

**CARACTERIZACIÓN DE DEMANDA PARA AJUSTAR LA
CONFIGURACIÓN DE UN MÉTODO DE PRONÓSTICOS EN LA
CADENA DE SUMINISTROS**

Trabajo de investigación

Para optar al título de Magíster en Ingeniería

Presentado por:

PAULA ANDREA ZAPATA FLÓREZ

Asesor:

CARLOS ALBERTO CASTRO ZULUAGA

Escuela de Ingeniería

Universidad EAFIT

Abril 2018

Resumen

Los pronósticos soportan la toma de decisiones reduciendo la sensación de incertidumbre. Particularmente en la planeación de operaciones de cualquier organización, los pronósticos asisten en la estimación de materiales, fuerza de trabajo, producción y otros procesos de relevancia. Este trabajo se da a la tarea de implementar una técnica de pronósticos de complejidad media, basada en el ajuste estacional de las series a través de descomposición clásica, y la estimación de pronósticos a través del método de suavización de *Holt*, en conjunto con un algoritmo de *clustering* de series de tiempo entrenado con base en características estadísticas que le permiten al modelo ofrecer un rango valores de alfa y beta acordes a la estructura de cada serie. Las características incluyen, *media*, *mediana*, *desviación estándar (Desvest)*, *coeficiente de variación (CV)*, *entropía* y *asimetría*, y son tomadas en ventanas de tiempo para permitir la comparación de las series de tiempo sin dependencia de la escala de los datos y para evitar realizar un juicio subjetivo basado en una mirada global. El modelo pretende facilitar su comprensión, posibilitar su reproducción y uso generalizado en series de datos que exponen diversos patrones de comportamiento, y además ofrecer una configuración adecuada de los parámetros de entrada de manera flexible, práctica y eficiente para ser usado en el entorno local. Este modelo fue aplicado a datos de la competencia NN₃ y mostró resultados positivos en comparación con un escenario base creado a partir de juicios de valor.

Agradecimientos

A todas las personas que estuvieron involucradas en este nuevo proceso, quienes en su medida contribuyeron a una feliz culminación.

A mi asesor, el profesor Carlos Alberto Castro por su paciencia, asesoría y cuidadosa revisión en el desarrollo del presente trabajo.

A Giovanni Espinal, por su apoyo incondicional y profesional sin el cual no hubiera podido terminar. Sus palabras de aliento transformaron una idea en una meta cumplida.

A mi familia por permanecer cerca.

A mis asesores personales, Pablo Maya y Daniel La Rotta, por prestarme sus conocimientos cuando los necesité.

A la Universidad EAFIT, por darme la oportunidad de enriquecer mi formación académica y personal.

Tabla de Contenido

| | |
|---|------------|
| Resumen..... | ii |
| Agradecimientos | iii |
| Introducción | 1 |
| 1. El Problema de la Investigación | 3 |
| 1.1. Planteamiento del Problema | 3 |
| 1.2. Enfoque de la Investigación | 5 |
| 1.3. Esquema del Trabajo | 6 |
| 2. Objetivos..... | 7 |
| 2.1. Objetivo General..... | 7 |
| 2.2. Objetivos específicos | 7 |
| 3. Marco teórico y revisión de literatura | 8 |
| 3.1. Métodos Cuantitativos y Cualitativos para Pronosticar | 8 |
| 3.2. Datos de corte transversal y series de tiempo en modelos cuantitativos | 9 |
| 3.3. Series de tiempo en lugar de modelos trasversales..... | 10 |
| 3.4. Componentes o patrones en series de tiempo | 11 |
| 3.5. Métodos Cuantitativos de Series de Tiempo | 13 |
| 3.6. Estimación de parámetros..... | 20 |
| 3.6.1. Constantes de suavización | 20 |
| 3.6.2. Otros enfoques de la estimación..... | 21 |
| 3.7. Hallazgos de la revisión de la literatura..... | 23 |
| 4. Metodología | 25 |
| 4.1. Revisión de la literatura..... | 25 |
| 4.2. Modelo integrado..... | 26 |
| 4.2.1. Etapa de Entrenamiento..... | 28 |

| | | |
|-----------|---|-----------|
| 4.2.1.1. | Ajuste estacional a través de descomposición de componentes | 28 |
| 4.2.1.2. | Suavizamiento de Holt | 29 |
| 4.2.1.3. | Optimización | 31 |
| 4.2.1.4. | Extracción de características y clustering o agrupamiento | 33 |
| 4.2.1.5. | Rango α y β | 34 |
| 4.2.2. | Etapas de Clasificación | 35 |
| 4.3. | Incorporación en software | 35 |
| 4.4. | Validación del software | 36 |
| 5. | Resultados | 37 |
| 5.1. | Consideraciones particulares de la experiencia | 37 |
| 5.2. | Rangos de α y β sugeridos según grupo | 38 |
| 5.3. | Resultados validación del modelo | 39 |
| 5.3.1. | Pruebas de normalidad de los errores | 41 |
| 5.3.2. | Comparación entre muestras pareadas | 42 |
| 5.3.2.1. | Pruebas bilaterales o de dos colas | 43 |
| 5.3.2.2. | Pruebas de una cola (Diferencia de medianas <0) | 43 |
| 5.4. | Análisis de datos atípicos | 43 |
| 5.5. | Relación entre características estadísticas y error de pronóstico | 44 |
| 5.6. | Análisis de los grupos formados | 46 |
| 5.6.1. | Análisis Individual | 48 |
| 5.6.1.1. | Grupo 1 | 48 |
| 5.6.1.2. | Grupo 2 | 49 |
| 5.6.1.3. | Grupo 3 | 50 |
| 5.6.1.4. | Grupo 4 | 51 |
| 5.6.1.5. | Grupos 5, 6, 7 y 8 | 51 |

| | |
|--|-----------|
| 5.6.1.6. Aspectos globales de las correlaciones | 54 |
| 6. Discusión..... | 57 |
| 6.1. Hallazgos fundamentales | 57 |
| 6.2. Conclusiones y limitaciones | 58 |
| 6.3. Trabajos futuros | 59 |
| Referencias..... | 61 |
| Anexo A. Métodos cuantitativos de series de tiempo de complejidad media ... | 69 |
| A.1. Métodos cuantitativos de series de tiempo simples | 69 |
| A.2. Regresión lineal..... | 69 |
| A.3. Promedios móviles..... | 70 |
| A.4. Suavización exponencial simple | 71 |
| A.5. Suavización exponencial lineal de <i>Holt</i> o doble suavización | 71 |
| A.6. Suavización exponencial doble con tendencia exponencial | 72 |
| A.7. Suavización exponencial doble con tendencia amortiguada | 73 |
| A.8. Suavización exponencial de <i>Holt-Winters</i> | 73 |
| Holt-winters aditivo | 74 |
| Holt-winters multiplicativo | 74 |
| A.9. Método de descomposición por componentes | 75 |
| Anexo B. Definición de características estadísticas extraídas de las series | 78 |
| B.1. Media aritmética..... | 78 |
| B.2. Mediana | 78 |
| B.4. Coeficiente de Variación | 79 |
| B.5. Asimetría | 79 |
| B.6. Entropía..... | 80 |

Introducción

Los seres humanos se ven obligados a tomar decisiones a diario y este proceso los lleva a considerar las consecuencias en el futuro, incluso de manera intuitiva. Las predicciones basadas en datos apoyan la toma de decisiones de una manera más objetiva ya que pueden ayudar a disminuir la sensación de incertidumbre y a prepararse para los efectos negativos o positivos, que acarrearán estas decisiones. Los pronósticos son empleados por una amplia gama de entidades, académicas y empresariales, y dada su preponderancia son usados en áreas como el análisis de decisiones, la planeación de la demanda, la comercialización y el mercadeo, y el análisis de información (*data analysis*), entre otras. Particularmente, en la administración de operaciones la información de demanda y ventas históricas es uno de los principales insumos para la obtención de pronósticos, que a su vez alimentan la planeación en la organización, para integrarse finalmente a la toma de decisiones.

Desde 1950 y hasta la actualidad, se vienen desarrollando métodos de pronósticos para la administración y gestión de las operaciones en todo tipo de organizaciones, impulsados por los desarrollos científicos, la sistematización de la información y como respuesta natural a las necesidades diarias que se presentan tanto en la manufactura como en los servicios.

Precisamente en el campo empresarial los gerentes se enfrentan a diversas situaciones que requieren de toma de decisiones con consecuencias significativas dentro de las operaciones de la organización. Es el caso de la programación de la producción donde son necesarias las predicciones de la demanda a corto plazo de una cantidad considerable de artículos para poder llevar a cabo la manufactura, ejercicio que demanda tiempo de trabajo, recursos escasos, e inmediatez de resultados, entre otras necesidades, por lo que sería poco práctico aplicar una técnica avanzada de pronósticos en cada uno de ellos. En vista de lo anterior, es posible encontrar en la literatura técnicas aplicables a un número elevado de artículos y que buscan proporcionar mejores resultados para las organizaciones. En ese sentido, este trabajo de investigación aborda las técnicas de proyección, específicamente las cuantitativas de series de tiempo, que correspondan con las necesidades de recursos expresadas

anteriormente y adicionalmente ofrece un apoyo alternativo en la configuración de los parámetros de entrada basada en las características de la información, esto con el fin de proponer un modelo único de pronósticos que brinda directrices para su uso en la gestión de la demanda de una organización. El modelo propuesto fue validado con la base de datos mensual de la *NN3 Neural Forecasting Competition*, una extensión de las competencias M3 cuyo uso es frecuente en el campo de inteligencia computacional, inicialmente como competencia en donde se prueban diferentes métodos y se validan con medidas de error o rendimiento, y abierta al público para futuras referencias (Crone, Hibon y Nikolopoulos, 2011).

1. El Problema de la Investigación

1.1. Planteamiento del Problema

Para el desarrollo de un negocio, sea este de manufactura o servicios, resulta imperativo el uso de un modelo de pronósticos de su demanda, ya que esta tarea incide en otros procesos de la organización, como la programación de la producción, la planeación de ventas y operaciones, y la gestión de inventarios, entre otros temas de interés.

Cuando cualquiera de estos u otros procesos de planeación se realizan bajo supuestos intuitivos de la demanda que no logran percibir aproximadamente su comportamiento real, como consecuencia, se puede incurrir en costos de sobreproducción o de faltantes, que son algunos de los gastos más importantes dentro de las empresas (Brueggen, Grabner y Sedatole, 2014), lo que lleva a incrementos en los niveles de inventario, retrasos en las entregas, problemas de almacenamiento o disminución de los niveles de servicio, entre otros indicadores que directa o indirectamente afectan la competitividad de una organización.

De esta manera se busca que los valores de la demanda sean pronosticados sobre la base de métodos y modelos formales que puedan replicarse, y monitorearse periódicamente para mejorar sus resultados. En este aspecto, las empresas de países en vía de desarrollo como Colombia pueden desistir de implementar modelos cuantitativos e inclinarse por otros cualitativos y de juicio, debido al esfuerzo que supone el uso de modelos consistentes, que requieren de control y monitoreo, o porque simplemente este proceso no se considera relevante, posiblemente porque se cree que no influye directamente en el núcleo de las operaciones de la organización (Albarune y Habib, 2015; Lewis-Beck, 2005; Gilliland, Sglavo, y Tashman, 2016). Si bien la ejecución de un modelo de pronósticos exige cierto conocimiento cuantitativo o metodológico, esto no debería ser un impedimento para la aplicación de métodos exhaustivos que medien su comprensión e impacto dentro del campo laboral. Sería de utilidad entonces para las empresas locales, que no disponen de un equipo dedicado exclusivamente al área de pronósticos o que no cuentan con las competencias para implementar modelos de predicción, un método robusto o una herramienta, que

permita obtener mejores pronósticos de diversos tipos de productos, a través de metodologías convencionales sin requerir de una comprensión profunda del tópico.

Para resolver el problema que supone estimar un pronóstico por medio de series de tiempo, existen diversos métodos clasificados principalmente como cualitativos y cuantitativos. Los cualitativos por lo general usan la opinión de expertos para dar una apreciación subjetiva de lo que serían los valores futuros de la demanda usando especialmente su experiencia para determinarlos (Bowerman y O'Connell, 2007). Los métodos cualitativos se adecuan a las situaciones en donde existe poca o ninguna información acerca de un producto, o cuando la situación cambia constantemente y el pronóstico debe adaptarse a estos cambios. Sin embargo, al estar basados principalmente en el juicio humano pueden verse afectados por factores externos como el estado emocional o social de la persona, o la memoria, lo que puede hacer el método no replicable y difícil de comprender (Hyndman y Athanasopoulos, 2013), hecho que puede entorpecer las operaciones en la gestión logística de una organización (Lee y Adam Jr, 1986).

Los métodos cuantitativos por su parte usan información histórica basada exclusivamente en el tiempo (series de tiempo) o variables asociadas a la que se pretende predecir (modelos causales), con el objetivo de pronosticar la variable de interés a través de metodologías formalizadas que requieren un conocimiento mayor de áreas como la estadística para comprender su funcionamiento e implementación. Dentro del marco teórico, se encuentran algunos métodos convencionales como promedios móviles, suavización exponencial, descomposición, *Box-Jenkins*, y regresión lineal y no lineal, entre otros (Bowerman y O'Connell, 2007). En la literatura se encuentran otros acercamientos más complejos, principalmente relacionados a modelos causales que son usados especialmente en ciencias de la computación y en tópicos como *Big Data* por la gran cantidad de variables que se analizan (Sapankevych y Sankar, 2009; Chen y Wang, 2007), pero que pueden suponer una inversión significativa en recursos financieros y mayor tiempo de dedicación para un desarrollo que finalmente será *ad hoc*, o en otras palabras, creado para atender las necesidades específicas de cada organización. Igualmente, en la literatura se indica que no existe

una relación directa entre la complejidad del modelo y la calidad de sus resultados (Green, y Armstrong, 2015), por tal razón se puede suponer que un modelo convencional de complejidad media, podría ser un adecuado soporte para las organizaciones.

1.2. Enfoque de la Investigación

Ante la variedad de métodos que se encuentran en la literatura, el enfoque del presente texto se centra en un modelo de complejidad media, que sea de utilidad para la administración de operaciones, que garantice su replicabilidad, que no sea influenciado por juicios de valor, que se encuentre dentro del alcance de la persona a cargo, que no demande un equipo exclusivo de investigación al respecto, y que pueda ser usado en una alta gama de productos con resultados apropiados para cada escenario en la industria, esto considerando los diferentes parámetros de entrada que dicho modelo pueda requerir. En resumen, el método debe ofrecer una relación costo-beneficio adecuada para los negocios, que brinda flexibilidad y facilita la inclusión de criterios de juicio, sin excluir la rigurosidad científica en su uso.

A la luz de estas necesidades se consideran los métodos cuantitativos de series de tiempo, porque permiten reproducibilidad y análisis debido a que se basan en formulaciones matemáticas formales. El trabajo se centra en los métodos convencionales los cuales pueden resultar más familiares para la industria dada su trayectoria (Gardner, 2006). En estos, es imperativo encontrar un modelo o una combinación que pueda ajustarse a una gama de series de datos minimizando el riesgo de que se conviertan en diseños *ad hoc*. Adicionalmente, es de importancia para el investigador y el empresario, especialmente para quienes se encuentran en una etapa temprana de entrenamiento, hacerse a una idea de la configuración de parámetros más conveniente para ellos. Finalmente teniendo esto en consideración se plantean las siguientes preguntas de investigación:

Pregunta 1. ¿Qué método o conjunto de métodos de complejidad media se pueden seleccionar de la literatura que sean adecuados para la gestión de pronósticos de demanda y que faciliten su implementación en un ambiente de pequeñas y medianas empresas?

Igualmente, se desea facilitar el uso de la herramienta para quienes se encargan de gestionar los pronósticos, pero no se encuentran dedicados exclusivamente a ellos, o bien para quienes no poseen de experticia suficiente para determinar los parámetros de entrada que los métodos requieren. Entonces, de acuerdo con esta necesidad se plantea una segunda pregunta:

Pregunta 2. ¿Cómo configurar adecuadamente el método de pronósticos, esto es, hallar los parámetros de entrada apropiados de acuerdo con las características de demanda del producto?

1.3. Esquema del Trabajo

El estudio se distribuye así: Los objetivos delimitan el alcance de la investigación, esto es, definir un modelo de pronósticos y sus parámetros que pueda ser aplicado a múltiples productos, basado en las características de la información. En el marco teórico se discuten los diferentes métodos de pronósticos encontrados en la literatura, de acuerdo con el camino descrito en los objetivos, los cuales ayudarán a responder la primera pregunta de investigación. Tras su lectura, se enuncian diferentes acercamientos académicos orientados a la estimación de parámetros, proceso necesario para implementar las técnicas de proyección expuestas. A partir de este contenido se resaltan los principales hallazgos de la literatura en cuanto a método y metodología, acción que conlleva a la segunda parte del estudio: el desarrollo del modelo integral, sus resultados y posterior análisis.

2. Objetivos

2.1. Objetivo General

Definir un modelo de pronósticos y sus parámetros que pueda ser aplicado a múltiples productos, basado en las características de la información dentro de cada serie de tiempo.

2.2. Objetivos específicos

- Analizar la literatura correspondiente a métodos de pronósticos aplicados de complejidad media que permitan adaptarse a múltiples situaciones.
- Facilitar la tarea de configuración o selección de parámetros en el método elegido.
- Incorporar el método de pronósticos elegido y la configuración de parámetros en una aplicación de computador.
- Validar el programa en comparación con un escenario base.

3. Marco teórico y revisión de literatura

Para la toma de decisiones a corto, mediano y largo plazo en las organizaciones, se han desarrollado desde hace más de seis décadas modelos de predicción, en un principio estadísticos, para determinar el comportamiento de los procesos en el futuro. Por mucho tiempo se llegó a creer que los avances tecnológicos llevarían a desarrollar modelos completamente exactos, sin embargo, estos mismos avances permitieron a la comunidad científica comprender sus limitaciones y las del objeto a pronosticar, dando paso a diferentes vertientes cuyas aplicaciones se dieron en campos como la estadística, la economía, la investigación de operaciones, la psicología, la mercadotecnia y la meteorología, entre otras áreas.

Diversos autores han contribuido a la formación de teorías y herramientas para la predicción de la demanda con efectos en la gestión de la cadena de suministro para varios tipos de organización. El objetivo de este inciso corresponde entonces al primer ítem del numeral 2.2., y pretende enunciar las principales contribuciones encontradas en la literatura, comenzando con algunas generalidades para concentrarse en aquellas que coinciden con las necesidades enunciadas en la primera pregunta de investigación, es decir, se presta particular interés a métodos cuantitativos de series de tiempo de complejidad media, disponibles para un gerente. De igual forma, se exploran diversas maneras de estimar parámetros, los cuales son elementos de entrada en el uso de la gran mayoría de métodos, para lo cual se revisaron las bases de datos Google Scholar® y Scopus®, así como la literatura básica en pronósticos para la cadena de suministro.

3.1. Métodos Cuantitativos y Cualitativos para Pronosticar

Los modelos y métodos de pronósticos están divididos principalmente en dos vertientes, cualitativos y cuantitativos. Los métodos cuantitativos están basados en modelos matemáticos e información numérica para determinar los pronósticos (Armstrong, 2001). De esta forma, se obtiene un resultado robusto en la medida en que, si se sigue el mismo procedimiento sobre el mismo conjunto de datos, se obtendrá el mismo resultado, una característica del pensamiento científico que los diferencia de los métodos cualitativos (Sale, Lohfeld, y Brazil, 2002). Estos últimos,

aunque estructurados, dependen del juicio humano y por ello son potencialmente subjetivos, hecho que podría dificultar la tarea de replicabilidad que se pretende para la presente investigación.

Para usar los métodos cuantitativos, debe haber información numérica disponible, y se considera que los datos pueden tener un comportamiento o patrones que continuarán en el futuro (Makridakis, Wheelwright y Cortés Fregoso, 2009). Si bien es posible usar solo uno de los enfoques, cualitativo o cuantitativo, diversos autores (Webby y O'Connor, 1996; Lawrence, Edmundson y O'Connor 1986), han discutido las ventajas de una mixtura entre ellos, debido a las fluctuaciones de los datos que dependen del contexto, y que pueden ser perceptibles fácilmente por el juicio humano. Seifert et al. (2015) por ejemplo, indican que el uso combinado de pronósticos de juicio y estadísticos para campos tan volátiles como el de la moda, puede ser de utilidad si las decisiones humanas están limitadas a detectar comportamientos no lineales en los datos, mientras que los lineales se tratan mediante un método estadístico. De modo que los métodos cuantitativos, pueden detectar patrones que el juicio humano puede ignorar, sin embargo, cuando la ocasión así lo requiere este juicio puede ser de utilidad para realizar ajustes.

