

Reto 2b La interpolación y el clima

1st Juan José Bolaños Melo
Ingeniería de sistemas
Pontificia Universidad Javeriana
Cali, Colombia
bolanos.jj@javeriana.edu.co

2nd David Andres Duarte Clavijo
Ingeniería de sistemas
Pontificia Universidad Javeriana
Bogotá, Colombia
david-duartec@javeriana.edu.co

3rd David Santiago Savedra
Ingeniería de sistemas
Pontificia Universidad Javeriana
Bogotá, Colombia
david_saavedrar@javeriana.edu.co

Abstract—In the work we will develop the second problem for challenge 2 where a group of values associated with climatic variables of Santa Quiteria and Itatira in Brazil is not given, where a detailed analysis is carried out. The data is related in time and space, the dataset has several weather stations close to the proposed areas in Brazil. We are going to guide ourselves in the use of Spline functions (polynomials), in addition to linear regression in the course of the work, as well as the different errors that can be found in the interpolations that are going to be carried out..

Keywords—Error, spline, regression, interpolation, polynomial.

I. INTRODUCCIÓN

En el trabajo desarrollaremos el segundo problema para el reto 2 en donde se no da un grupo de valores asociados a variables climáticas de Santa Quiteria y de Itatira en el Brasil, en donde se procede al análisis detallado. Los datos están relacionados en el tiempo y en el espacio, el dataset posee varias estaciones climáticas cercanas a las zonas propuestas en Brasil. Nos vamos a encaminar en el uso de funciones Spline(polinomiales), además de regresión lineal en el transcurso del trabajo, así como también los diferentes errores que se pueden encontrar en las interpolaciones que se van a realizar.



Imagen 1. Clima

II. PROYECCIÓN DE DATOS

Como lo mencionamos anteriormente lo que buscamos con esta proyección de datos es predecir y adelantarnos al comportamiento, la forma y las magnitudes de algunas variables, esto gracias a técnicas descritas más adelante.

Y luego con base en estas poder hacer correcciones o si se desea replantear algunas soluciones.

Para la realización de esta investigación profunda nos brindan un documento en formato abierto .xls, el cual es un tipo de archivo de Microsoft Excel que nos facilita el manejo de datos mediante hojas de cálculo. Dentro de este documento encontramos 17 hojas de cálculo. cada una destinada a un municipio brasileño, dentro de estas hojas de cálculo encontraremos diferentes variables que abarcan factores climatológicos importantes relacionadas a estos municipios, como por ejemplo lo es la velocidad del viento, la dirección del viento , la precipitación y radiación solar. Entre otras variables también destaca el día en juliano, la hora y la temperatura en grados centígrados.

A Continuación en la imagen [2] se muestra el formato que se utilizó en cada una de las hojas de cálculo.

