Práctica 9: Modelos de Media Móvil y modelos Auto regresivos

Profesor: Lauro Reyes Cocoletzi

Diego Castro
dcastroe2100@alumno.ipnx.mx

UPIIT: Unidad Profesional Interdisciplinaria en Ingeniería Campus Tlaxcala Instituto Politécnico Nacional, Tlaxcala, Tlaxcala, México 9000

Ingeniera en Inteligencia Artificial

8 de enero del 2024

***Resumen*— Los modelos de media móvil, como la Media Móvil Simple (SMA) y la Media Móvil Exponencial (EMA), son herramientas estadísticas para suavizar fluctuaciones a corto plazo y destacar tendencias en series temporales. La elección del tamaño de la ventana y el factor de suavizado en EMA afectan la sensibilidad del modelo. Los Modelos Autoregresivos (AR) predicen valores futuros basándose en valores pasados, con coeficientes autoregresivos (ϕ) que cuantifican la contribución de esos valores. La estacionariedad y la identificación del orden son características clave. Los modelos ARIMA combinan AR, I (integración) y MA, siendo esenciales en la predicción de series temporales. La construcción del modelo ARIMA implica diferenciación, identificación de componentes, ajuste y evaluación. Es crucial considerar las limitaciones, como la captura de estacionalidad y tendencias, y realizar validación cruzada para evaluar la generalización del modelo.**

***Palabras clave* — SMA, EMA, ARIMA, sigma, integración**

I. MARCO TEORICO

A. MODELOS DE MEDIA MÓVIL

- Definición: Los modelos de media móvil son una técnica estadística utilizada en el análisis de series temporales para suavizar las fluctuaciones a corto plazo y resaltar las tendencias a largo plazo. Estos modelos son útiles para entender y predecir patrones en datos que varían con el tiempo.

- Tipos:
  - Media Móvil Simple (SMA): La SMA es la media aritmética de un conjunto de datos en un período de tiempo específico. Se calcula sumando los valores de los datos en un intervalo de tiempo y dividiendo el resultado por el número de observaciones en ese intervalo.

$$\mathrm{SMA} = \frac{A_1 + A_2 + \cdots + A_n}{n}$$

    donde A es el valor en el tiempo t y n es el número de períodos

  - Media Móvil Exponencial (EMA): La EMA da más peso a los valores más recientes, lo que la hace más sensible a las tendencias recientes. Es especialmente útil cuando se necesita reaccionar rápidamente a cambios en los datos.

$$\mathrm{EMA}(t) = \propto \cdot X(t) + (1 - \propto) \cdot \mathrm{EMA}(t-1)$$

    donde α es el factor de suavizado, usualmente entre 0 y 1. Este determina la importancia relativa de los datos más recientes frente a los antiguos. Un α más alto da más peso a los datos más recientes, haciendo que la EMA sea más sensible a cambios recientes.

- Consideraciones:
  - Elección del Tamaño de la Ventana: En SMA, la elección del tamaño de la ventana afecta la sensibilidad del modelo. Ventanas más grandes suavizan más los datos, pero responden más lentamente a los cambios.
  - Factor de Suavizado en EMA: Ajustar el factor α en EMA afecta la sensibilidad del modelo. Experimentar con diferentes valores puede ser necesario para adaptarse a las características específicas de los datos.

B. MODELOS AUTOREGRESIVOS

- Serie temporal: Una serie temporal es un conjunto de observaciones ordenadas en función del tiempo

- Modelo Autoregresivo (AR):
  - Un modelo autoregresivo es un tipo de modelo estadístico que predice el valor futuro de una variable en función de sus valores pasados.

- La notación AR(p) se utiliza para representar un modelo autoregresivo de orden p, donde "p" es el número de períodos de tiempo anteriores considerados.

- Ecuación AR(p):

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t$$

- Y_t es el valor de la serie temporal en el momento t
- ϕ_0 es la constante.
- $\phi_0, \phi_1, \ldots, \phi_p$ son los coeficientes autoregresivos.
- $Y_{t-1}, Y_{t-2}, Y_{t-p}$ son los valores anteriores de la serie temporal.
- ϵ_t es el término de error en el momento t

- Características y Propiedades:
  - Estacionariedad: Los modelos AR suelen suponer que la serie temporal es estacionaria, lo que significa que las propiedades estadísticas no cambian con el tiempo.
  - Identificación del Modelo (Orden p): Determinar el orden p del modelo AR es crucial. Se pueden utilizar herramientas como la función de autocorrelación (ACF) para identificar la dependencia entre los valores pasados y presentes.
  - Coeficientes ϕ: Los coeficientes autoregresivos (ϕ) cuantifican la contribución de los valores anteriores al valor actual de la serie temporal.