3.2. Datos de corte transversal y series de tiempo en modelos cuantitativos

Cuando se aplican métodos cuantitativos es posible encontrar dos tipos o categorías en las que están organizados los datos: de corte transversal (*cross-sectional*) o de series de tiempo. La primera categoría se refiere a información que se recolecta a través de múltiples observaciones en un mismo punto en el tiempo, o cuando las diferencias en el tiempo son despreciables (Wang, y Jain, 2003). Con esta información se quiere predecir algo que no ha sido observado, a partir de los datos de casos que sí han sido observados. De acuerdo con Hyndman y Athanasopoulos (2013), los modelos transversales, también llamados explicativos o causales, se usan cuando se percibe una relación entre la variable que se va a predecir con una o más variables predictoras. El objetivo es predecir la variable con base en una ponderación de sus predictores, asumiendo que las relaciones entre ellos permanecerán constantes. Muchas de las técnicas usadas en ciencia de datos y aprendizaje de máquina (*machine learning*),

tratan exclusivamente con modelos transversales, sin embargo, también se observan enfoques en series de tiempo (Ahmed, Atiya, Gayar, y El-Shishiny, 2010). La ecuación 1 presenta un ejemplo de un modelo simplificado de datos de corte transversal para determinar la demanda de un producto. El valor de cada coeficiente establece una relación entre la demanda del producto (predicción) y variables predictoras como el producto interno bruto (PIB), el mes de año y el crecimiento de la población.

$$\begin{aligned} & \text{Demanda producto} = \\ & 8.15 \text{ PIB} - 0.05 \text{ mes del año} + 0.15 \text{ crecimiento población, ... , error} \end{aligned} \quad 1$$

Por otra parte, en las series de tiempo se modela el comportamiento de una sola variable a través del tiempo. En el caso particular del problema en cuestión, la variable de interés es la demanda de uno o varios productos en el transcurso del tiempo, y este tiempo transcurre en periodos regulares (semanas, meses, etc.). El objetivo es pronosticar el comportamiento para la secuencia de las observaciones en el futuro. La ecuación 2 es una simplificación de un modelo de serie de tiempo para la demanda de un producto del periodo $t+1$.

$$DB_{t+1} = f(DB_t, DB_{t-1}, \dots, DB_0, error) \quad 2$$

También hay modelos que pueden combinar información tanto de corte transversal como de series de tiempo, estos se encuentran en la literatura como *pooling data*, *datos de panel* o longitudinales, cuyo uso es muy notable en econometría (Morales y Moura, 2013; Davies y Lahiri, 1995; Holtz-Eakin, Newey y Rosen, 1988).

3.3. Series de tiempo en lugar de modelos trasversales

Según Hyndman y Athanasopoulos (2013), un administrador, usuario final de la presente investigación, podría elegir modelos de pronósticos de series de tiempo sobre los transversales por varias razones. En primer lugar, porque pueden resultar más fáciles de entender y modelar, esto se debe a que establecer las relaciones entre las variables de un modelo trasversal puede ser complicado (economía nacional ~ demanda producto), porque los efectos de los predictores sobre la variable son difíciles de conseguir u observar, o porque la información de todas las variables no se encuentra disponible. En segundo lugar, porque en general se desea saber qué va a

sucedir en el futuro y no las razones de este comportamiento. Sin embargo, el trabajo desarrollado con modelos transversales puede ser de gran utilidad para cualquier organización, si se dispone de la información y si las relaciones entre la variable a predecir y los predictores están planteadas correctamente y no cambian en el futuro (Armstrong, 2001). En este sentido, el trabajo de investigación se inclina hacia los modelos de series de tiempo por las razones antes mencionadas, y la dificultad en obtener información de tipo transversal que será diferente para cada situación específica, en contraste con el objetivo generalidad que se quiere alcanzar con el método propuesto.

3.4. Componentes o patrones en series de tiempo

En las series de tiempo es posible identificar cuatro patrones: horizontal, tendencial, estacional y/o cíclico (Makridakis, Wheelwright & Cortés Fregoso, 2009). El factor horizontal, nivel o estadísticamente llamado estacionario, representa el promedio local alrededor del cual fluctúa la serie. Este patrón por lo general se observa en productos que tienen ventas estables o en un conjunto de datos tomado por cortos periodos de tiempo, así que un producto que a mediano plazo tiene tendencia puede ser estable en un periodo más corto de tiempo y en este caso se considera horizontal. La tendencia (T) es el cambio entre puntos sucesivos de la serie (pendiente) de forma creciente o decreciente. Las series que exhiben este comportamiento pueden ser datos de ventas de las organizaciones, sus precios y otros aspectos económicos de largo plazo (Makridakis, Wheelwright, y Hyndman, 2008). La estacionalidad (S), es representada por los picos o valles asociados a estaciones dentro del periodo de tiempo analizado, tienen duración y frecuencia aproximadamente fija. Estas series se asocian a factores como las estaciones climáticas y los periodos de pagos de empleados, por ejemplo, y por lo general se analizan en periodos menores a un año. El patrón cíclico (C) es similar al estacional debido a que la serie fluctúa mostrando picos y valles, sin embargo, su duración es mayor a un año y no es fija al igual que la frecuencia con que se repiten. Es importante resaltar que algunos métodos cuantitativos pueden tener cierta ventaja sobre otros dependiendo del patrón que

exhiban los datos. En la figura 1 se observa una representación gráfica de cada uno de los cuatro patrones revisados.

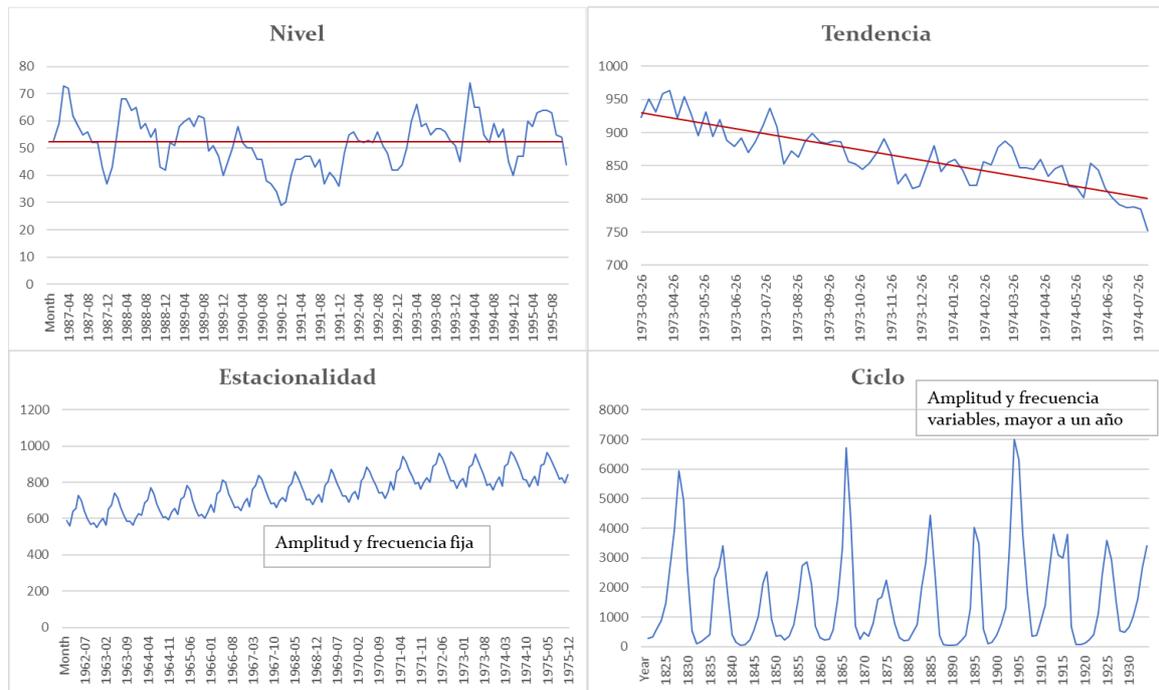


Figura 1. Representación gráfica de patrones. La figura muestra el comportamiento de una serie de datos con diferentes tipos de patrones. Elaboración propia.

Adicionalmente, se consideran dos posibles interacciones entre tres de estos patrones (S, T y C), una aditiva y una multiplicativa. Una interacción aditiva ($S+C+T+R$), considerando un factor de error o aleatorio R, se presenta cuando un cambio en el nivel de la serie no afecta al cambio en el componente estacional, es decir, este permanece aproximadamente constante a través del tiempo. En otras palabras, la diferencia entre dos meses de un año es aproximadamente la misma en el año siguiente y la amplitud del efecto estacional es el mismo (Figura 2).

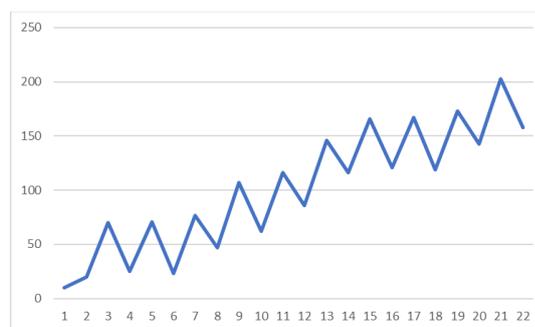


Figura 2. Estacionalidad aditiva. Los picos y valles no aumentan incluso cuando el nivel de la serie lo hace. Elaboración propia.

Por otro lado, un modelo multiplicativo ($S \times C \times T \times R$) se observa cuando, por efectos económicos, por ejemplo, existe un cambio en el nivel de cada año y también se presenta un aumento en la amplitud de los picos y valles de las estaciones, como en la figura 3. Este tipo de series es muy común en economía (Makridakis, Wheelwright, y Hyndman, 2008), porque por lo general los cambios en el nivel afectan de manera proporcional al componente estacional (Por ejemplo, marzo se lleva un 10% de las ventas y enero 5% en todos los años, en lugar de plantearse una diferencia constante se plantea un porcentaje). En términos generales los problemas que surgen en la cadena de suministro son modelados por una estacionalidad multiplicativa (Christou, 2011). Existen transformaciones matemáticas para establecer equivalencias entre los modelos aditivos y multiplicativos debido a sus propiedades logarítmicas, sin embargo, no son recomendadas para el usuario que no está familiarizado con estos tratamientos matemáticos (Hyndman y Athanasopoulos, 2013). Algunas de estas transformaciones hacen referencia a teorías del control en Ingeniería electrónica, como las funciones de transferencia de Laplace (Disney y Lambrecht, 2008), transformadas de Fourier (Hollingsworth y Lönnberg, 1986) y Box-Cox (Nelson y Granger, 1979), entre otras.

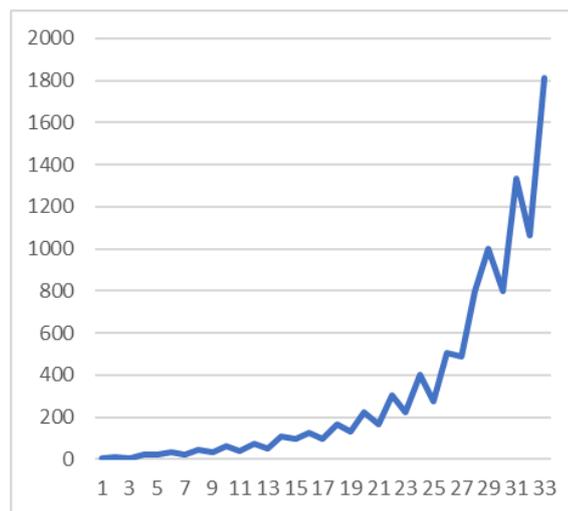


Figura 3. Estacionalidad multiplicativa. Las fluctuaciones aumentan en proporción al nivel.

3.5. Métodos Cuantitativos de Series de Tiempo

Para responder a la pregunta de investigación se revisan principalmente los modelos cuantitativos de series de tiempo **de complejidad media** que no signifiquen un deterioro en el rendimiento de los pronósticos. Esta revisión se orienta por el

principio de parsimonia y el marco referencial de Edmonds (1999), en el que el término complejidad se alude a diversas definiciones dependiendo del objetivo del autor, particularmente para este trabajo, la complejidad de los métodos se relaciona con la cantidad de variables a estimar y de información necesaria, respecto al total de métodos revisados. En la elección de este se considera también su generalidad o, dicho de otra forma, la amplia gama de patrones de datos a los que se aplique, todo esto desde una perspectiva pragmática.

A continuación, se presenta una tabla resumen (tabla 1) de los métodos cuantitativos de series de tiempo revisados, junto con las observaciones de interés para el presente trabajo. La descripción detallada de algunos de los métodos se encuentra en el anexo A.

Tabla 1 Resumen de métodos de pronósticos cuantitativos de series de tiempo de complejidad media

| Métodos | Ecuación | Parámetros | Ventajas | Desventajas | Patrón que se ajusta |
|---------------------------|--|---------------------|--|---|--|
| Ingenuo (Naïve) | $y_t = y_{t-1}$ | Ninguno | Método simple, punto de referencia | Las condiciones para que resulten eficientes son difícilmente posibles | Nivel altamente estable, sin fluctuaciones |
| Desviación (Drift) | $y_T + \frac{h}{T-1} \sum_{t=2}^T (y_t - y_{t-1})$ $= y_T + h \left(\frac{y_T - y_1}{T-1} \right)$ | Ninguno | Método simple, punto de referencia | Las condiciones para que resulten eficientes son difícilmente posibles | Tendencia estable |
| Regresión simple | $y_t = \beta_0 + \beta_1 x + e$ | $\beta_0 + \beta_1$ | Se obtiene un modelo aproximado del comportamiento de los datos de acuerdo con las relaciones de x e y | Más útiles con datos transversales, presenta inconvenientes con series de tiempo | Nivel o tendencia estable, comportamiento lineal |
| Promedio móvil | $y_n = \frac{\sum_{i=1}^n x_i}{n}$ | Ninguno | Método simple, punto de referencia | A mayor n, mayor será el número de valores que se tienen que almacenar y toda la información, incluso la más antigua, tiene el mismo peso en la generación del pronóstico | Nivel o tendencia estable |

| | | | | | |
|--|--|---|---|--|---|
| Suavización exponencial Simple (SES) | $\hat{y}_{t+h t} = \alpha y_t + (1 - \alpha)\hat{y}_{t-1}$ | α , patrón de suavizado del nivel Fo, valor de inicialización | Simple, prioridad a datos más recientes, y no necesita almacenar mucha información | Elección correcta de α , no es útil para tendencia o estacionalidad | Solo series con patrón horizontal o de nivel |
| Suavización exponencial doble o de Holt | $\begin{aligned}\hat{y}_{t+h t} &= l_t + hb_t \\ l_t &= \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1}) \\ b_t &= \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1}\end{aligned}$ | α , β de la tendencia lo y bo, valores de inicio | Prioridad a datos recientes, extensión de SES usa dos filtros para obtener estimaciones del factor horizontal y de la tendencia lineal. Muy frecuente en la literatura. | Elección correcta de α y β , no es útil para estacionalidad. | Patrones de nivel y tendencia lineal o aditiva |
| Suavización doble con tendencia exponencial | $\begin{aligned}\hat{y}_{t+h t} &= l_t b_t^h \\ l_t &= \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1}) \\ b_t &= \beta \frac{l_t}{l_{t-1}} + (1 - \beta)b_{t-1}\end{aligned}$ | α , β de la tendencia lo y bo, valores de inicio | Útil en situaciones donde la tasa de crecimiento es constante y no su pendiente. | Elección correcta de α y β , no es útil para estacionalidad. No es frecuente que una tendencia crezca exponencialmente (Hyndman y Athanasopoulos, 2013) | Patrones de nivel y tendencia exponencial o multiplicativa (Hyndman et al., 2002) |
| Suavización doble con tendencia | $\begin{aligned}\hat{y}_{t+h t} &= l_t + (\phi + \phi^2 + \dots + \phi^h)b_t \\ l_t &= \alpha y_t + (1 - \alpha)(l_{t-1} + \phi b_{t-1}) \\ b_t &= \beta(l_t - l_{t-1}) + (1 - \beta)\phi b_{t-1}.\end{aligned}$ | α , β , ϕ , nivel, tendencia y amortiguación, | Amortigua gradualmente la tendencia hasta llegar | Elección correcta de α , β y ϕ . Mayor complejidad, no útil para | Patrones de nivel y tendencia amortiguada aditiva |

| | | | | | |
|---|--|---|---|--|--|
| amortiguada aditiva | | respectivamente. lo y bo, valores de inicio | a una asíntota, imitando el comportamiento del ciclo de vida de algunos productos. | estacionalidad. | |
| Suavización doble con tendencia amortiguada multiplicativa | $\hat{y}_{t+h t} = \ell_t b_t^{(\phi + \phi^2 + \dots + \phi^h)}$ $\ell_t = \alpha y_t + (1 - \alpha) \ell_{t-1} b_{t-1}^\phi$ $b_t = \beta \frac{\ell_t}{\ell_{t-1}} + (1 - \beta) b_{t-1}^\phi$ | α, β, ϕ , nivel, tendencia y amortiguación, respectivamente. lo y bo, valores de inicio | Amortigua gradualmente la tendencia hasta llegar a una asíntota. Han demostrado buenos resultados (Fildes, 2001; Gardner & McKenzie, 1988; Taylor & Bunn, 1999) | Elección correcta de α, β y ϕ . No muy usado en gestión de operaciones, adición de otro parámetro, tiempo de cómputo alto para hallar intervalos de predicción, no útil para estacionalidad. | Patrones de nivel y tendencia amortiguada multiplicativa |
| Suavización Exponencial triple o Holt-Winters aditiva | $\hat{y}_{t+h t} = \ell_t + hb_t + s_{t-m+h}^*$ $\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m}$ | α, β, γ , nivel, tendencia y estacionalidad, respectivamente. m, periodo de estacionalidad lo y bo, valores de inicio | Estima los componentes para los tres patrones presentes en un año. | Elección correcta de α, β y γ , adición de otro parámetro. Complejidad mayor debido al número de parámetros a estimar. Modelo aditivo no muy frecuente (Christou, 2011). | Patrones de nivel tendencia y estacionalidad aditiva, es decir cuando las fluctuaciones de las estaciones no se ven afectadas por el nivel |

| | | | | | |
|--|---|--|---|--|---|
| <p>Suavización Exponencial triple o Holt-Winters multiplicativa</p> | $\hat{y}_{t+h t} = (\ell_t + hb_t)s_{t-m+h_m^*}$ $\ell_t = \alpha \frac{y_t}{s_{t-m}} + (1 - \alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1}$ $s_t = \gamma \frac{y_t}{(\ell_{t-1} + b_{t-1})} + (1 - \gamma)s_{t-m}$ | <p>α, β, γ, nivel, tendencia y estacionalidad, respectivamente. m, periodo de estacionalidad ℓ y b, valores de inicio</p> | <p>Estima los componentes para los tres patrones presentes en un año. Uso frecuente en la cadena de suministros (Christou, 2011).</p> | <p>Elección correcta de α, β y γ, adición de otro parámetro. Complejidad mayor debido al número de parámetros a estimar.</p> | <p>Patrones de nivel tendencia y estacionalidad multiplicativa, es decir cuando las fluctuaciones de las estaciones se ven afectadas por el nivel</p> |
| <p>Descomposición de componentes</p> | <p>Modelo no paramétrico</p> | <p>m, periodo de estacionalidad</p> | <p>No paramétrico, fácil de comprender, altamente usado y con este se puede “remover” el patrón estacional de una serie.</p> | <p>Se asume que el componente estacional se repite en cada año. Se “pierden” algunos valores estimando la tendencia. No recomendado para series erráticas.</p> | <p>Patrones de estacionalidad, la tendencia se debe estimar con un método paramétrico como suavización de Holt o regresión.</p> |

En la tabla 1 se observan, en primer lugar, los métodos básicos (*ingenuo y drift*), los cuales, en lugar de usarse para pronósticos y toma de decisiones, son considerados como un punto de referencia o comparación frente a otros, porque las condiciones para que resulten eficientes son difícilmente posibles. Luego, se encuentra la regresión simple, la cual si bien puede usarse en modelos de series de tiempo presenta mayor rendimiento con datos de corte transversal, además solo engloba patrones de tipo lineal, ya sea estacionario o tendencial, por lo que otros métodos de regresión ya sea múltiple o dinámica, son usados para modelos causales, y no de series de tiempo. En un tercer fragmento de la tabla se encuentran los métodos de suavización. Partiendo del promedio móvil, en donde los datos son suavizados a través de su media, y continuando con los de suavización exponencial, los cuales atenúan las series de datos de acuerdo con los parámetros α , β y γ , para estimar el nivel, la tendencia y la estacionalidad, respectivamente. Adicionalmente, se muestran variaciones de los métodos de suavización para ajustarse a otro tipo de tendencias. Así, la tendencia exponencial o multiplicativa, por ejemplo, pretende adaptarse a las situaciones en donde la tendencia crece progresivamente. Sin embargo, en Gardner (2006), se observa que, de las 65 publicaciones exhibidas con estudios empíricos, ninguna aborda este tipo de modelo, en contraste con la tendencia aditiva o lineal, del enfoque original de Holt, cuya presencia en dicha revisión se observa con mayor frecuencia. Por su parte, los modelos de tendencia amortiguada han demostrado buenos resultados (Fildes, 2001; Gardner y McKenzie, 1988; Taylor y Bunn, 1999), pero su uso en la gestión de operaciones de la cadena de suministro no ha sido tan amplio (Gardner, 2006) de forma que garantice que la adición de un parámetro más pueda ser beneficioso. Finalmente, el método de descomposición se usa principalmente para “desestacionalizar” una serie de datos separando sus patrones y obtiene los pronósticos a través de una combinación con otros métodos paramétricos de estimación de la tendencia como regresión simple, o suavización exponencial de Holt. Los métodos de Box-Jenkins también son considerados métodos convencionales, sin embargo, debido a la cantidad de parámetros a estimar y a que su uso se presenta mayormente en economía y finanzas (Ariyo, Adewumi, & Ayo, 2014) no fueron considerados en la revisión.

