灰色系统理论应用丛书

罗庆成　徐国新　编著

江苏科学技术出版社

灰色关联分析与应用

第一章 引 论

§1.1 什么叫关联分析

1.1.1 定 义

所谓关联分析，就是系统的因素分析。它回答的问题是，某个包含多种因素的系统中，哪些因素是主要的，哪些是次要的；哪些因素影响大，哪些影响小；哪些因素是明显的，哪些是潜在的；哪些需要发展，哪些需要抑制……。

在宏观经济系统中，社会总产值取决于工业、农业、交通邮电、建筑业和商业等。在这些因素中，哪些影响大，哪些影响小，哪些发展趋势与社会总产值一致，哪些不一致，哪些是发展快的，哪些是发展慢的。要回答这些问题，就有必要作因素分析。

在社会系统中，人口是一个子系统。影响人口发展的因素很多，诸如计划生育政策和措施、社会道德风尚、民族文化素质、医疗水平、社会福利、社会治安等。为了有效地控制人口，也要进行因素分析。

在农业系统中，粮食生产是一个重要的子系统。影响粮食产量的因素很多。比如科技、教育、价格、贷款、气候、农用生产资料供应、农田水利、林业和畜牧业的发展、劳动力素质、作物品种、耕作、土壤、交通等等。为了促进粮食生产，也有必要进行因素分析。

因素分析方法，过去主要是统计的方法，如单因素回归，线性回归，多因素回归，非线性回归等。但统计分析大都是少因素、线性的，而多因素、非线性的回归分析难度很大。

1982年我国著名学者邓聚龙教授首次提出灰色系统理论，灰色关联分析应运而生。

1.1.2 特点

灰色关联分析作为一种系统分析技术，是分析系统中各因素关联程度的方法，或者说是对系统动态过程发展态势的量化比较分析的方法。其基本思路是根据系统动态过程发展态势，即系统历年有关统计数据的几何关系及其相似程度，来判断其关联程度。

例1：已知某地国民生产总值、工业和农业产值，如表1-1：

表1-1　某地历年国民生产总值与工农业产值

单位：千万元

项目名称	年份 1980	1981	1982	1983	项目代号
国民生产总值	40	44	48	60	X_0（参考数列）
工业产值	24	38	40	50	X_1（比较数列1）
农业产值	16	22	18	20	X_2（比较数列2）

将表1-1作成曲线，如图1-1。曲线②与曲线①相似程度，大于曲线③与曲线①的相似程度，因此可以认为该地区国民生产总值与工业产值关系密切，也就是说该地工业发展对该地国民生产总值的增长影响较大，应当注重发展工业。

这种分析比较，实质上是几种曲线间几何形状的比较

分析，即认为几何形状越相似，则发展变化态势就越接近，其关联程度也就越大。这种几何形状的判断虽然比较直观，但不能量化。如果有若干条曲线形状相似，或者虽有差别但各区段情况不一，有的区段接近，有的区段不太相似，很难用直接观察的方法来判断各条曲线间的关联程度。在这种情况下，通过关联分析，就可以把若干曲线相似程度进行量化比较分析。

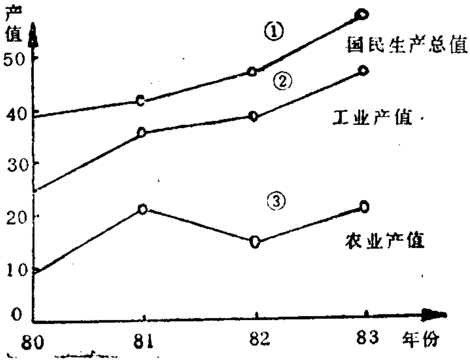

图1-1　曲线间相似程度比较分析

这种方法不同于其他因素分析法，其主要特点是：

（1）对数据要求不那么严。不象统计分析那样，要求大量数据；也不要求数据有典型分布规律（线性的、指数的或对数的）。

（2）计算方法简便。即使是多因素比较分析，计算工作量也不象统计分析那样复杂，有计算机当然好，没有条件手算也可以完成。

1.1.3 意　义

世间一切事物都是以系统的形式存在和发展，而社会、

经济、生态等系统都是由多种要素组成的。它们彼此之间的关系错综复杂，使人们在认识、分析、预测、决策和控制时，得不到全面的、足够的信息。在某种意义上可以说，它们彼此之间的关系是灰的，即有一部分关系是确知的，另一部分关系是不确知的。即使是确知的关系，也多属定性分析和判断，不一定数量化、精确化。因此，就难以抓住主要矛盾，发现主要特征或主要关系。

