

灰色系统理论与方法

曹 军 胡万义 编著

东北林业大学出版社

灰色系统理论与方法

曹军 胡万义 编著

东北林业大学出版社

(黑) 新登字第 10 号

灰色系统理论与方法

、 曹 军 胡万义 编著

东北林业大学出版社出版发行

(哈尔滨市和兴路 8 号)

东北林业大学印刷厂印刷

开本 787×1092 毫米 1/16 印张 10.75 字数 238 千字

1993 年 5 月第 1 版 1993 年 5 月第 1 次印刷

印数 1—1000 册

ISBN7—81008—397—x/TH·15

定价：5、50 元

前 言

1982年,我国著名学者、华中理工大学的邓聚龙教授创立了灰色系统理论。提出灰色系统理论是用来解决信息不完备系统的数学方法。它把控制论的观点和方法延伸到复杂的大系统中,将自动控制和运筹学的数学方法相结合,用独树一帜的有效方法和手段,去研究广泛存在于客观世界中的具有灰色性的问题。目前,其理论研究有了新的进展,应用研究已渗透到自然科学、社会经济、农业、气象、军事等众多领域,其研究已经形成一个范围广阔、发展迅速的新兴学科。

本书是在吸收近几年来众多学者发表的大量学术论文、出版的若干教材和专著等的基础上,结合作者的教学经验和科研成果,经过提炼和补充而编写的。

本书在选材、结构、叙述方法等方面侧重于介绍这一学科的基本观点和主要方法及新思想等,尤其注重从应用角度介绍运用灰色系统理论分析问题和解决问题,尽量做到即具有基本概念、计算方法,又具有结果分析。注重实用性、科学性和可读性。全书共分五章,内容包括:灰色系统基本概念、灰色关联度分析、灰色预测、灰色系统预测与应用、灰色聚类分析与决策方法。本书可作为经济、管理、理、工、农、医、天、地、生各专业大学生和研究生教材和自学参考书,亦可供其他科技工作者参阅。

由昌富同志在百忙中担任本书主审。本书在编写过程中参考使用了许多著作和文献,在此特向这些著作、文献的作者致谢。

由于编者水平有限,书中不足之处在所难免,敬请读者批评指正。

编者

1992年12月 于哈尔滨

目 录

1	概论	
1.1	灰色系统理论的学科概念	1
1.2	灰色系统理论的产生背景及形成	4
1.3	灰色系统理论的研究内容	6
2	灰色关联度分析	8
2.1	灰色关联分析的一般概念	8
2.2	灰色关联分析方法	8
2.3	灰色关联优势分析	15
2.4	灰色关联分析与应用	18
3	灰色预测	32
3.1	预测概论	32
3.2	灰色模块的建立	36
3.3	灰色预测模型的建立	38
4	灰色系统预测与应用	59
4.1	数列预测	59
4.2	灰色灾变预测	71
4.3	灰色季节灾变预测	86
4.4	灰色拓扑预测	93
4.5	灰色系统预测	99
4.6	灰色马尔柯夫概率预测	105
5	灰色聚类分析与决策方法	114
5.1	灰关联聚类	114
5.2	灰色权系数聚类分析	122
5.3	灰色多元聚类	129
5.4	灰色局势决策	136
5.5	灰色层次分析决策方法	146
5.6	灰色二值数据的贝叶斯方法	156
	参考文献	166

1 概 论

1.1 灰色系统理论的学科概念

1.1.1 系统的一般概念

世界上一切处于永恒运动中的物质都不是孤立存在的，而是通过相互联系而形成整体的结构。同样，世界上各种形式的物质运动不是孤立进行的，而是通过其互相作用来发挥整体功能的。在相互联系、相互作用的物质运动中，有一个描述物质世界整体结构与功能的普遍概念，这就是系统。因此，系统是自然界存在的普遍形态，自然界的一切物质都构成系统，自然界存在的一切都是系统。

1.1.1.1 系统的定义

系统是由若干相互作用和相互依赖的并各具功能的组分（要素）组成的具有特定整体功能的有机整体。

1.1.1.2 系统与组分

组分是构成系统的基本要素。由系统定义可知，单个组分是不能构成系统的，要构成一个系统，必须具有两个以上的组分。

一般说来，系统与组分的区别是相对的。一系统只有相对于构成它的组分而言才是系统。同样，一组分只有相对于由它和其它组分构成的系统而言才是组分。没有系统无所谓组分；没有组分也无所谓系统。系统和组分是相互依存的。任一系统都是较高一级系统的一个组分；任一系统中某一组分的本身，通常又是较低一级的系统。

