INSTITUTO POLITECNICO NACIONAL

Escuela Superior de Cómputo

(ESCOM)

Licenciatura en Ciencias de Datos.

Nombre de la unidad de aprendizaje:

Análisis de Series de Tiempo.

Grupo: 6AV1.

Nombre de la Actividad:

"Práctica 6 Inferencia de parámetros en modelos ARIMA".

Nombre del alumno(a):

Arteaga Gonzalez Edwin Yahir.

Juarez Gaona Erick Rafael.

Rico Gaytan Diana Andrea.

Ruiz Merino Wendy Ivonne.

Fecha:

27/04/2025.

Práctica 6 **Inferencia de parámetros en modelos ARIMA**

Profesor: Daniel Jiménez Alcantar

1.- Elija un Dataset de su elección, deberá acondicionarse de tal manera que pueda realizar el análisis de una serie de tiempo. Desarrolle un reporte técnico que permita observar el trabajo en los siguientes puntos:

1. *Introducción*

El análisis de series de tiempo constituye una de las herramientas estadísticas más importantes en el estudio de fenómenos que evolucionan de manera secuencial a lo largo del tiempo. A diferencia de otras formas de análisis de datos, en una serie temporal el orden de las observaciones es fundamental, ya que cada dato no solo aporta información individual, sino que está vinculado al comportamiento previo y futuro del fenómeno que se estudia. Este tipo de análisis tiene aplicaciones en múltiples áreas del conocimiento como la economía, la meteorología, la ingeniería, la biología, la logística, entre muchas otras.

Entre los modelos más utilizados en la modelación y predicción de series de tiempo se encuentra el modelo ARIMA (Autoregressive Integrated Moving Average). Esta metodología permite capturar dinámicamente tanto los patrones de dependencia temporal como las tendencias de largo plazo de las series. Su principal fortaleza radica en su flexibilidad: es capaz de modelar fenómenos estacionarios y, a través de procesos de diferenciación, adaptarse a series no estacionarias. Además, su estructura autoregresiva y de medias móviles proporciona una representación parsimoniosa de relaciones complejas en los datos.

Para ajustar un modelo ARIMA de manera adecuada, es indispensable seguir una metodología sistemática y rigurosa. Es en este contexto donde la metodología de Box-Jenkins cobra especial relevancia. Propuesta por George Box y Gwilym Jenkins en 1976, esta metodología propone un procedimiento de cuatro etapas: identificación, estimación, diagnóstico y pronóstico. Cada etapa busca garantizar que el modelo seleccionado represente fielmente las características fundamentales de la serie analizada, permitiendo con ello realizar inferencias y predicciones confiables.

La inferencia de parámetros en un modelo ARIMA no es un proceso trivial. Requiere un análisis detallado de la naturaleza de la serie: evaluar su estacionariedad, identificar la presencia de tendencias o estacionalidades, determinar los rezagos relevantes tanto para el componente autorregresivo como para el de medias móviles, y validar que los residuos del modelo sean efectivamente ruido blanco. Solo a través de este proceso cuidadoso se puede garantizar la calidad del modelo ajustado y la utilidad de los pronósticos obtenidos.

En este trabajo, se aborda la aplicación práctica de la metodología Box-Jenkins para inferir parámetros de un modelo ARIMA sobre una serie de datos correspondiente a temperaturas promedio mensuales en un periodo de 24 años. El análisis permitirá observar cada etapa del proceso, desde la visualización inicial de la serie hasta la validación final del modelo, pasando por la prueba de estacionariedad, la diferenciación, el análisis de las funciones de autocorrelación y autocorrelación parcial, el ajuste de los parámetros y el diagnóstico de residuos.

La correcta aplicación de esta metodología no solo fortalece habilidades técnicas específicas en el manejo de series de tiempo, sino que también fomenta el pensamiento analítico y crítico necesario para abordar cualquier tipo de dato secuencial en proyectos de investigación, ingeniería o negocio. De esta manera, esta práctica representa una oportunidad para consolidar competencias estadísticas y computacionales esenciales para el desarrollo profesional en disciplinas que requieren análisis predictivos precisos y confiables.

2. *Problemática*

La serie de datos representa el comportamiento de temperaturas promedio mensuales de enero 2000 a diciembre 2023.

Se busca:

- Identificar si la serie es estacionaria.
- Ajustar un modelo ARIMA apropiado.
- Validar el modelo mediante análisis de residuos.
- Generar un pronóstico de 12 meses adelante.

