

Tarea 1

EYP3907 - Series de Tiempo

Sebastián Celaya

Camila Echeverría

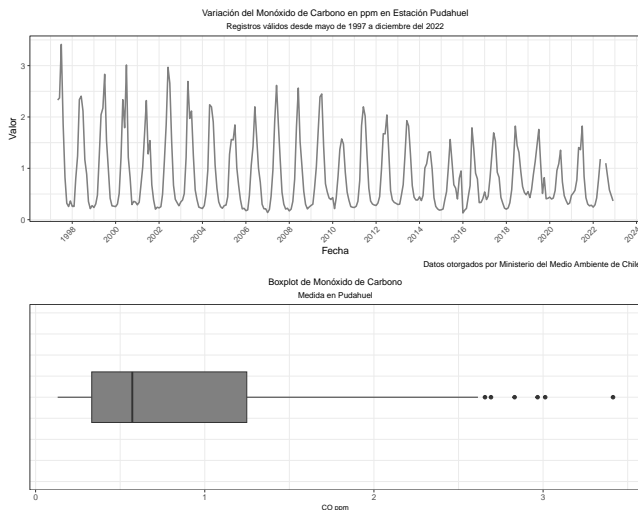
Francisca Vilca

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.

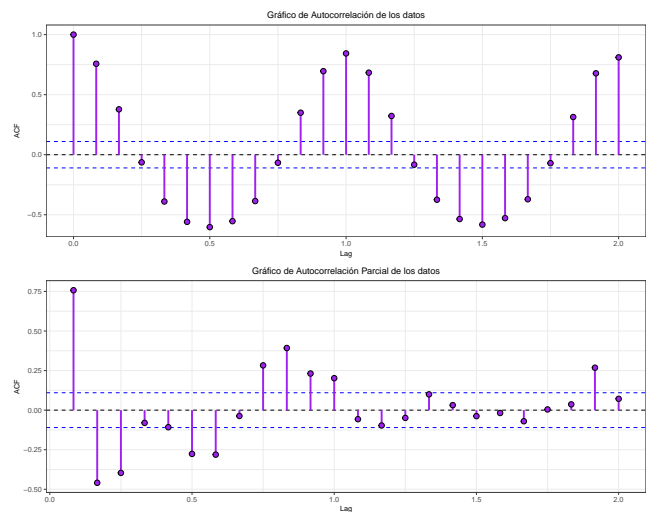
Análisis Exploratorio

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



Del gráfico anterior se puede ver que la mitad de los datos presenta un nivel de CO cercano al 0.5 y la otra mitad tiene un rango de valores mucho mayor. De igual forma se aprecia que los datos presentan una gran varianza.

Luego, si evaluamos los gráficos de autocorrelación y autocorrelación parcial de los datos presentados anteriormente, es claro notar que estos presentan una alta estacionalidad y correlación, a través del tiempo, tal como se observa a continuación:



Ajuste del modelo

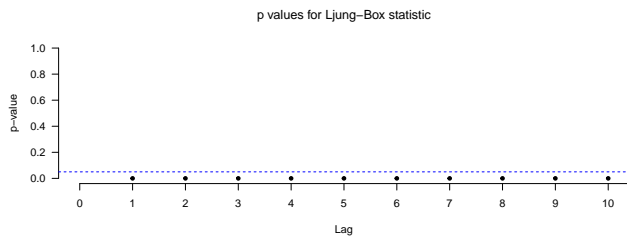
Como se pudo ver en el gráfico de autocorrelación, presentamos estacionalidad en los datos. Es por esto que utilizaremos la transformación de box-cox en los ajustes propuestos:

Ajuste modelo de regresión lineal con dummies

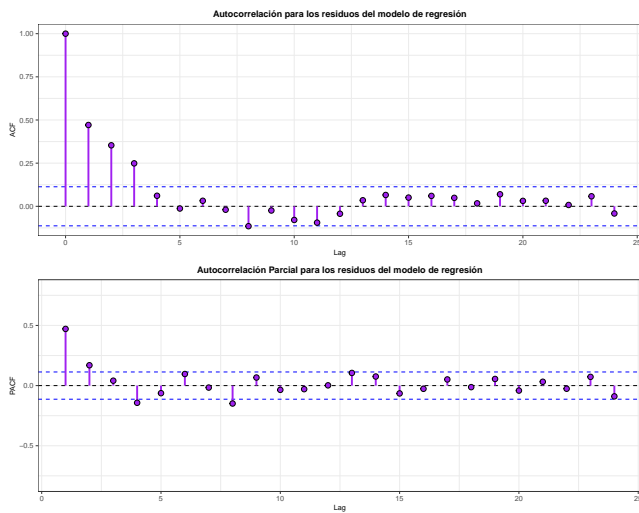
Al tener un período de estacionalidad igual a 12, ajustaremos un modelo regresión lineal con 12 variables dummies. Consideramos, además, la transformación de box-cox utilizando $\lambda = 0.22$ para este

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.
- Si se revisa el gráfico de Ljung-Box, es claro notar que se rechaza el test de blancura, por lo que se concluye que los datos podrían provenir de un ruido blanco.



