigmático Reyes (1993, pág. 236) expresa que Goethe disertando sobre A. Humboldt decía que este trazaba los itinerarios como el propio Cook y que “Al fin salió a medir con sus pasos los datos de las estadísticas, a recorrer la tierra siguiendo el camino de los guarismos”. Este ilustrado explorador visitó cinco países de América latina haciendo sus famosas descripciones por las regiones equinociales del nuevo continente con trabajos estadísticos, etnográficos y naturalistas. Época igualmente de explosión imaginaria, de cuentos de grandes viajes al centro de la tierra o a países de gigantes y enanos y en donde se crea al “otro” en ciencia evolutiva (observacional) como las extrañas descripciones que se hicieron sobre los habitantes de América, Soriano (2011) expone el tema cómo imaginaron cinocéfalos, anencéfalos, entre otras miradas evolutivas en este continente. Por su parte Morales Moya (1984) describe los grandes viajes de estos ilustrados por España. Tales viajes tenían como finalidad,

Mostrar lo que había sido y lo que era España, tal fue el propósito de los esforzados eruditos y fervorosos peregrinos que recorrían toda España, explorando archivos, reuniendo documentos, medallas, inscripciones, midiendo monumentos, caminos y puentes, observando el campo y sus plantíos, los pueblos y sus habitantes, apuntando todo lo que veían para luego darlo a conocer (Morales Moya, 1984, pág. 32) citando a (Helman, 1953)

La cita es elocuente respecto de todo el proceso que Goethe llamó “nuestros tiempos de Estadística”. Con el nombre de *Grand Tour* se conocieron estos periplos de acaudalados. El mismo autor clasifica los viajes en, 1. Económicos, pues deseaban conocer las formas de comercio de otros lugares; 2. Naturalistas,

pues recogían muestras botánicas y de animales; 3. Artísticos, para conocer las artes de otros pueblos; 4. Históricos-arqueológicos, por similares razones respecto de la historia local; 5. Literario-sociológicos y 6. Rigurosamente políticos (Morales Moya, 1984); hay que destacar que estos viajes formaban parte igualmente de la formación que debían tener los nobles y burócratas para el desempeño de sus actividades. En términos de Bacca (1984) estas descripciones se corresponderían con ciencia Natural-Observacional, hecha a escala humana, pues eran clasificaciones de todo, el mundo se catalogaba. Esto es importante porque la probabilidad era aún cosa de eruditos matemáticos, además que defectuosa aún, el universo laplaciano primaba “Se decía que el mundo podría parecer a menudo fortuito, pero solo porque no conocíamos el inevitable operar de sus resortes internos” (Hacking, 1995, pág. 18).

Las discusiones sobre las formas de medición censal involucraban el problema de la generalización inductiva (esta forma de inferencia estaba muy poco desarrollada) ya que, debido a la vastedad territorial, la cantidad de elementos por escrutar y los escasos medios de comunicar y centralizar resultados, hacía imposible una enumeración completa. Moreau De Jonnes (1857) quien es protagonista de estos eventos discute dos métodos usados en la operación de censos, por *inducción* entiende generalizar una característica con base en una proporción que se considera constante, relata que J. Necker (1732-1804)⁸ en 1784 emprende el censo de Francia y al no poder abarcar el territorio “dedujo el número de nacimientos, el número de habitantes, adoptando una proporción de 1 a 25,75” (pág. 48), admite a regañadientes este método pues es partidario del método expositivo, que no es otro que un censo por enumeración completa. La necesidad

de los censos estaba acompañada del catastro. Relata Moreau De Jonnes (1857, pág. 319) que para 1817 el cálculo impositivo de cuotas catastrales se dificultó mucho por problemas de acceso y registro de la totalidad de las propiedades “Faltó poco para que lo suprimiera la Restauración”, relata que Hennet, quien dirigía los trabajos expresaría en su Memoria que no se pudo catastrar al setenta y cuatro por cien del territorio, acudiendo a las no deseadas *técnicas de inducción* para determinar que habrían 10 millones de cuotas en el país, pese a que, según Da Costa Carballo (1998, pág. 240), desde 1784 el gobierno francés había encargado a algunos matemáticos, entre ellos Legendre (1752-1833) para hacer las *Tablas de Catastros*, dado que tenían las de logaritmos y trigonometría. Estos antecedentes tienen otros anteriores en Prusia, Leibniz (1646-1716) que promovía la creación de oficinas de estadística calcula en 1685 la población de este país usando un multiplicador de 30 (Hacking, 1995, pág. 41).