	A	B	C	D	E	F	G	H	I	J	K	L	M
	Año	Día Juliano	Hora	Temp. Interna (°C)	Presión Atmosférica (mmHg)	Temp. de Ar. (°C)	Velocidad del Viento (km/h)	Dirección del Viento (°)	Velocidad del Viento Máxima (km/h)	Dirección del Viento Máxima (°)	Precipitación (mm)	Radiación Solar Total (kJ/m²)	
1	2013	91	100	25.01	965.74	25.58	84.4	25.63	25.36	1.578	237.2	2.151	232.7
2	2013	91	200	24.88	964.99	25.44	85.1	25.56	25.32	1.229	223.6	2.532	219.3
3	2013	91	300	24.78	965.01	25.24	87.3	25.5	25.19	1.076	216.3	1.728	219.1
4	2013	91	400	24.67	964.89	25.11	74.5	25.51	25.11	2.085	94.7	4.391	69.86
5	2013	91	500	23.39	965.34	23.84	84.7	25.14	23.84	2.752	97.1	4.507	91.8
6	2013	91	600	22.97	965.82	24.28	79.5	24.38	23.67	3.511	93.6	5.263	91.5
7	2013	91	700	25.26	965.6	26.67	65.15	26.67	24.43	4.577	88.8	6.726	73.5
8	2013	91	800	30.02	965.97	27.79	61.6	27.98	26.76	6.18	89.2	8.14	91.4
9	2013	91	900	32.04	966.26	29.16	56.43	29.17	27.62	6.443	108	8.61	95
10	2013	91	1000	32.27	965.84	30.28	52.9	30.33	27.88	5.021	114.8	7.47	83.9
11	2013	91	1100	35.43	965.52	31.08	55.96	31.8	29.93	4.585	108.4	6.15	89.1
12	2013	91	1200	36.48	963.95	32.48	44.23	31.11	30.93	3.899	112.6	6.586	83.6
13	2013	91	1300	39.77	962.84	33.91	39.58	34.27	32.31	3.949	98.2	7.02	94.3
14	2013	91	1400	42.85	961.97	34.63	36.85	35.58	33.23	3.625	90.6	5.945	122
15	2013	91	1500	42.1	961.76	34.57	36.16	36.17	33.55	3.804	114.6	6.924	75.8
16	2013	91	1600	40.78	961.51	34.07	36.47	35.08	33.77	3.38	118.7	6.009	101.6
17	2013	91	1700	36.6	961.84	33.33	31.72	34.42	33.28	3.803	121	5.303	125.7
18	2013	91	1800	33.1	962.55	32.24	40.85	33.3	32.24	3.814	120.2	5.531	100
19	2013	91	1900	30.91	963.13	31.05	44.13	32.22	31.05	3.028	93.3	4.547	105.3
20	2013	91	2000	29.36	963.71	29.71	48.74	31.08	29.71	2.984	105.3	4.616	101
21	2013	91	2100	27.68	964.16	27.78	61.11	29.68	27.78	2.186	82.2	3.223	63.47
22	2013	91	2200	26.49	964.8	26.74	71	27.95	26.74	2.075	86.1	3.707	80.6
23	2013	91	2300	25.27	964.88	25.64	75.8	26.69	25.64	2.687	93.3	4.249	110
24	2013	91	2400	24.01	964.19	24.35	78.8	26.62	24.35	1.538	62.31	2.678	81.6
25	2013	92	100	22.93	964.4	24.51	69.9	24.6	23.17	1.646	90.6	3.923	105.6
26	2013	92	200	22.93	964.06	24.31	82.8	24.51	24.1	1.673	66.97	2.981	84.8
27	2013	92	300	23.12	964.27	24.69	82.9	24.69	24.1	1.121	36.57	2.527	129.2
28	2013	92	400	22.69	963.69	23.65	67.6	24.78	23.65	2.366	99.4	3.414	87.2
29	2013	92	500	22.03	964.22	22.93	90	23.63	22.91	1.349	70.5	2.063	69.28
30	2013	92	600	21.75	964.77	23.15	86.9	23.43	22.89	1.904	106	3.208	105.3
31	2013	92	700	26.33	965.28	25.84	75.8	25.9	23.17	1.095	77.1	1.813	119.4
32	2013	92	800	30.11	965.08	26.48	72.7	27.92	25.74	4.325	97	7.43	100.9
33	2013	92	900	27.06	965.71	26.06	61.7	26.78	25.53	4.488	111.7	7.06	109.3
34	2013	92	1000	31.62	965.16	29.61	59.84	30.13	26.3	4.32	113	6.679	128
35	2013	92	1100	34.27	964.46	30.66	55.42	31.36	29.75	4.693	123.7	7.56	124.8
36	2013	92	1200	35.67	963.47	33.08	45.58	33.35	30.62	4.151	121.2	6.874	102.2
37	2013	92	1300	38.44	962.34	32.96	44.57	33.98	32.54	4.167	115.2	6.185	95.9
38	2013	92	1400	41.2	961.21	34.08	44.7	34.7	32.89	4.515	116.7	5.57	106.7
39	2013	92	1500	42.3	960.8	34.08	44.7	34.7	32.89	4.515	116.7	5.57	106.7

Imagen 2. Formato .xls con variables climatológicas

a. Lectura y selección de datos:

Para llevar a cabo la lectura y selección de datos se hizo uso del archivo .xls y en especial se trabajó sobre los dos municipios mencionados en el enunciado. La estación finalmente elegida fue.

