

La evolución de los métodos de análisis multivariante en el campo de las ciencias de la administración

EVOLUTION OF MULTIVARIATE ANALYSIS METHODS IN THE FIELD OF MANAGEMENT SCIENCES

Karla María Alarcón Sánchez^{1*}  
Universidad de Guadalajara

José Luis Soriano Sandoval¹  
Universidad de Guadalajara

* Autora corresponsal.

1 Centro Universitario de Ciencias Económico Administrativas, Universidad de Guadalajara, Periférico Norte 799 Núcleo Universitario, C. Prol. Belenes, 45100 Zapopan, Jalisco, MÉXICO

Resumen

Propósito: Exponer la evolución de los métodos de análisis multivariante utilizados en el campo de las ciencias de la administración y que, además, permite identificar el estado actual y las tendencias futuras que posibiliten un mayor desarrollo en el estudio de los negocios y la administración.

Metodología: Revisión sistemática de literatura referente al tema y se discuten los puntos centrales de los métodos predominantes.

Resultados: Se identifican tres principales métodos: Análisis de Regresión, Modelado de Ecuaciones Estructurales e Inteligencia Artificial (IA).

Implicaciones: El estudio explora las tendencias futuras dentro de la IA como el uso de los bosques aleatorios, las redes neuronales y la red elástica.

Originalidad: Se presenta la evolución de los métodos de análisis multivariante específicamente para el campo de las ciencias de la administración e incorpora las tendencias de los últimos 25 años

INFORMACIÓN ARTÍCULO

Recibido: 17 de Enero 2024
Aceptado: 1 de Abril 2024

Palabras Claves:

Ciencias de la Administración
Técnicas Multivariantes
Regresión Lineal
Modelado de Ecuaciones Estructurales
Inteligencia Artificial

Abstract

Purpose: The evolution of multivariate analysis methods in the field of management sciences is presented, identifying both the current state and future trends that enable further development in business and administration studies.

Methodology: A systematic literature review was conducted on the topic, and the key aspects of the predominant analysis methods are discussed.

Results: the predominant methods are Regression Analysis, Structural Equation Modeling and Artificial Intelligence (AI). The results validate previous studies and highlight future trends.

Implications: The study explores future trends within AI such as the use of random forests, neural networks and the elastic network.

Originality: The evolution of multivariate analysis methods is presented, specifically for the field of management sciences and incorporates the trends of the last 25 years.

ARTICLE INFO

Received: 17 January 2024

Accepted: 1 April 2024

Keywords:

Management Sciences
Multivariate Techniques
Linear Regression
Structural Equation Modeling
Artificial Intelligence

INTRODUCCIÓN

El desarrollo de la ciencia está condicionado por los avances tecnológicos que permiten o imposibilitan mejorar las investigaciones, experimentos, estudios empíricos, teóricos o simplemente para validar estudios previos con mayor precisión o reducción de sesgos. Las investigaciones científicas parten de una hipótesis y dicha hipótesis debe responderse mediante una metodología previamente establecida. Por lo tanto, la metodología es parte fundamental. En el campo de las ciencias de la administración, la metodología predominante en los últimos 50 años ha sido el método de regresión lineal, el cual permite identificar relaciones entre variables por medio de la visualización de correlaciones, sin embargo, la tendencia está por cambiar, producto de los avances tecnológicos y de las mejoras en las metodologías estadísticas relacionadas al análisis multivariante.

El objetivo de este artículo es discutir la evolución de las diversas técnicas y métodos de análisis multivariante en el campo de las ciencias de la administración, permitiendo ser una guía para quienes se encuentren interesados en incorporar técnicas de análisis multivariante en sus investigaciones, al clarificar las fortalezas y debilidades de las principales técnicas disponibles en la actualidad.

En tanto, las técnicas cuantitativas son un medio muy poderoso a través de las cuales es posible resolver la incertidumbre mediante la toma de decisiones y potencializar con la posibilidad de generar proyecciones. Por consiguiente, estas técnicas multivariantes son importantes para el

campo de la administración porque permiten evaluar los factores de planificación y, cuando surgen, brindan una explicación significativa de los fenómenos relacionados a la administración de los negocios. He aquí la relevancia de implementar estas técnicas en las organizaciones empresariales.

Métodos multivariantes

Conforme a lo mencionado por Verma y Sharma (2017), las técnicas cuantitativas incluyen métodos o herramientas que se centran en la medición objetiva y el análisis de números para sacar conclusiones sobre problemas dados. Es un método o técnica científica utilizada por el mundo empresarial y académico para la resolución de problemas y la toma de decisiones. Asimismo, el estudio de las técnicas cuantitativas ha sido un concepto relativamente nuevo, que tiene grandes aplicaciones en la academia y otros campos de la vida. Estos son más relevantes para problemas de situaciones complejas, siendo un complemento del juicio y la intuición.

El análisis cuantitativo ahora se extiende a varias áreas, y que se designa ampliamente como ciencias de la administración, análisis de sistemas, proceso de toma de decisiones o ciencia de decisiones y métodos estadísticos, entre otros. A su vez, la técnica cuantitativa es un enfoque científico para la toma de decisiones gerenciales y su propia investigación, por tanto, el uso exitoso de la técnica cuantitativa para el desarrollo de la ciencia permite que la gestión de proyectos y estudios resuelvan problemas complejos a tiempo, con mayor precisión y de la manera más económica. En la actualidad, las técnicas

de administración científica están disponibles para resolver problemas científicos y mejorar la gestión, por lo que el uso de estas, ayuda a los investigadores a ser explícitos sobre sus áreas problemáticas. Además, las técnicas cuantitativas facilitan el proceso de toma de decisiones, proporcionan herramientas para la investigación científica, ayudan a seleccionar una estrategia adecuada para la reducción de costos en la empresa, y a obtener un adecuado despliegue de recursos y minimizar el tiempo necesario para completar una tarea.