1.1.1.3 系统功能

人们通常把系统与外界环境相互联系和相互作用过程的秩序和能力称为系统的功能。贝塔朗菲在把结构称为“部分的秩序”的同时，把功能称为“过程的秩序”。如果说，系统的结构是揭示系统内部各要素相互作用的秩序的话，那么，系统的功能则是揭示了系统对外界作用过程的秩序，系统的功能体现了系统与外部环境之间的物质、能量和信息的输入与变换的关系。系统的功能与结构之间存在着辨证的关系，表现为种种复杂的情形：①结构是功能的基础，功能是结构的表现。由不同要素按某种联系方式和相互作用而形成的系统整体。它既以一定的结构层次为基础，又表现为一定的功能。就系统整体来说，结构是功能的基础，如果没有要素并按一定的结构方式组织起来，那么，就无所谓系统的功能。另一方面，功能是结构的表现，具有一定结构层次的系统整体，必须表现为一定的功能。这就是结构与功能的辨证关系；②组成系统结构的要素不同，一般说来，系统的功能也就相应地不同；③组成系统结构的要素相同，但结构不同，系统的功能也就不同；④组成系统的要素与结构都不同，也能获得相同的功能；⑤同一结构系

统往往不只有一种功能，而可能有多种功能。

综上所述，系统的功能划分是相对的，功能与结构之间存在着辩证关系，在一定条件下功能与结构是可以相互转化的。

1.1.1.4 系统的特性

(1) 整体性：系统的整体性在于，首先从宏观上看，就一定的意义说来，系统是一种普遍现象，不论自然界、人类社会还是思维领域都具有系统性。其次，系统是各要素之间、要素与整体之间相互联系、相互作用的矛盾的统一体，具有从要素的量的组合达到系统整体的质的飞跃的总效应。这就是说，系统是由相互联系、相互制约的诸要素组成的具有特定功能的综合的整体，这种系统整体具有新的性质，而这种新的性质绝不是组成该系统的诸要素的个体性质的代数和，其整体大于它的代数和。

(2) 层次性：作为综合整体的系统是由不同层次结构组成的。系统内部各要素的组成是按一定的联系方式和作用方式分层次组织而成的有机的整体，而不是各要素的杂乱堆集。

首先，从结构上看，由于客观世界是无限的，因而系统的层次结构也是无限的。在系统结构的无限层次中，高一级的系统内部结构的要素，包含着低一级系统的结构层次。复杂的大系统内部结构中的要素，本身就可能是一个简单的系统或比较大的系统。系统的结构与要素是相对于系统的等级和层次而言的，需要对它作具体的、辩证的考察，而不能凝固化和绝对化。

(3) 制约性：从功能上看，一般地说，高一级的结构层次对低一级的结构层次具有重大的制约作用，而低一级的结构层次也不是消极地受高级结构层次的制约，而是表现为积极的相互联系和相互作用。它们之间通过信息的传输和反馈等方式，有着辩证的联系。

1.1.1.5 系统的动态性

系统的动态性是指考察系统时不仅应当研究系统的静态结构、结构与功能等方面的复杂情形，而且更重要的是应当研究系统的历史演化，研究系统从无序到有序、低序向高序和从有序向无序的演化过程并形成某种稳定的动态结构机制。

1.1.2 灰色系统

客观世界是由物质、能量和信息所组成，它们决定了事物的性质。因为物质的新陈代谢和能量的相互转换，必须依靠信息才能进行，所以信息虽然不是物质，也不是能量，但它不能脱离物质和能量而存在。所以我们说，客观世界是信息世界，在信息世界中，既有大量的已知信息，也有不少未知信息和非确知的信息。

在系统论与控制论中，人们常常用颜色深浅来表示系统信息的完备程度。因此我们把内部特性已知的信息系统称为白色系统；把内部特性未知的或非确知的信息系统称为黑色系统；把内部特性既含有已知的又含有未知的或非确知的信息系统称为灰色系统。

事物是客观存在的，而人类的认识是逐步发展、深化的，人们不可能一下子认识事物的客观规律，只能通过实践来修正主观认识并不断提高对客观事物的认识。所以，客观事物对人类来说，并不是“白的”（即系统中的全部信息确定或确知），也不是“黑的”（全部信息不确定或不确知），而是“灰的”（系统的信息部分确定、部分不确定）。因

此,人类的思维是“灰的”,行动依据是“灰的”,人们不得不在“灰的”环境里思考、决策和行动。通常,只是由于人们对一些不确定因素忽略不计,才把某些灰色系统当作白色系统来认识和处理,可见,严格地说,灰色系统是绝对的,而白色系统与黑色系统是相对的。

