

Reto 2:

La interpolación y el clima (12 de mayo 2021)

Nicolas Puerto, Tomas Rivera, Juan Andrés Mejia

Resumen— En el siguiente artículo se lleva a cabo la explicación del análisis realizado por un algoritmo que tiene como entrada, información de índole climatológica recogida en el año 2013 de la ciudad Fortaleza ubicada en la costa noroeste colindante con el mar atlántico sur (Brasil). El análisis se realiza por medio de diferentes metodologías de interpolación ya propuestas en el enunciado que, buscan predecir y estimar el error y comportamiento de ciertas variables(meteorológicas).

Palabras clave— Interpolacion, Clima, Splines, Hermite,

Paquete	Descripción:
Readr	El paquete readr facilita la lectura y separación de archivos de tipo Excel(CSV)
Polynomf	El paquete Polynomf alberga una colección de funciones polinómicas, esto permite su manipulación en el lenguaje R
Pracma	El paquete pracma al igual que el polynom, nos propicia una serie de funciones y operaciones algebraicas y diferenciales

Humedad, Temperaturas.

I. INTRODUCCIÓN

Dentro de el campo de las matemáticas, más específicamente en el del análisis numérico, la interpolación es y ha sido una herramienta muy practica que ha facilitado diversos tipos de procesos dentro de diferentes campos de la vida cotidiana, estos pueden ser; estadísticas, medio ambiental y hasta el medio audiovisual. El método de interpolación busca realizar una predicción de puntos nuevos con base a unos ya existentes dados. Las interpolaciones son clasificadas de manera genérica de diferentes formas: Una de ellas se conoce como interpolaciones globales, esta utiliza todos los puntos de muestra para determinar la forma de una función matemática aplicada en un área de interés normalmente. Otra se conoce como interpolación local, la cual a diferencia de la global, utiliza puntos dados que están cercanos a un punto mas exacto desconocido para así poderlos encontrar. En el mundo de la interpolación, podemos también, diferenciarlas por dos palabras claves; estocástico y determinísticos. La diferenciación entre estas dos hace referencia a la exactitud de los parámetros usados en la interpolación, se le llama determinísticos cuando se tiene certeza sobre los valores de los datos de entrada para poder llevar a cabo la interpolación y, se les llama estocásticos cuando el valor de esos datos no es

preciso y pueden llegar a variar.

En este caso, la interpolación nos ayudara de forma determinista a predecir algunas variables meteorológicas como pueden ser la humedad, temperatura entre otras. Para esto se utiliza un estudio realizado en Brasil, más específicamente en fortaleza, una ciudad costera ubicada en el noroeste de el país.

II. METODOLOGIAS UTILIZADAS.

El algoritmo utilizado para la solución de este problema este hecho en el lenguaje R y compilado por el IDE Rstudio. Esta decisión fue tomada puesto a que el lenguaje R permite mejor la manipulación de archivos tipo Excel o CSV por medio de librerías y los cuáles serán los portadores de la información que se tendrá que analizar. Dentro de R, tenemos librerías que nos facilitaran el uso de los diferentes métodos de interpolación que se utilizaran.

Cuadro de los paquetes implementados:

Código de las librerías:

```
library(readr)
library(PolynomF)
library(pracma)
```

Breve descripción de las metodologías:

- Hermite: Esta interpolación se realiza por medio de la función: `splinefun()`, esta recibe arreglos con la información para después interpolarla por medio del método de interpolación de Hermite.
- Splines: Este método se realiza por medio de la función `spline()` y, al igual que `splinefun()`, recibe los arreglos con los datos para realizar la interpolación por splines regulares.
- Lineal: La interpolación lineal se realiza con el método

`approx()`. Este recibe los mismos datos que `spline` y `splinefun` además de un parámetro especificando el método por el cual se realizara, en este caso “linear”.

Para ver la efectividad de los tres métodos diferentes se considera una variable que mira el criterio de aceptación (`criterioDeAceptacion`), este criterio lo hemos definido después de haber probado el programa con los valores aleatorios, se considera que, en base a los resultados y la cantidad de datos, el hecho de que el error cuadrático medio sea mayor que 90% nos da un indicador de efectividad aceptable.