Estas formas de proporcionalidad tenían de fondo la búsqueda de regularidades, los cálculos proporcionales de suicidios de Emile Durkheim (1858-1917), los de T. Malthus (1766-1834) sobre la producción y consumo de alimentos se expresaban como leyes inductivas, hoy se diría *regularidades nómicas* en términos de Moulines (1999, pág. 128) y que fallaban muchas veces a la luz de los acontecimientos. Dufau (1795-1877) en 1840 escribe su *Tratado de Estadística o Teoría del estudio de las leyes, según las cuales se desarrollan los hechos sociales* anunciando lo que sería una estadística metodológica, viniendo de ser meramente descriptiva, Relata Loyo (1971, pág. 407) que G. Rümelin (1815-1889) distinguió la Estadística que estudia las características de la sociedad humana y las del Estado, señala, además, que Gini (1884-1965) dice que este inicia un período ecléctico de la Estadística dividiéndola en metodológica, social y descriptiva, la primera se refiere a un método general de abordar la

8. Señala que fue banquero protestante varias veces ministro de Luis XVI.

realidad, la segunda, trata de hallar las leyes que rigen el comportamiento de la sociedad y la tercera la que estudia a la nación.

El procedimiento de obtener resúmenes de datos en proporciones y leyes inductivas parecía justificado a la luz de los medios para el procesamiento de datos. La *Estadística*, que implicaba la construcción del territorio tenía una pobre *Informática*, teniendo razón Da Costa Carvalho (1998) de clasificar este largo período como *Prehistoria* de la Informática. En otras palabras, la infraestructura sobre la cual edificar al país estaba en fase temprana o los elementos que configurarían el *sistema país* eran aún muy pobres para montar sobre este uno de comunicación, control y registro de datos.

Los s. XVII y XVIII son del mecanicismo y de la dinámica (Industry 1.0) en donde la filosofía cartesiana y el empirismo baconiano se imponían, la idea de matematizar la Naturaleza ronda en las mentes “Naturaleza a la que se imaginaba como un gigantesco autómatas” (Da Costa Carballo, 1998, pág. 239); sin embargo, estos intentos eran exiguos y contradictorios, por 1614 Neper (1550-1617) hizo unos intentos de *mecanizar* logaritmos en tablas para ayudar a los astrónomos o la máquina de Wilhelm Schikard en 1623 de 12 ruedas usada para hacer multiplicaciones. B. Pascal (1601-1662), que trabajaba en teoría de juegos y probabilidad, diseñó en 1642 la primera máquina moderna de sumar, la pascalina (en su nombre) para ayudar a su padre comerciante y luego, treinta años más tarde, Leibniz (1646-1716) mejoraría introduciendo multiplicación y división; son máquinas de cálculo, no de registro de datos. Estos desarrollos mecánicos no se concretaron comercialmente pues en esa época “no se consideraba que estas máquinas fueran necesarias” (Ceruzzi, 2008, pág. 112), de manera que el modo de *mecanizar* procesos matemáticos fue mediante tablas. Aun en el XIX “las ciencias que necesitaban realizar cálculos, como

la astronomía, se las arreglaban con las tablas impresas y las *calculadoras* humanas (así se llamaban quienes realizaban esta tarea) que trabajaban con papel y lápiz, libros de tablas matemáticas y, quizás, alguna máquina de sumar” [*cursivas del autor*] (Ceruzzi, 2008, pág. 113) práctica que se extendió hasta el s. XX. Podemos rememorar el uso de tablas en nuestra secundaria para ayudarnos con los cálculos antes de la llegada de la *regla de cálculo*.

Como vemos los procesos de organización de las unidades de análisis, la recolección de datos y el procesamiento informático eran manuales y muy lentos, esto es, en términos actuales de Estadística e Informática, la primera, como herramienta sociopolítica y ciencia de Estado, intentaba lograr cobertura horizontal clasificando los territorios y controlando la actividad social, la segunda, implicaba procesos manuales de cómputo que consumían papel, en Estadística se intentan los métodos de procesamiento resumidos o *Estadística para Datos Agrupados*, que duraron hasta entrado el s. XX, no pasando de la varianza y perdiendo toda la riqueza de los datos al agruparlos.

8. Estadística en Ciencia Experimental. Diseño de Constructos que intervengan el mundo.

Las técnicas observacionales se diferencian de las experimentales en que existe intervención de la mano o incremento de intencionalidad, los planes de observación y experimentación cambian, por lo tanto, lo hacen las técnicas usadas, digamos, para no profundizar en este arduo tema, que los instrumentos van adquiriendo mayor sofisticación debido a que lo observado, igualmente va pasando de observable a inobservable, la diferencia que hay entre aparejos, microscopio óptico, microscopio electrónico, escáner, telescopios Hubble y James Web, son formas de manipulación que hacen muy difícil a veces separar estrictamente y en ciertos niveles observación y experimen-

tación. Ahora bien, si en la observación hay *frecuencia* de apariciones de los fenómenos, en la experimentación tiene que haber replicación de los resultados y en esto cambian las técnicas estadísticas, aunque la *replicación* es una forma de *frecuencia*. De modo que en muchos casos deberá haber acuerdo en la ciencia. Así, ambas han existido desde siempre, pues el dominio del fuego en la antigüedad puede entenderse como experimentación. Ahora bien, estas formas de observación y experimentación tienen un evidente carácter histórico, aunque la superposición de ellas evite marcar períodos de rupturas como para hablar de *revolución* notándose esta más en lo copioso del desarrollo de tales formas que en la propia ruptura. Esto mismo ha pasado con la Estadística.

