

ANÁLISIS SIMULTÁNEO DE LOS INGRESOS DE MÚLTIPLES AGENCIAS GUBERNAMENTALES

Otoniel Arruda Costa



Tribunal de Cuentas del Estado de São Paulo (TCE/SP), São Paulo (SP), Brasil.

Henrique Raymundo Góia



Universidade de São Paulo (USP), São Paulo (SP), Brasil.

RESUMEN

La planificación presupuestaria es obligatoria por ley, pero muchos gestores aún enfrentan desafíos en la previsión de ingresos. A pesar de los avances en ciencia de datos, hay lagunas en la aplicación de modelos predictivos para este fin. Este estudio propone un modelo computacional en R que permite analizar series temporales de ingresos presupuestarios de múltiples órganos recaudadores simultáneamente, comparando diferentes métodos – Media Móvil Integrada Autorregresiva, en inglés Autoregressive Integrated Moving Average (ARIMA) y Suavizado Exponencial – de forma automatizada. El algoritmo desarrollado posibilita que los analistas verifiquen toda la información de los métodos con un único código, eliminando la necesidad de utilizar las mismas funciones repetidamente. El algoritmo fue desarrollado con datos de los portales del Tribunal de Contas del Estado de São Paulo (TCE-SP) y de la Fundação Sistema Estadual de Análise de Dados (Seade). Se seleccionaron dos ciudades con características económicas similares al 31 de diciembre de 2020, Sorocaba y São José dos Campos, sobre las cuales se generaron modelos de previsión de series temporales de los ingresos de los ejercicios de 2022 y 2023. Como metodología, se compararon los modelos por medio del Criterio de Información de Akaike – Akaike Information Criterion (AIC), de la versión corregida del AIC – Corrected Akaike Information Criterion (AICc) y del Criterio de Información Bayesiano (BIC), además de la estadística de Error Porcentual Absoluto Medio (MAPE). El método ARIMA presentó el mejor criterio de información para ambas ciudades y el

menor MAPE en São José dos Campos (9,75 %). Con el método de Suavizado Exponencial, Sorocaba tuvo el menor MAPE (12,31 %). Los resultados demuestran la viabilidad del modelo propuesto para facilitar análisis comparativos entre municipios y apoyar decisiones de gestores públicos, auditores y otros interesados en las informaciones económicas gubernamentales.

Palabras clave: business analytics; covid-19; Fable; previsión presupuestaria; X13-ARIMA-SEATS.

1 INTRODUCCIÓN

Uno de los factores cruciales para el equilibrio entre ingresos y gastos públicos – equilibrio fiscal – es la adopción de un proceso adecuado de planificación presupuestaria. Se trata de una constatación que se obtiene de la lectura de los arts. 11 y 12 de la Ley Complementaria 101, de 4 de mayo de 2000 – Ley de Responsabilidad Fiscal (LRF).

Para cumplir este objetivo, los entes federativos deben elaborar una planificación gubernamental a mediano plazo mediante el Plan Plurianual (PPA), instituido por ley ordinaria y con vigencia de cuatro años. En consonancia con este instrumento de planificación, están obligados a promulgar, anualmente, la Ley de Directrices Presupuestarias (LDO) y la Ley Presupuestaria Anual (LOA), que orientan la elaboración y ejecución del presupuesto público.

Estudios recientes demostraron que aún existen dificultades en la previsión presupuestaria. Oliveira Junior y Levino (2024) destacaron que la Unión prestó auxilio financiero a los municipios durante la pandemia, sin un análisis técnico robusto, evidenciando la necesidad de modelos predictivos más precisos para respaldar decisiones fiscales en escenarios inciertos.

En otro estudio, Nascimento y Boente (2022, p. 11), al estudiar los más de cinco mil municipios brasileños entre 2015 y 2018, demostraron que los errores de previsión de

ejercicios fiscales pasados (en promedio, 19,7 %) afectan los errores de previsión de ejercicios futuros (en promedio, 17,5 %), al alza o a la baja. Toledo Júnior (2010, p. 36), por su parte, aseveró que, muchas veces, los ingresos presupuestarios se ajustan al “tamaño deseado del gasto”.

Por otro lado, afortunadamente, en la actualidad, las ciencias de datos proporcionan herramientas cada vez más poderosas en análisis, incluso para el sector público. Fávero y Belfiore (2021, p. 14) destacaron que las aplicaciones profesionales “además de ofrecer una enorme capacidad de procesamiento de bases de datos, son capaces de elaborar las más diversas pruebas y modelos apropiados y robustos”. Un ejemplo reciente de esta aplicación es el estudio de la similitud de la evolución del covid-19 en los estados brasileños, conducido por Faveret *et al.* (2021). Los autores utilizaron las técnicas de agrupamiento de datos y análisis de series temporales para examinar patrones de ocurrencia de esta enfermedad. En otra aplicación, el gobierno del estado de Mato Grosso (Sá *et al.*, 2020) preconizó el uso de redes neuronales como metodología predictiva.

A pesar de los avances, el análisis de datos aún es subutilizado por los gestores públicos. Araújo, Behr y Schiavi (2023) señalaron que muchas herramientas utilizadas por los auditores no incorporan técnicas modernas de análisis de datos, limitando la precisión de las previsiones. Además, los autores también señalaron que el encapsulamiento de las herramientas de análisis de datos disponibles hace que los profesionales desconozcan el algoritmo en uso. Convergen con estas constataciones las conclusiones de las investigaciones de Batóg y Batóg (2021) y Liu, Liu y Chen (2023), quienes propusieron la ampliación del estudio de métodos, técnicas y algoritmos predictivos.

Considerando este panorama, el objetivo de la investigación es proponer un modelo que facilite el análisis de múltiples series temporales de órganos recaudadores por medio de un algoritmo computacional en lenguaje R. La finalidad del modelo es permitir que los usuarios de este tipo de información económica obtengan ahorro de tiempo en la detección de la técnica más adecuada para diferentes series temporales. El estudio también pretende contribuir al aumento de la capacidad de comparabilidad

entre instituciones y a la obtención de economías de escala en la cantidad de datos examinados.

Supóngase el caso hipotético de un técnico presupuestario que compare la serie temporal de los ingresos presupuestarios de la alcaldía en la que actúa con la de otro municipio de características económicas similares. En esta situación, las previsiones generadas por el algoritmo podrían servir como punto de partida para investigar, por ejemplo, si el otro ente se estuviera beneficiando de alguna legislación o política pública que también pudiera aplicarse a su municipio, en caso de que se identificaran diferencias significativas en la evolución de los ingresos presupuestarios para un mismo período.

Otro caso sería el de un auditor que, al comparar los ingresos de dos órganos o entes con características similares, pudiera identificar indicios de irregularidades o fraudes a partir de la detección de patrones atípicos de recaudación de ingresos.

En esta perspectiva, la solución propuesta busca proporcionar a los diversos interesados en las informaciones económico-financieras medios para acompañar, de forma proactiva, la evolución futura de los ingresos públicos, utilizando estas previsiones como herramienta de apoyo a la gestión, al control y a la toma de decisiones.

Entre los posibles interesados en la solución, se destacan los órganos de control, los órganos de ingresos públicos, las instituciones de investigación, los gestores públicos, los ciudadanos y otros interesados en el seguimiento de la gestión fiscal.

2 REVISIÓN DE LITERATURA

Esta sección está subdividida en tres subsecciones, para la presentación de la literatura sobre los conceptos de ingresos públicos, series temporales y procesos y técnicas predictivas.

2.1 Ingresos públicos

La planificación presupuestaria constituye, según el Manual de Contabilidad Aplicada al Sector Público (MCASP), un aspecto relevante de la gestión pública (Brasil, 2024). Sin embargo, incluso ante los avances tecnológicos y metodológicos, este sigue siendo un trabajo desafiante para los gobiernos. Diversos países enfrentan dificultades en este proceso y buscan alternativas para mejorar la previsión de los ingresos públicos.

Batóg y Batóg (2021), al analizar los factores de riesgo relacionados con la planificación de ingresos regionales en Polonia, concluyeron que los procesos de previsión exitosos no solo deben basarse en la estructura de ingresos, sino también en la selección de un método de previsión adecuado. En el estudio, los autores propusieron dos modelos: uno basado en series temporales y otro basado en la recaudación como una variable dependiente del Producto Interno Bruto (PIB).

Koniagina (2020) analizó problemas de implementación de soluciones digitales para los procesos de previsión de ingresos en gobiernos locales de la Federación Rusa. Entre los principales obstáculos para la adopción de algoritmos de previsión y planificación, la autora destacó la “ausencia de una herramienta única de planificación y previsión fiscal, integrada en el sistema consolidado de planificación y previsión, para los niveles regional y local” (Koniagina, 2020, p. 6).

En Estados Unidos, Levy, Chasalow y Riley (2021) analizaron el empleo de algoritmos en el proceso de toma de decisiones de la gestión pública bajo diferentes perspectivas, incluyendo las dimensiones computacional, jurídica, social y ética. Los autores observaron que las restricciones impuestas por los gobiernos estatales a los ingresos de los gobiernos locales pueden limitar la capacidad financiera de estos entes y comprometer su actuación. En este contexto, argumentan que la adopción de herramientas tecnológicas se convierte en una opción considerable para subsidiar la toma de decisiones complejas en el ámbito de las políticas públicas.

En España, Muñoz, Bolívar y Arellano (2022) destacaron que las dificultades enfrentadas por los gobiernos regionales y locales en el control de las finanzas públicas favorecieron la adopción de soluciones tecnológicas orientadas al perfeccionamiento de la gestión fiscal y del proceso de toma de decisiones.

En Brasil, el estado de Mato Grosso incorporó, en su Metodología utilizada en las estimaciones de ingresos y gastos (Sá *et al.*, 2020), el concepto de red neuronal para la estimación de los ingresos públicos. De acuerdo con el documento, en este método, los valores de la serie temporal son los datos de entrada (variables independientes o regresores) a los cuales se les asignan pesos ponderados automáticamente por el algoritmo durante el proceso de entrenamiento. En cada nueva iteración, los parámetros del modelo se recalibran con base en los errores observados en las simulaciones anteriores, lo que permite el perfeccionamiento de las previsiones.

Más allá de los desafíos tradicionales en el proceso de previsión de los ingresos de órganos recaudadores, la situación de calamidad internacional en salud pública derivada del covid-19 trajo nuevos desafíos. El principal de ellos fue la gran variabilidad de los datos, notablemente en los años 2020 y 2021, derivada de la paralización parcial de la actividad económica en todo el mundo. En este sentido, Oliveira Junior y Levino (2024), al aplicar la prueba de Correlación de Spearman, mostraron que la transferencia de auxilio financiero por parte de la Unión a los municipios afectados por el covid-19 no presentó una correlación evidente con la caída de recaudación en aquellos entes, lo que perjudicó el análisis técnico en el proceso de descentralización de recursos a los entes subnacionales. Por su parte, Nascimento y Boente (2022, p. 11) aseveraron que es necesaria la adopción de técnicas predictivas “más realistas” por los municipios. También propusieron que futuras investigaciones consideren externalidades, como períodos de crisis económicas.

