

# Tarea 1 - Series de Tiempo EYP3907

Diego Aravena - Alonso Campos

13 de Octubre de 2023

## Introducción

En la actualidad, la preocupación por el impacto ambiental y la calidad del aire es un tema de creciente interés a nivel nacional. La exposición al monóxido de carbono (CO), un gas incoloro e inodoro producido principalmente por la combustión de combustibles fósiles, puede tener graves consecuencias para el medio ambiente y la salud de las personas. Con el objetivo de comprender mejor la dinámica de este gas tóxico en la atmósfera, se ha recopilado mediciones de éste a través del tiempo, así como variables climáticas relevantes.

Este breve artículo tiene como objetivo presentar un enfoque analítico que utiliza modelos de regresión para investigar la relación entre los niveles de monóxido de carbono y las variables climáticas. Esto permitirá identificar patrones, tendencias y relaciones significativas que pueden ayudar a comprender mejor los factores que influyen en la concentración del contaminante en la atmósfera. Luego, se pretende ajustar una serie de tiempo, lo que nos permitirá evaluar si existen patrones temporales y estacionales en la concentración de monóxido de carbono que no han logrado ser explicados en el modelo de regresión.

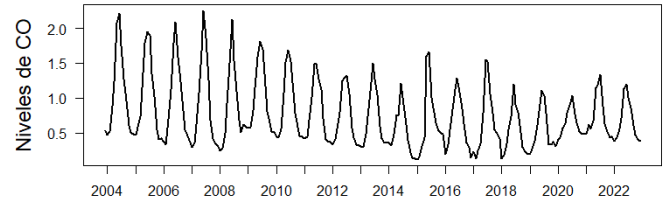
## Metodología

### Datos

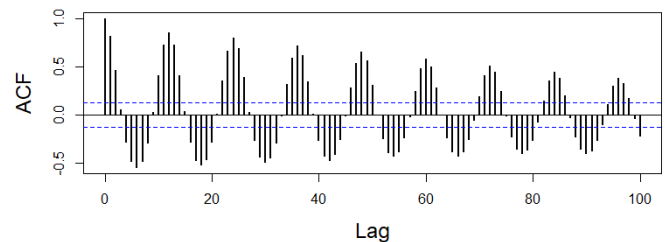
En este estudio se utilizó como referencia la estación de monitoreo de la comuna de La Florida de la Región Metropolitana. Los datos se encuentran disponibles en el sitio web del Sistema de Información Nacional de Calidad del Aire (SINCA) <sup>1</sup>. De este sitio fue posible obtener los promedios mensuales del contaminante Monóxido de Carbono (CO) y las variables climáticas humedad, velocidad del viento y temperatura.

### Primeras impresiones

En los datos originales (véase Figura 1 (a)) es posible notar que no se cumplen los supuestos de media y varianza constante, pues a lo largo del tiempo disminuye la tendencia y variabilidad de los niveles de monóxido de carbono. Asimismo, se logra apreciar estacionalidad en los datos, lo cual explica aquellos patrones en el gráfico de autocorrelación mostrados en la Figura 1 (b).



(a) Serie temporal de los niveles de monóxido de carbono



(b) Gráfico de autocorrelación

Figura 1: Comportamiento de los niveles de CO entre diciembre del año 2003 y diciembre de 2022

## Modelo de Regresión

Con el objetivo de implementar un modelo adecuado que logre capturar tanto la tendencia como la variabilidad en los datos, se evaluaron variables climáticas y variables relacionadas con el tiempo. Observando una tendencia cuadrática en las mediciones de CO a lo largo de los años, se tomó la decisión de incorporar esta covariable en el modelo. Además, se incluyó la variable relacionada con el mes para explicar el patrón estacional. Posteriormente, se descubrió que algunos meses eran estadísticamente significativos, en particular, los meses entre Marzo y Agosto, que coinciden con la temporada de temperaturas más bajas. A partir de esto, se creó una variable categórica, el cual toma valores distintos para meses entre Marzo y Agosto, mientras que el resto se agrupó en una única categoría. El modelo implementado también considera una interacción entre el promedio de la temperatura y el tiempo. Tanto humedad como velocidad del viento no resultaron ser significativos. A continuación, se presenta el modelo ajustado

$$CO^{0.6} = Tiempo^2 + Mes_c + Temperatura * Tiempo$$

La variable respuesta tuvo que ser transformada a la potencia de 0.6 para que cumpliera el supuesto de homocedasticidad, el cual se explica en la siguiente sección.

