

数学建模模型算法精讲课——

# 灰色关联分析

—— 江北老师

人若有志，  
万事可为！

## 灰色关联分析

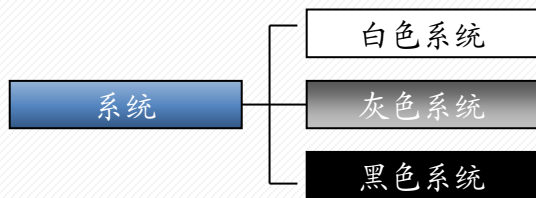
- 模型引出
- 模型原理
- 典型例题
- 相关代码





## ➤ 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



## ➤ 什么是关联分析？

- 所谓关联分析，就是**系统地分析因素**。回答的问题是：某个包含多种因素的系统，哪些因素是主要的，哪些是次要的；哪些因素影响大，哪些因素影响小；哪些因素是明显的，哪些因素是潜在的；哪些是需要发展的，那些需要抑制……
- 现有因素分析的量化方法，大都是数理统计法如回归分析、方差分析、主要成分分析等，这些方法都有下述弱点：
  - (1) 要求大量数据，数据量少难以找到统计规律
  - (2) 要求分布是典型的（线性的、指数的或对数的），即使是典型的并非都能处理
  - (3) 计算工作量大，一般需要计算机帮助
  - (4) 有时可能出现反常情况，如正相关则断为负相关，以至正确现象受到歪曲和颠倒
- 尤其是我国统计数据十分有限，而且现有数据灰度较大，许多数据都出现几次大起大落，没有典型的分布规律。因此，采用数理统计方法往往难以奏效。



## ➤ 灰色关联分析

- 灰色关联度分析 (Grey Relation Analysis, GRA)，是一种多因素统计分析的方法。灰色关联分析方法弥补了采用数理统计方法作系统分析所导致的缺憾。它对**样本量的多少和样本有无规律都同样适用**，而且计算量小，十分方便，更不会出现量化结果与定性分析结果不符的情况。
- 灰色关联分析的基本思想是**根据序列曲线几何形状的相似程度**来判断其联系是否紧密。曲线越接近，相应序列之间的关联度就越大，反之就越小。
- 对一个抽象的系统或现象进行分析，首先要选准反映**系统行为特征的数据序列**，称为找系统行为的映射量，用映射量来间接地表征系统行为。例如，用国民平均接受教育的年数来反映教育发达程度，用刑事案件的发案率来反映社会治安面貌和社会秩序，用医院挂号次数来反映国民的健康水平等。有了系统行为特征数据和相关因素的数据，即可作出各个序列的图形，从直观上进行分析。

曲线几何形态越相似



发展变化态势就越接近



其关联程度就越大



## ➤ 关联分析步骤

- 1、母序列(又称参考序列、母指标)

能反映系统行为特征的数据序列, 类似于因变量  $Y$ , 记为  $Y=[y_1, y_2, \dots, y_n]^T$

- 2、子序列(又称比较序列、子指标)

影响系统行为的因素组成的数据序列, 类似于自变量  $X$ , 记为

$$X_{nm} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}$$

- 3、数据预处理

由于不同要素具有不同量纲和数据范围, 因此我们要对他们进行预处理去量纲, 将他们统一到近似的范围内, 先求出每个指标的均值, 在用指标中的元素除以其均值

$$\widetilde{y_k} = \frac{y_k}{y_i}, \overline{y_i} = \frac{1}{n} \sum_{k=1}^n y_k \quad \widetilde{x_{ki}} = \frac{x_{ki}}{x_i}, \overline{x_i} = \frac{1}{n} \sum_{k=1}^n x_{ki} \quad (i=1, 2, \dots, m)$$



## ➤ 关联分析步骤

- 4、计算灰色关联系数

计算子序列中各个指标与母序列的关联系数

$$\text{记为: } a = \min_i \min_k |x_0(k) - x_i(k)|, \quad b = \max_i \max_k |x_0(k) - x_i(k)|$$