De esta forma, las técnicas cuantitativas proporcionan una base científica para afrontar la incertidumbre del futuro. Sin duda, estas incertidumbres no se pueden eliminar, pero las técnicas cuantitativas ayudan a minimizar los problemas del negocio y su estudio, porque refuerzan el pensamiento científico disciplinado sobre los problemas organizacionales, proporcionan una descripción precisa de la relación de causa y efecto y la eliminación del riesgo. En definitiva, reemplazan el enfoque subjetivo e intuitivo con el enfoque analítico y objetivo.

Dentro de las técnicas cuantitativas se encuentran las técnicas de análisis multivariantes, que se destacan por permitir el análisis de más de una variable. El análisis multivariado se utiliza para estudiar conjuntos de datos más complejos que los que pueden manejar los métodos de análisis univariante. En la realidad, interactúa "n" número de variables que al final, cuando se busca explicar dicha realidad, es importante considerar una gran cantidad de ellas, permitiendo ser el análisis multivariante una opción para mejorar los estudios en el campo de las ciencias sociales, así como en las ciencias de la administración.

A continuación, se exponen los conceptos más relevantes, seguido de una revisión de literatura presentada por Hair et al. (1979) en los libros relacionados con las técnicas de análisis multivariante. La evolución de los temas se contrasta con el número de publicaciones que se utiliza como metodología, para elaborar estudios en el área de las ciencias de la administración. Lo anterior, permite identificar tendencias futuras respecto al uso de las técnicas de análisis multivariante. Posteriormente, en el apartado de discusión se presentan las técnicas particulares junto a sus ventajas y debilidades. Finalizando con las conclusiones sobre el futuro de la técnica y a sus áreas de oportunidad.

El análisis de datos multivariante se ha convertido en uno de los textos metodológicos más utilizados en las ciencias sociales. La principal fuerza impulsora detrás de este esfuerzo fue Joseph Hair Jr., quien ha actuado como un visionario constante al expandir el texto a lo largo de los años para incorporar técnicas emergentes, por ejemplo, el modelado de ecuaciones estructurales. En tanto, el análisis de datos multivariante es un elemento básico entre los usuarios de estadísticas tanto comunidades académicas y profesionales.

De acuerdo con Dodge (2008), el análisis multivariado se utiliza para estudiar conjuntos de datos más complejos que los que pueden manejar los métodos de análisis univariante. Este tipo de análisis casi siempre se realiza con softwares: SPSS o SAS, pues trabajar con los conjuntos de datos más pequeños puede resultar abrumador sin el apoyo de estos. Igualmente, Cavanaugh (2007) menciona que el análisis multivariado puede reducir la probabilidad de errores de tipo I. De acuerdo con diferentes estudios, en ocasiones se prefiere el análisis univariado, porque las técnicas multivariadas pueden generar dificultades para interpretar los resultados de la prueba. Por ejemplo, las diferencias de grupo en una combinación lineal de variables dependientes en MANOVA pueden no estar claras. Además, el análisis multivariado no suele ser adecuado para pequeños conjuntos de datos.

METODOLOGÍA

Cuando se hace una revisión bibliográfica, el factor de impacto es la variable que se considera relevante al momento de elegir los textos, donde quizás la evidencia más directa del impacto sea a través del análisis de citas, que durante mucho tiempo se ha utilizado como un enfoque objetivo para evaluar la influencia y el impacto sobre el cuerpo de la investigación publicada. Con el tiempo, se han desarrollado varios métodos para recopilar citas, por ejemplo, Google Scholar, Web of Science y Scopus, cada uno con sus propias ventajas y limitaciones.

Se han realizado esfuerzos para comparar entre las diversas fuentes de citas y, aunque se identifican problemas con cada enfoque, sus advertencias de uso apuntan a una similitud entre los métodos. Para la presente investigación se utiliza la plataforma de Scopus y Web of Science

(WoS), porque proporcionan algunas perspectivas únicas, como la posibilidad de clasificar las publicaciones por área científica, permitiendo delimitar la búsqueda de estudios exclusivos pertenecientes al área de las ciencias de la administración, además, son las bases de datos más amplias dentro de la comunidad académica.

El creciente interés en el análisis de citas ha proporcionado a los analistas una serie de perspectivas sobre la influencia de cualquier artículo, libro o incluso investigador específico. En la presente investigación se emplean varias de esas perspectivas para comprender de manera objetiva las contribuciones del análisis de datos multivariante. En primer lugar, se hace una comparación entre los artículos y textos más citados en ciencias de la administración. La búsqueda inicial consiste en identificar todos los artículos pertenecientes al área de las ciencias de la administración, por lo que se utiliza el filtro “Campo de Investigación” para reconocer todos los artículos pertenecientes a las “Ciencias de la Administración”.

Luego, una vez realizado el filtro, se realiza una categorización para separar investigaciones cuantitativas y cualitativas. Una vez realizada la separación, se clasifican los métodos cuantitativos y se delimita el análisis a los métodos multivariantes. Una vez realizada la categorización, se identifica que el mayor número de investigaciones utiliza tres tipos de instrumentos, lo que representa por lo menos los dos primeros cuartiles (Q1, Q2) equivalente al 50%. Los tres tipos de instrumentos son regresiones lineales, ecuaciones estructurales y técnicas de Inteligencia Artificial (IA) como redes neuronales.