III. PROBLEMÁTICA 1

Para la primera problemática presentada, hemos optado por escoger Fortaleza como ciudad para llevar a cabo el análisis de los datos. En esta se nos presentan 15 datos de diferentes estaciones climáticas cercanas a la ciudad. Para la realización del punto optamos por escoger la variable de **temperatura interna**. En el siguiente grafico se aprecia el comportamiento de esta variable a lo largo de poco más de 700 días.

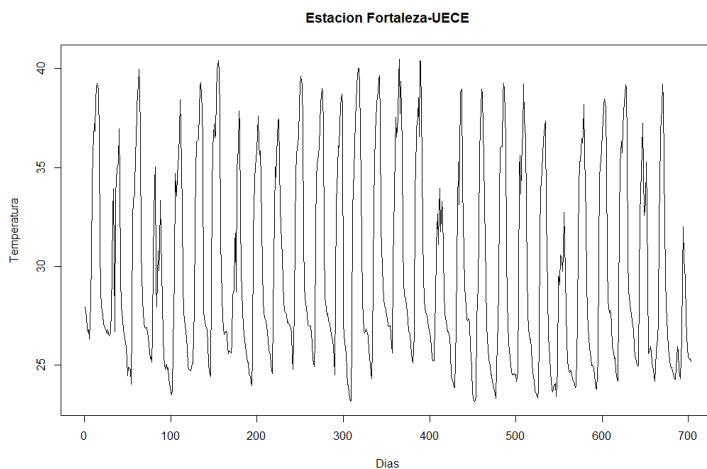


Imagen 1: Comportamiento de la variable

Con fin de no ser sesgado por los valores y tener la mayor precisión en cuanto a valores se refiere hemos escogido valores al azar que representan el 80% de la información original presentada en el documento. A continuación se muestran los resultados de los tres diferentes métodos de interpolación expuestos arriba:

1. Spline():

Respecto a `spline()`, como se muestra en la imagen 2, la representación de lo obtenido (curva azul) no contrasta en gran medida con los valores originales (curva negra); esto se debe al hecho de que el proceso intenta acercarse de manera cuadrática a cantidades recopiladas en forma aleatoria.

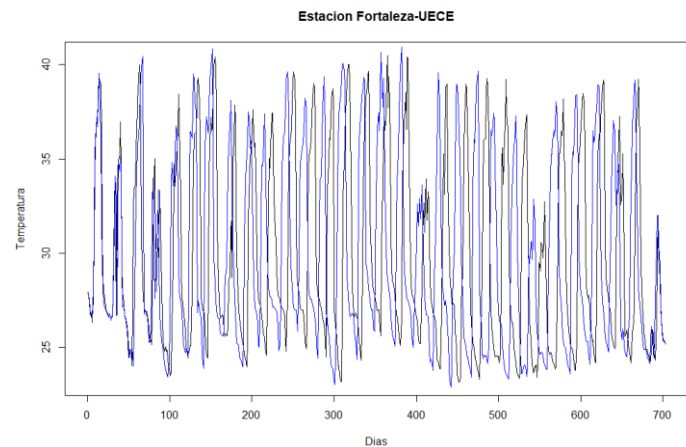


Imagen 2: Interpolación Spline

Se aplica el cálculo de errores y se determinan los valores expuestos en la imagen 3, esto permite entender la gran diferencia evidenciada en el proceso previo.

```
error EMC del spline 7.327524
> cat("error medio del spline ",(sum(espline))/tam,'\n')
error medio del spline 5.823641
> cat("error minimo del spline ",min(espline),'\n')
error minimo del spline 0
> cat("error maximo del spline ",max(espline),'\n')
error maximo del spline 16.14286
```

2. Splinefun()

Imagen 3: Error en Interpolación Spline

En el caso de `splinefun()` se puede observar en la imagen 4, una gran semejanza con los datos iniciales, en la mayoría de la trayectoria pasando sobre los mismos puntos. Siendo la mejor de las aproximaciones, lo demuestra con su bajo valor de error, en la imagen 5.