灰色关联分析正是适应灰色系统因素分析的这种客观需要，通过对灰色系统动态过程（即系统历年有关统计数据）发展态势的量化比较分析，把系统有关因素之间的各种关系展现在人们面前，为系统预测、决策、控制提供有用信息和比较可靠的依据。由于这种方法能使灰色系统各因素之间的“灰”关系“白”化（清晰化），所以把它称之为灰色关联分析，简称关联分析。

自从邓聚龙教授1982年创立灰色关联分析方法之后，很快渗透到了社会、经济、生态等各个领域，得到广大科技、教育和实际工作者的赞赏，农业、工业、教育、卫生、体育、环保、军事、地质、水文、气象、生物等各方面的关联分析。如雨后春笋般地涌现出来，在系统态势分析、投资效益分析、方案评审、产业结构调整等方面，均显示了这种系统分析的重要作用。

§1.2 关联分析的步骤

1.2.1 确定参考序列

关联分析首先要确定参考序列（即数列）。所谓参考序列，就是作比较的“母序列”。常记为x_0，它由不同时刻的

统计数据构成。记第一个时刻的值为$x_0(1)$，第二个时刻的值为$x_0(2)$，第k个时刻的值为$x_0(k)$。这样，参考序列x_0可表示为：

$$x(k)=\{x_0(1),\ x_0(2),\ \cdots,\ x_0(n)\}$$
$$(k=1,\ 2,\ \cdots,\ n)$$

对于例1（表1-1）的数据，其中的“国民生产总值”历年统计数据就是参考序列。

第1个时刻x_0的值为$x_0(1)=40$

第2个时刻x_0的值为$x_0(2)=44$

第3个时刻x_0的值为$x_0(3)=48$

第4个时刻x_0的值为$x_0(4)=60$

将$x_0(1)$到$x_0(4)$集合起来，就得到参考序列：

$$x_0(k)=\{x_0(1),\ x_0(2),\ x_0(3),\ x_0(4)\}$$
$$=\{40,\ 44,\ 48,\ 60\}$$

关联分析中与参考序列作关联程度比较的“子数列”，称之为比较序列，记为x_1，x_2，…，x_n。若第一个比较序列如表1-1“工业产值”，即得：

$$x_1(k)=\{x_1(1),\ x_1(2),\ x_1(3),\ x_1(4)\}$$
$$=\{24,\ 38,\ 40,\ 50\}$$

第二个比较序列为“农业产值”，即：

$$x_2(k)=\{x_2(1),\ x_2(2),\ x_2(3),\ x_2(4)\}$$
$$=\{10,\ 22,\ 18,\ 20\}$$

确定参考序列和比较序列时，要注意正确地表示，如：

$x_0(k)$中的 0 表示参考序列下标，k表示不同时刻。

$x_1(k)$中的 1 表示比较序列下标，k表示不同时刻。

1.2.2 无量纲化

由于系统中各因素的计量单位不同，所以数据的量纲也

不一致。有的是实物计量，如钢产量为吨、万吨，木材为立方米，棉布为米、万米，拖拉机为台、万台；有的是货币计量，工农业产值为元、万元、亿元；还有的是以人数、工时为计量单位。不仅量纲不同，而且数值的数量级也各异。有的以元为单位，有的以万元为单位，有的以亿元为单位；有的以公斤为单位，有的以吨为单位。不同量纲、不同数量级之间不便于比较，或者在比较时难以得到正确的结论。因此，在进行灰色关联分析时，一般都要进行无量纲化的数据处理。