Una vez identificados los instrumentos, se realiza la búsqueda específica en las plataformas Scopus y WoS con la finalidad de encontrar las tendencias a lo largo del tiempo, la búsqueda se hace por medio de operadores booleanos, con los tres principales métodos de análisis multivariante, “regresión lineal”, “SEM” e “IA”, con sus respectivos sinónimos y traducciones al idioma inglés, buscando evitar sesgos en la recolección de datos. Una vez realizada la búsqueda, se hace el análisis de las tendencias por medio de un gráfico de líneas a través del tiempo y, a partir de los resultados se identifican los temas y las tendencias futuras de las herramientas multivariantes dentro de los últimos 5 años.

A continuación, se exponen los resultados obtenidos mediante el análisis bibliográfico en las plataformas Scopus y WoS.

RESULTADOS Y DISCUSIÓN

En la Figura 1 se muestra el número de publicaciones que utiliza una técnica de análisis multivariante en el campo de las ciencias de la administración. Se observa un crecimiento exponencial en el uso del método SEM en los últimos cinco años, cerrando la brecha con el método de regresión. Por su parte, la IA ha incrementado su participación en los últimos tres años, lo que se puede asociar al desarrollo tecnológico que permite el uso de Big data y programas computacionales, desde pequeños o limitados grupos de trabajo. Es decir, la accesibilidad a la tecnología ha permitido la aplicación de la herramienta con mayor frecuencia.

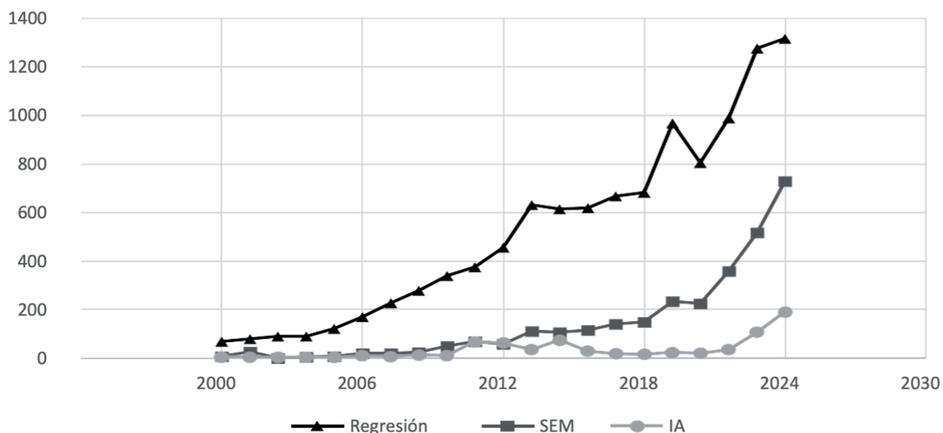


Figura 1. Número de publicaciones que utiliza una técnica de análisis multivariante en el campo de las ciencias de la administración. Fuente: *Elaboración propia.*

Por su parte, aunque la IA ha sido componente del campo de la informática durante muchas décadas, solo recientemente ha sido aplicada a diferentes áreas de las ciencias sociales y del comportamiento. Conforme a lo mostrado en la Figura 1, se observa el comienzo del Big Data y la IA, por lo que a continuación se exponen los conceptos y las técnicas particulares que se utilizan para lograr implementar dichas técnicas en un estudio científico.

De acuerdo con Robila y Robila (2020), las metodologías de IA se han aplicado con éxito a tres áreas principales de las ciencias sociales de la cual se desprenden las ciencias de la administración. En primer lugar, (1) para aumentar la eficacia del diagnóstico y la predicción de diferentes condiciones sociales como lo pueden ser las organizaciones empresariales; (2) para aumentar la comprensión del desarrollo y el funcionamiento humano, y (3) para aumentar la eficacia de la gestión de datos en diferentes servicios sociales y humanos.

A partir de lo descrito, los bosques aleatorios, las redes neuronales y la red elástica se encuentran entre las técnicas de IA más frecuentes que se utilizan para la predicción, complementando los enfoques tradicionales, mientras que el procesamiento del lenguaje natural y la robótica continúan aumentando su papel en la comprensión del funcionamiento humano y mejorando los servicios sociales.

En efecto, la IA también se ha expandido significativamente más allá de intentar simular funciones cerebrales. Uno de los más importantes y componentes recientes de IA, Machine Learning (ML), se refiere al proceso de construcción de un modelo de datos en el que los pasos algorítmicos se determinan y luego se refinan en función de un análisis de datos. En otras palabras, el diseño del algoritmo se “aprende” conforme a las propiedades de los datos que se espera procesar. En muchos ML, los algoritmos están anclados en estadísticas, a saber: un grupo utiliza varios enfoques de análisis de regresión para el ajuste de modelos; LASSO (operador de selección y contracción mínima absoluta) y las redes elásticas se han empleado ampliamente. Específicamente, las redes elásticas han ganado popularidad como modelo de ajuste, que permite situaciones en las que el número de puntos de datos es pequeño en comparación con la cantidad de características.

Asimismo, las técnicas de ML se pueden clasificar según el tipo de resultado esperado o de algorítmico subyacente. En términos de métodos algorítmicos, los enfoques donde se han visto avances importantes incluyen redes neuronales, máquinas de vectores de soporte, redes bayesianas y algoritmos genéticos (Robila & Robila, 2020). Respecto al resultado esperado, varios algoritmos ML dan como resultado una decisión / clasificación, extracción o selección de características, detección o anomalías, reglas de asociación de identificación, entre otros. Como ejemplo, los árboles de decisión o clasificación son algoritmos que, comenzando con un conjunto de valores de características para una observación, devuelven un valor específico, es decir, una decisión.