3.6. Estimación de parámetros

3.6.1. Constantes de suavización

Un factor común en la mayoría de los métodos revisados es la necesidad de estimación de parámetros, lo cual también supone una preocupación para el usuario. Específicamente, para los métodos de suavizamiento se deben estimar aquellos asociados a nivel, tendencia y estacionalidad, y esta elección afecta el rendimiento de los pronósticos (McCracken, 2004). Inicialmente estos parámetros eran elegidos arbitrariamente o través de búsquedas en malla o *grid search* (Gardner, 1985). En la actualidad se tiene una amplia disponibilidad de optimizadores como *Open solver* o el *Solver* de Microsoft Excel® (Gardner, 2006), el cual cuenta con popularidad porque provee soluciones rápidas para problemas de pronósticos frecuentes. Sin embargo, autores como Ravinder (2013) advierten que un uso inexperto de esta herramienta puede llevar a resultados erróneos y aconsejan la configuración de múltiples puntos de partida para mejorar los resultados, aunque esto puede extender el tiempo de cómputo del proceso. Igualmente, el valor fijo de las constantes proporcionado por el optimizador resta alguna libertad en la configuración del método, que podría contar con mayor flexibilidad si se tratara de un rango de valores y no de uno puntual. En la literatura relativa a la determinación de constantes de suavización en pronósticos se observan diversos aportes. Concretamente, Rasmussen (2004), minimiza la medida de error de pronóstico, el error cuadrático medio (*MSE*, por sus siglas en inglés), para encontrar los valores óptimos de las tres constantes de suavización y de los valores de inicio, mostrando mejoras en el método convencional y resaltando la facilidad con que los educadores pueden hacer uso de esta técnica. Asimismo, Quintana y Leung (2007), presentan un modelo orientado a demandas amortiguadas en la cadena de suministro a través de suavización exponencial en la que por medio de un filtro de *Kalman* se estiman las constantes actualizadas usando como fuentes de entrada el inventario en el tiempo de reposición y la demanda real, en donde la medida de rendimiento es también el *MSE*. También se observan otros enfoques como el de Tratar (2010), el cual realiza una optimización conjunta del pronóstico y el control de inventarios, para ello se minimizan los costos de la cadena de suministro y se optimizan las constantes de

suavización y los valores de inicio. También, Kristianto, Helo, Jiao y Sandhu (2012), proponen un control difuso adaptativo, en donde la demanda se actualiza considerando los cambios que experimenta en un periodo de tiempo y el inventario disponible, entre otros factores, estos sirven como elementos de entrada en la estimación de la constante del modelo de suavización exponencial simple. Igualmente, algunos autores ofrecen sugerencias en cuanto a los rangos adecuados para dichas constantes, Schroeder, Rungtusanatham, y Goldstein (2013) y Chase y Jacobs (2013), cómo se citó en Ravinder, 2013, sugieren valores de α entre 0.1 y 0.3, otros autores como Heizer y Render (2011) presentan un rango de 0.05 a 0.5, de cualquier forma, se ha determinado que los valores apropiados varían de acuerdo a la información que se está analizando, por tanto, contar con un solo rango puede resultar en una mala práctica.

En la consecución de dichas constantes es reiterativo el uso de medidas de rendimiento de los pronósticos como la *desviación absoluta media* o *MAD* y *MSE*, las cuales, ilustran sobre la diferencia entre lo pronosticado y lo sucedido en la realidad y son de utilidad en la comparación de métodos y en su optimización. No obstante, no todas las medidas son comparables para cualquier serie de tiempo y deben ser usadas bajo el conocimiento de la naturaleza de la información, dado que sus resultados pueden variar de acuerdo con la escala o tamaño de los datos, arrojando un resultado sesgado (Shcherbakov et al., 2013).

3.6.2. Otros enfoques de la estimación

La estimación de parámetros no es exclusiva de los pronósticos, es una necesidad común en cualquier modelo de ingeniería y matemáticas porque entre otros beneficios, ofrece herramientas necesarias para la gestión de la información (Beck & Arnold, 1977) y para su predicción en un problema conocido como problema inverso (Tarantola, 2013). La literatura brinda algunas soluciones al respecto desde áreas como estadística, ingeniería de control, ingeniería eléctrica y visión por computador, a través de técnicas tales como máxima verosimilitud, mínimos cuadrados y simulación de Monte Carlo, cuyos beneficios podrían replicarse en la estimación de parámetros para los métodos de demanda agregada. Particularmente, Bahl, Brown, De Souza y Mercer

(1986) estima los parámetros de un modelo de reconocimiento de voz a través de un método llamado maximización de mutua información el cual mide la dependencia mutua entre dos variables. Este método es comparado con el de máxima verosimilitud, de uso frecuente en estadística, obteniendo menores tiempos de cómputo y errores de reconocimiento, con un grado avanzado de rigurosidad estadística. Por otro lado, Moradkhani, Sorooshian, Gupta y Houser (2005), presentan un enfoque de estimación de parámetros para un modelo hidrológico a través de filtro de *Kalman* para conjuntos, ó *Ensemble Kalman Filter*, que, grosso modo, le permite estimar un par de parámetros de manera simultánea a través de la probabilidad de datos previamente obtenidos. Este método es una extensión del filtro de Kalman para una cantidad abundante de datos y se usa especialmente en las geociencias. En el área de visión por computador los autores Wachs, Stern y Edan (2005), proponen un algoritmo de agrupación de los datos para ajustar o sintonizar (*tunning*) los parámetros de un proceso de reconocimiento de gestos, de forma que esta configuración sea determinada de acuerdo con las similitudes existentes entre los grupos. La propuesta partió de un concepto de agrupamiento simple y presentó resultados de tiempo y rendimiento positivos en comparación con otros aportes académicos mencionados. Por otro lado, Zhang (1997), presenta varios enfoques en la estimación de parámetros para ajustarse a la ecuación cónica, problema de gran aplicación en cartografía y análisis de imágenes, en los que incluye técnicas como mínimos cuadrados, filtro de *Kalman*, regresiones y en particular la transformada de *Hough*, un agrupamiento que sirve para detectar ciertas figuras en las imágenes a través de grupos formados por características similares. El análisis de los datos por grupos o *clustering* es el “arte de encontrar grupos en los datos”, en este se pretende agrupar en k clases o grupos, las diferentes observaciones de acuerdo con su similitud, definida por uno o varios criterios (Rousseeuw y Kauffman, 1990). Este método ha sido usado principalmente en análisis de datos, estadística, aplicaciones de clasificación matemática, y en los últimos años en aprendizaje de máquina (*machine learning*) y *big data*, por su capacidad de manejar grandes cantidades de información (Hans-Hermann, 2008). Dicho método ha cobrado gran importancia por tratarse de una técnica de aprendizaje no supervisado que permite obtener resultados cuando no se posee a priori información sobre cómo

pueden darse las agrupaciones y es considerada de gran utilidad para la identificación de patrones dentro de un conjunto de datos (Aghabozorgi, Shirkorshidi y Wah, 2015). Esta tarea se lleva a cabo por medio de algunos algoritmos específicos como *k-means*, un algoritmo de fácil comprensión que resulta popular debido a la poca carga computacional que requiere su ejecución en contraste con su alta capacidad de almacenamiento de datos (Bradley, Fayyad y Reina, 1998; Aghabozorgi et al., 2015). Una búsqueda de literatura orientada a encontrar publicaciones respecto al uso de *clustering* para la estimación de parámetros en métodos de pronósticos no obtuvo ningún resultado similar. En dicha revisión los resultados encontrados se referían a algoritmos que fueron usados exclusivamente para clasificar la información, lo que podría indicar una oportunidad de investigación aprovechando las propiedades de esta metodología y las necesidades identificadas en la implementación de métodos de pronósticos para usuarios de negocios.

3.7. Hallazgos de la revisión de la literatura

A partir de la revisión de métodos cuantitativos para los pronósticos de demanda, cabe resaltar la alta implementación de los métodos de suavizamiento exponencial. Dentro del conocimiento de la autora del presente texto, estos métodos de complejidad media en la literatura han obtenido resultados comparables con modelos más sofisticados (Green, y Armstrong, 2015), han sido ampliamente ejecutados en la práctica (Lewis-Beck, 2005) y pueden ser implementados, de forma paramétrica o no, en los diferentes patrones de la demanda. En particular, se resalta el método de tendencia de *Holt* o doble suavizado, debido a la simplicidad de su formulación, apelando al principio de parsimonia y a su rendimiento, especialmente si se trata con series ajustadas estacionalmente. En este sentido, la autora opta por realizar una descomposición multiplicativa de componentes para obtener una serie ajustada estacionalmente que pueda ser usada por el método de *Holt*. Sin embargo, con el ánimo de cumplir con las características previamente enunciadas en el planteamiento del problema y a partir de la literatura encontrada respecto a la estimación de parámetros, se observa una posible oportunidad de investigación en el uso de algoritmos de agrupamiento (*clustering*) que apoyen la identificación de los

parámetros adecuados de α y β , propios del modelo de *Holt*, con el objetivo de ofrecer un valor que sirva de guía para el uso de dichos métodos y simultáneamente posea soporte teórico robusto.

4. Metodología

Con el fin de responder a las preguntas de investigación, y de acuerdo con lo ilustrado en el marco teórico y revisión de literatura, se adoptó la siguiente metodología:

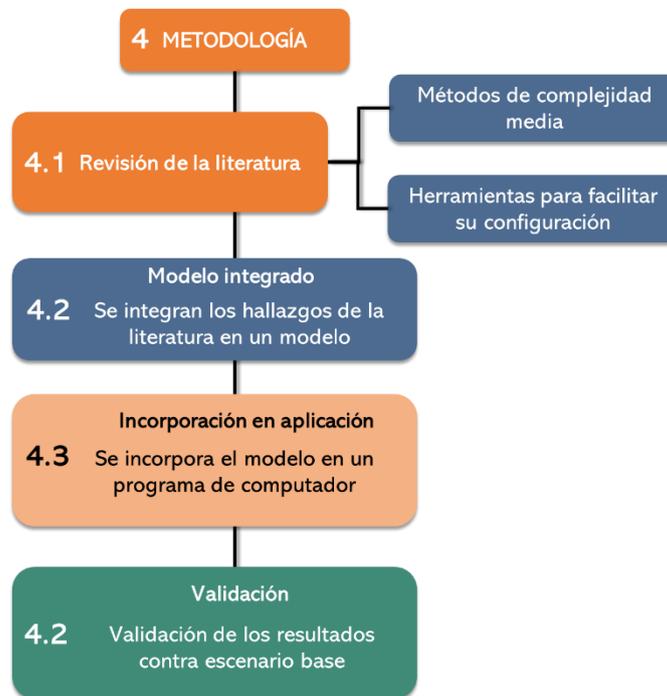


Figura 4. Metodología de la investigación. Se muestran las fases para llevar a cabo la investigación. Elaboración Propia.

Esta metodología permitió cumplir con los objetivos planteados para el proyecto de investigación.

4.1. Revisión de la literatura

Esta etapa es considerada como la base de toda investigación, y en este caso hace parte tanto del marco teórico como de la metodología. Tomando en cuenta lo enunciado en los hallazgos del marco teórico y la revisión de literatura, se planteó un modelo integrado descomposición-*Holt*, en donde se hallan rangos de valores para los parámetros α y β de acuerdo con una caracterización de las series basada en un algoritmo de *clustering*. A partir de esta propuesta, se revisó la literatura para identificar estudios similares de los cuales se pudieran extraer fortalezas y oportunidades de investigación. Se revisaron dos bases de datos académicas, *Google Scholar* y *Scopus*[®], los términos para la búsqueda fueron *clustering* y *exponential*

smoothing, en título, resumen o palabras clave, y se encontraron 36 contribuciones en las cuales no se encontró evidencia de un enfoque similar para estimar las constantes de suavización usando algoritmos de *clustering*. Sin embargo, se apreciaron otras posiciones en cuanto al uso combinado de estas dos metodologías. Joshi, Lingras, Wani, & Zhang (2013) por ejemplo, proponen agrupar las series de tiempo de demanda para luego aplicar un método de pronósticos apropiado para cada clúster, los métodos de pronósticos comparados incluyen suavización exponencial simple (SES), suavización de *Holt*, y ARIMA y los de *clustering* comprenden *k-means* y esperanza-maximización. Este estudio se realiza específicamente para políticas de inventario y considera el problema como uno específico de cada situación y no de manera generalizada.

Por otro lado, Aldhyani y Joshi (2017), usaron la información resultante de aplicar suavización exponencial simple a los centroides de un algoritmo de *clustering*, para pronosticar el tráfico de redes de telecomunicación. Otros autores como Laouafi, Mordjaoui, Laouafi, & Boukelia (2016), usan *Holt-Winters-Taylor*, un enfoque de pronósticos que incorpora en su modelación dos estacionalidades, para hacer un primer pronóstico. Luego se usan una variedad de técnicas de clúster para clasificar la información, y por último los resultados de todos estos pronósticos se concentran en un resultado global de los picos de energía eléctrica.

4.2. Modelo integrado

La segunda etapa comprende la integración de los métodos seleccionados y el agrupamiento para obtener los parámetros guía, lo cual se convierte en un modelo cuyos lineamientos se pueden asociar a las técnicas de aprendizaje estadístico. Estas técnicas, dentro de las cuales se incluye *clustering*, comprenden por lo general dos etapas: Entrenamiento y clasificación.

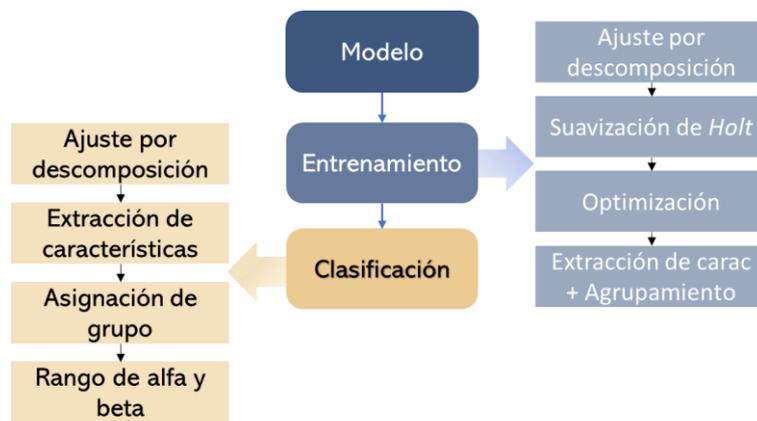


Figura 5. Modelo integrado. Muestran las dos etapas del modelo entrenamiento y la clasificación, así como sus sub-etapas. Elaboración propia.

En la primera, a partir de una base de datos se asocian las características de las series que corresponden a cada grupo. En la segunda, se toman datos nuevos o que no fueron usados en la etapa anterior, y según sus características se asignan a un grupo específico. El modelo integrado, detallado en los siguientes numerales, se presenta en la figura 5. En la etapa de entrenamiento se usó el 80% de las series de la base de datos de *NN₃ Competition* (Crone, Hibon y Nikolopoulos, 2011), de acuerdo con los lineamientos de Hyndman y Athanasopoulos (2013), equivalente a 89 columnas o series de tiempo, y el resto se usó en la etapa de clasificación o pruebas.

Resumidamente, estas series son ajustadas estacionalmente a través de descomposición, para luego usar el método lineal de *Holt*, cuyas constantes de suavización se obtienen a través de optimización. Para estas series se extraen características estructurales o descriptores como *media*, *mediana*, *desviación estándar* (*desvest*), *coeficiente de variación* (CV), *entropía* y *asimetría*, las cuales apoyan el proceso de agrupamiento encontrando similitudes entre ellas (Anexo B). Como resultado del entrenamiento se obtiene un número predeterminado de grupos de series, el cual es asignado a criterio del usuario. Dentro de cada grupo se encuentran series consideradas similares y los valores óptimos correspondientes encontrados en el pronóstico. La etapa de clasificación consiste en tomar una serie nueva, identificar el grupo al cual pertenece y sugerir con esto un rango para las constantes de suavización. En los apartados 4.2.1 a 4.2.4., se describe con mayor detalle el contenido del modelo.

4.2.1. Etapa de Entrenamiento

4.2.1.1. Ajuste estacional a través de descomposición de componentes

Se sigue un procedimiento de descomposición clásica de componentes cuya técnica es posible observar en la figura 6, y con mayor detalle en el anexo A. En el ajuste estacional de la serie, se toman 12 meses como periodos estacionales bajo el supuesto de que cualquier organización se ve afectada por factores que cambian cada año. Igualmente, se supone para este trabajo que a medida que aumenta la tendencia de la demanda, la magnitud de los picos y valles de la estacionalidad también aumenta, en otras palabras, se consideró una estacionalidad multiplicativa, hecho denotado por la representación $TxCxSxR$, siendo cada uno de estos los componentes de tendencia, ciclo, estacionalidad y aleatoriedad, respectivamente.

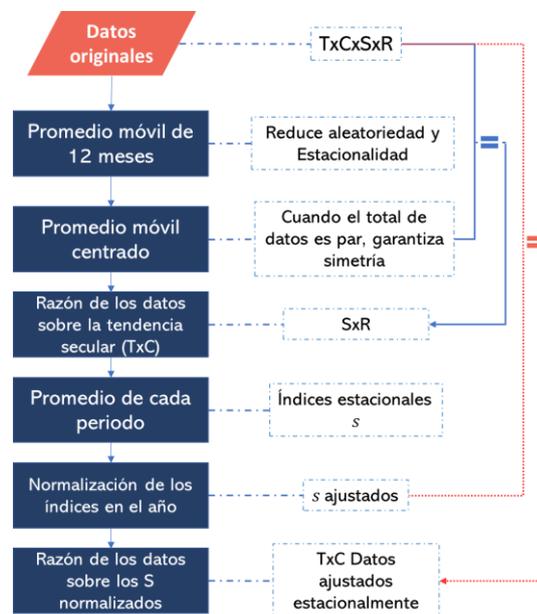


Figura 6. Ajuste estacional. Procedimiento para ajuste estacional a través de descomposición. Elaboración propia.

En la figura 6, también se describen los pasos de la descomposición y la salida o el resultado de cada uno, representado por los recuadros en guiones. En primer lugar, se realiza un promedio móvil de 12 meses con el objetivo de reducir estacionalidad y aleatoriedad al tiempo que se aísla la tendencia-ciclo o tendencia secular. En la descomposición se separan los patrones de tendencia y ciclo como conjunto, ya que, en muchos casos, no es de interés conocer el movimiento cíclico de manera individual

ya que este se orienta a decisiones con un horizonte de tiempo de largo plazo, como las de tipo estratégico y económico, y adicionalmente se requiere de series de tiempo prolongadas para poder separarlo. Seguidamente, es necesario realizar un nuevo promedio móvil centrado de solo dos periodos en caso de que el periodo estacional sea par, esto para garantizar que el valor refleje de manera simétrica el verdadero promedio de la serie. Después de este ajuste, se separan los componentes de estacionalidad y aleatoriedad (SxR), realizando una división de los datos originales o “sin tratar” (TxCxSxR) sobre el promedio móvil centrado (TxCxSxR/TxC). Para obtener los índices estacionales, se promedian los valores de cada estación, esto es, *media* de los meses enero, *media* de los meses febrero, y así sucesivamente, suponiendo que cada mes del año es una estación. Estos índices deben ser normalizados, por lo que son ajustados de manera que la suma de 12 índices sea igual a 12 (número de estaciones dentro de un año). Luego, debido a que se trata de un modelo multiplicativo, se dividen los datos originales y los índices estacionales normalizados obteniendo una serie “sin componente estacional”. Como resultado de aplicar el modelo de descomposición se obtiene una serie desestacionalizada, es decir, con patrones de nivel, tendencia-ciclo y aleatoriedad, y de manera independiente los índices estacionales que se usarán para “revertir” el proceso de descomposición y devolver la estacionalidad a la serie.

4.2.1.2. Suavizamiento de Holt

Para aplicar el suavizamiento lineal, se usa la información resultante del método de descomposición, que por principio se toma como no estacional. Para ello se usan las ecuaciones de la tabla 1, las cuales requieren de parámetros de inicialización de nivel (l_0) y tendencia (b_0), y de suavizamiento α y β . En este caso, se sigue a Hyndman y Athanasopoulos (2013), en donde los valores iniciales serán como se describen en las ecuaciones 3 y 4:

$$l_0 = y_1 \quad 3$$

$$b_0 = y_2 - y_1 \quad 4$$

Posteriormente, se obtienen los pronósticos de un periodo hacia adelante para valores dentro de la muestra (*in-sample*), y se calcula el error del pronóstico para cada periodo. Este procedimiento se presenta en la figura 7.

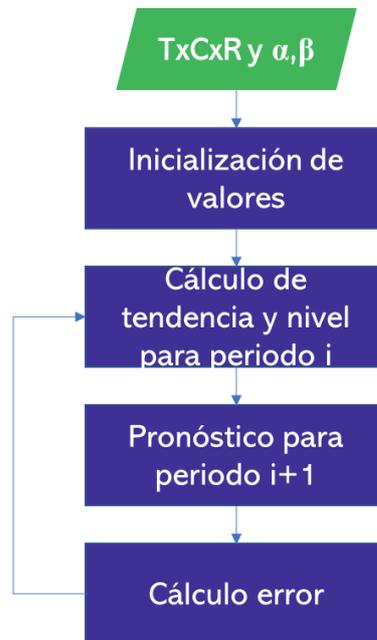


Figura 7. Suavizamiento de Holt. Pasos generales para el suavizamiento de Holt. Elaboración propia.