- Estimación de Parámetros:
  - Método de Mínimos Cuadrados: Los parámetros del modelo AR se estiman típicamente utilizando métodos de mínimos cuadrados para minimizar la suma de los cuadrados de los errores.

- Pronóstico y Evaluación:
  - Pronóstico: Después de estimar los parámetros, se pueden hacer pronósticos para valores futuros de la serie temporal.
  - Evaluación del Modelo: Se utilizan métricas como el error cuadrático medio (MSE) o el error absoluto medio (MAE) para evaluar el rendimiento del modelo.

- Limitaciones y Consideraciones:
  - Estacionalidad y Tendencia: Los modelos AR pueden no capturar patrones estacionales o tendencias en los datos, y pueden requerir ajustes adicionales.
  - Validación Cruzada: Es importante realizar validación cruzada para evaluar la capacidad de generalización del modelo.

C. ARIMA (AUTOREGRESSIVE INTEGRATED MOVING AVERAGE):

- Definición: ARIMA es un modelo estadístico utilizado para el análisis y la predicción de series temporales. La metodología ARIMA combina tres componentes principales: autoregresión (AR), integración (I) y media móvil (MA). La idea es ajustar la serie temporal para que sea estacionaria y, luego, modelar la relación entre las observaciones y los retrasos temporales. La notación general para un modelo ARIMA es ARIMA(p, d, q).

- Componente Autoregresiva (AR):
  - En este componente, la predicción se basa en combinaciones lineales de valores pasados de la propia serie temporal.
  - Se denota como AR(p), donde "p" es el orden autoregresivo, es decir, la cantidad de valores pasados considerados.

- Componente de Integración (I):
  - La integración implica tomar diferencias entre observaciones sucesivas para hacer que la serie temporal sea estacionaria.
  - Se denota como $I(d)$, donde "d" es el orden de integración, es decir, el número de diferencias necesarias para lograr la estacionariedad.

- Componente de Media Móvil (MA):
  - En este componente, la predicción se basa en combinaciones lineales de términos de error pasados.

- Pasos para construir un modelo ARIMA
  1. Diferenciación: Si la serie temporal no es estacionaria, se aplica diferenciación para lograr estacionariedad ($I(d)$).
  2. Identificación de Componentes AR y MA: Se utilizan gráficos de autocorrelación (ACF) y autocorrelación parcial (PACF) para identificar los órdenes "p" y "q".
  3. Ajuste del Modelo: Se ajusta el modelo ARIMA(p, d, q) utilizando datos históricos.
  4. Evaluación y Diagnóstico: Se evalúa la bondad de ajuste del modelo y se realizan diagnósticos para verificar la validez de las suposiciones del modelo.

- Consideración:
  - ARIMA asume que los patrones en los datos son lineales y que la serie temporal es estacionaria después de aplicar las diferencias necesarias. Si hay patrones no lineales o cambios estructurales en los datos, se pueden considerar modelos más avanzados.

II. DESARROLLO

A. MODELO MEDIA MÓVIL

1. Generación de serie temporal con ruido blanco

Antes de aplicar el modelo de media móvil, debemos tener la serie temporal. El ruido blanco es una secuencia de variables aleatorias independientes e idénticamente distribuidas con media cero y varianza constante.

La adición de ruido blanco a la serie temporal tiene varios propósitos:

- *Realismo*: En muchos casos del mundo real, los datos observados contienen cierto grado de variabilidad aleatoria o ruido. Al agregar ruido blanco, se está introduciendo esta variabilidad para hacer que la serie temporal sea más realista y representativa de situaciones prácticas.
- *Complejidad*: La introducción de ruido blanco puede aumentar la complejidad de la serie temporal, haciendo que sea más desafiante modelar y predecir.

2. Series MA(1)

Generar dos series MA(1) (modelo MA de orden q=1)

$$y(t) = w(t) + \theta_1 w_{t-1}$$

Donde considere $\theta_1 = 0{,}6$ y posteriormente $\theta_1 = -0{,}6$

B. COMPONENTE AUTORREGRESIVO AR(1)

1. Series AR(1)

Generar dos series AR (1)

$$y(t) = \phi_1 y_{t-1} + w_t$$

Donde considere $\theta_1 = 0{,}6$ y posteriormente $\theta_1 = -0{,}6$
La diferencia con el código anterior es que estamos modelando la autocorrelación en función de los valores previos de la propia serie temporal (AR) en lugar de solo la media móvil.