1.1.2.1 灰数

信息不完全的数称为灰数。比如“某人的年龄 18 岁左右”。这“18 岁左右”便是灰数,可记为 $\otimes=18$,或 $18\in\otimes$ 。 $\otimes$ 表示 $\otimes$ 的白化数(值)。又比如今天的气温是 $10\sim 20^{\circ}\text{C}$ 之间,这 $10\sim 20^{\circ}\text{C}$ 之间”也是灰数,可记为 $\otimes=[10, 20]$ 。

1.1.2.2 灰元

信息不完全或者内涵难以穷尽的元素称为灰元。比如货币是灰元,无论把货币作为流通的手段还是把它作为价值的尺度,它的功能都是不确定的。一张有价钞票,它代表多少社会价值、代表多少商品、代表多少工资并不确定,它可因人、因时、因地而异。

1.1.2.3 灰元素

信息不完全或机制不明确的关系称为灰关系。比如,某公司的经营效益与商品利税额、商品纯销售额、商品流通费用率、资金周转次数、人均销货额等有关,可是具体关系很难完全明确,更不容易量化,这便是灰关系。具有灰关系的因素是灰因素,灰因素之间的量化作用称为灰关联。

显然,灰色系统中往往包含灰数、灰元、灰关系。

1.1.2.4 本征灰系统与非本征灰系统

含灰数或者灰元、灰关系的系统是信息不完全的系统,即为灰系统。如按客体的性质分,灰色系统可分为本征灰色系统与非本征灰色系统。

(1) 本征灰系统:社会、经济、农业、生态、农业等是本征性灰系统。这类系统的基本特点是:没有物理原型,缺乏建立确定关系的信息,系统的基本特征是多个相互依存、相互制约的部分、部门、元素,按照一定的序关系组合,且具有一种或多种功能。象这样客观存在的抽象系统,可视为主观的本征性灰色系统,简称为本征灰系统。

(2) 非本征灰系统:工程技术系统、人体系统等,具有客观实体的、实际的物理系统,若有些信息暂时还不确知,或尚未获得,则称为非本征性灰色系统,简称非本征灰系统。

1.1.2.5 灰思想

灰色系统理论的哲学基础,是认识论的反映论,是认识论的信息论。人们对客观世界的认识,是客体以信息形式在大脑中引起反映而开始的。由于对客观世界的认识是无穷尽的,因此人们获得的信息也总是不完全的;由于认识是可深化的,因此信息不完全的客体是可研究的;由于感性认识是可升华为理性认识的,基于理性认识,对客体是可进行处置的,因此信息不完全的客体是可以处置的。人体是一个复杂的灰系统,信息多,其中尚有许多信息尚未获得。如医生根据可获得的有限信息,对疾病进行诊断和治疗,就是一个灰系统处置的实例。

灰色系统理论的技术思想是解的非唯一性。由非唯一性可引伸为集合性、范畴性、可比性、可序化性、可量化性、可选择性、可优化性。影响系统行为特征的因素是非唯一

的。按可比性准则可由全体灰关联因素，构成关联空间。关联空间是对系统因子作量化与序化分析的基础。描述系统行为数据列往往是离乱的。数据的离乱削弱了其可比性和可量化性。但系统行为的数据列的表现形式是非唯一的。从一种表现形式转化为另一种表现形式是数据生成，因此数据生成是非唯一的。生成方式的全体构成生成空间。根据系统量化的要求，在生成空间里可确定一种有效方式，为分析与建模提供较好的基础。同一数据列，可按某种准则构成子序列，一般子序列是非唯一的。因此由子序列构成的子模型是非唯一的，子模型的全体构成子模型群。在子模型群中，可按信息利用程度以及模型与实际状态的吻合程度来确定满意模型。面对一个事件，其处置的方法、途径、手段、对策是非唯一的。在多种对策中挑选一种效果较好的对策，对付事件的发生，便是决策。如果决策的数量表现用一个范围来约束，这个范围便称为灰靶。在把握事件、对策与目标的全局基础上，从灰靶中确定一个满意的对策，便是灰决策。

灰色系统理论属于软科学的范畴，具有多学科的综合性。它把控制论的观点和方法延伸到复杂的大系统中，将自动控制和运筹学的数学方法相结合，用独树一帜的有效方法和手段，来研究广泛存在于客观世界中具有灰色性的问题。在短短的时间里，灰色系统理论有了飞速的发展，它的应用已渗透到自然科学和社会经济等许多领域，显示出这门学科的强大生命力，具有广阔的发展前景。

1.2 灰色系统理论的产生背景及形成

1.2.1 系统理论的发展简史

1.2.1.1 控制论的发展

控制论的诞生是以 1948 年维纳 (N. Wiener) 的《控制论》一书出版为标志的。在维纳的著作中，他首先提出了控制论这一名词。指出了控制工程和通信工程的密切关系，强调了信息和反馈的普遍意义，预言控制论不仅可用于物理系统，而且还可以推广到生物系统、乃至经济领域和社会过程中去。维纳在总结了许多科学家跨学科的共同研究的基础上，创立了控制论这门学科。