Itatira:

Municipio pertenecientes al país de Brasil y ubicado específicamente en el estado de Ceara. Este cuenta al día de hoy con una población de aproximadamente 21.836 habitantes. [4]



Imagen 3. Ubicación geográfica Itatira [4]

En relación a la hoja de cálculo perteneciente a este municipio, encontramos que posee un total de 720 datos, datos que fueron extraídos en el año 2013, dentro de este municipio encontramos variables importantes que fueron utilizadas como lo son:

- Dia Juliano: esta variable lleva el número de días desde el inicio del año, este método de recuento de días es principalmente usado en temas relacionados con la astronomía y software.
- Hora: para esta variable se hace uso del formato de convección del tiempo empleado en ámbitos militares. cabe aclarar que este formato no hace uso de ningún signo de dos puntos. por ende la hora 17:00 es equivalente a 1700.

- Temp. Interna (°C): Esta variable será llevada a cabo con referencia a la medida termométrica Celsius (°C). y se implementa con el fin de llevar un registro de la temperatura del aire, indicando el calentamiento o enfriamiento de este. todo esto como resultado del intercambio de calor entre la atmósfera y la tierra.

III. INTERPOLACIÓN

Con el fin de realizar una correcta interpolación de los diferentes datos recolectados primeramente de la estación de Itetira, se tomaron en cuenta el conjunto de datos que representaban los días (30 días en total desde el Día Juliano 91 al 120) y las temperaturas obtenidas durante las 24 horas del día con el objetivo de crear una variable ideal, a la cual llamamos “Indicador”, este índice o variable depende de la introducción de los datos de entrada, debido a que Itatira nos brinda 720 datos diferentes, el rango de este índice se estableció en un rango de 1 hasta 720. Después de establecer este índice, el siguiente paso fue obtener la gráfica correspondiente:

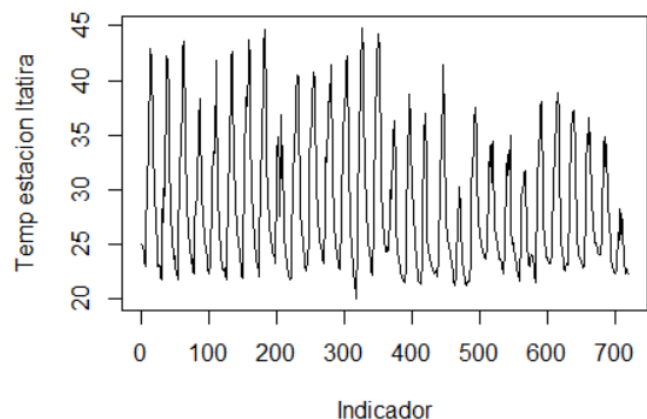


Imagen 4. Temperaturas estación de Itatira

Teniendo en cuenta la gráfica y sus datos originales, además de su respectivo rango, el siguiente paso fue tomar una muestra del 80% de las temperaturas, es decir se removi6 en total un 20% de los datos recolectados; con esta muestra anteriormente recolectada y el rango establecido se logró obtener el modelo de interpolación, este modelo tiene como función calcular las temperaturas que faltan en el modelo original, el cual en este caso fue Itatira.

Para realizar este modelo se procedió a realizar una serie de pasos, en primer lugar, se utilizó la función `Splinefun()` de R, esta función nos devuelve una lista donde se encuentra diferentes puntos resultados de una interpolación, esto, a partir de un conjunto de datos previamente dado.