Sin embargo, el uso del modelo de *Holt* implica la elección apropiada de parámetros α y β , que en general está motivada por una técnica de optimización que obtenga aquellos valores que minimicen una medida global de error del pronóstico. En este caso, dicha medida será el error porcentual medio absoluto (MAPE), que al tratarse de una proporción permite comparar errores de series con distinta escala o tamaño, a diferencia de medidas como MAD o MSE, las cuales dependen del tamaño de los datos (Shcherbakov et al., 2013). Se encuentran desventajas con medidas como MAPE, al tratarse con demandas intermitentes en los que se presentan valores de cero, lo que matemáticamente resultaría en valores tendientes al infinito, sin embargo, la base de datos no posee esta clase de estructura. De cualquier modo, es posible usar otra medida de error, siempre que esta sea consistente a lo largo del trabajo y útil para comparar el rendimiento del modelo.

4.2.1.3. Optimización

El problema de optimización consiste en encontrar los α y β que minimicen la medida de error MAPE, dentro de los rangos establecidos para estos parámetros (0-1). Para resolverlo se usó un método de mínimos cuadrados, una técnica convencional que se usa con frecuencia para la estimación, minimizando la suma de los errores cuadrados. En este caso, se minimiza el MAPE obtenido del 20% de los últimos datos de cada serie. El código relacionado con esta técnica se escribió en el lenguaje *Python*, utilizando el paquete *Scipy* (`scipy.optimize.least_squares`), y un algoritmo de mínimos cuadrados por región de confianza. Los algoritmos de optimización por región de confianza presentan beneficios en comparación con los de búsqueda lineal, ya que por su naturaleza de cotas pueden usar modelos de aproximación no convexos y aun así presentar una solución “global”. Asimismo, son confiables y robustos, y tienen buenas propiedades de convergencia a partir de cualquier punto inicial (Yuan, 2000). Particularmente, el algoritmo *Trust Region Reflective*, o TRF, permite resolver problemas con fronteras (*bounds*) como en el escenario actual de los rangos de las constantes de suavización y su mayor beneficio se observa en los problemas de alta dimensión, en comparación con el método *dogbox*, cuya utilidad es exclusiva para problemas más pequeños (Scipy.org., 2001-). Igualmente, para complementar la solución al problema relacionado con el punto inicial elegido, se realiza una búsqueda de rejilla (*grid search*) plantando varias semillas, en un esquema similar a la figura 8:

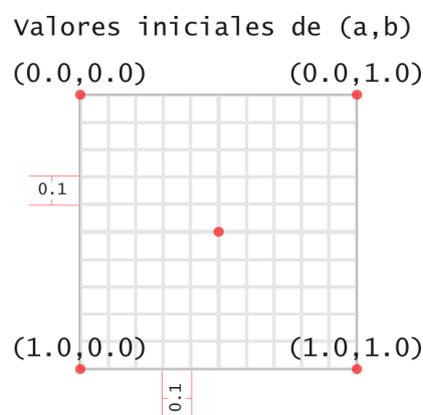


Figura 8. Búsqueda en malla. Se ejemplifica el avance de la búsqueda en la malla y la ubicación de las diferentes semillas. Elaboración propia.

Esto implica que el algoritmo trf se ejecuta 100 veces por cada serie y en cada ejecución se encuentran los valores óptimos de α y β de la sección de la malla correspondiente y se compara respecto al resultado de las demás secciones con el objetivo de seleccionar la de menor error. El uso de la división del área en 100 celdas permite al algoritmo realizar una búsqueda exhaustiva, ya que heurísticamente, se observó que la inclusión de estas semillas disminuyó el error respecto a una única inicialización al azar. No obstante, se incrementó también el tiempo de búsqueda de minutos a horas, por lo que se requiere de futuras investigaciones que permitan determinar el número óptimo de divisiones respecto al tiempo de cómputo que estas implican y que se considere adecuado para una organización. El incremento en el tiempo de búsqueda no resulta un problema a largo o mediano plazo, porque es un procedimiento que se realiza una única vez, durante la etapa de entrenamiento, y que se realizará de nuevo solo si el usuario considera que el algoritmo debe ser entrenado con nuevas series. Una vez se tienen los resultados de todas las series se almacenan sólo las constantes que presentaron el menor error. Este algoritmo es iterativo, es decir, va avanzando por pasos de 0.005 y se seguirá ejecutando hasta que se cumpla un criterio de parada, que puede ser un valor específico o hasta que no se produzca un cambio significativo en la respuesta óptima. Debido a que los rangos sugeridos en la literatura fueron del orden de las décimas (por ejemplo, 0.3-0.4 (Ravinder, 2013)), se eligió un paso del orden de las centésimas con el objetivo de encontrar un valor más preciso durante la búsqueda. Para el caso particular el algoritmo finaliza una vez se cumpla uno de los siguientes criterios de parada:

1. Si después de dos iteraciones el cambio entre las variables independientes es de menos de $1e-8$.
2. Si después de dos iteraciones la norma del gradiente, encargado de calcular la dirección de búsqueda, presenta cambios de menos de $1e-8$.
3. Si después de dos iteraciones el cambio en la función de salida, el error, presenta cambios de menos de $1e-8$.

Como resultado de este aparte entonces, se obtienen los α y β óptimos de cada serie dentro del conjunto de entrenamiento, las cuales serán almacenadas para ser usadas en la definición de los rangos sugeridos en la etapa de prueba.

4.2.1.4. Extracción de características y clustering o agrupamiento

Para realizar el agrupamiento se extrajeron operaciones o momentos estadísticos de primer orden que abarcan características subyacentes de la serie como *media*, *mediana*, *desvest*, *CV*, y *asimetría*, así como medidas como *entropía*, la cual es de utilidad en la teoría de la información (Ver Anexo B). Algunas de estas medidas han sido usadas ya como descriptores de la información en algoritmos de redes neuronales para series de tiempo y se han obtenido resultados positivos en cuanto facilidad de cómputo, insensibilidad a transformaciones y rendimiento en agrupación, en contraste con otras características de mayor complejidad y menos robustas (Nanopoulos, Alcock y Manolopoulos, 2001). En el enfoque de agrupamiento estas medidas pretenden capturar rasgos globales de los datos para establecer relaciones de similitud entre las series, con la intención adicional de hacer el método menos sensible a datos faltantes o ruido (Wang, Smith & Hyndman, 2006). Sin embargo, estas pueden ser sensibles a la escala o tamaño de los datos, es decir, aunque dos series posean comportamientos similares pueden terminar en diferentes grupos, porque las características como *media* o *mediana*, muy disimiles, así lo determinaron. Frente a esto, en lugar de hacerse de manera total, las medidas se extraen por ventanas de tiempo, funciones de ventana, o ventana deslizante (*sliding window*), en donde cada serie se divide por el mismo número de ventanas, en este caso seis. Entonces, para cada serie se extrajeron las seis características estructurales antes mencionadas por cada ventana de tiempo, resultando en un total de 36 características a comparar para formar los grupos. En el agrupamiento de series de tiempo hay dos principales categorías, “*whole clustering*” y “*subsequence clustering*” (Aghabozorgi et al., 2015). En la primera se ejecuta el algoritmo en distintas series de tiempo individuales y se agrupan en diferentes clases, mientras que la última extrae las características o descriptores por una ventana que se desliza. El objetivo es encontrar similitudes o diferencias entre una misma serie de tiempo (Keogh y Lin, 2005). El enfoque del presente texto es una mezcla entre las dos

categorías dado que se busca asociar diferentes series, sin embargo, la extracción de los descriptores se realiza por cada ventana de tiempo, por la razón antes mencionada.

El agrupamiento a través de algoritmos de partición como *k-means* es uno de los más utilizados y esto puede deberse a la poca carga computacional que requiere su ejecución (Bradley et al., 1998), y a la facilidad de comprender su funcionamiento. Por tal razón se tomó como un método adecuado para el presente enfoque, sin embargo, para mejorar su rendimiento se implementó también el algoritmo de *spectral clustering*. Este algoritmo aprovecha las funcionalidades de *k-means*, presentando mejores resultados que los obtenidos únicamente con este último, y además es simple y eficiente en tiempo de cómputo (Von Luxburg, 2007; Zelnik-Manor y Perona, 2005).

En resumen, se extraen las características (operaciones estadísticas de primer orden) por ventana de tiempo en cada serie, se comparan entre series y se forman grupos, estos grupos contienen series que comparten similitudes, así como los valores correspondientes de α y β obtenidos en el proceso de optimización anterior. La elección del número de grupos en los que se dividirá todo el conjunto de datos es una particular para cada caso de estudio. Algunos autores han cuestionado este enfoque, sin embargo, la velocidad de ejecución del método sigue siendo una ventaja en cuanto a que se pueden comparar los resultados usando diferentes valores para los k grupos (Aghabozorgi et al., 2015). Para el presente trabajo, las pruebas se efectúan con ocho clases o grupos por defecto, pero finalmente es un valor libre para cada usuario.

4.2.1.5. Rango α y β

En cada grupo específico se encuentran los valores de alfa y beta óptimos de cada serie. Para calcular el rango correspondiente a este grupo se establece el promedio y la desviación estándar de las constantes. El rango se define entonces como el promedio de la constante, más su desviación estándar por un coeficiente que se ha denominado de “tolerancia”, t , que le permitirá al usuario encontrar un intervalo más amplio o más angosto según su criterio. Para este proceso se presentan las ecuaciones 5 y 6:

$$\alpha_1 < \alpha < \alpha_2, \bar{\alpha} - \sigma t < \alpha < \bar{\alpha} + \sigma t \quad 5$$

$$\beta_1 < \beta < \beta_2, \bar{\beta} - \sigma t < \beta < \bar{\beta} + \sigma t$$

Este sería el rango sugerido para α y β , en la etapa de clasificación.

4.2.2. Etapa de Clasificación

Finalmente, en la etapa de prueba o clasificación, se introduce una nueva serie, se realiza la descomposición, pero no se repiten algunos de los procesos de la etapa anterior. En su lugar el modelo, extrae las características de la serie, identifica automáticamente a que grupo pertenece, e indica respectivamente el intervalo sugerido para α y β de acuerdo con lo encontrado en el numeral 4.2.1.5.

4.3. Incorporación en software

El lenguaje de programación utilizado en este proyecto fue Python. Este lenguaje que se ha posicionado en los primeros puestos de los rankings, y su uso es frecuente en la ciencia de datos (Muenchen, 2017). El modelo se tradujo a la distribución Anaconda con el entorno de desarrollo integrado (IDE) Spyder (Scientific Python Development EnviRonment). De la distribución seleccionada se hizo uso de las librerías:

- *Openpyxl*, utilizada para el acceso a la información de los archivos de con formato xlsx.
- *Scipy*, usada en la extracción de descriptores y la optimización por mínimos cuadrados.
- *Numpy*, permitió el manejo de arreglos dinámicos.
- *Sklearn*, contiene todas las funciones asociadas al Clustering.
- *Pyplotlib*, usado para la generación de gráficos estadísticos.

Los datos de entrada se obtuvieron de la base de datos *NN3 Competition* cuya extensión es propia de Microsoft Excel (.xlsx), de forma que el modelo puede ejecutarse en cualquier otro archivo con la misma estructura de datos. Se programaron entonces cada uno de los procesos para ejecutar el modelo: lectura de datos; descomposición de componentes; suavizamiento de *Holt*; optimización; entrenamiento del clúster; prueba y salida de datos.

4.4. Validación del software

Como se mencionó anteriormente, el modelo planteado se desarrolló en la plataforma Python Anaconda. Para la etapa de entrenamiento se tomaron aproximadamente el 80% del total de series equivalente a 89 series, y los datos de clasificación o prueba correspondieron al 20% restante. Se obtuvieron los resultados de MAPE para alfas y betas sugeridos por el método propuesto, y también para alfas y betas aleatorios con la intención de emular una decisión humana entre los rangos de las constantes. Se generaron valores de las constantes 100 veces y a partir de esto se obtuvo el promedio de la medida de error MAPE, cuya ecuación se presenta a continuación:

$$MAPE = \frac{\sum_{t=1}^n \frac{|y_t - \hat{y}_t|}{|y_t|}}{n} \quad 7$$

Ambos resultados se compararon con ayuda de técnicas estadísticas para muestras pareadas. Debido a que no se conoce la naturaleza de las series de prueba con antelación, no es posible añadir una clasificación o etiqueta a cada grupo, esto es, no es posible aseverar que las series de cierto grupo pertenezcan a negocios específicos, o tengan economías distintivas. Esta es una característica general de los algoritmos de aprendizaje no supervisado, que se compensa por la generalización del modelo (aplicable a diversas series). Por esta razón, no se considera dentro del estudio una validación de la clasificación del algoritmo.

5. Resultados

Para cada una de las etapas descritas en el aparte de metodología, se presentan sus correspondientes resultados obtenidos en la realización del presente trabajo de investigación. De la revisión de literatura, se encontró que la combinación del método de descomposición (ajuste estacional) y la suavización de *Holt*, cumple con las características descritas en el planteamiento del problema como reproducibilidad, análisis y ajuste a una gama amplia de series de datos que incluyen los patrones de nivel, tendencia y estacionalidad. También se obtuvo de esta revisión una posible oportunidad de investigación en un método para facilitar la configuración de las constantes de suavización: el agrupamiento de información a través del algoritmo de *spectral clustering*. La combinación de estos métodos dio lugar a un modelo integrado, ilustrado en la figura 5, un resultado que a su vez fue implementado en un aplicativo de computador cuyos parámetros pueden ser modificados por el usuario. Finalmente se presentan los resultados obtenidos de la validación del software contra un escenario base.

5.1. Consideraciones particulares de la experiencia

Se reitera el uso del conjunto de datos de la *NN3 Competition*, el cual contiene series mensuales de diversos orígenes y tamaños, sin embargo, el algoritmo puede entrenarse, si así se considera, con otras bases de datos que posean la misma estructura de una serie en cada columna. Para la ejecución del algoritmo se hace necesario enunciar algunos parámetros y características particulares:

- Ruta del archivo informático, que en este caso es una hoja de cálculo de Microsoft Excel.
- La ubicación específica de los datos dentro del archivo.
- El número de series que se usarán para el entrenamiento, en este caso serán 89 que equivalen aproximadamente al 80% del total, un porcentaje usual en el tratamiento de datos (Hyndman y Athanasopoulos, 2013)
- El número de periodos estacionales m , que en este caso será igual a 12, ya que la frecuencia de los valores en la base de datos es mensual.

- El número de clases o grupos k en los que se clasificarán las series. Particularmente se determina como 8, no obstante, este es un parámetro flexible que el usuario puede cambiar.
- El número de ventanas en las que se dividirán las series, para hacer una comparación de estas, independiente del tamaño que posean.
- El coeficiente de tolerancia, que multiplicado por la desviación estándar acorta o dilata el rango de alfa y beta que será sugerido. Se determina un valor de 0.5, pero es un valor que puede cambiarse a criterio del usuario.

5.2. Rangos de α y β sugeridos según grupo

Como resultado de la aplicación del modelo y con base en los parámetros anteriormente enunciados se obtienen los rangos de α y β de acuerdo con el grupo al cual fueron asignados (Tabla 2). Adicionalmente se muestra la carga o el porcentaje de cada grupo en el conjunto de datos de entrenamiento, de prueba y del total de 111 series.

Tabla 2. Valores de alfa y beta para cada grupo, junto con su porcentaje (total series).

| Grupo | α | β | %Entrena (89) | %prueba (22) | % del total |
|-------|------------|------------|---------------|--------------|-------------|
| 1 | 0.32, 0.75 | 0.04, 0.39 | 5,62% | 31,8% | 10.81% |
| 2 | 0.5, 0.8 | 0.07, 0.38 | 22,47% | 45,5% | 27.03% |
| 3 | 0.08, 0.1 | 0.11, 0.47 | 11,24% | 0,0% | 9.01% |
| 4 | 0.05, 0.09 | 0.28, 0.55 | 14,61% | 0,0% | 11.71% |
| 5 | 0.09, 0.11 | 0.11, 0.18 | 11,24% | 0,0% | 9.01% |
| 6 | 0.17, 0.52 | 0.01, 0.13 | 17,98% | 18,2% | 18.02% |
| 7 | 0.07, 0.1 | 0.28, 0.54 | 12,36% | 0,0% | 9.91% |
| 8 | 0.24, 0.41 | 0.03, 0.06 | 4,49% | 4,5% | 4.50% |

Específicamente para las series consideradas de prueba o clasificación se tienen los siguientes resultados en cuanto al porcentaje de series incluidas (Figura 9).

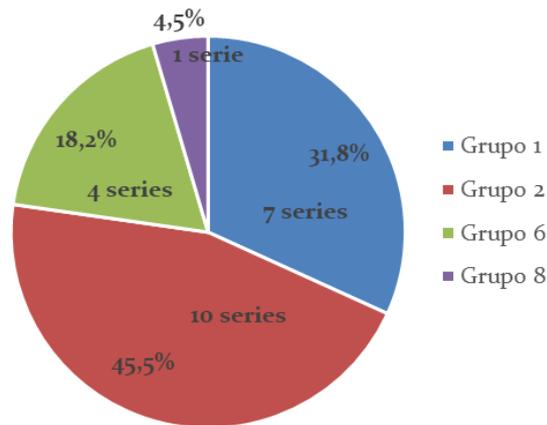


Figura 9. Distribución de las series de prueba dentro de los grupos. Elaboración propia.

A algunos grupos no les fue asignado ninguna serie de datos, como los grupos 3, 4, 5 y 7, situación que podría no repetirse con otro conjunto de datos. Cada serie nueva que se ingrese será clasificada en un grupo, el cual tiene asignado un rango para las constantes de suavización. Se reitera que el número de grupos es un valor que puede ser modificado por el usuario del aplicativo, sin embargo, un número muy reducido de estos puede estar “obligando” a las series a agruparse aun cuando no poseen características similares. Es posible evidenciar que los valores de alfa y beta son aproximadamente bajos, a excepción del rango de alfa en el grupo 2, con valores de 0.5-0.8. Los mayores valores para beta se encuentran en el grupo 4 y corresponden a (0.28-0.55). Cabe precisar que los valores de los parámetros de suavización dependerán de las series que sean utilizadas en la etapa de entrenamiento.

5.3. Resultados validación del modelo

En la etapa de validación del método propuesto, con el objetivo de medir su rendimiento, se calculó 100 veces el MAPE de las series de prueba con diferentes valores de alfa y beta dentro del rango sugerido, adicionalmente como medida global se calculó el promedio de estos MAPE para cada serie. El mismo procedimiento se realizó una segunda vez, sin embargo, en este caso se eligieron valores aleatorios dentro del rango 0-1, en lugar de considerar los rangos sugeridos por el modelo, este se consideró el escenario base de comparación. En la tabla 3 se muestran los MAPE

promedio para cada serie de prueba, tanto del modelo propuesto como del escenario base.

Tabla 3 Medida de error promedio de la propuesta y el escenario base

| Series de prueba | MAPE promedio propuesta | MAPE promedio base |
|------------------|-------------------------|--------------------|
| Serie 90 | 3.79 | 5.1 |
| Serie 91 | 0.6 | 2.23 |
| Serie 92 | 3.62 | 8.74 |
| Serie 93 | 211.52 | 245.72 |
| Serie 94 | 4.65 | 23.08 |
| Serie 95 | 23.94 | 43.98 |
| Serie 96 | 146.75 | 147.62 |
| Serie 97 | 11.87 | 48.35 |
| Serie 98 | 25.81 | 28.35 |
| Serie 99 | 26.59 | 57.62 |
| Serie 100 | 5.13 | 8.88 |
| Serie 101 | 2.02 | 5.8 |
| Serie 102 | 11.45 | 10.09 |
| Serie 103 | 30.86 | 37.06 |
| Serie 104 | 6.87 | 9.77 |
| Serie 105 | 2.17 | 2.44 |
| Serie 106 | 5.31 | 12.73 |
| Serie 107 | 3.02 | 8.29 |
| Serie 108 | 20.75 | 76.43 |
| Serie 109 | 5.63 | 6.84 |
| Serie 110 | 29.17 | 78.97 |
| Serie 111 | 14.77 | 60.06 |

Para comparar el rendimiento del modelo respecto al escenario base se pueden contrastar sus *medias* y determinar de este modo, si hay una diferencia estadísticamente significativa entre ellas y cual tuvo un mejor desempeño. Debido a que los datos se refieren a errores, un mejor desempeño equivale a una *media* menor. Para realizar esta comparación, en primer lugar, es necesario realizar una prueba de normalidad, ya que esta es una característica imperativa en las pruebas paramétricas de comparación de *medias*.

5.3.1. Pruebas de normalidad de los errores

Se realizan pruebas de normalidad y gráficas de probabilidad normal, a través del software Statgraphics Centurion XVI, para validar los supuestos de normalidad de los datos en la tabla 3. A continuación, se presentan los resultados.

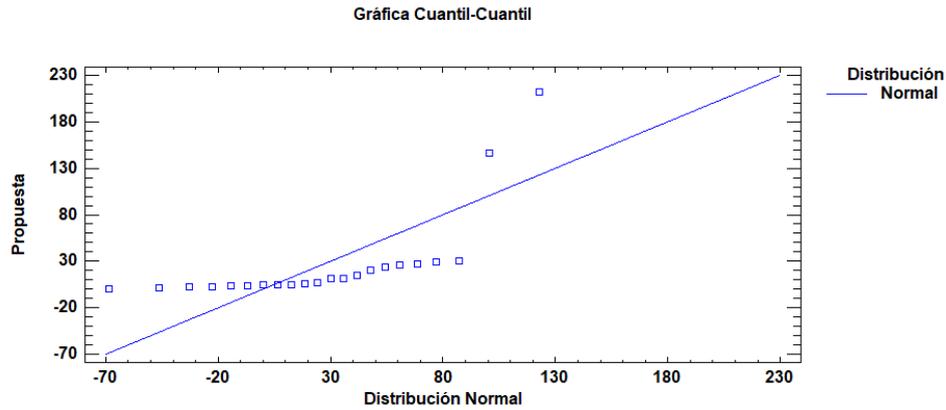


Figura 10. Gráfico cuantil-cuantil del MAPE promedio del modelo propuesto. Elaboración propia.