2.2 Series temporales

El establecimiento de un modelo predictivo presupone el entendimiento del concepto de series temporales. Una definición adecuada para ellas es “cualquier conjunto de observaciones ordenadas en el tiempo”. Según Morettin y Toloí (2018, p. 1), hay más

aspectos interesantes en la realización de series temporales, como describir el comportamiento del ingreso/gasto e investigar sus mecanismos generadores, que en la realización solo de previsiones.

Uno de los métodos ampliamente utilizados para la previsión de series temporales es el Suavizado Exponencial, en inglés Error, Trend and Seasonal (ETS), cuya principal característica consiste en reducir el impacto de las oscilaciones aleatorias presentes en los datos mediante la asignación de un mayor peso a las observaciones más recientes. Para los autores (2018, p. 89), las oscilaciones presentes en los extremos representan fuentes de “aleatoriedad”, lo que justifica el uso frecuente del método, por su “simplicidad, eficiencia computacional y razonable precisión”. Los modelos ARIMA, según Sartoris (2003), se basan en datos pasados de la propia serie, es decir, funcionan como una regresión de sí mismos.

Morettin y Tolo (2018) clasifican el Suavizado Exponencial en tres modalidades: la simple, indicada para series que no poseen tendencia ni estacionalidad; la de Holt, recomendada para series con tendencia y sin estacionalidad; y la de Holt-Winters, apropiada para series que presentan tendencia y estacionalidad simultáneamente.

Ferreira y Duca (2020, p. 75) explican que la tendencia es el crecimiento o decrecimiento normalmente existente en una serie, “no necesariamente lineal”. También instruyen que la estacionalidad es la repetición de valores en períodos anuales.

Otro concepto importante es el de la estacionariedad de las series temporales. Sartoris (2003, p. 342) conceptúa que una serie temporal dada por Y_t es estacionaria cuando su valor histórico tiende a cero, de modo que un eventual choque se “disipa a lo largo de los años”. El autor también preconiza que, en este tipo de serie, la esperanza y la varianza matemáticas son constantes. Por el contrario, cuando hay un componente de error sumado al valor del período anterior de la serie, se tiene el llamado paseo aleatorio (random walk).

Para Morettin y Tolo (2018, p. 5), “una de las suposiciones más frecuentes que se hace respecto a una serie temporal es la de que es estacionaria, es decir, que se

desarrolla en el tiempo aleatoriamente alrededor de una media constante”. Para los autores, las series temporales generalmente presentan tendencias, cuando se refieren a series de datos económicos y financieros. Resaltan, además, que la mayor parte de los métodos posee la premisa de que los patrones de comportamiento de las series temporales se basan en datos pasados. Además, advierten que las series no estacionarias necesitan sufrir un proceso de transformación a la estacionariedad, conocido como diferenciación. Para ello, con el fin de detectar la existencia de esta no estacionariedad (tendencias deterministas o estocásticas con raíces unitarias), los autores enumeran diversas pruebas. Una de ellas es la de Dickey-Fuller, que utiliza el método de los mínimos cuadrados para detectar raíces unitarias y remover las tendencias de las series que, cuando son estocásticas, se realizan mediante diferenciaciones.

En la literatura hay otros dos puntos de preocupación en el análisis de series temporales. El primero es la necesidad de estabilización de su varianza, para lo cual “una transformación logarítmica puede ser adecuada” (Morettin; Toloí, 2018, p. 9). El otro punto a ser examinado es si los datos están apropiadamente representados por el modelo (Souza *et al.*, 2022). Para ello, se aplica la prueba de sobreajuste. De acuerdo con Morettin y Toloí (2018), la técnica consiste en añadir parámetros al modelo. Si no hay reducción de la varianza de los residuos, el modelo encontrado inicialmente es el más adecuado. Alternativamente a esta prueba, es posible la utilización de la técnica de detección de correlaciones entre los residuos, como la prueba de Ljung-Box, descrita por Ferreira y Mattos (2020). En relación con la unidad que mide la diferencia entre los valores previstos y los reales, se destaca el MAPE, conforme fue discutido por Kim y Kim (2016).

Para el presente estudio, se adoptó la Escala de Precisión de la Previsión presentada por Klimberg, Sillup, Boyle y Tavva (2010). En ella, el MAPE por debajo del 10 % posibilita concluir que la previsión es altamente precisa; entre 10 % y 20 %, se dice que la previsión es buena; y entre 21 % y 50 %, la previsión se considera razonable. Por su parte, un MAPE por encima del 50 % indica un modelado sin precisión.

2.3 Procesos y técnicas predictivas

El proceso de previsión de ingresos basado en series temporales está cada vez más vinculado a la evolución tecnológica. En los últimos años, un concepto que ha ganado fuerza en el análisis de datos es el de Business Analytics (BA), entendido como la aplicación de métodos estadísticos y matemáticos para extraer conocimiento de los datos y subsidiar el proceso de toma de decisiones.

Según Araújo, Behr y Schiavi (2023, p. 2), este concepto fue desarrollado “para representar el principal componente analítico del Business Intelligence [BI]”, mientras que el BI está más relacionado con la producción y la disponibilidad de informes destinados a apoyar la toma de decisiones.

En la investigación sobre la literatura en BA, Liu, Liu y Chen (2023) identificaron cuatro perspectivas para la definición de este concepto: técnica, de procesos, práctica y de gestión. Para este estudio, se destacan las dos primeras, por ser las que mejor se relacionan con el desarrollo y la aplicación de algoritmos de previsión de ingresos públicos.

Desde el punto de vista de la técnica, se consideran BA cualesquiera aplicaciones en ciencia o análisis de datos. Ya bajo el prisma del proceso, los autores definen BA como el “encapsulamiento de herramientas para convertir datos en insights accionables por medio de un proceso científico/matemático/inteligente”.

Desde el punto de vista técnico, Hyndman y Rostami-Tabar (2024) analizaron estrategias de series temporales con cambios sustanciales en las series, como la interrupción de datos en cierto período de tiempo. Los autores propusieron soluciones cuando ocurren, por ejemplo, intervenciones gubernamentales específicas en determinados períodos de tiempo; eventos disruptivos, como la situación de la pandemia del covid-19; o, incluso falta de datos. Para ello, y con base en datos reales de turismo en Australia, entre 2019 y 2022, analizaron, entre otros, los siguientes tipos de modelado: adaptativo; conjunto; de intervención; o lo que denominaron estimación de lo que podría ser (Hyndman; Rostami-Tabar, 2024, p. 7). Los autores concluyeron,

al final del estudio, que cada situación puede demandar una solución distinta. Como ejemplo, para una intervención (gubernamental) compleja, es más indicado el método adaptativo. Ya el enfoque conjunto es útil “cuando hay incertidumbre sobre qué enfoque usar y generalmente resulta en una mayor precisión debido al poder del promedio” (Hyndman; Rostami-Tabar, 2024, p. 19). En el presente estudio, las series examinadas no poseían datos interrumpidos, ni siquiera en el período de situación de emergencia del covid-19.

Aún en relación con la pandemia, una herramienta importante en el proceso de modelado fue el programa de ajuste estacional X13-ARIMA-SEATS (Sax; Eddelbuettel, 2018). Esta aplicación fue desarrollada en 2012 por la U.S. Census Bureau (Mattos; Ferreira, 2020) y trabaja en conjunto con el paquete Seasonal, en lenguaje R. Uno de los puntos fuertes del programa, según estos autores, es la herramienta de preajuste, que permite el modelado de las series con el tratamiento de datos relativos a días hábiles, feriados y puntos extremos de la serie (outliers), entre otros aspectos. Antes de los ajustes automáticos, los usuarios pueden refinar el modelado con la posibilidad de ajustes manuales, ya que hay diferenciaciones en el calendario de cada país y en cada año.

También sobre el X-13-ARIMA-SEATS, se destaca el estudio de Abeln y Jacobs (2022), quienes investigaron el impacto de la pandemia en el PIB de los Países Bajos. En la investigación, analizaron esta crisis como un outlier en el contexto del análisis de datos. En Brasil, el programa X13-ARIMA-SEATS es utilizado por varios órganos que trabajan con índices oficiales, como el Banco Central do Brasil (Brasil, 2021) y el Ministerio de Economía (Brasil, 2020).

3 METODOLOGÍA

En el presente estudio, se adoptó como metodología, en cuanto a su naturaleza, la investigación aplicada, ya que el objetivo es su utilización en la práctica (Gerhardt; Silveira, 2009). En cuanto al propósito, el trabajo es una investigación descriptiva, pues examina la aplicación de un algoritmo para la realización del proceso de

comparación de los ingresos de dos municipios distintos. En cuanto al enfoque, la investigación es cuantitativa, pues los fenómenos examinados se refieren a la aplicación estadística sobre datos numéricos. En cuanto al procedimiento metodológico, se trata de un estudio de caso, ya que se busca, a partir de la elaboración y la aplicación de un algoritmo, solucionar un problema de comparación de datos presupuestarios.

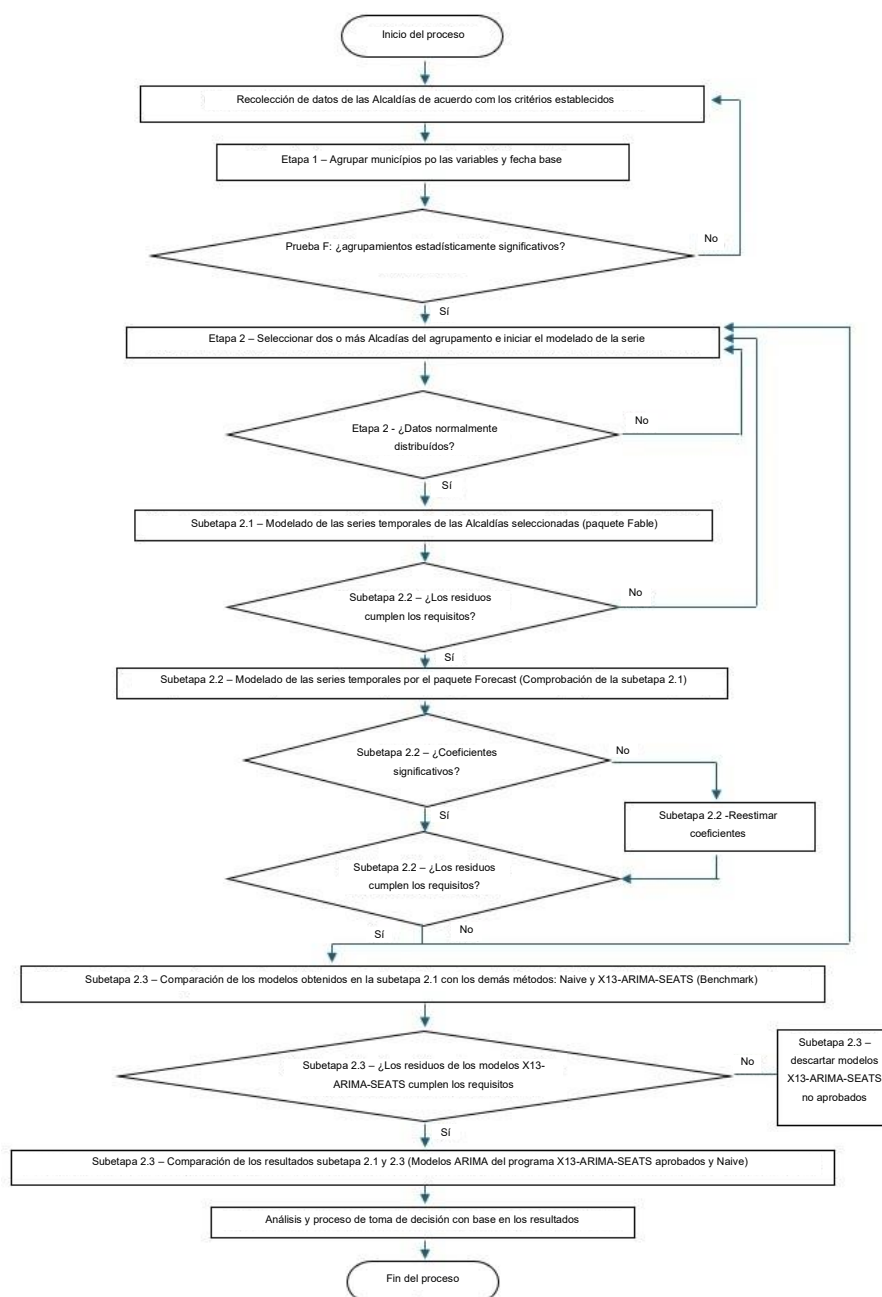
Para la organización lógica del contenido y la comprensión gradual de los procesos realizados, la sección fue subdividida en las subsecciones “caracterización del sujeto de la investigación”, “recolección de datos”, “análisis de datos”, según la propuesta de Volpato (2015), con adaptaciones.

