综合评价和预测方法简介

陈　磊　副主任，博导

西安交通大学能源与动力工程学院

数据中心节能与低碳技术重点实验室

热流科学与工程教育部重点实验室

2005年全国大学生数学建模竞赛A题–长江水质的评价和预测

“长江水质的评价和预测”问题的第一部分给出了17个观测站（城市）的最近28个月的实际检测指标数据，包括反映水质污染程度的最主要的四项指标：溶解氧（DO）、高锰酸盐指数（CODMn）、氨氮(NH3-N）和PH值，要求综合这四种污染指标的28个月的检测数据对17个城市的水质情况做出综合评价。

发布日期:2003-06

序号	点位名称	断面情况	主要监测项目(单位:mg/L)				水质类别		主要污染指标
			pH*	DO	CODMn	NH3-N	本月	上月	
1	四川攀枝花	干流	7.6	6.8	0.2	0.1	Ⅱ	Ⅱ	
2	重庆朱沱	干流（川-渝省界）	7.63	8.41	2.8	0.34	Ⅱ	Ⅱ	
3	湖北宜昌南津关	干流（三峡水库出口）	7.07	7.81	5.8	0.55	Ⅲ	Ⅲ	
4	湖南岳阳城陵矶	干流	7.58	6.47	2.9	0.34	Ⅱ	Ⅱ	
5	江西九江河西水厂	干流（鄂-赣省界）	7.34	6.19	1.7	0.13	Ⅱ	Ⅱ	
6	安徽安庆皖河口	干流	7.52	6.54	3.2	0.22	Ⅱ	Ⅱ	
7	江苏南京林山	干流（皖-苏省界）	7.78	6.9	3.1	0.11	Ⅱ	Ⅱ	
8	四川乐山岷江大桥	岷江(与大渡河汇合前)	7.66	4.2	5.8	0.53	Ⅳ	Ⅳ	溶解氧
9	四川宜宾凉姜沟	岷江（入长江前）	8.01	7.63	2.4	0.25	Ⅱ	Ⅱ	
10	四川泸州沱江二桥	沱江（入长江前）	7.63	4.02	3.6	1.06	Ⅳ	Ⅳ	溶解氧、氨氮
11	湖北丹江口胡家岭	丹江口水库（库体）	8.63	10.2	1.8	0.1	Ⅰ	Ⅰ	
12	湖南长沙新港	湘江（洞庭湖入口）	7.42	6.45	4.3	0.99	Ⅲ	Ⅲ	
13	湖南岳阳岳阳楼	洞庭湖出口	7.73	6.26	1.4	0.21	Ⅱ	Ⅲ	
14	湖北武汉宗关	汉江（入长江前）	8	6.43	2.4	0.17	Ⅱ	Ⅱ	
15	江西南昌滁槎	赣江（鄱阳湖入口）	6.64	5.18	1.1	0.92	Ⅲ	Ⅲ	
16	江西九江蛤蟆石	鄱阳湖出口	7.28	6.87	2.7	0.15	Ⅱ	Ⅱ	
17	江苏扬州三江营	夹江（南水北调取水口）	7.29	6.9	1.6	0.15	Ⅱ	Ⅱ	

发布日期:2003-07

表（1）：《地表水环境质量标准》（GB3838—2002）中 4 个主要项目标准限值　　　单位：mg/L

指　标	Ⅰ类	Ⅱ类	Ⅲ类	Ⅳ类	Ⅴ类	劣Ⅴ类
溶解氧(DO)	[7.5,∞)	[6,7.5)	[5,6)	[3,5)	[2,3)	[0,2]
高锰酸盐指数(CODMn)	(0,2]	(2,4]	(4,6]	(6,10]	(10,15]	(15, ∞)
氨氮（NH3-N）	(0,0.15]	(0.15,0.5]	(0.5,1]	(1,1.5]	(1.5,2]	(2, ∞)
PH 值（无量纲）	[6 , 9]					

根据国标（GB 3838—2002）的规定，关于地表水的水质可分为Ⅰ类、Ⅱ类、Ⅲ类、Ⅳ类、Ⅴ类、劣Ⅴ类共六个类别，每一个类别对每一项指标都有相应的标准值（区间），只要有一项指标达到高类别的标准就算是高类别的水质，所以实际中不同类别的水质有很大的差别，而且同一类别的水在污染物的含量上也有一定的差别。

在对17个城市的水质做综合评价时，要充分考虑这些指标值不同类别水的“质的差异”和同类别水的“量的差异”,在此简称为“质差”和“量差”。因此，这是一个较复杂的多因素多属性的综合评价问题。

综合评价:对被评价对象所进行的客观、公正、合理的全面评价。通常的综合评价问题都是有若干个同类的被评价对象(或系统)，每个被评价对象往往都涉及到多个属性（或指标）。

综合评价的目的:根据系统的属性判断确定这些系统的运行（或发展）状况哪个优，哪个劣，即按优劣对各被评价对象进排序或分类。这类问题又称为**多属性（或多指标）的综合评价问题**。

综合评价的应用：在政治、经济、社会及军事管理、工程技术及科学决策等领域都有重要的应用价值研究，解决这类问题在实际中是很有意义的。

一、 构成综合评价问题的五个要素

构成综合评价问题的五个要素分别为:被评价对象、评价指标、权重系数、综合评价模型和评价者。

（1）被评价对象

被评价对象就是综合评价问题中所研究的对象，或称为系统。通常情况下，在一个问题中被评价对象是属于同一类的，且个数要大于 1，不妨假设一个综合评价问题中有n 个被评价对象（或系统），分别记为$S_1, S_2, \cdots, S_n (n>1)$。

（2）评价指标

评价指标是反映被评价对象(或系统)的运行(或发展)状况的基本要素。通常的问题都是有多项指标构成，每一项指标都是从不同的侧面刻画系统所具有某种特征大小的一个度量。

一个综合评价问题的评价指标一般可用一个向量表示，其中每一个分量就是从一个侧面反映系统的状态，即称为综合评价的指标体系。

评价指标体系应遵守的原则：系统性、科学性、可比性、可测性（即可观测性）和独立性。这里不妨设系统有 m 个评价指标（或属性），分别记为 $x_1, x_2, \cdots, x_m (m>1)$，即评价指标向量为 $\boldsymbol{x} = (x_1, x_2, \cdots, x_m)^T$。

（3）权重系数

每一综合评价的问题都有相应的评价目的，针对某种评价目的，各评价指标之间的相对重要性是不同的，评价指标之间的这种相对重要性的大小可以用权重系数来刻画。如果用 w_j 来表示评价指标 $x_j (j=1,2,\cdots,m)$ 的权重系数，则应有 $w_j \geq 0 (j=1,2,\cdots,m)$，且 $\sum_{j=1}^{m} w_j = 1$。

权数的确定方法

- **主观赋权法**

■ **德尔菲法（专家法）**——实际上各个专家可以根据自己的理解选择不同的方法

■ **相邻指标比较法**；先按重要性将全部评价指标排序，再将相邻指标的重要性进行比较

■ 层次分析法（ＡＨＰ）——互反式两两比较构权法。

- **客观赋权法——从指标的统计性质来考虑，它是由客观数据决定。**
- **客观定权法包括模糊定权法、秩和比法、熵权法和相关系数法等**

⑩ 权数的特性（指主观权数、人工权数）

- 重要性——权数是一种重要性程度的量化值。指对合成值的影响程度大小。重要性本身是个综合的概念，表现在多个方面，如可以是“价值判断取向”上的重要性，也可以是合成时“分辨能力（信息含量）高低”的重要性，或“可靠度大小”的重要性。
- 模糊性——重要性本身就是个模糊的概念。
- 人工性——没有绝对的正确错误标准；只能尽可能选择相对科学合理的权数。
- 主观性——受评权者主观意识的影响

（4）综合评价模型

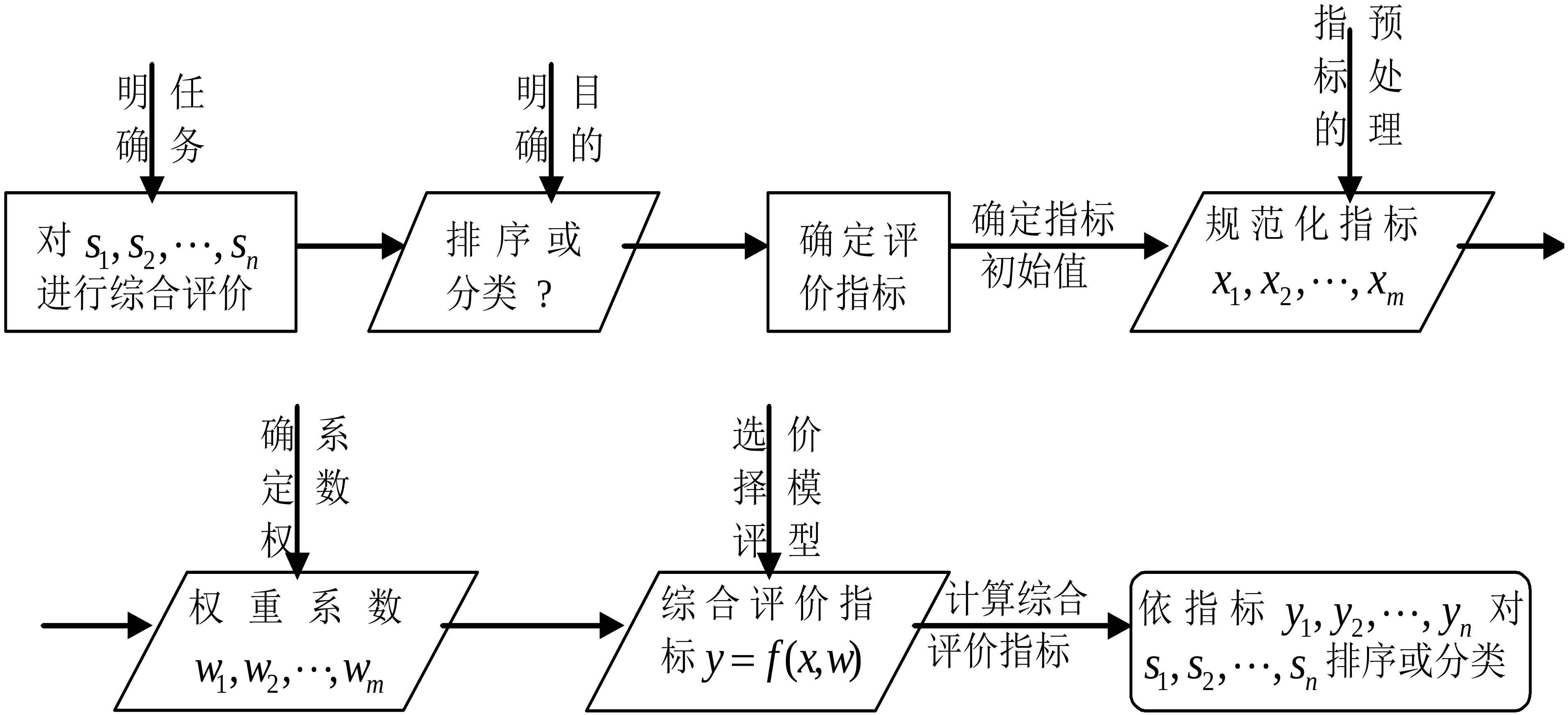
对于多指标（或多因素）的综合评价问题,就是要通过建立合适的综合评价数学模型将多个评价指标综合成为一个整体的综合评价指标，作为综合评价的依据，从而得到相应的评价结果。