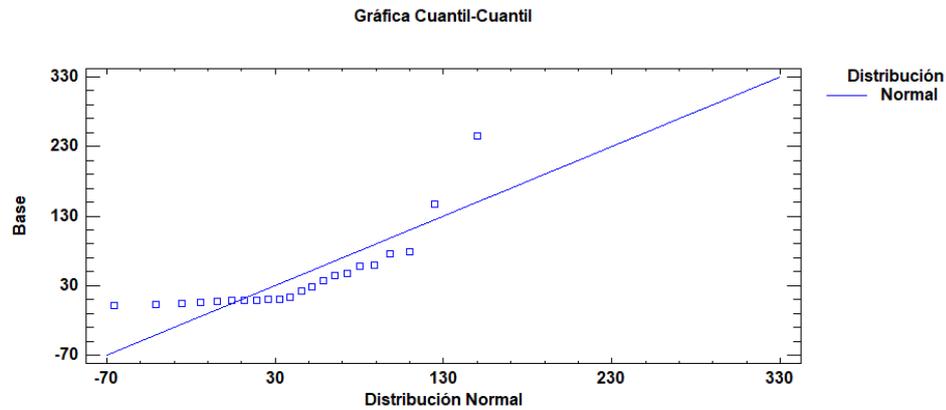


Figura 11. Gráfico cuantil-cuantil del MAPE promedio del escenario base. Elaboración propia.

Es posible observar al inicio de la figura 10, una desviación de la línea que podría indicar un comportamiento no normal en los errores tanto del método propuesto como del escenario base (figura 11). También se observan valores muy alejados de la línea que representa una distribución normal, lo que puede indicar la presencia de datos atípicos.

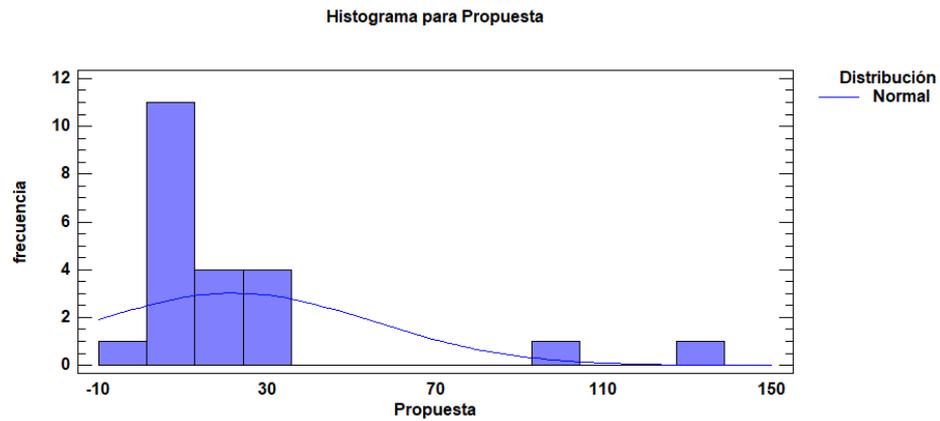


Figura 12. Histograma para los MAPE promedio del modelo propuesto. Elaboración propia

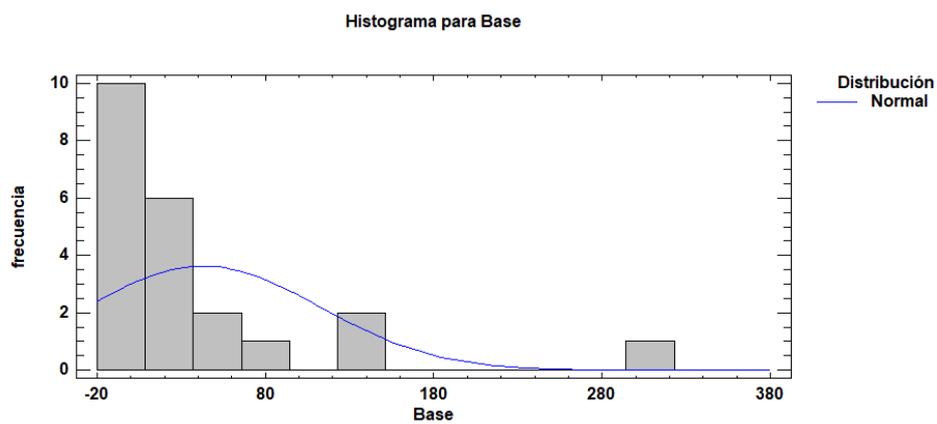


Figura 13. Histograma para los MAPE promedio del escenario base. Elaboración propia.

Igualmente, el histograma de ambos escenarios presenta un sesgo notable de los errores hacia la izquierda. Para reafirmar esta suposición se realizan algunas pruebas estadísticas de normalidad, como la prueba *Shapiro Wilks*. Para el modelo propuesto, esta prueba presentó un valor p de 1.55018×10^{-8} , menor al nivel de significancia de 5%, por tal razón es factible suponer que los errores no provienen de una distribución normal. Esta misma prueba, para los errores del escenario base, arrojó un valor p de 3.60073×10^{-6} , también menor al nivel de significancia. En vista de estos resultados, se recurre a una prueba no paramétrica para la comparación de medias.

5.3.2. Comparación entre muestras pareadas

Debido a que las pruebas de normalidad sugieren que los errores en ambos métodos no provienen de una distribución normal una alternativa de comparación es realizar la prueba de signos y de rangos de signos de *Wilcoxon* para comparar la

diferencia entre las medianas. Estas alternativas no paramétricas pueden resultar más exigentes en la comparación con significancia estadística, pero es una alternativa adecuada cuando se tiene información no normal (Healy, 1994; Sainani, 2012).

5.3.2.1. Pruebas bilaterales o de dos colas

La prueba de signos, realizada también en Statgraphics Centurion XVI, arrojó un valor-p de 0.0000510655 y la prueba de rangos con signo de Wilcoxon uno de 0.0000855685, a dos colas. De acuerdo con estas pruebas no paramétricas es posible considerar que los métodos sean significativamente diferentes en cuanto a sus medianas.

5.3.2.2. Pruebas de una cola (Diferencia de medianas <0)

A la luz de estos resultados se repiten las pruebas, esta vez para establecer si una mediana es potencialmente mayor o menor que la otra. La prueba de signos arrojó un valor-p de 0.0000255328, al igual que la prueba de rangos con signo, con la cual se obtuvo un valor de 0.0000427842, ambos menores que el nivel de significancia de 0.05. Dado que la hipótesis alternativa para esta prueba es que la diferencia de medianas es menor a 0, hay un indicio estadístico de que la mediana del primer método es menor que la del segundo, por tanto, no se rechaza la idea de que el método propuesto por el presente texto posea un mejor rendimiento que el escenario base planteado.

5.4. Análisis de datos atípicos

A partir de las figuras 14 y 15, es posible distinguir dos puntos notablemente alejados del resto de las observaciones, que podrían ser valores atípicos. Tomando en cuenta que estos valores podrían causar el comportamiento no normal, se realiza la prueba de Grubb, obteniendo valores-p menores a 0.05 y la prueba de Dixon obteniendo resultados similares que indican la presencia de valores atípicos para ambos conjuntos de datos. En las figuras 14 y 15, también se pueden observar datos alejados de la media muestral, correspondientes a los valores de las series 93 y 96 de la tabla 3 en el caso del escenario propuesto y escenario base.

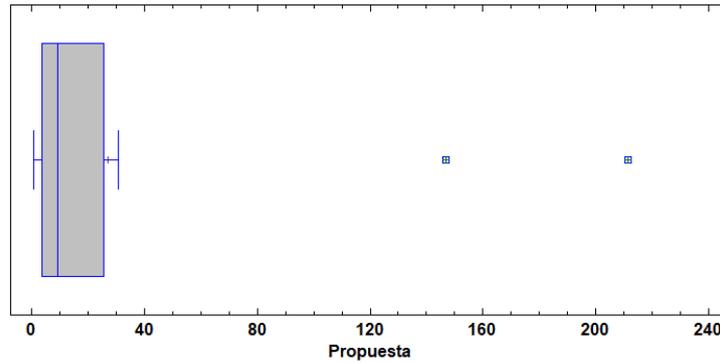


Figura 14. Gráfico de caja y bigotes para los MAPE promedio del modelo propuesto. Elaboración propia

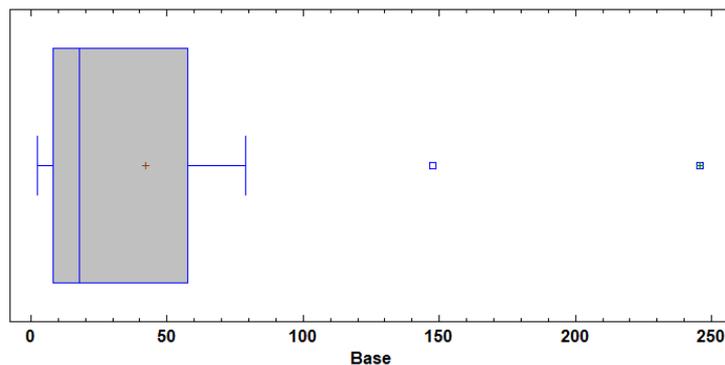


Figura 15. Gráfico de caja y bigotes para los MAPE promedio del escenario base. Elaboración propia.

Aun cuando se detecta la presencia de estos datos atípicos, se observa en el conjunto de datos un marcado sesgo a la izquierda, situación que se resalta cuando se eliminan dichas observaciones del conjunto de datos, indicador visual de no normalidad.

5.5. Relación entre características estadísticas y error de pronóstico.

El hecho de poseer errores de tal magnitud en estas dos series (93 y 96) en ambos escenarios, motiva la idea de una posible relación entre las características de estas series y su desempeño en ambos métodos, en otras palabras, es posible que las características estadísticas propias de estas series sean tales que impidan u obstaculicen el desempeño de este y otros métodos de pronósticos de series de tiempo.

Debido a que los datos no son normales, se exploran técnicas no paramétricas para estimar una posible relación entre el MAPE promedio y las características que fueron extraídas de los datos, *media*, *mediana*, *desvest*, *CV*, *entropía* y *asimetría*.

A continuación, se presentan los diagramas de dispersión como una representación de la relación entre cada una de las características y el MAPE promedio.

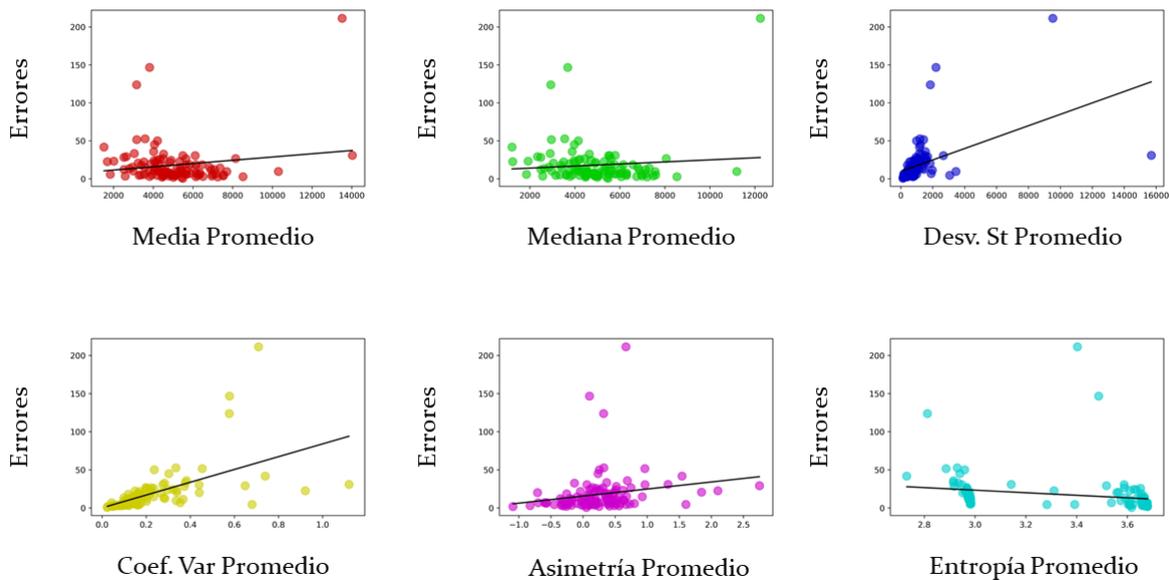


Figura 16. Diagrama de dispersión entre los MAPE promedio de los errores de la propuesta y las características extraídas de las series. Elaboración propia.

En la figura 16, es posible observar una línea de tendencia que sirve de apoyo para identificar la posible inclinación de cada una de las variables. Particularmente, es posible notar una relación débil o nula para la media y la mediana, dado que su tendencia no parece tener alguna inclinación. De igual forma, la asimetría, parece exhibir una relación débil, aunque en este caso la línea de tendencia muestra mayor inclinación. De todas las características, se observa que la entropía es la única que parece tener una correlación negativa, ya que a medida que se avanza en el eje x, disminuye el valor en el eje y. En contraste, los diagramas del coeficiente de variación y la desviación estándar, muestran un comportamiento positivo de magnitud significativa. Para cuantificar la fuerza de estas posibles relaciones se usa una alternativa no paramétrica, el coeficiente de Spearman.

Los coeficientes de Spearman para la *media* ($r_s = -0.2703$), *mediana* ($r_s = -0.2787$), *desviación estándar* ($r_s = 0.7027$), *coeficiente de variación* ($r_s = 0.8089$), *asimetría* ($r_s = 0.3407$) y *entropía* ($r_s = -0.6704$), parecen confirmar lo observado en los diagramas de dispersión. Las características de media, mediana y entropía exhiben una posible correlación negativa, siendo la entropía la que presenta mayor fuerza en la relación,

que podría clasificarse como una correlación negativa moderada, esto es, a medida que la *entropía* aumenta, el error disminuye de forma regular. Esta misma afirmación podría hacerse para la mediana ya que posee un valor similar. El valor de la asimetría indica una posible correlación débil, entre el error y esta variable, corroborando lo encontrado en la figura 16. Para el coeficiente de variación y desviación estándar, se encontró una correlación positiva fuerte, coincidiendo también con las suposiciones realizadas con base en la figura 16 en este caso, a medida que aumentan las variables, aumenta el error aproximadamente proporcional.

Se tiene entonces, que las variables de coeficiente de variación, desviación estándar y entropía fueron las que exhibieron mayores correlaciones con el MAPE promedio, de forma negativa y positiva. Estos resultados sustentan la noción de correlación entre estas características y el rendimiento expresado con el MAPE promedio, con la particularidad de que se tratan de características que miden el grado de dispersión o desorden en los datos. En el cálculo de estas correlaciones se toman los promedios de las características de cada ventana de tiempo para las 111 series que componen la base de datos. Este promedio es el que se relaciona con la medida de error, MAPE promedio del modelo propuesto.

5.6. Análisis de los grupos formados

Se analiza también el comportamiento de las características en cada uno de los grupos formados por el algoritmo para las 111 series en su totalidad. En este análisis se consideraron las series sin ser normalizadas, con el fin de apreciar el comportamiento real de las características en su interior.

Se tomaron los promedios de las características de cada grupo y ventana, y se obtuvieron los resultados que muestran una diferencia de escala notable entre las tres primeras características y el resto, en todos los grupos. Por tal razón se revisan de manera separada de acuerdo con la escala de sus valores. En la figura 17, se encuentran las medidas de *media*, *mediana* y *desvest* de todos los grupos.

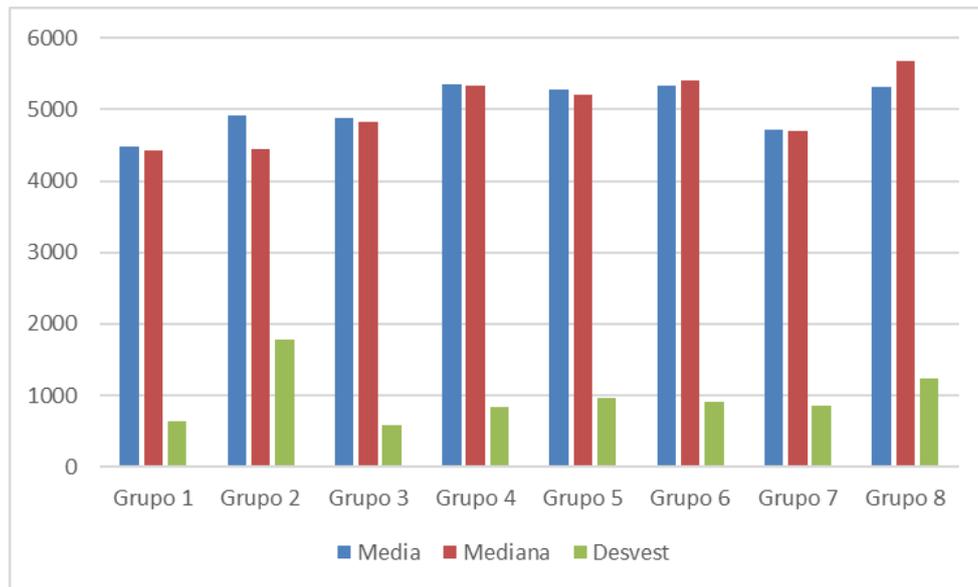


Figura 17. Media, mediana y desviación promedio para cada grupo (total de series). Elaboración propia.

Es de notar la estabilidad de los valores de media y mediana, los cuales oscilan entre 4480 y 5355 aproximadamente. La desviación estándar también expone un comportamiento aproximadamente estable, que oscila entre 593 y 1000, exceptuando a los grupos 2 y 8, que presentan valores de 1773 y 1245, respectivamente. Cuando se considera la estabilidad del comportamiento de estos valores a lo largo de los grupos, se podría pensar que estas características no resultaron de gran influencia en la formación de los grupos, aunque valores elevados para la desviación estándar como el del grupo 2 podría indicar que este alberga series con una aparente mayor dispersión en la población. De igual forma se resalta que el algoritmo de *clustering*, al ser un enfoque multivariado evalúa la similitud de las características de las series simultáneamente, análisis que resulta difícil de simplificar considerando limitaciones como la visualización de más de tres descriptores, y la compleja interpretación de la influencia individual o combinada de los mismos.

El mismo análisis se realizó para las otras tres características, debido a la escala tan pequeña de sus datos, los cuales se observan en la figura 18.

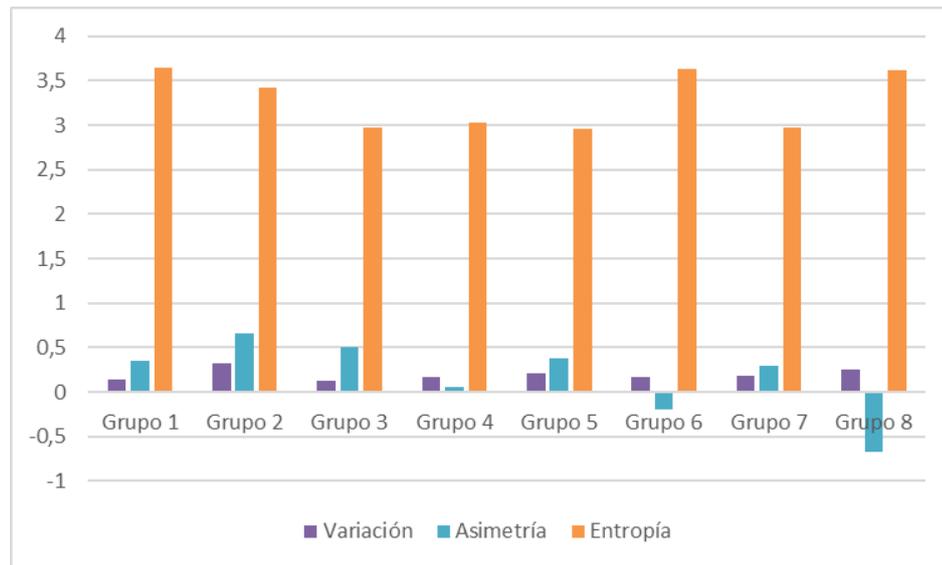


Figura 18. Coeficiente de variación, asimetría y entropía para cada grupo (total series). Elaboración propia.

Al igual que con las medidas anteriores, el valor de la entropía es más bien estable, el cual varía entre 3 y 3,6, y por tal razón se podría pensar que su influencia en la agrupación es de menor relevancia. Por otro lado, nuevamente el grupo 2 presenta mayores valores para el coeficiente de variación, y la asimetría, reforzando la idea, junto con la desviación estándar, de que en este grupo se encuentran series con alto grado de dispersión. Este hecho podría ser un factor común en el resto de las series, es decir la agrupación podría estar motivada principalmente por las medidas de dispersión (desviación, asimetría y coeficiente de variación).

A partir de los resultados anteriores, se inspeccionan las relaciones entre las características y el rendimiento del modelo propuesto al interior de cada grupo, para las 111 series. El objetivo es identificar si existe una característica predominante en todos los grupos, y observar si es un patrón de relevancia para realizar las agrupaciones.