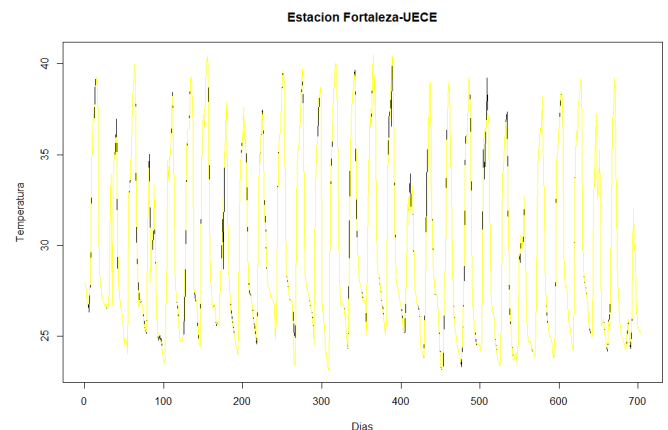


Imagen 4: Interpolación Splinefun

```
error EMC del splinefun(hermite) 0.4261986
> cat("error medio del splinefun(hermite) ",(sum(ec))/tam,'\n')
error medio del splinefun(hermite) 0.115004
> cat("error minimo del splinefun(hermite)",min(ec),'\n')
error minimo del splinefun(hermite) 0
> cat("error maximo del splinefun(hermite)",max(ec),'\n')
error maximo del splinefun(hermite) 4.929595
```

Imagen 5: Error en Interpolación Splinefun

3. Interpolación lineal:

Por último la **interpolación lineal** nos proporciona un rendimiento similar a **spline()**, sin embargo muestra una pequeña mejoría, imagen 6. Con un margen de error semejante, imagen 7.

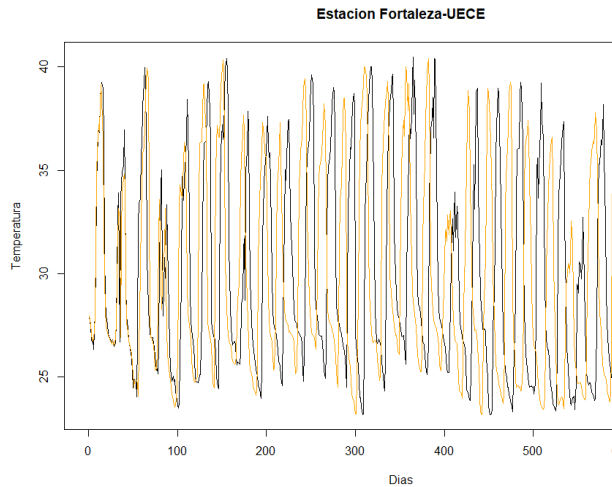


Imagen 6: Interpolación Lineal

```
error EMC de la lineal 7.262965
> cat("error medio de la lineal ",(sum(e1))/tam,'\n')
error medio de la lineal 5.761659
> cat("error minimo de la lineal ",min(e1),'\n')
error minimo de la lineal 0
> cat("error maximo de la lineal ",max(e1),'\n')
error maximo de la lineal 16.10393
```

Imagen 7: Error en Interpolación Lineal

En síntesis, lo mostrado por **slipnefun()** es el mejor resultado posible superando el criterio de aceptación con un valor del 99.518 %, lo cual da un comportamiento casi idéntico a lo expuesto en la imagen 1; por su parte **spline()** y la **interpolación lineal** no lograron llegar al objetivo, con 75.633% y 75.892% respectivamente.

IV. PROBLEMÁTICA 2

Con la información de las estaciones climáticas se realiza un gráfico con su ubicación respectiva, gracias a sus coordenadas en la imagen 8, tenemos su distribución en el territorio.