C. COMPONENTE AUTORREGRESIVO AR(2)

Para un modelo AR(2), se puede escribir como

$$y(t) = \phi_1 y_{t-1} + \phi_2 y_{t-2} + w_t$$

Generemos cuatro series AR(2) diferentes para revisar sus estructuras auto correlacionales:

- Modelo AR(2) I: $\phi 1 = 0{,}5$ y $\phi 2 = 0{,}3$
- Modelo AR(2) II: $\phi 1 = -0{,}5\ y\ \phi 2 = 0{,}3$
- Modelo AR(2) III: $\phi 1 = 1\ y\ \phi 2 = -0{,}5$
- Modelo AR(2) IV: $\phi 1 = -0{,}5\ y\ \phi 2 = -0{,}3$

D. AR(1)

Siguiendo con otro ejemplo, uno más básico es: teniendo las siguientes autocorrelaciones:

- $p_0 = 1$
- $p_1 = \phi$
- $p_2 = \phi^2$
- ...
- $p_s = \phi^s$

Se obtienen a partir de:

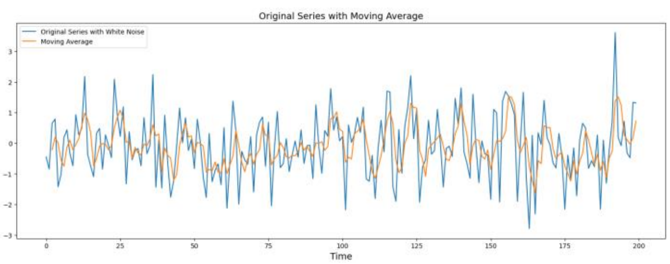
- $p_0 = \frac{Y_0}{Y_0}$
- $p_1 = \frac{Y_1}{Y_0}\ \ Y_1 = \phi Y_0$
- ...

Podemos experimentar con diferentes phis: $0.1, -0.5, 0.7, 0.9$

III. RESULTADOS

A. MODELO MEDIA MÓVIL

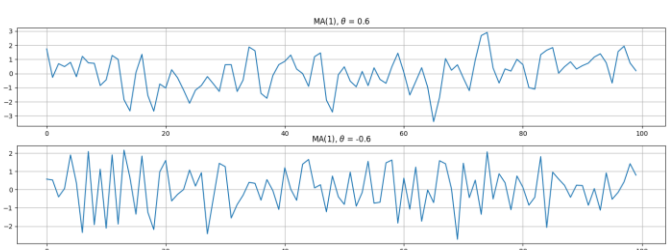
1. Generación de serie temporal con ruido blanco

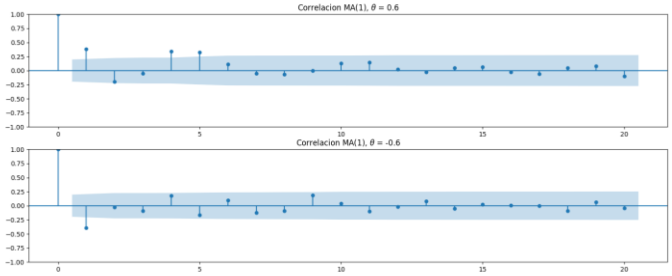


La serie temporal cuenta con las siguientes características:

- Se genera una serie de 200 valores generados aleatoriamente a partir de una distribución normal con media y desviación estándar. Estos valores representan el ruido blanco.
- Se utiliza la función $rolling$ de pandas para aplicar una ventana móvil de tamaño 3 a la serie temporal. Luego, se calcula la media de cada ventana para obtener las medias móviles