不妨假设 n 个被评价对象的m 个评价指标向量为 $\boldsymbol{x}=(x_1,x_2,\cdots,x_m)^T$，指标权重向量为 $\boldsymbol{w}=(w_1,w_2,\cdots,w_m)^T$，由此构造综合评价函数为 $y=f(\boldsymbol{w},\boldsymbol{x})$ 。

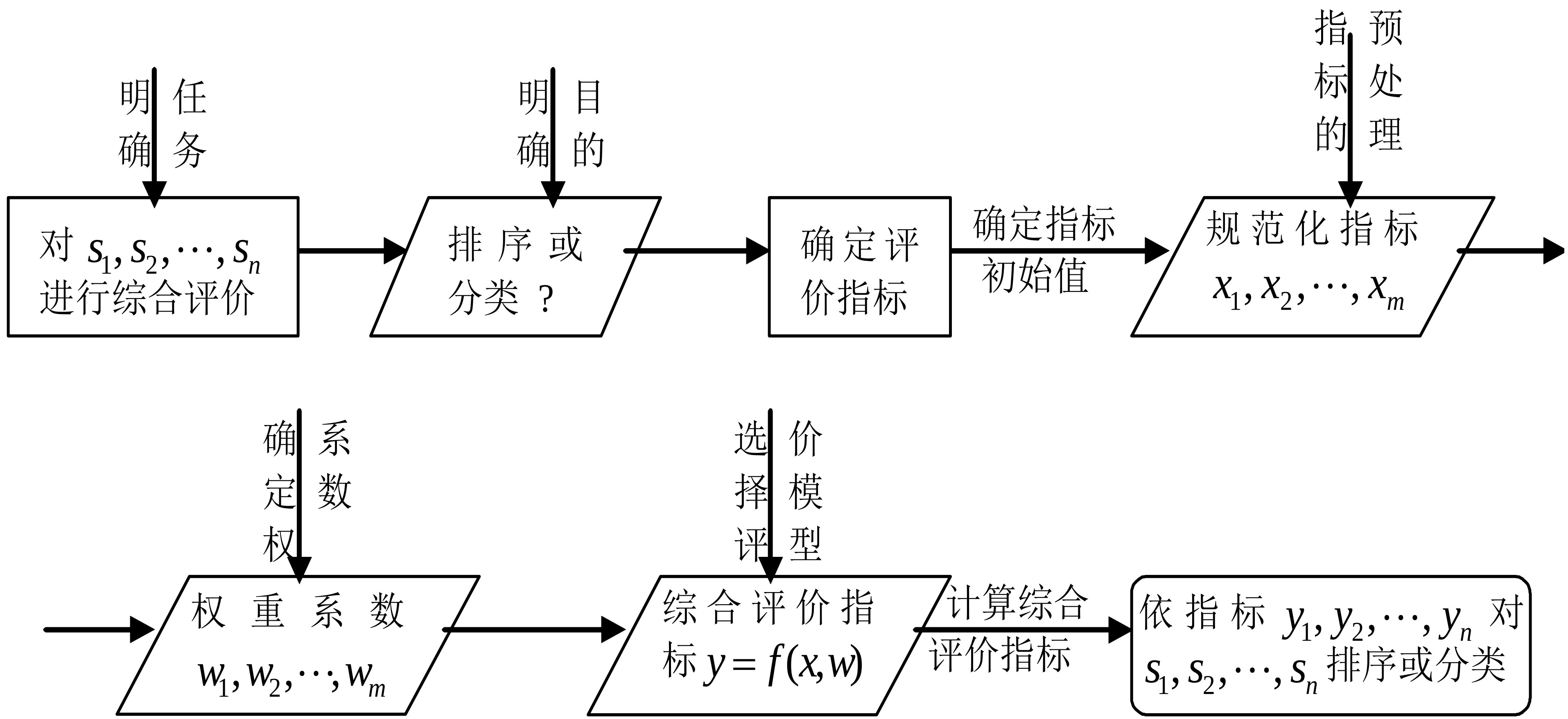
如果已知各评价指标的 n 个观测值为 $\{x_{ij}\}(i=1,2,\cdots,n;\ j=1,2,\cdots,m)$，则可以计算出各系统的综合评价值 $y_i=f(\boldsymbol{w},\boldsymbol{x}^{(i)})$，$\boldsymbol{x}^{(i)}=(x_{i1},x_{i2},\cdots,x_{im})^T\ (i=1,2,\cdots,n)$ 。根据 $y_i(i=1,2,\cdots,n)$ 值的大小将这 n 个系统进行排序或分类，即得到综合评价结果。

（5）评价者

评价者是直接参与评价的人，可以是某一个人，也可以是一个团体。对于评价目的选择、评价指标体系确定、评价模型的建立和权重系数的确定都与评价者有关。



综合评价过程的流程


明确任务
对 $s_1, s_2, \cdots, s_n$ 进行综合评价
明确目的
排序或分类？
确定评价指标
确定指标初始值
指标的预处理
规范化指标 $x_1, x_2, \cdots, x_m$
确定权系数
权重系数 $w_1, w_2, \cdots, w_m$
选择评价模型
综合评价指标 $y = f(x, w)$
计算综合评价指标
依指标 $y_1, y_2, \cdots, y_n$ 对 $s_1, s_2, \cdots, s_n$ 排序或分类

二、评价指标的规范化处理

1. 评价指标类型的一致化

一般说来，在评价指标 $x_1, x_2, \cdots, x_m (m > 1)$ 中可能包含有“极大型”指标、“极小型”指标、“中间型”指标和“区间型”指标。

极大型指标:总是期望指标的取值越大越好;

极小型指标:总是期望指标的取值越小越好;

中间型指标:总是期望指标的取值既不要太大，也不要太小为好，即取适当的中间值为最好;

区间型指标:总是期望指标的取值最好是落在某一个确定的区间内为最好。

（1）极小型指标：对于某个极小型指标x，则通过变换$x'=\frac{1}{x}(x>0)$，或变换$x'=M-x$，其中M为指标x的可能取值的最大值，即可将指标x极大化。

（2）中间型指标：对于某个中间型指标x，则通过变换

$$x'=\begin{cases}\dfrac{2(x-m)}{M-m}, & m\le x\le\dfrac{1}{2}(M+m)\\[2ex] \dfrac{2(M-x)}{M-m}, & \dfrac{1}{2}(M+m)\le x\le M\end{cases}$$

其中M和m分别为指标x的可能取值的最大值和最小值，即可将中间型指标x极大化。

（3）区间型指标

对于某个区间型指标 x，则通过变换

$$x' = \begin{cases} 1 - \dfrac{a-x}{c}, & x < a \\ 1, & a \le x \le b \\ 1 - \dfrac{x-b}{c}, & x > b \end{cases}$$

其中 $[a,b]$ 为指标 x 的最佳稳定的区间，$c = \max\{a-m, M-b\}$，M 和 m 分别为指标 x 的可能取值的最大值和最小值。即可将区间型指标 x 极大化。

2. 评价指标的无量纲化

在实际中的评价指标 $x_1, x_2, \cdots, x_m (m>1)$ 之间，往往都存在着各自不同的单位和数量级，使得这些指标之间存在着不可公度性，这就为综合评价带来了困难，尤其是为综合评价指标建立和依据这个指标的大小排序产生不合理性。

如果不对这些指标作相应的无量纲处理，则在综合评价过程中就会出现“大数吃小数”的错误结果，从而导致最后得到错误的评价结论。

无量纲化处理又称为指标数据的标准化,或规范化处理。

常用方法:标准差方法、极值差方法和功效系数方法等。

假设m 个评价指标$x_1, x_2, \cdots, x_m$，在此不妨假设已进行了类型的一致化处理，并都有n组样本观测值$x_{ij}(i=1,2,\cdots,n; j=1,2,\cdots,m)$，则将其作无量纲化处理。

（1）标准差方法：令 $x'_{ij}=\dfrac{x_{ij}-\bar{x}_j}{s_j}\ (i=1,2,\cdots,n; j=1,2,\cdots,m)$，

其中 $\bar{x}_j=\dfrac{1}{n}\sum_{i=1}^{n}x_{ij},\ s_j=[\dfrac{1}{n}\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2]^{1/2}(j=1,2,\cdots,m)$。

显然指标 $x'_{ij}(i=1,2,\cdots,n; j=1,2,\cdots,m)$ 的均值和均方差分别为 0 和 1，称之为 x_{ij} 的标准观测值。

（2）极值差方法：令 $x'_{ij}=\dfrac{x_{ij}-m_j}{M_j-m_j}\ (i=1,2,\cdots,n; j=1,2,\cdots,m)$，

其中 $M_j=\max\limits_{1\le i\le n}\{x_{ij}\}, m_j=\min\limits_{1\le i\le n}\{x_{ij}\}(j=1,2,\cdots,m)$。则 $x'_{ij}\in[0,1]$ 是无量纲的指标观测值。

（3）功效系数法：令 $x'_{ij}=c+\dfrac{x_{ij}-m_j}{M_j-m_j}\times d\ \ (i=1,2,\cdots,n; j=1,2,\cdots,m)$，