5.6.1. Análisis Individual

5.6.1.1. Grupo 1

En primer lugar, se observa la figura 19 correspondiente a las posibles relaciones entre las características de *media*, *mediana*, *desvest*, *CV*, *asimetría* y *entropía*, y error

del grupo 1. Este grupo está compuesto por 12 series de las 111 series pertenecientes a la base de datos *NN3 Competition*.

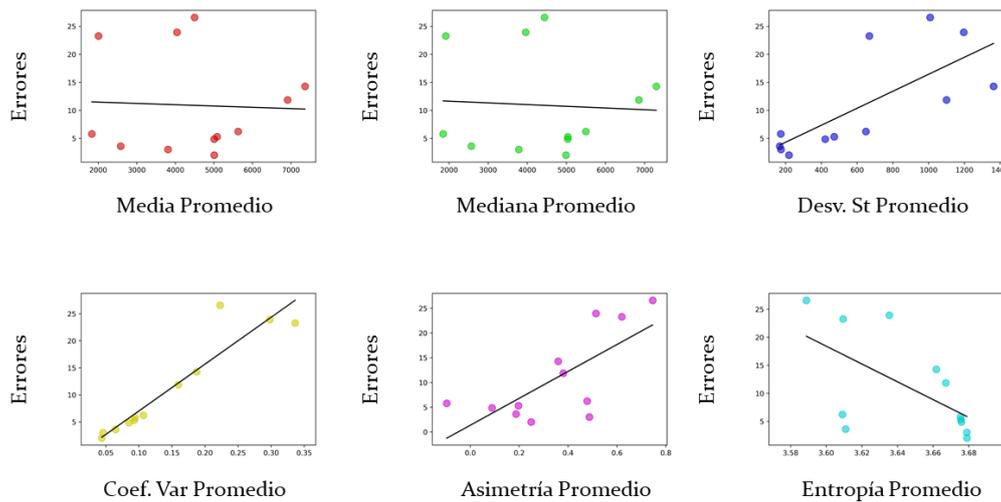


Figura 19. Correlación Características promedio y MAPE promedio para Grupo 1. Elaboración propia.

En este grupo, la media y mediana, parecen no tener relación con el error MAPE del modelo propuesto. Las medidas de desviación estándar, coeficiente de variación y asimetría, exhiben una relación positiva, contrario a la entropía. En este caso las características con mayor magnitud respecto al coeficiente de Spearman fueron coeficiente de variación ($rs=0.972$), desvest ($rs=0.7972$) y entropía ($rs=-0.7483$).

5.6.1.2. Grupo 2

En este caso particular, el grupo estuvo compuesto por 30 series de las 111 disponibles. Como se observa en la figura 20, la línea de tendencia casi horizontal de la asimetría da cuenta de una relación casi inexistente entre esta y el MAPE. La media, mediana, desvest y CV, manifiestan una relación positiva, y la entropía una negativa. Los tres factores con mayor valor para el coeficiente de Spearman son CV ($rs= 0.8029$), desvest ($rs=0.7811$) y entropía ($rs=-0.6271$).

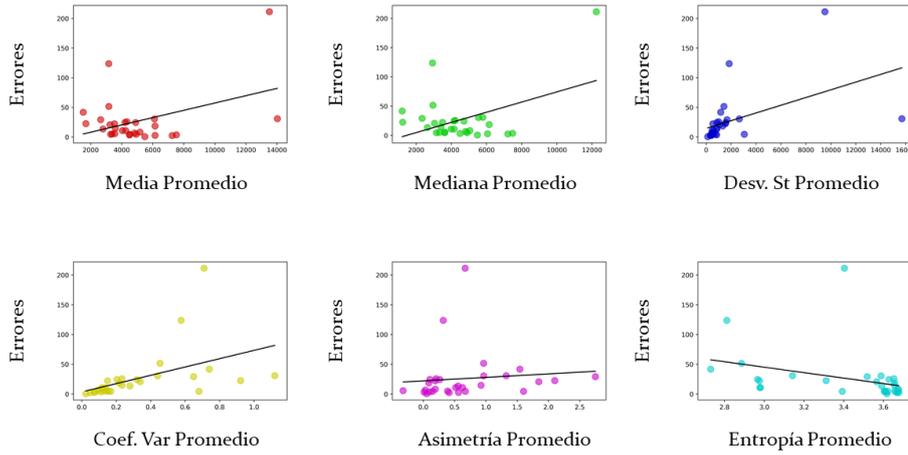


Figura 20. Correlación Características promedio y MAPE promedio para Grupo 2. Elaboración propia.

5.6.1.3. Grupo 3

Se observan para este grupo, efectos positivos para la desviación estándar, y coeficiente de variación. Relaciones negativas para la media, mediana y entropía, y una relación débil para la asimetría. En este caso se contó con 10 series.

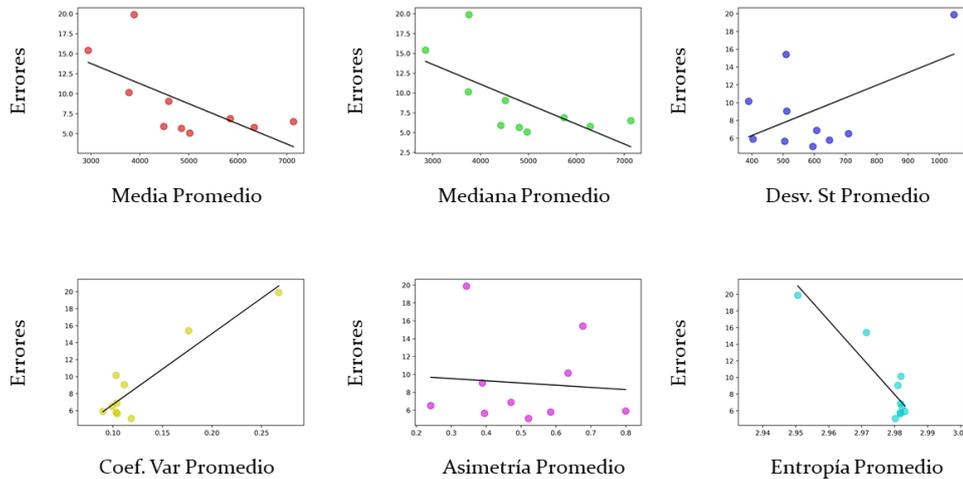


Figura 21. Correlación Características promedio y MAPE promedio para Grupo 3. Elaboración propia.

Particularmente para este grupo, las características con mayores coeficientes de Spearman fueron la media y mediana (ambos $rs=-0.6364$), y el coeficiente de variación ($rs=0.3333$). En este caso los valores de la media no distan en gran magnitud de los de la mediana, por esa razón los valores para Spearman son muy similares. Esta similitud también da cuenta de una distribución aproximadamente simétrica de los datos en las 10 series que componen el grupo. Los valores para coeficiente de variación en este

grupo fueron relativamente bajos, al igual que su correlación con el error, esto puede deberse a que este depende de la desviación estándar que fue baja en relación con la media.

5.6.1.4. Grupo 4

En este grupo (13 series) también se puede apreciar una simetría de las series de tiempo, debido a la similitud de las figuras de media y mediana. Las figuras muestran relaciones negativas para la media, mediana y entropía, y en contraste, la desviación estándar, el coeficiente de variación y la asimetría, exhiben relaciones positivas, indicando que al aumentar su valor el MAPE también parece haber aumentado. La magnitud de la relación también se puede observar a medida que los valores se acercan más a la línea de tendencia, en este caso el coeficiente de variación parece cumplir con esta propiedad. De forma análoga, se presentan los factores con mayor coeficiente de Spearman: CV ($r_s= 0.9231$), entropía (0.9176) y desvest ($r_s=0.8022$).

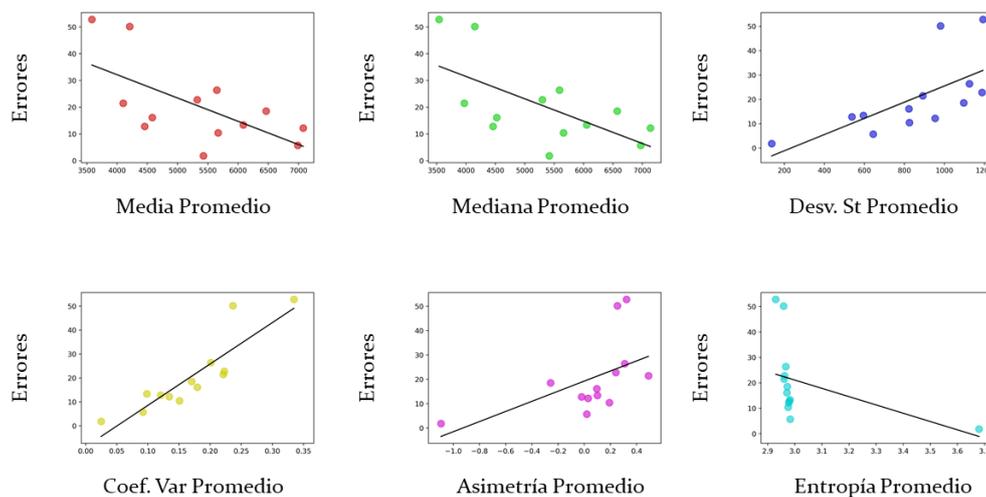


Figura 22. Correlación Características promedio y MAPE promedio para Grupo 4. Elaboración propia.

5.6.1.5. Grupos 5, 6, 7 y 8.

Los análisis del resto de los grupos se realizan de forma análoga. Se revisa la línea de tendencia de las series en cada característica, para determinar una posible correspondencia positiva o negativa entre las variables. La cercanía de los datos a esta línea de tendencia es un indicador de la magnitud de las relaciones, a mayor cercanía mayor relación puede existir. Igualmente se presentan las tres

características que obtuvieron mayores coeficientes de Spearman, y por tanto reflejan mayor relación de la característica en cada grupo.

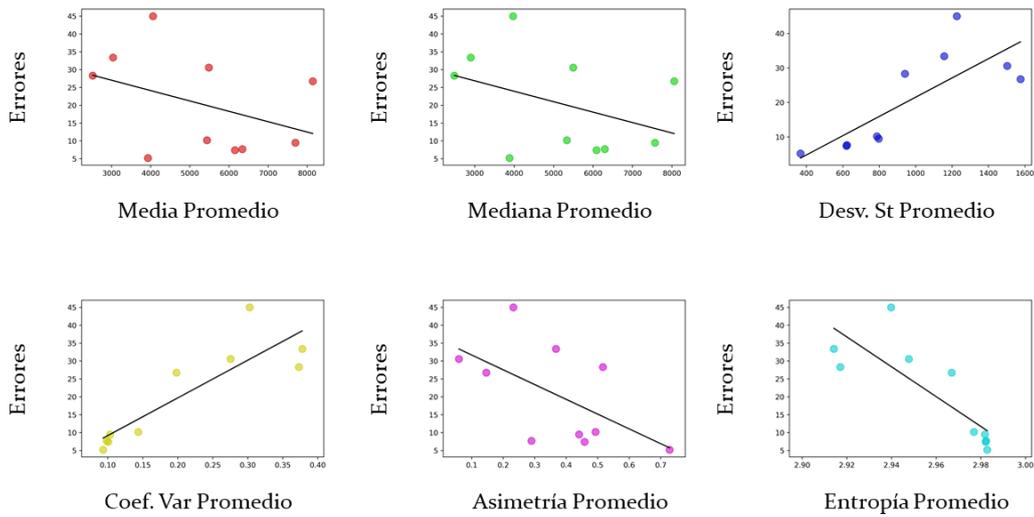


Figura 23. Correlación Características promedio y MAPE promedio para Grupo 5. Elaboración propia.

Para el grupo 5 (figura 23) se tiene que el coeficiente de variación ($rs=0.9273$), entropía ($rs=-0.9273$) y desviación estándar ($rs=0.8303$), son las características con mayores valores para Spearman. El grupo contiene 10 series de tiempo.

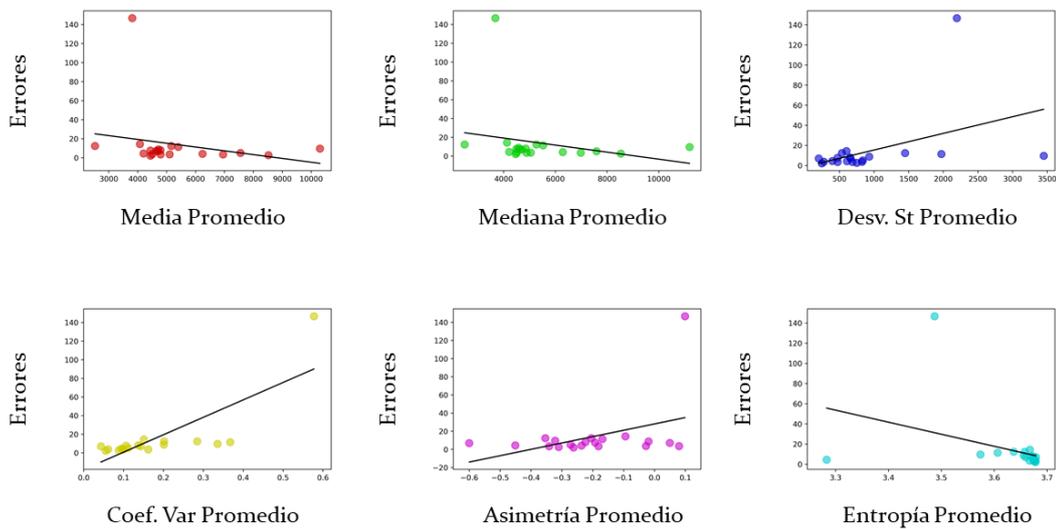


Figura 24. Correlación Características promedio y MAPE promedio para Grupo 6. Elaboración propia.

En el grupo 6 (figura 24), los primeros puestos respecto al coeficiente de Spearman los ocupan las características de CV ($rs=0.7820$), entropía ($rs=-0.6842$) y

desvest ($rs=0.4466$). La serie está compuesta por 20 series de tiempo de la base de datos.

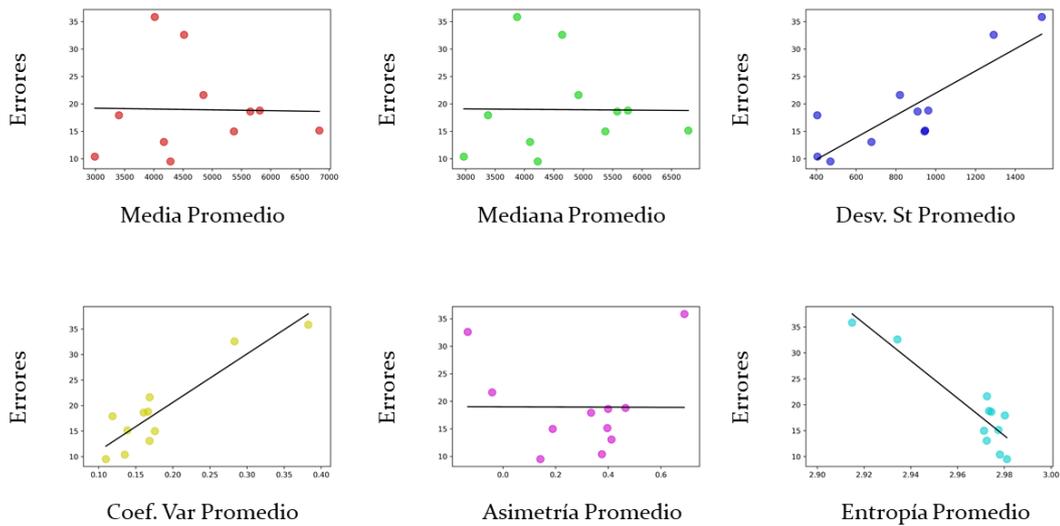


Figura 25. Correlación Características promedio y MAPE promedio para Grupo 7. Elaboración propia.

En este grupo se encuentran la desviación estándar ($rs=0.7$), el coeficiente de variación ($rs=0.6909$) y entropía ($rs=-0.6364$), como las características con mayores coeficientes de Spearman, evidenciando mayor relación con el MAPE. En este grupo se encuentran 11 series.

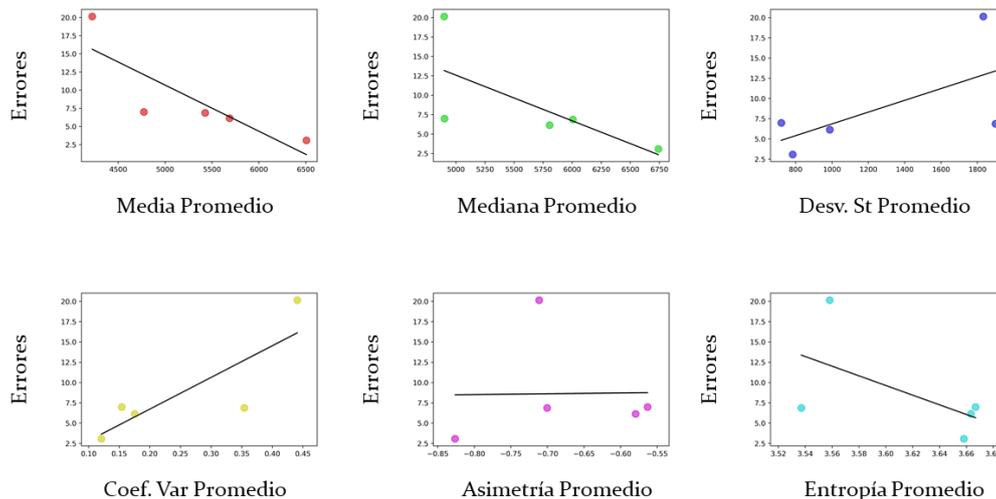


Figura 26. Correlación Características promedio y MAPE promedio para Grupo 8. Elaboración propia.

En la figura 26 se aprecian las posibles relaciones entre las características de las 5 series que se encuentran en el grupo 8. En este caso las características con mayores

magnitudes según *Spearman*, fueron la media ($rs=-1$), mediana ($rs=-0.9$) y coeficiente de variación ($rs=-0.7$).

5.6.1.6. Aspectos globales de las correlaciones

Es posible identificar una posible característica predominante en todas las agrupaciones, en donde el coeficiente de variación estuvo entre las tres magnitudes más grandes para el coeficiente de Spearman en los 8 grupos, siendo este valor positivo en todos los casos. Otras dos características presentaron patrones consistentes a lo largo de los grupos: la desviación estándar estuvo presente con gran dimensión en 6 de los grupos de manera positiva, y la entropía que obtuvo valores significativos también en 6 de los grupos, pero con una relación negativa. Estos valores pueden dilucidar una posible preponderancia de estas variables en las agrupaciones, hecho que se revalida con los resultados globales obtenidos en el numeral 5.5, en el cual el coeficiente de variación estuvo altamente relacionado con el error del modelo propuesto. La figura 27 resume los valores antes presentados en cuanto a la correlación existente (coeficiente de Spearman) entre cada característica de cada grupo y el MAPE promedio.

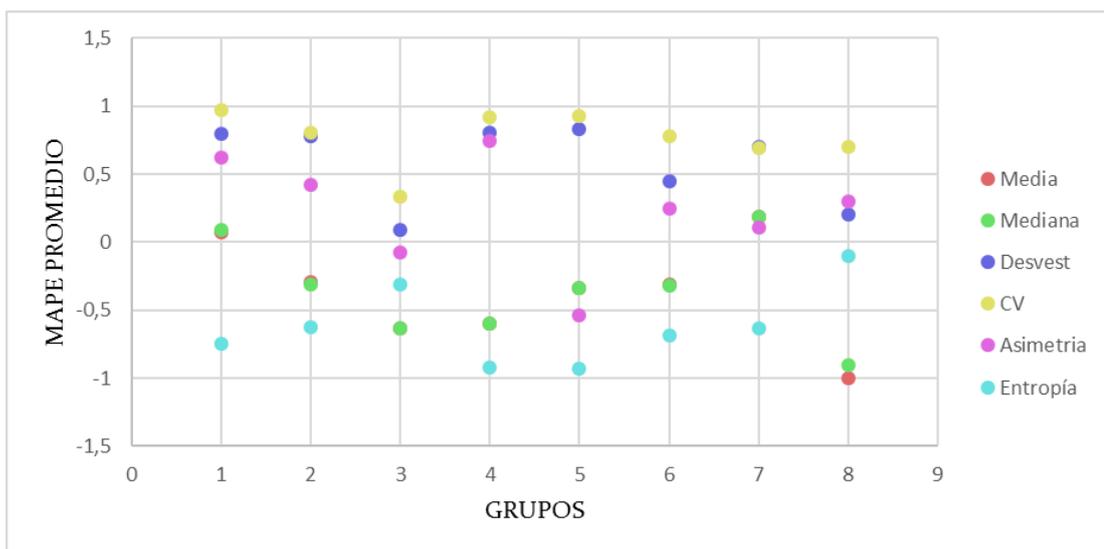


Figura 27. Resumen Spearman Características vs MAPE promedio por grupo. Elaboración propia.

Con base en los coeficientes presentados con anterioridad y de acuerdo con la figura 27, el coeficiente de variación es un factor predominante en la mayoría de los grupos, sin embargo, dado que en algunos grupos otras características como la

entropía fueron más “influyentes”, se puede pensar que esta agrupación requiere de uno o más descriptores o características para que se afecte el error del modelo propuesto. Para el caso particular dichos descriptores “influyentes” se trataron de medidas de dispersión de las series: coeficiente de variación, entropía y desviación estándar.

