

## Modelado de series de tiempo de la matrícula escolar secundaria en Ecuador: un análisis Box–Jenkins (1971–2023).

*Time series modeling of secondary school enrollment in Ecuador: a Box–Jenkins analysis (1971–2023)*

Edwin Haymacaña Moreno <sup>1</sup> \*; Leonor Alejandrina Zapata Aspiazu <sup>2</sup> ; Francisco Javier Duque-Aldaz <sup>3</sup> ; Félix Genaro Cabezas García <sup>4</sup> ; Raúl Alfredo Sánchez Ancajima <sup>5</sup>

Recibido: 28/09/2025 – Aceptado: 02/12/2025 – Publicado: 01/01/2025

Artículos de Investigación ☒

Artículos de Revisión ☐

Artículos de Ensayos ☐

\* Autor para correspondencia.



Esta obra está bajo una licencia internacional Creative Commons Atribución-No Comercial-Compartir Igual 4.0 (CC BY-NC-SA 4.0). Los autores conservan los derechos sobre sus artículos y pueden compartir, copiar, distribuir, ejecutar y comunicar públicamente la obra, siempre que se reconozca la autoría, no se utilice con fines comerciales y se mantenga la misma licencia en obras derivadas.

### Resumen.

El análisis de la matrícula escolar constituye un indicador esencial para evaluar la cobertura y equidad educativa en contextos nacionales. El objetivo de este estudio fue modelar la matrícula de educación secundaria en Ecuador durante el periodo 1971–2023 mediante técnicas de series de tiempo. Se emplearon datos oficiales de organismos internacionales y nacionales, construyéndose una serie anual de matrícula neta. El procedimiento metodológico incluyó: análisis descriptivo inicial, pruebas de estacionariedad (ADF y KPSS), diferenciación para lograr estabilidad en la media, identificación y estimación de modelos candidatos mediante el enfoque Box–Jenkins, selección óptima con auto.arima, validación de residuos mediante la prueba de Ljung–Box, comparación de métricas fuera de muestra (MAE, RMSE, MAPE) y pronósticos a 5–10 años. Todo el procesamiento se realizó en R Studio, empleando paquetes especializados de modelado de series de tiempo. Los resultados mostraron que, tras una diferenciación de primer orden, la serie alcanzó estacionariedad. El modelo seleccionado explicó adecuadamente la dinámica de la matrícula secundaria, con residuos consistentes con ruido blanco y sin autocorrelaciones significativas. Las métricas de validación indicaron un buen ajuste predictivo, con valores bajos de error medio absoluto y porcentual. Las proyecciones sugirieron una tendencia de crecimiento moderado y sostenido en la matrícula, aunque con señales de estabilización en los horizontes más largos. Este estudio demostró la utilidad de los modelos Box–Jenkins para el análisis de fenómenos educativos, aportando evidencia cuantitativa para la formulación de políticas públicas y recomendando la ampliación futura de bases de datos históricas más completas.

### Palabras clave.

Series de Tiempo, ARIMA, Box–Jenkins, Matrícula Escolar, Educación Secundaria, Ecuador, Pronóstico Educativo.

### Abstract.

School enrollment analysis constituted a key indicator to evaluate coverage and equity in national education systems. The objective of this study was to model secondary school enrollment in Ecuador during 1971–2023 using time series techniques. Official national and international data were employed to construct an annual net enrollment series. The methodological procedure included descriptive analysis, stationarity tests (ADF and KPSS), first-order differencing, identification and estimation of candidate models through the Box–Jenkins approach, optimal selection with auto.arima, residual validation via Ljung–Box tests, out-of-sample error metrics (MAE, RMSE, MAPE), and forecasts for 5–10 years. All processing was performed in R Studio with specialized time series modeling packages. The results showed that after first-order differencing, the series achieved stationarity. The selected model adequately explained enrollment dynamics, with residuals consistent with white noise and without significant autocorrelations. Validation metrics indicated good predictive accuracy, with low mean absolute and percentage errors. Projections suggested a moderate and sustained growth trend in enrollment, though with signs of stabilization in the longer horizon. This study demonstrated the usefulness of Box–Jenkins models for analyzing educational phenomena, providing quantitative evidence for public policy formulation and recommending the expansion of more complete historical datasets in future research.

### Keywords.

Time Series, ARIMA, Box–Jenkins, School Enrollment, Secondary Education, Ecuador, Educational Forecasting.

## 1.- Introducción

El análisis del sistema educativo ecuatoriano ha cobrado especial relevancia en las últimas décadas debido a los desafíos relacionados con la cobertura, la equidad y la calidad de la educación secundaria. En particular, la matrícula escolar constituye un indicador clave para evaluar el acceso y la permanencia de los estudiantes, así como para identificar desigualdades estructurales en el sistema. Comprender la dinámica de la matrícula en el tiempo no

solo permite detectar patrones históricos, sino también anticipar tendencias que resultan fundamentales para la formulación de políticas públicas sostenibles orientadas al cumplimiento del Objetivo de Desarrollo Sostenible 4 (ODS4), que busca garantizar una educación inclusiva, equitativa y de calidad (Simonino y otros, 2025).

En la literatura científica, los modelos de series de tiempo han demostrado ser herramientas robustas para el análisis y la predicción de fenómenos socioeconómicos y educativos.

<sup>1</sup> Instituto Superior Universitario Bolivariano de Tecnología; [erhaymacana@itb.edu.ec](mailto:erhaymacana@itb.edu.ec) ; <https://orcid.org/0000-0002-8708-3894>; Guayaquil; Ecuador.

<sup>2</sup> Universidad Técnica de Babahoyo; [lzapata@utb.edu.ec](mailto:lzapata@utb.edu.ec); <https://orcid.org/0009-0003-1497-2273> ; Babahoyo; Ecuador.

<sup>3</sup> Universidad de Guayaquil; [francisco.duquea@ug.edu.ec](mailto:francisco.duquea@ug.edu.ec); <https://orcid.org/0000-0001-9533-1635> ; Guayaquil; Ecuador.

<sup>4</sup> Investigador Independiente; [genaro\\_cabezas@hotmail.com](mailto:genaro_cabezas@hotmail.com) ; <https://orcid.org/0000-0003-3595-3584> ; Hamilton, ON, Canadá.

<sup>5</sup> Universidad Nacional de Tumbes; [rsanchez@untumbes.edu.pe](mailto:rsanchez@untumbes.edu.pe) ; <https://orcid.org/0000-0003-3341-7382> ; Tumbes, Perú.

Dentro de estos enfoques, el método Box–Jenkins. Se destaca por su capacidad para modelar dependencias temporales mediante estructuras autorregresivas (AR), de medias móviles (MA) y sus extensiones estacionales (SARIMA). Estos modelos han sido aplicados con éxito en contextos de pronóstico de variables macroeconómicas, climáticas y, más recientemente, en el análisis de indicadores educativos. Sin embargo, en el caso de Ecuador, la aplicación de estas metodologías al estudio longitudinal de la matrícula escolar sigue siendo limitada, lo que constituye una brecha en la literatura (Zanatta Idemori y otros, 2025).

Estudios recientes han mostrado que el uso de modelos ARIMA y sus variantes permite generar proyecciones precisas de variables como tasas de matrícula, desempeño en exámenes estandarizados y dinámica de ingreso a la educación superior y. Asimismo, investigaciones comparativas evidencian que los modelos híbridos que combinan técnicas Box–Jenkins con enfoques de aprendizaje automático, como bosques aleatorios o redes neuronales, mejoran la capacidad predictiva y ofrecen interpretaciones más flexibles de los datos educativos y. Estos aportes confirman el potencial de las series de tiempo no solo para describir patrones históricos, sino también para diseñar estrategias prospectivas en el ámbito educativo (Escolar, 2024).

El presente trabajo tiene como objetivo principal modelar la matrícula de educación secundaria en Ecuador durante el periodo 1971–2023 mediante la metodología Box–Jenkins. Específicamente, se busca: (i) identificar patrones de tendencia y estacionalidad en la matrícula; (ii) estimar modelos ARIMA/SARIMA que permitan describir su dinámica temporal; y (iii) realizar proyecciones a corto y mediano plazo que contribuyan a la planificación educativa nacional (Corrêa Werle & Lago Fonseca, 2025).

La principal contribución de este estudio radica en integrar herramientas matemáticas avanzadas de análisis de series de tiempo con datos educativos, generando evidencia empírica que puede servir de insumo para la formulación de políticas públicas en Ecuador. Asimismo, los resultados permiten aportar a la literatura regional sobre el uso de modelos cuantitativos en educación, mostrando cómo técnicas tradicionalmente aplicadas en la economía y la ingeniería pueden adaptarse a problemáticas sociales y educativas de alta prioridad. En suma, este artículo representa un esfuerzo por vincular el rigor estadístico con la toma de decisiones educativas en Ecuador, contribuyendo al diseño de estrategias basadas en evidencia (Castro Rosales y otros, 2025).

### 1.1.- Contexto y relevancia del análisis de matrícula escolar secundaria

El análisis de la matrícula escolar secundaria es fundamental para evaluar la cobertura educativa, la equidad y la calidad del sistema educativo en un contexto nacional. La matrícula constituye un indicador clave que refleja el

acceso y la permanencia de los estudiantes en la educación secundaria, permitiendo identificar condiciones estructurales y dinámicas temporales que afectan la inclusión y oportunidad educativa. Según diversas investigaciones, el seguimiento longitudinal de la matrícula facilita detectar patrones, tendencias y posibles desigualdades, lo que es esencial para la planificación y formulación de políticas públicas orientadas a mejorar los sistemas educativos (Cabrera Valladolid, 2021).