其中c, d均为确定的常数。c表示"平移量"，d表示"旋转量"，即表示"放大"或"缩小"倍数，则$x'_{ij}\in[c, c+d]$。

譬如若取$c=60, d=40$，则$x'_{ij}\in[60,100]$。

三、综合评价的数学模型

- 1、计分法
- 2、综合指数法
- 3、Topsis法
- 4、秩和比(RSR)法
- 5、层次分析(AHP)法
- 6、模糊评价方法
- 7、多元统计分析方法
- 8、灰色系统评价方法
- 9、线性加权综合评价
- 10、非线性加权综合评价
- 11、动态加权综合评价方法

- 12、主成分分析法
- 13、数据包络分析法
- 14、人工神经网络评价方法
- 15、专家评价法
- 16、熵权法
- 17、简单指标比对法
- 18、随机前沿分析法
- 19、可拓物元法
- 20、泰勒展开法
- 21、混合方法
- … …

1) 线性加权综合法

线性加权综合法：用线性加权函数 $y=\sum_{j=1}^{m} w_j x_j$ 作为综合评价模型，对 n 个系统进行综合评价。

2) 非线性加权综合法

非线性加权综合法：用非线性函数 $y=\prod_{j=1}^{m} x_j^{w_j}$ 作为综合评价模型，对 n 个系统进行综合评价。其中 w_j 为权系数，且要求 $x_j \geq 1$。

非线性加权综合法适用于各指标间有较强关联的情况。

3、Topsis法

TOPSIS（Technique for order preference by similarity to ideal solution）法，即逼近理想解排序法，意为与理想方案相似性的顺序选优技术，是系统工程中有限方案多目标决策分析的一种常用方法。

- 它是基于归一化后的原始数据矩阵，找出有限方案中最优方案和最劣方案（分别用最优向量和最劣向量表示），然后分别计算诸评价对象与最优方案和最劣方案的距离，获得各评价对象与最优方案的相对接近程度，以此作为评价优劣的依据。

3、Topsis法

1. 设有*n*个评价对象、*m*个评价指标，原始数据可写为矩阵$X=(X_{ij})_{n\times m}$

2. 对高优、低优指标分别进行同向化、归一化变换

$$Z_{ij}=\frac{X_{ij}}{\sqrt{\sum_{i=1}^{n}X_{ij}^{2}}} \qquad Z_{ij}=\frac{1/X_{ij}}{\sqrt{\sum_{i=1}^{n}(1/X_{ij})^{2}}}$$

3、Topsis法

3. 归一化得到矩阵$Z=(Z_{ij})_{n\times m}$，其各列最大、最小值构成的最优、最劣向量分别记为

$$Z^{+}=(Z_{\max 1}\ \ Z_{\max 2}\ \ \cdots\ \ Z_{\max m})$$

$$Z^{-}=(Z_{\min 1}\ \ Z_{\min 2}\ \ \cdots\ \ Z_{\min m})$$

4. 第i个评价对象与最优、最劣方案的距离分别为

$$D_i^{+}=\sqrt{\sum_{j=1}^{m}(Z_{\max j}-Z_{ij})^2}\qquad D_i^{-}=\sqrt{\sum_{j=1}^{m}(Z_{\min j}-Z_{ij})^2}$$

5. 第i个评价对象与最优方案的接近程度C_i为

$$C_i=D_i^{-}\big/(D_i^{+}+D_i^{-})$$

3、Topsis法

例 某儿童医院1994～1998年7项指标的实际值，用Topsis法比较该医院这5年的医疗质量

年份	出院人数	病床使用率	平均住院日	病死率	抢救成功率	治愈好转率	院内感染率
1994	21584	76.7	7.3	1.01	78.3	97.5	2.0
1995	24372	86.3	7.4	0.80	91.1	98.0	2.0
1996	22041	81.8	7.3	0.62	91.1	97.3	3.2
1997	21115	84.5	6.9	0.60	90.2	97.7	2.9
1998	24633	90.3	6.9	0.25	95.5	97.9	3.6

3、Topsis法

平均住院日、病死率、院内感染率为低优指标，其余为高优指标，同向化、归一化变换

$$Z_{12} = 76.7\Big/\sqrt{76.7^2 + 86.3^2 + 81.8^2 + 84.5^2 + 90.3^2} = 0.4081$$

变换后，得到矩阵

$$\boldsymbol{Z} = \begin{pmatrix} 0.4234 & 0.4081 & 0.4380 & 0.2024 & 0.3916 & 0.4464 & 0.5612 \\ 0.4781 & 0.4592 & 0.4321 & 0.2556 & 0.4556 & 0.4487 & 0.5612 \\ 0.4324 & 0.4353 & 0.4380 & 0.3298 & 0.4556 & 0.4455 & 0.3508 \\ 0.4142 & 0.4496 & 0.4634 & 0.3408 & 0.4511 & 0.4473 & 0.3871 \\ 0.4833 & 0.4805 & 0.4634 & 0.8178 & 0.4776 & 0.4482 & 0.3118 \end{pmatrix}$$

3、Topsis法

计算各列最大、最小值构成的最优、最劣向量分别为

$$Z^{+}=(0.4833\ \ 0.4805\ \ 0.4634\ \ 0.8178\ \ 0.4776\ \ 0.4487\ \ 0.5612)$$

$$Z^{-}=(0.4142\ \ 0.4081\ \ 0.4321\ \ 0.2024\ \ 0.3916\ \ 0.4455\ \ 0.3118)$$

计算各年与最优、最劣向量的距离（以94年为例）

$$D_1^{+}=\sqrt{(0.4833-0.4234)^2+\cdots+(0.5612-0.5612)^2}=0.6289$$

$$D_1^{-}=\sqrt{(0.4142-0.4234)^2+\cdots+(0.3118-0.5612)^2}=0.2497$$

计算接近程度（以94年为例）

$$C_1=0.2497/(0.6289+0.2497)=0.2842$$

三、Topsis法

年份	D^+	D^-	C_i	排序
1994	0.6289	0.2497	0.2842	3
1995	0.5640	0.2754	0.3281	2
1996	0.5369	0.1514	0.2200	5
1997	0.5141	0.1762	0.2552	4
1998	0.2494	0.6302	0.7164	1

可以看出，1998 年综合效益最好，其次为 1995年，随后为 1994年、1997年，1996 年最差

4) 动态加权综合评价

现设有 n 个被评价对象（或系统），分别记为 $S_1,S_2,\cdots,S_n(n>1)$，每个系统都有 m 属性（或评价指标），分别记为 $x_1,x_2,\cdots,x_m(m>1)$，对于每一个属性 x_i 都可以分为 K 个等级，记为 $p_1,p_2,\cdots,p_K$ $(K>1)$ 。而对于每一个等级 p_k 都包含一个区间范围，记为 $[a_k^{(i)},b_k^{(i)})$ ，且 $a_k^{(i)}<b_k^{(i)}$ $(i=1,2,\cdots,m;k=1,2,\cdots,K)$，即当属性 $x_i\in[a_k^{(i)},b_k^{(i)})$ 时，则属性 x_i 属于第 k 类 p_k $(1\le k\le K)$。也就是对于每一个属性而言，既有不同类别的差异，同类别的又有不同量值的差异。对于这种既有“质差”，又有“量差”的问题，如果用通常的定常权综合评价法做综合评价显然是不合理的，然而合理有效的方法是动态加权综合评价方法。

根据标准化后的各评价指标值，不妨仍用 x_i 表示，以及相应的动态权函数 $w_i(x)(i=1,2,\cdots,m)$，建立综合评价模型来对 n 个被评价对象做出综合评价。在此，取综合评价模型为各评价指标的动态加权和，即

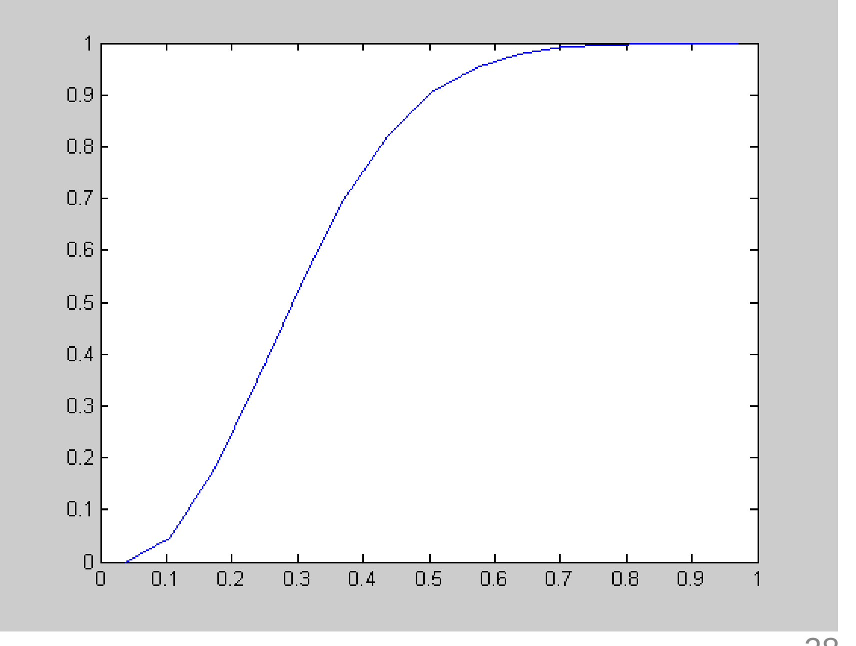
$$X=\sum_{i=1}^{m} w_i(x_i)\cdot x_i 。$$

以此作为问题的综合评价指标函数，如果每个被评价对象的 m 个属性都有 N 组样本观测值 $\{x_{ij}\}(i=1,2,\cdots,m;\ j=1,2,\cdots,N)$，代入上式计算，则每一个被评价对象都有 N 个综合评价指标值 $X_k(j)\,(k=1,2,\cdots,n;\ j=1,2,\cdots,N)$。由此按其大小排序，可以给出 n 个被评价对象的 N 个排序方案。

