de

A la vista de estos resultados, se podría considerar un modelo sencillo para la descripción de los datos; pues se observa que cuando la temperatura está entre 0 y 75 los estimadores no paramétricos de la regresión se comportan de forma más o menos lineal con poca pendiente; entre los valores de temperatura 75 y 91 los estimadores se comportan de una forma más o menos lineal pero con mucha más pendiente; finalmente, cuando la temperatura toma valores entre 91 y 96 los estimadores presentan un comportamiento constante. En este último tramo cabe notar que hemos eliminado el último valor de la temperatura, 97, pues provoca un efecto frontera. Por todo esto, hemos probado estimar la función de regresión mediante la regresión lineal simple a trozos.



Variables que se usan en cada tramo de la regresión lineal a trozos.

|  |  |
| --- | --- |
| x | 57 58 59 59 61 61 61 62 62 63 64 64 65 65 66 66 66 67 67 67 68 68 68 68 69 69 70 71 71 71 72 72 72 73 73 73 73 74 74 74 75 75 |
| y |  6 18 19 1 8 4 32 18 11 14 14 13 23 20 28 11 34 41 23 18 14 30 24 20 16 7 30 9 13 23 36 37 9 11 12 10 24 12 7 16 21 14 |
| x | 75 76 76 76 76 76 76 77 77 77 77 78 78 78 78 79 79 79 80 80 80 81 81 81 81 81 81 81 81 81 81 82 82 82 82 82 82 82 83 83 83 84 84 84 85 85 85 85 86 86 86 86 86 86 87 87 87 87 88 88 89 89 90 90 90 91 |
| y | 14 37 13 23 21 13 18 21 22 21 28 31 44 23 46 115 59 45 7 65 20 45 32 27 48 59 39 9 168 44 36 29 23 35 20 16 28 16 64 40 64 135 61 32 49 63 108 35 80 52 50 78 44 73 39 79 66 47 77 82 85 122 71 89 110 96 |
| x | 91 92 92 92 93 93 94 94 96 |
| y | 96 97 97 78 73 91 118 85 84 |

También se podría haber realizado esta aproximación mediante splines.



Como observamos en la representación gráfica, es suficiente con el ajuste lineal un ajuste cuadrático o cúbico es exagerado.

Parte de una base de (*K+p*) polinomios B-splines (de Boor, 1978)



donde

*K* = número de nodos interiores.

*p* = grado de los polinomios (p.e. cúbicos, *p=*3).

**Modelo de regresión B-spline**:



Con este modelo, se ha convertido el problema original en un problema de regresión lineal múltiple.

• **Grado de flexibilidad del modelo:**

*K* = número de nodos interiores.

*df* = *K+p+1* (ó *K+p* sin la constante).

**VENTAJAS /DESVENTAJAS DE LA REGRESIÓN SPLINE**

**Ventajas:**

• Es una regresión **paramétrica LOCAL**.

• Puede realizarse fácilmente con R, usando **lm** ó **glm**,

pudiéndose aplicar a **cualquier tipo de respuesta** (continua,

binaria, poisson,…).

**Desventajas:**

• Debemos fijar de antemano:

1. Nº de knots ***K*** (ó equivalentemente los ***df*** del modelo).

2. Localización de los knots.

• Todavía no existe un criterio automático totalmente óptimo para

la selección y localización de los knots.

• Si el criterio es subjetivo, el investigador debe guiarse por el

 conocimiento del problema en cuestión.