“A principios del siglo XIX, cuando Quetelet inicia su trabajo en estadística, básicamente el cálculo de probabilidades y la administración estadística (el recuento de la población a través de los censos) estaban separadas” (Sánchez Carrión, 1999, pág. 295), por su parte Ríos (1994) va a decir que entre 1800 y 1950 se produjeron unos eventos desde la Estadística que marcaron a la propia ciencia calificándolos como *Revolución Probabilística*, igualmente Hacking (1995, pág. 22) dice que la probabilidad, fue el triunfo de la primera mitad del s. XX, cuádruple triunfo metafísico, epistemológico, lógico y ético. Veamos esta consecuencia.

W. Heisenberg (1901-1976) es famoso por su sentencia de eliminar todos los inobservables de las teorías (como las elipses de los átomos, el núcleo, etc.) por la experimentación⁹ (las leyes del átomo como metáfora del sistema solar, eran vacías de sensación, en términos kantianos) haciendo a la física probabilística; por otra parte las corrientes evolucionistas con Charles Darwin (1809-1882) dan un duro gol-

pe a la naturaleza humana al introducir la aleatoriedad en la deriva genética (ver entre otros Monleón-Getino (2010)) de modo que esto abre varias vías de desarrollo para la Estadística al incorporar la contingencia como variabilidad, F. Galton (1822-1911) y K. Pearson (1857-1936) se dejan influir por estas nuevas tendencias comenzando estudios sobre Herencia.¹⁰ Este último declararí en 1889 que la probabilidad “reina con serenidad y completamente inadvertida en medio de la más profunda confusión” (Hacking, 1995, pág. 18) y en 1896 el estudio sobre la correlación es publicado (Arribas Macho, 2004, pág. 251) por Pearson, quien años más tarde en Oxford en 1907 señala que “En física, la base última del conocimiento es estadística, una correlación, no una relación causal” [el autor cita estas palabras de Pearson] (Arribas Macho, 2004, pág. 252). Es decir, la experimentación que había sido cosa prevalente de los laboratorios físicos y químicos y guiados por principios de la mecánica, ahora se enriquecía con la probabilidad posibilitándose fuera del laboratorio. Aun cuando se les dice a estas técnicas *pseudoexperimentales*, pueden verse como complejización de la observación. La importancia de esta estriba en que es necesaria para tomar cursos de acción con base en datos empíricos sin esperar a que estos se presenten. Por su parte, los desarrollos de R. Fisher (1890-1962) y otros, profundizan estas técnicas en estudios sobre genética introduciendo nuevas formas como el Análisis de Varianza, la Regresión y la Correlación.

Pero lidiar con la contingencia no es cosa sencilla pues el azar no es objetivable por lo que debe existir causa probable y asegurar estos caminos de acción. La famosa disputa Fischer vs. Neyman-Pearson ofrece una alternativa, el primero anticipa un modo *falsacionista* (distinto al de Popper, según algunos) al

9. De hecho, el origen de la experimentación como método se atribuye a Bacon, pero acá nos referimos a la introducción de la probabilidad.

10. Pearson fue discípulo de Maxwell (1831-1879) y Ernst Mach (1838-1916) y Galton es primo de Darwin fundando la Eugenesia

plantear que la hipótesis nula, H_0 , no puede probarse solo refutarse, ver por ejemplo a (Silva Aycaguer, Benavides Rodríguez, & Almenara Barrios, 2002, pág. 121) y llama a esto *prueba de significación*, sus oponentes piensan que si se refuta una hipótesis se deben estar aceptando otras y llaman a este procedimiento *contraste de hipótesis o comportamiento inductivo* (aclarando que es totalmente deductivo, más bien sería hipotético-inductivo ya que hipotético implica, de hecho, probar *deducción*). Este último proceso se toma como *regla de decisión* por lo que cogen caminos distintos, si se falsa H_0 se debe tener una alternativa, sobre todo en ciencia social y económica. Sin embargo, el enfoque anterior, basado en frecuencias de ocurrencia, da la probabilidad de que una hipótesis se realice, ocurra dadas las premisas, pero no dice nada de la probabilidad como explicación de un evento previo, la probabilidad inversa, muy útil en sistemas y juegos, de modo que el viejo *enfoque bayesiano* toma relevancia; Thomas Bayes (1702-1761) había construido su teoría inferencial considerando la probabilidad inversa o probabilidad de las causas, los frecuentistas la directa, la diferencia es sustantiva, su importancia reside en que no es prospectiva sino causal, por lo tanto, de amplio uso en teoría de sistemas y juegos. Así, la Estadística puede lidiar con la causalidad y con la contingencia usando probabilidad y variabilidad. De modo que lo contingente es aceptado como inevitable, pero lo será en tanto y en cuanto el azar sea domado y aquí hay algo controversial ¿qué significa domar al azar? No es solo explicarlo a través de modelos de distribución.¹¹ En primer lugar, de él se sabe solo que viene y sorprende, la explicación laplaciana del universo se ve obstruida por la indeterminación de Heisenberg, en el sentido según el cual no es posible conocer las coordenadas espacio-temporales de los fenómenos atómicos en tiempo real, pero introducirá el relativismo y fortalecerá el pensamiento experimental.

