数学建模模型算法精讲课——

灰色预测GM(1,1)模型

—— 江北老师

水激石则鸣，
人激志则宏

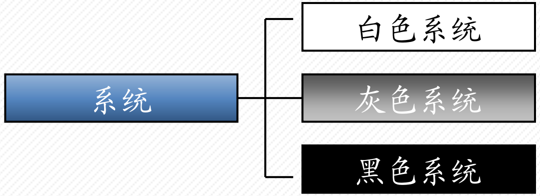
灰色预测模型

- 模型引出
- 模型求解
- 模型检验
- 具体代码

灰色预测模型——模型引出

➢ 什么是灰色系统?

- **灰色系统理论**是1982年由邓聚龙创立的一门边缘性学科(interdisciplinary)。灰色系统用颜色深浅**反映信息量的多少**。说一个系统是黑色的，就是说这个系统是黑洞洞的，信息量太少；说一个系统是白色的，就是说这个系统是清楚的，信息量充足。这种处于黑白之间的系统，或说信息不完全的系统，称为灰色系统或简称灰系统。



“信息不完全”是灰的基本含义，一般指：
（1）系统因素不完全明确
（2）因素关系不完全清楚
（3）系统的结构不完全知道
（4）系统的作用原理不完全明了

	白	黑	灰
从表象看	明朗	暗	若明若暗
从过程看	新	旧	新旧交替
从性质看	纯	不纯	多种成分
从信息看	完全	不完全	部分完全
从结果看	唯一的解	无数的解	非唯一性
从态度看	肯定	否定	扬弃
从方法看	严厉	放纵	宽容

灰色预测模型——模型引出

数模加油站

什么是灰色预测?

- 所谓灰色预测，就是对既含有已知信息又含有不确定信息的系统进行预测，就是对在一定范围内变化的、与时间有关的灰色过程进行预测。

灰色预测适用于什么情况?

- （1）数据是**以年份度量**的非负数据（如果是月份或者季度数据一般要用时间序列模型），比如定时求量的题目，即已知一些年份数据，预测下一年的数据，常见有GDP、人口数量、耕地面积、粮食产量等；或者定量求时，已知一些年份数据和某灾变的阈值，预测下次灾变时间。
- （2）数据能经过准指数规律的检验（除了前两期外，后面至少90%的期数的光滑比要低于0.5）。
- （3）数据的**期数较短**且和**其他数据之间的关联性**不强（小于等于10，也不能太短了，比如只有3期数据），要是数据期数较长，一般用传统的时间序列模型比较合适。

灰色预测模型——模型引出

- **GM（1,1）模型**
  - *GM*： *Grey model* 灰色模型，
  - (1,1)：只含有一个变量的一阶微分方程模型
  - 如何用*GM*(1,1)进行灰色预测？
    - ✓ 根据原始的离散非负数据列，通过累加等方式削弱随机性、获得有规律的**离散数据列**
    - ✓ 建立相应的**微分方程模型**，得到离散点处的解
    - ✓ 再通过累减求得的**原始数据的估计值**，从而对原始数据预测
- **长江水质污染预测**
  - 长江在1995-2004年废水排放总量如下，如果不采取保护措施，请对今后水质污染发展趋势做出预测。

年份	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
排污总量	174	179	183	189	207	234	220.5	256	270	285

灰色预测模型——模型求解

➢ 长江水质污染预测

年份	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
排污总量	174	179	183	189	207	234	220.5	256	270	285

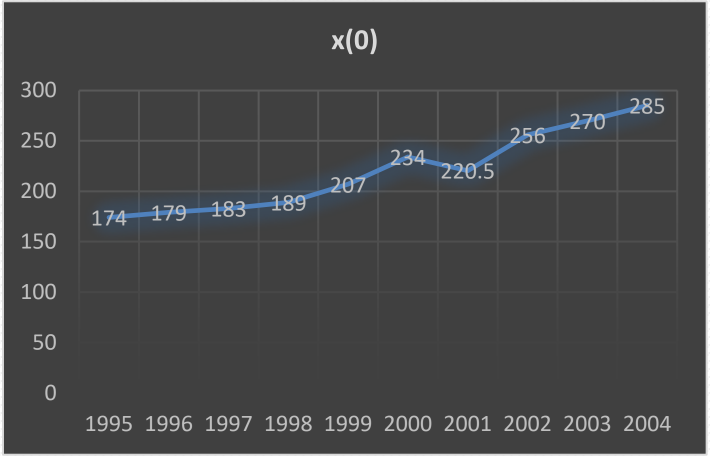
- 观察发现，没有明显规律，数据也比较少，而且是以年份度量的，可以考虑用灰色预测