Teniendo en cuenta los resultados globales y por grupo en cuanto a la correlación de las características, se puede acordar que el coeficiente de variación, la desviación estándar y la entropía son variables influyentes. A partir de estos resultados se revisan los valores de los promedios de cada característica en cada grupo (Tabla 4), y se muestra la posición de cada grupo, cuando estas cifras son ordenados de menor a mayor (Tabla 5).

Tabla 4. Promedio de las características por grupo junto con alfas y betas. Elaboración propia.

| Grupo | Media | Mediana | Desvest | CV | Asimetría | Entropía | α | β |
|-------|---------|---------|---------|------|-----------|----------|------------|------------|
| 1 | 4480,32 | 4435,26 | 634,34 | 0,14 | 0,35 | 3,65 | 0,32-0,75 | 0,04- 0,39 |
| 2 | 4910,23 | 4444,82 | 1772,84 | 0,32 | 0,66 | 3,42 | 0,5- 0,8 | 0,07- 0,38 |
| 3 | 4886,10 | 4824,75 | 592,63 | 0,13 | 0,51 | 2,98 | 0,08- 0,1 | 0,11-0,47 |
| 4 | 5355,50 | 5334,31 | 845,31 | 0,17 | 0,05 | 3,02 | 0,05-0,09 | 0,28- 0,55 |
| 5 | 5280,67 | 5208,13 | 959,47 | 0,21 | 0,37 | 2,96 | 0,09- 0,11 | 0,11-0,18 |
| 6 | 5341,53 | 5399,80 | 910,55 | 0,17 | -0,20 | 3,63 | 0,17-0,52 | 0,01- 0,13 |
| 7 | 4717,70 | 4690,47 | 851,14 | 0,18 | 0,29 | 2,97 | 0,07- 0,1 | 0,28- 0,54 |
| 8 | 5323,23 | 5670,57 | 1244,92 | 0,25 | -0,68 | 3,62 | 0,24-0,41 | 0,03- 0,06 |

Tabla 5. Posición de los grupos cuando los promedios de variables se ordenan de menor a mayor.

| CV | Desvest | Entropía | Mediana | Media | Asimetría |
|----|---------|----------|---------|-------|-----------|
| 3 | 3 | 5 | 1 | 1 | 8 |
| 1 | 1 | 7 | 2 | 7 | 6 |
| 4 | 4 | 3 | 7 | 3 | 4 |
| 6 | 7 | 4 | 3 | 2 | 7 |
| 7 | 6 | 2 | 5 | 5 | 1 |
| 5 | 5 | 8 | 4 | 8 | 5 |
| 8 | 8 | 6 | 6 | 6 | 3 |
| 2 | 2 | 1 | 8 | 4 | 2 |

De acuerdo con este ordenamiento, es posible efectuar algunas suposiciones, especialmente si se considera que la desviación y el CV fueron las características con mayor correlación. En este respecto, el grupo 3 pudo haber albergado las series con menor dispersión o variación en sus datos, y también se observa que ocupó una posición media en cuanto a la media y mediana, variables comúnmente asociadas a la escala o nivel de los datos. En contraste se tiene al grupo 2, que según la tabla 4 al estar ubicado en la última posición, posee en promedio los valores de CV y *desvest* más altos. Precisamente en este grupo se ubica una de las series identificadas como problemáticas en un análisis anterior, debido a la magnitud de su error, la serie 93. Se observa también que los menores valores de alfa parecen concordar con los menores de la variable entropía. Esta variable, referente a la teoría de la información, es una medida de “uniformidad” en la distribución de los datos, contrario al concepto inicial de entropía, una entropía “alta”, significa que x proviene de una distribución “uniforme” o estable, y un valor “bajo”, significa que proviene de una distribución muy variable, con picos y valles (Shannon, 1948). En este sentido, y de acuerdo con la teoría de la suavización exponencial, cuando alfa toma valores cercanos a uno, el efecto del suavizado es menor sobre la serie de tiempo, el pronóstico se convierte en uno ingenuo, en otras palabras, es igual al valor inmediatamente anterior, hecho que indica poca variación de un periodo a otro. Cuando se reduce el valor de la constante se está suavizando la serie en mayor magnitud, esto puede ser una respuesta a la alta variación que puede tener reflejada en valores bajos de la entropía.

6. Discusión

6.1. Hallazgos fundamentales

El trabajo se centró en facilitar el uso de suavización exponencial de Holt, al proporcionar los parámetros alfa y beta de acuerdo con un algoritmo de *spectral clustering*. Se definió un escenario base en donde un usuario puede escoger dichos parámetros de forma aleatoria en el intervalo de 0 a 1, y se comparó con la metodología propuesta. La medida de comparación usada para este fin fue el MAPE promedio de los errores, encontrando que, en general, el método propuesto arroja buenos resultados en comparación con el escenario de *benchmark*. Estos resultados se compararon a través de pruebas no paramétricas debido a que los errores exhibieron comportamientos no normales, y se obtuvieron valores-p menores al nivel de significancia de 5%, de 0.0000255328 para la prueba de signos y de 0.0000427842 para la de rangos con signo de Wilcoxon. El modelo reduce el sesgo al introducir rangos de alfa y beta, basados en metodologías cuantitativas formales, pero al mismo tiempo brinda un espacio para la experiencia del usuario al permitírsele mover dentro de este rango.

Adicionalmente, debido a la presencia de errores muy notables, se estudió la posible relación entre el rendimiento del modelo y las características estadísticas de las series. Con base en los diagramas de dispersión y los coeficientes de Spearman, se encontró una alta correlación ($r_s = 0.8089$) entre el coeficiente de variación y el MAPE, lo que podría traducirse en una influencia mayor de esta característica tanto a nivel global como en la agrupación. Otras medidas de dispersión de los datos, como la desviación estándar ($r_s = 0.7027$) y la entropía ($r_s = -0.6704$) obtuvieron valores altos. Con esto en consideración la agrupación de las series estaría motivada principalmente por el coeficiente de variación, sin embargo, al no tratarse del único descriptor de influencia, no se descarta la importancia de incluir varios descriptores con el fin de obtener una separación correcta de los grupos. Por otro lado, las características de media ($r_s = -0.2703$), mediana ($r_s = -0.2787$) y asimetría ($r_s = 0.3407$), no obtuvieron en general valores significativos en cuanto a la correlación, sugiere un efecto muy modesto en la agrupación de los datos, sin embargo, la media al ser una medida

resumen que indica el nivel o la escala de los datos, puede estar altamente relacionada con el valor de alfa, asociado también a la ecuación que suaviza el nivel en el método de *Holt*, por tal razón no se debe descartar en el análisis porque puede resultar importante en la agrupación de las series, aunque no tenga directa influencia en el error del modelo propuesto.

6.2. Conclusiones y limitaciones

El presente trabajo propone un modelo integrado, traducido en un aplicativo computacional, que se puede emplear en demandas de diversas industrias, y a series de tiempo de diferentes escalas, sin un conocimiento extenso en el área de estadística. Debido a la naturaleza de los métodos de pronósticos, el modelo puede ser implementado en series que exhiban patrones de nivel, tendencia o estacionalidad y está encaminado hacia el principio de parsimonia al contar con un número reducido de parámetros, los cuales pueden ser modificados por el usuario.

El modelo pretende facilitar la tarea de aquellas personas encargadas de las estimaciones de demanda en los negocios que, no poseen aún experticia en la predicción, información a priori de la naturaleza de los datos o dedicación exclusiva al análisis de las series de tiempo. Todo esto, porque ayuda a parametrizar un método que abarca una amplia gama de patrones regulares que manifiestan las series de tiempo, tarea que resulta compleja, especialmente si se considera la gran cantidad de artículos con patrones diferentes que se puede tener una organización y el efecto de estos en la toma de decisiones y planeación. El modelo también contribuyó a la mejora de las predicciones en el escenario planteado. En promedio se observa una mejoría de 43% en el error, en cuanto al uso de la combinación descomposición-suavización de *Holt* propuesto, valor de gran importancia para las compañías, especialmente para aquellos productos de alta prioridad, que constituyen grandes ganancias o pérdidas para las mismas

A pesar de que en la literatura y en la oferta comercial se encuentran diversas soluciones para el problema de pronósticos de demanda en los negocios, aún se considera en estado de poco desarrollo en el marco colombiano, la falta de literatura al respecto puede ser muestra de esta situación. Usar un único modelo y parámetros para

predecir la demanda de varios productos, es una estrategia arriesgada debido a las particularidades que cada uno posee, y si bien el enfoque propuesto hace uso de un único modelo, tiene en consideración los patrones más relevantes de las series de tiempo, con múltiples parámetros que se basan en las características de la información. Con la realización del presente trabajo se contribuye en la práctica y la academia de manera interdisciplinaria, ilustrando acerca de la implementación de modelos no complejos, como el suavizamiento *Holt*, y técnicas de uso principal en la minería de datos como el *clustering*, en un enfoque novedoso no hallado en la revisión de literatura realizada.

6.3. Trabajos futuros

El trabajo propuesto involucra una serie de características estadísticas conocidas y un algoritmo de *clustering* popular por sus buenos resultados. Aunque los parámetros necesarios para implementar el algoritmo son pocos, la sensibilidad del modelo frente a su elección no ha sido investigada, por tal razón las investigaciones futuras deben incluir diferentes escenarios en comparación con los resultados del modelo. Se resalta la influencia del coeficiente de variación en el proceso de agrupación y también el error del modelo propuesto, futuras investigaciones podrían dedicarse a aislar los efectos de esta variable en el comportamiento estable o errático de una serie de tiempo. En el futuro se propone la implementación de otros métodos de pronósticos de mayor complejidad como los de Box-Jenkins, con el fin de encontrar mayores ventajas o desventajas en su uso, e igualmente se pueden explorar otros algoritmos de clasificación, sin embargo, se resalta que el *clustering* al tratarse de un método de aprendizaje no supervisado se ajusta a las situaciones en donde no se conoce el tipo de clasificación de los datos y debido a que la intención del modelo es que abarque una gama amplia de series de las que no se conoce su naturaleza, esta característica resulta beneficiosa. Sin embargo, una organización en conocimiento de su demanda podría especular en cuanto a la razón por la que las series se agruparon en la forma en que lo hicieron, es posible que los productos de una misma línea se encuentren en un grupo, o aquellos dirigidos a un público que posee características de

consumo similares. Este tipo de validación se considera como una extensión del presente trabajo.

Referencias

- Adshead, N., y Price, D. (1987). Demand forecasting and cost performance in a model of a real manufacturing unit. *International Journal of Production Research*, 25(1), 1251- 1265.
- Aghabozorgi, S., Shirkhorshidi, A. S., y Wah, T. Y. (2015). Time-series clustering–A decade review. *Information Systems*, 53, 16-38.
- Ahmed, N., Atiya, A., Gayar, N., y El-Shishiny, H. (2010). An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, 29(5-6), 594-621.
- Albarune, A., y Habib, D. (2015). A study of forecasting practices in supply chain management. *International Journal of Supply Chain Management*, 4(2), 55-61.
- Aldhyani, T., & Joshi, M. (Septiembre de 2016). Integration of time series models with soft clustering to enhance network traffic forecasting. En *Research in Computational Intelligence and Communication Networks (ICRCICN), 2016 Second International Conference on* (pp. 212-214). IEEE.
- Ali, M., Babai, M., Boylan, J., y Syntetos, A. (2017). Supply chain forecasting when information is not shared. *European Journal of Operational Research*, 260(3), 984-994
- Ariyo, A., Adewumi, A., & Ayo, C. K. (Marzo de 2014.). Stock price prediction using the ARIMA model. En *16th International Conference on Computer Modelling and Simulation (UKSim), 2014 UKSim-AMSS*. (pp. 106-112). IEEE.
- Armstrong, J. S. (2001). *Principles of forecasting: a handbook for researchers and practitioners* (Vol. 30). Springer Science & Business Media.
- Bahl, L., Brown, P., De Souza, P., & Mercer, R. (April de 1986). Maximum mutual information estimation of hidden Markov model parameters for speech recognition. En *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP'86*. (pp. 49-52). IEEE.
- Beck, J. V., & Arnold, K. J. (1977). Parameter estimation in engineering and science. New York: John Wiley & Sons.

- Belalia, Z., & Ghaiti, F. (Mayo, 2016). The impact of three forecasting methods on the value of vendor managed inventory. En *3rd International Conference on Logistics Operations Management (GOL), 2016* (pp. 1-7). IEEE.
- Bon, A., & Ng, T. (Enero, 2017). An Optimization of Inventory Demand Forecasting in University Healthcare Centre. En *IOP Conference Series: Materials Science and Engineering* (Vol. 166, No. 1, p. 012035). IOP Publishing.
- Bowerman, B. L. y O'Connell, R. T. (2007). *Pronósticos, series de tiempo y regresión: un enfoque aplicado*. México: International Thomson Editores.
- Bradley, P. S., Fayyad, U. M., & Reina, C. (Agosto de 1998). Scaling Clustering Algorithms to Large Databases. In *KDD*. (pp. 9-15).
- Brueggen, A., Grabner, I., y Sedatole, K. L. (2014). The Folly of Forecasting: The Effects of Sales Forecast Accuracy and Bias on Inventory and Production Decisions under Aggregated and Disaggregated Forecasting Regimes.
- Chase, R. B., & Jacob, F. R. (2013). *Operations and Supply Chain Management: The Core* (pp. 59). McGraw-Hill Higher Education.
- Chen, K. Y., y Wang, C. H. (2007). Support vector regression with genetic algorithms in forecasting tourism demand. *Tourism Management*, 28(1), 215-226.
- Christou, I. T. (2011). *Quantitative methods in supply chain management: models and algorithms*. Springer Science & Business Media.
- Chu, F. (1998). Forecasting tourism: a combined approach. *Tourism Management*, 19(6), 515-520.
- Crone, S. F., Hibon, M., y Nikolopoulos, K. (2011). Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time series prediction. *International Journal of Forecasting*, 27(3), 635-660.
- Davies, A., y Lahiri, K. (1995). A new framework for analyzing survey forecasts using three-dimensional panel data. *Journal of Econometrics*, 68(1), 205-227.
- Disney, S. M., y Lambrecht, M. R. (2008). On replenishment rules, forecasting, and the bullwhip effect in supply chains. *Foundations and Trends® in Technology, Information and Operations Management*, 2(1), 1-80.

- Edmonds, B. M. (1999). *Syntactic measures of complexity*. Manchester, UK: University of Manchester.
- Everitt, B (1998). *The Cambridge Dictionary of Statistics*. Cambridge, UK New York: Cambridge University Press.
- Farnum, N. (1992). Exponential smoothing: Behavior of the ex-post sum of squares near 0 and 1. *Journal of Forecasting*, 11(1), 47-56.
- Fildes, R. (2001). Beyond forecasting competitions. *International Journal of Forecasting*, 17, 556- 560.
- Flores, B., Olson, D., y Pearce, S. (1993). Use of cost and accuracy measures in forecasting method selection: A physical distribution example. *International Journal of Production Research*, 31(1), 139- 160.
- Gardner Jr., E. S., y McKenzie, E. (1988). Model identification in exponential smoothing. *Journal of the Operational Research Society*, 39, 863- 867.
- Gardner, E. S. (1985). Exponential smoothing: The state of the art. *Journal of forecasting*, 4(1), 1-28.
- Gardner, E. S. (2006). Exponential smoothing: The state of the art—Part II. *International journal of forecasting*, 22(4), 637-666.
- Gilliland, M., Sglavo, U., y Tashman, L. (2016). *Business Forecasting: Practical Problems and Solutions*. Hoboken, New Jersey: John Wiley & Sons.
- Green, K. C., y Armstrong, J. S. (2015). Simple versus complex forecasting: The evidence. *Journal of Business Research*, 68(8), 1678-1685.
- Hans-Hermann, B. (2008). Origins and extensions of the k-means algorithm in cluster analysis. *Journal Electronique d'Histoire des Probabilités et de la Statistique Electronic Journal for History of Probability and Statistics*, 4(2).
- Healy, M. J. (1994). Statistics from the inside. 12. Non-normal data. *Archives of disease in childhood*, 70(2), 158.
- Heizer, J., y Render, B. (2011). *Operations Management* (pp. 113). Prentice Hall.

- Hollingsworth, A., y Lönnberg, P. (1986). The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus A*, 38(2), 111-136.
- Holt, C. (1957). Forecasting seasonals and trends by exponentially weighted moving averages, *ONR Memorandum (Vol. 52)*, Pittsburgh, PA: Carnegie Institute of Technology. Available from the Engineering Library, University of Texas at Austin.
- Holtz-Eakin, D., Newey, W., y Rosen, H. S. (1988). Estimating vector autoregressions with panel data. *Econometrica: Journal of the Econometric Society*, 1371-1395.
- Hyndman, R., Koehler, A., Snyder, R., y Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18(3), 439-454.
- Hyndman, R.J. y Athanasopoulos, G. (2013) *Forecasting: principles and practice*. Recuperado de <http://otexts.org/fpp/2/3>.
- Joshi, M., Lingras, P., Wani, G., & Zhang, P. (2014). Clustering-Based Stability and Seasonality Analysis for Optimal Inventory Prediction. En *Global Trends in Intelligent Computing Research and Development* (pp. 1-18). IGI Global.
- Keogh, E., y Lin, J. (2005). Clustering of time-series subsequences is meaningless: implications for previous and future research. *Knowledge and information systems*, 8(2), 154-177.
- Kristianto, Y., Helo, P., Jiao, J. R., y Sandhu, M. (2012). Adaptive fuzzy vendor managed inventory control for mitigating the Bullwhip effect in supply chains. *European Journal of Operational Research*, 216(2), 346-355.
- Laouafi, A., Mordjaoui, M., Laouafi, F., & Boukelia, T. E. (2016). Daily peak electricity demand forecasting based on an adaptive hybrid two-stage methodology. *International Journal of Electrical Power & Energy Systems*, 77, 136-144.

- Lawrence, M. J., Edmundson, R. H., y O'Connor, M. J. (1986). The accuracy of combining judgemental and statistical forecasts. *Management Science*, 32(12), 1521-1532.
- Lee, T. S., y Adam Jr, E. E. (1986). Forecasting error evaluation in material requirements planning (MRP) production-inventory systems. *Management Science*, 32(9), 1186-1205.
- Lewis-Beck, M. S. (2005). Election forecasting: Principles and practice. *The British Journal of Politics & International Relations*, 7(2), 145-164.
- Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (2008). *Forecasting methods and applications*. John Wiley & sons.
- Makridakis, S., Wheelwright, S., & Cortés Fregoso, H. (2009). *Métodos de pronósticos* (1st ed., pp. 72-73). México: Limusa.
- McCracken, M. W. (2004). Parameter estimation and tests of equal forecast accuracy between non-nested models. *International Journal of Forecasting*, 20(3), 503-514.
- Moradkhani, H., Sorooshian, S., Gupta, H., y Houser, P. (2005). Dual state-parameter estimation of hydrological models using ensemble Kalman filter. *Advances In Water Resources*, 28(2), 135-147.
- Morales, L., y Moura, G. (2013). Adaptive forecasting of exchange rates with panel data. *International Journal of Forecasting*, 29(3), 493-509.
- Muenchen, R. A. (2017). The popularity of data analysis software [Mensaje en un Blog]. Recuperado de <http://r4stats.com/popularity>.
- Nanopoulos, A., Alcock, R., y Manolopoulos, Y. (2001). Feature-based classification of time-series data. *International Journal of Computer Research*, 10(3), 49-61.
- Nelson, H. L., y Granger, C. W. J. (1979). Experience with using the Box-Cox transformation when forecasting economic time series. *Journal of Econometrics*, 10(1), 57-69.
- Pegels, C.C. (1969) Exponential Forecasting: Some New Variations. *Management Science*, 15, 311-315.

- Quintana, R., y Leung, M. T. (2007). Adaptive exponential smoothing versus conventional approaches for lumpy demand forecasting: case of production planning for a manufacturing line. *International Journal of Production Research*, 45(21), 4937-4957.
- Rasmussen, R. (2004). On time series data and optimal parameters, *Omega*, 32, 111-120.
- Ravinder, H. V. (2013). Determining The Optimal Values Of Exponential Smoothing Constants-Does Solver Really Work? *American Journal of Business Education (Online)*, 6(3), 347.
- Rousseeuw, P. J., y Kaufman, L. (1990). *Finding Groups in Data* (pp. 1). Wiley Online Library.
- Sainani, K. L. (2012). Dealing with non-normal data. *PM&R*, 4(12), 1001-1005
- Sale, J. E., Lohfeld, L. H., y Brazil, K. (2002). Revisiting the quantitative-qualitative debate: Implications for mixed-methods research. *Quality and quantity*, 36(1), 43-53.
- Sapankevych, N. I., y Sankar, R. (2009). Time series prediction using support vector machines: a survey. *IEEE Computational Intelligence Magazine*, 4(2).
- Schroeder, R., Rungtusanatham, M. J., y Goldstein, S. (2012). *Operations management in the supply chain* (pp. 261). McGraw-Hill Higher Education.
- Scipy.org. (2001-). Scipy.optimize.least_squares. Recuperado de https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.least_squares.html#id14
- Seifert, M., Siemsen, E., Hadida, A. L., y Eisingerich, A. B. (2015). Effective judgmental forecasting in the context of fashion products. *Journal of Operations Management*, 36, 33-45.
- Shannon, C. E. 1948. "A Mathematical Theory of Communication," *Bell System Technical Journal*, 27(3/4), 379-423.

