

Laboratorio 3.

Series de Tiempo.

INSTRUCCIONES:

Utilice el conjunto de datos que se le provee de la importación de combustibles desde enero de 2001 hasta marzo de 2020. Trabaje únicamente con los datos de gasolinas super, regular y el diesel. Considere como las importaciones de diesel desde 2018, los datos del diesel con bajo contenido de azufre (dieselLS).

DESCRIPCIÓN DEL DATASET

El conjunto de datos contiene 231 observaciones de 25 variables. Representa el volumen de importaciones de distintos tipos de combustibles desde enero de 2001 hasta marzo de 2020. El volumen está dado en barriles de 42 galones. El documento lo puede encontrar en el siguiente vínculo: <https://www.mem.gob.gt/wp-content/uploads/2020/03/C01-Importaci%C3%B3n-de-combustibles-VOLUMEN-2020-03.pdf>

Algunos vínculos interesantes:

- <https://otexts.com/fpp2/>
- <https://otexts.com/fpp2/arima.html>
- <https://otexts.com/fpp2/accuracy.html>
- <https://robjhyndman.com/papers/mase.pdf>

EJERCICIOS

1. Haga un análisis exploratorio de los datos que se le presentan, se sugiere explorar el comportamiento de las variables y si están distribuidas normalmente, en caso de ser continuas. Meses en los que más importaciones hay, picos en importaciones por año por tipo de combustibles, comportamiento en los últimos x años, etc.
2. Haga una serie univariante por cada una de las variables (gasolina regular, gasolina super, diesel). De cada serie:
 - a. Especifique Inicio, fin, y frecuencia.
 - b. Haga un gráfico de la serie y explique qué información puede obtener a primera vista.
 - c. Descomponga la serie en sus componentes. Teniendo en cuenta el diagrama de la serie y sus componentes discuta si es posible hablar de estacionariedad en media y en varianza.
 - d. Determine si es necesario transformar la serie. Explique.
 - e. Explique si no es estacionaria en media. Para esto:
 - i. Haga el gráfico de autocorrelación y úselo para explicar la no estacionariedad en media.
 - ii. Básese en los valores de estadísticos como la prueba de Dickey-Fuller Aumentada para corroborar la no estacionariedad en media. ¿Qué es necesario hacer para hacerla estacionaria en media en caso de que no lo sea?

- f. Una vez analizada la serie, elija los parámetros p , q y d del modelo ARMA o ARIMA que utilizará para predecir. Explique en qué se basó para darle valor a estos parámetros, basándose en las funciones de autocorrelación y autocorrelación parcial. Si usa la función `autoarima` de R, determine y explique si tiene sentido el modelo propuesto.
 - g. Haga varios modelos ARIMA, y diga cuál es el mejor de ellos para estimar los datos de la serie.
 - h. Haga un modelo usando el algoritmo prophet de Facebook. Compárelo con los modelos del inciso anterior. ¿Cuál funcionó mejor?
3. Haga una predicción de los valores de la serie para los años 2018, 2019 y 2020. Explique qué tan buena es su predicción.
 4. Haga una predicción del año 2020 usando los años anteriores y determine qué tan apegada es a la realidad teniendo en cuenta los meses pasados del año actual.

EVALUACIÓN

(15 puntos) Análisis exploratorio:

- Se elaboró un análisis exploratorio en el que se explican los cruces de variables, hay gráficos explicativos y análisis que permiten comprender el conjunto de datos.
- Se crearon las series de tiempo correspondiente a los datos de las variables que representan las importaciones de diesel, gasolina superior y gasolina regular.
- Para cada una de las series se informa inicio, fin y frecuencia.

(15 puntos) Análisis de las series de tiempo

- Para cada una de las series creadas se analiza:
 - o El gráfico de la serie y su descomposición en componentes.
 - o Si la serie presenta estacionalidad o no y que implica que sí tenga.
 - o Si la serie presenta tendencia o no y esto que significa.

(25 puntos) Determinación de Estacionariedad.

- Para cada una de las series creadas:
 - o Se analiza si es estacionaria en varianza y en caso de no serlo se aplica una transformación adecuada.
 - o Se analiza si es estacionaria en media, para esto se basa en la función de autocorrelación y en la prueba de Dickey-Fuller aumentada. Se determina la cantidad de diferenciaciones que hay que hacer en caso de que no sea estacionaria en media.

(25 puntos) Generación de modelos

- Para cada una de las series creadas:
 - o Se determinan los valores de los parámetros p , q , y d . Para esto se basa en las funciones de autocorrelación y autocorrelación parcial.
 - o Se explica la elección de los parámetros y de los modelos en caso de que se propongan varios modelos. Se deben explicar los parámetros, aunque sean propuestos de forma automática por R (en caso de usar este lenguaje).
 - o Se genera un modelo con el algoritmo prophet.

(20 puntos) Predicción con los modelos generados.

- Para cada una de las series creadas:
 - o Se crean los conjuntos de entrenamiento y prueba siguiendo las instrucciones.

- Se explica que tan bueno es el modelo prediciendo los volúmenes de importación de combustible para el conjunto de prueba.
- Se comparan los modelos generados con ambos algoritmos.

MATERIAL A ENTREGAR

- Archivo .pdf con el informe que contenga, los resultados de los análisis y las explicaciones.
- Link de Google drive donde trabajó el grupo.
- Script de R (.r o .rmd) o de Python que utilizó para responder las preguntas con el código utilizado o archivo de flujo de trabajo de KNime
- Link del repositorio usado para versionar el código.

FECHAS DE ENTREGA

- **AVANCE:** Análisis Exploratorio, análisis del diesel y primer modelo ARIMA con su predicción y toda la explicación: jueves 30 de julio 18:55.
 - **DOCUMENTO FINAL COMPLETO:** lunes 3 de agosto de 2020 23:59
- NOTA:** Solo se calificará el Documento Final si está entregado el avance con todo lo que se pide.