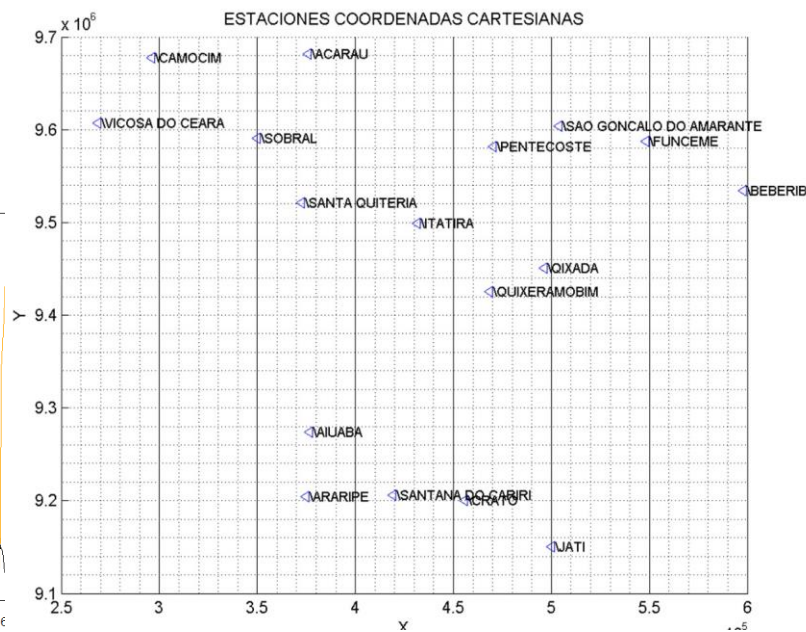


Imagen 8: localización de las estaciones

En este caso, para realizar el proceso se escoge como estación base **Acaraú** y se utilizan los mismos procedimientos anteriores con el fin de generar la estimación de la variable **temperatura interna (°C)** en la estación de Aiauba.

Al igual que antes se dará una selección aleatoria del 80% de los datos iniciales, lo cual va a producir mejores resultados.

- Gracias a **splines()** encontramos una aproximación muy cercana a los valores de la estación base, sin embargo su comportamiento está dirigido hacia puntos medios; se mantiene lejos de los extremos, como se muestra en la imagen 9. Con relación a su error, la imagen 10 expone un error cuadrático medio no muy alto.

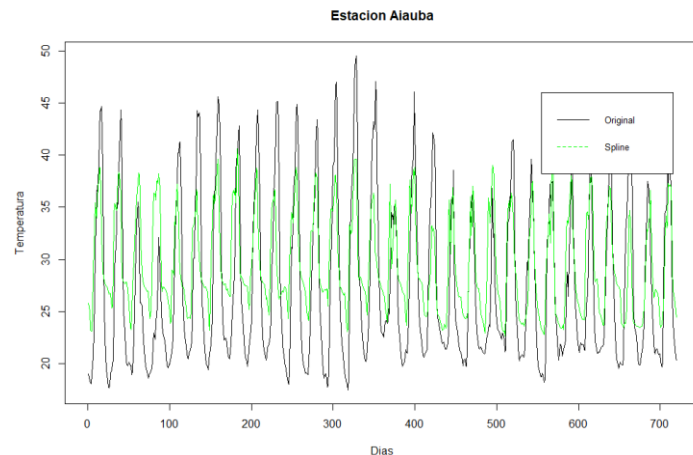


Imagen 9: Interpolación Spline 2

```

error EMC del spline 5.513738
> cat("error medio del spline ",(sum(espline))/tam,'\n')
error medio del spline 4.656548
> cat("error minimo del spline ",min(espline),'\n')
error minimo del spline 0.00789158
> cat("error maximo del spline ",max(espline),'\n')
error maximo del spline 17.38569

```

Imagen 10: Error en Interpolación Spline 2

- Con *splinefun()* conseguimos valores aún más cerca que en caso anterior, imagen 11. Su error es un poco menor, imagen 12, gracias al método utilizado.