3.1 Caracterización del sujeto de la investigación

El objeto de la investigación son los municipios brasileños del estado de São Paulo (SP). Para obtener recursos, estos entes obedecen a la Ley 4.320, de 17 de marzo de 1964, y a la LRF. Por ello, deben promulgar, anualmente, la LDO y la LOA, normativas que guían el reconocimiento de ingresos y la realización de gastos. A pesar de que el estudio tiene como enfoque los municipios, la metodología puede ser aplicada a otros entes. En esta investigación, se consideró, en la selección de las muestras, el concepto de ente federativo municipal, según la Constitución Federal de 1988. Por lo tanto, también se incluyeron como órganos recaudadores los institutos de previsión, las autarquías, entre otros entes públicos municipales que, por ley, poseen capacidad recaudatoria en el ámbito local. Así, la municipalidad es el conjunto de todos estos órganos.

A continuación, se describe el proceso de desarrollo del modelo, que se caracterizó por la elaboración del algoritmo. El diagrama de flujo de la Figura 1, a continuación, representa este proceso de forma sintética.

Figura 1 — Diagrama de flujo del proceso de modelado de múltiples series temporales



Fuente: elaboración propia.

3.2 Recolección de datos

Los datos se recolectaron en los sitios del TCE-SP y de la Fundación Seade. Del TCE-SP, se extrajeron informaciones sobre los ingresos municipales totales, entre 2008 y

2023 (São Paulo, 2020c). Ya el PIB municipal de 2020 (São Paulo, 2020a) y la población estimada de 2020 (São Paulo, 2020b) se obtuvieron en el Portal de la Seade.

En la hoja de cálculo de ingresos totales, están contenidas las codificaciones, de acuerdo con la Tabla-Resumen del MCASP (Brasil, 2024). Los datos se tabularon en una hoja de cálculo electrónica y, posteriormente, tratados en lenguaje R. Los valores se encuentran en la Tabla 1.

3.3 Análisis de datos

A partir de la obtención de los datos y de su necesario tratamiento, se procedió al desarrollo del algoritmo, también conocido como script, que creó los modelos predictivos. Los algoritmos se desarrollaron en lenguaje R (R Core Team, 2023) y se encuentran disponibles en el repositorio Zenodo (<https://zenodo.org/records/21877515>).

3.3.1 Etapa 1

A partir de estos tres tipos de datos, se realizó el análisis de agrupamiento, denominado Etapa 1 del experimento, de acuerdo con los conceptos y las técnicas de agrupamiento (cluster) impartidas por Fávero y Belfiore (2021), descritos a continuación.

- Estandarización de las variables (PIB, ingresos y población).
- Cálculo de las distancias (o disimilitudes) entre las observaciones por medio de la medida denominada “euclidiana”.
- Adopción del esquema de aglomeración jerárquico y del método de encadenamiento único (single linkage).
- Generación del dendrograma, gráfico que permite la visualización de los agrupamientos.

- Realización del análisis de varianza entre grupos (prueba F). La prueba busca verificar si la variabilidad entre grupos es mayor que la variabilidad intragrupo (Fávero; Belfiore, 2021, p. 338), es decir, averigua si cada variable elegida como parámetro de agrupamiento (PIB, ingreso y población) contribuyó a la formación de los agrupamientos. Se realiza a partir de la elaboración del análisis de la varianza, también conocida como analysis of variance (Anova). Si, en la tabla con los resultados, el valor de F es elevado y el p-valor encontrado es menor que 0,05 (nivel de significancia adoptado en esta investigación), entonces se puede rechazar la hipótesis nula de que “la variable en análisis presenta la misma media en todos los grupos formados” (Fávero; Belfiore, 2021, p. 338). Ya la hipótesis alternativa (H1) ocurre cuando al menos un grupo difiere de los demás y, cuando esto sucede, se concluye que las variables en cuestión contribuyeron a la formación de agrupamientos de la muestra. Según los autores, la prueba es impactada fuertemente por el tamaño de la muestra y, en caso de que existan variables que no posibiliten la formación de los agrupamientos, se deben excluir. Por ello, es necesario retroceder a la etapa de selección de municipios en el caso de no rechazo de la hipótesis nula.

3.3.2 Etapa 2

A partir del agrupamiento de municipios, se pasó a la Etapa 2, iniciada con el examen de los municipios que compusieron las etapas seleccionadas. A continuación, se relacionan las actividades realizadas en esta fase.

- Selección de dos municipios, cuyas series temporales de ingresos del período de 2008 a 2021 estaban normalmente distribuidas en el nivel de significancia del 5,0 % (Morettin; Toloi, 2018). Para este diagnóstico, se adoptó la prueba de Kolmogorov-Sminorv (Método KS), mediante el acceso al sitio Statistic Kingdom (2017). A partir de esta elección, se utilizaron los datos de los ingresos de cada municipio seleccionado, teniendo como base la clasificación de los ingresos presupuestarios por naturaleza (Brasil, 2024), notablemente

por las llamadas “categorías económicas” y “orígenes”. Un aspecto importante es que se retiraron de la columna “categoría” los ingresos corrientes y de capital intrapresupuestarios, dado que en ellos están incluidos los ingresos que pasan de las alcaldías a otros órganos de la misma estructura presupuestaria del ente. De esta forma, se evitó la doble contabilidad de valores. Además, los ingresos intrapresupuestarios provenientes de otros entes también se excluyeron. Se destaca, también, que, en las hojas de cálculo de ingresos obtenidas en el sitio del TCE-SP, la columna que se refiere a la clasificación “origen” se denomina “subcategoría”, tal como se preconiza en las tablas de escrituración contable del TCE-SP (São Paulo, 2007, p. 9).

- Importación de las tablas a la aplicación RStudio y carga de paquetes, notablemente el Forecast (Hyndman *et al.*, 2024) y el Fable (O’Hara-Wild *et al.*, 2023), que posibilita el “apilamiento” de series temporales para un análisis concomitante y múltiple. Además, el paquete Fable (O’Hara-Wild *et al.*, 2023) adopta parametrizaciones del paquete Forecast y, por lo tanto, utiliza las estadísticas AIC y BIC para calcular el modelo más adecuado para la serie temporal en cuestión (O’Hara-Wild *et al.*, 2023).
- Separación de las muestras en entrenamiento (enero de 2008 a diciembre de 2021) y prueba (enero de 2022 a diciembre de 2023).

3.3.2.1 Subetapa 2.1 – Modelado por el paquete Fable

Comprendió las acciones relacionadas a continuación.

- Generación de los modelos ARIMA y de Suavizado Exponencial.
- Verificación de los coeficientes del modelo y selección de los menores valores de AIC, AICc y BIC.
- Realización de la prueba de Ljung-Box para atestiguar si existía o no correlación entre los residuos.

- Aplicación de la prueba de KS para verificar su normalidad (Lopes; Branco; Soares, 2013).
- Realización de la prueba de ARCH (ArchTest) para averiguar si en ellos había o no estacionariedad de la varianza.
- Cálculo del MAPE de las previsiones.

Ferreira y Mattos (2020) muestran que, si los presupuestos de la ausencia de autocorrelación, de la ocurrencia de normalidad y de la homocedasticidad (varianza constante) de los residuos no son cumplidos, el modelo no estará apto para realizar previsiones, pues su parte lineal no estará modelada. Así, en todas las etapas del algoritmo en que los residuos no sean aprobados en estas tres pruebas, será necesaria la selección de nuevos municipios.

3.3.2.2 Subetapa 2.2 – Validación del modelado por el paquete Forecast

Para la comprobación de la exactitud de los modelos obtenidos en la subetapa 2.1, se realizó el modelado por el paquete Forecast, siguiéndose las fases de identificación, estimación, diagnóstico y previsión, tal como fue presentado por Ferreira y Mattos (2020). Se realizaron los análisis relacionados a continuación.

- Identificación y estimación de los modelos ARIMA: se realizó la identificación y estimación por medio de la función `auto.arima`, obteniéndose los componentes autorregresivos “p” (AR), integrados “d” (I) y de medias móviles “q” (MA) y de las análogas componentes estacionales “P”, “D” y “Q”.
- Verificación de la significancia estadística de los coeficientes: se evaluaron la significancia estadística de los coeficientes estimados y la adecuación del modelo.

- Diagnóstico de los residuos: se verificaron la ausencia de autocorrelación de los residuos, la normalidad y la homocedasticidad.
- Previsión y evaluación de la precisión: se generaron las previsiones y se calculó la estadística de error MAPE.
- Modelado por Suavizado Exponencial: para este método, se utilizó la función `forecast.ets`, seleccionándose el modelo con el menor valor de AIC (Hyndman; Khandakar, 2008).
- Comparación de los criterios de información: se verificó si los AIC, AICc y BIC obtenidos eran iguales a los generados en la subetapa 2.1.
- Diagnóstico de los residuos del modelo de Suavizado Exponencial: se verificaron la hipótesis de no existencia de autocorrelación de los residuos del modelo generado, la normalidad y la estacionariedad de la varianza de los residuos.
- Previsión por el método de Suavizado Exponencial: finalmente, se generaron las previsiones y se calculó la estadística de error MAPE.

3.3.2.3 Subetapa 2.3 – Comparación de los resultados con métodos alternativos

Se realizó la comparación de los resultados obtenidos en las subetapas 2.1 y 2.2 con otros métodos para verificar la adecuación del modelo.

