

Reto II

Juan Jose Camacho

j_camacho@javeriana.edu.co

Gabriela Maria Camacho

gabriela.m.camacho@javeriana.edu.co

20 de abril de 2020

1. Proyección de datos

Para la realización de este punto se nos entrego la información de 15 diferentes estaciones climáticas cercanas la zona de Fortaleza en Brasil, este documento contiene el año en que se realizo esta medición, el día juliano, la hora, la temperatura en grados centígrados, entre otros, como presión atmosférica y precipitación meteorológica.

■ Pasos que se llevaron acabo:

1. Lectura y selección de datos:

Para llevar acabo exitosamente este reto se selecciono a la estación del Itatira, un municipio de Brasil, este contiene 720 datos todos tomados en el 2013, se selecciono este debido a que es la estación con los datos mas completos, a partir de esto se tomaron las columnas a evaluar:

- **Día Juliano:** Es una manera de referirse a los días del año, en este se toma numero de días que han pasado desde el inicio de año siendo este el número 1 de 365.
- **Hora:** Se medirán las horas por medio del formato militar siendo 100 la primera hora del día y 24 la ultima hora del día.
- **Temperatura:** Esta es medida en Celsius ($^{\circ}\text{C}$) y hace referencia a la temperatura del aire.

2. Interpolación de los datos:

Para realizar la interpolación adecuada a partir de los datos de la estación de Itatira, se tuvieron en cuenta los "Índices ideales."^{es} decir, ya que esta estación cuenta con un conjunto de datos completos, con 30 días y 24 horas de temperaturas obtenidas, se procedió a asignar un "Índice ideal."^{esto} a partir del numero de entradas. Por lo que un índice ideal abarca un rango de 1 hasta 720, en el caso de la estación Itatira. Por lo que después de tener los índices correctos, se procedió a obtener la siguiente gráfica de datos originales:

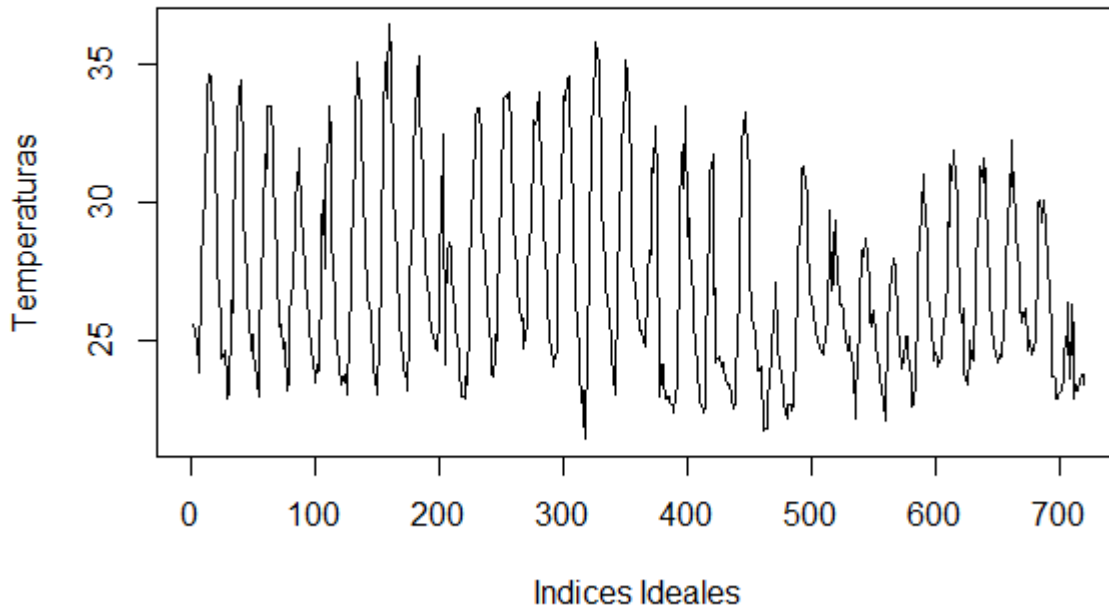


Figura 1: Temperaturas de Itatira

A partir de los datos originales, haciendo uso de los índices se procede a remover un 20 % de estos datos, por lo que se tomo una muestra del 80 % de las temperaturas originales, a partir de esta muestra con ayuda de los índices ideales se pudo obtener el modelo de interpolación para satisfacer la necesidad de calcular las temperaturas faltantes del modelo original.