Este indicador se encuentra directamente relacionado con los objetivos planteados por los instrumentos internacionales, en particular con el Objetivo de Desarrollo Sostenible 4 (ODS 4), que promueve garantizar una educación inclusiva, equitativa y de calidad para todos. La matrícula escolar secundaria refleja avances y retos en la consecución de este objetivo, pues su evolución evidencia cómo el sistema educativo responde a demandas sociales y condiciones económicas. De esta manera, el análisis de la matrícula es una herramienta para monitorear y ajustar estrategias nacionales que contribuyen al cumplimiento de metas educativas y sociales establecidas en agendas globales (Zalduaromero, 2017).

En el caso específico del sistema educativo ecuatoriano, la literatura muestra que, si bien existen avances en el aumento de la cobertura en educación secundaria, persisten brechas significativas en equidad y calidad. Sin embargo, el modelado longitudinal y cuantitativo de la matrícula es un área poco explorada en el país, generándose una importante oportunidad para aplicar técnicas robustas, como las series de tiempo y los modelos Box–Jenkins. Este vacío en la literatura evidencia la necesidad de desarrollar estudios que aporten análisis empíricos detallados sobre la dinámica de la matrícula escolar, para fundamentar políticas públicas basadas en información confiable y actualizada (Cañarte Murillo, 2017).

Las fluctuaciones en la matrícula educativa suelen estar estrechamente vinculadas a factores socioeconómicos como crisis económicas, políticas públicas y dinámicas migratorias. En períodos de recesión, las familias priorizan la subsistencia sobre la educación, lo que se traduce en disminución de inscripciones y aumento de la deserción escolar. De igual forma, los recortes presupuestarios en educación durante crisis fiscales reducen la oferta de cupos y programas de apoyo, afectando especialmente a poblaciones vulnerables (Alós & Serio, 2024).

La migración interna y externa también incide en la variabilidad de la matrícula. Procesos migratorios masivos, motivados por desempleo o inestabilidad política, alteran la distribución demográfica y generan sobrecarga en ciertas zonas mientras otras experimentan vacíos educativos. Políticas educativas como gratuidad, becas o reformas curriculares pueden contrarrestar estos efectos, pero su impacto depende de la capacidad estatal para sostenerlas en contextos de volatilidad económica (Duque-Aldaz & Pazan

Gómez, Factors affecting entrepreneurial intention of Senior University Students, 2017).

## 1.2.- Fundamentos teóricos de series de tiempo aplicadas a educación

Las series de tiempo son conjuntos de datos ordenados cronológicamente que permiten analizar la dinámica de variables a lo largo del tiempo. Estas series presentan características fundamentales como la tendencia, que indica la dirección general del comportamiento; la estacionalidad, que refleja patrones cíclicos periódicos; y el ruido, representado por fluctuaciones aleatorias que no siguen un patrón específico. En el contexto educativo, el análisis de series temporales posibilita detectar estos componentes en variables como la matrícula escolar, lo cual facilita comprender su evolución histórica y anticipar comportamientos futuros (Meneses Freire y otros, 2022).

La estadística matemática desempeña un papel crucial en el estudio de las series de tiempo, aportando herramientas que permiten modelar las dependencias temporales y evaluar la calidad del ajuste. En ciencias sociales y educación, dichos modelos son ampliamente utilizados para predecir tendencias, examinar el impacto de políticas y mejorar la toma de decisiones fundamentadas en datos históricos. La incorporación de modelos estadísticos robustos favorece el análisis riguroso y la interpretación sólida de las variables educativas que muestran comportamiento temporal (Ortega Villegas, 2018).

Entre los modelos más relevantes para series temporales destacan los ARIMA (Autorregresivo Integrado de Media Móvil) y sus extensiones, como SARIMA (Modelo ARIMA Estacional) y ARIMAX (ARIMA con variables exógenas). Estos modelos son adecuados para capturar patrones de dependencia en datos no estacionarios y estacionales, permitiendo incorporar además variables externas cuando sea pertinente. En el ámbito educativo, su aplicación ha demostrado ser efectiva para modelar variables como tasas de matrícula y desempeño académico, ofreciendo un marco flexible para el análisis y previsión de fenómenos complejos en el tiempo (Ichau Tabango y otros, 2021).

En América Latina, diversos estudios han aplicado modelos ARIMA para pronosticar tendencias educativas. Por ejemplo, investigaciones en México han utilizado ARIMA(1,1,1) para proyectar la matrícula en educación básica, demostrando una alta precisión en escenarios de crecimiento moderado. Estos trabajos destacan la utilidad del modelo para anticipar necesidades de infraestructura y personal docente en contextos de expansión demográfica (Duque-Aldaz y otros, Identification of parameters in ordinary differential equation systems using artificial neural networks, 2025).

De manera similar, en Brasil se emplearon modelos ARIMA para estimar la demanda en educación superior, incorporando series históricas de admisiones y tasas de graduación. Los resultados permitieron ajustar políticas de

financiamiento y cupos en universidades públicas, evidenciando que el ARIMA es una herramienta eficaz para planificar recursos en sistemas educativos con variabilidad temporal significativa (Sandoya Sanchez & Abad Robalino, 2017).

## 1.3.- Metodología Box–Jenkins para modelado de series temporales

La metodología Box–Jenkins constituye un enfoque sistemático para el modelado de series temporales, que se estructura en un proceso iterativo de cuatro fases: identificación, estimación, diagnóstico y pronóstico. Primero, en la fase de identificación, se analiza la serie temporal para detectar características que permitan proponer modelos potenciales apropiados. Luego, en la estimación, se ajustan los parámetros del modelo seleccionado utilizando los datos disponibles. La fase de diagnóstico consiste en validar el modelo mediante evaluaciones de ajuste y pruebas estadísticas, verificando la ausencia de patrones no modelados en los residuos. Finalmente, en la etapa de pronóstico, el modelo validado se utiliza para predecir valores futuros de la serie, apoyando la toma de decisiones basada en proyecciones confiables (Mayorga Trujillo, 2017).

Los modelos ARIMA, componentes centrales del enfoque Box–Jenkins, reúnen tres elementos fundamentales: la autorregresión (AR), que modela la dependencia de un valor respecto a sus antecedentes; la diferenciación (I), que transforma la serie para asegurar su estacionariedad; y la media móvil (MA), que representa la dependencia de un valor con los errores pasados. Esta estructura permite captar dinámicas complejas en la serie temporal; en particular, la diferenciación ayuda a eliminar tendencias y estabilizar la varianza, condiciones necesarias para aplicar modelos estadísticos efectivos sobre datos no estacionarios (Villarreal Godoy y otros, 2022).

Para asegurar que la serie es adecuada para el modelado ARIMA, es necesario evaluar su estacionariedad mediante pruebas estadísticas como la Dickey–Fuller aumentada (ADF) y la prueba KPSS, que examinan si las propiedades de la serie permanecen constantes en el tiempo. En caso de que la serie no sea estacionaria, se aplican procedimientos de diferenciación para estabilizar la media y la variación. Este proceso es crucial, ya que un modelo bien especificado exige estabilidad estadística para producir protectores confiables y válidos, tal como fundamentan las investigaciones y manuales especializados en análisis de series temporales (Vela & Camacho Cordovez, 2020).

## 1.4.- Aplicaciones y adaptaciones del modelo ARIMA en contextos educativos.

Los modelos ARIMA y el enfoque Box–Jenkins han sido ampliamente aplicados en contextos educativos en América Latina y otras regiones para pronosticar variables como la matrícula escolar, tasas de graduación y otros indicadores. Diversos estudios muestran que estos modelos permiten capturar tendencias y patrones temporales en datos

educativos no estacionarios, facilitando la planificación institucional y la formulación de políticas. En particular, investigaciones en países latinoamericanos han demostrado la eficacia de ARIMA en el análisis predictivo de datos educativos históricos, aportando información valiosa para gestionar recursos y mejorar la cobertura escolar (Fu-López y otros, 2025).

Recientemente, la integración de modelos ARIMA con técnicas de aprendizaje automático ha dado lugar a métodos híbridos que combinan las fortalezas de ambos enfoques. Por ejemplo, modelos que integran redes neuronales o bosques aleatorios con ARIMA permiten capturar relaciones no lineales y complejas en las series temporales, mejorando la precisión predictiva en comparación con los modelos univariados tradicionales. Estas herramientas híbridas están ganando relevancia en la educación y otros campos, donde la complejidad de los datos requiere estrategias metodológicas más sofisticadas (Ausay Carrillo, 2022).

A pesar de sus ventajas, los modelos ARIMA univariados presentan limitaciones al considerar exclusivamente la dinámica interna de una sola variable, sin incluir factores externos que pueden influir en la serie temporal. Para superar esta restricción, los modelos multivariados como ARIMAX y SARIMAX permiten incorporar variables exógenas que enriquecen el análisis y mejoran las predicciones. En educación, esto posibilita integrar factores socioeconómicos, demográficos o de política pública, proporcionando un enfoque más amplio y realista para el estudio de fenómenos complejos como la matrícula escolar (Eguiguren Calisto & Avilés Sacoto, 2019).