2. Series MA(1):

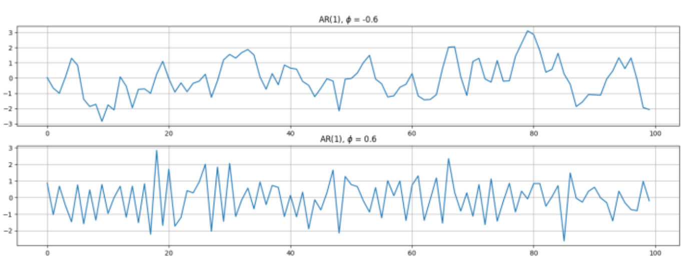


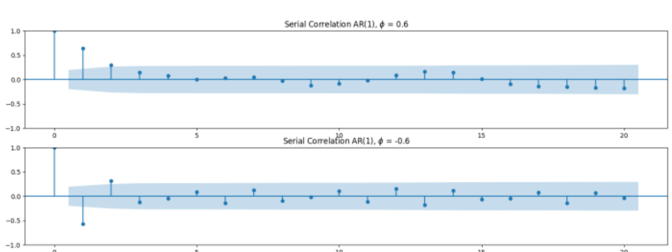


- Dirección de la Autocorrelación:
  - Con $\theta = 0.6$ (positivo): La serie temporal tiene una autocorrelación positiva, lo que significa que las observaciones sucesivas estarán positivamente correlacionadas. Esto implica que valores altos tienden a ser seguidos por valores altos y viceversa.
  - Con $\theta = -0.6$ (negativo): La serie temporal tiene una autocorrelación negativa, indicando que las observaciones sucesivas estarán negativamente correlacionadas. Esto sugiere que valores altos tienden a ser seguidos por valores bajos y viceversa.
- Suavizado y Amplitud de Variación:
  - Con $\theta = 0.6$: La serie temporal tiene una variación más suave y menos volatilidad en comparación con el ruido blanco original. El efecto de suavizado se debe a la autocorrelación positiva que ayuda a atenuar las fluctuaciones.
  - Con $\theta = -0.6$: La serie temporal tiene una variación más abrupta y mayor volatilidad en comparación con el ruido blanco original. La autocorrelación negativa puede amplificar las fluctuaciones y generar cambios más rápidos.

B. COMPONENTE AUTORREGRESIVO (AR)

1. Series AR(1)



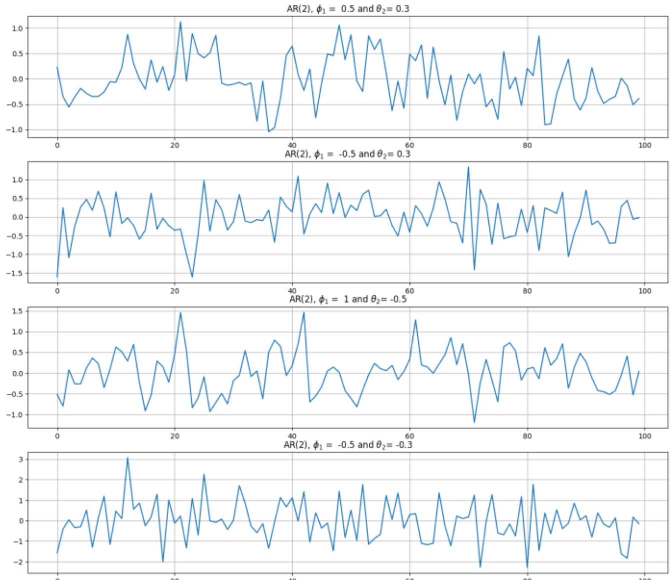


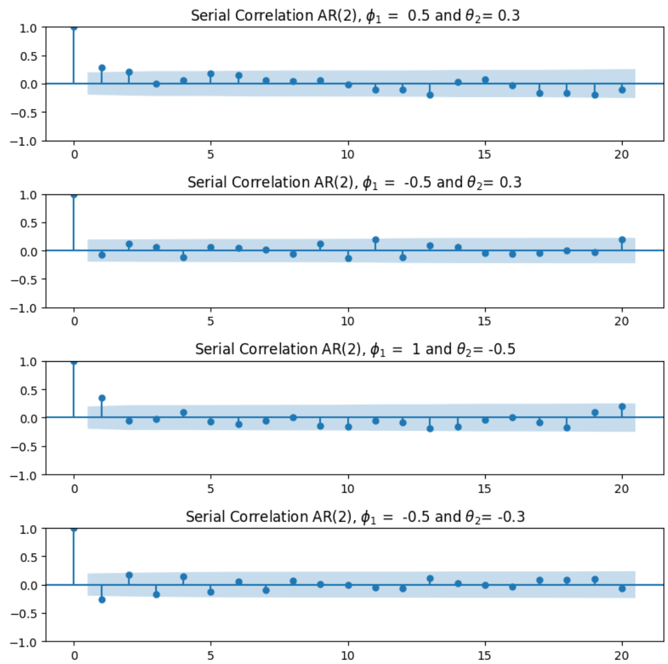
- Amplitud de la Autocorrelación:
  - El valor absoluto de ϕ afecta la fuerza de la autocorrelación. Los valores absolutos cercanos a 1 (ya sea positivo o negativo) indica que tiene una autocorrelación más fuerte y persistente.
- Efecto en la Forma de la Serie Temporal:
  - Con $\phi = 0.6$: La serie temporal podría muestra patrones de aumento o disminución suave, ya que los valores tienden a seguir la dirección de los valores anteriores.
  - Con $\phi = -0.6$: La serie temporal muestra oscilaciones y cambios rápidos, ya que los valores tienden a cambiar de dirección después de cada observación.
- Impacto en la Predicción:
  - Un ϕ positivo implica que los modelos AR(1) tenderán a prever valores que sigan la dirección de la tendencia actual.
  - Un ϕ negativo indica que los modelos AR(1) tenderán a prever valores que se mueven en la dirección opuesta a la tendencia actual.