11. Que es hasta donde llegan los historiadores de la Estadística

Y vamos a lo trascendental, anteriormente señalamos que al instaurarse el Estado-Nación el espacio territorial se hizo sistémico, controlable, y el único modo de hacerlo es construir sistemas con baja incertidumbre. Veámoslo en palabras de Hacking (1995, pág. 17) “El concepto cardinal de la psicología de la ilustración había sido sencillamente la naturaleza humana. Al terminar el siglo XIX ese concepto fue remplazado por algo diferente: la persona normal” ¿qué significa esto? El sentido de Foucault de Biopolítica implica “racionalizar los problemas propios de los seres humanos como la salud, higiene, sexualidad, natalidad, longevidad, razas... por parte de la práctica gubernamental” (Aguilera Portales, 2010, pág. 33), la famosa curva normal de probabilidad concentra alrededor de la media a más del 70% de los casos, distribuyendo *equitativamente* los extremos, de modo tal que *normalizar* implicará reducir la variabilidad, lo extremo, disciplinar al cuerpo y va a ser A. Quetelet (1796-1874) quien introducirá las primeras mediciones censales sobre este, “Así, aunque pueda afirmarse que determinados desvíos de la media indican una patología, no es simple comprender la estructura explicativa que define para una infinita variedad de indicadores el momento exacto en que se inicia el proceso patológico” (Caponi, 2013, pág. 832), de modo que se evalúa una masa de indicadores biológicos que permiten modelar la sociedad y esta curva de Gauss (1777-1855) será una guía, ya Quetelet creía en un Hombre Medio, que no existiendo en términos materiales podrá tender a él en sus características vitales. El control experimental será útil para estandarizar todo y en adelante miraremos intervalos de normalidad. Para que la Biopolítica de Foucault fuese posible el pensamiento probabilístico y la explicación de la variación tendría que intervenir.

Sin embargo, este paradigma de la variabilidad fue igualmente controvertido en el siglo XIX, como búsqueda de patrones es interesante, pero se olvida del individuo, Porter (1986,

pág. 158 ss.) expone que Risueño d' Amador en 1836 objeta el uso de la probabilidad en medicina como anticientífico pues se opone precisamente a la variabilidad, al determinarse la prevalencia de una patología pueden descuidarse los casos particulares dada la similaridad sintomática, señala que “la ley de la mayoría no tiene autoridad sobre los casos rebeldes” y justamente en esta actualidad de pandemia muchos médicos tratan toda afección gripal, primero colectivamente como Covid-19 y solo al descartarla se ocupan de lo específico, estos procesos de *descarte* son experimentales en tanto que debe ocurrir una tormenta de datos (como síntomas) que definan la enfermedad (rellenen el concepto de sensación intervenida por instrumentos); por su parte Gustav Rümelin cambia la visión que se venía teniendo con la variabilidad, Galton y Pearson estaban tan impresionados por la estabilidad de los grupos y su propensión a la media o regresión que vieron en la Eugenesia la salida al desarrollo de la humanidad, esta visión le pareció a Rümelin contraria a la variabilidad, pues la anula, expresó que “la Estadística era más útil precisamente en relación a esos dominios donde la individualidad es más prevalente”, el ser humano es tremendamente típico en las ciencias de la naturaleza, pero la sociedad es el dominio de la diversidad. Aunque Rümelin con esto desmarca las ciencias sociales sin negar la importancia de la Estadística sino precisamente reafirmando, su posición tendrá consecuencias en Investigación Operativa y en Ciencia de Datos. De modo que la ciencia experimental tanto en laboratorios como en campo han reclamado hoy un amplio espectro de la ciencia en general, asistimos a un ciudadano que se automedica por conocer protocolos y sus costumbres se hacen cada vez más apegadas a ciencia experimental como la culinaria, manejando una probabilidad no solo subjetiva sino con conocimiento de regularidades de eventos personales y a unos sistemas de salud que regulan aspectos orgánicos de la vida. Igualmente sucede en el terreno empresarial.

9. Estadística en Ciencia Sintética. Del dato a la información y su transferencia a la máquina

La diferencia entre las ciencias experimental y sintética es, digamos en palabras de Bacca (1984) la transferencia aun mayor de racionalidad a los dispositivos, de modo que logren independencia o que resuelvan problemas sin la intervención de los humanos y esto se logró al transferir a la máquina un lenguaje siendo ejemplar los casos de la robótica y la inteligencia artificial.