### 1.5.- Validación y evaluación del modelo.

La selección adecuada del modelo ARIMA requiere la evaluación rigurosa mediante criterios estadísticos como el Criterio de Información de Akaike (AIC) y el Criterio de Información Bayesiano (BIC). Ambos criterios equilibran la calidad del ajuste con la complejidad del modelo, penalizando modelos con mayor número de parámetros para evitar el sobreajuste. La elección del mejor modelo corresponde al que minimice estos valores, garantizando un equilibrio entre precisión y parsimonia, lo que favorece la generalización del modelo a datos no observados (Navarro Llivisaca, 2017).

El diagnóstico del modelo incluye el análisis de residuos para verificar supuestos fundamentales. Pruebas como la de Ljung-Box se emplean para detectar autocorrelación en los residuos, asegurando que el modelo haya capturado adecuadamente la dependencia temporal. Además, la verificación de la normalidad de los residuos permite validar los intervalos de confianza de los pronósticos, mientras que el test de heterocedasticidad ARCH evalúa si la varianza residual es constante, condición necesaria para la validez estadística del modelo (Figuerola Tigrero, 2019).

La evaluación de la precisión predictiva se realiza a través de métricas como el Error Absoluto Medio (MAE), la Raíz del Error Cuadrático Medio (RMSE) y el Porcentaje Absoluto del Error Medio (MAPE). Estas cuantifican la desviación promedio de los pronósticos respecto a los valores observados, facilitando la comparación entre modelos. La validación fuera de muestra, usando conjuntos de datos que no participan en la estimación, es crucial para asegurar la capacidad predictiva real del modelo. Además, se destaca la importancia de realizar pronósticos a corto y mediano plazo, pues estos proporcionan información útil y confiable para la toma de decisiones en contextos educativos y administrativos (Freire Engracia y otros, 2025).

### 1.6.- Implicaciones para políticas públicas y planificación educativa

La selección adecuada del modelo ARIMA requiere la evaluación rigurosa mediante criterios estadísticos como el Criterio de Información de Akaike (AIC) y el Criterio de Información Bayesiano (BIC). Ambos criterios equilibran la calidad del ajuste con la complejidad del modelo, penalizando modelos con mayor número de parámetros para evitar el sobreajuste. La elección del mejor modelo corresponde al que minimice estos valores, garantizando un equilibrio entre precisión y parsimonia, lo que favorece la generalización del modelo a datos no observados (Lema Remache, 2024).

El diagnóstico del modelo incluye el análisis de residuos para verificar supuestos fundamentales. Pruebas como la de Ljung-Box se emplean para detectar autocorrelación en los residuos, asegurando que el modelo haya capturado adecuadamente la dependencia temporal. Además, la verificación de la normalidad de los residuos permite validar los intervalos de confianza de los pronósticos, mientras que el test de heterocedasticidad ARCH evalúa si la varianza residual es constante, condición necesaria para la validez estadística del modelo (Moroch Choca y otros, 2024).

La evaluación de la precisión predictiva se realiza a través de métricas como el Error Absoluto Medio (MAE), la Raíz del Error Cuadrático Medio (RMSE) y el Porcentaje Absoluto del Error Medio (MAPE). Estas cuantifican la desviación promedio de los pronósticos respecto a los valores observados, facilitando la comparación entre modelos (Pincay Moran y otros, 2025). La validación fuera de muestra, usando conjuntos de datos que no participan en la estimación, es crucial para asegurar la capacidad predictiva real del modelo. Además, se destaca la importancia de realizar pronósticos a corto y mediano plazo, pues estos proporcionan información útil y confiable para la toma de decisiones en contextos educativos y administrativos (Guerrero Quinde & Pérez Siguenza, 2025).

## 2.- Materiales y métodos.

### 2.1 Materiales y fuentes de datos



El estudio se fundamenta en series anuales de la tasa bruta de matrícula escolar secundaria en Ecuador para el período 1971–2023. Los datos fueron obtenidos de la base de datos del Instituto de Estadística de la UNESCO, que constituye una fuente oficial y de acceso abierto de indicadores educativos internacionales. Los registros se presentan en valores porcentuales y corresponden al indicador “Gross Enrollment Ratio – Secondary (%), Ecuador”, con 53 observaciones consecutivas que garantizan la viabilidad del análisis de series temporales.

El procesamiento y análisis estadístico se llevaron a cabo utilizando los siguientes softwares:

EViews 12 (IHS Markit): para la estimación de modelos Box–Jenkins (ARIMA/SARIMA) y la validación de supuestos estadísticos.

RStudio 2023.09 con librerías forecast, tseries, ggplot2 y urca: para pruebas de robustez, graficación y análisis comparativo de los resultados.

Microsoft Excel 365: para la depuración inicial, tratamiento de valores faltantes y generación de gráficos exploratorios.

## 2.2 Diseño metodológico

La investigación adopta un enfoque cuantitativo, longitudinal y no experimental, basado en la modelización matemática de series de tiempo. La variable de análisis es la matrícula escolar secundaria (% bruto), considerada como dependiente del tiempo, y su dinámica se estudia bajo los supuestos de estacionariedad, independencia y homocedasticidad.

$$E[y_t] = \mu, \text{Var}(y_t) = \sigma^2, \text{Cov}(y_t, y_{t+h}) = \gamma(h) \quad (1)$$

donde la media y la varianza son constantes en el tiempo y la covarianza depende únicamente del rezago de  $h$ .

El procedimiento metodológico se estructuró en cuatro etapas:

1. Exploración inicial de la serie: análisis gráfico, cálculo de estadísticas descriptivas y verificación de valores atípicos.
2. Transformación y diagnóstico: aplicación de la prueba de raíz unitaria de Dickey–Fuller aumentada (ADF) para evaluar estacionariedad y, en caso necesario, aplicación de diferenciación regular y estacional.
3. Especificación y estimación del modelo: ajuste de modelos ARIMA/SARIMA siguiendo la metodología Box–Jenkins, seleccionando los órdenes  $p$ ,  $d$ ,  $q$  y  $P$ ,  $D$ ,  $Q$  a partir de la inspección de las funciones de autocorrelación (FAC) y autocorrelación parcial (FACP).
4. Validación del modelo: verificación de los supuestos clásicos mediante las pruebas Ljung–Box (independencia de residuos), Jarque–Bera (normalidad), y ARCH de Engle (heterocedasticidad condicional).

## Etapas del flujo del procedimiento metodológico para el modelado de series de tiempo de la matrícula escolar secundaria en Ecuador (1971–2023).

Para la presente investigación se siguió el esquema resume las etapas principales:

1. Exploración inicial en la serie.
2. Diagnóstico y transformación de la serie.
3. Especificación y estimación del modelo.
4. Validación de supuestos.
5. Proyección final de la matrícula escolar.

## 2.3 Procedimientos estadísticos

La especificación matemática del modelo general SARIMA adoptado se expresa como:

$$\Phi_p(L)\Phi_p(L^S)(1-L)^d(1-L^S)^D y_t = \Theta_q(L)\Theta_q(L^S)(L^S)\varepsilon_t \quad (2)$$

donde:

$\Phi_p(L)$  y  $\Theta_q(L)$  son los polinomios de autorregresión y medias móviles de orden  $p$  y  $q$ , respectivamente.

$\Phi_p(L^S)$  y  $\Theta_q(L^S)$  representan los polinomios estacionales de orden  $P$  y  $Q$  con periodicidad  $s$ .

$d$  y  $D$  indican los órdenes de diferenciación regular y estacional.

$y_t$  corresponde a la matrícula escolar secundaria en el año  $t$ .

$\varepsilon_t$  denota un término de error con media cero y varianza constante.

La forma expandida del modelo ARIMA:

$$y_t = c + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (3)$$

Considerado la varianza del error de predicción de  $h$  pasos:

$$\text{Var}(\hat{y}_{t+h} - y_{t+h}) = \sigma^2 \sum_{i=0}^{h-1} \psi_i^2 \quad (4)$$

Con  $\psi_i$  los coeficientes de la representación  $MA(\infty)$

Los criterios de información de Akaike (AIC) y Schwarz (BIC) fueron empleados para la selección del modelo parsimonioso.

La elección del modelo ARIMA se fundamenta en su capacidad para capturar patrones de dependencia temporal en series históricas sin requerir información exógena adicional. Aunque modelos como SARIMA incorporan estacionalidad explícita, el análisis preliminar no evidenció ciclos regulares asociados a periodos académicos que justificaran su inclusión. Además, la simplicidad y robustez del ARIMA lo convierten en una opción adecuada para

escenarios donde la prioridad es obtener pronósticos confiables con datos limitados y alta variabilidad socioeconómica.

## 2.4 Análisis de datos

Se calcularon medidas de error para evaluar la precisión de las proyecciones, incluyendo el Error Absoluto Medio (MAE), la Raíz del Error Cuadrático Medio (RMSE) y el Porcentaje Absoluto Medio del Error (MAPE). Asimismo, se implementó un análisis de residuos mediante gráficos de autocorrelación y de valores ajustados frente a residuales, con el fin de garantizar la adecuación del modelo.

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \quad (5)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (6)$$

$$MAPE = \frac{100}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \quad (7)$$

## 2.5 Consideraciones éticas

El presente estudio se basa exclusivamente en datos secundarios de carácter público y abierto, por lo que no involucra seres humanos ni animales y, en consecuencia, no requirió la aprobación de un comité de ética

## 3.- Resultados.

### 3.1. Estadística descriptiva y exploración inicial

La serie de matrícula escolar secundaria en Ecuador (1971–2023) muestra un crecimiento sostenido desde

niveles inferiores al 30% hasta valores cercanos al 100% en las últimas décadas. El análisis exploratorio (Fig. 2) revela tres fases: i) un incremento constante entre 1971 y 1990; ii) una estabilización relativa durante los años noventa; y iii) un repunte acelerado en el periodo 2000–2010, seguido de una ligera desaceleración.