- Aplicación de métodos de referencia (benchmarks): primeramente, se realizaron las previsiones de ingresos por los métodos de elaboración menos compleja Seasonal Naive, Naive, Drift y Mean, conforme Dhakal (2017).
- Evaluación de la precisión: se calcularon las medidas de la precisión MAPE de cada uno de ellos.

- Modelado por el X13-ARIMA-SEATS: se generó un modelo del método ARIMA por el programa X13-ARIMA-SEATS.
- División de la muestra: se realizó la división de la serie temporal en muestras de entrenamiento y prueba para la validación de las previsiones.
- Generación y ajuste de los modelos: se generaron múltiples modelos de series temporales y se realizaron los ajustes de cada serie individualmente, por medio de la función `seas`.

Sobre el programa X13-ARIMA-SEATS, Mattos y Ferreira (2020, p. 146) aseveran que una de las funcionalidades que “merece destacarse es el preajuste de la serie temporal, es decir, una corrección antes de hacerse, de hecho, el ajuste estacional”. En ella, se encuentra el preajuste de eventuales valores discrepantes que no son estacionales, pero que afectan la estacionalidad de las series. Se advierte que una de las limitaciones del estudio es que no se ha realizado el preajuste de días hábiles (entre otras fechas específicas) antes del ajuste automático. Por ello, la inversión del orden del uso de los paquetes `Fable` y `Seasonal`, en la metodología propuesta, debe ser objeto de futuros estudios. Importa señalar que, si, por un lado, esta función modela solo una serie a la vez (Mattos; Ferreira, 2020), por otro lado, la ventaja es que todos los análisis de residuos ya se hacen en esta misma función. Finalmente, las previsiones se realizaron por medio de la función `series` (“`basededatos_ajustada`”, `forecast.forecasts`).

Se resalta, en este contexto, que el proceso de modelado debe ser iterativo, es decir, los órganos y las respectivas series temporales a ser seleccionadas deben pasar por pruebas constantes, notablemente en el análisis de los residuos. Además, todo este mecanismo debe estar fundamentado en la “teoría subyacente y en la experiencia del investigador” (Fávero; Belfiore, 2021, p. 511) y ser examinado a la luz de cada caso específico de quien lo está utilizando.

4 RESULTADOS Y DISCUSIÓN

Esta sección está subdividida en dos subsecciones: la primera presenta los resultados del proceso de agrupamiento de municipios, y la segunda demuestra los resultados del modelado de las series temporales de los ingresos de los municipios analizados.

4.1 Resultados obtenidos en la Etapa 1 – Agrupamiento económico de municipios

Los 15 municipios del estado de São Paulo que alcanzaron los mayores valores del PIB están dispuestos en la Tabla 1, en orden descendente.

Tabla 1 — Municipios con los PIB más altos del estado de São Paulo (2020)

Municipio	PIB (BRL)	Ingresos (BRL)	Población
Osasco	63.377.281.663,00	2.855.078.622,94	680.964
Guarulhos	54.753.119.623,00	4.261.460.107,26	1.351.275
Campinas	53.364.538.386,00	5.905.702.404,37	1.175.501
São Bernardo do Campo	39.856.385.608,00	4.363.769.356,61	812.086
Barueri	39.482.997.035,00	3.226.248.352,97	264.390
Jundiaí	39.278.730.712,00	2.284.188.239,98	407.016
São José dos Campos	34.197.969.838,00	3.013.041.783,68	710.654
Paulínia	32.690.552.549,00	1.347.499.218,87	105.037
Ribeirão Preto	31.423.776.118,00	2.949.437.391,30	683.777
Sorocaba	29.554.630.061,00	2.957.377.427,20	658.547
Santo André	25.828.979.709,00	2.638.303.559,53	693.867
Piracicaba	21.130.510.915,00	1.780.043.134,66	389.873
Santos	18.513.709.213,00	2.960.581.983,75	428.703
São José do Rio Preto	16.848.532.921,00	1.996.643.669,15	447.924
Cajamar	16.523.071.678,00	594.102.652,04	77.627

Fuente: Fundação Seade (São Paulo, 2020a, 2020b) y TCE-SP (São Paulo, 2020c).

Posteriormente, de acuerdo con lo propuesto por Fávero y Belfiore (2021), a partir de los datos retrotabulados, se obtuvieron los resultados del proceso de agrupamiento, cuyos cinco primeros niveles están dispuestos ordinalmente en la Tabla 2. Los números negativos representan las posiciones de las observaciones (orden alfabético). Los números sin signo representan los agrupamientos, de acuerdo con el nivel en que ocurrieron. Así, el primer agrupamiento ocurrió con los municipios de Ribeirão Preto (posición 9 en el orden de la hoja de cálculo) y de Sorocaba (posición 15) en el nivel 1. Y, en el nivel 2, las ciudades agrupadas en el nivel 1 se sumaron al municipio de São José dos Campos (observación 12). A continuación, se unió al agrupamiento el municipio de Santo André (observación 13).

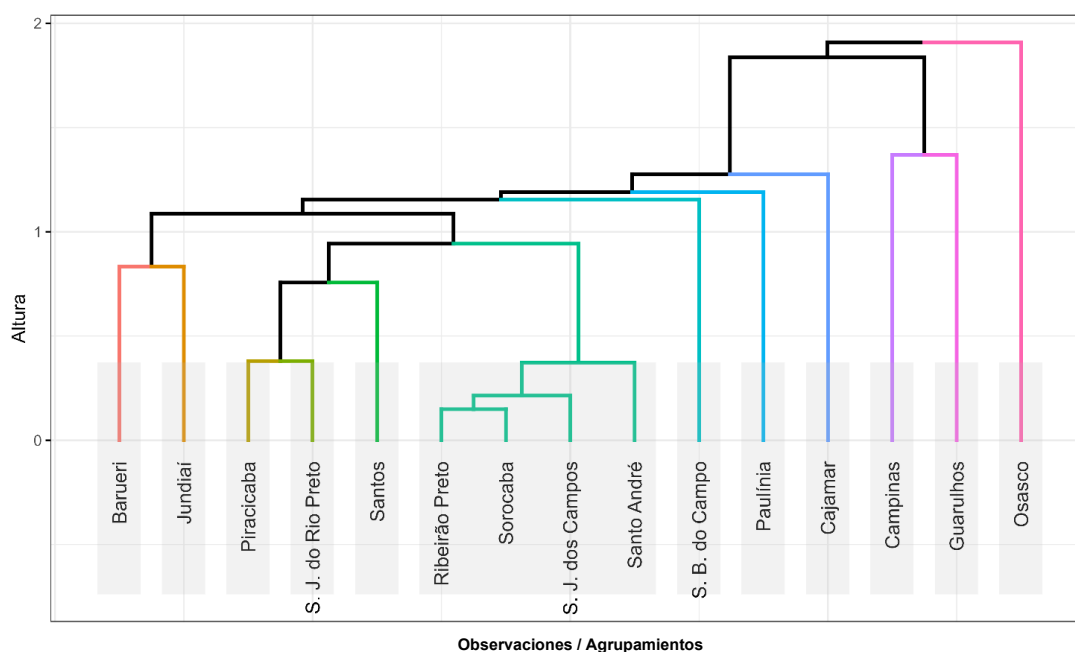
Tabla 2 — Niveles de agrupación de los municipios con los PIB más altos del estado de São Paulo (2020)

Nivel	Observación 1 / Agrupamiento 1	Observación 2 / Agrupamiento 2	Coeficientes
1	-9 (Ribeirão Preto)	-15 (Sorocaba)	0,149
2	1	-12 (São José dos Campos)	0,214
3	2	-13 (Santo André)	0,372
4	-8 (Piracicaba)	-11 (São José do Rio Preto)	0,380
5	4	-14 (Santos)	0,756

Fuente: elaboración propia, con base en cálculos en el lenguaje R (R Core Team, 2023).

El dendrograma contenido en la Figura 2, a continuación, sintetiza gráficamente este resultado.

Figura 2 — Dendrograma del agrupamiento de los 15 municipios con los PIB más altos del estado de São Paulo (2020)



Fuente: elaboración propia, con base en cálculos en el lenguaje R (R Core Team, 2023).

Los resultados de la aplicación de la prueba F se presentan en la Tabla 3.

Tabla 3 — Análisis de la variabilidad intragrupal e intergrupal del agrupamiento de los 15 municipios con los mayores PIB del estado de São Paulo (2020)

Variable de agrupamiento	Valor de la prueba F	P-valor
PIB	20,8	0,01470
Ingreso	73,44	0,00231
Población	332,9	0,00024

Fuente: elaboración propia, con base en cálculos en el lenguaje R (R Core Team, 2023).

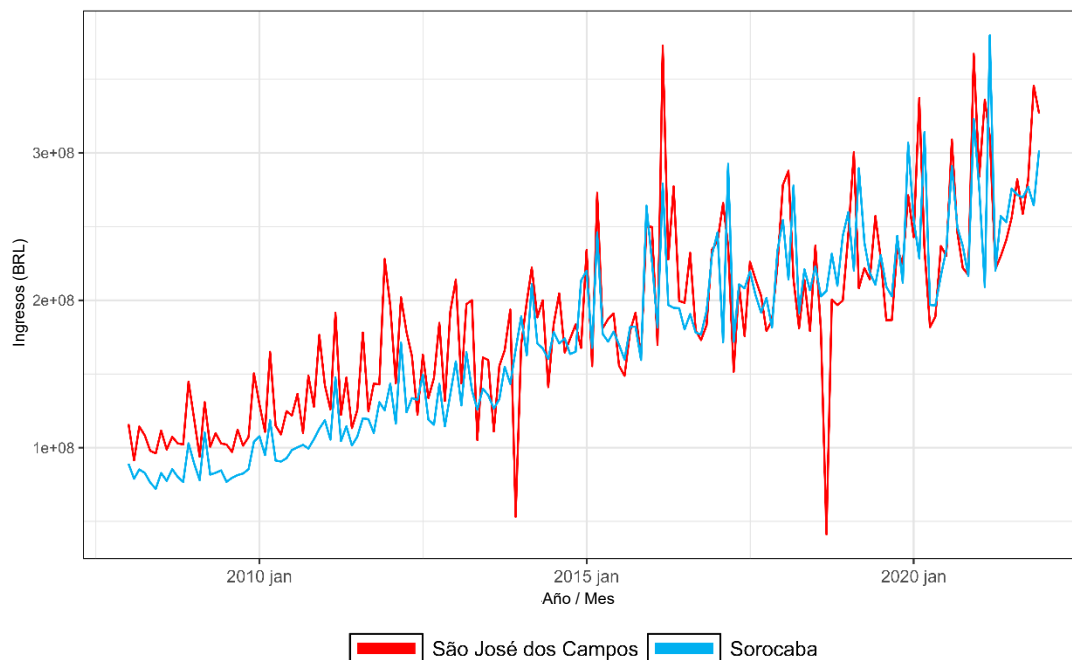
Dado que el p-valor fue inferior a 0,05, se demostró, conforme a lo previsto en Fávero y Belfiore (2021), que las variables seleccionadas justifican la idoneidad de los agrupamientos obtenidos.