Este valor se obtiene siguiendo una secuencia de pruebas de decisión aplicadas a los valores de las características de entrada. El diseño específico para un árbol puede ser hecho usando regresión, enfoques de gradiente y las pruebas de decisión específicas se determinan en función de la muestra. Una variación de este enfoque son los árboles de decisión difusos que reemplazan el resultado único por un conjunto de resultados, junto con puntuaciones de confianza para cada valor. Este enfoque ofrece más flexibilidad en la interpretación de los resultados y mejor alineación con el proceso de pensamiento humano. Como ejemplo, un difuso sistema de árbol de decisión destinado a clasificar las emociones humanas por medio de las caras, proporcionan una mejor comprensión que un sistema binario, lo que manifiesta las complejas emociones humanas. Finalmente, en un refinamiento adicional de bosques aleatorios (colecciones de decisiones árboles), cada árbol de la colección se construye usando subconjuntos de datos o subconjuntos de características y el resultado de la clasificación se entrega utilizando el promedio de los resultados del árbol individual o por mayoría de votos.

La Red Neuronal (NN, por sus siglas en inglés) se refiere a un marco algorítmico que tiene como objetivo simular una versión simplificada de conexiones neuronales en el cerebro (Sze et al., 2017). Las NN están formadas por una colección de nodos conectados, es decir, neuronas. Cada nodo recibe uno o más valores como entrada y genera un valor de salida. Usando una formulación matemática (como un promedio

ponderado), los valores de entrada se combinan en un solo valor que luego se procesa a través de un umbral de cálculo (función de activación) para generar la salida.

Múltiples neuronas se combinan en capas y con la salida de una de ellas se convierte en entrada para las neuronas en la siguiente. Si el objetivo del algoritmo ha sido la clasificación, luego la salida de un NN se utiliza con el valor de salida objetivo que sugiere un etiquetado para una clase específica. Los pesos específicos para cada uno de los insumos se adaptan en función de un proceso de formación, que utiliza casos de ejemplo de los datos. Existen varios enfoques para la realización de este proceso. Otra variante es la mencionada por Gil de Zúñiga y Diehl (2016), y se relaciona con una Red Neuronal Profunda (DNN, por sus siglas en inglés), además, es una NN con una gran cantidad de capas. El entrenamiento de DNN permite modelar entornos de datos complejos. Sin embargo, el proceso de aprendizaje y de adaptar los pesos es computacionalmente costoso y requiere de grandes cantidades de datos. Por eso, aunque se han utilizado NN durante décadas, han aumentado en importancia solo recientemente a partir de cómo el hardware avanza y según el aumento o facilidad en la recolección de datos.

Por su parte, el procesamiento del lenguaje natural (PNL) es otra importante área de la IA. Y se refiere a la construcción de herramientas computacionales que analizan y representan el lenguaje humano en la comunicación a niveles complejos. Conceptos como voz a texto, traducción automática de semántica léxica y la segmentación, han logrado un progreso significativo, lo que permite la recopilación y procesamiento de datos que se realizarán en vastas colecciones, tales como foros de discusión. De acuerdo con Chen et al. (2016), la capacidad de combinar el lenguaje humano, el procesamiento y acceso a grandes bases de datos de información, ha permitido el diseño de entornos inteligentes complejos como Watson de IBM, que sirve como base tecnológica para sistemas de preguntas y respuestas.

También, Siegart et al. (2011) mencionan que la robótica se refiere a la investigación y desarrollo de Robots, es decir, sistemas que pueden realizar un conjunto complejo de tareas de forma autónoma. Mientras que la robótica tiene gran variedad de aplicaciones que incluyen la automatización industrial (por ejemplo, la auto-

matización de líneas de producción), desastres y asistencia (los robots se emplean en entornos no adecuados para humanos). A menudo, el aspecto más visible es el que tiene como objetivo que los sistemas robóticos realicen actividades humanas que se consideran inteligentes.

Los ejemplos recientes incluyen avances en la conducción autónoma o diseño de plataformas robóticas humanoides. Como muchas de estas actividades implican percepción, reconocimiento y planificación, la robótica está fuertemente relacionada con la IA. Con el avance en el aprendizaje profundo, la robótica sigue avanzando en la realización de actividades a nivel humano o por encima de él. Si bien, no es el tema central de este documento, es importante señalar que la robótica está ganando importancia en la conexión a las ciencias sociales y, a su vez, en las ciencias de la administración.

Más importante aún, y más allá de los avances en el campo, la IA se ha convertido en un enfoque popular para el procesamiento de datos debido a la creciente disponibilidad de herramientas como el idioma R o Python, que permite que los principiantes se introduzcan a la programación desde un enfoque fácil de usar. Esto se ve reflejado en publicaciones actuales dentro del campo de las ciencias de la administración. En la Tabla 1 se exponen las investigaciones actuales que hacen uso de las técnicas de análisis multivariante bajo el concepto de IA.

En la Tabla 1, se observan investigaciones recientes elaboradas en el campo de las ciencias de la administración y utilizando técnicas relativamente novedosas, que permiten planteamientos más complejos. La técnica con mayor uso dentro de la IA se concentra en redes neuronales, seguido de bosque aleatorio y red elástica. Por lo tanto, los resultados coinciden con lo propuesto por Robila y Robila (2020) respecto a los métodos que comienzan a incrementar su uso dentro de las investigaciones científicas, enfocadas al área de las ciencias de la administración.