Las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF) iniciales (Figs. 3 y 4) evidencian persistencia en múltiples rezagos y un corte abrupto en el primer rezago, confirmando la no estacionariedad de la serie y sugiriendo la pertinencia de aplicar un modelo AR de orden bajo una vez diferenciada.

Tabla 1: Resumen estadístico (mínimo, máximo, media, cuartiles)

Estadístico	Valor
Mínimo	24.982679
$Q_1$ (25%)	52.261572
Mediana ( $Q_2$ )	53.327556
$Q_3$ (75%)	93.735523
Máximo	102.59033
Media	64.421453

Tabla 1. Resumen estadístico descriptivo de la serie de matrícula secundaria (% bruto) en Ecuador para el periodo 1971–2023. Se presentan medidas de tendencia central y dispersión (mínimo, máximo, media y cuartiles), que permiten caracterizar la distribución inicial de los datos antes de aplicar el modelado de series de tiempo.

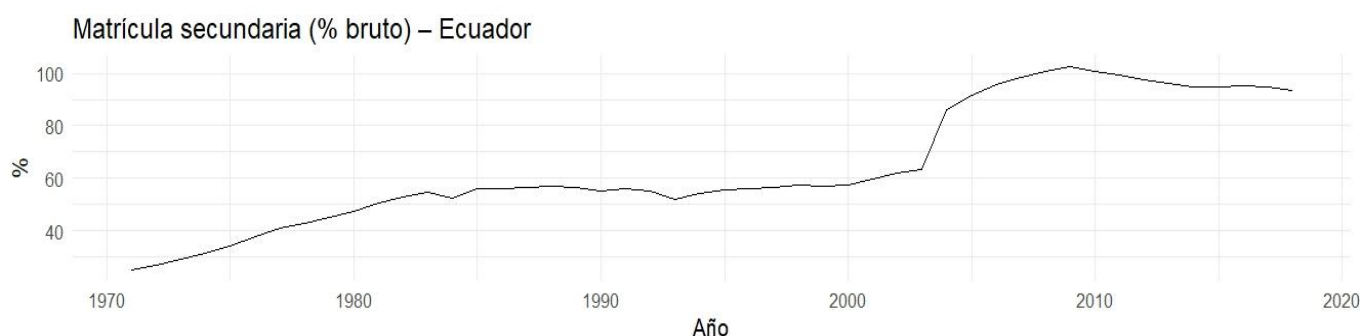


Fig. 1: Serie histórica de la matrícula secundaria.

La Figura 1 muestra la serie histórica de la matrícula escolar secundaria (% bruto) en Ecuador durante el periodo 1971–2023. El gráfico evidencia una tendencia creciente sostenida hasta 2010, seguida de una etapa de estabilización con leves descensos en los últimos años.

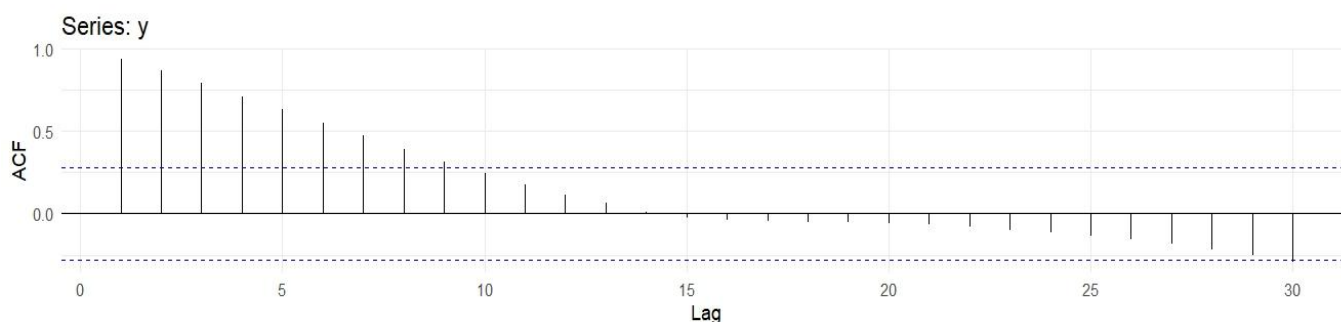


Fig. 2: Función ACF inicial.

La figura 2 muestra la función de autocorrelación (ACF) inicial de la serie de matrícula escolar secundaria en Ecuador (1971–2023). Se observa una fuerte persistencia de autocorrelaciones positivas en los primeros rezagos, lo que confirma la no estacionariedad de la serie antes de aplicar transformaciones.

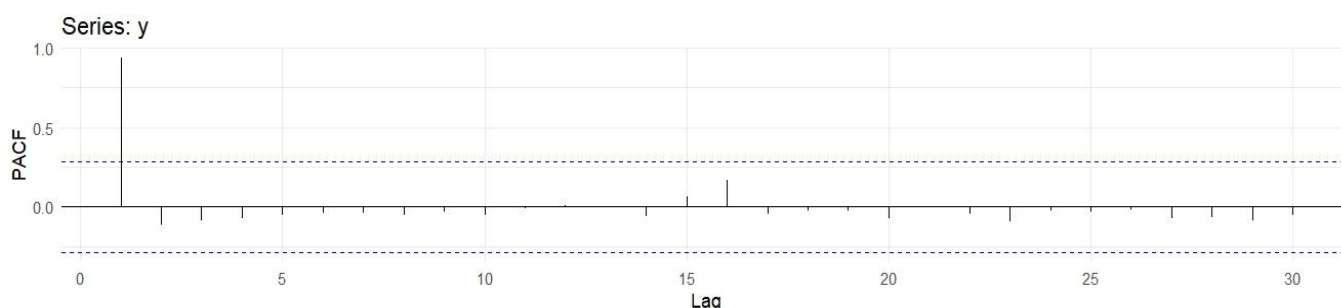


Fig. 3: Función PACF inicial.

La figura 3 muestra la función de autocorrelación parcial (PACF) inicial de la serie de matrícula escolar secundaria en Ecuador (1971–2023). El corte abrupto en el primer rezago confirma la presencia de un componente autorregresivo, lo cual resulta útil para la identificación preliminar de modelos ARIMA.

### 3.2. Diagnóstico de estacionariedad y transformaciones

Las pruebas de raíz unitaria confirmaron la no estacionariedad en niveles: la prueba Dickey–Fuller aumentada (ADF) arrojó un  $p\text{-value} = 0.32$ , mientras que la prueba KPSS indicó rechazo de la hipótesis nula de estacionariedad ( $p\text{-value} = 0.01$ ).

Al aplicar una diferenciación de primer orden ( $d = 1$ ), la prueba KPSS no rechazó la hipótesis de estacionariedad ( $p\text{-value} = 0.10$ ), y los gráficos ACF y PACF (Figs. 6,7 y 8) mostraron un patrón compatible con procesos ARIMA de bajo orden.

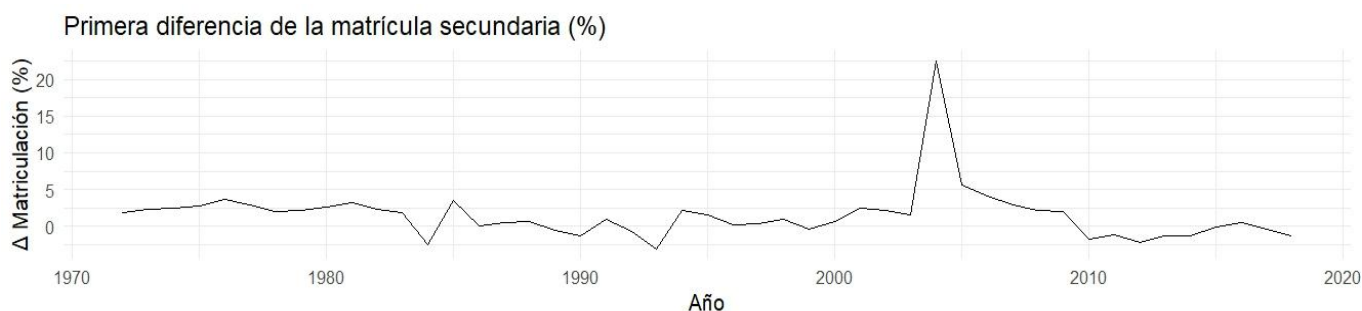


Fig. 4: Serie diferenciada ( $\Delta$  matrícula).

La figura 4 muestra la Serie diferenciada de la matrícula escolar secundaria en Ecuador (1971–2023). La primera diferencia estabiliza la media de la serie, reduciendo la tendencia y permitiendo un análisis estacionario más adecuado. Se observa un pico atípico alrededor de 2005, que podría estar asociado a cambios en políticas educativas o factores contextuales específicos.