原始数列无量纲化的方法，有初值化、均值化、中值化、公值化、区间相对值化等。常用的方法，主要是前两种。

1. 初值化

初值化，是指同一数列的所有数据，均除以第一个数据，得到一个新的数列。这个新的数列，即是各个时刻的数值相对于第一个时刻的数值的倍数的数列。例1（表1-1）中的“国民生产总值”数列为：

$$\{x_0(k)\}=\{x_0(1), x_0(2), x_0(3), x_0(4)\}$$
$$=\{40, 44, 48, 60\}$$

作初值化处理后得：

$$\left\{x_0(k)\right\}=\left\{\frac{x_0(1)}{x_0(1)}, \frac{x_0(2)}{x_0(1)}, \frac{x_0(3)}{x_0(1)}, \frac{x_0(4)}{x_0(1)}\right\}$$
$$=\left\{\frac{40}{40}, \frac{44}{40}, \frac{48}{40}, \frac{60}{40}\right\}$$
$$=\{1.0, 1.1, 1.2, 1.5\}$$

这个新的数列表明，1981年为1980年的1.1倍，1982年为1980年的1.2倍，1983年为1980年的1.5倍。

「そうかしら。機械にこしょうはつきものよ。[1]そんな、マンガみたいなわけにはいかないのよ」

ママはかなりの懐疑派だから、なかなかあとにひかない[2]。

「まったくいやんなっちゃうな[3]。未来は科学がもっと発達するんだよ。自動車だけじゃないよ、家のなかだって——おそうじや、おせんたくだって、みんなロボットがやってくれるようになるから、いまよりずっとらくになるんだ」

「あら、ママはお料理もおそうじもすきだから、じぶんでするわ。ロボットなんかに、家のなかをごっとんがったん動きまわられちゃ、気もちわるくって、いられないわ[4]」

ママは、ほんとうに気もちわるそうに、まゆをひそめた。

わたしは、おもしろくなって、ふたりのやりとりを、もうすこし聞いていることにした[5]。

「ママは、未来なんて[6]、あんまりいい時代だとは思わな

---

1. 〔機械にこしょうはつきものよ〕机器发生故障是难免的啊!
2. 〔あとにひかない〕不后退；不罢休。
3. 〔まったくいやんなっちゃうな〕感到很厌烦。"な"是终助词。这里表示使对方理解自己的想法。
4. 〔ロボットなんかに、…いられないわ〕让机器人在家里咯噔咯噔到处转来转去，有多讨厌，怎能受得了！"ちゃ"是接续助词"ては"的变音。接续助词"て"与形容词连用形连接时可变成"って"。
5. 〔もうすこし聞いていることにした〕决定再稍微听一听。"…ことにする"接动词连体形后，表示"决定…""一定…"的意思。△この字引を買うことにする。/决定买这本字典。
6. 〔未来なんて〕"なんて"，副助词。含有轻视的语气，相当于汉语的"等等""之类"的意思。

关联程度，实质上是曲线间几何形状的差别。因此，曲线间差值大小，可以作为关联程度的衡量尺度。

一般情况下的关联分析，是对于一个参考序列x_0，有若干个比较序列x_1，x_2，…，x_n。各比较序列（即比较曲线）与参考序列（即参考曲线）在各个时刻（即曲线的各点）的差，可用下述关系式表示：

$$\xi_{0i}(k)=\frac{\min\limits_{i}\min\limits_{k}\left|x_0(k)-x_i(k)\right|+P\cdot\max\limits_{i}\max\limits_{k}\left|x_0(k)-x_i(k)\right|}{\left|x_0(k)-x_i(k)\right|+P\cdot\max\limits_{i}\max\limits_{k}\left|x_0(k)-x_i(k)\right|} \tag{1-1}$$

式中，$\xi_{0i}(k)$是第k个时刻比较序列x_i与参考序列x_0的相对差值。这种形式的相对差值，就称为x_i对x_0在k时刻的关联系数。

$$\min\limits_{i}\min\limits_{k}\left|x_0(k)-x_i(k)\right|$$

称为两个层次（即两级）的最小差：

第一个层次（第一级）的最小差为：

$$\min\limits_{k}\left|x_0(k)-x_i(k)\right|=\Delta_{0i}\binom{(\min)}{k}$$

即在参考序列x_0与第i个比较序列的绝对差值（取绝对值）中，选出一个最小的差值。简记为$\Delta_{0i}\binom{(\min)}{k}$。

第二个层次（第二级）的最小差：

$$\min\limits_{i}\left(\min\limits_{k}\left|x_0(k)-x_i(k)\right|\right)=\Delta_{0i}(\min)$$