Conforme a lo anterior y desde la perspectiva de Vogt y Johnson (2005), hay más de 20 formas diferentes de realizar análisis multivariante. El que se elija dependerá del tipo de datos que se tienen y de los objetivos propios del investigador. Por ejemplo, si tiene un solo conjunto de datos, tiene varias opciones: los árboles aditivos, el escalado multidimensional y el análisis de conglomerados son apropiados cuando las fi-

Tabla 1. Investigaciones recientes que hacen uso de las técnicas de análisis multivariante bajo el concepto Inteligencia Artificial dentro del campo de las ciencias de la administración.

Autor	Artículo	Metodología	Año
Rombaut y Guerry (2020)	The effectiveness of employee retention through an uplift modeling approach	Bosque Aleatorio	2020
Świecka et al. (2021)	Transaction factors' influence on the choice of payment by Polish consumers	Bosque Aleatorio	2021
Choudhury et al. (2021)	Machine learning for pattern discovery in management research	Bosque Aleatorio	2021
Sarkar y De Bruyn (2021)	LSTM Response Models for Direct Marketing Analytics: Replacing FeatureEngineering with Deep Learning	Redes Neuronales	2021
Ray et al. (2021)	Exploring the drivers of customers' brand attitudes of online travel agency services: Atext-mining based approach	Redes Neuronales	2021
Aakash et al. (2021)	How features embedded in eWOM predicthotel guest satisfaction: an application of artificial neural networks	Redes Neuronales	2021
Bock et al. (2021)	A simulation-based approach to business model design and organizational Change	Redes Neuronales	2021
Muniain y Ziel (2020)	Probabilistic forecasting in day-ahead electricity markets: Simulating peak andoff-peak prices	Red Elástica	2020
Rutz et al. (2011)	Modeling indirect effects of paid search advertising: Which keywords lead to more future visits?	Red Elástica	2011
Carstensen et al. (2020)	Predicting ordinary and severe recessionswith a three-state Markov-switching dynamic factor model: An application to the German business cycle	Red Elástica	2020

Fuente: *Elaboración propia.*

las y columnas de la tabla de datos representan las mismas unidades y la medida es una similitud o una distancia.

En tanto, el análisis de componentes principales (PCA) descompone una tabla de datos con medidas correlacionadas en un nuevo conjunto de medidas no correlacionadas. El análisis de correspondencia es similar al PCA. Sin embargo, se aplica a las tablas de contingencia. Aunque existen límites bastante claros con un conjunto de datos, por ejemplo, si tiene un único conjunto de datos en una tabla de contingencia, sus opciones se limitan al análisis de correspondencia. En la mayoría de los casos podrá elegir entre varios métodos.

En el estudio de Black y Babin (2019), se hace una revisión sobre la bibliografía presentada por Hair sobre las publicaciones relacionadas a los métodos afines al análisis multivariante, la primera edición de Hair (Hair et al., 1979) incluyó la información básica de métodos multi-

variantes y fue construida a partir de un marco de regresiones lineales. El análisis de cluster se agregó en la segunda edición, y gran parte del resto del libro permaneció igual en estructura. La tercera edición agregó un capítulo introductorio sobre SEM, pero también fue la primera edición en proporcionar una sintaxis de software completa (IBM-SPSS) y salidas para todos los capítulos. Y aunque otros libros han incorporado posteriormente esta característica, en ese momento fue un avance notable y un desafío de proporcionar, dadas las capacidades de descarga del día.

Por su parte, Felt (2016) menciona que, en la cuarta edición, se agregó el exitoso proceso de decisión de seis pasos utilizado en la discusión de cada técnica. De esta manera, el investigador tiene un marco coherente que puede aplicar a cualquier situación de investigación, y los seis pasos proporcionan una "lista de verificación" básica para garantizar que se hayan considerado

todos los problemas necesarios. La quinta edición, que duró 8 años (1998-2006) sin revisión, reconoce los temas avanzados y amplió el capítulo introductorio sobre el modelado de ecuaciones estructurales (SEM). En ese momento, se completó el cambio hacia el uso del análisis factorial como técnica introductoria, lo que abrió el camino para el mayor cambio en la tabla de contenido en la historia del libro.

La sexta edición amplió la cobertura de SEM a tres capítulos completos. Por lo tanto, el libro proporcionó un enfoque fácil de usar para una técnica que no se consideraba simple de usar en ese momento. La sexta edición también vio una formalización de las Reglas de Pulgar en cada área temática, proporcionando un conjunto de pautas para ayudar al investigador en decisiones prácticas como estadísticas. Se creó un nuevo conjunto de datos, más apropiado para SEM: HBAT-SEM, para ilustrar la técnica.

En la séptima edición publicada en 2010, un cuarto capítulo sobre SEM parece proporcionar descripciones de técnicas más avanzadas. En 2018 llegó la octava edición, ahora con la

editorial Cengage International. La adición de un capítulo para SEM de mínimos cuadrados parciales, que se introdujo originalmente en la sexta edición, proporciona el mayor cambio en la tabla de contenido. Sin embargo, la edición también integra principios de ciencia de datos (Big Data) en cada capítulo, proporcionando una perspectiva ampliada sobre la aplicabilidad de las técnicas no solo a problemas estadísticos clásicos, sino también a la situación emergente de análisis que involucra conjuntos de datos muy grandes en términos de casos y variables. También se amplía la cobertura de los tipos de análisis multinivel.

El objetivo general en todos estos desarrollos ha sido proporcionar al investigador, ya sea en el ámbito académico o profesional, una “práctica”. A continuación, se expone en la Figura 2 la evolución de las publicaciones relacionadas a explicar los métodos estadísticos multivariantes que se pueden utilizar en el campo de las ciencias de la administración, basada en los libros publicados por Hair desde 1979.