动态加权函数设定——偏大型正态分布函数

如果某项指标 x_i 对于综合评价效果的影响大约是随着类别 $p_k(k=1,2,\cdots,K)$ 的增加，先是缓慢增加，中间有一个快速增长的过程，随后平缓增加趋于最大，相应的图形呈正态分布曲线（左侧）形状。那么，此时对指标 x_i 的变权函数可以设定为偏大型正态分布函数。即

$$w_i(x)=\begin{cases}0 \quad , & \text{当 } x\le\alpha_i\text{时}, \\ 1-e^{-\left(\frac{x-\alpha_i}{\sigma_i}\right)^2}, & \text{当 } x>\alpha_i\text{时},\end{cases}$$



其中参数 α_i 可取 $[a_1^{(i)},b_1^{(i)})$ 中的某定值，在此不妨取 $\alpha_i=(b_1^{(i)}-a_1^{(i)})/2$，$\sigma_i$ 由 $w_i(a_K^{(i)})=0.9(1\le i\le m)$ 确定。

针对长江水质的综合评价这一问题，采用动态加权综合评价方法来解决。假设 17 个城市为被评价对象$S_1, S_2, \cdots, S_{17}$，共有四项评价指标（或属性）DO、CODMn、NH3-N 和 PH 值，分别记为x_1, x_2, x_3 和 x_4，前三项指标都有 6 个等级$p_1, p_2, \cdots, p_6$，相应的分类区间值如表（1）所示，而 PH 值没有等级之分。

表（1）：《地表水环境质量标准》（GB3838—2002）中 **4** 个主要项目标准限值　　　　单位：**mg/L**

指　标	Ⅰ类	Ⅱ类	Ⅲ类	Ⅳ类	Ⅴ类	劣Ⅴ类
溶解氧(**DO**)	[7.5,∞)	[6,7.5)	[5,6)	[3,5)	[2,3)	[0,2]
高锰酸盐指数(CODMn)	(0,2]	(2,4]	(4,6]	(6,10]	(10,15]	(15, ∞)
氨氮（NH3-N）	(0,0.15]	(0.15,0.5]	(0.5,1]	(1,1.5]	(1.5,2]	(2, ∞)
PH 值（无量纲）	[6 , 9]					

1. 指标数据的标准化处理

（1）溶解氧（DO）的标准化

注意到溶解氧（DO）为极大型指标，首先将数据指标作极小化处理，即令倒数变换 $x_1' = \dfrac{1}{x_1}$，相应的分类标准区间变为

$$(0, \frac{1}{7.5}], (\frac{1}{7.5}, \frac{1}{6}], (\frac{1}{6}, \frac{1}{5}], (\frac{1}{5}, \frac{1}{3}], (\frac{1}{3}, \frac{1}{2}], (\frac{1}{2}, \infty),$$

然后通过极差变换 $x_1'' = \dfrac{x_1'}{0.5}$ 将其数据标准化，对应的分类区间随之变为

$$(0, 0.2667], (0.2667, 0.3333], (0.3333, 0.4], (0.4, 0.6667], (0.6667, 1], (1, \infty)$$

（2）高锰酸盐指数（CODMn）的标准化

高猛酸盐指数本身就是极小型指标，即由极差变换将其数据标准化，

即令 $x_2' = \dfrac{x_2}{15}$， 对应的分类区间随之变为

$$(0,0.1333],(0.1333,0.2667],(0.2667,0.4],(0.4,0.6667],(0.6667,1],(1,\infty)$$

（3）氨氮（NH3-N）的标准化

氨氮也是极小型指标，对指标数据作极差变换将其数据标准化，

即令 $x_3' = \dfrac{x_3}{2}$，对应的分类区间随之变为

$$(0,0.075],(0.075,0.25],(0.25,0.5],(0.5,0.75],(0.75,1],(1,\infty)$$

（4）PH 值的处理

酸碱度（PH 值）的大小反映出水质呈酸碱性的程度，通常的水生物都适应于中性水质，即酸碱度的平衡值（PH 值略大于 7），在这里不妨取正常值的中值 7.5。当 PH<7.5 时水质偏碱性，当 PH>7.5 时偏酸性，而偏离值越大水质就越坏，PH 值属于中间型指标。为此，对所有的 PH 值指标数据作均值差处理，即令

$$x_4' = \frac{|x_4 - 7.5|}{1.5} = \frac{2}{3}|x_4 - 7.5|,$$

则将其数据标准化。

2. 动态加权函数的确定

根据对这一实际问题的分析，不妨取动态加权函数为偏大型正态分布函数，即

$$w_i(x)=\begin{cases}0\quad, & \text{当}\ x\le\alpha_i\text{时},\\ 1-e^{-\left(\frac{x-\alpha_i}{\sigma_i}\right)^2}, & \text{当}\ x>\alpha_i\text{时},\end{cases}$$

其中 α_i 在这里取指标 x_i 的 I 类水标准区间的中值，即 $\alpha_i=(b_1^{(i)}-a_1^{(i)})/2$，$\sigma_i$ 由 $w_i(a_4^{(i)})=0.9(i=1,2,3)$ 确定。

由实际数据经计算可得 $\alpha_1=0.1333,\alpha_2=0.0667,\alpha_3=0.0375$，$\sigma_1=0.1757,\sigma_2=0.2197,$ $\sigma_3=0.3048$，则代入上式可以得到 DO、CODMn 和 NH3-N 三项指标的动态加权函数。

3.综合评价指标函数的确定

考虑到对实际评价效果影响差异较大的是前三项指标，以及指标 PH 值的特殊性，这里取前三项指标的综合影响权值为 0.8，而 PH 值的影响权值取 0.2。因此，根据综合评价模型，某城市某一时间的水质综合评价指标定义为

$$X = 0.8\sum_{i=1}^{3} w_i(x_i)x_i + 0.2x_4\ 。$$

根据 17 个城市的 28 组实际检测数据，经计算可得各城市的水质综合评价指标值，即可得到一个17×28 阶的综合评价矩阵$(X_{ij})_{17\times 28}$ 。

4.各城市水质的综合评价

由 17 个城市 28 个月的水质综合评价指标 $X_{ij}(i=1,2,\cdots,17;\ j=1,2,\cdots,28)$，根据其大小（即污染的程度）进行排序，数值越大说明水质越差。由此可得反映 17 个城市水质污染程度的 28 个排序结果，根据 Borda 数的计算方法则得到第 i 个城市（被评价对象）S_i 的 Borda 数为

$$B(S_i)=\sum_{j=1}^{28} B_j(S_i)\quad (i=1,2,\cdots,17)$$

。

经计算可得到各城市的 Borda 数及总排序结果如表(2)所示。

表(2)：按各城市的水质污染总排序结果

城市 \ 排序	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}	S_{13}	14	15	16	17
Borda 数	203	136	143	234	106	139	138	378	232	271	60	357	277	264	438	214	217
总排序	11	15	12	7	16	13	14	2	8	5	17	3	4	6	1	10	9

数学建模竞赛中用到综合评价方法

1）1993-B:足球队排名问题；
2）2001-B:公交车调度问题；
3）2002-B:彩票中的数学问题；
4）2004-D:公务员招聘问题；
5）2005-A:长江水质的评价和预测问题；
6）2005-C:雨量预报方法评价问题；
7）2006-B:艾滋病疗法评价与预测问题；
8）2007-C:手机“套餐”优惠几何问题；
9）2008-B:高教学费标准探讨问题；
10）2008-D:NBA赛程的分析与评价问题；
11）2009-D:会议筹备问题。
12）2010-B：上海世博会影响力的定量评估
13）2010-D题 对学生宿舍设计方案的评价
14）2010-C题 输油管的布置
15）2011-B题 交巡警服务平台的设置与调度
16）2015-B题 “互联网+”时代的出租车资源配置
17）2016-B题 小区开放对道路通行的影响
18）2017-B题 “拍照赚钱”的任务定价

2003—A题 SARS的传播
案例：SARS疫情对某些经济指标影响

- 1 问题的提出：2003年的SARS疫情对中国部分行业的经济发展产生了一定的影响，特别是对帮分疫情较严重的省市的相关行业所造成的影响是明显的，经济影响主要分为直接经济影响和间接影响．直接经济影响涉及到商品零售业、旅游业、综合服务等行业．很多方面难以进行定量地评估，现仅就SARS疫情较重的某市商品零售业、旅游业和综合服务业的影响进行定量的评估分析．
- 究竟SARS疫情对商品零售业、旅游业和综合服务业的影响有多大，已知该市从1997年1月到2003年10月的商品零售额、接待旅游人数和综合服务收入的统计数据如下表1、表2、表3．

表1 商品的零售额(单位：亿元)

年代	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月
1997	83.0	79.8	78.1	85.1	86.6	88.2	90.3	86.7	93.3	92.5	90.9	96.9
1998	101.7	85.1	87.8	91.6	93.4	94.5	97.4	99.5	104.2	102.3	101.0	123.5
1999	92.2	114.0	93.3	101.0	103.5	105.2	109.5	109.2	109.6	111.2	121.7	131.3
2000	105.0	125.7	106.6	116.0	117.6	118.0	121.7	118.7	120.2	127.8	121.8	121.9
2001	139.3	129.5	122.5	124.5	135.7	130.8	138.7	133.7	136.8	138.9	129.6	133.7
2002	137.5	135.3	133.0	133.4	142.8	141.6	142.9	147.3	159.6	162.1	153.5	155.9
2003	163.2	159.7	158.4	145.2	124	144.1	157	162.6	171.8	180.7	173.5	176.5

表2 接待海外旅游人数(单位：万人)

年代	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	l1月	12月
1997	9.4	11.3	16.8	19.8	20.3	18.8	20.9	24.9	24.7	24.3	19.4	18.6
1998	9.6	11.7	15.8	19.9	19.5	17.8	17.8	23.3	21.4	24.5	20.1	15.9
1999	10.1	12.9	17.7	21.0	21.0	20.4	21.9	25.8	29.3	29.8	23.6	16.5
2000	11.4	26.0	19.6	25.9	27.6	24.3	23.0	27.8	27.3	28.5	32.8	18.5
2001	11.5	26.4	20.4	26.1	28.9	28.0	25.2	30.8	28.7	28.1	22.2	20.7
2002	13.7	29.7	23.1	28.9	29.0	27.4	26.0	32.2	31.4	32.6	29.2	22.9
2003	15.4	17.1	23.5	11.6	1.78	2.61	8.8	16.2	20.1	24.9	26.5	21.8