即在参考序列x_0与所有比较序列x_i的最小绝对差值中，再选出一个最小的差值，简记为$\Delta_{\min}$。

$$\max_i \max_k \left| x_0(k) - x_i(k) \right|$$

为两个层次（即两级）的最大差。

第一个层次（第一级）的最大差为：

$$\max_k \left| x_0(k) - x_i(k) \right| = \Delta_{0i} \left(\max_k\right)$$

即在参考序列x_0与第i个比较序列的绝对差值中，选出一个最大的差值，简记为 $\Delta_{0i}\left(\max_k\right)$。

第二个层次（即第二级）的最大差为：

$$\max_i \left(\max_k\right) \left| x_0(k) - x_i(k) \right| = \Delta_{0i}(\max)$$

简记为Δ_{max}。

下式(1-2)为分辨系数，是为了削弱最大绝对差值因为过大而失真的影响，以提高关联系数之间的差异显著性，而人为给定的系数，在一般情况下可取0.1～1.0。

$$|x_0(k) - x_i(k)| = \Delta_{0i}(k)$$

为参考序列x_0与各比较序列x_i，在第k个时刻的绝对差值，简记为$\Delta_{0i}(k)$。

以上（1-1）式，可简化为：

$$\xi_{0i}(k) = \frac{\Delta_{min} + P \cdot \Delta_{max}}{\Delta_{0i}(k) + P \cdot \Delta_{max}} \tag{1-2}$$

关联系数ξ_{0i}是定性分析和定量分析相结合的产物，其抽象思维的准则（公理）是：

1. 规范性

ξ_0 大于0，小于或等于1。因为没有绝对不关联的两个事物，关联是绝对的，所以$\xi > 0$；虽然没有绝对关联的两个事物，但自己和自己比总是相同的(在同一时间、地点和条件

下），所以$\xi \leqslant 1$。

2． 偶对对称性

偶对即“两两”，对称即“彼此”，两两彼此关联是关联的基础。关联分析把不同曲线的距离空间转变为关联空间。

3． 整体性

关联系数是在保留最小差值和最大差值的条件（环境）下，计算第k个时刻比较序列（比较曲线）x_i与参考序列（参考曲线）x_0的相对差值。

4． 接近性

P这个分辨系数实际上是人为给定的（定性分析的人为系数）。

$P \in [0, 1]$

分辨系数非唯一，可在0～1之间取值，一般为0.5或1.0。

利用例1（表1-1）作初值化后的数值，计算关联系数的过程如下：

记经过初值化处理的参考序列为$\{x_0(k)\}$、比较数列为$\{x_i(k)\}$。

$\{x_0(k)\} = \{1.0, 1.1, 1.2, 1.5\}$ （国民生产总值）

$\{x_1(k)\} = \{1.0, 1.6, 1.7, 2.1\}$ （工业产值）

$\{x_2(k)\} = \{1.0, 2.2, 1.8, 2.0\}$ （农业产值）

1． 求绝对差

$$\Delta_0(k) = |x_0(k) - x_i(k)|$$
$$(k = 1, 2, \cdots, n)$$

$$\Delta_{01}(k) = |x_0(k) - x_1(k)|$$
$$(k = 1, 2, 3, 4)$$

$k = 1 \quad \Delta_{01}(1) = |x_0(1) - x_1(1)| = |1 - 1| = 0$

たが、お気にめさないらしく[1]、いちいち、ひにくなことをいう。

「なまけ者の世界じゃないよ、レジャー時代っていうんだよ[2]」

たかしも、まけてはいず[3]、ぎゃくしゅうした。

「未来になると、いろんなものがオートメーションになるから、はたらく時間もすくなくてすむし、ひまができるでしょう。だから、スキーだって、登山だって、水泳だって、ナイター[4]なんかにも、しょっちゅう、つれていってもらえるようになるんだよね、パパ。いまはパパはいそがしくて、めったに、旅行にもつれていってくれないけどね」

わたしはママと顔を見あわせてにがわらいした。たかしは、そんなわたしたちを横目でみやると、また、にっとわらった。

「でも、いいんだ。未来になれば、ぼくもおとなだから、宇宙船で、宇宙ステーションに行って地球をながめたり、月にできるムーンシティ[5]へ行って、月見物をやったりできるんだから。なにしろ、宇宙時代だからね」

