

Tarea 1

EYP3907 - Series de Tiempo

Sebastián Celaya

Camila Echeverría

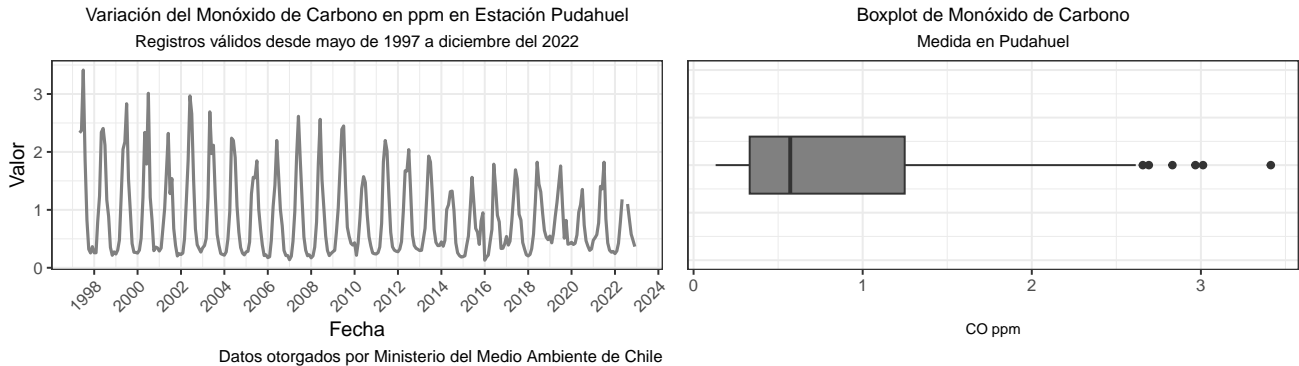
Francisca Vilca

Introducción

En esta tarea, utilizaremos los datos mensuales de concentración de monóxido de carbono (CO), recopilados en la estación de monitores de Pudahuel, para ajustar un modelo de regresión que nos permita predecir la concentración de este contaminante a lo largo del tiempo.

Análisis Exploratorio

Los niveles de Monóxido de Carbono (CO) registrados en Pudahuel presentan los siguientes valores:

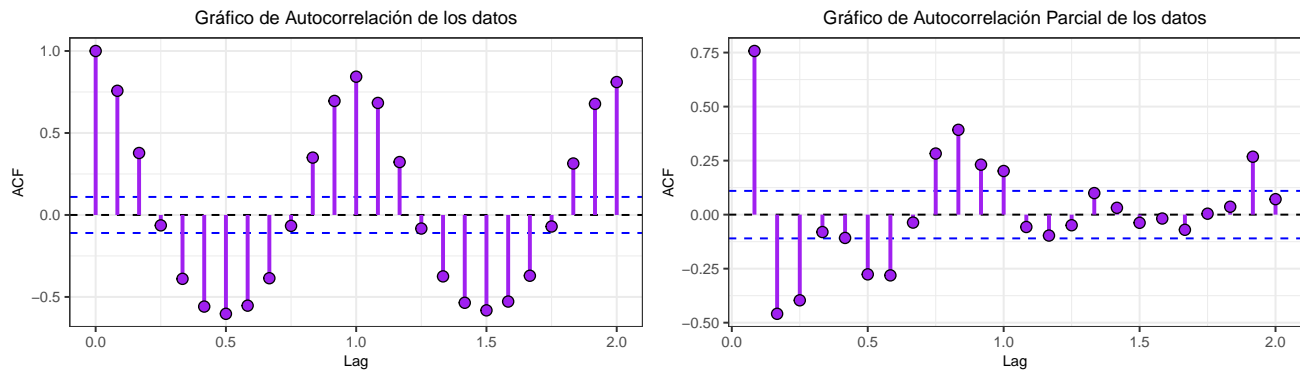


Del gráfico anterior se puede ver que los niveles de CO son valores cercanos a 0.5, igualmente se observa que hay una gran varianza. A modo de resumen, se ven los valores del gráfico anterior en la Tabla 1.

| Mínimo | 1° Cuartil | Mediana | Promedio | 3° Cuartil | Máximo | Valores NA |
|--------|------------|---------|----------|------------|--------|------------|
| 0.1305 | 0.3318 | 0.5716 | 0.8599 | 1.2484 | 3.4134 | 3 |

Tabla 1: Tabla con Cuartiles

Los datos temporales anteriormente presentados presentan los siguientes gráficos de autocorrelación y de autocorrelación parcial:



Ajuste del modelo

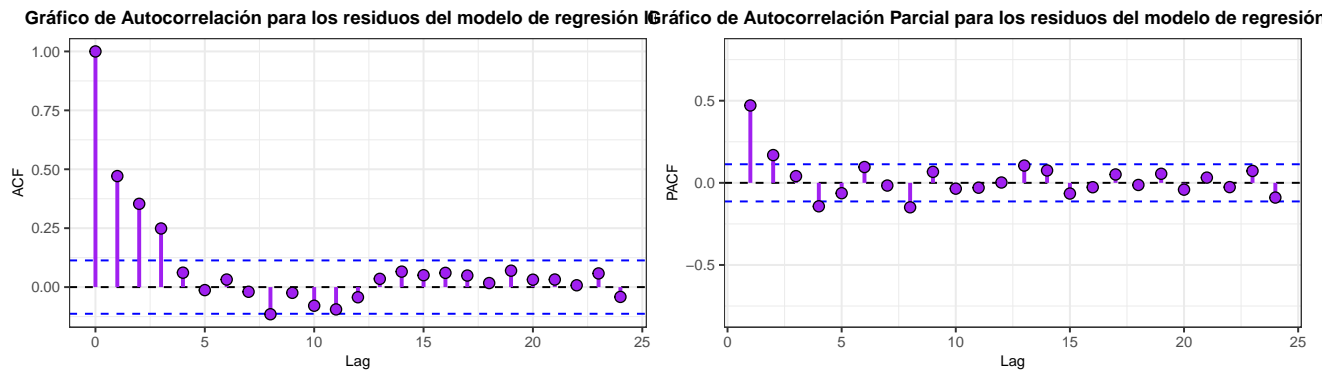
Como se pudo ver en el gráfico de autocorrelación, presentamos estacionalidad en los datos. Es por esto que utilizaremos la transformación de box-cox en ambos ajustes.