Para poder hacer este modelo se hizo uso de la función *splinefun()* de R, la cual provee de un modelo de interpolación con los datos a ser graficados. Para adaptar los datos y hacer uso de ellos en la función, se procedió a llenar un arreglo de tamaño de los datos originales con 1s y 0s, de tal manera que los 1s son aquellos

datos que están presentes en la muestra y 0s los que han sido eliminados, para así hacer una comparación con los datos originales esto con el fin de llenar dos nuevos arreglos, que serian las nuevas coordenadas X y Y.

Con las nuevas coordenadas se genero el siguiente gráfico, siendo así la linea negra los datos originales, y la roja los datos interpolados;

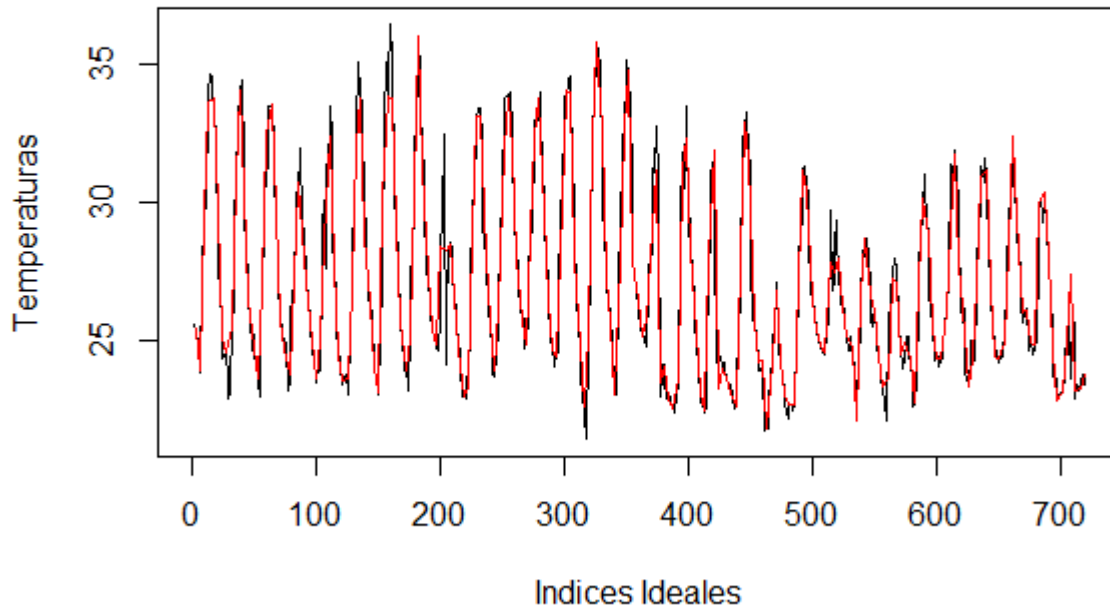


Figura 2: Interpolación resultante

3. Error del modelo:

Para finalizar esta parte del reto, se tomo el dato del error absoluto de la interpolación en comparación a los datos generales, debido a que se tienen 720 datos para obtener el error se mostrara la media, el máximo y el mínimo error que se hallo al momento de comparar los datos, también para tener una referencia del error se seleccionaron los 5 primeros datos y los últimos 5 datos dados en la siguiente tabla.