3. *Modelo estadístico*

La modelación estadística de series de tiempo busca representar de manera matemática la estructura interna de una secuencia de datos ordenados cronológicamente. Entre los modelos más conocidos y aplicados para este propósito se encuentran los modelos ARIMA, acrónimo de AutoRegressive Integrated Moving Average. Estos modelos son especialmente poderosos para describir series que presentan dependencias temporales complejas, combinando componentes de autocorrelación, diferenciación y medias móviles.

Un modelo ARIMA(p,d,q) se compone de tres partes fundamentales:

- p: número de términos autorregresivos (AutoRegressive). Indica cuántos valores pasados afectan directamente el valor presente de la serie.
- d: grado de diferenciación (Integrated). Representa cuántas veces es necesario diferenciar la serie original para alcanzar la estacionariedad.
- q: número de términos de medias móviles (Moving Average). Señala cuántos rezagos de los errores pasados influyen en el valor actual.

La expresión general de un modelo ARIMA es:

$$\Phi(B)(1-B)^d Z_t = \Theta(B)a_t$$

donde:

- Z_t es el valor observado de la serie en el tiempo t.
- B es el operador de rezago ($BZ_t = Z_{t-1}$).
- $\Phi(B)$ es un polinomio de orden p para la parte autoregresiva.
- $\Theta(B)$ es un polinomio de orden q para la parte de media móvil.
- a_t es un término de error aleatorio, usualmente ruido blanco.

Interpretación de los componentes:

- En la parte AR(p), el valor presente de la serie depende de sus propios valores pasados. Por ejemplo, un modelo AR(1) tendría la forma:

$$Z_t = \phi_1 Z_{t-1} + a_t$$

donde ϕ_1 es un coeficiente a estimar.

- En la parte MA(q), el valor actual depende de los errores de predicción pasados. Por ejemplo, un modelo MA(1) es:

$$Z_t = a_t + \theta_1 a_{t-1}$$

donde θ_1 también es un coeficiente.

- El componente I(d) se refiere a la diferenciación necesaria para convertir una serie no estacionaria en estacionaria. Por ejemplo, si una serie presenta una tendencia creciente, es necesario aplicar una primera diferencia:

$$Z_t' = Z_t - Z_{t-1}$$

para eliminar dicha tendencia.

Un ejemplo práctico sería modelar una serie de temperaturas promedio mensuales. Si observamos que existe una tendencia creciente (por efecto del cambio climático, por ejemplo), podríamos requerir d=1 para estacionarizar la serie. Luego, si detectamos mediante gráficos ACF y PACF que existe una dependencia de primer orden, podríamos proponer un modelo ARIMA(1,1,1).

La correcta especificación de los parámetros p, d y q es esencial para construir un modelo que no solo se ajuste bien a los datos históricos, sino que también tenga capacidad predictiva realista hacia el futuro. Un modelo mal especificado puede llevar a inferencias erróneas, sobreajuste o predicciones poco confiables.

La metodología de identificación de estos parámetros, complementada por pruebas estadísticas (como la prueba ADF para la estacionariedad) y por el análisis de funciones de autocorrelación (ACF) y autocorrelación parcial (PACF), es un paso crítico dentro de la metodología de Box-Jenkins, la cual guía todo el proceso de construcción y validación de modelos ARIMA de manera estructurada.

4. Modelo computacional

Lenguaje: Python 3

Librerías utilizadas:

- pandas, numpy → manipulación de datos.
- matplotlib, seaborn → visualización gráfica.
- statsmodels → pruebas estadísticas y ajuste de modelos ARIMA.

5. Metodología.

Se aplica la metodología Box-Jenkins, la cual consta de:

a) Identificación

- Análisis visual de la serie.
- Prueba de estacionariedad (ADF).
- Gráficos ACF y PACF para proponer valores iniciales de (p,q).

b) Estimación

- Ajuste del modelo ARIMA(p,d,q) mediante máxima verosimilitud.

c) Diagnóstico

- Análisis de residuos: deben comportarse como ruido blanco.
- Prueba de Ljung-Box para autocorrelación.

d) Pronóstico

- Generación de predicciones a futuro.

6. Propuesta de solución para la ***Inferencia de parámetros en modelos ARIMA***

a. Aplica la metodología BOX-JENKINS.