4.2 Resultados obtenidos en la Etapa 2 – Elaboración de los modelos predictivos

Del agrupamiento realizado en la Etapa 1, se seleccionaron los municipios de São José dos Campos y Sorocaba para un análisis múltiple de series temporales. La omisión del municipio de Ribeirão Preto se debió a que esta ciudad, aunque poseía una distribución normal de los ingresos, no presentó normalidad en la distribución de los residuos (Morettin; Toloí, 2018). Este requisito es esencial para que los modelos generados capten toda la información de la base de datos, minimizando la pérdida en los residuos (Ferreira; Mattos, 2020).

En la continuidad del análisis, se verificó que los ingresos de los dos municipios poseían valores relativamente cercanos, cuando se comparaban con los de otros municipios. En el mes de diciembre de 2020, São José dos Campos tuvo un ingreso mensual de BRL 367 millones y Sorocaba, BRL 322 millones (en valores redondeados). Con base en los panoramas económicos de los dos municipios, se observó la evolución de los ingresos de enero de 2008 a diciembre de 2021. El análisis gráfico de las series temporales (Figura 3) indica una tendencia de crecimiento en ambos municipios. En 2020, se observa una caída en los ingresos, posiblemente asociada a la crisis económica causada por la pandemia del covid-19, seguida de una recuperación en los años siguientes.

Figura 3 — Series de tiempo de los ingresos de los municipios de São José dos Campos y de Sorocaba (2008-2021) (en BRL)



Fuente: elaboración propia, con base en cálculos en el lenguaje R (R Core Team, 2023).

A partir de la presentación gráfica de la evolución del ingreso de 2008 a 2021, se generaron los modelos, presentados en las secciones a continuación.

4.2.1 Subetapa 2.1 – Elaboración de los modelos ARIMA y de Suavizado Exponencial

Como resultado de la ejecución del paquete Fable, se obtuvieron los modelos listados en la Tabla 4.

Tabla 4 — Modelos ARIMA y de Suavizado Exponencial de las series temporales de los ingresos de São José dos Campos y de Sorocaba (2008-2021)

Continúa

Municipio	Tipo de modelado	Identificación del modelo
São José dos Campos	ARIMA	(0,1,1) (0,0,1) [12]

Conclusión

Municipio	Tipo de modelado	Identificación del modelo
São José dos Campos	Suavizado Exponencial	Multiplicativo, aditivo, multiplicativo
Sorocaba	ARIMA	(1,0,1) (0,1,1) [12] con constante
Sorocaba	Suavizado Exponencial	Multiplicativo, aditivo, multiplicativo

Fuente: elaboración propia, con base en cálculos en el lenguaje R (R Core Team, 2023).

En los modelos ARIMA, se observa que, en São José dos Campos, la serie es integrada de orden 1 en la parte no estacional. En Sorocaba, la serie es integrada de orden 1 en la parte estacional. Esto significa que es necesario realizar una diferenciación en la serie original para eliminar raíces unitarias, tanto en la parte no estacional como en la estacional, a fin de que se pueda estabilizar la media y la varianza de los valores de las series y, con ello, obtener previsiones más precisas. A continuación, se obtuvieron, por medio del algoritmo, los menores valores de AIC, AICc y BIC, necesarios para la identificación y especificación del modelo (O'Hara-Wild *et al.*, 2023). Los resultados se presentan en la Tabla 5.

Tabla 5 — Valores de AIC, AICc y BIC para los modelos ARIMA y de Suavizado Exponencial de São José dos Campos y de Sorocaba (2008-2021)

Continúa

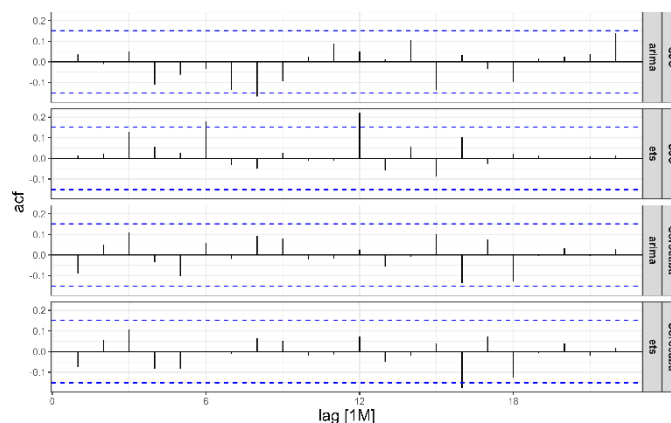
Municipio	Tipo de modelado	AIC	AICc	BIC	Parámetros
São José dos Campos	ARIMA	6313	6313	6322	Ma1=-8,58e-1; Sma1=3,83e-1
São José dos Campos	Suavizado Exponencial	6678	6682	6732	Alpha=1,00e-2; Beta=1,04e-4; Gama=1,00e-4.
Sorocaba	ARIMA	5650	5651	5666	Ar1=9,17e-1; Ma1=-8,32e-1; Sma1=-5,78e-1; Constante = 1,22e+6

Sorocaba	Suavizado Exponencial	6379	6383	6432	Alpha=1,28e-1; Beta=7,35e-3; Gama=1,77e-4.
----------	--------------------------	------	------	------	--

Fuente: elaboración propia, con base en cálculos en el lenguaje R (R Core Team, 2023).

En la secuencia de la verificación, el correlograma (Figura 4) mostró que hay ruptura de las líneas de significancia estadística de los residuos de los modelos ARIMA para los dos municipios. En otras palabras, hubo indicativo de correlación entre los residuos (Ferreira, 2020).

Figura 4 — Correlograma de los residuos de las series temporales de los modelos ARIMA y de Suavizado Exponencial (2008-2021)



Fuente: elaboración propia, con base en cálculos en el lenguaje R (R Core Team, 2023).

Ante esta situación, para averiguar si esta hipótesis se confirmaría o no, se procedió a las pruebas sobre los residuos. En la Tabla 6, todos los valores en la columna “Prueba Ljung-Box” son superiores a 0,05. Por lo tanto, los residuos no están correlacionados. A continuación, la prueba Kolmogorov-Smirnov mostró que hay normalidad en la distribución de los residuos. Y, por último, la prueba de ARCH, con un nivel de significación del 5,0 %, mostró que no se puede rechazar la hipótesis de homocedasticidad de la varianza de los residuos. Así, no fue posible, para São José dos Campos, rechazar ni la hipótesis nula de ausencia de autocorrelación lineal, ni la hipótesis de normalidad ni la hipótesis de homocedasticidad de la varianza de los residuos. Para Sorocaba, la única diferencia fue el nivel de significancia estadística

para la hipótesis de no rechazo de la hipótesis de homoscedasticidad de la varianza, del 1,0 %. Los valores constan en la Tabla 6.

Tabla 6 — Resultados del diagnóstico de los residuos de las series temporales para los modelos ARIMA y de Suavizado Exponencial (2008-2021)

Municipio	Tipo de modelado	Prueba Ljung-Box	Prueba de Kolmogorov - Smirnov	ArchTest
São José dos Campos	ARIMA	0,11	0,08	0,99
São José dos Campos	Suavizado Exponencial	0,57	0,48	0,99
Sorocaba	ARIMA	0,31	0,06	0,02
Sorocaba	Suavizado Exponencial	0,38	0,88	0,25

Fuente: elaboración propia, con base en cálculos en el lenguaje R (R Core Team, 2023).

Ante estos valores, todos por encima del 1,0%, tal como instruyen Mattos y Ferreira (2020), se concluyó que el componente lineal de la serie temporal había sido modelado, lo que permitió la realización de las previsiones. Así, el cuarto y último procedimiento del modelado del paquete Fable fue el de las previsiones para 2022 y 2023. En la Tabla 7, están dispuestos los valores de las estadísticas de error MAPE.

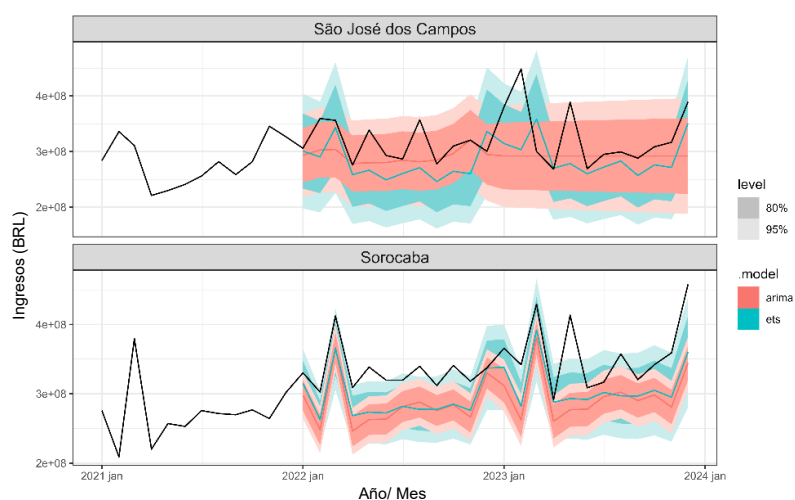
Tabla 7 — Comparación de los errores de previsión (MAPE) entre los modelos ARIMA y de Suavizado Exponencial para São José dos Campos y para Sorocaba (2022-2023)

Municipio	Tipo de modelado	MAPE
São José dos Campos	ARIMA	9,75%
São José dos Campos	Suavizado Exponencial	13,2%
Sorocaba	ARIMA	15,3%
Sorocaba	Suavizado Exponencial	12,3%

Fuente: elaboración propia, con base en cálculos en el lenguaje R (R Core Team, 2023).

A partir de la Tabla 7, es posible concluir que São José dos Campos posee un modelo de previsión de ingresos más preciso con el método ARIMA, ya que la estadística de error MAPE fue de 9,75 %. Ya con relación a Sorocaba, el 12,3 % de MAPE indica que el método de Suavizado Exponencial es el más indicado para realizar previsiones. Gráficamente, la evolución de las previsiones está expuesta en la Figura 5. En la línea negra, constan los valores reales de las series (los datos anteriores a 2021 se removieron de la figura para proporcionar mayor nitidez).

Figura 5 — Previsiones de las series temporales de São José dos Campos y de Sorocaba para los modelos ARIMA y de Suavizado Exponencial (2022-2023)



Fuente: elaboración propia, con base en cálculos en el lenguaje R (R Core Team, 2023).

Después de todos estos procedimientos, se verificó que el paquete Fable dispone de todas las funcionalidades para la realización del proceso de modelado de múltiples series temporales. De este modo, se concluyó la subetapa 2.1.

4.2.2 Subetapa 2.2 – Comprobación y validación de resultados

En la subetapa 2.2, ya en la primera verificación – tras la ejecución de la función `auto.arima`, se constató que los valores de AIC, AICc y BIC obtenidos en el paquete Forecast fueron los mismos del paquete Fable. A continuación, la comprobación de la significancia estadística de los coeficientes (Ferreira; Mattos, 2020), por medio de la

función `BETS:t_test`, del paquete `BETS` (Ferreira *et al.*, 2022), mostró que todos los coeficientes del modelo ARIMA cumplían con esta exigencia, conforme se observa en la Figura 6. Se nota que el valor crítico “t Crit. Value”, de 1,97, está comprendido en el intervalo de las estadísticas obtenidas para los errores estándar “Std. Errors”, lo que impide el rechazo de la hipótesis nula de que los coeficientes de los modelos ARIMA sean estadísticamente iguales a cero. Por ello, el resultado presentado es “True”.