⑩ **试根据这些历史数据建立预测评估模型，评估2003年SARS疫情给该市的商品零售业、旅游业和综合服务业所造成的影响.**

按时态分，数据可分为历史数据和现实数据；按预测对象分，可分为内部数据和外部数据；就收集的手段分，可分为第一手数据和第二手数据。

第一手数据，包括以各种形式初次收集的数据。收集第一手数据的途径包括：抽样调查，连续调查，或全面调查。

第二手数据多为已经公布和发表的资料，易于获取，代价低，数据精度也有一定的保证。其缺点是数据可能不能直接适用于预测情况。因此，常常需要对已公布的数据进行修正和处理，使其适应于预测需要。

数据处理方法

（1）判别法

通过对历史数据的判断，选择其中可代表整个预测过程中很可能发生的模式的数据作为建模数据；

（2）剔除法

如果数据量比较大，且非必须具备连续的数据量，这时可剔除数据中受随机干扰的异常值；

（3）平均值法

在数据比较少或需要连续数据时，则可采取平均值法对数据进行处理。对于时间序列数据，可用异常值前后两期数据的算术平均值或几何平均值对异常值进行修正。

$$\overline{x_t} = \frac{x_{t-1} + x_{t+1}}{2} \quad \text{或}$$

$$\overline{\mathrm{x}_{\mathrm{t}}} = \sqrt{x_{t-1} \bullet x_{t+1}}$$

通常当历史数据的发展趋势呈线性时，取算求平均值，当发展趋势呈非线性时，取几何平均值。

（4）拉平法

由于条件发生变化，常常使一些历史数据不能反映现时的情况，例如，大型钢铁厂、化肥厂、或油气田的建成投产或开发，可以使产量猛增，这时历史数据将发生突变，出现一个转折，如用这类数据建模，则需要处理。这时拉平法是一种较好的方法。它的原理是对转折点前的数据加一个适当的量值，使其与折点后的数据走向一致。

1、时间序列预测模型

时间序列模型主要研究事物的自身发展规律，借以预测事物的未来趋势。主要方法有移动平均、指数平滑、分解预测、鲍克斯詹金斯模型、多变量模型以及类推法等。

特点和应用范围时间序列一般指一组按时间顺序排列的数据，展示了研究对象在一定时期的发生变化过程。时间序列模型，就是根据预测对象时间变化特征，研究事物自身的发展规律，探讨未来发展趋势，是一种重要的定量预测方法，包括多种模型，主要适用于经济预测、商业预测、需求预测、库存预测等，预测期限主要为中、短期，不适用于有拐点的长期预测。

2、微分方程预测模型

当我们描述实际对象的某些特性随时间（或空间）而演变的过程、分析它的变化规律、预测它的未来性态、研究它的控制手段时，通常要建立对象的动态微分方程模型。微分方程大多是物理或几何方面的典型问题，假设条件已经给出，只需用数学符号将已知规律表示出来，即可列出方程，求解的结果就是问题的答案，答案是唯一的，但是有些问题是非物理领域的实际问题，要分析具体情况或进行类比才能给出假设条件。作出不同的假设，就得到不同的方程。比较典型的有：传染病的预测模型、经济增长预测模型、正规战与游击战的预测模型、药物在体内的分布与排除预测模型、人口的预测模型、烟雾的扩散与消失预测模型以及相应的同类型的预测模型。其基本规律随着时间的增长趋势是指数的形式，根据变量的个数建立初等微分模型：

$$\begin{cases} \dfrac{\mathrm{d}x}{\mathrm{d}t} = rx \\ x(0) = x_0 \end{cases}$$ （如传染病预测，经济增长预测，人口预测等）

或者微分方程组

$$\begin{cases} (\dot{x}(t)) = A(x(t)) \\ (\dot{x}(0)) = B \end{cases}$$ （如正规战与游击战、药物的分布与排除预测等）

其中$(\dot{x}(t))$，$(x(t))$均是列向量，A、B是矩阵。

而后由于实际问题的改变，会出现外在的干预等，例如传染病模型，只有健康人才可能被传染为病人，病人治愈后仍有可能成为病人或者治愈后有免疫力，政府卫生部门的干预等，都会使得所建立的初等模型失败。为此根据情况可以适当地一步步改进所建立的初等模型，从而达到我们所需要的微分方程预测模型。改进包括：

（1）常系数的改进 $\begin{cases} \dfrac{dx}{dt} = r(t)x \\ x(0) = x_0 \end{cases}$ 和 $\begin{cases} (\dot{x}(t)) = A(t)(x(t)) \\ (\dot{x}(0)) = B \end{cases}$

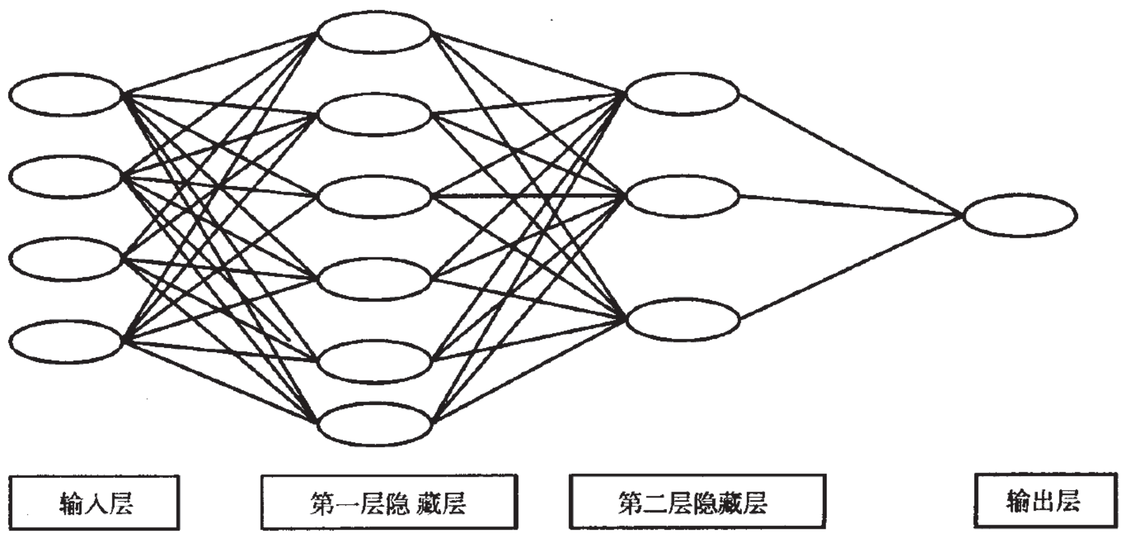
（2）增加一个控制函数 $\begin{cases} \dfrac{dx}{dt} = \mathrm{r}x + f(t) \\ x(0) = x_0 \end{cases}$ 和 $\begin{cases} (\dot{x}(t)) = A(x(t)) + (f(t)) \\ (\dot{x}(0)) = B \end{cases}$

（3）综合 $\begin{cases} \dfrac{dx}{dt} = r(t)x + f(t) \\ x(0) = x_0 \end{cases}$ 和 $\begin{cases} (\dot{x}(t)) = A(t)(x(t)) + (f(t)) \\ (\dot{x}(0)) = B \end{cases}$

以及一些相应的根据情况而进行的改进模型。得到模型后可以用**Matlab**软件来求解画出散点图，并比较拟合度。

3、神经网络预测模型

BP神经网络模型是目前神经网络学习模型中最具代表性、应用最普遍的模型。BP神经网络架构是由数层互相连结的神经元组成，通常包含了输入层、输出层及若干隐藏层，各层包含了若干神经元。



适用于中长期的预测；优点是逼近效果好，计算速度快，不需要建立数学模型，精度高；具有强非线性拟合能力。缺点是无法表达和分析被预测系统的输入和输出间的关系，预测人员无法参与预测过程；收敛速度慢，难以处理海量数据，得到的网络容错能力差，算法不完备（易陷入局部极小）。

4、灰色预测模型

（1）灰色系统、白色系统和黑色系统

- **白色系统是指一个系统的内部特征是完全已知的，即系统的信息是完全充分的。**

- **黑色系统**是指一个系统的内部信息对外界来说是一无所知的，只能通过它与外界的联系来加以观测研究。

- **灰色系统**是指一个系统内的一部分信息是已知的，另一部分信息是未知的，系统内各因素间有不确定的关系。

（2）灰色预测法

- 灰色预测法是一种对含有不确定因素的系统进行预测的方法。

- 灰色预测是对既含有已知信息又含有不确定信息的系统进行预则，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。

- 灰色预测通过鉴别系统因素之间发展趋势的相异程度，即进行关联分析，并对 原始数据进行生成处理来寻找系统变动的规律，生成有较强规律性的数据序列，然后建立相应的微分方程模型，从而预测事物未来发展趋势的状况。

灰色系统理论认为，尽管客观表象复杂，但总是有整体功能的，因此必然蕴含某种内在规律。关键在于如何选择适当的方式去挖掘和利用它。灰色系统是通过对原始数据的整理来寻求其变化规律的，这是一种就数据寻求数据的现实规律的途径，即为灰色序列的生成。一切灰色序列都能通过某种生成弱化其随机性，显现其规律性。数据生成的常用方式有累加生成、累减生成和加权累加生成。

（1）累加生成

把数列各项（时刻）数据依次累加的过程称为累加生成过程（AGO）。由累加生成过程所得的数列称为累加生成数列。设原始数列为

$$x^{(0)}=(x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n))$$

令

$$x^{(1)}(k)=\sum_{i=1}^{k}x^{(0)}(i),k=1,2,\cdots,n,$$

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

称所得到的新数列为数列$x^{(0)}$的1次累加生成数列。类似地有

$$x^{(r)}(k)=\sum_{i=1}^{k}x^{(r-1)}(i),k=1,2,\cdots,n,r\geq 1$$

称为 $x^{(0)}$的r次累加生成数列。

（2）累减生成

对于原始数据列依次做前后相邻的两个数据相减的运算过程称为累减生成过程IAGO。如果原始数据列为

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

令

$$x^{(0)}(k) = x^{(1)}(k) - x^{(1)}(k-1), k = 2,3,\cdots,n,$$

称所得到的数列$x^{(0)}$为 $x^{(1)}$的1次累减生成数列。

注：从这里的记号也可以看到，从原始数列 $x^{(0)}$，得到新数列 $x^{(1)}$，再通过累减生成可以还原出原始数列。实际运用中在数列 $x^{(1)}$ 的基础上预测出 $\hat{x}^{(1)}$，通过累减生成得到预测数列 $\hat{x}^{(0)}$ 。