为两极最小差和最大差

$$\text{构造: } \xi_i(k) = y(x_0(k), x_i(k)) = \frac{a + \rho b}{|x_0(k) - x_i(k)| + \rho b} \quad \text{其中 } \rho \text{ 为分辨系数, 一般取 } 0.5$$

- 5、计算关联度

$$r_i = \frac{1}{n} \sum_{k=1}^n \xi_i(k) = \frac{1}{n} \sum_{k=1}^n y(x_0(k), x_i(k))$$



## ➤ 关联分析例题

- 1、已知某地国民生产总值，工业和农业生产总值，原始数据的形式及来源见下表，分析工业农业哪个对国民生产总值影响大

项目名称	年份			
	2016	2017	2018	2019
国民生产总值	55	65	75	100
工业产值	24	38	40	50
农业产值	10	22	18	20

1) 定义母序列及子序列如下

项目名称	年份				项目代号
	2016	2017	2018	2019	
国民生产总值	55	65	75	100	$X_0$ (母序列)
工业产值	24	38	40	50	$X_1$ (子序列)
农业产值	10	22	18	20	$X_2$ (子序列)





## ➤ 关联分析例题

### 2) 数据预处理

对数据进行均值化  $\widetilde{y}_k = \frac{y_k}{\overline{y_i}}, \overline{y_i} = \frac{1}{n} \sum_{k=1}^n y_k$   $\widetilde{x}_{ki} = \frac{x_{ki}}{\overline{x_i}}, \overline{x_i} = \frac{1}{n} \sum_{k=1}^n x_{ki} (i=1, 2, \dots, m)$

项目名称	年份				项目代号
	2016	2017	2018	2019	
国民生产总值	0.75	0.88	1.02	1.36	(母序列)
工业产值	0.63	1.00	1.05	1.32	(子序列)
农业产值	0.57	1.26	1.03	1.14	(子序列)

### 3) 求关联系数

$$a = \min_i \min_k |x_0(k) - x_i(k)|$$

$$b = \max_i \max_k |x_0(k) - x_i(k)|$$

$$\xi_i(k) = \gamma(x_0(k), x_i(k)) = \frac{a + \rho b}{|x_0(k) - x_i(k)| + \rho b}$$

$k$	$X_0$	$X_1$	$X_2$	$ x_0(k) - x_1(k) $	$ x_0(k) - x_2(k) $
1	0.75	0.63	0.57	<b>0.12</b>	0.18
2	0.88	1	1.26	0.12	<b>0.38</b>
3	1.02	1.05	1.03	<b>0.03</b>	<b>0.01</b>
4	1.36	1.32	1.14	0.04	0.22

$$a = 0.01$$

$$b = 0.38$$



## ➤ 关联分析例题

3) 求关联系数

$$a = 0.01 \quad b = 0.38$$

$$\begin{aligned}\xi_i(k) &= y(x_0(k), x_i(k)) = \frac{0.01 + 0.5 \times 0.38}{|x_0(k) - x_i(k)| + 0.5 \times 0.38} \\ &= \frac{0.2}{|x_0(k) - x_i(k)| + 0.19}\end{aligned}$$

$k$	$ x_0(k) - x_1(k) $	$ x_0(k) - x_2(k) $	$\xi_1$	$\xi_2$
1	0.12	0.18	0.645	0.541
2	0.12	0.38	0.645	0.351
3	0.03	0.01	0.909	1.000
4	0.04	0.22	0.870	0.488

3) 求关联度

$$r_i = \frac{1}{n} \sum_{k=1}^n \xi_i(k) = \frac{1}{n} \sum_{k=1}^n y(x_0(k), x_i(k))$$

$r_1 > r_2$  工业产值关联度更大

$k$	$\xi_1$	$\xi_2$
1	0.645	0.541
2	0.645	0.351
3	0.909	1.000
4	0.870	0.488
$r$	<b>0.767</b>	<b>0.595</b>