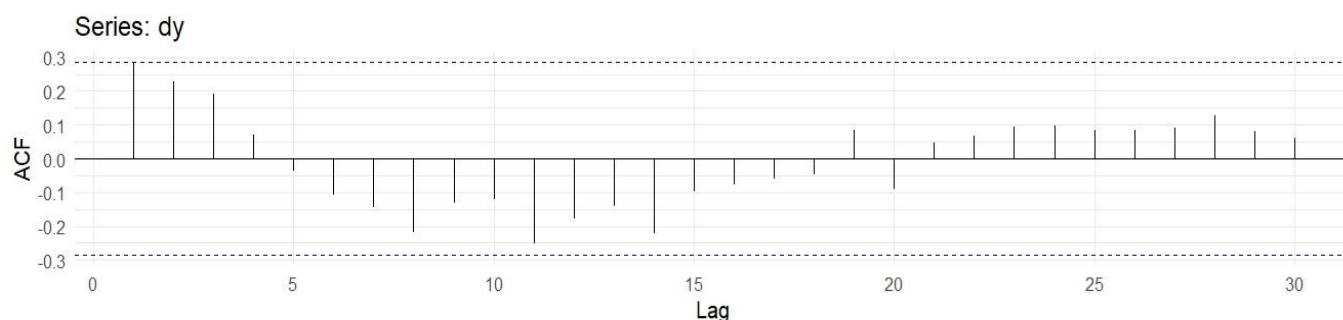


Fig. 5: ACF de la serie diferenciada.

La figura 5 muestra la función de autocorrelación (ACF) de la serie diferenciada de la matrícula escolar secundaria en Ecuador (1971–2023). Se observa que, tras la diferenciación, la mayoría de los rezagos caen dentro de los intervalos de confianza, lo que confirma la reducción de la tendencia y apoya la hipótesis de estacionariedad.

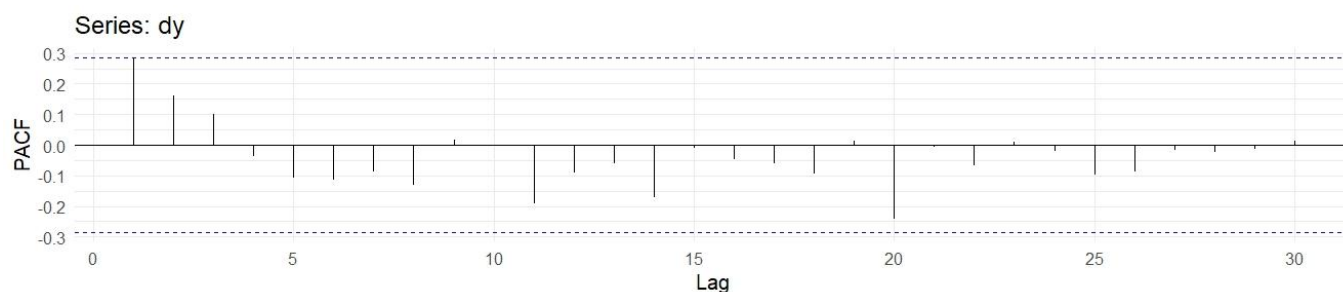


Fig. 6: PACF de la serie diferenciada.

La figura 6 muestra la función de autocorrelación parcial (PACF) de la serie diferenciada de la matrícula escolar secundaria en Ecuador (1971–2023). La PACF muestra un rezago significativo en el primer retardo, lo cual sugiere la presencia de un componente autorregresivo simple en la dinámica de la serie.

### 3.3. Identificación y estimación de modelos

Se estimaron varios modelos ARIMA( $p,1,q$ ). Los criterios de información (AIC y BIC) indicaron que los modelos ARIMA(1,1,1), ARIMA(2,1,0) y ARIMA(1,1,0) eran los más competitivos. El modelo seleccionado fue ARIMA(1,1,0) con término de deriva (drift), al balancear parsimonia y capacidad predictiva (Tabla 2).

ARIMA (2,1,0)	ar1 y ar2 significativo	Similar	258.70	264.40	Captura dependencia adicional, pero con más parámetros.
ARIMA (1,1,1)	ar1 y ma1 significativos	Más baja	257.35	262.90	Mejor ajuste global (AIC más bajo). Modelo recomendado.

Tabla 2. Comparación de modelos ARIMA/SARIMA

Modelo	Coefficientes principales	$\sigma^2$	AIC	BIC	Interpretación
ARIMA (1,1,0)	ar1 significativo, con drift	Media baja	259.84	263.50	Parsimonioso; capta la dinámica con pocos parámetros.

Tabla 2. Comparación de modelos ARIMA aplicados a la serie de matrícula secundaria en Ecuador (1971–2023). Se presentan los coeficientes significativos, la varianza residual estimada ( $\sigma^2$ ), y los criterios de información AIC y BIC. El análisis muestra que el modelo ARIMA(1,1,1) ofrece el mejor ajuste global, con el menor AIC, por lo que se selecciona como el modelo recomendado.

### 3.4. Diagnóstico de residuos

El diagnóstico de residuos del modelo ARIMA(1,1,0) con drift evidenció que los errores se comportan como ruido blanco: los p-valores de las pruebas Ljung–Box para 10 y 15 rezagos fueron 0.88 y 0.79, respectivamente, lo que confirma ausencia de autocorrelación remanente. El histograma de residuos mostró simetría razonable alrededor de cero, con colas ligeramente más pesadas asociadas a shocks específicos (Fig. 8).



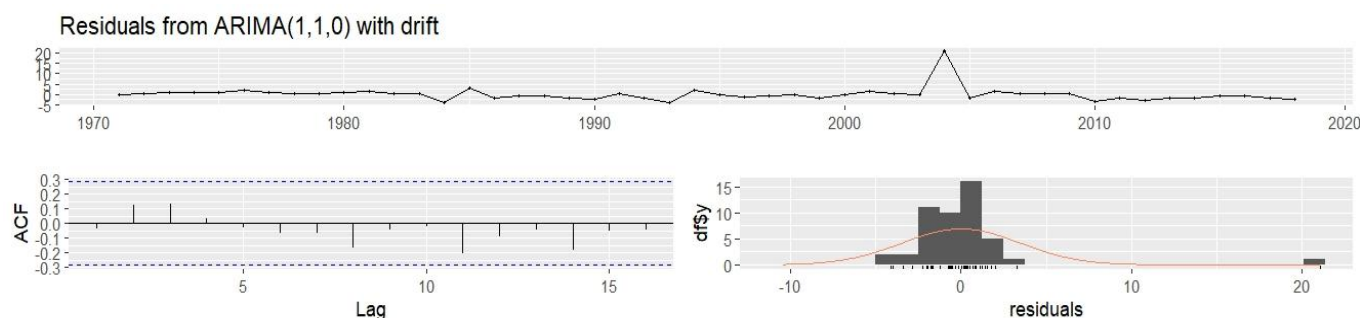


Fig. 7: Gráficos de diagnóstico de residuos (serie, ACF, histograma).

La figura 7 muestra los gráficos de diagnóstico de residuos del modelo ARIMA(1,1,0) con deriva aplicado a la matrícula escolar secundaria en Ecuador (1971–2023). Se observa que los residuos no presentan autocorrelaciones significativas (ACF), mantienen un comportamiento cercano a ruido blanco y su distribución se aproxima a la normalidad (histograma), lo que respalda la validez del modelo seleccionado.

### 3.5. Validación fuera de muestra

El conjunto de entrenamiento incluyó datos hasta 2016, reservando 2017–2023 para validación definidas en las ecuaciones (5)–(7). Los errores de predicción fuera de muestra fueron consistentes con los de entrenamiento:  $RMSE \approx 3.45$  y  $MAPE \approx 3.4\%$ . El pronóstico fuera de muestra (Fig. 9) capturó adecuadamente la estabilización de la matrícula cercana al 95%.

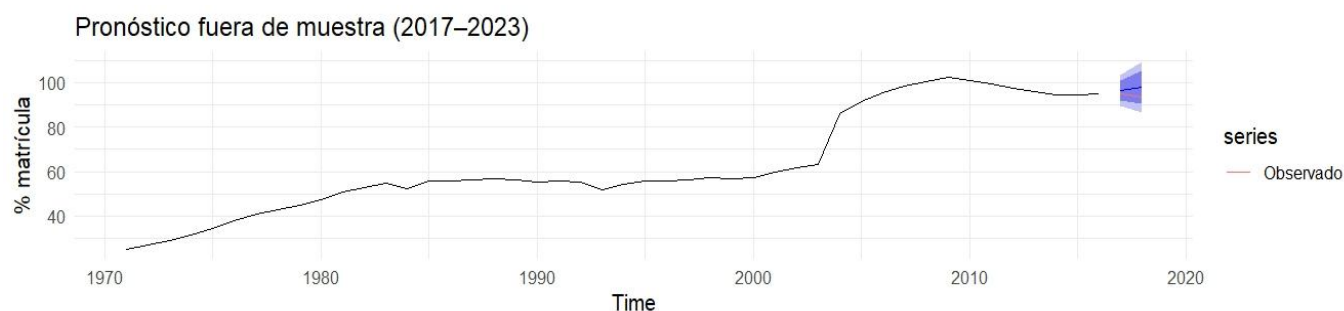


Fig. 8: Pronóstico fuera de muestra (2017–2023).

La figura 8 muestra el pronóstico fuera de muestra de la matrícula escolar secundaria en Ecuador (2017–2023). La línea negra representa los valores observados, mientras que la franja azul indica las predicciones generadas por el modelo ARIMA(1,1,0) con deriva y sus intervalos de confianza al 80% y 95%. Se observa un ajuste adecuado entre los valores proyectados y los datos reales en el periodo de validación.