（3）加权邻值生成

设原始数列为 $x^{(0)}=(x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n))$

称 $x^{(0)}(k-1),x^{(0)}(k)$ 为数列 $x^{(0)}$ 的邻值。

$x^{(0)}(k-1)$ 为后邻值，$x^{(0)}(k)$ 为前邻值，对于常

数 $\alpha\in[0,1]$，令 $z^{(0)}(k)=\alpha x^{(0)}(k)+(1-\alpha)x^{(0)}(k-1),k=2,3,\cdots,n,$

由此得到的数列 $x^{(0)}$ 称为数列 $z^{(0)}$ 在权 α 下的邻值生成数，权 α 也称为生成系数。

特别地，当生成系数 $\alpha=0.5$ 时，则称

$$z^{(0)}(k)=0.5x^{(0)}(k)+0.5x^{(0)}(k-1),k=2,3,\cdots,n,$$

为均值生成数，也称等权邻值生成数。

累加生成计算示例

例： $x^{(0)}=(x^{(0)}(k)\ |\ k=1, 2, 3, 4, 5)$

$=x^{(0)}(1), x^{(0)}(2), x^{(0)}(3), x^{(0)}(4), x^{(0)}(5)$

$=(3.2, 3.3, 3.4, 3.6, 3.8)$

求 $x^{(1)}(k)$

解：

$$k=1, x^{(1)}(1)=x^{(0)}(1)=3.2$$

$$k=2, x^{(1)}(2)=\sum_{i=1}^{2}x^{(0)}(i)=x^{(0)}(1)+x^{(0)}(2)=3.2+3.3=6.5$$

$$k=3, x^{(1)}(3)=\sum_{i=1}^{3}x^{(0)}(i)=x^{(1)}(2)+x^{(0)}(3)=6.5+3.4=9.9$$

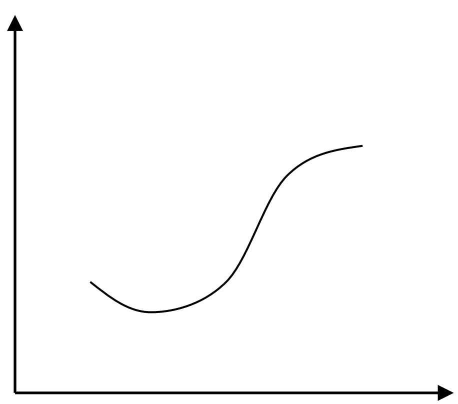
$$k=4, x^{(1)}(4)=\sum_{i=1}^{4}x^{(0)}(i)=x^{(1)}(3)+x^{(0)}(4)=9.9+3.6=13.5$$

$$k=5, x^{(1)}(5)=\sum_{i=1}^{5}x^{(0)}(i)=x^{(1)}(4)+x^{(0)}(5)=13.5+3.8=17.3$$

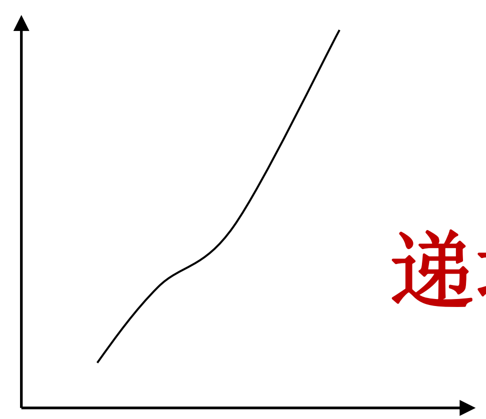
累加生成的特点

一般经济数列都是非负数列。累加生成能使任意非负数列、摆动的与非摆动的，转化为非减的、递增的。

原始数列作图



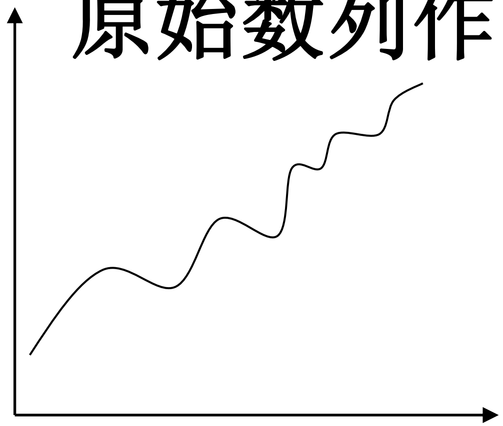
1—AGO作图



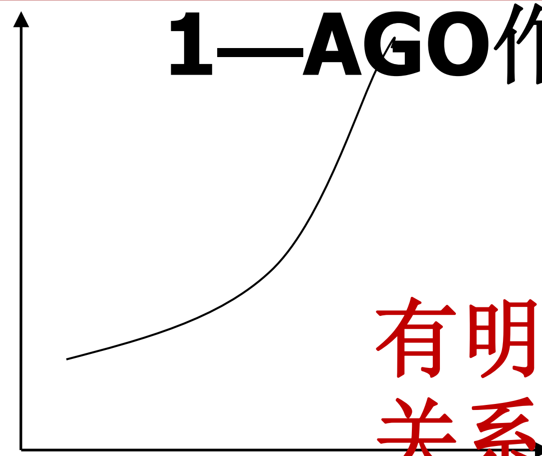
递增的规律

某市的汽车销售量

原始数列作图

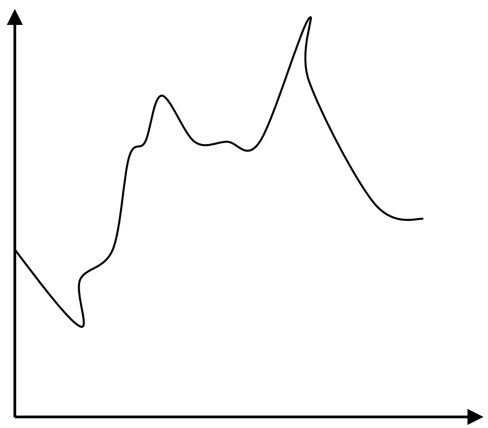


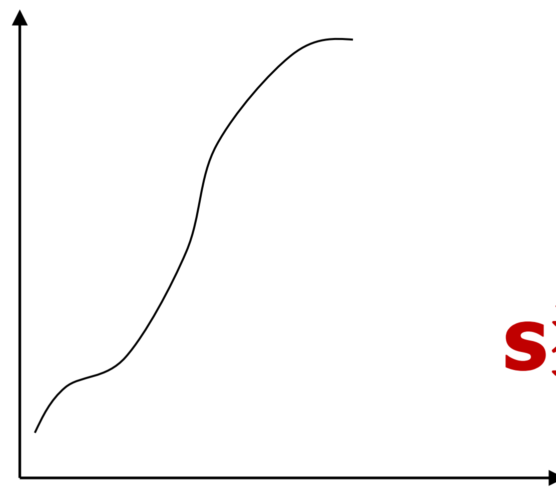
1—AGO作图



有明显的指数关系的规律

某地区作物产量





s型变化规律

某钢厂产量

累减生成计算示例

$x^{(1)}=(x^{(1)}(1),x^{(1)}(2),x^{(1)}(3),x^{(1)}(4),x^{(1)}(5),x^{(1)}(6))=(5,9,14,24,35,46)$

解： $x^{(0)}(k)=x^{(1)}(k)-x^{(1)}(k-1)$

若$k=0,x^{(1)}(0)=0$

$k=1,x^{(0)}(1)=x^{(1)}(1)-x^{(1)}(0)=x^{(1)}(1)=5$

$k=2,x^{(0)}(2)=x^{(1)}(2)-x^{(1)}(1)=4$

$k=3,x^{(0)}(3)=x^{(1)}(3)-x^{(1)}(2)=5$

$k=4,x^{(0)}(4)=x^{(1)}(4)-x^{(1)}(3)=10$

$k=5,x^{(0)}(5)=x^{(1)}(5)-x^{(1)}(4)=11$

$k=6,x^{(0)}(6)=x^{(1)}(6)-x^{(1)}(5)=11$

从而有： IGAO（$x^{(0)}$）$=(5,4,5,10,11,11)$

不难看出，累减生成具有求导性质，这是因为

$$\frac{dx(k)}{dt}=\lim_{\Delta t\to 0}\frac{x(k)-x(k-\Delta t)}{\Delta t}$$

而$\alpha^{(1)}(x(k))=x(k)-x(k-1)$,相当于$\Delta t=1$

灰色模型G M(1,1)

灰色系统理论是基于关联空间、光滑离散函数等概念定义灰导数与灰微分方程，进而用离散数据列建立微分方程形式的动态模型，即灰色模型是利用离散随机数经过生成变为随机性被显著削弱而且较有规律的生成数，建立起的微分方程形式的模型，这样便于对其变化过程进行研究和描述。

- **G**表示**grey**（灰色），**M**表示**model**（模型）

- 设 $x^{(0)}=(x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n))$ 为原始数列，其1次累加生成数列为

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

- 其中

$$x^{(1)}(k)=\sum_{i=1}^{k}x^{(0)}(i),k=1,2,\cdots,n,$$

- 定义 $x^{(1)}$ 的灰导数为

$$d(k)=x^{(0)}(k)=x^{(1)}(k)-x^{(1)}(k-1).$$

令 $z^{(1)}$ 为数列 $x^{(1)}$ 的邻值生成数列，即

$$z^{(1)}(k)=\alpha x^{(1)}(k)+(1-\alpha)x^{(1)}(k-1),$$

于是定义GM（1，1）的灰微分方程模型为

$$d(k)+az^{(1)}(k)=b,$$

即或 $x^{(0)}(k)+az^{(1)}(k)=b,$ **(1)**

在式（1）中，$x^{(0)}(k)$ **称为灰导数，a称为发展系数，**

$z^{(1)}(k)$ **称为白化背景值，b称为灰作用量。**

将时刻表 $k=2,3,\cdots,n$ **代入（1）式有**

$$\begin{cases} x^{(0)}(2)+az^{(1)}(2)=b, \\ x^{(0)}(3)+az^{(1)}(3)=b, \\ \cdots\cdots\cdots\cdots \\ x^{(0)}(n)+az^{(1)}(n)=b, \end{cases}$$