Veamos este recorrido, el cálculo en ciencias observacionales y experimentales eran muy tediosos por manuales, solo para calcular la varianza había que crear varias columnas con múltiples operaciones simples o procesos interminables de diagonalización de matrices en álgebra lineal. El proceso manual era tremendamente improductivo, los censos de millones de personas y sus variables eran imposibles de procesar, de modo que la alternativa fue la agrupación de datos en el conteo. La graficación por histogramas suponía hacer fuertes supuestos como uniformidad de la distribución de los datos en el intervalo, ojivas, media agrupada, entre otros escasos elementos. Tratar los datos de modo agrupado supuso crear una estructura de la masa de datos en centro, partes y dispersión (medidas centrales, centiles y coeficientes de dispersión), estructura que desemboca en la modelización como gran resumen de datos pues hallar uno en la masa implicaba especulación matemática, de modo que este camino la llevó a críticas epistemológicas. Metodológicamente lo primero que se hacía con los datos era la formación de intervalos de clases para los histogramas y suponer normalidad aun cuando la Estadística ya tenía desarrollos importantes y un camino andado por el engoroso tema de la inducción. Por otra parte, las aplicaciones se ralentizaron pues la Informática (como procesamiento) estaba aún en fase de una precaria modernidad, esto es,

eran asincrónicas la técnica de procesamiento de datos con los desarrollos de la estadística matemática. La solución tendría que ver con el paso de lo analógico a lo digital.

En 1880 Hollerith (1860-1929) es contratado por la Oficina de Censos de EEUU para preparar el análisis estadístico del censo, proceso que duró 7 años y medio dando muchos errores, pero que mejoraría para el de 1887 cuando introdujera una máquina tabuladora basada en la de Charles Babage (1791-1871) usando tarjetas perforadas y papel continuo, de modo que para el censo de 1890 el procesamiento se redujo a 2 años y medio (Da Costa Carballo, 1998, pág. 246). El autor narra que este proceso continúa hasta convertirse en asunto de Estado, Hollerith interviene en los censos de Inglaterra en 1911 fundando una empresa que sería antecesora de IBM creada en 1924 en la Antigüedad de la Informática según el autor. La *Revolución Probabilística* hallaría en la *Informática* un aliado mucho más que técnico. Aunque en 1944 se inicia la etapa Contemporánea de la Informática (o diríamos para dar sincronía plenamente moderna) con la primera computadora digital (ENIAC) este procesamiento no fue masivo y habría que esperar por el software, el lenguaje. En términos de procesamiento masivo, en 1961 apareció el concepto de Hoja de Cálculo en el artículo *Budgeting and System Simulation* de Richard Mattessich, pero tendría que esperar un poco más para que en 1982 se desarrollara la hoja Multiplan, conocida como *Electronic Paper* desarrollada por Microsoft en 1982 para competir con Visicalc, luego saldría Excel, estos desarrollos tenían objetivos contables y financieros, las aplicaciones estadísticas salieron inicialmente para mainframe como el BMDP (Biomedical Computers Programs) o el SPSS (Statistical Package for the Social Science) y luego se desarrollaron para PC.

En este proceso uno de los pioneros es Tukey (1915-2000) quien, siendo estadístico y viniendo del área informática, habría interpre-

tado a Rümelin, los datos son la variabilidad, lo que se calcula como tal es lo típico. Esto es, si σ (sigma) de dispersión en la curva de Gauss agrupa alrededor del 70% de los datos, se habla más de lo típico que de lo que varía, la varianza fue pensada como una media de desvíos. Prensa (2013) cita de Tukey (en su libro *El futuro del Análisis de Datos* de 1962) que “El análisis de datos, y las partes de las estadísticas que se adhieren a él, deben... asumir las características de la ciencia en lugar de las de las matemáticas... el análisis de datos es intrínsecamente una ciencia empírica”. El Análisis Agrupado de Datos que hiciera gruesos supuestos como que *al agrupar los datos se pierde la individualidad del dato* se anulaba en el computador, los diagramas de tallos y hojas, caja y bigotes con la obtención de los valores atípicos, los extremos y los valores perdidos hacen una Estadística de casos de variabilidad (*outliers*) y los diagramas de control al centrarse en la tendencia de los casos muestra el efecto de la variabilidad no como resumen de datos sino caso por caso. Hay una gran diferencia entre un histograma y un diagrama de tallos y hojas, el primero supone uniformidad de los datos en el intervalo, el segundo los muestra a todos en su verdadera posición y permite trabajar con ellos de modo individual, esto abrió las puertas a la *Minería de Datos*. De modo que la realidad se redescubre en la nube, no siendo posible fuera. “La minería de datos es, en realidad, una prolongación de una práctica estadística de larga tradición, la del Análisis de Datos” (Aluja, 2001, pág. 492). La *Data Science* revoluciona la Estadística pues potencia las simulaciones digitales que quedaban en el papel en cálculos manuales en el computador recreando la realidad en la nube potenciando toda la modelación como los casos arriba mencionados, Darlington (2021) lo expresa respecto de la inferencia bayesiana “Lo que más me llamó la atención de aquella técnica fue el hecho de que una fórmula matemática pudiera imitar y mejorar en muchos casos procesos de toma de decisiones de expertos humanos”, agregamos, al transferir el método

al computador usando su lenguaje en un beneficio inmediato “La democratización de la ciencia de datos es la noción de que cualquier persona, con poca o ninguna experiencia, puede hacer ciencia de datos si se le proporcionan amplios datos y herramientas de análisis fáciles de usar” Matheson (2019) cambiando *dato* por *información* que implica un cambio ontológico pues si el primero es significado el segundo le da posibilidad de transferencia en machine learning.