Figura 6 — Análisis de la significancia estadística de los coeficientes de los modelos ARIMA

```
> BETS::t_test(city1.treino.autoarima) # # Análise dos Coeficientes
      Coeffs Std.Errors      t Crit.Values Rej.H0
ma1  -0.8579742 0.04284415 20.025470      1.974446  TRUE
sma1  0.3828907 0.06756902  5.666661      1.974446  TRUE
> BETS::t_test(city2.treino.autoarima) # # Análise dos Coeficientes
      Coeffs Std.Errors      t Crit.Values Rej.H0
ar1    9.168114e-01 7.113143e-02 12.888977      1.974535  TRUE
ma1   -8.324925e-01 9.058996e-02  9.189677      1.974535  TRUE
sma1  -5.783509e-01 7.489816e-02  7.721830      1.974535  TRUE
drift  1.220164e+06 1.031064e+05 11.834029      1.974535  TRUE
```

Fuente: elaboración propia, con base en cálculos en el lenguaje R (R Core Team, 2023).

Se debe hacer una salvedad importante sobre la constante de la función lineal obtenida (término de drift). Es posible que el modelo generado en el paquete `Fable` presente un valor diferente al generado por el paquete `Forecast`. Esta inconsistencia se explica en el propio Manual del paquete `Fable` (2023, p. 59): “Forecast standard errors allow for uncertainty in estimating the drift parameter (unlike the corresponding Forecasts obtained by fitting an ARIMA model directly)”. Además, cuando alguno de los coeficientes no es estadísticamente significativo, es posible la reestimación del modelo por medio de la función `ARIMA` (Mattos; Ferreira, 2020), bastando suprimir tantas unidades de coeficientes como sean no significativas.

A continuación, en cuanto a la normalidad de los residuos, se realizaron las pruebas de ausencia de autocorrelación lineal, normalidad y homocedasticidad de la varianza. Con la utilización de la función `accuracy` (paquete `Forecast`), se verificó que São José dos Campos presentó un MAPE de 9,75 % y Sorocaba, 12,3 % para los modelos ARIMA. Estos resultados convergieron con los valores encontrados por el modelado del paquete `Fable`. Para la verificación de la idoneidad de los modelos de Suavizado

Exponencial, el proceso fue más simple. Se generó el modelo a partir de la función *ets*. A continuación, se procedió a las pruebas y diagnósticos sobre los residuos.

Como fue posible observar, tanto el modelo generado como las previsiones obtenidas por medio del paquete *Fable* resultaron en los mismos valores y en las conclusiones obtenidas por medio del paquete *Forecast*, lo que corroboró el hecho de que los dos paquetes utilizan la misma parametrización (O'Hara-Wild *et al.*, 2023).

4.2.3 Subetapa 2.3 – Benchmark y otros métodos

En la subetapa 2.3 del algoritmo generado en el RStudio, los modelos “ingenuos” de previsiones *Drift*, *Mean*, *Naive* y *Seasonal Naive* presentaron los valores de error (MAPE), en el bienio 2022 y 2023, conforme se muestra en la Tabla 8.

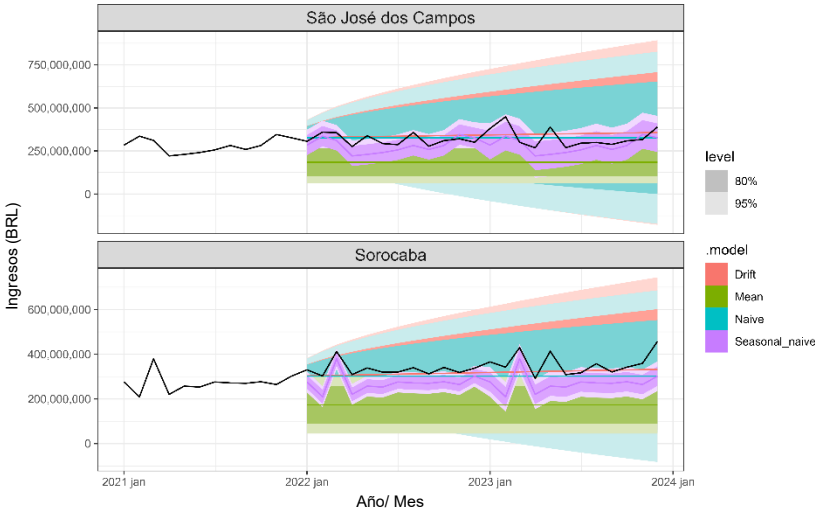
Tabla 8 — Estadística de error MAPE de las series de tiempo de los ingresos de São José dos Campos y de Sorocaba por los métodos Naive (2022-2023)

Municipio	Drift	Mean	Naive	Seasonal Naive
São José dos Campos	14,4	41,6	11,7	14,4
Sorocaba	8,69	49,2	11,7	21,3

Fuente: elaboración propia, con base en cálculos en el lenguaje R (R Core Team, 2023).

Considerando los resultados de esta tabla, para São José dos Campos, el método *Naive*, con MAPE de 11,7 %, fue el más cercano al resultado presentado por el método *ARIMA*, con 9,75 %. El método *Naive* es aquel que considera el valor real más reciente de la serie como la previsión de los períodos siguientes. Para Sorocaba, el método *Drift*, con MAPE de 8,69 %, fue el más cercano al resultado presentado por el método de Suavizado Exponencial, con 12,3 %. Este es el método que toma en consideración la tendencia de la serie para realizar las previsiones, de acuerdo con lo que se observa en la Figura 7.

Figura 7 — Previsiones de las series de tiempo de los ingresos de los municipios de São José dos Campos y de Sorocaba por los métodos Naive (2022-2023)



Fuente: elaboración propia, con base en cálculos en el lenguaje R (R Core Team, 2023).

Como se puede constatar, los modelos ARIMA y de Suavizado Exponencial son de elaboración más compleja que los modelos Naive, lo que corrobora los estudios de Dhakal (2017). En cuanto a la utilización del programa X13-ARIMA-SEATS, por medio del paquete Seasonal, los resultados están dispuestos en la Tabla 9.

Tabla 9 — Modelado de las series de tiempo de los ingresos de São José dos Campos y de Sorocaba por el programa X13-ARIMA-SEATS (2008-2021)

Continúa

Elemento del modelado	São José dos Campos	Sorocaba
Identificación del modelo	(0,1,1) (0,1,1)	(0,1,1) (0,1,1)
Estimación – Coeficientes	No estacional: MA=0,9 Estacional: MA=0,61	No estacional: MA=0,71 Estacional: MA=0,73
Criterios de información AIC; AICc; BIC	-; 5793; 5830	-; 5498; 5527
QS Prueba – ¿La serie ajustada presentó estacionalidad?	No	No

Elemento del modelado	São José dos Campos	Sorocaba
Ljung-Box – ¿La serie posee autocorrelación residual?	No	No
Shapiro Wilk – ¿Los residuos poseen distribución normal?	No	Sí
Previsiones MAPE (2022 y 2023)	7,54%	11,78%

Fuente: elaboración propia, con base en cálculos en el lenguaje R (R Core Team, 2023).

Como se nota, hay una diferencia entre los modelos identificados en los modelos generados por el paquete Fable (o Forecast) y por el Seasonal (X13-ARIMA-SEATS). Primero, porque, en este, se constató que se realizaron dos diferenciaciones, estacional y no estacional para cada una de las dos ciudades, mientras que, en el paquete Fable, hubo apenas una diferenciación para la serie temporal de cada municipio. Segundo, porque el componente autorregresivo presentó valor cero en el modelo referente a Sorocaba. En cuanto a las estadísticas de error MAPE, para el paquete Fable, São José dos Campos tuvo MAPE de 9,75 % y Sorocaba 15,3 %. Bajo la prueba de Shapiro Wilk, los residuos de São José dos Campos no son normales al 5 % de significancia, lo que demandaría ajustes adicionales bajo el método X13-ARIMA-SEATS. En Sorocaba, el paquete Seasonal realizó una transformación logarítmica. De esta forma, se verificó que el paquete Fable, en el método ARIMA, presentó MAPEs y criterios de información superiores a los del paquete Seasonal (programa X13-ARIMA-SEATS) para ambas ciudades. A pesar de esto, el modelo obtenido para São José dos Campos en el programa X13-ARIMA-SEATS no pudo ser utilizado para la comparación con el generado en el paquete Fable, por no cumplir con el presupuesto de normalidad de los residuos. Tampoco pudo ser utilizado para realizar previsiones o calcular el MAPE. Para que esto se pudiera hacer, serían necesarios ajustes manuales (Mattos; Ferreira, 2020), no previstos en este estudio.

Ante tales resultados, se resalta que cada una de estas técnicas posee presupuestos; por ejemplo, la calidad y la cantidad de los datos; la necesidad de ajustes de los datos en relación con posibles impactos de datos discordantes (outliers); la complejidad de

los algoritmos a ser desarrollados. Se verificó, además, en la presente investigación, que los modelos desarrollados en el X13-ARIMA-SEATS utilizaron menos líneas de comando que aquellos implementados por el paquete Fable. Sin embargo, dependiendo de la serie analizada, la complejidad del modelado puede aumentar. Además, para São José dos Campos, aunque el método ARIMA implementado en el paquete Fable haya presentado menor precisión que aquel implementado en el paquete Seasonal, fue el único que cumplió con todos los presupuestos del modelado. En Sorocaba, a su vez, se observó mayor precisión en las previsiones obtenidas con el paquete Seasonal. Como se verificó, el algoritmo generó las series de dos municipios al mismo tiempo y en dos métodos distintos – ARIMA y Suavizado Exponencial, lo que posibilitó mayor celeridad en el modelado y en la comparación de los órganos en cuestión. En cuanto a la precisión de los resultados obtenidos, considerando la escala de precisión presentada por Klimberg, Sillup, Boyle y Tavva (2010), se verificó que São José dos Campos, con MAPE de 9,75 %, presentó una previsión de alta precisión, mientras que Sorocaba, con MAPE de 12,31 %, presentó precisión buena. Además, los valores de MAPE de los modelos ARIMA y de Suavizado Exponencial encontrados en el modelo generado en esta investigación fueron inferiores al promedio del 17,5 % de error por ejercicio, encontrado en los estudios de Nascimento y Boente (2022).

5 CONSIDERACIONES FINALES

Este estudio demostró la viabilidad de un modelo predictivo para el análisis simultáneo de múltiples series temporales de ingresos públicos, facilitando la planificación presupuestaria y la toma de decisiones en órganos recaudadores. Los resultados mostraron que hay plena compatibilidad entre los paquetes Forecast y Fable, con la ventaja de que, en este, existe la posibilidad de generación de múltiples modelos (Suavizado Exponencial, ARIMA), simultáneamente. Así, los interesados en la información contable, en vez de efectuar el modelado por el método de Suavizado Exponencial y por el método ARIMA, separadamente, para verificar cuál presenta los menores valores de AIC, AICc, BIC o el menor MAPE para los datos en análisis, pueden hacerlo con un único algoritmo en el paquete Fable. Y esto puede realizarse

con más de un municipio en el mismo código. Otra ventaja es la transparencia en el algoritmo, que posibilita a los analistas la comprensión de cada etapa y de cada resultado parcial obtenido.