Para hacer uso de esta función se tuvieron que adaptar los datos, esto se hizo mediante el uso de un arreglo, este arreglo contaba con un tamaño igual al tamaño exacto de los datos y en el cual se introdujeron unos y ceros, los unos representan un dato que estaba incluido en la muestra y los ceros un dato que fue removido de la muestra, luego de esto y haciendo uso de este arreglo, se hizo una comparación con los datos originales con el fin de descartar los datos removidos. Nuevamente, con el uso del anterior arreglo y de la muestra, se crearon dos arreglos que representan las coordenadas X y Y, en el arreglo `CoordenadaX` se introdujeron los índices y en el arreglo `CoordenadaY` se pusieron las diferentes temperaturas, siendo estos datos pertenecientes a la muestra. Con estas coordenadas se generó una gráfica donde la línea azul representa los datos interpolados y la negra los datos originales. La gráfica es el siguiente:

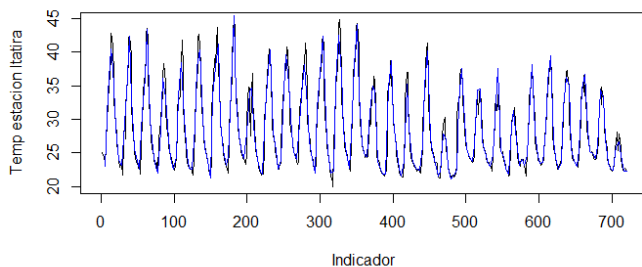


Imagen 5. Resultado interpolación 1

IV. ERROR DE MODELO

Para terminar este punto del segundo reto, se obtuvo el error absoluto de la interpolación en comparación de los datos generales. El valor absoluto se define como la diferencia entre el valor real y el valor aproximado:

$$e = |\text{Valor real} - \text{Valor aproximado}|$$

Imagen 4. Fórmula del error absoluto

Donde:

- El valor real es el dato que mide la magnitud a medir.
- El valor aproximado es la media de diversas medias.

Al tener una gran cantidad de datos como lo son los 720 que se utilizaron para obtener el error, se calculó el promedio de cada uno de los valores que se extrajeron y se obtuvieron datos de la media, el mínimo y el máximo error.

Error absoluto obtenido = NA.

Error máximo: Es el cálculo de un intervalo de confianza para la diferencia de medias.

$$E = z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}$$

Imagen 6. Fórmula del error máximo

Error máximo obtenido = 0.19355933

Error mínimo obtenido = 0

Error media:

$$SE = \frac{\sigma}{\sqrt{N}}$$

Imagen 7. Fórmula con la desviación estándar de la población

Se usa esta fórmula si se conoce la desviación estándar de la población.

$$SE = \frac{s}{\sqrt{N}}$$

Imagen 8. Fórmula con la desviación estándar de la muestra

Y se usa esta forma si cuenta con la desviación estándar de la muestra.

En donde cada uno de esos datos es:

- **SE:** El error estándar.
- σ : La desviación estándar de la población.
- **s:** Desviación estándar de la muestra.
- **N:** Número de observación de la muestra.

Error media obtenido = 0.003997017

V. PROYECCIÓN DE DATOS A PARTIR DE OTRO ESTACION

Para llevar a cabo la segunda parte de nuestra investigación se procede a simular aproximadamente los datos de una estación y la otra otra.

Para nuestro ejemplo y siguiendo con los pasos descritos en el enunciado, se hace uso de la información pertinente de los municipios de Santa Quiteria y Itatira.

Para nuestro ejemplo vamos a estimar la información de la estación de Santa Quiteria para generar la simulación a partir de los datos de la estación de Itaquera.

b. Interpolación:

Ahora, el siguiente paso está relacionado con la estación de Santa Quiteria, para poder generar los datos de esta estación se utilizaron los índices encontrados al inicio mediante el uso de los datos de la estación de Itatira; estos índices nos sirvieron dado que representan los días en los que se tomaron los datos que hicieron parte de la muestra original. Es necesario comentar que entre la cantidad de datos que tiene Itatira y Santa Quiteria hay una gran diferencia, ya que Itatira cuenta con un total de 720 datos y Santa Quiteria tiene un total de 297 datos.