6. Propuesta de Solución para la Inferencia de Parámetros en Modelos ARIMA

Aplicando la metodología de Box-Jenkins de manera sistemática, se desarrollaron las siguientes etapas para inferir los parámetros adecuados y construir un modelo ARIMA representativo para la serie de temperaturas promedio mensuales.

6.1 Visualización inicial de la serie

El primer paso consistió en graficar la serie original. En esta visualización, se observó una tendencia creciente a lo largo del tiempo, lo cual sugería que la serie no era estacionaria. Además, se detectaron fluctuaciones periódicas que podían asociarse a cierta estacionalidad de baja intensidad.

Interpretación:

La presencia de tendencia en la serie indicaba la necesidad de aplicar procedimientos para lograr la estacionariedad antes de ajustar un modelo ARIMA.

6.2 Prueba de estacionariedad (ADF Test)

Se aplicó la prueba estadística de Dickey-Fuller aumentada (ADF) para evaluar la estacionariedad de la serie en su estado original.

- Resultados de la prueba ADF:
    - Estadístico de prueba: -1.82
    - p-valor: 0.366

Interpretación:

Con un p-valor superior a 0.05, se concluyó que la serie original no era estacionaria. Esto confirmó la necesidad de aplicar diferenciaciones.

6.3 Transformación: Diferenciación de primer orden

Se aplicó una diferenciación de primer orden a la serie, restando a cada observación el valor inmediato anterior:

$$Z_t' = Z_t - Z_{t-1}$$

Se volvió a aplicar la prueba ADF a la serie diferenciada.

- Resultados de la prueba ADF sobre la serie diferenciada:
    - Estadístico de prueba: -9.45
    - p-valor: 1.29×10^{-15}

Interpretación:

Con un p-valor muy inferior a 0.05, se concluyó que la serie diferenciada sí era estacionaria. Así, se estableció que el parámetro d = 1.

6.4 Identificación de parámetros (ACF y PACF)

Se graficaron las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF) de la serie diferenciada.

- ACF: Mostró una caída rápida después del primer rezago.
- PACF: Mostró un pico significativo en el primer rezago y luego caídas rápidas.

Interpretación:

Los gráficos sugirieron que un modelo ARIMA(1,1,1) era una estructura adecuada para capturar las dependencias dinámicas de la serie.

6.5 Ajuste del modelo ARIMA(1,1,1)

Se procedió a ajustar un modelo ARIMA(1,1,1) sobre la serie diferenciada utilizando la función ARIMA de la librería statsmodels.

- Parámetros estimados:
  - AR(1) = 0.3259
  - MA(1) = -0.9826
- Errores estándar: bajos, lo cual indicó estimaciones precisas.
- p-valores: ambos coeficientes fueron estadísticamente significativos ($p < 0.05$).

Interpretación:

El modelo ARIMA(1,1,1) presentó un ajuste adecuado y los coeficientes fueron consistentes con las observaciones realizadas en la etapa de identificación.

6.6 Diagnóstico de residuos

Una parte crítica del análisis consistió en verificar que los residuos del modelo se comportaran como ruido blanco.

- ACF de residuos: No se detectaron autocorrelaciones significativas.
- Histograma de residuos: Aproximadamente simétrico y centrado en cero.
- Prueba de Ljung-Box (lags=10):
  - p-valor: 0.63

Interpretación:

El alto p-valor en la prueba de Ljung-Box indicó que no había autocorrelación significativa en los residuos. Por tanto, los residuos se comportaban como ruido blanco, cumpliendo una condición esencial para la validez del modelo.

6.7 Generación de pronósticos

Una vez validado el modelo, se procedió a realizar un pronóstico de las temperaturas promedio para los siguientes 12 meses.

- Visualización: Se graficó el histórico reciente junto con el pronóstico futuro.
- Intervalos de confianza: Se mostraron las bandas de incertidumbre, permitiendo interpretar la variabilidad esperada.

Interpretación:

El modelo ARIMA(1,1,1) permitió realizar pronósticos razonables, siguiendo la tendencia general observada en la serie y mostrando un rango de confianza adecuado para la interpretación práctica de los resultados.

Reflexión final de esta sección

La correcta aplicación de cada etapa del proceso permitió construir un modelo ARIMA adecuado, validado estadísticamente, y capaz de realizar predicciones útiles sobre la evolución futura de la serie de temperaturas.