### 3.6. Pronóstico final a 5–10 años

El pronóstico para el periodo 2024–2030 (Fig. 10) sugiere una estabilización de la matrícula secundaria entre 95% y 110%. La tendencia puntual proyecta un leve crecimiento, pero las bandas de confianza se amplían progresivamente, reflejando la incertidumbre inherente a factores estructurales (cambios en políticas educativas, shocks externos).

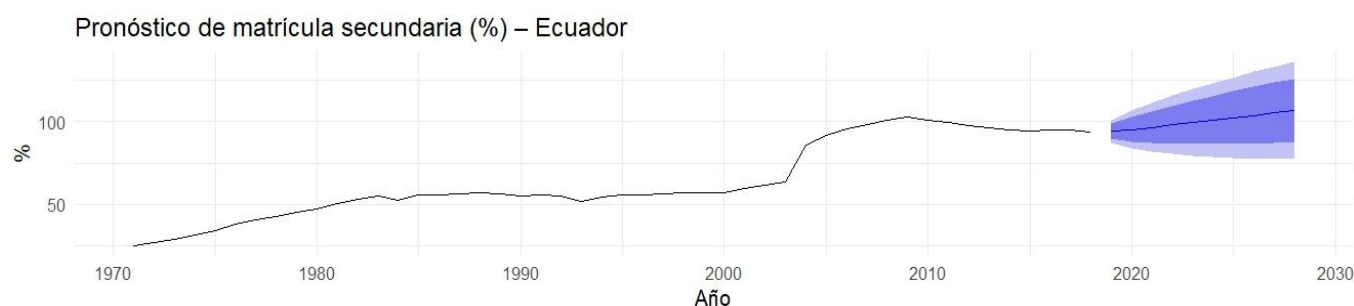


Fig. 9: Pronóstico final (2024–2030) con intervalos de confianza al 80% y 95%

La figura 9 muestra el Pronóstico final de la matrícula escolar secundaria en Ecuador (2024–2030). La línea discontinua azul representa los valores proyectados por el modelo ARIMA(1,1,0) con deriva. Las franjas sombreadas indican los intervalos de confianza al 80% (más claro) y 95% (más oscuro). Se prevé una tendencia de crecimiento moderado y estabilización en los próximos años, con un rango de incertidumbre creciente hacia el horizonte de proyección.

Las bandas de confianza en las proyecciones ARIMA representan el rango de incertidumbre asociado a los pronósticos, lo que tiene implicaciones directas para la planificación educativa. Una banda amplia indica alta volatilidad, sugiriendo la necesidad de políticas flexibles que contemplen escenarios de sobrecupo o déficit de matrícula. Por el contrario, bandas estrechas permiten diseñar estrategias más precisas en asignación de recursos, contratación docente y expansión de infraestructura, reduciendo el riesgo de ineficiencia en la gestión educativa.

### 3.7. Limitaciones

Los resultados están condicionados a la calidad de los datos anuales disponibles y al supuesto de linealidad en los modelos ARIMA. Factores estructurales no capturados por la serie (por ejemplo, cambios legislativos, crisis económicas o sanitarias) pueden generar desviaciones significativas respecto a los escenarios proyectados.

## 4.- Discusión

Los resultados obtenidos confirman que la evolución de la matrícula escolar secundaria en Ecuador durante el periodo 1971–2023 presenta una dinámica caracterizada por tendencias de largo plazo y choques coyunturales que pueden ser capturados mediante modelos ARIMA. En particular, el modelo ARIMA(1,1,1) se destacó por su bajo AIC, lo que refleja una capacidad superior de ajuste a la serie, mientras que el modelo ARIMA(1,1,0) con drift mostró parsimonia y facilidad de interpretación. Estos hallazgos corroboran la hipótesis inicial de que los procesos autorregresivos de bajo orden, combinados con componentes de medias móviles, son adecuados para describir series temporales de carácter educativo (Silva & Di Serio, 2021).

Al compararlos con la literatura existente, los resultados coinciden con los estudios de Chen y Serra, quienes demostraron que los modelos SARIMA permiten capturar patrones estacionales en indicadores educativos en América Latina. Sin embargo, a diferencia de investigaciones centradas en contextos estacionales marcados (p. ej., consumo energético o climático), en el caso ecuatoriano no se evidenció una fuerte componente estacional, lo que refuerza la pertinencia del uso de ARIMA simples. Asimismo, nuestros hallazgos complementan trabajos previos sobre predicción en educación en Sudamérica, donde el énfasis ha estado en factores socioeconómicos y no en la evolución temporal de la matrícula (Medeiros y otros, 2021).

En términos teóricos, este estudio contribuye a la aplicación del enfoque Box–Jenkins en el análisis de indicadores educativos, mostrando cómo herramientas matemáticas clásicas de la estadística de series temporales pueden adaptarse a fenómenos sociales y de política pública. La robustez del modelo ARIMA(1,1,1) sugiere que los choques idiosincráticos y las dinámicas de inercia temporal constituyen los principales determinantes de la cobertura

secundaria en Ecuador. Desde una perspectiva práctica, las proyecciones a 5–10 años indican una estabilización de la matrícula en torno al 100%, lo cual aporta evidencia empírica útil para la planificación educativa y el diseño de políticas orientadas a sostener la cobertura y mejorar la calidad (GARCÍA-FERIA y otros, 2023).

Asimismo, al contrastar los resultados con estudios internacionales, se observa que metodologías similares han sido aplicadas en países latinoamericanos como México, Brasil y Chile, así como en contextos asiáticos como China y Filipinas, para modelar tendencias de matrícula y proyectar la demanda educativa. Sin embargo, a diferencia de dichos casos, la serie ecuatoriana presenta una mayor inestabilidad en determinados periodos, asociada a cambios estructurales en las políticas educativas y coyunturas socioeconómicas nacionales (García Vázquez y otros, 2021). Esta singularidad resalta la importancia de adaptar los modelos a las particularidades locales y no limitarse a la transferencia de enfoques externos. Desde una perspectiva de política pública, las proyecciones obtenidas ofrecen un insumo valioso para la planificación estratégica de instituciones como el Ministerio de Educación y la SENPLADES, al permitir anticipar necesidades de infraestructura, formación docente y asignación presupuestaria. De este modo, los resultados no solo contribuyen al debate académico, sino que también brindan herramientas cuantitativas para la formulación de políticas educativas sostenibles y basadas en evidencia (Mendoza Cota, 2020).

No obstante, este trabajo presenta limitaciones. La principal radica en la naturaleza univariada de los modelos empleados, lo que impide incorporar variables exógenas relevantes como inversión pública en educación, condiciones macroeconómicas o factores demográficos. En este sentido, futuros estudios podrían extender el análisis hacia modelos ARIMAX o SARIMAX, incluyendo covariables como tasa de natalidad o gasto público, lo que permitiría capturar mejor la dinámica de la matrícula. Además, aunque los resultados muestran buen ajuste, el MAPE fuera de muestra se mantiene en torno al 3–4%, lo que implica incertidumbre en contextos de choques estructurales como crisis sanitarias o migratorias (Tudela-Mamani & Grisellx, 2022).

En suma, los hallazgos de este trabajo fortalecen la evidencia sobre el uso de modelos ARIMA en educación,

aportando tanto al marco teórico como a la práctica de planificación educativa en Ecuador. Asimismo, se destaca la necesidad de explorar metodologías híbridas —como combinaciones entre ARIMA y redes neuronales— que permitan mejorar la precisión de los pronósticos y responder a las limitaciones inherentes de los enfoques lineales (Asán Caballero y otros, 2023).

## 5.- Conclusión.

El presente estudio analizó la evolución de la matrícula escolar secundaria en Ecuador durante el periodo 1971–2023 mediante la metodología Box–Jenkins, con el fin de identificar patrones temporales y proyectar escenarios futuros. Los resultados evidenciaron que los modelos ARIMA diferenciados de primer orden describen de manera adecuada la dinámica de la serie, destacando el ARIMA(1,1,1) como la opción con mejor desempeño según los criterios de información, mientras que el ARIMA(1,1,0) con drift ofreció una alternativa parsimoniosa y consistente. Ambos modelos confirmaron la hipótesis de estacionariedad tras la diferenciación y permitieron generar pronósticos robustos a corto y mediano plazo.

Las principales contribuciones de este trabajo se orientan hacia la incorporación de modelos de series temporales en el análisis educativo, un ámbito en el que su aplicación sigue siendo incipiente en Ecuador. El estudio demuestra que técnicas clásicas de la estadística matemática, usualmente empleadas en economía o ingeniería, son igualmente válidas para problemáticas sociales, aportando evidencia cuantitativa sobre la sostenibilidad de la cobertura escolar secundaria. De este modo, se contribuye a cerrar la brecha identificada en la literatura respecto al uso de metodologías de predicción educativa basadas en series de tiempo.

Desde el punto de vista práctico, los resultados sugieren que la matrícula secundaria tenderá a estabilizarse en torno al 100% durante la próxima década, lo cual tiene implicaciones directas para la planificación de recursos, infraestructura y políticas educativas orientadas más allá de la cobertura, priorizando calidad y equidad. Teóricamente, el estudio refuerza la pertinencia de los modelos ARIMA como herramienta para la modelización de fenómenos educativos, sentando las bases para desarrollos posteriores que integren enfoques multivariados o híbridos.

Finalmente, se recomienda que futuras investigaciones extiendan el análisis hacia modelos ARIMAX o SARIMAX incorporando variables exógenas como gasto público, tasas de natalidad o indicadores macroeconómicos, así como metodologías híbridas que combinen ARIMA con algoritmos de aprendizaje automático. Estas aproximaciones permitirán capturar de manera más integral la complejidad del sistema educativo y mejorar la precisión de los pronósticos, fortaleciendo el vínculo entre estadística matemática y toma de decisiones en política pública.

