
Fisica Computacional - Series de tiempo



Gustavo Alberto Medina Ferrer¹

A219223438¹

Numero(s):(81) 8309 6131¹

gustavomedinaferrer@gmail.com¹

Universidad de Sonora

Licenciatura en Fisica

Fisica Computacional

12/02/2021

Resumen

En este documento se hará un reporte de la actividad 5 de la materia de *Física Computacional*, impartida por el profesor C. Lizarraga en la Universidad de Sonora.

La actividad trata acerca de analizar los datos previamente elegidos de la base de datos de CONAGUA en línea (en este caso, de la estación 26068-Presa Alvaro Obregón), en series de tiempo. Se hizo uso de conceptos tales como *promedios móviles* y la descomposición de una serie de tiempo.

1. INSTRUCCIONES

1.1. Actividad 1: Explora tus datos de Precipitación los últimos años 1990-fin de datos.

- Grafica todo el rango de datos utilizando la función `df.plot()`
- Analiza un año específico completo, utiliza la función `df.loc()` para seleccionar el rango de datos.
- Grafica la estacionalidad de tus datos por mes utilizando la función `sns.barplot()`

1.2. Actividad 2: Explora tus datos de Temperaturas Máxima y Mínima de los últimos 30 años datos a partir de 1990. Trata de construir dos sub-gráficas contiguas.

- Grafica todo el rango de datos utilizando la función `df.plot()`, incluye una línea punteada mostrando el promedio.
- Analiza un año específico completo para las dos Temperaturas, con apoyo de la función `df.loc()`
- Grafica la estacionalidad filtrando tus datos por mes y utilizando la función `sns.boxplot()`, incluye una línea punteada mostrando el promedio.

- 1.3. **Actividad 3:** Realiza el mismo análisis que en la Actividad 2, pero para la variable de Evaporación.
- 1.4. **Actividad 4:** Crea nuevos DataFrames de las Temperaturas y la Evaporación con promedios móviles de 7, 30 y 365 días, utilizando la función `df.rolling()`. Explora tus datos:
 - Para un año de datos, contrastando los datos diarios con el promedio móvil de 7 y 30 días.
 - Explora para 10 o más años de datos, contrastando los datos diarios con las gráficas de promedio móvil de 365 días.
- 1.5. **Actividad 5:** Utiliza la biblioteca `statsmodels.tsa` de Python, para realizar una descomposición de una serie de tiempo: $\text{Serie Observada} = \text{Tendencia} + \text{Estacionalidad} + \text{Residuo}$. Aplica la función `seasonal_decompose` para lograr esto, a los datos de Temperaturas Tmax y Tmin, y si tienes datos de Evaporación, selecciona un periodo que haya datos. La función de `seasonal_decompose` aplica su propio promedio móvil a las series de datos para su análisis.
- 1.6. **Actividad 6:** Por favor explora visualmente si se observa un cambio en la distribución de Temperaturas (histogramas y kde), sobreponiendo dos histogramas de dos periodos de 30 años contiguos. Utiliza la función de Seaborn: `sns.distplot`. Agrega una línea vertical punteada indicando el valor del promedio de cada grupo de datos.
- 1.7. **Producto a entregar**

En tu repositorio en Github deberá haber una carpeta Actividad5 que contenga:

1. Una copia de tu cuaderno de trabajo Jupyter elaborado en Google Colab.
2. Una copia en PDF de tu reporte en LaTeX elaborado en Overleaf, de la actividad desarrollada.

El reporte debe incluir:

1. Título del reporte, Autor, Fecha.

2. Una breve introducción de la Actividad desarrollada y algunas características que hayas notado en el análisis inicial del sitio de estudio.
3. Comentarios generales de la información analizada y realizada en Google Colab.
4. Agrega la sección con tus primeras impresiones de la Actividad desarrollada. Se pide un poco de retroalimentación de parte tuya para analizar los resultados de la Actividad. ¿Qué te pareció el tema de Análisis Exploratorio de Datos?, ¿Cómo estuvo el reto?, ¿Qué se te dificultó más?, ¿Qué te aburrió?, ¿Qué recomendarías para mejorar la primera Actividad?. ¿Que grado de complejidad le asignarías a esta Actividad? (Bajo, Intermedio, Avanzado).

Habilidades a demostrar

- Autogestión de tiempo.
- Habilidades computacionales en Internet.
- Elaboración de documentos técnicos en LaTeX
- Uso de Notebooks de Jupyter en Google Colab.
- Manipulación de DataFrames en Pandas.
- Manejo básico de funciones de visualización con Matplotlib y Seaborn.
- Introducción al Análisis de Series de Tiempo (Promedios móviles, tendencias, estacionalidad)
- Organización y manejo de repositorios en Github.
- Lectura en Inglés
- Otros ...

2. ACTIVIDAD

En la primera actividad se nos pidió graficar las variables de Precipitación, Evaporación, TMax y TMin en periodos de 30 años. Los resultados (como los de todas las actividades) se pueden observar en el archivo ipnyb que se subirá junto con este reporte.

Algunas cosas a destacar de esta primera actividad es que se observa que en el año 2009-2010, aproximadamente, hay un hoyo en los datos. Esto puede ser debido a algún fenómeno meteorológico o de cualquier otro tipo que dejó a la estación sin la capacidad de registrar datos.

En la siguiente actividad de la parte 1, se puede ver el patrón senoidal del que he hablado en los últimos reportes. Se generó una gráfica con los tres periodos de 30 años para analizar de manera general los datos, donde también se muestra el mismo patrón.

En la parte 2, se obtuvieron graficas de las temperaturas desde 1990 hasta el 2016. En esta parte en especifico se tardo mucho tiempo, ya que se intento por mucho tiempo el poder graficar las subplots en diferentes graficas con un loop, sin embargo no se pudo, por lo que se observan las dos en una misma grafica.

Con la parte de los boxplot me diverti encontrando una paleta de colores.

Lo mas interesante fue en la parte 4, ya que fue lo que me dio mas desafio, sin embargo es la parte que mas rapido resolvi. El problema fue en formular un algoritmo para un loop que me imprimiera las graficas de cada variable de un solo. Sin embargo, despues de formularlo, no tuve que hacer nada mas que ejecutarlo. Pero surgio el problema de que cada vez que ejecutaba esas celdas me daba el error de que mi sesion habia sido cerrada por un problema desconocido. Para resolverlo tenia que reiniciar el entorno de ejecucion cada vez que queria volver a ejecutar esas celdas.

Como ultimo analisis, podemos observar en los histogramas contiguos de 60 años en total que los datos de 1979 a 2009 superan a los de 1949 a 1979 en las temperaturas mas altas, mientras que los de 1949 a 1979 superan a los otros en las temperaturas mas bajas, lo que quiere decir que hay un incremento de la frecuencia de las temperaturas altas en los ultimos años. Se adjuntan las graficas.