En la Figura 2, se observa que a partir de 1979

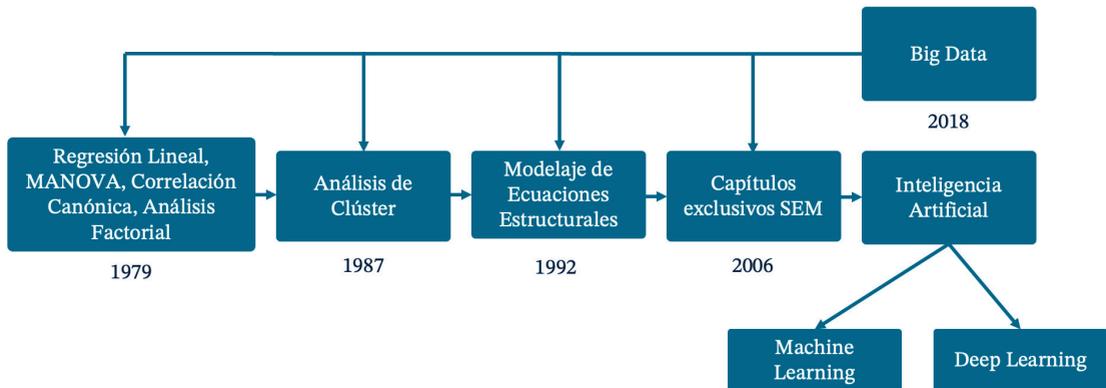


Figura 2. Evolución de los tópicos en el análisis multivariante. Fuente: Elaboración propia con base a Hair et al. (1979;1987;1992;2010;2014;2019) y Hair (2006).

las técnicas multivariantes predominantes se enfocaban en la regresión lineal: Análisis de varianza (MANOVA), Correlación Canónica y Análisis factorial. Para 1987 se incorpora el Análisis Clúster, seguido del modelado de ecuaciones estructurales (SEM) en 1992. Se observa que a inicios de la década de los 90 e inicios de los años 2000, se integran capítulos exclusivos del método SEM, derivados de la gran acepta-

ción en el ámbito académico para la realización de estudios complejos, donde la necesidad de involucrar un gran número de variables latentes y observables era evidente. Posteriormente, para el año 2018 en la edición más reciente, se observa la integración de una sección enfocada al Big data, una herramienta que se puede utilizar para mejorar los métodos previos, como el análisis de clúster, SEM y principalmente, la IA,

la cual se puede clasificar en dos grandes ramas: el Machine Learning y el Deep Learning, donde se puede hacer uso de redes neuronales para su desarrollo.

Sobre la base de lo anterior, se observa un desarrollo considerable de las principales técnicas multivariantes que pueden ser utilizadas en el ámbito de las ciencias de la administración, con la finalidad de contrastar el desarrollo de las publicaciones de Hair, conforme al desarrollo de la producción científica.

Conforme a lo mencionado por Hesse et al. (2015), uno de los desafíos asociados al alto volumen y diversos conjuntos de datos, en sí la síntesis de datos abiertos y las corrientes pueden traducirse en conocimientos prácticos. El concepto de traducir “Big data” es importante para las ciencias sociales y ciencias de la administración. Primero, porque es un cambio general hacia un uso intensivo de datos donde la ciencia ejercerá una influencia en todas las disciplinas científicas, pero particularmente en el comportamiento social, dada la riqueza del comportamiento y los constructos relacionados y capturados por grandes fuentes de datos. Segundo, la ciencia en sí misma es una empresa social, que aplica principios de las ciencias sociales a la realización de la investigación, por lo cual, debe ser posible mejorar algunos de los problemas sistémicos que plagan a la empresa científica en la era de Big data.

De acuerdo con Robila y Robila (2020), las técnicas de IA ahora son frecuentemente empleadas en diferentes campos debido a la creciente disponibilidad de datos y el aumento de la potencia y precisión computacional.

Lo presentado anteriormente se considera una mejora metodológica, no obstante, las técnicas contienen limitaciones, por ejemplo, en el caso de las redes neuronales hay cuatro razones principales por las que el aprendizaje profundo es complicado en cuanto a su correcta aplicación (Yehudai & Shamir, 2019). Podría decirse que la desventaja más conocida de las redes neuronales es su naturaleza de “caja negra”. En pocas palabras, no se sabe cómo o por qué su NN obtuvo un resultado determinado. Por ejemplo, cuando pones la imagen de un gato en una red neuronal y predice que será un automóvil, es muy difícil entender qué causó que llegara a esta predicción. Cuando tiene características que son interpretables por humanos, es mucho más fácil

comprender la causa del error. En comparación, los algoritmos como los árboles de decisión son muy interpretables. Esto es importante porque en algunos dominios, la interpretabilidad es fundamental.

Lo descrito es la razón por la que muchos bancos no usan redes neuronales para predecir si una persona es solvente, porque necesitan explicar a sus clientes por qué no obtuvieron el préstamo, de lo contrario, la persona puede sentirse tratada injustamente. Lo mismo ocurre con sitios como Quora, si un algoritmo de aprendizaje automático decidiera eliminar la cuenta de un usuario, se le debería dar una explicación del porqué. Otros escenarios serían decisiones comerciales importantes, por ejemplo, el director ejecutivo de una gran empresa tomando una decisión sobre millones de dólares sin entender por qué debería tomar dicha decisión ¿Solo porque la “computadora” dice que necesita hacerlo? En las investigaciones pasa un problema similar, en el caso de que un estudio se quiera replicar con la finalidad de validar los resultados, difícilmente se llegaría al mismo resultado, y en caso de hacerlo, no se tendría el 100% de certeza del porqué se llegó al mismo resultado.