---

1. 〔お気にめさないらしく〕好象并不在乎。
2. 〔レジャー時代っていうんだよ〕那是个悠闲的时代。"って"是助词"と"的变形。
3. 〔まけてはいず〕不服输，不甘休。"ず"是助动词"ぬ"的连用形，接在动词或某些助动词的未然形下面，表示否定。△一晩中ねむらずに患者の世話を見る。/照顾病人，一夜没合眼。
4. 〔ナイター〕夜间球类比赛。
5. 〔ムーンシティ〕月球(上的)城市。

由于初值化处理后，第1个差值均为1（其他无量纲化方法所得的第1个差值就不一定为1），所以最小差值均为0。

第二级最小差

$$\Delta_{min}=\Delta_{0i}(\min_{i})=\min_{i}(\min_{k}|x_0(k)-x_i(k)|$$

$$=\min_{i}(0,\ 0)=0$$

第一级最大差

$$\Delta_{01}(\max_{k})=\max_{k}|x_0(k)-x_1(k)|$$

$$=\max_{k}\{0,\ 0.5,\ 0.5,\ 0.6\}$$

$$=0.6$$

$$\Delta_{02}(\max_{k})=\max_{k}|x_0(k)-x_2(k)|$$

$$=\max\{0,\ 1.1,\ 0.6,\ 0.5\}$$

$$=1.1$$

第二级最大差

$$\Delta_{max}=\max_{i}(\max_{k}|x_0(k)-x_i(k)|)$$

$$=\max_{i}(0.6,\ 1.1)$$

$$=1.1$$

3. 决定分辨系数

取P=0.5

4. 求关联系数

$$\because\quad \xi_{0i}(k)=\frac{\Delta_{min}+P\cdot\Delta_{max}}{\Delta_{0i}(k)+P\cdot\Delta_{max}}$$

$$= \frac{0 + 0.5 \times 1.1}{\Delta_{0i}(k) + 0.5 \times 1.1}$$

$$= \frac{0.55}{\Delta_{0i}(k) + 0.55}$$

$\{\Delta_{01}(k)\} = \{0,\ 0.5,\ 0.5,\ 0.6\}$

$\therefore\quad k = 1 \quad \xi_{01}(1) = \frac{0.55}{0 + 0.55} = 1$

$k = 2 \quad \xi_{01}(2) = \frac{0.55}{0.5 + 0.55} = 0.52$

$k = 3 \quad \xi_{01}(3) = \frac{0.55}{0.5 + 0.55} = 0.52$

$k = 4 \quad \xi_{01}(4) = \frac{0.55}{0.6 + 0.55} = 0.48$

$\{\xi_{01}(k)\} = \{\xi_{01}(1),\ \xi_{01}(2),\ \xi_{01}(3),\ \xi_{01}(4)\}$
$= \{1,\ 0.52,\ 0.52,\ 0.48\}$

同理：

$\because\quad \{\Delta_{02}(k)\} = \{0,\ 1.1,\ 0.6,\ 0.5\}$

$\therefore\quad k = 1 \quad \xi_{02}(1) = \frac{0.55}{0 + 0.55} = 1$

$k = 2 \quad \xi_{02}(2) = \frac{0.55}{1.1 + 0.55} = 0.33$

$k = 3 \quad \xi_{02}(3) = \frac{0.55}{0.6 + 0.55} = 0.48$

$k = 4 \quad \xi_{02}(4) = \frac{0.55}{0.5 + 0.55} = 0.52$

$\{\xi_{02}(k)\} = \{\xi_{02}(1),\ \xi_{02}(2),\ \xi_{02}(3),\ \xi_{02}(4)\}$
$= \{1,\ 0.33,\ 0.48,\ 0.52\}$

1.2.4 求关联度

因为关联系数是比较曲线与参考曲线在第k个时刻的相对差值，所以它的数不止一个，信息过于分散，不便于从整体上进行比较。因此，就有必要将各个时刻的关联系数集中为一个值，也就是求其平均值，作为关联程度的数量表示。