En el aspecto estadístico, los resultados mostraron que incluso los municipios con datos económicos similares (agrupamientos) pueden resultar en valores de AIC, AICc, BIC y MAPEs distintos y, por lo tanto, demandar métodos de modelado diferentes. Esta situación puede ser el punto de partida para investigaciones sobre el motivo por el cual municipios con trayectorias de ingresos semejantes pueden presentar resultados futuros diferentes, como en el caso de una auditoría financiera.

Se proponen nuevos estudios conducidos por otras hipótesis, como la utilización de series temporales que no posean distribución normal. Del mismo modo, desde el punto de vista de la formación de agrupamientos, otras variables pueden ser sumadas a los análisis, además de diversificar el tamaño de las muestras.

Se recuerda que el proceso de modelado es útil no solo para realizar previsiones, sino también para entender el comportamiento de las series temporales. Además, es un proceso de apoyo a la toma de decisiones. Por ello, la utilización de estas herramientas tecnológicas de modelado debe estar asociada a la experiencia profesional, al marco conceptual de las ciencias económicas, a la ciencia de datos, así como al cumplimiento de la legislación, en el caso de uso por órganos gubernamentales.

REFERENCIAS

ABELN, Barend; JACOBS, Jan P. A. M. Covid-19 and seasonal adjustment. **Journal of Business Cycle Research**, v. 18. p. 159-169, jul. 2022. DOI: <https://doi.org/10.1007/s41549-022-00071-z>. Disponible en: <https://link.springer.com/article/10.1007/s41549-022-00071-z>. Consultado el: 23 sept. 2024.

ARAÚJO, Letícia; BEHR, Ariel; SCHIAVI, Giovana Sordi. Adoption of business analytics in accounting. **Revista Contabilidade & Finanças**, São Paulo, v. 34, n. 93, e1771, dez. 2023. DOI: <https://doi.org/10.1590/1808-057x20231771.en>. Disponible en: <https://www.scielo.br/j/rcf/a/h8BSKTCgthvCfwCf9dNSCkH/?lang=en>. Consultado el: 23 sept. 2024.

BATÓG, Barbara; BATÓG, Jacek. Regional government revenue forecasting: risk factors of investment financing. **Risks**, v. 9, n. 12, p. 1-15, nov. 2021. DOI: <https://doi.org/10.3390/risks9120210>. Disponible en: <https://www.mdpi.com/2227-9091/9/12/210>. Consultado el: 23 sept. 2024.

BRASIL. Banco Central do Brasil. **Efeitos da pandemia sobre o ajuste sazonal dos indicadores econômicos**. Estudo Especial 108/2021. Brasília, DF: BCB, 2021. Disponible en: https://www.bcb.gov.br/conteudo/relatorioinflacao/EstudosEspeciais/EE108_efeitos_pandemia_sobre_ajuste_sazonal.pdf. Consultado el: 23 sept. 2024.

BRASIL. [Constituição (1988)]. **Constituição da República Federativa do Brasil de 1988**. Brasília, DF: Presidência da República, 1988. Disponible en: http://www.planalto.gov.br/ccivil_03/Constituicao/Constituicao.htm. Consultado el: 27 jul. 2026.

BRASIL. **Lei n. 4.320, de 17 de março de 1964**. Estatui normas gerais de Direito Financeiro para elaboração e controle dos orçamentos e balanços da União, dos estados, dos municípios e do Distrito Federal. Brasília, DF: Presidência da República, 1964. Disponible en: https://www.planalto.gov.br/ccivil_03/leis/l4320.htm. Consultado el: 28 jul. 2026.

BRASIL. **Lei Complementar n. 101, de 4 de maio de 2000**. Estabelece normas de finanças públicas voltadas para a responsabilidade na gestão fiscal e dá outras providências. Brasília, DF: Presidência da República, 2000. Disponible en: https://www.planalto.gov.br/ccivil_03/leis/lcp/lcp101.htm. Consultado el: 28 jul. 2026.

BRASIL. Ministério da Economia. **Metodologia**: ajuste sazonal. Brasília, DF: ME, 2020. Disponible en: https://balanca.economia.gov.br/balanca/IPQ/arquivos/metodologia_ajuste_sazonal.pdf. Consultado el: 23 sept. 2024.

BRASIL. Secretaria do Tesouro Nacional. **Manual de contabilidade aplicado ao setor público**. 11. ed. Brasília, DF: STN, 2024. Disponible en: <https://www.gov.br/tesouronacional/pt-br/contabilidade-e-custos/manuais/manual-de-contabilidade-aplicada-ao-setor-publico-mcasp-1>. Consultado el: 6 feb. 2026.

DHAKAL, Chuda Prasad. A naïve approach for comparing a forecast model. **International Journal of Thesis Projects and Dissertations (IJTPD)**, v. 5, n. 1. p. 1-3, jan./mar. 2017. Disponible en: <https://www.researchpublish.com/papers/a-naive-approach-for-comparing-a-forecast-model>. Consultado el: 23 sept. 2024.

FAVERET, Guilherme; AGUILAR, Soraida; BAIÃO, Fernanda; SIMÕES, Paulo Henrique; MAÇAIRA, Paula. Estudo de similaridade da evolução da covid-19 nos estados brasileiros. In: Simpósio Brasileiro de Pesquisa Operacional, 53, João Pessoa, 2021. **Anais [...]**, v. 53. João Pessoa: Sociedade Brasileira de Pesquisa Operacional (SOBRAPO), 2021. Disponible en: <https://proceedings.science/sbpo/sbpo-2021/trabalhos/estudo-de-similaridade-da-evolucao-da-covid-19-nos-estados-brasileiros?lang=pt-br>. Consultado el: 23 sept. 2024.

FÁVERO, Luiz Paulo; BELFIORE, Patrícia. **Manual de análise de dados**: estatística e modelagem multivariada com Excel®, SPSS® e Stata®. 1. ed. Rio de Janeiro: LTC, 2021.

FERREIRA, Pedro Costa; SPERANZA, Talitha; COSTA, Jonatha; TEIXEIRA, Fernando; MARCOLINO, Daiane. Pacote 'BETS: Brazilian Economic Time Series'. **The comprehensive R archive network**. Documentation. Reference manual. Versão 0.4.9. sept. 2018. Disponible en: <https://cran.r-project.org/src/contrib/Archive/BETS/>. Consultado el: 23 sept. 2024.

FERREIRA, Pedro Guilherme Costa; DUCA, Victor Eduardo Leite de Almeida. Modelos de suavização exponencial. In: FERREIRA, Pedro Guilherme Costa (org.). **Análise de séries temporais em R**: curso introdutório. 1. ed. São Paulo: GEN Atlas; FGV IBRE, 2020. p. 75-93.

FERREIRA, Pedro Guilherme Costa. **Processos não estacionários**. In: FERREIRA, Pedro Guilherme Costa (org.). In: FERREIRA, Pedro Guilherme Costa (org.). **Análise de séries temporais em R**: curso introdutório. 1. ed. São Paulo: GEN Atlas; FGV IBRE, 2020. p. 95-120.

FERREIRA, Pedro Guilherme Costa; MATTOS, Daiane Marcolino de. **Modelo SARIMA**. In: FERREIRA, Pedro Guilherme Costa (org.). **Análise de séries temporais em R**: curso introdutório. 1. ed. São Paulo: GEN Atlas; FGV IBRE, 2020. p. 121-143.

GERHARDT, Tatiana Engel; SILVEIRA, Denise Tolfo (org.). **Métodos de Pesquisa**. Série Educação a distância. Universidade Aberta do Brasil (UAB); Universidade Federal do Rio Grande do Sul (UFRGS) (coord.). Porto Alegre: Editora da UFRGS, 2009. Disponible en: <https://www.ufrgs.br/cursopgdr/downloadsSerie/derad005.pdf>. Consultado el: 27 jul. 2026.

HYNDMAN, Rob; ATHANASOPOULOS, George; BERGMEIR, Christoph; CACERES, Gabriel; CHHAY, Leanne; KUROPTEV, Kirill; MÜCKE, Maximilian; O'HARA-WILD, Mitchell; PETROPOULOS, Fotios; RAZBASH, Slava; WANG, Earo; YASMEEN, Farah; GARZA. Forecast: forecasting functions for time series and linear models. **The comprehensive R archive network**. Documentation. Reference manual. Versão 8.21.1. 2024. DOI: <https://doi.org/10.32614/CRAN.package.forecast>. Disponible en: <https://cran.r-project.org/web/packages/forecast/index.html>. Consultado el: 23 sept. 2024.

HYNDMAN, Rob J.; KHANDAKAR, Yeasmin. Automatic time series forecasting: the forecast package for R. **Journal of Statistical Software**, v. 27, n. 3, p. 1-22, jul. 2008. DOI: <https://doi.org/10.18637/jss.v027.i03>. Disponible en: <https://www.jstatsoft.org/article/view/v027i03>. Consultado el: 23 sept. 2024.

HYNDMAN, Rob. J.; ROSTAMI-TABAR, Bahman. Forecasting interrupted time series. **Journal of the Operational Research Society**, v. 76, n. 4, p. 790-803, sept. 2024. DOI: <https://doi.org/10.1080/01605682.2024.2395315>. Disponible en: <https://www.tandfonline.com/doi/full/10.1080/01605682.2024.2395315>. Consultado el: 23 sept. 2024.

KIM, Sungil; KIM, Heeyoung. A new metric of absolute percentage error for intermittent demand forecasts. **International Journal of Forecasting**, v. 32, n. 3, p. 669-679, jul./sept. 2016. DOI: <https://doi.org/10.1016/j.ijforecast.2015.12.003>. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0169207016000121?via%3Dihub>. Consultado el: 14 mar. 2025.

KLIMBERG, Ronald K.; SILLUP, George P.; BOYLE, Kevin J.; TAVVA, Vinay. Forecasting performance measures: what are their practical meaning? In: LAWRENCE, Kenneth D.; KILMBERG, Ronald K (eds.). **Advances in business and management forecasting**. 7. ed. Leeds: Emerald Books, 2010. p. 137-147. DOI: [https://doi.org/10.1108/S1477-4070\(2010\)0000007012](https://doi.org/10.1108/S1477-4070(2010)0000007012). Disponible en: <https://www.emerald.com/books/edited-volume/11151/chapter-abstract/80949903/Forecasting-performance-measures-what-are-their?redirectedFrom=fulltext>. Consultado el: 14 mar. 2025.

KONIAGINA, Maria N. Forecast of budget revenues from taxes in the context of economy digitalization. **IOP Publishing**, IOP conference series: materials science and engineering, v. 940, 012040. 2020. DOI: <https://doi.org/10.1088/1757-899X/940/1/012040>. Disponible en: <https://iopscience.iop.org/article/10.1088/1757-899X/940/1/012040>. Consultado el: 23 sept. 2024.