Rao (1995) al preguntarse ¿qué es la estadística? ¿es ciencia, tecnología, lógica o arte? Se responde más adelante “La Estadística tal como es entendida actualmente es la lógica a través de la cual podemos subir un peldaño en la escalera que nos lleva de los *datos* a la *información*” [*cursivas suyas*] y bajo este concepto hace explosión en las Redes o en Internet. La realidad de la tecnología informática implica la creación de un complejo mundo de datos en las Redes, accesibles, pero luciendo *despegados* del mundo cotidiano pues se le llama *nube* de la que hay que *bajar* información o procesarla por métodos que solo en ella tienen sentido, no siendo el mundo de los sentidos de Frege o el de los conceptos de Platón o el tercer mundo de Popper, pues es real. Ese es el mundo que ahora se explora.

10. Discusión y Conclusiones

A manera de síntesis la Estadística ha pasado por los tres paradigmas anteriores etapas observacional, teórica y experimental estando ahora en el cuarto, la Ciencia de Datos, en una diacronía que las superpone sin anularlas mutuamente llegando a este no sin los atascos propios de los agotamientos paradigmáticos de las visiones de la realidad y su coajuste entre Ciencia y Técnica. Tales superposiciones se ven escaladas *grosso modo*, desde el mundo analógico al mundo digital en una transformación que va de *dato interpretado* (creación de significación) en ciencia observacional, *dato como constructo* (leyes, regularidades) en

ciencia experimental y *dato como lenguaje* (información) en ciencia sintética (tecnología). La Data Science no es sino la transferencia de racionalidad a los dispositivos, en el sentido de la filosofía de la técnica, Bacca (1984). Las preocupaciones propias de Kendall (1968) y recientemente de Efron (1988) por el futuro de la Estadística debido a los agotamientos de las aplicaciones y hallazgos técnicos, hoy se disipa.

En términos trascendentes diríamos que la Estadística corre paralela con el desarrollo de la relación Estado-Sociedad y Ciencia-Sociedad. En la Antigüedad, admitiendo ausencia de *sociedad* en términos de ciudadanía (Derechos Humanos, Estado democrático y libertades económicas), la Estadística era de dominio exclusivo del Estado principalmente para propósitos de tributación y reclutamiento en procesos de expansión, no extrañando que, con la emergencia del estado-nación, su dominio epistémico se expandiera del Estado a la Nación (Demografía, Epidemiología, Contabilidad Nacional, Administración) apoyándose mutuamente con la Etnografía y que fuese definida como ciencia política (s. XVII y XVIII). François Simiand (1873-1935) expresó que hasta los primeros decenios del siglo XIX la disciplina se refería al Estado, pero que en adelante se expandiría hacia otros dominios mediante mediciones métricas (Loyo, 1971, pág. 401), aparente ruptura que implica realmente que el giro científico de la política, operado en esos siglos, haga que el Estado pase a fondo contextual pues, en ciencia ahora se describe a la Administración y a la Nación (sociedad) transversalizando a toda la ciencia empírica en su costado cuantitativo. Así, esta disciplina ha tenido muchos apellidos, administrativa, matemática, metodológica; apartando la propias de las aplicaciones (a: economía, ingeniería, administración, mercadeo, sicología, sociología, además de las que ya traía). El paso del paradigma observacional a los teórico y experimental tiene que ver con el paso, en ciencia, de lo observable a lo inobservable (de hecho, o por principio), teorías socio-económi-

cas, leyes naturales y en general, regularidades epistémicas y conceptuales siempre en pugna entre lo digital de la Estadística Matemática y lo analógico de la ciencia, interpretado como cuantitativo vs. cualitativo. Testigo de esto son las verbalizaciones conceptuales que ha tenido a lo largo de su historia y que dan muestra de su transversalización, puede verse, por ejemplo, en Loyo (1971), un intento de síntesis de esto desde el s. XVII hasta 1934 como recopilación de 117 verbalizaciones conceptuales sobre Estadística que hizo Wilcox (1869-1949) y que continuó Nalimov (1981) (1910-1997) hasta 1971 con otras 158 aproximadamente, zigzagueando entre ciencia política, matemática, probabilidad y metodología. En este medio siglo habría que incluir las que ahora se producen como Ciencia de Datos.