引入矩阵向量记号：

$$u=\begin{bmatrix} a \\ b \end{bmatrix} \qquad Y=\begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix} \qquad \mathbf{B}=\begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix}$$

于是GM（1，1）模型可表示为 $Y = \mathbf{B}u.$

现在问题归结为求 a, b 值。用一元线性回归，即最小二乘法求它们的估计值为

$$\hat{u} = \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = (B^T B)^{-1} B^T Y.$$

注：实际上回归分析中求估计值是用软件计算的，有标准程序求解，如matlab等。

GM（1，1）的白化型

对于GM（1，1）的灰微分方程（1），如果将灰导数 $x^{(0)}(k)$ 的时刻 $k = 2,3,\cdots,n$ 视为连续变量 *t*，则 $x^{(1)}$ 视为时间 *t* 的函数 $x^{(1)}(t)$ ，于是 $x^{(0)}(k)$ 对应于导数量级 $\frac{dx^{(1)}(t)}{dt}$ ，白化背景值 $z^{(1)}(k)$ 对应于导数 $x^{(1)}(t)$ 于是GM（1,1）的灰微分方程对应于的白微分方程为

$$\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = b, \qquad (2)$$

GM（1,1）灰色预测的步骤

1.数据的检验与处理

为了保证GM（1，1）建模方法的可行性，需要对已知数据做必要的检验处理。

设原始数据列为了 $x^{(0)}=(x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n))$ ，计算数列的级比:

$$\lambda(k)=\frac{x^{(0)}(k-1)}{x^{(0)}(k)},k=2,3,\cdots,n.$$

如果所有的级比都落在可容覆盖区间 $X=(e^{\frac{-2}{n+1}},e^{\frac{2}{n+1}})$ 内，则数据列 $x^{(0)}$ 可以建立GM（1，1）模型且可以进行灰色预测。否则，对数据做适当的变换处理，如平移变换：

取C使得数据列

$$y^{(0)}(k)=x^{(0)}(k)+c,k=1,2,\cdots,n,$$

的级比都落在可容覆盖内。

2. 建立GM（1,1）模型

不妨设 $x^{(0)}=(x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n))$ **满足上面的要求，以它为数据列建立GM（1,1）模型**

$$x^{(0)}(k)+az^{(1)}(k)=b,$$

用回归分析求得a, b的估计值，于是相应的白化模型为

$$\frac{dx^{(1)}(t)}{dt}+ax^{(1)}(t)=b,$$

解为 $$x^{(1)}(t)=(x^{(0)}(1)-\frac{b}{a})e^{-a(t-1)}+\frac{b}{a}. \qquad (3)$$

于是得到预测值

$$\hat{x}^{(1)}(k+1)=(x^{(0)}(1)-\frac{b}{a})e^{-ak}+\frac{b}{a},k=1,2,\cdots,n-1,$$

从而相应地得到预测值：

$$\hat{x}^{(0)}(k+1)=\hat{x}^{(1)}(k+1)-\hat{x}^{(1)}(k),k=1,2,\cdots,n-1,$$

3. 检验预测值

（1）残差检验：计算相对残差

$$\varepsilon(k)=\frac{x^{(0)}(k)-\hat{x}^{(0)}(k)}{x^{(0)}(k)},k=1,2,\cdots,n,$$

如果对所有的 $|\varepsilon(k)|<0.1$ **，则认为达到较高的要求；**

否则，若对所有的 $|\varepsilon(k)|<0.2$ **，则认为达到一般要求。**

（2）级比偏差值检验：计算

$$\rho(k)=1-\frac{1-0.5a}{1+0.5a}\lambda(k),$$

如果对所有的 $|\rho(k)|<0.1$**，则认为达到较高的要求；**

否则，若对所有的 $|\rho(k)|<0.2$ **，则认为达到一般要求。**

单击此处编辑母版标题样式

- 例 南方某城市2000～2006 年道路交通噪声平均声级数据见右表

序号	年份	eq L
1	2000	71.1
2	2001	72.4
3	2002	72.4
4	2003	72.1
5	2004	71.4
6	2005	72.0
7	2006	71.6

第一步: 级比检验
建立交通噪声平均声级数据时间序列如下：

$$x^{(0)}=(x^{(0)}(1),x^{(0)}(2),\cdots x^{(0)}(7))$$

=(71.1, 72.4, 72.4, 72.1, 71.4, 72.0, 71.6)

- （1）求级比λ (k) $\lambda(k)=\dfrac{x^{(0)}(k-1)}{x^{(0)}(k)}$
- $\lambda=(\lambda(2),\lambda(3),\cdots,\lambda(7))$

=(0.982,1,1.0042,1.0098, 0.9917,1.0056)

（2）级比判断

由于所有的λ (k)∈[0.982,1.0098]，k = 2,3,…,7，故可以用x(0)作满意的GM（1，1）建模。

第二步: GM（1，1）建模

（1）对原始数据 $x^{(0)}$作一次累加，即

$x^{(1)}$= (71.1,143.5,215.9,288,359.4,431.4,503）

（2）为求a，b值，构造数据矩阵B 及数据向量Y

$$B=\begin{bmatrix} -\frac{1}{2}(x^{(1)}(1)+x^{(1)}(2)) & 1 \\ -\frac{1}{2}(x^{(1)}(2)+x^{(1)}(3)) & 1 \\ \vdots & \vdots \\ -\frac{1}{2}(x^{(1)}(6)+x^{(1)}(7)) & 1 \end{bmatrix}, Y=\begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(7) \end{bmatrix}$$

（3）采用最小二乘法计算

$$\hat{u}=(\hat{a},\hat{b})^T=(B^T\cdot B)^{-1}B^TY=\begin{pmatrix}0.0023\\72.6573\end{pmatrix}$$

于是得到a = 0.0023，b = 72.6573。

（4）建立模型

$$\frac{dx^{(1)}}{dt}+0.0023x^{(1)}=72.6573$$

⑩ 求解得

$$x^{(1)}(k+1)=(x^{(0)}(1)-\frac{b}{a})e^{-ak}+\frac{b}{a}=-30929e^{-0.0023k}+31000$$

（5）求生成数列值 $\hat{x}^{(1)}(k+1)$ 及模型还原值 $\hat{x}^{(0)}(k+1)$ ：

令k = 1,2,3,4,5,6,由上面的时间响应函数可算得 $\hat{x}^{(1)}$ ，其中取

$$\hat{x}^{(1)}(1)=\hat{x}^{(0)}(1)=x^{(0)}(1)=71.1,$$

由 $\hat{x}^{(0)}(k)=\hat{x}^{(1)}(k)-\hat{x}^{(1)}(k-1)$, 取k = 2,3,4, … ,7，得

$$\hat{x}^{(0)}=(\hat{x}^{(0)}(1),\ \hat{x}^{(0)}(2),\cdots,\ \hat{x}^{(0)}(7))=$$

(71.1, 72.4, 72.2, 72.1, 71.9, 71.7, 71.6)

- 第三步: 模型检验
- 模型的各种检验指标值的计算结果见表 7.
- 表7 GM(1,1)模型检验表

序号	年份	原始值	模型值	残差	级比偏差
1	2000	71.1	71.1	0	
2	2001	72.4	72.4	-0.0057	0.0023
3	2002	72.4	72.2	0.1638	0.0203
4	2003	72.1	72.1	0.0329	-0.0018
5	2004	71.4	71.9	-0.4984	-0.0074
6	2005	72.0	71.7	0.2699	0.0107
7	2006	71.6	71.6	0.0378	-0.0032

- 经验证，该模型的精度较高，可进行预测和预报。

2003-SARS传播问题 模型的分析

- 根据所掌握的历史统计数据可以看出，在正常情况下，全年的平均值较好地反映了相关指标的变化规律，这样可以把预测评估分成两部分：
- (i)利用灰色理论建立灰微分方程模型，由1997~2002年的平均值预测2003年平均值；
- (ii)通过历史数据计算每个月的指标值与全年总值的关系，从而可预测出正常情况下2003年每个月的指标值，再与实际值比较可以估算出SARS疫情实际造成的影响.

模型假设

- 给出下面两条假设：
- (1)假设该市的统计数据都是可靠准确的；
- (2)假设该市在SARS疫情流行期间和结束之后，数据的变化只与SARS疫情的影响有关，不考虑其他随机因素的影响.

建立灰色预测模型GM(1, 1)

- 由已知数据，对于1997~2002年某项指标记为矩阵 $A=(a_{ij})_{6\times 12}$ 计算每年的年平均值，记为 $x^{(0)}=(x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(6))$
- 并求级比 $\sigma(i)=x^{(0)}(i-1)/x^{(0)}(i)\in(0.7515,1.3307)(i=2,3,\cdots,6)$
- 对 $x^{(0)}$ 作一次累加，则
- $x^{(1)}(1)=x^{(0)}(1),x^{(1)}(i)=\sum_{k=1}^{i}x^{(0)}(k)(i=2,3,\cdots,6),$
- $$x^{(1)}=(x^{(1)}(1),x^{(1)}(2),\cdots,x^{(1)}(6))$$
- 取 $x^{(1)}$ 的加权均值，则

$$z^{(1)}(k)=\alpha x^{(1)}(k)+(1-\alpha)x^{(1)}(k-1)(k=2,3,\cdots,6),\alpha$$

为确定参数 ,于是GM(1,1)的白化微分方程模型为

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

其中a是发展灰度,b是内生控制灰度.