Temperatura Original	Temperatura Interpolada	E.Absoluto
25.44	26.05629	0.02422521
25.24	25.47172	0.009180627
25.11	24.50629	0.02404264
29.16	28.66825	0.01686392
30.28	29.75502	0.0173375
22.95	22.78287	0.007282302
23.24	23.21329	0.001149279
23.28	23.3678	0.003771439
25.29	23.26343	0.08013326
26.33	24.91969	0.05356287

Cuadro 1: Tabla de errores absolutos

- **Error Mínimo:** 0
- **Error Máximo:** 0.1308487
- **Error Media:** 0.003900002

2. Proyección de datos a raíz de otra estación

Al igual que con el punto anterior se tiene la misma información. Pero en este caso el objetivo final es diferente a raíz de una estación del clima se obtendrá la información de otra.

- Pasos que se llevaron a cabo:

1. Lectura y selección de datos:

Con el fin de simular aproximadamente el segundo punto, se seleccionó la estación de Santa Quiteria, para generar la simulación a partir de los datos de la estación de Itaquira. La principal razón por la cual se seleccionó es la cercanía que estas dos manejan, como se evidencia en la imagen a continuación.

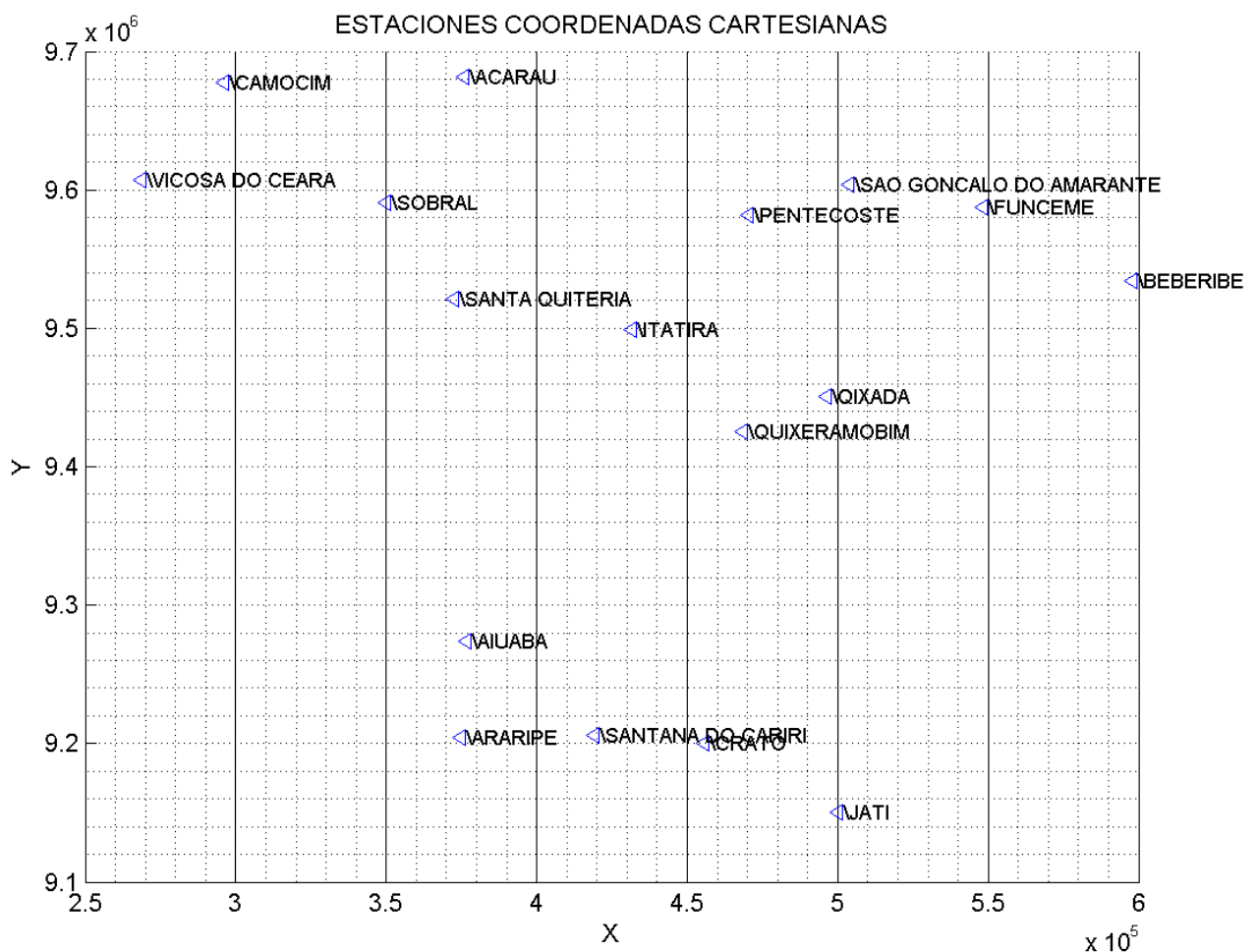


Figura 3: Distancia entre estaciones