## ➤ 灰色关联综合评价

- 2、我们来换个方法给明星K选对象，还是A、B、C三位候选人（K觉得身高165是最好，体重在90-100斤是最好）

候选人	颜值	脾气 (争吵次数)	身高	体重
A	9	10	175	120
B	8	7	164	80
C	6	3	157	90

1) 数据正向化处理（可看TOPSIS一节）

将原始矩阵正向化，就是要将所有的指标类型统一转化为**极大型指标**。

$$X_{nm} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}$$

候选人	颜值	脾气 (争吵次数)	身高	体重
A	9	0	0	0
B	8	3	0.9	0.5
C	6	7	0.2	1



## ➤ 灰色关联综合评价

### 2) 正向化后数据预处理

每个指标的元素除以该指标元素的平均值

$$Z_{nm} = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1m} \\ z_{21} & z_{22} & \cdots & z_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{nm} \end{bmatrix}$$

候选人	颜值	脾气 (争吵次数)	身高	体重
A	1.17	0.00	0.00	0.00
B	1.04	0.90	2.45	1.00
C	0.78	2.10	0.55	2.00

### 3) 构造母序列

$$Y = [y_1, y_2, \dots, y_n]^T \quad \text{其中 } y_i = \max(z_{i1}, z_{i2}, \dots, z_{im})$$

候选人	Y	颜值	脾气 (争吵次数)	身高	体重
A	1.17	1.17	0.00	0.00	0.00
B	2.45	1.04	0.90	2.45	1.00
C	2.10	0.78	2.10	0.55	2.00



## ➤ 灰色关联综合评价

### 4) 计算关联系数

记差值矩阵为  $K$

$$K_{nm} = \begin{bmatrix} k_{11} & k_{12} & \cdots & k_{1m} \\ k_{21} & k_{22} & \cdots & k_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ k_{n1} & k_{n2} & \cdots & k_{nm} \end{bmatrix} = \begin{bmatrix} |z_{11} - y_1| & |z_{12} - y_1| & \cdots & |z_{1m} - y_1| \\ |z_{21} - y_2| & |z_{22} - y_2| & \cdots & |z_{2m} - y_2| \\ \vdots & \vdots & \ddots & \vdots \\ |z_{n1} - y_n| & |z_{n2} - y_n| & \cdots & |z_{nm} - y_n| \end{bmatrix}$$

$$a = \min_i \min_k |x_0(k) - x_i(k)| \quad b = \max_i \max_k |x_0(k) - x_i(k)|$$

候选人	$Y$	$Z_1$	$Z_2$	$Z_3$	$Z_4$	$ z_{k1} - y_k $	$ z_{k2} - y_k $	$ z_{k3} - y_k $	$ z_{k4} - y_k $
A	1.17	1.17	0.00	0.00	0.00	<b>0.00</b>	1.17	1.17	1.17
B	2.45	1.04	0.90	2.45	1.00	<b>1.41</b>	<b>1.55</b>	<b>0.00</b>	<b>1.45</b>
C	2.10	0.78	2.10	0.55	2.00	1.32	<b>0.00</b>	<b>1.55</b>	<b>0.10</b>

易得  $a = 0$ ,  $b = 1.55$



## ➤ 灰色关联综合评价

4) 计算关联系数

$$a = 0, \quad b = 1.55 \quad \xi_i(k) = y(x_0(k), x_i(k)) = \frac{0 + 0.5 \times 1.55}{|x_0(k) - x_i(k)| + 0.5 \times 1.55} = \frac{0.775}{|x_0(k) - x_i(k)| + 0.775}$$

候选人	$ z_{k1} - y_k $	$ z_{k2} - y_k $	$ z_{k3} - y_k $	$ z_{k4} - y_k $	$\xi_1$	$\xi_2$	$\xi_3$	$\xi_4$
A	0.00	1.17	1.17	1.17	1.000	0.398	0.398	0.398
B	1.41	1.55	0.00	1.45	0.355	0.333	1.000	0.348
C	1.32	0.00	1.55	0.10	0.370	1.000	0.333	0.886

5) 计算关联度

$$r_i = \frac{1}{n} \sum_{k=1}^n \xi_i(k) = \frac{1}{n} \sum_{k=1}^n y(x_0(k), x_i(k))$$