Debido a esta diferencia lo que se hizo fue tomar el arreglo inicial de índices con el fin de verificar si se encontraban datos de Santa Quiteria de acuerdo con los días y las horas, luego, a estos datos se les asignó un índice que fuera adecuado para poder graficar su comportamiento, la gráfica que resultó de los datos originales fue la siguiente:

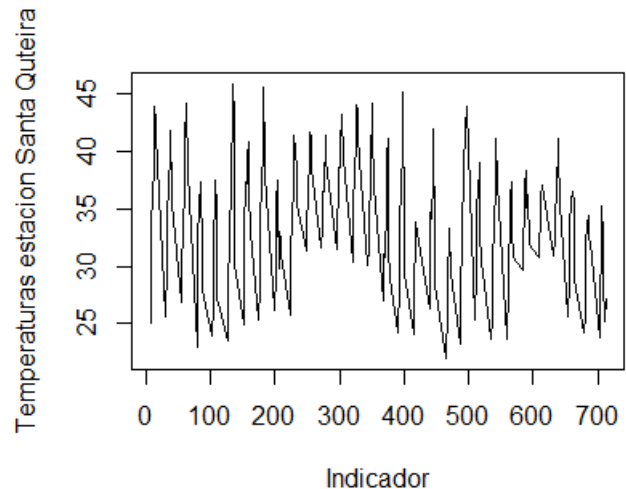


Imagen 9. Temperaturas estación de Santa Quiteria

Para finalizar, se repitió el procedimiento para generar los datos de la estación Santa Quiteria partiendo de los datos de la estación de Itatira, se realizó una gráfica donde se reflejan dos líneas, una morada y una negra, la negra son los datos originales y la morada representa la interpolación:

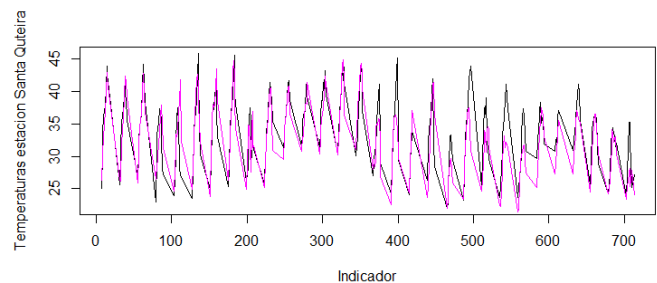


Imagen 10. Resultado interpolación 2

Para concluir esta sección de resultados se hizo uso de una función en R llamada `qqnorm()`, la cual nos brinda un análisis que consiste en saber cuál fue la precisión con la que se realizó nuestro modelo en comparación con una media real; en la gráfica se muestra diferentes puntos que representan la precisión de esto, entre más cercanos sean los puntos, representa mayor precisión, la gráfica muestra el resultado de la simulación del modelo, mostrada a continuación:

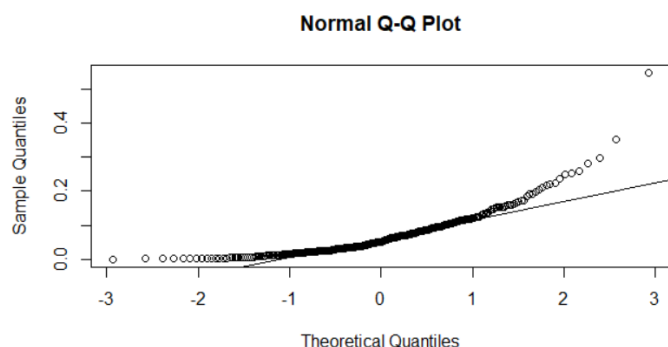


Imagen 11. Análisis función qqnorm

c. Error del modelo:

Para finalizar con la segunda parte del reto, se realizó el cálculo del error absoluto de la interpolación en comparación con los datos de entrada, con la idea de comprar lo que se obtuvo en la simulación realizada.

