Mathematical Modeling

数学建模

同济大学数学科学学院

第十三章

数据分析建模

目录/Contents ▌第十三章 数据分析建模

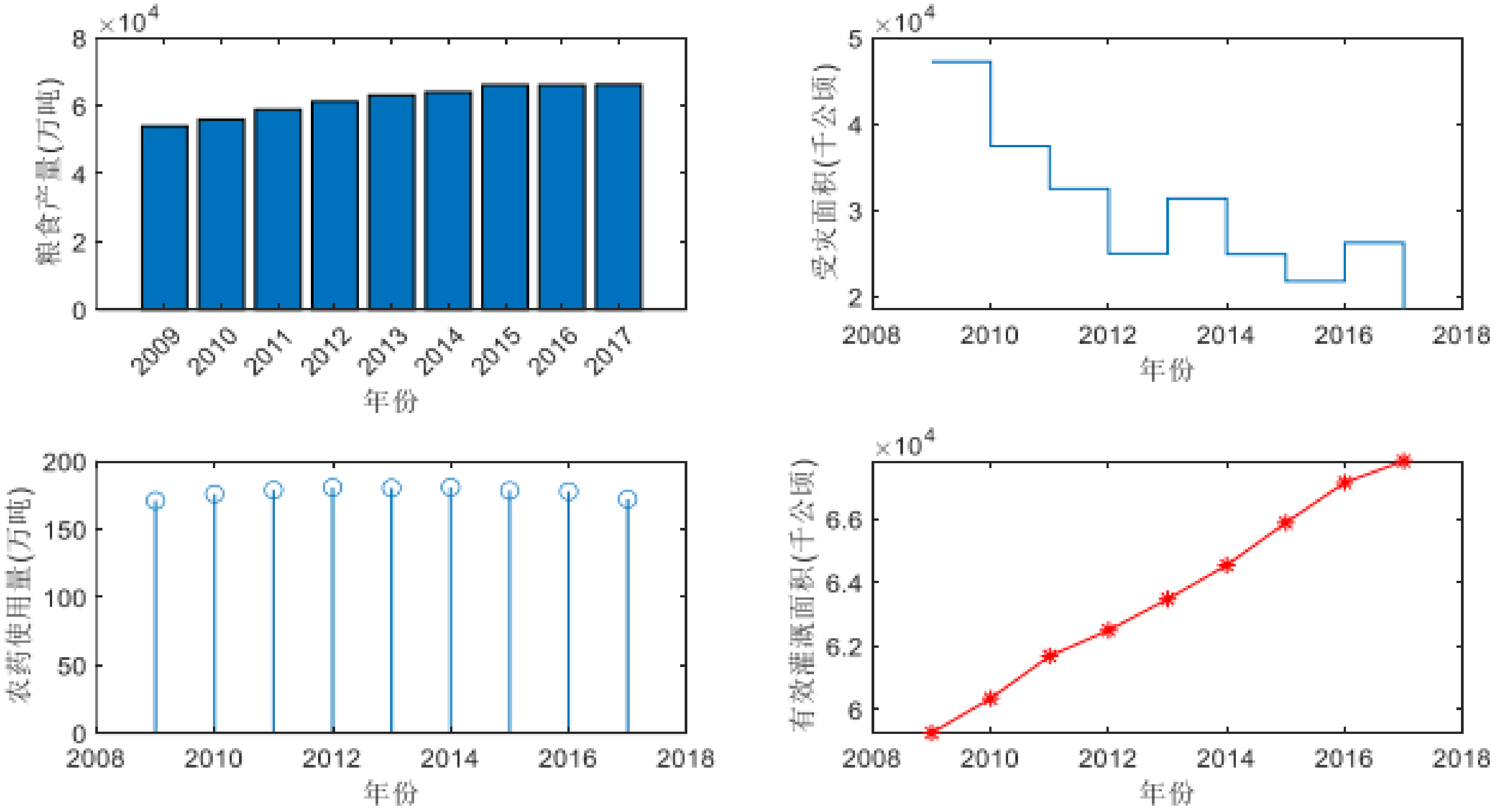
数据的可视化

2009年到2017年我国农业生产情况的统计数据

	2009年	2010年	2011年	2012年	2013年	2014年	2015年	2016年	2017年
粮食产量(万吨)	53940.86	55911.31	58849.33	61222.62	63048.2	63964.83	66060.27	66043.52	66160.72
受灾面积(千公顷)	47214	37426	32471	24962	31350	24891	21770	26221	18478
农药使用量(万吨)	170.9	175.82	178.7	180.61	180.19	180.69	178.3	177.6	172
有效灌溉面积(千公顷)	59261.4	60347.7	61681.56	62490.52	63473.3	64539.53	65872.64	67140.62	67815.57
农用化肥施用折纯量(万吨)	5404.4	5561.68	5704.24	5838.85	5911.86	5995.94	6022.6	5984.1	5859.41
农村用电量(亿千瓦时)	6104.44	6632.35	7139.62	7508.46	8549.52	8884.4	9026.92	9238.3	9524.42