- 那看不出规律怎么办，可以制造规律：

设$x^{(0)}=(x^{(0)}(1),x^{(0)}(2),\dots,x^{(0)}(n))$是最初的非负数据列，我们可以对其累加，得到新的数据列$x^{(1)}$

$$x^{(1)}=(x^{(1)}(1),x^{(1)}(2),\dots,x^{(1)}(n))$$

其中：$x^{(1)}(m)=\sum_{i=1}^{m}x^{(0)}(i)\,,m=1,2,\dots,n$



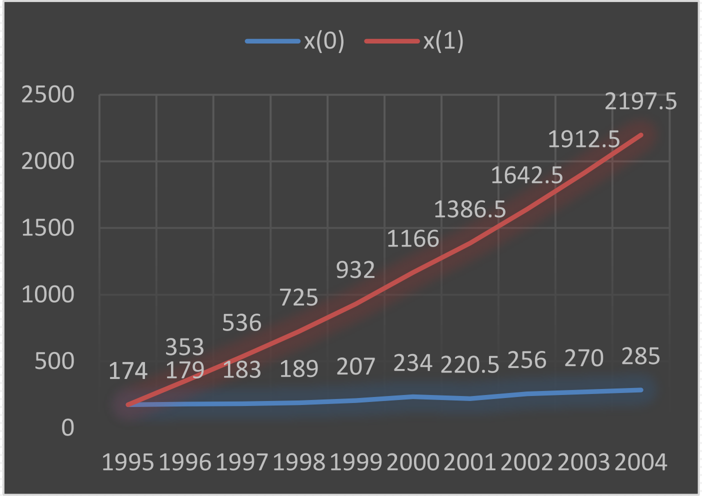
长江水质污染预测

年份	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
排污总量 $x^{(0)}$	174	179	183	189	207	234	220.5	256	270	285
排污总量 $x^{(1)}$	174	353	536	725	932	1166	1386.5	1642.5	1912.5	2197.5

- 观察可知，新序列$x^{(1)}$曲线像一个指数曲线（直线）
- 我们可以用指数曲线的表达式来逼近序列$x^{(1)}$，相应可以构建一阶常微分方程来求解拟合指数曲线的函数表达式
- 一阶常微分方程

$$\frac{dx^{(1)}}{dt}+ax^{(1)}=u$$

- 我们是想求出$x^{(1)}$的表达式，就需要解出常微分方程，所以要先知道参数a和u



求解微分方程

- 一阶常微分方程

$$\frac{dx^{(1)}}{dt}+ax^{(1)}=u$$

- 已知，我们的数据是离散的，所以$\frac{dx^{(1)}}{dt}$等同于$\frac{\Delta x^{(1)}}{\Delta t}=\frac{\Delta x^{(1)}}{t-(t-1)}=\Delta x^{(1)}=x^{(1)}(t)-x^{(1)}(t-1)=x^{(0)}(t)$
- 则微分方程变为$x^{(0)}(t)+ax^{(1)}(t)=u$
- 上式为常见的一元线性方程，为了消除数据随机性，定义$z^{(1)}=(z^{(1)}(1),z^{(1)}(2),\dots,z^{(1)}(n))$

  其中：$z^{(1)}(m)=\delta x^{(1)}(m)+(1-\delta)x^{(1)}(m-1),m=2,3,\dots,n$ 且$\delta=0.5$，即为前后时刻的均值
- 则微分方程改为$x^{(0)}(t)=-az^{(1)}(t)+u$
- 我们已知$x^{(0)}(t)$，$z^{(1)}(t)$的数据，结合线性规划的知识，可利用线性规划或者用最小二乘法求解参数

灰色预测模型——模型求解

求解微分方程

- $x^{(0)}(t) = -az^{(1)}(t) + u$，用矩阵表示$Y = X\beta + \epsilon$
- 利用回归分析或者最小二乘法求解，在回归分析一节我们已经介绍过最小二乘法的原理和求解
- 最小二乘法即使因变量的观察值与估计值之间的离差平方和**达到最小**来求解

  即$minQ = \sum\left(x^{(0)} - \hat{x}^{(0)}\right)^2$
- 最小二乘法通解形式为： $\hat{\beta} = (X^TX)^{-1}(X^TY)$，解得$\hat{\beta} = [156.6162\,,0.0624]$
- 即$\hat{a} = -0.0624, \hat{u} = 156.6162$
- 把$\hat{a}$和$\hat{u}$代入微分方程$\frac{dx^{(1)}}{dt} + ax^{(1)} = u$，并求解可得

