



"El saber de mis hijos
hará mi grandeza"

UNIVERSIDAD DE SONORA

DEPARTAMENTO DE FÍSICA

ACTIVIDAD 6

Autor:

Sebastián Browarski Ruiz

19/02/21

1. Introducción

Previamente empezamos a trabajar con series de tiempo, ahora nos adentramos al pronóstico de series de tiempo y para esto requerimos algún modelo que nos permita predecir valores futuros en base a los datos ya obtenidos en el pasado. Antes de usar un modelo, se deben de hacer algunos ajustes previos, primero que todo hay que revisar la estacionaridad. Una serie de tiempo es estacionaria si sus propiedades estadísticas (promedio, varianza, covarianza) permanecen constantes en el tiempo, y con esto podemos prever que las propiedades de la serie de tiempo se van a comportar de la misma forma en el futuro. Para saber si la serie de tiempo es estacionaria o no, primero creamos la función proporcionada por el profesor que nos facilitará el análisis y luego utilizamos la Prueba Aumentada de Dicker-Fuller (ADF), que viene incluida en la biblioteca Statsmodels. Después habrá que hacer otros cambios como eliminar la tendencia y la estacionalidad, y finalmente empezaremos con el pronóstico de series de tiempo. Para realizar esto, emplearemos el método de modelación estadística llamado ARIMA (Auto Regresive Integrated Moving Average).

2. Comentarios Generales de la Información Analizada y realizada en Google Colab

Como actividades anteriores, utilizamos los datos de las estaciones meteorológicas por todo México, en mi caso yo use los de la estación 2038 en Tijuana, BC. Al trabajar con mis datos encontré varios espacios vacíos, por lo que tuve que ir buscando un buen intervalo de tiempo hasta que encontré uno. Mi intervalo de tiempo estaba completo para la Tmax pero no para la Tmin, por lo que a los datos de Tmin tuve que aplicarles la función `.dropna(inplace=True)`. También en la serie Tmax no tuve ninguna complicación para conseguir un valor de $p < 0.05$, pero para Tmin hubo que hacerle unos cambios de escala con $\log(x)$ y luego cambiar el tipo de promedio móvil a Exponencial (Exponential Weighted Moving) con la función de pandas `pd.ewm()`.

3. ¿Qué puedes agregar de las series de temperatura que analizaste?

1. Respecto a la estacionaridad de tus series. ¿Qué procedimientos seguiste para obtener una serie estacionaria?

Primero se debe definir una función para aplicar la Prueba Aumentada de Dickey-Fuller(ADF), a veces se requiere hacer cambios porque la prueba ADF no admite datos faltantes. Luego se debe de crear los dataframes con los cuales se trabajarán. El objetivo es conseguir que el valor p de ADF sea menor a 0.05 para así afirmar que la serie de tiempo es estacionaria con 99 % de confianza.

Para Tmax no tuve problema, pero para Tmin tuve que cambiar la escala y hacer cambios por los datos faltantes, y ya con esto obtuve las series estacionarias.

2. Respecto a las tendencias de Tmax y Tmin. ¿Qué se observa?

En temperaturas, se ha notado a lo largo de los últimos años que la tendencia es creciente debido a la contaminación humana y a la destrucción de la capa de ozono. Es por eso que la tendencia en todos los casos no es constante, sino que va ligeramente en aumento. Como este es el caso, hubo que hacer ajustes a la tendencia para lograr que la serie fuera estacionaria.

3. ¿Qué puedes decir en general sobre los datos de la estación que estuviste analizando?

Con los datos de temperaturas se puede observar que las más altas están en veranos, aproximadamente entre julio y agosto, y las más bajas se encuentran en inviernos, en diciembre y enero. La estación fue bastante consistente, en comparación a los miles de datos hay relativamente pocos huecos.

4. ¿Qué limitaciones encontraste en tus datos?, ¿Vacíos?, ¿Cuál fue el periodo más largo que pudiste encontrar?

Mi intervalo de tiempo fue de 1999-01 a 2002-12, con la temperatura máxima no hubo problema ya que no hubo vacíos pero en la temperatura mínima sí hubo entonces eso fue algo que tuve que trabajar. Escogí ese intervalo porque cuatro años se me hizo considerable y el más largo que encontré, también por la temperatura máxima.

5. Otros comentarios generales

Las funciones que aplicamos se ven bastante complejas y dan mucha información que probablemente es útil pero que no sabría emplear.

4. Primeras Impresiones y Retroalimentación

La actividad fue desafiante ya que usamos funciones y aplicaciones con mucha información que es difícil de interpretar y llegó a ser sofocante pero supongo que así es analizar series de tiempo; siento que entendí la base de lo que aplicamos pero no el todo en sí, y eso es algo que me gustaría mejorar, que la actividad hubiera sido algo más suave y no tanta información de golpe. Como los datos pueden llegar a ser inconsistentes, fue algo tedioso tener que estar averiguando qué intervalo nos servía o qué cambios había que hacer. A pesar de estas cosas, con bastante trabajo creo que logré el objetivo y mis resultados fueron los esperados. Le pondría un nivel avanzado de dificultad.