数据的可视化

2009年到2017年我国农业生产情况的统计数据



数据的预处理

极大型， 极小型，居中型数据

$$d_{ij}=\begin{cases}\dfrac{x_{ij}}{\max_{1\le i\le n} x_{ij}}, & x_j\in I_1\\[2ex] \dfrac{\min_{1\le i\le n} x_{ij}}{x_{ij}}, & x_j\in I_2\\[2ex] \dfrac{\min_{1\le i\le n}|x_{ij}-\alpha_j|}{|x_{ij}-\alpha_j|}, & x_j\in I_3\end{cases}$$

目录/Contents ▍第十三章 数据分析建模

集中趋势的描述　数值平均数

- 算术平均数 $$\bar{x}=\frac{x_1+x_2+\cdots+x_n}{n}=\frac{1}{n}\sum_{i=1}^{n}x_i \qquad \bar{x}=\frac{\omega_1x_1+\omega_2x_2+\cdots+\omega_nx_n}{\omega_1+\omega_2+\cdots+\omega_n}$$
- 调和平均数 $$\bar{x}=\frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\cdots+\frac{1}{x_n}}$$
- 几何平均数 $$\bar{x}=\left(x_1x_2\cdots x_n\right)^{1/n}$$

集中趋势的描述　位置平均数

- 中位数　M_e
- 四分位数　$Q_1=\frac{n+1}{4},\frac{2(n+1)}{4},Q_3=\frac{3(n+1)}{4}$
- 十分位数，百分位数
- 众数　M_o

离散程度的描述

- 极差 $R = \max\{x_i\} - \min\{x_i\}$
- 四分位差 $Q_r = Q_3 - Q_1$
- 平均离差 $M_D = \frac{1}{n}\sum_{i=1}^{n} |x_i - \bar{x}|$
- 标准差 总体数据 $\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$

  样本数据 $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$

目录/Contents ▌第十三章 数据分析建模

- 多项式插值 $\varphi_n(x) = a_0 + a_1 x + \cdots + a_n x^n$，$\varphi_n(x_i) = y_i$
- Lagrange插值 $L_n(x) = \sum_{i=0}^{n} y_i \left(\prod_{j \neq i, j=0}^{n} \frac{x - x_j}{x_i - x_j} \right)$
- Hermite插值 $\varphi_n(x) = a_0 + a_1 x + \cdots + a_n x^n$，$\varphi_n(x_i) = y_i$，$\varphi'_n(x_i) = y_i'$
- 分段线性插值 $I_n(x) = \sum_{i=0}^{n} y_i l_i(x)$ $\quad l_i(x) = \begin{cases} \frac{x - x_{i-1}}{x_i - x_{i-1}}, & x \in [x_{i-1}, x_i] \\ \frac{x - x_{i+1}}{x_i - x_{i+1}}, & x \in [x_i, x_{i+1}] \end{cases}$
- 三次样条插值 分段三次多项式，二阶光滑