Finalmente siguiendo la diacronía que muestra Jin Gray, respecto de la Estadística, asistimos a una transformación técnica sin precedentes en su recorrido, como ciencia observacional la Estadística, a partir del dato, ha coadyuvado con la interpretación del mundo, en ciencia experimental a acompañado la construcción y objetivación de las regularidades de la ciencia y ahora en ciencia sintética apoya la recreación del mundo en espacios digitales reales.

7. Referencias

- Aguilera Portales, R. E. (enero de 2010). "Biopolítica, Poder y Sujeto en Michel Foucault". *Universitas. Rev de Filosofía, Derecho y Política*(11), 27-42.
- Aluja, T. (2001). "La Minería de Datos, entre la Estadística y la Inteligencia Artificial". *Questiiló*, 25(3), 479-498.
- Aretxaga Burgos, R. (1998). La filosofía de la técnica de Juan David García Bacca. Tesis Doctoral, Universidad de Bilbao, Dto de Filosofía, Bilbao.
- Arribas Macho, J. M. (2004). "Karl Pearson. The scientific life in a statistical age". *Empiria. Revista de ciencias sociales*(8), 249-254.
- Buccellato, M. (2021). "Sombras de antepasados olvidados. Una reflexión sobre el origen de la etnografía y la etnología en la obra de Han Vermeulen". *Runa*, 27-42.
- Caponi, S. (jul-sep de 2013). "Quetelet, el hombre medio y el saber médico". *Historia, Ciencia, Saúde-Manguinhos*, 20(3), 831-847.
- Carreras y Gonzalez, M., & Piernas Hurtado, J. M. (1873). *Tratado Elemental de Estadística*. (M. Guijarro, Ed.) Madrid: Imprenta y Librería de Miguel Guijarro.
- Ceruzzi, P. E. (2008). "Historia de la Informática". En BBVA, *Fronteras del Conocimiento* (págs. 102-197). Madrid, España: Turner.
- Colopy, G. W. (April de 2021). "Welcome to the Philosophy og Data Science Section". *Journal of Data Science*, 19(2), 173-177. doi:10.6339/21-JDS192EDI
- Da Costa Carballo, C. M. (1998). "Los Origenes de la Informática". *Revista General de Información y Comunicación*, 8(1).
- Darlington, K. (15 de septiembre de 2021). "Inferencia Bayesiana, aprendizaje y desarrollo de sistemas de IA". Recuperado el 22 de diciembre de 2021, de OpenMind BBVA.Tecnología-Inteligencia Artificial: <https://www.bbvaopenmind.com/tecnologia/inteligencia-artificial/inferencia-bayesiana-aprendizaje-ia/>
- Díez, J. A., & Moulines, C. U. (1999). *Fundamentos de Filosofía de la Ciencia*. Barcelona, España: Ariel, S.A.
- Efron, D. B. (1988). Discurso de Investidura como Dr. Honoris Causa uc3m. Recuperado el Diciembre de 2020, de Universidad Carlos III uc3m: <https://www.uc3m.es/about-uc3m/honoris-causa/professor-bradley-efron>
- Foucault, M. (2006). *Seguridad, Territorio, Población. Curso en el college de France (1977-1978)* (Vol. 1ra reimpresión). Buenos Aires: Fondo de Cultura Económica.
- Fox, P., & Hendler, J. (2009). "Semantic eScience: Encoding Meaning in Next-Generation Digitally Enhanced Science". En T. Hey, S. Tansley, & K. Tolle, *The Fourth Paradigm* (págs. 147-152). EEUU: Microsoft Corporation.