Un segundo inconveniente es la duración del desarrollo, aunque hay bibliotecas como *Keras* que hacen que el desarrollo de redes neuronales sea bastante simple, en ocasiones se necesita más control sobre los detalles del algoritmo, como cuando se intenta resolver un problema difícil con el aprendizaje automático que nadie ha hecho antes. En ese caso, puede usar *Tensor flow* que brinda más oportunidades, pero también es más complicado y el desarrollo lleva mucho más tiempo (dependiendo de lo que se desee compilar). Entonces, surge una pregunta práctica para cualquier individuo que pretenda utilizar la técnica: ¿realmente vale la pena que los ingenieros pasen semanas desarrollando algo que se puede resolver mucho más rápido con un algoritmo más simple?

El tercer inconveniente se centra en la cantidad de datos, porque las redes neuronales suelen requerir muchos más datos que los algoritmos tradicionales de aprendizaje automático, en al menos miles, sino millones de muestras etiquetadas. Este no es un problema fácil de resolver y muchos problemas de aprendizaje automático se pueden resolver con menos datos si se utili-

zan otros algoritmos.

Por último, se hace mención que el uso de la técnica puede ser computacionalmente caro. Por lo general, las redes neuronales también son más costosas desde el punto de vista informático. Los algoritmos de aprendizaje profundo de última generación, que realizan un entrenamiento exitoso de redes neuronales realmente profundas, pueden tardar varias semanas en entrenarse completamente desde cero. Por el contrario, la mayoría de los algoritmos tradicionales de aprendizaje automático tarda mucho menos en entrenarse, desde unos minutos hasta unas pocas horas o días.

La cantidad de potencia computacional necesaria para una red neuronal depende en gran medida del tamaño de sus datos, pero también de la profundidad y complejidad de su red. Por ejemplo, una red neuronal con una capa y 50 neuronas será mucho más rápida que un bosque aleatorio con 1000 árboles. En comparación, una red neuronal con 50 capas será mucho más lenta que un bosque aleatorio con solo 10 árboles. Por lo tanto, en regiones económicamente limitadas, las investigaciones pueden prescindir de la técnica, por no contar con el acceso al equipo necesario para su correcta aplicación.

CONCLUSIONES

Mediante la revisión sistemática de la literatura se identifica una evolución en las diversas técnicas de análisis multivariante, principalmente en el uso de las técnicas basadas en regresiones y en el modelaje por medio de ecuaciones estructurales. Las técnicas mencionadas son técnicas robustas que continúan vigentes, lo cual se ve reflejado en su uso dentro de las investigaciones actuales. Sin embargo, se observa un crecimiento en el uso de técnicas relacionadas con la IA, particularmente, bosques aleatorios, las redes neuronales y la red elástica, porque son las técnicas que comienzan un ascenso considerable en su uso, como metodología central para contrastar hipótesis.

Una vez identificadas las tendencias, es importante mencionar las limitaciones de dichas técnicas, porque, aunque se consideran estadísticamente robustas, la misma complejidad con la que interactúan, genera confusión en su adecuado uso. Por un lado, las técnicas se asocian

más con las ciencias computacionales, y es una barrera técnica que debe de superar el investigador, desde el uso del software hasta la disponibilidad del hardware. Una segunda complicación es comprender con claridad los supuestos y algoritmos que se encuentran detrás de las herramientas, con la finalidad de entender sus propias implicaciones respecto a los datos adecuados para la herramienta y la coincidencia, con la lógica relacionada al planteamiento de la investigación, es decir, mantener la congruencia en el planteamiento de la investigación, la técnica que se utiliza para contrastar la hipótesis y el correcto uso de los datos.

Un punto relevante se centra en las publicaciones de las redes sociales, puesto que están llenas de potencial para la minería y el análisis de datos. Reconociendo este potencial, los proveedores de plataformas restringen cada vez más el libre acceso a dichos datos. Este cambio ofrece nuevos desafíos para los científicos sociales y otras organizaciones sin fines de lucro, que buscan generar investigaciones para el análisis de temas públicos con el propósito de comprender mejor la interacción humana y mejorar la condición.

Dentro de las limitaciones más importantes se destaca el “área temática”, al abarcar únicamente el campo de las ciencias de la administración, y analizar estudios previamente en otras áreas donde posiblemente han sido utilizadas más técnicas de análisis multivariantes y que, de igual forma, pueden incorporarse al catálogo de las herramientas para el estudio de las ciencias de la administración.

Declaración de conflicto de interés:

Los autores declaran que no presentan conflictos de interés.

Financiamiento:

Los autores declaran que no recibieron financiamiento para el desarrollo de esta investigación.

Uso de Inteligencia Artificial (IA):

Los autores declaran que no recibieron asistencia de una IA durante el proceso de investigación, ni durante la escritura de este documento.