- Shcherbakov, M. V., Brebels, A., Shcherbakova, N. L., Tyukov, A. P., Janovsky, T. A., y Kamaev, V. A. E. (2013). A survey of forecast error measures. *World Applied Sciences Journal*, 24, 171-176.
- Tarantola, A. (2013). *Inverse Problem Theory*. Burlington: Elsevier Science, p. xi.
- Taylor, J. (2003). Exponential smoothing with a damped multiplicative trend. *International Journal of Forecasting*, 19, 715-725.
- Taylor, J. W., & Bunn, D. W. (1999). A quantile regression approach to generating prediction intervals. *Management Science*, 45, 225- 237.
- Tratar, L. F. (2010). Joint optimisation of demand forecasting and stock control parameters. *International Journal of Production Economics*, 127(1), 173-179.
- Von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and computing*, 17(4), 395-416.
- Wachs, J. P., Stern, H., y Edan, Y. (2005). Cluster labeling and parameter estimation for the automated setup of a hand-gesture recognition system. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 35(6), 932-944.
- Wang, G. C., y Jain, C. L. (2003). Regression analysis: modeling & forecasting. *Institute of Business Forec.* pp. 91.
- Wang, X., Smith, K., y Hyndman, R. (2006). Characteristic-based clustering for time series data. *Data mining and knowledge Discovery*, 13(3), 335-364.
- Webby, R., y O'Connor, M. (1996). Judgemental and statistical time series forecasting: a review of the literature. *International Journal of Forecasting*, 12(1), 91-118.
- Winters, P. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, 6, 324-342.
- Yuan, Y. X. (2000). A review of trust region algorithms for optimization. En *ICIAM* (Vol. 99, pp. 271-282).
- Zelnik-Manor, L., & Perona, P. (2005). Self-tuning spectral clustering. En *Advances in neural information processing systems* (pp. 1601-1608).

Zhang, Z. (1997). Parameter estimation techniques: A tutorial with application to conic fitting. *Image and vision Computing*, 15(1), 59-76.

Zwillinger, D. y Kokoska, S. (2000). *CRC Standard Probability and Statistics Tables and Formulae*. Chapman & Hall: New York. 2000.

Anexo A. Métodos cuantitativos de series de tiempo de complejidad media

A.1. Métodos cuantitativos de series de tiempo simples

En un primer lugar se encuentran los métodos simples, considerados como un punto de referencia o comparación frente a otros, porque las condiciones para que resulten eficientes son difícilmente posibles. Entre estos se encuentran el método ingenuo (*naïve*), en el que se asume que el dato justamente anterior se repetirá, y el método de desviación (*drift*) que es similar al ingenuo con la particularidad de ajustar el valor de acuerdo con el cambio promedio que haya experimentado la información entre un periodo y otro, según la ecuación 8 (Hyndman y Athanasopoulos, 2013). Estos se usan como comparación porque son económicos computacionalmente y para evaluar qué tan efectivos son otros métodos (Armstrong, 2001).

$$y_T + \frac{h}{T-1} \sum_{t=2}^T (y_t - y_{t-1}) = y_T + h \left(\frac{y_T - y_1}{T-1} \right) \quad 8$$

A.2. Regresión lineal

El modelo de regresión lineal simple encuentra la relación de dos variables mediante la siguiente ecuación:

$$y_t = \beta_0 + \beta_1 x + e \quad 9$$

En donde y_t es la variable por predecir, o la variable dependiente, x_t la variable independiente, β_0 es el intercepto con el eje y , es decir cuando $x=0$, y β_1 es el valor de la pendiente, que determina el valor de y por cada unidad de incremento en x . Para pronosticar usando regresión lineal se obtiene un modelo estimado a través de los parámetros de B_0 y B_1 , los cuales se calculan a partir de los valores observados de x .

Aunque el modelo de regresión lineal se puede utilizar para datos de series de tiempo, siendo y el valor a pronosticar y x las observaciones en los diferentes periodos de tiempo. Supone ciertos inconvenientes, ya que los valores futuros de y , son una entrada para estimar el modelo, y deben conocerse con anterioridad. Adicionalmente los datos deberían seguir un comportamiento más o menos lineal para obtener un pronóstico ajustado a la realidad. Sin embargo, se observan en la literatura enfoques

multivariados y combinaciones de métodos para los pronósticos que abordan este problema, Chu (1998), por ejemplo, presenta un enfoque combinado de un modelo de promedio móvil autorregresivo (ARIMA) junto con un modelo de regresión no lineal para predecir las llegadas de turismo internacional en Singapur obteniendo una medida de error (*MAE*) menor, comparada con otros estudios similares.

A.3. Promedios móviles

Es considerado como un método simple, en donde el pronóstico del valor futuro es igual al promedio de los n datos históricos que el administrador o encargado determine, este método también se puede usar para datos transversales y debido a que usa promedios se puede considerar un método de suavización. Mientras más valores históricos se usen, la curva que describen los datos será más suavizada. El método está dado por la siguiente ecuación:

$$y_n = \frac{\sum_{i=1}^n x_i}{n} \quad 10$$

En donde n es el número de valores incluidos en el promedio de y_n . Aunque el método se considera simple, presenta ciertas desventajas. Mientras más grande sea n , mayor será el número de valores que se tienen que almacenar para realizar el pronóstico, por otro lado, al tratarse de un promedio, cada observación tiene el mismo peso o efecto sobre el resultado, tomando en cuenta que las observaciones más recientes pueden tener información más actualizada del comportamiento de los datos, se puede preferir otorgar mayor peso a los valores más recientes (Makridakis, Wheelwright & Cortés Fregoso, 2009). Sin embargo, es usado en la práctica y es la base para otros métodos de suavización a revisar. Publicaciones recientes lo mencionan como referente para estudiar conceptos como el intercambio de información en la cadena de suministro, y el efecto látigo (Ali, Babai, Boylan y Syntetos, 2017) y además es mencionado en literatura relacionada con la gestión de inventarios (Bon & Ng, 2017; Belalia & Ghaiti, 2016).

A.4. Suavización exponencial simple

La suavización exponencial simple (SES) corrige las desventajas mencionadas en el promedio móvil, porque presta mayor importancia a las observaciones más recientes, y debido a que usa el pronóstico del valor anterior no necesita almacenar una cantidad de datos tan grande como en el promedio móvil. El modelo “suaviza” la serie compuesta por datos continuos, de acuerdo con un parámetro alfa α que va de 0 a 1, y es definido por el gerente o administrador, a mayor alfa, menor suavización del comportamiento de los datos. La siguiente ecuación muestra el procedimiento que se sigue para obtener el pronóstico.

$$\hat{y}_{t+h|t} = \alpha y_t + (1 - \alpha)\hat{y}_{t-1} \quad 11$$

Según la ecuación 11 se otorga mayor peso a x_t , es decir a la observación justamente anterior, y un remanente $(1 - \alpha)$ al pronóstico anterior, si alfa es igual a 1, entonces la línea de suavización se convierte en un pronóstico ingenuo, en el caso opuesto, cuando alfa es igual a 0, se convierte en una línea recta, igual al primer pronóstico o al valor de inicialización del modelo. El valor de inicialización es importante cuando los valores de alfa son muy pequeños o cuando la serie es corta porque de ellos puede depender el rendimiento de la serie (Gardner, 2006). Para SES, se consideran varios parámetros como el valor de inicio F_0 , que en general, es igual al último valor que se tiene de la serie, es decir $F_2 = y_1$, porque se asume que la serie podría continuar con un patrón similar al del último período. El valor alfa, es determinado por el usuario, aunque existen métodos de optimización del método para obtener estos valores (Bowerman y O’Connell, 2007; Rasmussen, 2004; Farnum, 1992; Ravinder, 2013). El método de SES es particularmente útil en series de tiempo que no presentan tendencia, ni estacionalidad, es decir son horizontales y solo tienen nivel y han sido usados ampliamente en la gestión de la cadena de suministros (Flores, Olson, y Pearce (1993); Adshead y Price, 1987).

A.5. Suavización exponencial lineal de Holt o doble suavización

Para extender el método de suavización a series con componentes de tendencia, Holt en 1957 hizo público su desarrollo sobre el suavizamiento exponencial doble

(Holt, 1957), que usa dos filtros para obtener estimaciones del factor horizontal y de la tendencia lineal de la serie, a partir de estas se puede obtener un pronóstico de un periodo hacia adelante. El pronóstico se obtiene a través de la ecuación 12:

$$y_{t+h|t} = \ell_t + hb_t \quad 12$$

Que a su vez es obtenido a través de las ecuaciones 13 y 14:

$$\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + b_{t-1}) \quad 13$$

$$b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1} \quad 14$$

El enfoque es muy parecido al SES, sin embargo, toma en cuenta la pendiente de la serie $(\ell_t - \ell_{t-1})$ y la incluye en la ecuación 14 del SES, para obtener la tendencia b_t . En este caso el pronóstico del periodo $t+h$ es una ecuación lineal del nivel ℓ_t más la tendencia para el horizonte de tiempo h . Estas estimaciones se consiguen, como en el método anterior a través de los parámetros de suavizado, *alfa* para el nivel, y *beta* para la tendencia. Estos pueden ser hallados a través de métodos de optimización, minimizando las medidas de error (*suma de los cuadrados de los errores*, *SSE* por sus siglas en inglés y *desviación media absoluta* o *MAD*, entre otras medidas).

A.6. Suavización exponencial doble con tendencia exponencial

Puede darse la situación de contar con una serie que muestra una tendencia exponencial y no lineal, para ello se hace una corrección a las ecuaciones originales, de manera que se estime un comportamiento aproximadamente exponencial, en donde la tasa de crecimiento es constante y no su pendiente como en el caso anterior. Las siguientes ecuaciones permiten hallar el pronóstico.

$$y_{t+h|t} = \ell_t b_t^h \quad 15$$

$$\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + b_{t-1}) \quad 16$$

$$b_t = \beta \frac{\ell_t}{\ell_{t-1}} + (1 - \beta)b_{t-1} \quad 17$$

En este caso la pendiente es calculada como una razón entre el nivel del periodo actual t y un nivel anterior, y se multiplica en lugar de sumarse (Hyndman y Athanasopoulos, 2013).

A.7. Suavización exponencial doble con tendencia amortiguada

Gardner & McKenzie (1988), añaden a la suavización lineal de Holt un parámetro *fi de* amortiguamiento junto a la tendencia b_t , siguiendo la ecuación 18. Este parámetro oscila entre 0 y 1, tal como alfa y beta, cuando es igual a 1, el modelo es igual al de suavizamiento doble, cuando es 0 es equivalente al SES.

$$\hat{y}_{t+h|t} = \ell_t + (\phi + \phi^2 + \dots + \phi^h)b_t \quad 18$$

$$\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1}) \quad 19$$

$$b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)\phi b_{t-1} \quad 20$$

Fi amortigua gradualmente la tendencia hasta llegar a una asíntota, es decir, hasta alcanzar un valor constante. El objetivo es que la tendencia se ajuste tanto a demandas con una tendencia evidente, como a aquellas en donde se observa un comportamiento errático. Lo anterior se presenta en la literatura como tendencia amortiguada aditiva, para el caso multiplicativo Taylor (2003), añade un parámetro a las ecuaciones de tendencia exponencial, traduciéndolas en las ecuaciones siguientes:

$$\hat{y}_{t+h|t} = \ell_t b_t^{(\phi + \phi^2 + \dots + \phi^h)} \quad 21$$

$$\ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1} b_{t-1}^\phi \quad 22$$

$$b_t = \beta \frac{\ell_t}{\ell_{t-1}} + (1 - \beta)b_{t-1}^\phi \quad 23$$

Los modelos de tendencia amortiguada suponen formulaciones adicionales en cuanto a su optimización, situación que se incrementa con modelos estadísticos para determinar intervalos de predicción, en donde los tiempos de computación son muy altos (Hyndman, Koehler, Snyder y Grose, 2002), y un administrador novicio puede encontrarlo abrumador.

A.8. Suavización exponencial de *Holt-Winters*

Los métodos de suavización anteriores fueron creados para series con presencia de nivel y tendencia, sin tener en cuenta en la predicción el componente estacional de la serie. Winters, en 1960, extendió el trabajo del modelo de Holt, para series de

tiempo cuyo comportamiento también exhibiera factores estacionales. Esto da lugar a una nueva ecuación asociada al factor estacional s_t (Ecuación 27) con un parámetro $gamma$ (γ), y otras expresiones derivadas de los dos métodos anteriores. Para este caso se considera también el parámetro m , que representa el periodo de estacionalidad, cuyo valor depende de la naturaleza del producto y el mercado local, entre otras características. Si la estacionalidad es mensual entonces $m=12$, debido a que existen 12 estaciones dentro de un año, si es trimestral $m=4$, por la misma razón. Este método presenta dos variantes principales que nacen debido a la naturaleza de los datos: aditiva y multiplicativa, y son derivadas de los métodos anteriores.

Holt-winters aditivo

$$\hat{y}_{t+h|t} = \ell_t + hb_t + s_{t-m+h_m^+} \quad 24$$

$$\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1}) \quad 25$$

$$b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1} \quad 26$$

$$s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m} \quad 27$$

Holt-winters multiplicativo

$$\hat{y}_{t+h|t} = (\ell_t + hb_t)s_{t-m+h_m^+} \quad 28$$

$$\ell_t = \alpha \frac{y_t}{s_{t-m}} + (1 - \alpha)(\ell_{t-1} + b_{t-1}) \quad 29$$

$$b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1} \quad 30$$

$$s_t = \gamma \frac{y_t}{(\ell_{t-1} + b_{t-1})} + (1 - \gamma)s_{t-m} \quad 31$$

En cuanto a los métodos de suavización exponencial, Pegels (1969) presentó una taxonomía que fue extendida luego por varios autores (Gardner, 1985; Taylor, 2003), en donde se presentan las diferentes derivaciones de los modelos de suavización.

A.9. Método de descomposición por componentes

Los métodos de suavizamiento usan el promedio para reducir el efecto aleatorio de las series y estimar un pronóstico para el patrón subyacente, sin llegar a la separación de estos (Makridakis, Wheelwright, & Cortés Fregoso, 2009). La descomposición clásica de series de tiempo separa dichos patrones y ayuda a entender el comportamiento de cada uno de ellos. Como se explica en el numeral 3.4, por lo general, se asume que las series poseen patrones o componentes que ayudan a describir su comportamiento: nivel, tendencia, ciclo y estacionalidad. El nivel o factor horizontal es común a todas las series de tiempo, de tal forma que los métodos de descomposición tratan con los tres factores restantes, cuya presencia puede variar. Estos métodos han sido usados desde la década de 1920, por economistas y administradores y aunque su base es empírica, han tenido buena aceptación debido a su planteamiento y resultados, los cuales son considerados prácticos para la planeación de operaciones (Makridakis, Wheelwright, & Cortés Fregoso, 2009). Para el planteamiento inicial del modelo se debe considerar el comportamiento de los datos, aditivo ($S+C+T+R$) o multiplicativo ($S \times C \times T \times R$), a manera de ejemplo se presentan los pasos, suponiendo un comportamiento:

1. **Una estimación inicial de tendencia secular (T y C), eliminando estacionalidad (S) y aleatoriedad (R).**

Se obtiene el promedio móvil del tamaño de la estacionalidad m (Si los períodos son trimestrales, $m = 4$, si son meses, $m = 12$, etc.). Este proceso pretende reducir o suavizar las fluctuaciones de los datos respecto a la estacionalidad, y por la definición de promedio también se reduciría la aleatoriedad, porque este fluctúa entre 0). Este promedio móvil es equivalente a extraer los componentes S y R, del resto, por tanto, cuando se realiza la serie resultante representa a los componentes de $T \times C$, en conjunto. En caso de que m sea par, es necesario realizar un nuevo promedio móvil centrado, tomando dos periodos, esto para garantizar que el valor refleje de manera simétrica el verdadero promedio de la serie.

2. Una estimación de la estacionalidad combinada con aleatoriedad cuando se extraen los componentes $T \times C$.

El conjunto de datos resultante del paso anterior es dividido sobre los datos originales, esto es:

$$\frac{S \times C \times T \times R}{T \times C} = S \times R \quad 32$$

De esta forma se obtienen la estacionalidad y la aleatoriedad de forma individual ($S \times R$).

3. Una estimación de índices de estacionalidad, reduciendo el componente aleatorio.

A partir de $S \times R$, se promedia cada periodo con su equivalente en cada año, en otras palabras, si se habla de meses sería, promedio de los enero, promedio de los febreros, y así sucesivamente, con el fin de hallar los índices estacionales. Estos índices deben ser normalizados, por lo que son ajustados de manera que la suma de 12 índices sea igual a 12 (número de estaciones dentro de un año). Así se asegura que los factores estacionales no capturen información residual de los ciclos, es decir información de más de un año y se reduce el error. Estos factores deben indicar las proporciones de aumento y disminución de cada estación de acuerdo con su tendencia media.

4. Dichos índices son extraídos de los datos originales, y de este modo se obtiene una serie ajustada estacionalmente o desestacionalizada.

Para desestacionalizar los datos o “ajustarlos estacionalmente”, se extraen los índices estacionales de la serie original, para esto, debido a que se trata de un modelo multiplicativo, se dividen los datos originales y los índices estacionales normalizados obteniendo una serie “sin componente estacional”.

$$\frac{S \times C \times T \times R}{S} = C \times T \times R \quad 33$$

5. Obtener el componente de tendencia de forma paramétrica para obtener un pronóstico.

Para poder presentar un pronóstico es necesario estimar la tendencia, pero de una forma paramétrica. Para ello la descomposición clásica puede valerse de métodos como regresión lineal o suavización exponencial de Holt, y luego “revertir” el proceso de descomposición multiplicando los índices estacionales en cada estación correspondiente.

Anexo B. Definición de características estadísticas extraídas de las series

B.1. Media aritmética

O promedio de un conjunto de números x_1, x_2, \dots, x_n es la sumatoria de estos dividido por el número de observaciones. La media es una medida de tendencia central, porque resume de cierta forma el comportamiento de la información y sus propiedades permiten realizar supuestos en cuanto a la suavización de los datos (Everitt, 1998).

- Si un conjunto como x_1, x_2, \dots, x_n tienen media \bar{x} , entonces $(x_1 - \bar{x}) + \dots + (x_n - \bar{x}) = 0$. En otras palabras, los números a la izquierda del promedio se balancean a la derecha del promedio.
- Su formulación está dada por:

$$\frac{1}{n} \sum_{i=1}^n x_i \quad 34$$

B.2. Mediana

La mediana, otra medida de tendencia central, es el valor que separa la mitad más grande de los datos de más baja cuantía. La mediana puede resultar una medida de resumen adecuada de los datos cuando la media aritmética es asimétrica hacia un lado, pues la mediana es una medida más robusta en cuanto a valores muy pequeños o muy grandes. Cuando una distribución es simétrica, la mediana y la media son similares, sin embargo, esto no es necesariamente verdadero en la dirección contraria (Zwillinger y Kokoska, 2000).

B.3. Desviación Estándar

La desviación promedio de los datos respecto a su media, es una métrica común que resume la dispersión de los datos. La desviación estándar es la raíz cuadrada positiva de la varianza con la ventaja de que se encuentra en las mismas unidades que los datos. Esta dada por la siguiente formulación (Zwillinger y Kokoska, 2000):

$$\sigma = \sqrt{\frac{\sum_i^N (x_i - \bar{x})^2}{N}} \quad 35$$

B.4. Coeficiente de Variación

El coeficiente de variación es una medida de dispersión estandarizada de la distribución de un conjunto de datos. Se obtiene a través de la proporción o razón de la desviación estándar y la medida. Esta dado por la siguiente fórmula generalizada:

$$cv = \frac{\sigma}{\mu} \quad 36$$

Debido a que se trata de una medida sin dimensiones, a menudo interpretada como un porcentaje, es más útil que la desviación estándar en la comparación de la volatilidad de conjuntos de datos cuando están medidos en diferentes unidades porque no depende de estas, o cuando las medias entre estos son muy disimiles (Everitt, 1998).

B.5. Asimetría

Mide la asimetría de una distribución de probabilidad y puede tener valores negativos, o positivos. Una asimetría negativa indica que la función de densidad de probabilidad a la izquierda es más larga o más amplia, y no hace distinción entre estos dos tipos de forma, análogamente se presenta esta situación para una asimetría positiva en el lado derecho de la distribución. Aunque el cálculo de la asimetría puede variar para una distribución de probabilidad específica, en este caso se usa el coeficiente de asimetría definido por la siguiente ecuación (Zwillinger y Kokoska, 2000):

$$g_1 = \frac{m_3}{m_2^{3/2}} \quad 37$$

En donde m_r es el momento r-ésimo alrededor de la media definido por:

$$m_r = \frac{1}{n} \sum_{i=1}^k (x_i - \bar{x})^r \quad 38$$

B.6. Entropía

La entropía está definida comúnmente como una señal de caos o desorden, principalmente en el área de la termodinámica. Sin embargo, algunos enfoques dirigidos a mecánica estadística y teoría de la información han permitido usar este concepto para medir la incertidumbre en una fuente de información. Claude Shannon, conocido como el padre de la teoría de la información desarrolló las siguientes formulaciones para la entropía entendida como la incertidumbre en la cantidad de información que produce una variable aleatoria discreta.

$$H(x) = -K \sum_{i=1}^n p_i \log_2 p_i \quad 39$$

En donde K es una constante positiva, usada como una opción de unidad de medida y p_i es la probabilidad de que un evento i ocurra en este conjunto de datos, que en el caso de este proyecto está definida por la distribución discreta de los datos. Contrario al concepto inicial de entropía, en el caso de la entropía de Shannon, una entropía “alta”, significa que x proviene de una distribución “uniforme” o estable, y un valor “bajo”, significa que proviene de una distribución muy variable, con picos y valles (Shannon, 1948).