LEVY, Karen; CHASALOW, Kyla E.; RILEY, Sarah. Algorithms and decision making in the public sector. **Annual review of law and social science**, v. 17, p. 309-334, out. 2021. DOI: <https://doi.org/10.1146/annurev-lawsocsci-041221-023808>. Disponible en: <https://www.annualreviews.org/content/journals/10.1146/annurev-lawsocsci-041221-023808>. Consultado el: 14 mar. 2025.

LIU, Shiyu; LIU, Ou; CHEN, Junyang. Review on business analytics: definitions, techniques, applications and challenges. **Mathematics**, v. 11, n. 4, p. 1-20, fev. 2023. DOI: <https://doi.org/10.3390/math11040899>. Disponible en: <https://www.mdpi.com/2227-7390/11/4/899>. Consultado el: 23 sept. 2024.

LOPES, Manuela de Mesquita; BRANCO, Verônica Teixeira Franco Castelo; SOARES, Jorge Barbosa. Utilização dos testes estatísticos de Kolmogorov-Smirnov e Shapiro-Wilk para verificação da normalidade para materiais de pavimentação. **Revista Transportes**, São Paulo, v. 21, n. 1, p. 59-66, abr. 2013. DOI: <https://doi.org/10.4237/transportes.v21i1.566>. Disponible en: <https://transportes.anpet.org.br/anpet/article/view/566>. Consultado el: 23 sept. 2024.

MATTOS, Daiane Marcolino de; FERREIRA, Pedro Guilherme Costa. Ajuste sazonal. In: FERREIRA, Pedro Guilherme Costa (org.). **Análise de séries temporais em R: curso introdutório**. 1. ed. São Paulo: GEN Atlas; FGV IBRE, 2020. p. 145-163.

MORETTIN, Pedro A.; TOLOI, Clélia M. C. **Análise de séries temporais: modelos lineares univariados**. Volume 1. 3. ed. São Paulo: Bluncher, 2018.

MORETTIN, Pedro A.; TOLOI, Clélia M. C. **Análise de séries temporais: modelos multivariados e não lineares**. Volume 2. 1. ed. São Paulo: Bluncher, 2020.

MUÑOZ, Laura Alcaide; BOLÍVAR, Manuel Pedro Rodríguez; ARELLANO; Cinthia L. Villamayor. Factors in the adoption of open government initiatives in Spanish local governments. **Government information quarterly**, v. 39, n. 4, p. 1-13, out. 2022. DOI: <https://doi.org/10.1016/j.giq.2022.101743>. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0740624X2200079X?via%3Dihub>. Consultado el: 5 mar. 2025.

NASCIMENTO, Maria de Nazaré Souza; BOENTE, Diego Rodrigues. Fatores associados aos erros de previsão orçamentária da receita do setor público. **Administração pública e gestão social**, Universidade Federal de Viçosa, v. 14, n. 2, p. 1-15, abr./jun. 2022. Disponible en: <https://doi.org/10.21118/apgs.v14i2.12945>. Consultado el: 23 sept. 2024.

O'HARA-WILD, Mitchell; HYNDMAN, Rob; WANG, Earo; CACERES, Gabriel; BERGMEIR, Christoph; HENSEL, Tim-Gunnar; HYNDMAN, Timothy. Fable: forecasting models for tidy time series. **The comprehensive R archive network**. Documentation. Reference manual. Versão 0.3.3. 2023. DOI: <https://doi.org/10.32614/CRAN.package.fable>. Disponible en: <https://cran.r-project.org/web/packages/fable/index.html>. Consultado el: 23 sept. 2024.

OLIVEIRA JUNIOR, João Paulo de; LEVINO, Natallya de Almeida. Impacto da pandemia covid-19 nas finanças dos municípios do Nordeste: uma análise da efetividade dos critérios de distribuição do auxílio financeiro. **Revista Brasileira de Planejamento e Desenvolvimento**, Curitiba, v. 13, n. 2, p. 369-400, maio/ago., 2024. Disponible en: <https://periodicos.utfpr.edu.br/rbpd/article/view/16231/10175>. Consultado el: 23 sept. 2024.

SÁ, Rogério de Oliveira; AMORIM, Junior José; SANTOS, Cleidiany Dias dos; ZAMBRIM, Dejan Arruda de Carli; GOMES, Vallência Maira. **Metodologia utilizada nas estimativas de receitas e despesas (2021-2025)**. Cuiabá: Secretaria de Estado de Fazenda, 2020. Disponible en: <https://www5.sefaz.mt.gov.br/documents/6071037/15333979/Metodologias+utilizada+s+nas+proje%C3%A7%C3%B5es+04.09.2020.pdf/38604051-d9da-f912-e9e0-e0150e1c1148>. Consultado el: 23 sept. 2024.

SÃO PAULO. Fundação Sistema Estadual de Análise de Dados. **Tabela – PIB 2020**. São Paulo, SP: Fundação Seade, 2020a. Disponível em: <https://repositorio.seade.gov.br/dataset/pib-municipal-ate-2021/resource/13af6a0f-e731-4fc7-8664-73e57de8f465>. Consultado el: 23 sept. 2024.

SÃO PAULO. Fundação Sistema Estadual de Análise de Dados. **Taxa de crescimento populacional – 2000 a 2050. São Paulo, SP: Fundação Seade**, 2020b. Disponível em: <https://repositorio.seade.gov.br/dataset/populacao-por-idade-sexo/resource/2a0551df-ec74-473c-b0c3-387f0f128523>. Consultado el: 23 sept. 2024.

SÃO PAULO. Tribunal de Contas do Estado de São Paulo. **Conjunto de dados**. São Paulo, SP: TCE-SP, 2020c. Disponível em: <https://transparencia.tce.sp.gov.br/conjunto-de-dados>. Consultado el: 23 sept. 2024.

SÃO PAULO. Tribunal de Contas do Estado de São Paulo. **Projeto Audeesp**. Tabela de escrituração contábil. São Paulo, SP: TCE-SP, 2007. Disponível em: https://www.tce.sp.gov.br/sites/default/files/audesp-documentacao/tabelas_escrituracao_contabil_sistema_audeesp_versao_janeiro_2007.pdf. Consultado el: 23 sept. 2024.

SARTORIS, Alexandre. **Estatística e introdução à econometria**. 1. ed. São Paulo: Saraiva, 2003.

SAX, Christoph; EDELBUETTEL, Dirk. Seasonal adjustment by X-13ARIMA-SEATS in R. **Journal of Statistical Software**, Áustria, v. 87. n. 11, p. 1-17, dez. 2018. DOI: <http://doi.org/10.18637/jss.v087.i11>. Disponível em: <https://www.jstatsoft.org/article/view/v087i11>. Consultado el: 23 sept. 2024.

SOUZA, Celso Correia; CARNEIRO JUNIOR, João Bosco Arbués; CRISTALDO, Márcia Ferreira; CASTELÃO, Raul Asseff; FRAINER, Daniel Massen; VIAGNÓ, Hevelynne Henn da Gama. Previsão de produção de cana-de-açúcar e seus derivados, através de ARIMA, no período de 2020 a 2025, na macrorregião Centro-Sul do Brasil. **Research, society and development**, Vargem Grande Paulista, v. 11, n. 4, e6211427048, mar. 2022. DOI: <https://doi.org/10.33448/rsd-v11i4.27048>. Disponível em: <https://rsdjournal.org/rsd/article/view/27048>. Consultado el: 5 mar. 2025.

STATISTICS KINGDOM. Multiple linear regression calculator. **Statistics Kingdom**, 2017. Aplicativo. Disponível em: https://www.statskingdom.com/410multi_linear_regression.html. Consultado el: 27 sept. 2024.

TOLEDO JÚNIOR, Flávio Corrêa de. Os conteúdos da Lei Orçamentária Anual. **Revista do TCU**, Brasília, v. 118, n. 1, p. 35-44, maio/ago. 2010. Disponível em: <https://revista.tcu.gov.br/ojs/index.php/RTCU/article/view/231>. Consultado el: 28 sept. 2024.

VOLPATO, Gilson Luiz. O método lógico para redação científica. **Revista Eletrônica de Comunicação Informação & Inovação em Saúde**, Rio de Janeiro, v. 9, n. 1, p. 1-14, jan./mar. 2015. DOI: <https://doi.org/10.29397/reciis.v9i1.932>. Disponible en: <https://www.reciis.iciict.fiocruz.br/index.php/reciis/article/view/932>. Consultado el: 1º nov. 2024.

CONTRIBUCIÓN DE AUTORÍA (CRediT)

Otoniel Arruda Costa: conceptualización; metodología; investigación; curación de datos; análisis formal; software; redacción – borrador original; administración del proyecto; recursos; y visualización.

Henrique Raymundo Góia: metodología; análisis formal; software; revisión y edición; administración del proyecto; supervisión; y validación.

FINANCIAMIENTO

Esta investigación no recibió financiamiento o apoyo específico de agencia de fomento, institución pública o privada, empresa u organización sin fines de lucro.

CONFLICTO DE INTERESES

Declaramos no existir conflicto de intereses relacionado al desarrollo, a los resultados o a la publicación de este trabajo.

DISPONIBILIDAD DE LOS DATOS

Los datos que subsidiaron la elección de los municipios objeto de la investigación están disponibles en dos repositorios. El primero de ellos es el Seade Repositório, del cual se obtuvieron los datos poblacionales de los municipios del estado de São Paulo en 2020, pudiendo ser accedidos por medio del enlace <https://repositorio.seade.gov.br/dataset/populacao-por-idade-sexo/resource/2a0551df-ec74-473c-b0c3-387f0f128523> (Tasa de Crecimiento poblacional – 2000 a 2050), y los datos de los PIB de los municipios paulistas en 2020,

por medio del enlace <https://repositorio.seade.gov.br/dataset/pib-municipal-ate-2021/resource/13af6a0f-e731-4fc7-8664-73e57de8f465> (Tabla PIB - 2020). El segundo fue el portal del Tribunal de Contas del Estado de São Paulo, por medio del cual se obtuvieron los datos de los ingresos anuales de los municipios paulistas, pudiendo ser accedidos en el enlace <https://transparencia.tce.sp.gov.br/conjunto-de-dados>. Los códigos de programación (script) del modelado propuesto en este artículo están en el repositorio Zenodo (<https://zenodo.org/records/21877515>).

APROBACIÓN ÉTICA

Esta investigación no involucra procedimiento, dato o participante que exija aprobación ética.

PUBLICACIÓN DE PREPRINT

Este trabajo no se ha puesto a disposición previamente en servidor o plataforma de preprints.

USO DE IA

No fue utilizada herramienta de inteligencia artificial generativa en el desarrollo de este trabajo.

HISTORIAL DE EVALUACIÓN DEL TRABAJO

Fecha de sumisión: 4/11/2024

Fecha de publicación: 24/9/2026

Editor responsable: Alessandra Caballero Brügger Freitas (editora ejecutiva)

[Enlace al historial de evaluación completo](#) (El historial de evaluación facilitado solo está disponible en portugués.)