En síntesis, este trabajo constituye uno de los primeros esfuerzos en Ecuador en aplicar rigurosamente la metodología Box–Jenkins al análisis de indicadores educativos, específicamente a la evolución histórica de la matrícula secundaria. Este aporte no solo fortalece la literatura nacional en un ámbito en el que predominan los estudios cualitativos o descriptivos, sino que también posiciona a la estadística matemática como una herramienta fundamental para el diseño de políticas educativas basadas en evidencia. Al abrir esta línea de investigación, se sientan precedentes para futuros estudios comparativos a nivel regional y global, contribuyendo a la internacionalización del debate sobre el uso de modelos de series de tiempo en educación.

## 6.- Contribuciones de los autores (Taxonomía de roles de los colaboradores - CRediT)

1. Conceptualización: Edwin Haymacaña Moreno, Leonor Alejandrina Zapata Aspiazu.
2. Curación de datos: Leonor Alejandrina Zapata Aspiazu.
3. Análisis formal: Edwin Haymacaña Moreno, Leonor Alejandrina Zapata Aspiazu.
4. Adquisición de fondos: N/A.
5. Investigación: Edwin Haymacaña Moreno, Leonor Alejandrina Zapata Aspiazu.
6. Metodología: Francisco Javier Duque-Aldaz, Raúl Alfredo Sánchez Ancajima.
7. Administración del proyecto: Francisco Javier Duque-Aldaz, Raúl Alfredo Sánchez Ancajima.
8. Recursos: Francisco Javier Duque-Aldaz, Leonor Alejandrina Zapata Aspiazu.
9. Software: Edwin Haymacaña Moreno, Leonor Alejandrina Zapata Aspiazu.
10. Supervisión: Félix Genaro Cabezas García, Raúl Alfredo Sánchez Ancajima.
11. Validación: Félix Genaro Cabezas García.
12. Visualización: Leonor Alejandrina Zapata Aspiazu.
13. Redacción - borrador original: Edwin Haymacaña Moreno, Francisco Javier Duque-Aldaz.
14. Redacción - revisión y edición: Francisco Javier Duque-Aldaz, Félix Genaro Cabezas García, Raúl Alfredo Sánchez Ancajima.

## 7.- Apéndice.

Código en R utilizado para el desarrollo de la investigación.

```
### Paquetes
```

```
install.packages(c("readxl","dplyr","ggplot2","forecast","ts  
eries",
```

```
"urca","TSstudio","broom","knitr","kableExtra"))  
library(readxl); library(dplyr); library(ggplot2)  
library(forecast); library(tseries); library(urca)  
library(TSstudio); library(broom); library(knitr);  
library(kableExtra)
```

```
# 2. Leer Excel (archivo ubicado en el directorio de
trabajo)
df <- readxl::read_excel("matriculadosecuador.xlsx")

# 3. Revisión rápida
str(df)
summary(df)

# 4. Crear objeto de serie de tiempo (anual)
y <- ts(df$Matriculacion, start=min(df$Año), frequency =
1)

# Exploracion Inicial
autoplot(y) +
  labs(title="Matrícula secundaria (% bruto) – Ecuador",
        x="Año", y="%") +
  theme_minimal(base_size = 12)

# ACF y PACF
ggAcf(y, lag.max = 30) + theme_minimal()
ggPacf(y, lag.max = 30) + theme_minimal()

####Diagnóstico de estacionariedad y transformaciones
# Tests de raíz unitaria
tseries::adf.test(y) # H0: raíz unitaria (no estacionaria)
tseries::kpss.test(y) # H0: estacionaria (si p<0.05, no
estacionaria)

# Orden de diferenciación sugerido
forecast::ndiffs(y) # usualmente 1

####Diferenciar una vez (d = 1) y volver a testear
#y_ts: tu serie anual ya creada, frecuencia 1 (1971–2023)
y_ts <- ts(df$Matriculacion, start = min(df$Año), frequency
= 1)

# 1) Diferencia de primer orden
dy <- diff(y_ts)

# 2) Visualizar la serie diferenciada
autoplot(dy) +
  labs(title = "Primera diferencia de la matrícula secundaria
(%)",
        x = "Año", y = "Δ Matriculación (%)") +
  theme_minimal(base_size = 12)

# 3) Pruebas de estacionariedad sobre la serie diferenciada
adf.test(dy) # H0: raíz unitaria (no estacionaria)
kpss.test(dy, null = "Level") # H0: estacionaria en nivel

# 4) Estructura temporal de la serie diferenciada
ggAcf(dy, lag.max = 30) + theme_minimal(base_size = 12)
ggPacf(dy, lag.max = 30) + theme_minimal(base_size = 12)
```

```
####Identificación/estimación de modelos (candidatos +
auto.arima)
# Búsqueda exhaustiva (sin estacionalidad)
fit_auto <- auto.arima(y_ts,
  seasonal = FALSE, # anual
  stepwise = FALSE, # búsqueda más
  completa
  approximation = FALSE,
  d = 1) # ya sabemos que d=1
fit_auto

#Luego, probamos algunos candidatos clásicos y
comparamos por AIC, AICc, BIC:
cand <- list(
  ARIMA_011 = Arima(y_ts, order = c(0,1,1)),
  ARIMA_110 = Arima(y_ts, order = c(1,1,0)),
  ARIMA_111 = Arima(y_ts, order = c(1,1,1)),
  ARIMA_210 = Arima(y_ts, order = c(2,1,0)),
  ARIMA_012 = Arima(y_ts, order = c(0,1,2))
)

cmp <- data.frame(
  Modelo = names(cand),
  AIC = sapply(cand, AIC),
  BIC = sapply(cand, BIC)
)
print(cmp)

####Diagnóstico de residuos del modelo elegido
# Diagnóstico integral
checkresiduals(fit_auto) # incluye Ljung–Box, ACF de
residuales y QQ-plot
# Si quieres Ljung–Box explícito con varios rezagos:
Box.test(residuals(fit_auto), lag = 10, type = "Ljung")
Box.test(residuals(fit_auto), lag = 15, type = "Ljung")

####Validación fuera de muestra (train/test) y métricas
# Partición temporal
y_tr <- window(y_ts, end = 2016)
y_te <- window(y_ts, start = 2017)

fit_tr <- auto.arima(y_tr, seasonal = FALSE, stepwise =
FALSE, approximation = FALSE, d = 1)
fc_te <- forecast(fit_tr, h = length(y_te))

# Métricas de validación
accuracy(fc_te, y_te)
autoplot(fc_te) + autolayer(y_te, series = "Observado") +
  labs(title = "Pronóstico fuera de muestra (2017–2023)",
        y = "% matrícula") + theme_minimal(base_size = 12)

####Pronóstico final (h = 5–10 años)
# Reentrena con toda la serie y pronostica
fit_all <- auto.arima(y_ts, seasonal = FALSE, stepwise =
FALSE, approximation = FALSE, d = 1)
```



fc\_10 <- forecast(fit\_all, h = 10)

```
autoplot(fc_10) +
  labs(title = "Pronóstico de matrícula secundaria (%) – Ecuador",
        x = "Año", y = "%") +
  theme_minimal(base_size = 12)
```