<sup>1</sup> <https://sinca.mma.gob.cl/>

Finalmente, se obtuvo un valor  $R^2$  de 89 %, además cada coeficiente es significativo al 5 %. Luego, se calculó el error porcentual absoluto medio (MAPE) para evaluar la bondad de ajuste del modelo. En la Figura 2 se muestran los niveles de monóxido de carbono junto con los valores ajustados hasta el año 2022, los cuales logran ajustarse razonablemente a los datos, en efecto, hasta esa fecha se reporta un MAPE de 0.11, lo cual significa que las predicciones están distancias en promedio un 11 % del valor real. Por otro lado, entre Enero y Septiembre del año 2023 se muestran las predicciones. Se observa que estos logran seguir la tendencia apropiadamente, de aquí se logra obtener un MAPE de 0.12.

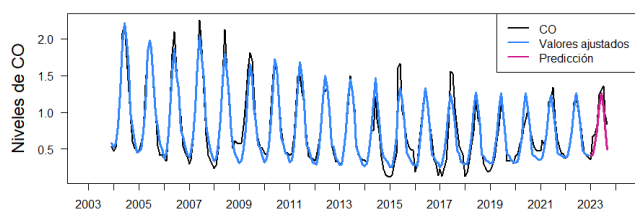


Figura 2: Valores ajustados y predicción del modelo de regresión

## Análisis de Supuestos Residuos Modelo de Regresión

**Media constante** Se implementa un modelo de regresión simple, cuya variable respuesta son los residuos y la covariable es el tiempo. De esta forma se obtiene que el valor-p del test asociado a la significancia del predictor es cercano a 1, por ende, se concluye que los residuos presentan una media constante con respecto al tiempo.

**Homocedasticidad de la varianza** Al ajustar el modelo con los datos sin transformación los residuos no cumplían con el supuesto de homocedasticidad. De esta forma, al implementar el método de Box-Cox, se propone transformar la variable respuesta a una potencia de 0.6.

Los resultados del test de Breusch-Pagan indican que, tras aplicar la transformación, el valor-p asociado es de 0.052. Aunque no se considera un valor alto, en un sentido estricto, no se rechaza la hipótesis de que los residuos son homocedásticos.

**Normalidad** Para comprobar normalidad en los residuos se considera el test de Shapiro-Wilks y el test de Kolmogorov-Smirnov para verificar una normalidad estándar.

Tabla 1: Valores-p para test de Normalidad

	Valor-p
Shapiro-Wilks	0.65
Kolmogorov-Smirnov	0.744

A partir de los resultados presentados en la Tabla 1, se observa que ambos test rechazan el supuesto de normalidad. Por ende, se concluye que los residuos distribuyen normalmente.

**Ruido Blanco** Para verificar este supuesto se utiliza el gráfico de función de autocorrelación (ACF) y el test de Box-Ljung.

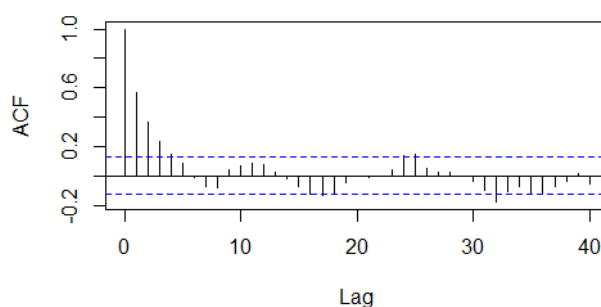


Figura 3: Gráfico de autocorrelación de los residuos del modelo de regresión

La Figura 3 evidencia que la autocorrelación de los residuos se escapan de las bandas de confianza. Del mismo modo, el test de Box-Ljung, que se muestra en la Figura 4, complementa lo expuesto anteriormente, ya que, los valores-p respectivos no logran superar el 5 %, lo cual lleva a concluir que los residuos están correlacionados.

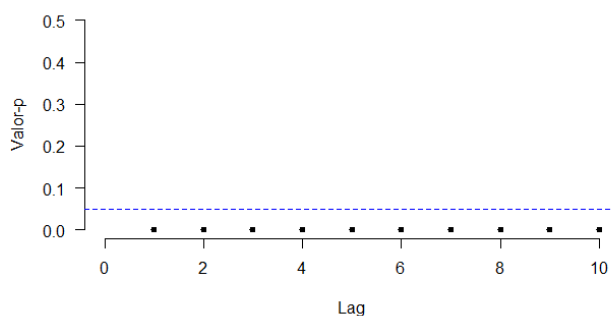


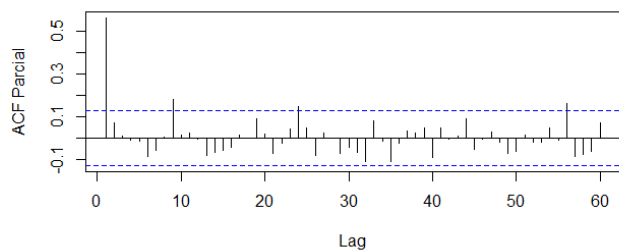
Figura 4: Valores-p para Test de Box-Ljung para los residuos del modelo lineal

## Modelo Series de Tiempo

En esta sección se procederá a modelar la dependencia serial de los residuos del modelo de regresión.