- García Bacca, J. D. (1984). *Teoría y Metateoría de la Ciencia (Vol. II)*. (U. C. Venezuela, Ed.) Caracas, Venezuela: Ediciones de la Biblioteca.
- Gil. (28 de mayo de 2013). “Una historia muy corta se la ciencia de datos”. Recuperado el 19 de 4 de 2021, de Forbes: <https://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/?sh=365of69755cf>
- Goethe, J. W. (1891). *Viaje a Italia (Vol. I)*. (F. G. Garrido de Rodríguez Mourelo, Trad.) Madrid, España: Librería de la Viuda de Hernando y Ca.
- Gray, J. (2009). “Jim Gray on Science: A Transformed Scientific Method”. En T. Hey, S. Tansley, & K. Tolle, *Fourth Paradigm. Data-Intensive Scientific Discovery* (Version 1.1, 2da impresión ed.). Whashington: Microsoft Corporation.
- Hacking, I. (1995). *La Domesticación del Azar* (Primera reimpresión ed.). (A. L. Bixio, Trad.). Gedisa.
- Helman, E. F. (1953). “Viajes de españoles por la España del siglo XVIII”. *Nueva Revista de Filología Hispana*, VII, 618-619.
- Hey, T., Tansley, S., & Tolle, K. (2009). *Fourth Paradigm. Data-Intensive Scientific Discovery*. Whashington, EEUU: Microsoft Corporation.
- INEGI. Instituto Nacional de Estadística y Geografía. (2014). *Época Moderna. Siglos XVII y XVIII. Historia de la Estadística Moderna (Vol. 3)*. Mexico.
- Jimenez, J. (16 de octubre de 2020). “Científico de datos: así es y así se forma uno en esta profesión cada vez más demandada”. Recuperado el 2021 de junio de 2021, de Xataka: <https://www.xataka.com/otros/cientifico-datos-asi-profesion-demandada>
- Kendall, M. (1968). “Sobre el futuro de la estadística”. Una segunda visión. *Estadística Española*(41), 5-29.
- Liu, A. (2015). “Data Science and Data Scientist”. (C. IBM, Ed.) Recuperado el 6 de junio de 2021, de <http://www.researchmethods.org/DataScienceDataScientists.pdf>
- Loyo, G. (Febrero de 1971). “Evolución de la Definición de Estadística”. (D. G. Estadística, Ed.) *Revista de Estadística*, 157-166.
- Madrid Casado, C. M. (2019). “Filosofía de la técnica y de la tecnología”. *Revista Iberoamericana de Ciencia, Tecnología y Sociedad - CTS*, 14(40), 309-312.
- Martín Jimenez, L. C. (2018). *Filosofía de la Técnica y de la Tecnología*. Pentalfa.
- Matheson, R. (15 de enero de 2019). “Democratizando la ciencia de datos”. Recuperado el 22 de diciembre de 2021, de MIT News. On campus and around the world: <https://news.mit.edu/2019/nonprogrammers-data-science-0115>
- Merayo, P. (s.f.). “Data Science”. Recuperado el 6 de junio de 2021, de Máxima Formación: <https://www.maximiformacion.es/blog-dat/que-es-la-ciencia-de-datos/>
- Monleón-Getino, T. (2010). “Importancia de Darwin en el desarrollo de la estadística moderna”. *Estadística Española*, 52(175), 371-391.
- Morales Moya, A. (1984). “El viaje ilustrado”. *Estudios turísticos*(83), 31-43.
- Moreau De Jonnes, M. A. (1857). *Elementos de Estadística*. Madrid: Imprenta de Francisco Abienzo.
- Moscolini, N. (2011). *Las nubes de datos. Métodos para analizar la complejidad*. UNR Editora. Universidad Nacional de Rosario.
- Nalimov, V. V. (1981). *In the Labyrinths of language: A Mathematician's Journey*. Philadelphia, USA: ISI Press.
- Porter, T. M. (1986). “Número y Diversidad: la fruición del pensamiento estadístico”. *LLULL. Revista de la Sociedad Española de Historia de las Ciencias y las Técnicas*, 9, 153-161.
- Prensa, G. (28 de Mayo de 2013). “Una historia muy corta de la Ciencia de Datos”. Recuperado el 3 de Noviembre de 2021, de Forbes: <https://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/?sh=365of69755cf>
- Rao, C. R. (1995). “El Estadística de Dominio Público”. *Questiiló*, 19(1,2,3), 361-380.
- Reyes, A. (1993). *Obras Completas de Alfonso Reyes XXVI*. México: Fondo de Cultura Económica.

- Ríos García, S. (1994). "Historia de la Matemática". Recuperado el 18 de Enero de 2022, de ICMAT Instituto de Ciencias Matemáticas: https://dmle.icmat.es/pdf/HISTORIADELAMATEMATICA_1994_00_00_06.pdf
- Sánchez Carrión, J. J. (1999). "Presentación Quetelet y la Sociología". (U. Complutense, Ed.) *Reis. Revista Española de Investigaciones Sociológicas*, 87, 291-303.
- Schwab, K. (2016). *La Cuarta Revolución Industrial* (Formato Digital. Foro Economico Mundial Todos los Derechos Reservados ed.). Barcelona, España: Penguin Random House Grupo Editorial.
- Silva Aycaguer, L. C., Benavides Rodríguez, A., & Almendra Barrios, J. (2002). El Péndulo Bayesiano: crónica de una polémica estadística. *LLULL*, 25, 109-128.
- Soriano Nieto, N. (2011). El viaje y lo monstruoso en el siglo XVIII. Por una Ética-Estética del Grand Tour. *Nómadas. Revista de Ciencias Sociales y Jurídicas*(32), 255-288.
- T21. (21 de enero de 2019). Una nueva herramienta democratiza la ciencia de datos. Recuperado el 15 de Noviembre de 2021, de Tendencias: https://tendencias21.levante-emv.com/una-nueva-herramienta-democratiza-la-ciencia-de-datos_a44987.html
- Toca Rey, G. (enero-marzo de 2015). Estadística: la abuela del "big data". *Historia y Vida*(395).
- Vazquez Brust, A. (31 de mayo de 2020). Ciencia de Datos para Gente Sociable. Recuperado el 5 de junio de 2021, de Ciencia de Datos para Gente Sociable: https://bitsandbricks.github.io/ciencia_de_datos_gente_sociable/
- Wills, J. (3 de mayo de 2012). Josh Wills. Recuperado el 7 de junio de 2021, de Twitter.com: https://twitter.com/josh_wills