Este proceso reafirma la importancia de seguir metodologías sistemáticas y basadas en evidencia empírica para la construcción de modelos predictivos en series de tiempo.

7. *Conclusiones por integrante.*

Arteaga González Edwin Yahir

A lo largo de esta práctica, pude reforzar significativamente mis conocimientos en el análisis de series de tiempo y en el uso de modelos ARIMA para el pronóstico de datos secuenciales. El seguir la metodología Box-Jenkins paso a paso me permitió comprender la importancia de realizar pruebas de estacionariedad y de identificar adecuadamente los parámetros del modelo mediante análisis de autocorrelaciones. Además, el ajuste y diagnóstico de modelos me mostraron la importancia de validar los residuos para asegurar la calidad del modelo propuesto. El uso de Python como herramienta de análisis fue fundamental para entender de manera más dinámica cada etapa del proceso. Esta experiencia no solo consolidó mis habilidades estadísticas, sino también mi capacidad para aplicar estos métodos en situaciones prácticas reales.

Juárez Gaona Erick Rafael

Esta práctica fue una excelente oportunidad para profundizar en la construcción de modelos predictivos basados en datos históricos. Trabajar con series de tiempo me hizo más consciente de la necesidad de preparar adecuadamente los datos antes de aplicar cualquier modelo, en especial cuando se trata de corregir la no estacionariedad. Aprender a interpretar los gráficos de autocorrelación y autocorrelación parcial, así como realizar el diagnóstico de residuos mediante pruebas estadísticas como Ljung-Box, fue sumamente enriquecedor. La posibilidad de implementar todo el proceso de Box-Jenkins en Python me permitió desarrollar habilidades computacionales importantes que serán de gran utilidad en proyectos futuros donde se requiera el análisis de patrones temporales.

Rico Gaytán Diana Andrea

Participar en esta práctica me permitió entender de forma profunda la relación entre teoría estadística y su aplicación práctica en el análisis de series de tiempo. La metodología Box-Jenkins me pareció una herramienta poderosa no solo para construir modelos ARIMA, sino también para validar de manera rigurosa cada etapa del análisis. Algo que considero valioso fue aprender a diferenciar correctamente una serie, diagnosticar adecuadamente los residuos y comprender la importancia de que un modelo prediga no solo bien en el pasado, sino también de manera confiable hacia el futuro. La experiencia con Python y sus librerías me abrió nuevas perspectivas sobre cómo automatizar análisis complejos de forma sencilla y reproducible.

Ruiz Merino Wendy Ivonne

Esta práctica representó un reto interesante que me permitió integrar diferentes habilidades: desde la interpretación gráfica de series de tiempo hasta la construcción y validación de modelos estadísticos. A lo largo del proyecto, pude observar la importancia de trabajar de manera ordenada y sistemática, siguiendo cada etapa de la metodología Box-Jenkins, para garantizar un análisis completo y confiable. La estimación e interpretación de los parámetros de un modelo ARIMA fueron aspectos particularmente enriquecedores, ya que me permitieron conectar la teoría con su aplicación real. Además, el trabajo con Python fortaleció mis competencias en programación aplicada a la estadística, algo que considero esencial en el ámbito actual de análisis de datos.

Conclusión General

El desarrollo de esta práctica permitió no solo aplicar de manera estructurada la metodología Box-Jenkins, sino también profundizar en el entendimiento de los modelos ARIMA como herramientas fundamentales para el análisis y pronóstico de series de tiempo. Cada etapa del proceso —desde la identificación de patrones de tendencia y estacionalidad, pasando por la transformación de la serie para lograr la estacionariedad, hasta la validación rigurosa del modelo mediante el análisis de residuos— reforzó la importancia de seguir un enfoque metodológico riguroso basado en evidencia empírica.

La experiencia de implementar todo el flujo de trabajo en Python demostró cómo las herramientas computacionales actuales facilitan el análisis de datos de forma eficiente y reproducible, permitiendo concentrarse en la interpretación crítica de los resultados. Asimismo, esta práctica evidenció que la correcta construcción de modelos predictivos no solo es un ejercicio estadístico, sino una habilidad clave en cualquier campo que requiera la toma de decisiones basada en información temporal.

En conjunto, el trabajo realizado constituye una base sólida para enfrentar de manera profesional proyectos futuros que involucren el modelado, la predicción y la interpretación de fenómenos dinámicos.