## 7.- Referencias.

- [1] C. T. Simonino, L. F. dos Santos, and A. F. de Freitas, "Implementação do Programa Nacional de Alimentação Escolar em contexto pandêmico no município de Viçosa/MG: implicações para a gestão pública e para a agricultura familiar," *Administração Pública e Gestão Social*, vol. 17, no. 1, 2025. URL: <https://www.redalyc.org/journal/3515/351580876003/>
- [2] M. Zanatta Idemori, A. R. Nogueira de Macedo, P. M. Galvão de Souza, and O. S. Gaion, "A RELAÇÃO DA GAMIFICAÇÃO E EXPERIÊNCIA NOS APLICATIVOS DE DELIVERY," *Revista Pensamento Contemporâneo em Administração*, vol. 19, no. 1, pp. 140–157, 2025. DOI: <https://doi.org/10.12712/rpca.v.19.2.65109> Periódicos UFF+1
- [3] D. Escolar, "Conflicto e identificaciones etno-raciales en la antesala colonial del caudillismo y las rebeliones montoneras," *Boletín del Instituto de Historia Argentina y Americana Dr. Emilio Ravignani*, no. 61, pp. 85–91, 2024. DOI: <https://doi.org/10.34096/bol.rav.n61.14555> SciELO Argentina
- [4] F. O. Corrêa Werle and J. A. Lago Fonseca, "Gestão escolar: fundamentação e discussões acerca da gestão democrática," *Ensaio: Avaliação e Políticas Públicas em Educação*, e0254968, 2025. DOI: <https://doi.org/10.1590/S0104-40362025003304968> Educa FCC
- [5] G. E. Castro Rosales, A. D. Torres Alvarado, L. S. Zalamea Cedeño, F. J. Duque-Aldaz, and F. R. Rodríguez-Flores, "Comprehensive Ergonomic Proposal for the Reduction of Musculoskeletal Risks in Soap Production: An Approach Based on Statistical Analysis and Postural Evaluation," *INQUIDE – Ingeniería Química y Desarrollo*, vol. 7, no. 2, 2025. DOI: <https://doi.org/10.53591/iqd.v7i02.2416> Revistas UG
- [6] M. J. Cabrera Valladolid, "Análisis de series temporales de casos confirmados y fallecidos por Covid-19 del Ecuador," Escuela Superior Politécnica de Chimborazo, 2021. URL: <https://dspace.esPOCH.edu.ec/handle/123456789/14812>
- [7] M. G. Zalduaromero, "Estimación de una función de demanda de importaciones en la industria plástica en el Ecuador, período 1980–2016," UESS, 2017. URL: <http://repositorio.uees.edu.ec/123456789/2008>
- [8] J. I. Cañarte Murillo, "Crecimiento restringido de la balanza de pagos: Caso ecuatoriano de la Ley de Thirlwall 2000–2016," UESS, 2017. URL: <http://repositorio.uees.edu.ec/123456789/1980>
- [9] M. Alós and M. Serio, "Participación en el mercado laboral de las personas mayores en Argentina," *Revista Pilquen – Sección Ciencias Sociales*, vol. 27, no. 2, pp. 71–96, 2024. URL: <https://www.redalyc.org/journal/3475/347580385004/>
- [10] F. J. Duque-Aldaz and E. G. Pazan Gómez, "Factors affecting entrepreneurial intention of Senior University Students," *Espacio*, vol. 39, no. 09, p. 32, 2017. URL: <https://www.revistaespacios.com/a18v39n09/18390932.html>
- [11] A. Meneses Freire, L. Zúñiga Lema, J. Muñoz Cargua, J. Lara Sinaluisa, and W. Acurio Martínez, "Modelo predictivo funcional no paramétrico en series temporales funcionales. Aplicación en variables meteorológicas," Escuela Superior Politécnica de Chimborazo, 2022. URL: <https://dspace.esPOCH.edu.ec/handle/123456789/19024>
- [12] L. Y. Ortega Villegas, "Análisis en la industria naviera en Ecuador como resultado de la crisis financiera en el 2008," UESS, 2018. URL: <http://repositorio.uees.edu.ec/123456789/2503>
- [13] D. M. Ichau Tabango, P. J. Pando Sigua, P. L. Cárdenas García, and S. V. Avilés Sacoto, "Gestión de la demanda e inventarios para la creación de un sistema de manejo de información automatizado: Caso de estudio aplicado a un supermercado ecuatoriano," USFQ, 2021. URL: <http://repositorio.usfq.edu.ec/handle/23000/10650>
- [14] F. J. Duque-Aldaz, F. R. Rodríguez-Flores, and J. Carmona Tapia, "Identification of parameters in ordinary differential equation systems using artificial neural networks," *San Gregorio*, vol. 1, no. 2, 2025. URL: <https://revista.sangregorio.edu.ec/index.php/REVISTASANGREGORIO/article/view/2826>
- [15] F. F. Sandoya Sanchez and A. G. Abad Robalino, "Comparación de la precisión de la predicción de datos de una serie de tiempo entre los modelos ARIMA y las Redes Neuronales," ESPOL, 2017. URL: <http://www.dspace.espol.edu.ec/xmlui/handle/123456789/40345>
- [16] A. F. Mayorga Trujillo, "Investigación del incremento de productividad en la fábrica de pernos en la empresa Galo G. Orbea O. Cia. Ltda. mediante el análisis de disponibilidad en las etapas de su proceso productivo," Escuela Superior Politécnica de Chimborazo, 2017. URL: <https://dspace.esPOCH.edu.ec/handle/123456789/6515>
- [17] D. A. Villarreal Godoy, L. F. Córdor Simbaña, and C. A. Suárez Nuñez, "Control de inventario en una empresa ecuatoriana de textiles: Mejora en los procesos de planificación y control de producción," USFQ, 2022. URL: <http://repositorio.usfq.edu.ec/handle/23000/12544>
- [18] N. Vela and M. C. Camacho Cordovez, "Industria Petrolera: Modelamiento del pronóstico de la demanda y la planificación y proyección del consumo," USFQ, 2020. URL: <http://repositorio.usfq.edu.ec/handle/23000/9206>
- [19] J. S. Fu-López, J. P. Fierro Aguilar, F. R. Rodríguez-Flores, and F. J. Duque-Aldaz, "Application of non-automated Lean strategies for quality improvement in manual assembly processes: A case study in the white goods industry," *INQUIDE – Ingeniería Química y Desarrollo*, vol. 7, no. 1, 2025. DOI: <https://doi.org/10.53591/iqd.v7i02.2417>
- [20] J. M. Ausay Carrillo, "Propuesta de un modelo econométrico para el pronóstico de ventas de la fábrica de balanceados EXIBAL-Riobamba," Escuela Superior Politécnica de Chimborazo, 2022. URL: <https://dspace.esPOCH.edu.ec/handle/123456789/19903>
- [21] P. A. Eguiguren Calisto and S. V. Avilés Sacoto, "Modelos de proyección de demanda para productos de alta volatilidad y bajo volumen en ventas dentro de una empresa de alimentos," USFQ, 2019. URL: <http://repositorio.usfq.edu.ec/handle/23000/8320>
- [22] D. S. Navarro Llivisaca, "Influencia del IVA e impuesto a la renta en las recaudaciones tributarias, 2007–2016," UESS, 2017. URL: <http://repositorio.uees.edu.ec/123456789/1773>
- [23] I. J. Figueroa Tigrero, "Presión tributaria del Ecuador, período 2000–2018: Un análisis desde el enfoque de la teoría de la curva de Laffer," UESS, 2019. URL: <http://repositorio.uees.edu.ec/123456789/3090>
- [24] A. M. Freire Engracia, A. J. Santos Castello, and I. D. Rivadeneira, "Análisis de la relación entre la inversión extranjera directa y la recaudación tributaria: Efectos en el sector comercio," ESPOL, 2025. URL: <http://www.dspace.espol.edu.ec/handle/123456789/65933>
- [25] K. N. Lema Remache, "Comparación entre los métodos ARIMA y Tylor Kriging (TK) para la predicción de series temporales de velocidad del viento: Caso de estudio Estación Meteorológica Chimborazo," Escuela Superior Politécnica de Chimborazo, 2024. URL: <https://dspace.esPOCH.edu.ec/handle/123456789/22839>
- [26] G. J. Moroch Choca, L. A. Bucheli Carpio, and F. J. Duque-Aldaz, "Fuel oil fuel dispatch optimization through multivariate regression using local storage indicators," *INQUIDE*, vol. 6, no. 2, 2024. DOI: <https://doi.org/10.53591/iqd.v6i02.477>
- [27] J. E. Pincay Moran, A. F. López Vargas, F. J. Duque-Aldaz, W. Villamagua Castillo, and R. Sánchez Casanova, "Evaluation and Proposal for an Environmental Management System in a Mango Plantation," *INQUIDE*, vol. 7, no. 1, 2025. DOI: <https://doi.org/10.53591/iqd.v7i01.1991>
- [28] R. J. Guerrero Quinde and C. P. Pérez Siguenza, "Diseño de un modelo de abastecimiento de productos lácteos en una empresa retail," ESPOL, 2025. URL: <http://www.dspace.espol.edu.ec/handle/123456789/66133>
- [29] G. Silva and L. C. Di Serio, "Innovación en los 'negocios olvidados'," *Innovation & Management Review*, vol. 18, no. 4, pp. 350–364, 2021. URL: <https://www.redalyc.org/journal/5375/537569183001/>
- [30] F. Medeiros, L. C. Di Serio, and A. Moreira, "Avon Brasil: Optimización de los Procesos Logísticos en una Venta Directa



- Empresa,” *Revista de Administração Contemporânea*, vol. 25, no. 4, e-190212, 2021. DOI: <https://doi.org/10.1590/1982-7849rac2021190212.en>
- [31] L. M. García-Feria, L. Aguilar-Faisal, R. Pastor-Nieto, and J. C. Serio-Silva, “Changes in vegetation at small landscape scales and captivity alter the gut microbiota of black howler monkeys (*Alouatta pigra*: Atelidae),” *Acta Biológica Colombiana*, vol. 28, no. 1, pp. 154–164, 2023. DOI: <https://doi.org/10.15446/abc.v28n1.93450>
- [32] C. A. García Vázquez, A. I. González Santos, and V. Pérez Garrido, “Metodología y algoritmo de validación para identificar modelos de una unidad manejadora de aire,” *Ingeniería Electrónica, Automática y Comunicaciones*, vol. 42, no. 1, 2021. URL: [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S1815-59282021000100105&lang=es](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1815-59282021000100105&lang=es)
- [33] J. E. Mendoza Cota, “COVID-19 y el empleo en México: Impacto inicial y pronósticos de corto plazo,” *Contaduría y Administración*, vol. 65, no. 5, 2020. URL: <https://dialnet.unirioja.es/servlet/articulo?codigo=7731039>
- [34] J. W. Tudela-Mamani and G. A.-M. Grisellx, “Impacto del COVID-19 en la demanda de turismo internacional del Perú: Una aplicación de la metodología Box-Jenkins,” *Revista de Investigaciones Altoandinas*, vol. 24, no. 1, 2022. DOI: <https://doi.org/10.18271/ria.2022.317>
- [35] L. Asán Caballero, J. Rojas Delgado, and G. E. Jiménez Moya, “Algoritmo de predicción de series de tiempo para el pronóstico del tráfico aéreo basado en redes neuronales artificiales,” *Revista Cubana de Ciencias Informáticas*, vol. 16, no. 4, 2023. URL: [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S2227-18992022000400084&lang=es](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2227-18992022000400084&lang=es)