Ajuste modelo de regresión lineal con dummies

Al tener un período de estacionalidad igual a 12, ajustaremos un modelo regresión lineal con 12 variables dummies. Consideramos, además, la transformación de box-cox utilizando $\lambda = 0.22$ para este modelo. Así, podemos obtener las siguientes conclusiones:

- Con un valor-p de 0.01, el modelo rechaza los test de normalidad.
- Por otro lado, con un valor-p de 0.12, el test de Breusch-Pagan no rechaza la homocedasticidad, por lo que podemos concluir que el modelo presenta residuos con varianza constante.

Luego, vemos los gráficos de autocorrelación y autocorrelación parcial de los residuos:



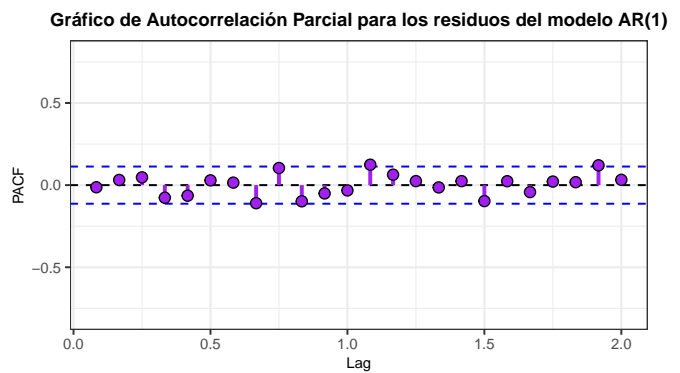
De esta manera, se puede apreciar que el modelo que sugiere esta estructura de correlación corresponde a un ARMA(2,4).

Ajuste con función `auto.arima()`

CAMBIAR LOS VALORES CON LOS DATOS DE ESTA WEÁ

Si se realiza el ajuste de la serie de tiempo usando la función `auto.arima()` de la librería `forecast`. Esta recomienda usar un modelo AR(1) en los datos y de los resultados del ajuste, es posible dar razones de porque es recomendable usar este modelo:

- Los residuos son homocedásticos, esto es sencillo de notar, ya que, el test de Breusch-Pagan no se rechaza a un nivel de confianza al 5%, por lo que se permite afirmar que los residuos son de varianza constante.
- No hay autocorrelación en los residuos: Como el interés es que no existan patrones en los residuos, es bastante bueno que el modelo AR(1) que recomienda la función `auto.arima()`, demuestre que no pasa:



- Los residuos no son normales: A pesar de que el modelo creado es bueno en muchos sentido, no logra normalizar los residuos, lo que se comprueba sencillamente con test de Kolmogorov-Smirnov, pues el valor-p asociado a dicho estadístico es extremadamente pequeño lo que nos hace rechazar la normalidad en los residuos.

Predicciones de los modelos:

Ahora que se tiene ambos modelos, una forma de poder determinar su funcionamiento es probandolos, para ello se usarán datos de testeo correspondiente a los meses de enero a septiembre del 2023, los cuales se pueden visualizar en la Tabla 2:

A simple vista de la Tabla 2, es claro observar, que hay valores que si se estiman bastante bien como el mes de abril o agosto, y otros que fallan rotundamente como mayo. Dado que de esta forma es difícil determinar cuál es mejor se usarán algunos valores para medir la calidad del ajuste como el MAPE o MSE, los cuales pueden ser observados en la Tabla 3:

Argumentar que volaita es cada uno y pq es mejor el `auto.arima` pero no se como sacar los valores del modelo 1

| Mes | Valor Real | Predicción Modelo 1 | Predicción Modelo AR(1) |
|------------|------------|---------------------|-------------------------|
| Enero | 0.37 | 0.35 | 0.29 |
| Febrero | 0.57 | 0.33 | 0.31 |
| Marzo | 0.56 | 0.39 | 0.43 |
| Abril | 0.66 | 0.60 | 0.73 |
| Mayo | 1.17 | 0.94 | 1.91 |
| Junio | 1.53 | 1.44 | 1.47 |
| Julio | 1.47 | 1.41 | 1.40 |
| Agosto | 0.98 | 0.90 | 0.88 |
| Septiembre | 0.73 | 0.57 | 0.56 |

Tabla 2: Comparación de predicciones

| | RMSE | MAE | MAPE | R2adj |
|------------------|------|------|-------|-------|
| Regresión Lineal | 0.15 | 0.12 | 16.35 | 0.90 |
| Modelo AR(1) | 0.24 | 0.14 | 17.27 | 0.70 |

Tabla 3: Medidas de calidad de ajuste

Conclusión y Discusión

Explicar por alguna razón pq el auto.arima es mejor en general q como van subiendo año a año las volaitas y no se alguna otra cosa que se nos ocurra para terminar esto yuju