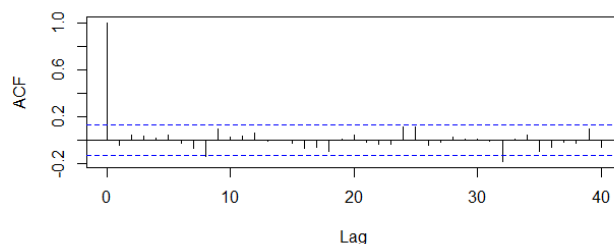
Como se puede observar en la Figura 3, los residuos presentan dependencia serial de hasta un desfase de 4 meses. Por otra parte, la correlación parcial mostrada en la Figura 5 presenta significancia hasta 8 meses. De esta forma, un posible modelo que explique la correlación temporal sería un modelo ARMA(8,4). Sin embargo, el modelo entregado por la función autoarima de la librería forecast corresponde a un AR(1), el cual nos dice que

el valor actual está influenciado principalmente por el valor inmediatamente anterior.



**Figura 5:** Gráfico de correlación parcial para residuos modelo de regresión

mayores al 5%, por ende, se concluye los residuos del modelo no están correlacionados.



**Figura 6:** Gráfico ACF modelo AR(1)

## Análisis de Supuestos Residuos Modelo AR(1)

**Media constante** Al igual que en la sección anterior, se implementa un modelo de regresión entre los residuos y el tiempo, donde el valor-p asociado al test de la significancia del predictor es 0.98. De esta forma se concluye que la media es constante con respecto al tiempo.

**Homocedasticidad** En la sección anterior se evidencio que el valor-p era cercano a 0.05. Ahora, el valor-p asociado al test de Breusch-Pagan corresponde a 0.31. Por ende, se mantiene la homocedasticidad en los datos.

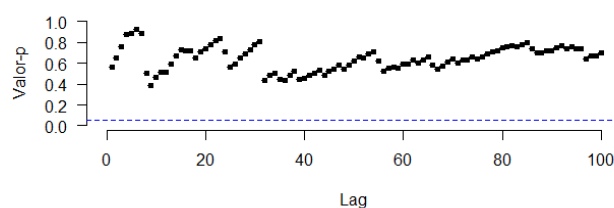
**Normalidad** Se aplicaron nuevamente los test Shapiro-Wilks y Kolmogorov-Smirnov a los residuos del modelo de series de tiempo.

**Tabla 2:** Valores-p para test de Normalidad

	Valor-p
Shapiro-Wilks	0.009
Kolmogorov-Smirnov	0.43

Los resultados de aplicar cada test se aprecia en la Tabla 2. Por un lado, el test de Shapiro-Wilks muestra evidencia que los datos no son normales, mientras que el test Kolmogorov-Smirnov sustenta el supuesto de normalidad en los residuos. La diferencia entre las conclusiones se debe a la presencia de un dato con error demasiado alto, sumado a la forma de determinar los estadísticos en cada test dependerá de la conclusión que se obtenga.

**Ruido Blanco** En la Figura 6 es posible notar que se elimina la tendencia en los primeros saltos de la correlación temporal, lo anterior presenta una mejora con respecto a la Figura 3, por lo que da indicios de buen ajuste del modelo de series de tiempo. Además se observa que en los desfases 7 y 32 las correlaciones son significativas. No obstante, como se puede observar en la Figura 7 todos los valores-p asociados el test de Box-Ljung son



**Figura 7:** Valores-p para Test de Box-Ljung para los residuos modelo AR(1)

## Conclusiones

Como se mencionó anteriormente, los niveles de monóxido de carbono no logran seguir una tendencia lineal ni constante a través del tiempo, por lo que se consideró un modelo cuadrático con respecto a esta covariable. Por otro lado, de las tres variables climáticas que se tenían a disposición, tan sólo fue significativa la temperatura, la que a su vez interactuaba con el tiempo. Asimismo, fue necesario considerar el mes en que se realizó cada medición, debido a que este factor ayudó a comprender de mejor manera el comportamiento estacional de los niveles de CO. Posterior al ajuste se evalúan los supuestos usuales para series de tiempo, donde se advierte la presencia de autocorrelación de los residuos del modelo de regresión.

Por consecuencia, se determinó ajustar un modelo AR(1), propuesto por la función autoarima, en base a los residuos del la regresión. Al igual que en el modelo anterior, se evalúan nuevamente los supuestos, donde se concluye que el modelo de series de tiempo logran capturar la dependencia serial de los residuos, aunque en el test de normalidad no se presentaron mejoras.