Al modelar la autocorrelación en función de los valores previos de la propia serie temporal (AR) en lugar de solo la media móvil, permite obtener resultados más claros e interpretables que los modelos basados solo en la media móvil (MA).

- Tendencias y Patrones Temporales: Los modelos AR son efectivos para capturar tendencias y patrones temporales en los datos. Si hay una relación significativa entre los valores pasados y futuros, un modelo AR puede capturar estas dependencias y proporcionar una representación más fiel del comportamiento temporal de la serie.
- Predicción de Tendencias a Largo Plazo: Los modelos AR permiten prever tendencias a largo plazo al tener en cuenta los valores pasados. Si la serie temporal muestra una tendencia creciente o decreciente, un modelo AR puede capturar y proyectar esa tendencia.
- Menor Sensibilidad al Ruido Aleatorio: Los modelos AR tienden a ser menos sensibles al ruido aleatorio en comparación con los modelos MA. Esto se debe a que los modelos AR utilizan una combinación lineal de valores pasados, lo que puede ayudar a suavizar el impacto del ruido en la serie temporal.

C. Componente Autorregresivo AR(2)





Análisis de las Gráficas de Series Temporales:

1. $\phi = 0.5, \theta = 0.3$:
   - La serie temporal parece tener una tendencia positiva y una variación que persiste. Hay una correlación positiva evidente entre los valores sucesivos.

2. $\phi = -0.5, \theta = 0.3$:
   - La serie temporal parece tener oscilaciones y una tendencia negativa. Hay una correlación negativa entre los valores sucesivos.

3. $\phi = 1, \theta = -0.5$:
   - La serie temporal debería mostrar un crecimiento exponencial, la autocorrelación es positiva y fuerte. Sin embargo, no se logra apreciar

4. $\phi = -0.5, \theta = -0.3$:
   - La serie temporal parece tener oscilaciones y una tendencia negativa. La correlación negativa es menos pronunciada que en el segundo caso.

Análisis de las Funciones de Autocorrelación (ACF):

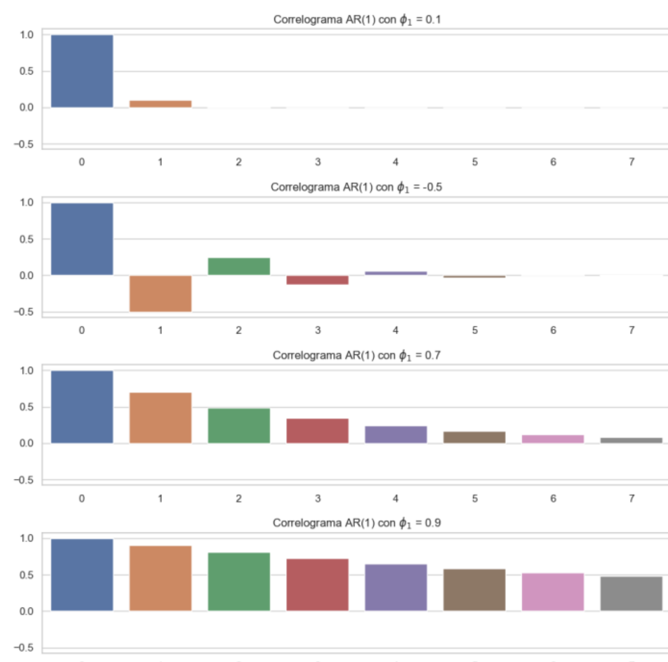
1. $\phi = 0.5, \theta = 0.3$:
   La ACF muestra una correlación positiva que decae gradualmente, indicando una dependencia positiva

2. $\phi = -0.5, \theta = 0.3$:
   - La ACF muestra una correlación negativa que decae rápidamente, indicando una dependencia negativa fuerte

3. $\phi = 1, \theta = -0.5$:
   - La ACF muestra una correlación positiva que decae lentamente, indicando una dependencia positiva persistente.