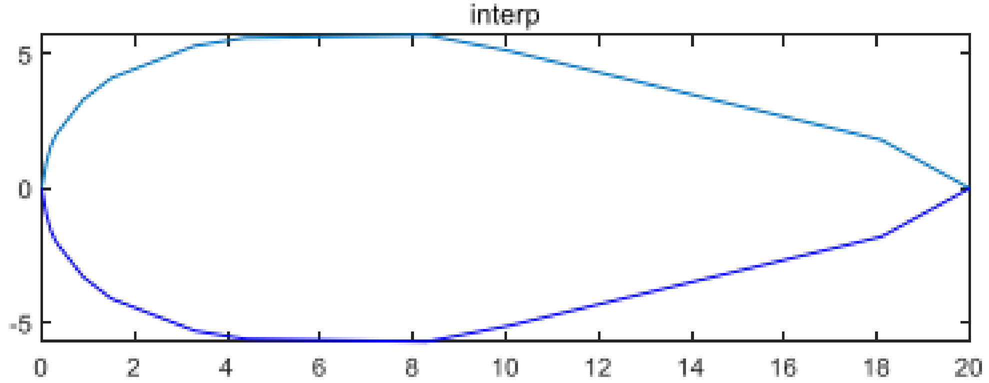
Matlab的插值方法

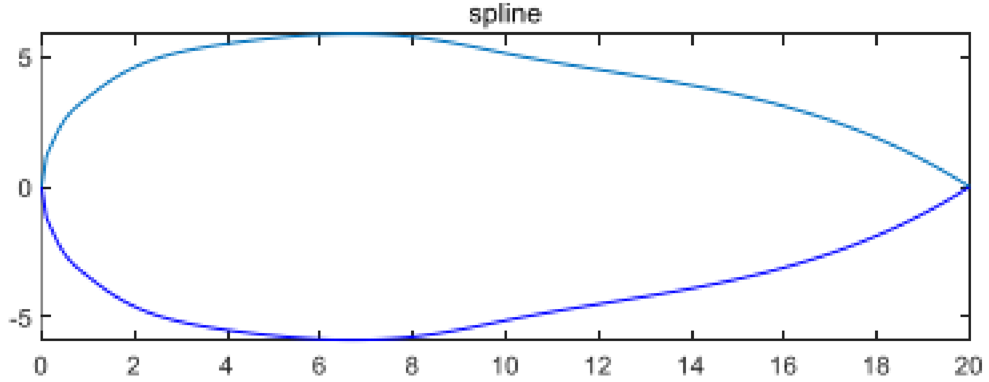
- 一维插值
  - `yi = interp1(x,y,xi,'method')`
  - method: nearest, linear, spline, cubic
- 二维插值
  - `z = interp2(x0,y0,z0,x,y,'method')`
  - method: 同一维插值
- 散乱节点插值
  - `cz = griddata(x,y,z,cx,cy,'method')`
  - method: 同一维插值

机翼断面的上下轮廓线

x	0	0.03	0.18	0.31	0.90	1.5	3.3	4.4	8.3	10.1	18.1	20.0
y_1	0	0.5	1.5	2.0	3.3	4.1	5.3	5.6	5.7	5.1	1.8	0
y_2	0	−0.5	−1.5	−2.0	−3.3	−4.1	−5.3	−5.6	−5.7	−5.1	−1.8	0

给出加工数据，求出面积





```
x0 = [ 0 0.03 0.18 0.31 0.90 1.50 3.30 4.40 8.30 10.10 18.10 20.00];
Y1 = [ 0 0.50 1.50 2.00 3.30 4.10 5.30 5.60 5.70 5.10 1.80 0 ];
Y2 = - Y1;
x = 0:0.1:20;
y1_in = interp1(x0,Y1,x);    % 分段线性插值
y2_in = interp1(x0,Y2,x);
y1_sp = spline(x0,Y1,x);    % 三次样条插值
y2_sp = spline(x0,Y2,x);
subplot(2,1,1)
   plot(x,y1_in,x,y2_in,'b');
   title('interp');
subplot(2,1,2)
   plot(x,y1_sp,x,y2_sp,'b');
   title('spline')
trapz(x,y1_in)-trapz(x,y2_in)   %分段线性插值积分值计算加工端面的面积
trapz(x,y1_sp)-trapz(x,y2_sp)   %三次样条插值积分值计算加工端面的面积
```

目录/Contents ▌第十三章 数据分析建模

- 线性拟合 $y = a_1\varphi_1(x) + \cdots + a_n\varphi_n(x)$ $\left(\varphi_j(x_i)\right)a = y$
- 多项式拟合 $y = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$ `p = polyfit(x,y,n)`
- 非线性拟合 `beta = nlinfit(X,Y, modelfun, beta0)`
- 矛盾方程 $Ax = b$ 法方程 $A^T Ax = A^T b$