Luego, si vemos los gráficos de autocorrelación y autocorrelación parcial de los residuos, vemos que aun existen patrones de estacionalidad.



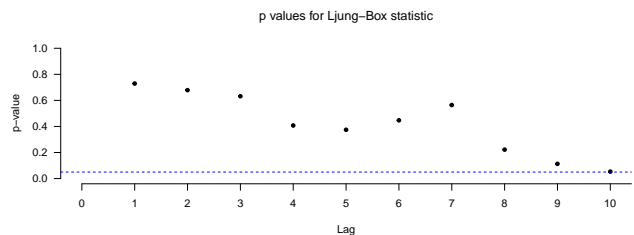
De esta manera, se puede apreciar que el modelo que sugiere esta estructura de correlación corresponde a un ARMA(2,4).

Ajuste con función `auto.arima()`

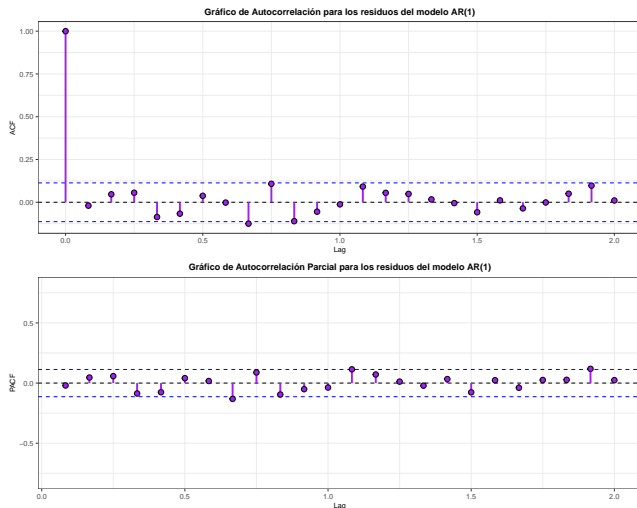
En base a las conclusiones obtenidas anteriormente, utilizaremos la función `auto.arima()` de la librería `forecast` para ajustar un modelo ARMA(2,4) a los datos. Algunos resultados importantes de este modelo son los siguientes:

Al usar en la función `autorima` un modelo ARMA(2,4), la función recomienda un modelo AR(1) que explique de mejor manera la estacionariedad de los datos, de lo que se puede concluir lo siguiente:

- Los residuos son homocedásticos, ya que el test de Breusch-Pagan no se rechaza a un nivel de confianza de 1%, pues el valor-p asociado es de 0.12
- Si se revisa el gráfico de Ljung-Box, es claro notar que no se rechaza el test de blancura, pues a pesar de tener datos muy cercanos a la banda de confianza ninguno está por debajo de la misma, por lo que se concluye que los datos no provienen de un ruido blanco.



- También, es posible observar que los residuos no presentan autocorrelación serial, tal como se muestra a continuación



- A pesar de que el modelo creado es bueno en muchos sentidos, el valor-p del test de Kolmogorov-Smirnov es mucho menor a 0.05, lo que nos indica que los residuos rechazan la hipótesis nula de normalidad.

Comparación de los modelos

Para poder probar ambos modelos ajustados, utilizaremos los datos de testeo para generar predicciones desde enero hasta septiembre del 2023. Estos resultados se presentan en la siguiente tabla:

Mes	Valor Real	Regresión	ARMA(4,2)
Enero	0.37	0.35	0.32
Febrero	0.57	0.33	0.33
Marzo	0.56	0.39	0.44
Abril	0.66	0.60	0.75
Mayo	1.17	0.94	1.21
Junio	1.53	1.44	1.63
Julio	1.47	1.41	1.64
Agosto	0.98	0.90	0.99
Septiembre	0.73	0.57	0.61

Tabla 1: Comparación de predicciones

Ambos modelos generan predicciones bastante cercanas a los valores reales para los meses correspondientes. Para complementar esto, en la siguiente

tabla compararemos algunas medidas de calidad de ajuste:

	RMSE	MAE	MAPE
Regresión Lineal	0.15	0.12	16.35
Modelo AR(1)	0.23	0.14	17.00

Tabla 2: Medidas de calidad de ajuste

Conclusión

A pesar de que el modelo de regresión lineal pareciera tener un mejor desempeño en base a las medidas de calidad de ajuste, no podemos olvidar que no cumple con el supuesto de normalidad y que posee autocorrelación serial, por lo que el modelo ARMA(2,4) representa un mejor ajuste para nuestros datos.