REFERENCIAS

- Aakash, A., Tandon, A., & Gupta Aggarwal, A. (2021). How features embedded in eWOM predict hotel guest satisfaction: An application of artificial neural networks. *Journal of Hospitality Marketing & Management*, 30(4), 486-507. <https://doi.org/10.1080/19368623.2021.1835597>
- Black, W., & Babin, B. J. (2019). Multivariate data analysis: Its approach, evolution, and impact. In *The Great Facilitator*, 121-130. Springer. https://doi.org/10.1007/978-3-030-06031-2_16
- Bock, A. J., Warglien, M., & George, G. (2021). A simulation-based approach to business model design and organizational change. *Innovation*, 23(1), 17-43. <https://doi.org/10.1080/14479338.2020.1769482>
- Carstensen, K., Heinrich, M., Reif, M., & Wolters, M. H. (2020). Predicting ordinary and severe recessions with a three-state Markov-switching dynamic factor model: An application to the German business cycle. *International Journal of Forecasting*, 36(3), 829-850. <https://doi.org/10.1016/j.ijforecast.2019.09.005>
- Cavanaugh, J. (2007). Encyclopedia of Statistical Sciences (2nd ed.). Samuel Kotz, Campbell B. Read, N. Balakrishnan, and Brani Vidakovic. *Journal of the American Statistical Association*. 1074-1075.
- Chen, Y., Argentinis, J. E., & Weber, G. (2016). IBM Watson: how cognitive computing can be applied to big data challenges in life sciences research. *Clinical Therapeutics*, 38(4), 688-701. <https://doi.org/10.1016/j.clinthera.2015.12.001>
- Choudhury, P., Allen, R. T., & Endres, M. G. (2021). Machine learning for pattern discovery in management research. *Strategic Management Journal*, 42(1), 30-57. <https://doi.org/10.1002/smj.3215>
- Dodge, Y. (2008). The Concise Encyclopedia of Statistics. Springer. <https://doi.org/10.1007/978-0-387-32833-1>
- Felt, M. (2016). Social media and the social sciences: How researchers employ Big Data analytics. *Big Data & Society*, 3(1), 1-15. <https://doi.org/10.1177/2053951716645828>
- Gil de Zúñiga, H., & Diehl, T. (2016). Citizenship, social media, and Big Data. *Social Science Computer Review*, 35(1), 3-9. <https://doi.org/10.1177/0894439315619589>
- Hair, J., Anderson, R., Tatham, R., & Grablovsky, B. (1979). *Multivariate Data Analysis*. PPC Books.
- Hair, J., Anderson, R., & Tatham, R. (1987). *Multivariate Data Analysis with Readings*. Macmillan.
- Hair, J., Anderson, R., Tatham, R. & Black, W. (1992). *Multivariate Data Analysis with Readings*. Macmillan.
- Hair, J. F. (2006). *Multivariate data analysis (6th Edition)*. Pearson Prentice Hall.
- Hair, J., Black, W., Babin, B., & Anderson, R. (2010). *Multivariate Data Analysis (7th Edition)*.
- Hair, J., Black, W., Babin, B., Anderson, R., & Tatham, R. (2014). *Multivariate data analysis: Global edition*. Pearson Education.
- Hair, J., Black, W., Babin, B., & Anderson, R. (2019). *Multivariate Data Analysis (8th Edition)*. Cengage.
- Hesse, B. W., Moser, R. P., & Riley, W. T. (2015). From Big Data to Knowledge in the Social Sciences. *The ANNALS of the American Academy of Political and Social Science*, 659(1), 16-32. <https://doi.org/10.1177/0002716215570007>
- Muniain, P., & Ziel, F. (2020). Probabilistic forecasting in day-ahead electricity markets: Simulating peak and off-peak prices. *International Journal of Forecasting*, 36(4), 1193-1210. <https://doi.org/10.1016/j.ijforecast.2019.11.006>
- Ray, A., Bala, P. K., & Rana, N. P. (2021). Exploring the drivers of customers' brand attitudes of online travel agency services: A text-mining based approach. *Journal of Business Research*, 128, 391-404. <https://doi.org/10.1016/j.jbusres.2021.02.028>

- Robila, M., & Robila, S. A. (2020). Applications of artificial intelligence methodologies to behavioral and social sciences. *Journal of Child and Family Studies*, 29, 2954-2966. <https://doi.org/10.1007/s10826-019-01689-x>
- Rombaut, E., & Guerry, M. A. (2020). The effectiveness of employee retention through an uplift modeling approach. *International Journal of Manpower*, 41(8), 1199-1220. <https://doi.org/10.1108/IJM-04-2019-0184>
- Rutz, O. J., Trusov, M., & Bucklin, R. E. (2011). Modeling indirect effects of paid search advertising: Which keywords lead to more future visits? *Marketing Science*, 30(4), 646-665. <https://doi.org/10.2307/23012017>
- Sarkar, M., & De Bruyn, A. (2021). LSTM response models for direct marketing analytics: Replacing feature engineering with deep learning. *Journal of Interactive Marketing*, 53(1), 80-95. <http://dx.doi.org/10.2139/ssrn.3601025>
- Sieglwart, R., Nourbakhsh, I. R., Scaramuzza, D., & Arkin, R. C. (2011). *Introduction to autonomous mobile robots*. MIT press. <https://doi.org/10.5860/eleccion.49-1492>
- Świecka, B., Terefenko, P., & Paprotny, D. (2021). Transaction factors' influence on the choice of payment by Polish consumers. *Journal of Retailing and Consumer Services*, 58, 102264. <https://doi.org/10.1016/j.jretconser.2020.102264>
- Sze, V., Chen, Y. H., Yang, T. J., & Emer, J. S. (2017). Efficient processing of deep neural networks: a tutorial and survey. *Proceedings of the IEEE*, 105(12), 2295-2329. <https://doi.org/10.1109/JPROC.2017.2761740>
- Verma, G., & Sharma, K. (2017). The Role of quantitative techniques in business and management. *Journal of Humanities Insights*, 1(01), 24-26. <https://doi.org/10.22034/JHI.2017.59559>
- Vogt, P., & Johnson, R. (2005). *Dictionary of Statistics & Methodology: A Nontechnical Guide for the Social Sciences*. SAGE. <https://doi.org/10.4135/9781071909751>
- Yehudai, G., & Shamir, O. (2019). On the power and limitations of random features for understanding neural networks. <https://doi.org/10.48550/arXiv.1904.00687>