目录/Contents ▌第十三章 数据分析建模

黄河小浪底调水调沙问题

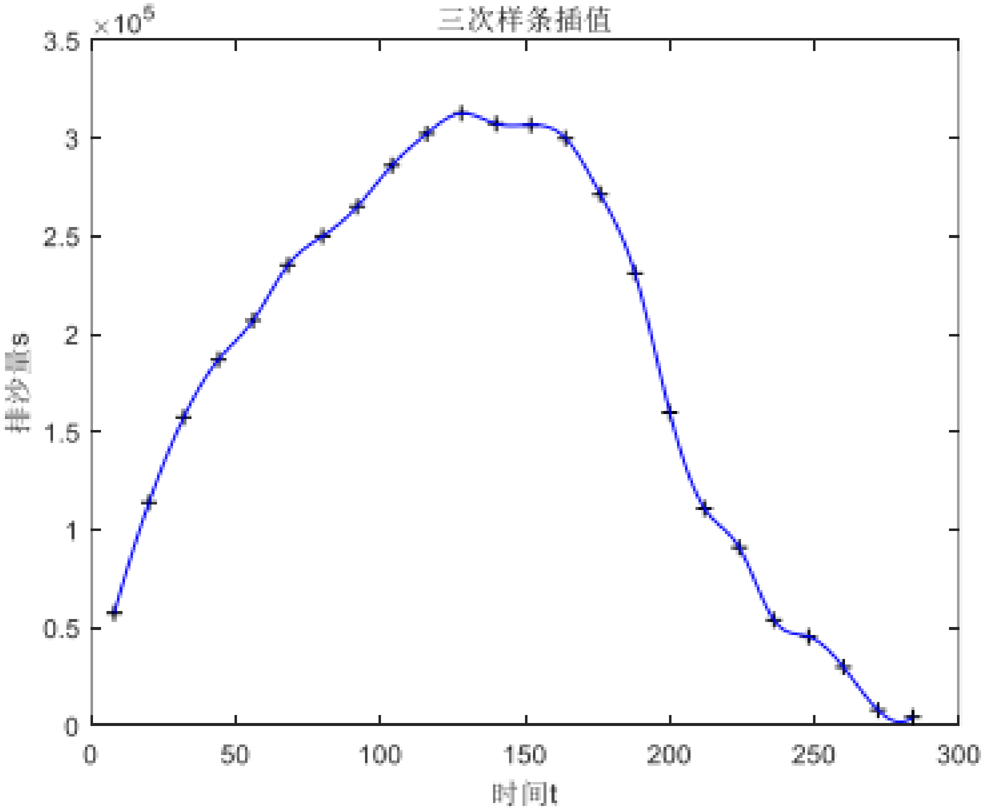
2004 年 6 月至 7 月黄河进行了第三次调水调沙试验，特别是首次由小浪底、三门峡和万家寨三大水库联合调度，采用接力式防洪预泄放水，形成人造洪峰进行调沙试验获得成功。整个试验期为20多天，小浪底从6月19日开始预泄洪放水，直到7月13日结束并回复成功供水。小浪底水利工程按设计拦沙量为 75.5 亿立方米，在这之前，小浪底共积沙达14.15亿吨。下表是由小浪底观测站从 6 月 29 日到 7 月 10 日检测到的试验数据。

日期	6.29		6.30		7.1		7.2		7.3		7.4	
时间	8:00	20:00	8:00	20:00	8:00	20:00	8:00	20:00	8:00	20:00	8:00	20:00
水流量	1800	1900	2100	2200	2300	2400	2500	2600	2650	2700	2720	2650
含沙量	32	60	75	85	90	98	100	102	108	112	115	116
日期	7.5		7.6		7.7		7.8		7.9		7.10	
时间	8:00	20:00	8:00	20:00	8:00	20:00	8:00	20:00	8:00	20:00	8:00	20:00
水流量	2600	2500	2300	2200	2000	1850	1820	1800	1750	1500	1000	900
含沙量	118	120	118	105	80	60	50	30	26	20	8	5

1. 给出估算任意时刻的排沙量及总排沙量的方法；
2. 确定排沙量与水流量的关系。

记时间 t_i， $x_i = 8 + 12i,\ (i = 0, 1, \cdots, 23)$， 排沙量为 y_i

构造三次样条 $S(x)$， 并计算积分
(不建议采用多项式拟合)



```
t0 = 8:12:24*12;        t = 8:284;
sll = [ 1800 1900 2100 2200 2300 2400 2500 2600 2650 2700 2720 2650 2600 2500 2300 ...
        2200 2000 1850 1820 1800 1750 1500 1000 900 ];
hsl = [ 32 60 75 85 90 98 100 102 108 112 115 116 118 120 118 105 80 60 50 30 26 20 8 5 ];
pslo = sll .* hsl;     % 计算相应时刻的排沙量
pp = spline(t0,pslo);
psl = ppval(pp,t);
plot(t0,pslo,'k+',t,psl,'b');
title('三次样条插值');
xlabel('时间t');
ylabel('排沙量s');
a = pp.breaks;
b = pp.coefs;
syms x c f
for i=1:23
   c = [(x-a(i))^3 (x-a(i))^2 (x-a(i)) 1];
   f = c*b(i,:)'
   s(i) = int(f,a(i),a(i+1));
end
zpsl = sum(s)*3600;
zpsl = vpa(zpsl)
```

任务

1. 水道测量问题 MCM1986A
2. 估计水箱的流水量 MCM1991A

Mathematical Modeling

数学建模

同济大学数学科学学院

学海无涯，祝你成功！

人民邮电出版社
POSTS & TELECOM PRESS