Los datos obtenidos son:

```
[1] 0.0483033932 0.0186335404 0.0529116169 0.1018647370 0.0879758935 0.0284953395
[7] 0.0204433498 0.0223591148 0.0105757932 0.0166385339 0.0257109466 0.0009973404
[13] 0.0205214232 0.0887608069 0.1032294650 0.0343800758 0.0065491184 0.0112359551
[19] 0.0703902066 0.0627883137 0.0973003375 0.0890868529 0.0647333926 0.0052566481
[25] 0.0217246888 0.0250330148 0.0013732491 0.0200881921 0.0296984811 0.0400104603
[31] 0.1864764785 0.0367647059 0.1542874396 0.1382605441 0.0812660393 0.0950335570
[37] 0.0087312415 0.0598483411 0.1313223644 0.2286185584 0.1429616088 0.0398038226
[43] 0.0554507733 0.0226810904 0.0159263005 0.0783384444 0.1056845476 0.0488497586
[49] 0.3508277563 0.5464642725 0.2026022305 0.0714506435 0.0333333333 0.0918554807
[55] 0.0223792697 0.0050223214 0.0209352909 0.0832259727 0.0009376465 0.1098013534
[61] 0.1230232558 0.0026170398 0.0614469772 0.0653530991 0.0853242321 0.0272407733
[67] 0.0238591054 0.0220134228 0.0472991131 0.0470053071 0.0719194894 0.1533406353
[73] 0.1650485437 0.0385978732 0.0120141343 0.0706473544 0.0253383242 0.0030136986
[79] 0.0427929207 0.1045501022 0.0320855471 0.0358084359 0.0135689093 0.1001314060
[85] 0.0350567040 0.0703296703 0.1729748850 0.0635308491 0.0280269058 0.1595092025
[91] 0.2210913203 0.1540013450 0.1412563288 0.0999240168 0.0161030596 0.0461912956
[97] 0.0515924685 0.0263750402 0.0399533392 0.0014667058 0.0237452779 0.0177035913
[103] 0.0232333011 0.0320947112 0.0173803526 0.0822130300 0.1124266612 0.0643514156
[109] 0.0057979860 0.0136970247 0.0109581343 0.0147219193 0.0277990309 0.0155679883
[115] 0.0024413692 0.0274538387 0.0691889090 0.0262991128 0.0104916067 0.0729968634
[121] 0.0817334065 0.0180970661 0.0067654977 0.0707144599 0.0208658284 0.0569167943
[127] 0.0378017789 0.0349733254 0.0527655484 0.0756672713 0.0502720912 0.0885430041
[133] 0.0247892910 0.0356729210 0.0008751708 0.0383054287 0.0059269015 0.0261360261
[139] 0.0430226145 0.0209564750 0.0060526316 0.0161725067 0.0025005683 0.0776069198
[145] 0.0337976707 0.1163468545 0.0205067396 0.0967536601 0.1152213463 0.0968112611
[151] 0.0578444748 0.0444058546 0.0330110823 0.0031746032 0.0205103744 0.0934452871
[157] 0.0408163265 0.0171998624 0.0215503377 0.0625802311 0.1212613784 0.0090625885
```

Imagen 12. Error absoluto

Además se desarrolló el cálculo del mínimo, máximo y la media de los resultados conseguidos.

Error mínimo obtenido = 0.0009376465

Error máximo obtenido = 0.5464643

Error medio obtenido = 0.0619291

REFERENCIAS

[1]. Error absoluto y error relativo: Qué son y cómo se calculan . [En línea]. Disponible en: <https://ekuatío.com/error-absolutos-y-error-relativos-que-son-y-como-se-calculan/>. [Último acceso: 7 05 2021]

[2]. El error estándar y su interpretación. [En línea]. Disponible en: <https://bookdown.org/dietrichson/metodos-cuantitativos/el-error-estandar-y-su-interpretacion.html> [Último acceso: 7 05 2021]

[3]. Error y tamaño de la muestra. [En línea]. Disponible en: http://e-ducativa.catedu.es/44700165/aula/archivos/repositorio/4750/4754/html/31_error_y_tamao_de_la_muestra.html#:~:text=El%20error%20m%C3%A1ximo%20que%20se,la%20amplitud%20del%20intervalo%2C%20E.&text=El%20Error%20m%C3%A1ximo%20admisible%20en,la%20amplitud%20del%20intervalo%2C%20E. [Último acceso: 7 05 2021]

[4]. Location itatira. [En línea]. Disponible en: <https://en.wikipedia.org/wiki/Itatira> [Último acceso: 7 05 2021]