- **由于** $x^{(1)}(k)-x^{(1)}(k-1)=x^{(0)}(k)$**，取** $x^{(0)}(k)$ **为灰导数，为** $z^{(1)}(k)$ **背景值，则建立灰微分方程为:**
- $$x^{(0)}(k)+az^{(1)}(k)=b(k=2,3,\cdots,6)$$

或 $$x^{(0)}(k)=-az^{(1)}(k)+b(k=2,3,\cdots,6)$$

- **其矩阵形式为** $Y^{(0)}=B\cdot(a,b)^T$**，**
- **其中**

$$Y^{(0)}=(x^{(0)}(2),x^{(0)}(3),\cdots,x^{(0)}(6))^T,B=\begin{pmatrix}-z^{(1)}(2) & -z^{(1)}(3) & \cdots & -z^{(1)}(6)\\ 1 & 1 & \cdots & 1\end{pmatrix}^T$$

用最小二乘法求得参数的估计值为

$$(\hat{a},\hat{b})^T=(B^T\cdot B)^{-1}\cdot B^T\cdot Y^{(0)}.$$

则灰微分方程模型(4)的解为 $\hat{x}^{(1)}(t+1)=(x^{(0)}(1)-\frac{b}{a})\cdot e^{-at}+\frac{b}{a},$

则 $\hat{x}^{(0)}(k+1)=\hat{x}^{(1)}(k+1)-\hat{x}^{(1)}(k)=(x^{(0)}(1)-\frac{b}{a})\cdot(e^{-ak}-e^{-a(k-1)})$

由(7)式可以得到2003年的平均值为 $\overline{x}$**，则预测2003年的总值为** $X=12\cdot\overline{x}.$ **根据历史数据，可以统计计算出2003年第i个月的指标值占全年总值的比例为** u_i **，即**

$$u_i=\sum_{j=1}^{6}a_{ij}\Big/\sum_{i=1}^{12}\sum_{j=1}^{6}a_{ij}\qquad(i=1,2,\cdots,12),$$

则 $u=(u_1,u_2,\cdots,u_{12})$**,于是可得2003年每一个月的指标值为Y=X •u.**

4 模型的求解

(i)商品零售额

- 由数据表1，计算可得每年月平均值、一次累加值分别为

$x^{(0)}$=(87.6167，98.5000，108.4750，118.4167，132.8083，145.4083)，

$x^{(1)}$=(87.6167，186.1167，294.5917，413.0083，545.8167，691.2250)．

- 显然 $x^{(0)}$ 的所有级比都在可行域内．经检验，在这里取参数 $\alpha=0.4$ 比较合适， 则有 $z^{(1)}=$ (127.0167，229.5067，341.9583，466.1317，603.9800)．

(i)商品零售额

- 由最小二乘法求得a=-0.0993，b=85.5985．可得2003年的月平均值为

$\overline{x}$ =162.8826亿元；年总值为 $X = 12 \cdot \overline{x}$.

=1954.6亿元.　　由(8)式得每月的比例为

u=(0.0794，0.0807，0.0749，0.0786，0.0819， 0.0818，0.0845，

0.0838，0.0872，0.0886， 0.0866，0.0920)

- 故2003年1—12月的预测值为
- Y=u•X =(155.2，157.8，146.4，153.6，160.1， 159.9，165.2，

163.8，17o.5，173.2， 169.3，179.9)(亿元)

⑩ 将预测值与实际统计值进行比较如下表所示.

月 份 1月 2月 3月 4月 5月 6月 7月 8月 9月 10月11月 12月

预测值 155.2 157.8 146.4 153.6 160.1 159.9 165.2 163.8 170.5 173.2 169.3 179.9

实际值 163.2 159.7 158.4 145.2 124.0 144.1 157.0 162.6 171.8 180.7 173.5 176.5

```
clc,clear
han1=[83.0 79.8 78.1 85.1 86.6 88.2 90.3 86.7 93.3 92.5 90.9 96.9
101.7 85.1 87.8 91.6 93.4 94.5 97.4 99.5 104.2 102.3 101.0 123.5
92.2 114.0 93.3 101.0 103.5 105.2 109.5 109.2 109.6 111.2 121.7 131.3
105.0 125.7 106.6 116.0 117.6 118.0 121.7 118.7 120.2 127.8 121.8 121.9
139.3 129.5 122.5 124.5 135.7 130.8 138.7 133.7 136.8 138.9 129.6 133.7
137.5 135.3 133.0 133.4 142.8 141.6 142.9 147.3 159.6 162.1 153.5 155.9
163.2 159.7 158.4 145.2 124.0 144.1 157.0 162.6 171.8 180.7 173.5 176.5];
han1(end,:)=[];m=size(han1,2);
x0=mean(han1,2);
x1=cumsum(x0);
alpha=0.4;n=length(x0);
z1=alpha*x1(2:n)+(1-alpha)*x1(1:n-1)
Y=x0(2:n);B=[-z1,ones(n-1,1)];
ab=B\Y
k=6;
x7hat=(x0(1)-ab(2)/ab(1))*(exp(-ab(1)*k)-exp(-ab(1)*(k-1)))
z=m*x7hat
u=sum(han1)/sum(sum(han1))
v=z*u
```

- x1 = [87.6167 186.1167 294.5917 413.0083 545.8167 691.2250]
- z1 = [127.0167 229.5067 341.9583 466.1317 603.9800]
- ab = [-0.0993 85.5985]
- x7hat = 162.8793 z = 1.9546e+003

- u = [0.0794 0.0807 0.0749 0.0786 0.0819 0.0818 0.0845 0.0838 0.0872 0.0886 0.0866 0.0920]

- v =[155.2152 157.7365 146.4023 153.5421
- 160.1400 159.8337 165.0649 163.7924
- 170.5317 173.1473 169.3064 179.8394]

5 模型的结果分析

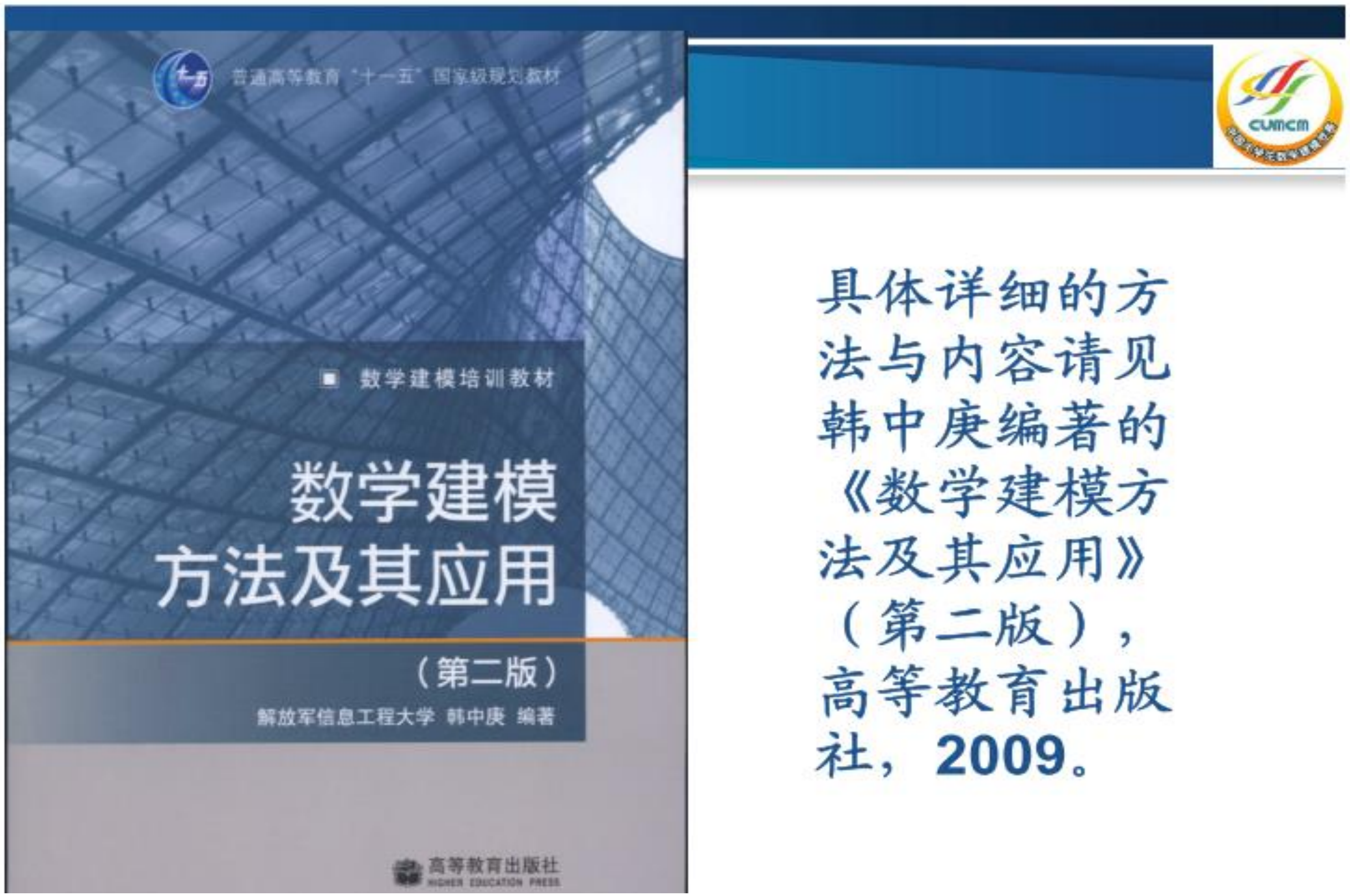
- 根据该市的统计报告显示，2003年4、5、6三个月的实际商品零售额分别为145.2，124、144.1亿元．在这之前，根据统计部门的估计4、5、6三个月份SARS疫情对该市的商品零售业的影响最为严重，这三个月估计大约损失62亿元左右．从我们的模型预测结果来计算，4、5、6三个月的损失为60.3亿元，这个数据基本与专家的估算值相符，8月份基本恢复正常，这也说明了模型的正确性和可靠性．

- 对于旅游业来说是受影响最严重的行业之一，最严重的4、5、6、7四个月就损失100多万人，按最新统计数据，平均每人消费1002美元计算，大约损失1O亿美元．全年大约损失160万人，约合16亿美元，到年底基本恢复正常．

- 对于综合服务业中的部分行业影响较大，如航空交通运输、宾馆餐饮等，但有些行业影响不大，如电信、通讯等，总平均来看，影响还不算太大，5、6、7、8四个月大约损失70亿元.
- 从预测结果可以看出，虽然下半年没有发生疫情，但人们一直担心SARS会卷土重来，所以，对这些行业还是有一定的影响，即SARS影响的延续性的作用.
- 该模型虽是就某经济指标的发展规律进行评估预测而建立的，但类似的也适用于其他方面的一些

  数据规律的评估预测问题，即该模型具有很广泛的应用性.

- 参考文献：
- [1] 韩中庚，数学建模方法及应用（第二版），2009，北京：高等教育出版社。



欢迎批评指正

谢　谢！