4. $\phi = -0.5, \theta = -0.3$:
   - La ACF muestra una correlación negativa que decae gradualmente, indicando una dependencia negativa persistente.

D. AR(1)



1. $\phi = 0.1$: En este caso, ϕ es positivo, lo que indica que hay una tendencia positiva en la serie temporal. La autocorrelación disminuye exponencialmente a

medida que aumenta el rezago, lo que sugiere una dependencia positiva de los valores pasados.

2. $\phi = -0.5$: Con ϕ negativo, la serie temporal tiende a oscilar y tiene una autocorrelación negativa, especialmente en los lags cercanos. Esto indica que los valores sucesivos están inversamente relacionados, lo que podría sugerir un patrón de reversión a la media.
3. $\phi = 0.7$: Aquí, ϕ es positivo y relativamente alto. La serie temporal muestra una autocorrelación positiva que decae más lentamente, lo que sugiere una fuerte dependencia positiva de los valores pasados. Esto podría indicar una tendencia a seguir una dirección específica.
4. $\phi = 0.9$: Un valor alto de ϕ positivo indica una fuerte dependencia positiva de los valores pasados. La serie temporal parece tener una tendencia clara y persistente en una dirección positiva.

IV. CONCLUSIONES

En los ejercicios realizados, exploramos la generación y análisis de series temporales utilizando modelos autoregresivos (AR) y de media móvil (MA). A través de la manipulación de datos sintéticos y la visualización de correlogramas, obtuvimos insights valiosos sobre el comportamiento temporal de estas series.

Los modelos AR y MA son herramientas fundamentales en la modelización de datos temporales, y su comprensión es crucial para diversos campos, incluyendo finanzas, economía, climatología, entre otros. Estos modelos permiten capturar patrones y estructuras subyacentes en datos secuenciales, proporcionando una base sólida para la predicción y la toma de decisiones.

En particular, los modelos AR describen cómo cada punto en la serie temporal está relacionado con sus valores pasados, lo que ayuda a identificar tendencias y patrones de comportamiento a lo largo del tiempo. Por otro lado, los modelos MA capturan las variaciones en la serie temporal mediante la incorporación de términos de ruido blanco, lo que les confiere la capacidad de modelar eventos aleatorios o impredecibles.

La visualización de correlogramas nos brindó una herramienta poderosa para comprender la autocorrelación en los datos y evaluar la influencia de diferentes parámetros, como el coeficiente ϕ en los modelos AR. La interpretación de estos gráficos es esencial para ajustar adecuadamente los modelos a los datos reales y entender cómo las observaciones pasadas afectan las futuras.

V. BIBLIOGRAFIA

[1] Modelo autorregresivo integrado de media móvil. (2023, 11 de noviembre). Wikipedia, La enciclopedia libre. Fecha de consulta: 21:39, noviembre 11, 2023 desde https://es.wikipedia.org/w/index.php?title=Modelo_autorregresivo_integrado_de_media_m%C3%B3vil&oldid=155258410.

[2] Modelo autorregresivo. (2023, 7 de octubre). Wikipedia, La enciclopedia libre. Fecha de consulta: 09:12, octubre 7, 2023 desde https://es.wikipedia.org/w/index.php?title=Modelo_autorregresivo&oldid=154375680.

[3] Modelo de medias móviles. (2020, 14 de marzo). Wikipedia, La enciclopedia libre. Fecha de consulta: 23:38, marzo 14, 2020 desde https://es.wikipedia.org/w/index.php?title=Modelo_de_medias_m%C3%B3viles&oldid=124262669.

[4] Andrés Navarro, B. (2020). Técnica de Previsión de Medias Móviles. https://riunet.upv.es/handle/10251/145255

[5] Villavicencio, J. (2010). Introducción a series de tiempo. Puerto Rico. https://www.academia.edu/download/38458362/manual_intro_series_tiempo.pdf