2. Interpolación de datos:

Para poder generar los datos de la estación de Santa Quiteria, se hizo uso de los índices ideales, ya que estos índices corresponden a los días completos en los cuales se tomaron los datos para así, poder tener una idea de como se podrían ver las entradas originales de la estación, cabe resaltar que las temperaturas obtenidas en la información de la estación, no están completos, ya que Itaquira cuenta con 720 mientras que la Santa Quiteria cuenta solamente con 297.

Por lo que se tomo el arreglo de índices ideales, para así comparar los días y las horas en los cuales si se encontraban datos de la estación de Santa Quiteria, y asignarles a estos datos un índice que pudiera ser adecuado para poder graficar su comportamiento, la gráfica de los datos originales es:

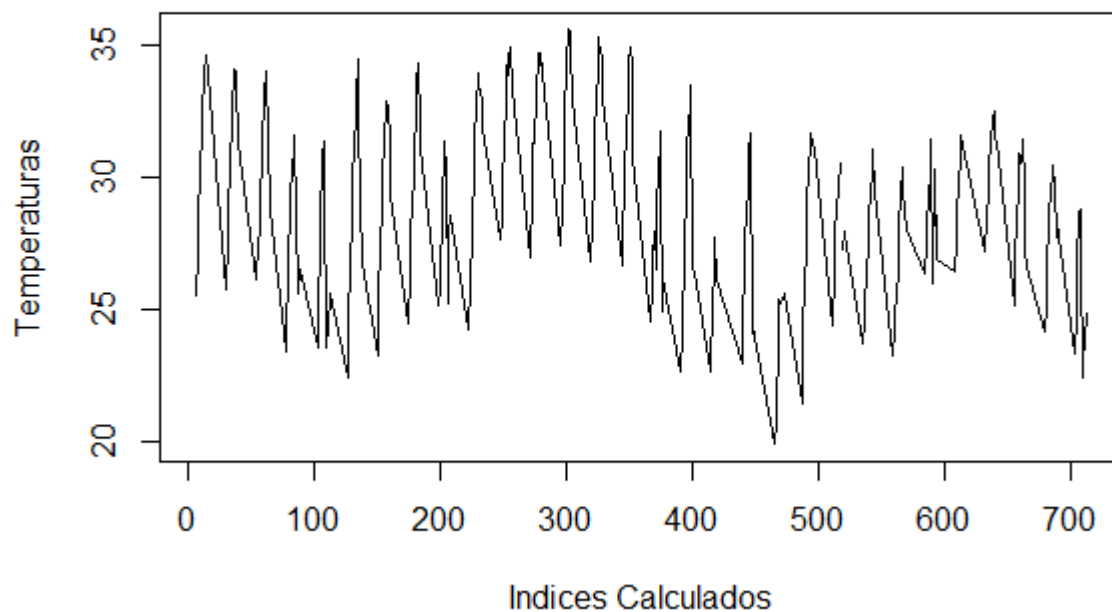


Figura 4: Temperatura Santa Quiteria

Por otra parte se hizo el mismo procedimiento para generar los datos de Santa Quiteria a partir de los datos de Itaquira, la siguiente gráfica refleja con la linea verde la interpolación y la linea negra con los datos reales:

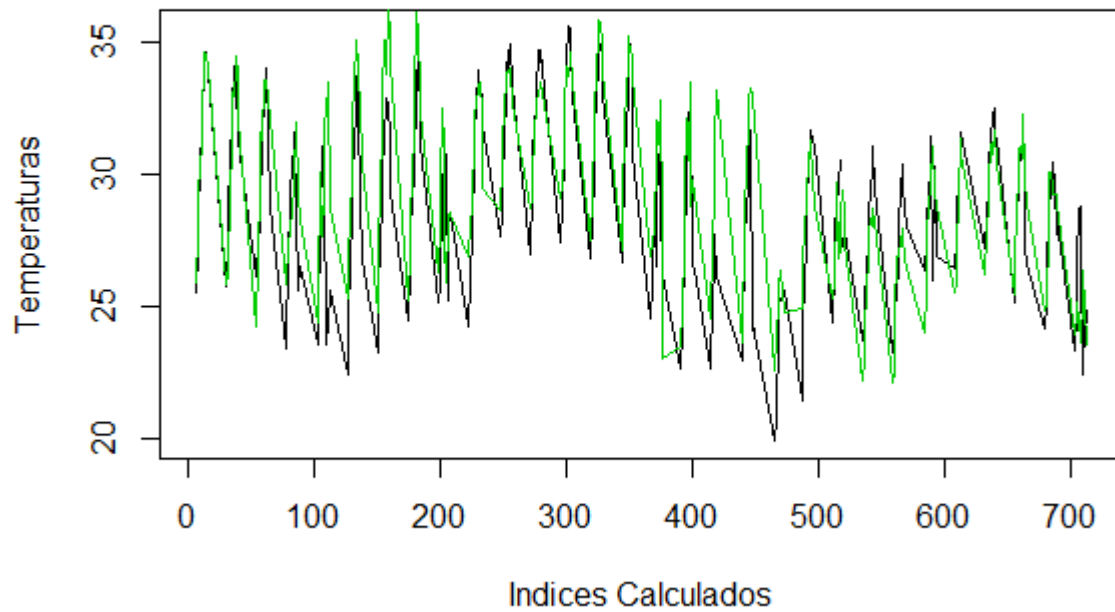


Figura 5: Interpolación generada

Para concluir esta parte de resultados hicimos uso de la función `qqnorm()`, esta función nos da un análisis de que tan preciso fue nuestro modelo en comparación a una media real, por lo que entre mas cerca estén los puntos a la recta de media el modelo sera mas preciso , por lo que la gráfica a continuación, muestra el resultado de la simulación del modelo de interpolación:

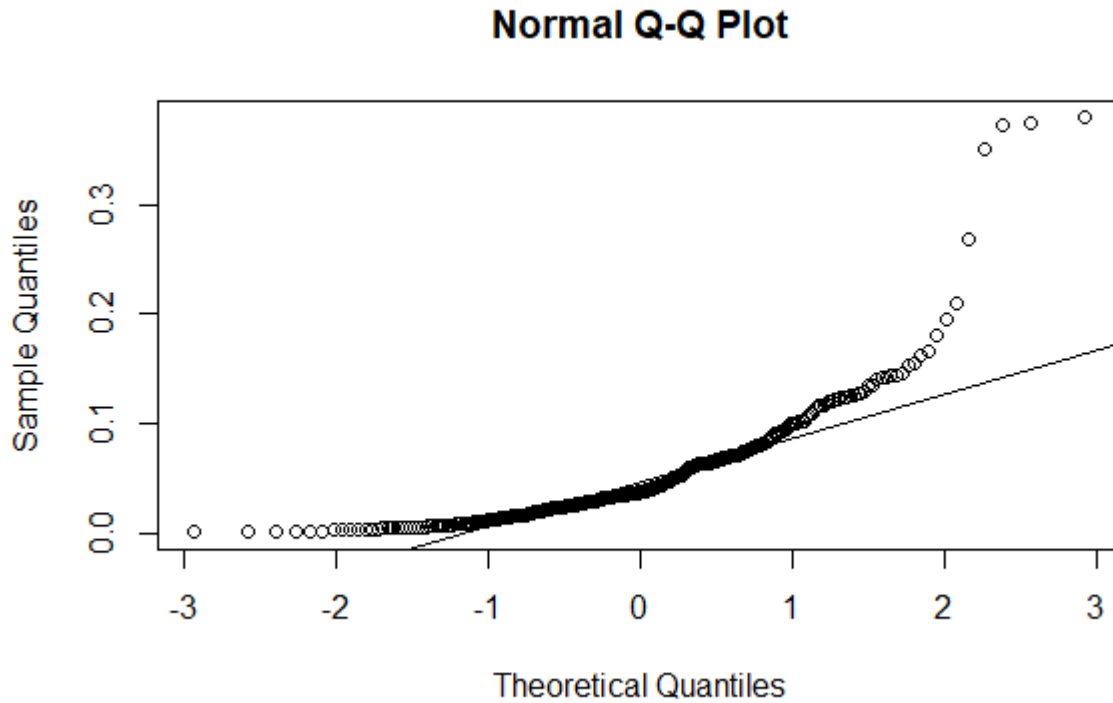


Figura 6: Análisis qqnorm

3. Error del modelo:

Para terminar esta parte del reto, se calculo el dato del error absoluto de la interpolación en comparación a los datos de entrada, esto con el fin de comparar la simulación obtenida, los datos que se obtuvieron son los siguientes:

[1]	0.0450626959	0.0288781933	0.0195804196	0.0103436770	0.0299625468	0.0058157331	0.0074272133	0.0197290931	0.0008670520	0.0046603920	0.0038850039
[12]	0.0406091371	0.0843288826	0.1267212819	0.1035230561	0.0470514970	0.0334310850	0.0029429076	0.0350276302	0.0678233438	0.0236346215	0.0317278287
[23]	0.0253531329	0.0397137746	0.0064428620	0.0261360729	0.0044958253	0.0129909366	0.0146886016	0.1159319681	0.1027718550	0.0155038760	0.0533895328
[34]	0.0516220513	0.0683453237	0.0340577360	0.0202467574	0.0994727592	0.1155726257	0.2009939311	0.0966767372	0.0713982150	0.0281800391	0.0797182054
[45]	0.0716066975	0.0173175685	0.0149729213	0.0702576112	0.3746813934	0.3790123457	0.1214843750	0.1299107143	0.0917618271	0.0628742515	0.0581724211
[56]	0.0592543276	0.0844072165	0.0486603018	0.0656136087	0.0031748335	0.0333230484	0.1271040879	0.1354790419	0.0806027010	0.0860299921	0.0686274510
[67]	0.1016416346	0.0995339547	0.0749165269	0.0285975053	0.1171060699	0.1529874866	0.1655290102	0.0293757650	0.0261010556	0.0010593220	0.0410382322
[78]	0.0232939633	0.0758400000	0.1074966724	0.0826444682	0.0306122449	0.0381369178	0.0378933847	0.0274244833	0.0905222437	0.1899973995	0.1173030616
[89]	0.0035495321	0.2316024212	0.1431404959	0.1176237624	0.0318103834	0.0063380282	0.0166020221	0.0622937294	0.0827241247	0.0323974082	0.0266106443
[100]	0.0351170569	0.0193466540	0.0137698898	0.0191345305	0.0023887728	0.0042029421	0.0139902676	0.0755959849	0.0350686913	0.0632022472	0.0579758713
[111]	0.0515129565	0.0193411907	0.0095392962	0.0014762327	0.0179554011	0.0266246779	0.0215956809	0.0660237389	0.0129207752	0.046453674	0.0079846694
[122]	0.0192438211	0.0084592145	0.0467128028	0.0139248819	0.0184597635	0.0522088353	0.0232074818	0.0075335735	0.0079926222	0.0076515597	0.0046756283
[133]	0.0511811024	0.0276134122	0.0118460872	0.0167071689	0.0624880276	0.0843036427	0.0605148780	0.0214578731	0.0230922380	0.0317367029	0.0101127448
[144]	0.0227731335	0.0098209128	0.0257575758	0.0316767705	0.0313601128	0.0810996564	0.0515665796	0.0448473282	0.0363691013	0.0238510762	0.0077363897
[155]	0.0078763127	0.0919727007	0.0954323002	0.0633255633	0.0482374768	0.0057245081	0.1544561934	0.0932377049	0.0425460637	0.0062407929	0.0262720319
[166]	0.0677259742	0.1210385643	0.0393283252	0.0369908562	0.0014609204	0.0876809512	0.1108921933	0.0321336761	0.0441176471	0.0238386308	0.0275647193
[177]	0.123773849	0.1059803179	0.0529456062	0.0427138058	0.1001511716	0.1235590778	0.1952617501	0.0309503051	0.0064231232	0.0244461421	0.0323392023
[188]	0.0612460401	0.0734160241	0.0580645161	0.0514131195	0.3732335827	0.3099630996	0.1346057258	0.1536974384	0.1269177127	0.0603595890	0.0129146249
[199]	0.0443564356	0.0340910541	0.1486540879	0.0998349835	0.0726364335	0.0644092749	0.0208768267	0.0183066362	0.0109748225	0.0164089618	0.0736808237
[210]	0.0356887176	0.0167109761	0.0531069364	0.0282661151	0.0373342638	0.0982172889	0.1095081967	NA	0.0778267254	0.0354077253	0.0036813214
[221]	0.0632644454	0.0628314973	0.0434954527	0.0243531202	0.0279289082	0.0205767931	0.0380434783	0.0699834163	0.0738233398	0.0404211957	0.0494623656
[232]	0.0690082645	0.0071856287	0.0674335581	0.0033532042	0.0309278351	0.0643952300	0.0805656034	0.0643631436	0.0663424800	0.0566844920	0.0909435392
[243]	0.0698110411	0.0349895031	0.0013807387	0.0357611549	0.0527821940	0.1961464355	0.0300131926	0.0703386677	0.0034940913	0.0315192399	0.0155871147
[254]	0.0125296309	0.0145489816	0.0240887480	0.0363569592	0.0252720253	0.0153165419	0.0038498764	0.0035256410	0.0082828926	0.0438024231	0.0273593606
[265]	0.0001941757	0.0396638439	0.0115354018	0.0287373090	0.0503459604	0.0311126141	0.0022624434	0.0180268764	0.0157455013	0.0260648442	0.0110849849
[276]	0.0942637925	0.0881024096	0.0307116971	0.0011942675	0.0361491629	0.0366236484	0.0293468680	0.0210383441	0.0091736410	0.0396981627	0.0141090501
[287]	0.0551948052	0.0187146893	0.0115532734	0.0158086745	0.0608663181	0.1425078589	0.0846634282	0.1247224278	0.1265033408	0.1452805568	0.0566948130

Figura 7: Error absoluto obtenido

Si hay casillas donde aparece NA esto hace referencia a valores no disponibles que viene directamente desde el documento recibido

Adicionalmente también se calculo el máximo, el mínima y la media, los resultados obtenidos fueron:

- **Error Mínimo:** 0.000867052
- **Error Máximo:** 0.3746814
- **Error Media:** 0.0411477

3. Conclusiones

Finalmente al realizar ambos parte podemos concluir que se logro llevara acabo el ejercicio gracias los métodos de interpolación elegido el spline y splinefun fueron acertados, esto lo comprobamos gracias a Q-Q Plot pues una gran parte los resultados se encuentran sobre la linea deseada. Sin embargo, no podemos ignorar que si se tiene error, sobre todo en el punto que nos pide encontrar los datos a partir de otra estación esto era esperado por la diferencia en cantidad de información que se tenia entre las ciudades, es decir de 423 datos y la distancia que se tiene entre estas.