候选人	$\xi_1$	$\xi_2$	$\xi_3$	$\xi_4$
A	1.000	0.398	0.398	0.398
B	0.355	0.333	1.000	0.348
C	0.370	1.000	0.333	0.886
<b>r</b>	<b>0.575</b>	<b>0.577</b>	<b>0.577</b>	<b>0.544</b>



## ➤ 灰色关联综合评价

6) 计算指标权重

$$w_i = \frac{r_i}{\sum_{k=1}^m r_k} (i=1, 2, \dots, m)$$

候选人	颜值	脾气 (争吵次数)	身高	体重
$r$	0.575	0.577	0.577	0.544
$w$	0.253	0.254	0.254	0.239

6) 计算得分并归一化

$$S_i = \sum_{k=1}^m Z_{ik} \cdot w_i \quad \tilde{S}_i = \frac{S_i}{\sum_{k=1}^m S_k} (i=1, 2, \dots, n)$$

候选人	颜值	脾气 (争吵次数)	身高	体重	得分	归一化得分
A	1.17	0.00	0.00	0.00	0.296	0.099
B	1.04	0.90	2.45	1.00	1.353	<b>0.451</b>
C	0.78	2.10	0.55	2.00	1.348	0.450

- 明星K依然选择了B~~~





## ➤ 关联分析例题代码

%% 灰色关联分析用于系统分析例题的讲解

```
clear;clc
```

```
A=input('请输入初始矩阵=') % 输入初始矩阵
```

```
Mean = mean(A); % 求出每一列的均值以供后续的数据预处理
```

```
A = A ./ repmat(Mean, size(A, 1), 1); %size(A, 1)=6
```

```
disp('预处理后的矩阵为: '); disp(A)
```

```
Y = A(:, 1); % 母序列
```

```
X = A(:, 2:end); % 子序列
```

```
absX0_Xi = abs(X - repmat(Y, 1, size(X, 2))) % 计算 $|X_0 - X_i|$ 矩阵(在这里我们把 $X_0$ 定义为了 $Y$ )
```

```
a = min(min(absX0_Xi)) % 计算两级最小差a
```

```
b = max(max(absX0_Xi)) % 计算两级最大差b
```

```
rho = 0.5; % 分辨系数取0.5
```

```
gamma = (a+rho*b) ./ (absX0_Xi + rho*b) % 计算子序列中各个指标与母序列的关联系数
```

```
disp('子序列中各个指标的灰色关联度分别为: ')
```

```
disp(mean(gamma))
```





## ➤ 灰色关联综合评价例题代码

%%1. 矩阵正向化 代码见TOPSIS一节

%% 对正向化后的矩阵进行预处理

Mean = mean(X); % 求出每一列的均值以供后续的数据预处理

Z = X ./ repmat(Mean, size(X, 1), 1);

disp('预处理后的矩阵为: '); disp(Z)

%% 2. 构造母序列和子序列

Y = max(Z, [], 2); % 母序列为虚拟的, 用每一行的最大值构成的列向量表示母序列

X = Z; % 子序列就是预处理后的数据矩阵

%% 3. 计算得分

absX0-Xi = abs(X - repmat(Y, 1, size(X, 2))) % 计算 $|X_0 - X_i|$ 矩阵

a = min(min(absX0-Xi)) % 计算两级最小差a

b = max(max(absX0-Xi)) % 计算两级最大差b

rho = 0.5; % 分辨系数取0.5

gamma = (a+rho\*b) ./ (absX0-Xi + rho\*b) % 计算子序列中各个指标与母序列的关联系数

weight = mean(gamma) / sum(mean(gamma)); % 利用子序列中各个指标的灰色关联度计算权重

score = sum(X .\* repmat(weight, size(X, 1), 1), 2); % 未归一化的得分

stand-S = score / sum(score); % 归一化后的得分

[sorted-S, index] = sort(stand-S, 'descend') % 进行排序

# 欢迎关注数模加油站

## THANKS



有兴趣的小伙伴可以关注微信公众号或加入建模交流群获取更多免费资料

公众号：数模加油站

交流群：709718660