  $\hat{x}^{(1)}(m+1) = \left[x^{(0)}(1) - \frac{\hat{b}}{\hat{a}}\right]e^{-\hat{a}m} + \frac{\hat{b}}{\hat{a}}, m = 1, 2, \ldots, n-1$
- 由于$x^{(1)}(m) = \sum_{i=1}^{m} x^{(0)}(i)\,, m = 1,2, \ldots, n$，所以我们可以得到：

  $\hat{x}^{(0)}(m+1) = \hat{x}^{(1)}(m+1) - \hat{x}^{(1)}(m) = \left(1 - e^{\hat{a}}\right)\left[x^{(0)}(1) - \frac{\hat{b}}{\hat{a}}\right]e^{-\hat{a}m}, m = 1, 2, \ldots, n-1$
- 当m取$0,1, \ldots, 9$时，得到的$\hat{x}^{(0)}$为拟合值，大于9时，得到的为预测值

➢ 指数规律检验（原始数据级比检验）

- 为了确定原始数据是否可使用灰色预测模型，需要对原始数据进行**级比检验**
- 要使用灰色预测数据首先应具有准指数规律:

  累加r次的序列为：$x^{(r)}=\left(x^{(r)}(1), x^{(r)}(2), \ldots, x^{(r)}(n)\right)$，定义级比$\sigma(k)=\frac{x^{(r)}(k)}{x^{(r)}(k-1)}, k=2,3,\ldots,n$

  对于$\forall k$，$\sigma(k)\in[a,\ b]$，且区间长度$\delta=b-a<0.5$，则称累加r次后的序列具有准指数规律

- 对于$GM(1,1)$模型中，我们只需要判断累加一次后的序列$x^{(1)}=\left(x^{(1)}(1), x^{(1)}(2), \ldots, x^{(1)}(n)\right)$是否具有准指数规律
- 根据上述公式：序列$x^{(1)}$的级比$\sigma(k)=\frac{x^{(1)}(k)}{x^{(1)}(k-1)}=\frac{x^{(0)}(k)+x^{(1)}(k-1)}{x^{(1)}(k-1)}=\frac{x^{(0)}(k)}{x^{(1)}(k-1)}+1$
- 定义$\rho(k)=\frac{x^{(0)}(k)}{x^{(1)}(k-1)}$为原始序列$x^{(0)}$的光滑比，注意到$\rho(k)=\frac{x^{(0)}(k)}{x^{(0)}(1)+x^{(0)}(2)+\cdots+x^{(0)}(k-1)}$，假设$x^{(0)}$为非负序列（生活中常见的时间序列几乎都满足非负性），那么随着k增加，最终$\rho(k)$会逐渐接近0，因此要使得具有$x^{(1)}$具有准指数规律，即$\forall k$，区间长度$\delta<0.5$，只需要保证$\rho(k)\in(0,\ 0.5)$即可，此时序列$x^{(1)}$的级比$\sigma(k)\in(1,\ 1.5)$。
- 实际建模中，我们要计算出$\rho(k)\in(0,\ 0.5)$的占比，占比越高越好（一般$\rho(2)$和$\rho(3)$可能不符合，重点关注后面的数据）

模型评价（检验模型对原始数据的拟合程度）

- 1）残差检验

绝对残差：$\epsilon(k) = x^{(0)}(k) - \hat{x}^{(0)}(k), k = 2, 3, \dots, n$

相对残差：$\epsilon_r(k) = \frac{|x^{(0)}(k) - \hat{x}^{(0)}(k)|}{x^{(0)}(k)} \times 100\%\ , k = 2, 3, \dots, n$

平均相对残差：$\bar{\epsilon}_r = \frac{1}{n-1}\sum_{k=2}^{n}|\epsilon_r(k)|$

如果$\bar{\epsilon}_r < 20\%$，则认为$GM(1,\ 1)$对原数据的拟合达到一般要求

如果$\bar{\epsilon}_r < 10\%$，则认为$GM(1,\ 1)$对原数据的拟合效果非常不错

- 2）级比偏差检验

首先计算原始数据级比$\sigma(k) = \frac{x^{(0)}(k)}{x^{(0)}(k-1)}(k = 2,3, \dots, n)$

再根据预测发展系数$(-\hat{a})$计算级比偏差和平均级比偏差

$\eta(k) = \left|1 - \frac{1-0.5\hat{a}}{1+0.5\hat{a}}\frac{1}{\sigma(k)}\right|,\ \bar{\eta} = \sum_{k=2}^{n}\eta(k)/(n-1)$

如果$\bar{\eta} < 0.2$, 则认为模型对元数据拟合达到一般要求， $\bar{\eta} < 0.1$,则认为模型对元数据拟合效果非常不错

灰色预测模型——具体代码

- **主代码：**

  详见$GM_main.m$文件

- **函数文件**

  详见$gm11.m$文件

欢迎关注数模加油站

THANKS