关联度记为r_{0i}，其表达式为：

$$r_{0i}=\frac{1}{N}\sum_{k=1}^{n}\xi_0(k) \tag{1-3}$$

或者说，r_{0i}是比较曲线x_i对参考曲线x_0的关联度。

式中，N为比较序列的数据数。

例1的关联度为：

$$\begin{aligned}r_{01}&=\frac{1}{4}\sum_{k=1}^{4}\xi_{01}(k)\\&=\frac{1}{4}\Big(\xi_{01}(1)+\xi_{01}(2)+\xi_{01}(3)+\xi_{01}(4)\Big)\\&=\frac{1}{4}(1+0.52+0.52+0.48)\\&=0.63\end{aligned}$$

同理：

$$\begin{aligned}r_{02}&=\frac{1}{4}(1+0.33+0.48+0.52)\\&=0.58\end{aligned}$$

以上说明x_1与x_0(即“工业产值”与“国民生产总值”)关联度$r_{01}=0.63$；x_2与x_0(即“农业产值”与“国民生产总值”)关联度为0.58。

从上例可以看出，关联度取决于参考数列和比较数列及其数列长度(数据多少)等因素。

1.2.5 排关联序

当比较序列有m个时，相对的关联度也有m个，按其值大小排列起来，即为关联序。关联度直接反映各个比较序列对于参考序列的优劣关系。

上述关联度，$r_{01}=0.63$，$r_{02}=0.58$，即：

$$r_{01}>r_{02}$$

因此，称比较序列x_1对于参考序列x_0，优于比较序列x_2对于x_0。或者说x_1（工业产值）对x_0（国民生产总值）的影响，大于x_2（农业产值）对于x_0，也就是x_1对x_0的影响较大，x_2对x_0的影响相对较小。

§1.3 优势分析

1.3.1 什么叫优势分析

一般的关联分析，参考序列只有一个。当参考序列和比较序列都不止一个时，这种关联分析，就叫做优势分析。

优势分析称参考序列（数列）为母序列（或母数列、母因素），比较序列为子序列（或子数列、子因素）。由母序列与子序列构成关联矩阵。通过关联矩阵，分析各因素之间的关系，找出优势因素和非优势因素。

进行优势分析，对于研究社会经济发展战略，合理分配使用人力、物力、财力资源，统筹安排各部门、各项生产的发展，做到保证重点，兼顾一般，扬长避短，提高社会、经济和生态效益等，有着重要的意义。

1.3.2 如何进行优势分析

1. 确定母序列和子序列

母序列有m个，记为y_1，y_2，y_3，…，y_m；子序列有

n个，记为x_1，x_2，x_3，…，x_n。

2． 无量纲化，方法同前（1.2.2）

3． 计算关联度，方法同前（1.2.3及1.2.4）

4． 构造关联度矩阵。

如果有5个母序列、6个子序列，则每个母序列对6个子序列，需要计算6个关联度。记第1个母序列对6个子序列的关联度为：

r_{11}＝（y_1与x_1的关联度）＝$r(y_1, x_1)$，

r_{12}＝（y_1与x_2的关联度）＝$r(y_1, x_2)$，

r_{13}＝（y_1与x_3的关联度）＝$r(y_1, x_3)$，

r_{14}＝（y_1与x_4的关联度）＝$r(y_1, x_4)$，

r_{15}＝（y_1与x_5的关联度）＝$r(y_1, x_5)$，

r_{16}＝（y_1与x_6的关联度）＝$r(y_1, x_6)$。

将母序列y_1对6个子序列的关联度按次序排成第一行，得：

$$[r_{11}, r_{12}, r_{13}, r_{14}, r_{15}, r_{16}]$$

同理，将母序列y_2对6个子序列的关联度，按次序排为第二行，得：

$$[r_{21}, r_{22}, r_{23}, r_{24}, r_{25}, r_{26}]$$

将y_3、y_4、y_5对母序列对6个子序列的关联度，分别排成第三行、第四行、第五行，便可得一个5×6的关联度矩阵R，

$$R=\begin{pmatrix} r_{11} & r_{12} & r_{13} & r_{14} & r_{15} & r_{16} \\ r_{21} & r_{22} & r_{23} & r_{24} & r_{25} & r_{26} \\ r_{31} & r_{32} & r_{33} & r_{34} & r_{35} & r_{36} \\ r_{41} & r_{42} & r_{43} & r_{44} & r_{45} & r_{46} \\ r_{51} & r_{52} & r_{53} & r_{54} & r_{55} & r_{56} \end{pmatrix}$$