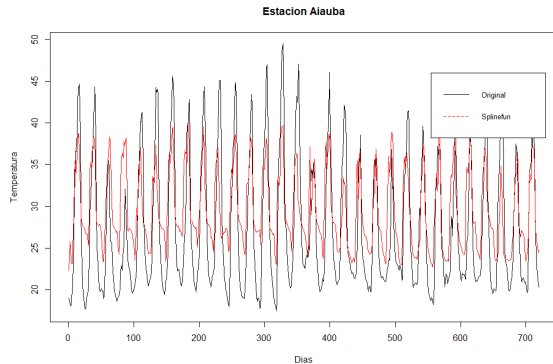


Imagen 11: Interpolación SplineFun 2

```

error EMC del splinefun(hermite) 5.363323
> cat("error medio del splinefun(hermite) ",(sum(ec))/tam,'\n')
error medio del splinefun(hermite) 4.522034
> cat("error minimo del splinefun(hermite) ",min(ec),'\n')
error minimo del splinefun(hermite) 0.005195383
> cat("error maximo del splinefun(hermite) ",max(ec),'\n')
error maximo del splinefun(hermite) 16.69

```

Imagen 12: Error en Interpolación SplineFun 2

- Para finalizar, la *interpolación lineal* arroja resultados muy parecidos a *splines()*, imagen 13. Con un error un poco mayor, imagen 14.

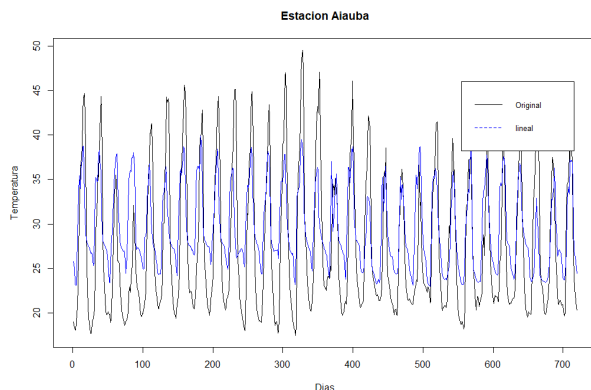


Imagen 12: Interpolación Lineal 2

```

error EMC del lineal 5.525794
> cat("error medio del lineal ",(sum(e1))/tam,'\n')
error medio del lineal 4.654935
> cat("error minimo del lineal ",min(e1),'\n')
error minimo del lineal 0.01057719
> cat("error maximo del lineal ",max(e1),'\n')
error maximo del lineal 17.29624

```

Imagen 13: Error en Interpolación lineal 2

En conclusión, para este ejercicio los tres métodos de interpolación exponen resultados muy similares, siendo la diferencia algo mínimo. Sin embargo ninguno de ellos superó el criterio de aceptación, teniendo como error cuadrático medio porcentajes no mayores al 85%, *splines*(83.099), *splinesfun*(83.587) e *interpolación lineal*(83.104).

V. CONCLUSIÓN

- Las interpolaciones tienen parámetros que con muy poco cambio nos generan una salida muy fluctuante entre si con aspectos tales como; la distancia o pequeños cambios de temperatura son muy notorios por que devuelven una lista de puntos que se comparan entre todos, entonces al ser por ejemplo, una spline cubica que utiliza un filtrado especial el rango de x llega a ser mucho mayor.
- Las splines cubicas se pueden construir en un espacio vectorial mientras que las splines de interpolación no. Esto ocasiona que, haya menos potencias truncadas se presente un mayor grado de oscilaciones que el polinomio consiga una regularidad global con mayor facilidad y con un dominio mucho menor y mejor establecido (mas certero)

REFERENCIAS

- [1] Burden, R. and Faires, J., 2005. *Numerical analysis*. Belmont, CA: Thomson Brooks/Cole, pp.106-171.
- [2] Rdocumentation.org. 2021. *splinefun function* - *RDocumentation*. [online] Available at: <<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/splinefun>> [Accessed 13 May 2021].
- [3] S. Iyengar and R. Jain, *Numerical Methods*. Daryaganj: New Age International Pvt. Ltd., Publishers, 2009.