控制论的发展大致可分为三个阶段：

第一阶段：50 年代末，称为古典控制理论阶段。古典控制理论主要用于线性定常系统，即可以采用常系数线性微分方程描述的系统。系统数学模型的基本形式是传递函数。对非线性系统则采用相平面法和描述函数法。对于存在多输入、多输出的复杂系统，应用古典理论就非常困难了。古典控制理论采用的是一种外部描述法，它只讨论系统的输入与输出之间的关系，而忽视了系统的内在特性。然而在不少情况下，还需要研究系统的内在特性。古典控制理论的主要贡献是在于建立了系统、信息、调节、反馈、控制和稳定性等控制论的基本概念和分析方法。由于其分析方法采用的是频率法和根轨迹法，而这两种方法带有试凑的特点，因而不能提供最优控制特性。

第二阶段：50 年代末到 70 年代初为现代控制理论阶段。为了适应各种更广泛、更复杂的控制要求，一种建立在线性代数的数学基础之上的现代控制理论迅速地发展起来。这

种理论所采用的状态变量法不仅能提供外部信息，而且能提供系统状态变量的信息。这种方法的一大优点是其通用性强。无论对于线性或非线性系统，定常或时变系统，单变量或多变量系统，它都是一种有效的方法。这种方法虽然需要繁琐费时的计算，但便于应用计算机。随着计算机技术的广泛应用，使现代控制理论的应用范围更加广泛，同时又出现了许多新的分支，如：最优控制、最佳估计、随机控制、系统辨识及自适应控制等。它们主要依靠概率论和统计数学及实变函数等数学方法来建立系统的数学模型，并且在一定条件下求得（估计）模型中的未知参数。

现代控制理论的发展是控制工程中的一个新的阶段，是一个跃变。它综合了应用控制技术（Control）、通信技术（Communication）和计算技术（Computation）的成就，开辟了一条通向各种广阔领域的通路。

第三阶段：从 80 年代开始控制论进入飞速发展阶段。在当今飞速发展的科学和技术中，控制理论处于数学、工程和计算机技术科学融合交叉的前沿。由于从这些学科中获取了养料，使得控制理论产生了强大的效能，并在冲击着先进的技术。从技术、医学到经济、社会及人类生活许多领域中的过程和系统均需要控制。然而实际系统的模型很少是完全已知的，即使已知，也很少是线性的。控制理论工作者现在所面临的挑战是：开阔视野，发展概念和方法，去研究模型不完全的系统，去研究其模型开始时不明确，但可在线改进的系统，以及离散事件动态系统和性质上完全断续的系统。为了适应更复杂的控制要求，智能控制理论将随着从计算机开始的智能机和工程控制实践的进一步要求而发展。其基本特征在于研究的主要目标不再是控制对象，而是控制器本身；控制器不再是单一数学解析型，而是包括有数学解析和直觉推理的知识型。

1986 年 9 月，美国国家科学基金会与 IEEE 控制系统学会（CSS）共同发起、召开了一次控制界的“高峰会议”，有美国控制界最著名的 50 余位专家学者参加，会议的目的是讨论控制科学的发展及当前面临的问题。经过集体讨论，写成了一份报告，题为“对控制的挑战”。报告中指出了今后研究的七大方向问题：（1）多变量鲁棒性、自适应及耐故障控制；（2）非线性及随机控制；（3）分布参数系统的控制；（4）含有离散变量和离散事件的系统控制；（5）信号处理与通讯；（6）分布信息处理及决策结构；（7）系统的集成、实验及执行。上述提法正好与美国工业与应用数学会（SIAM）成立的“控制理论未来方向”专题组提出的报告相吻合。

1.2.1.2 灰色系统理论的形成

实际系统的模型很少是完全已知的，即使已知，也很少是线性的。控制理论工作者现在所面临的挑战是：开阔视野，发展概念和方法，去研究模型不完全的系统，去研究其模型开始时不明确，但可在线改进的系统以及离散事件动态系统和性质上完全断续的系统。我国学者邓聚龙教授 1982 年 3 月在国际上首次提出灰色系统理论，并于 1982 年在北荷兰出版公司的“系统与控制通讯”杂志上正式发表了奠基性论文“Control Problems of Grey Systems”，从而创立了灰色系统理论。

灰色系统理论的产生，从思维逻辑的观点看，是由“黑箱”到“灰箱”再到“灰色系统”。“黑箱”指系统内部结构、特性、参数全部未知的对象，只从对象外部研究它。“灰箱”指对象中有了部分明确的信息，但其框架没有打破，箱的约束依然存在，因此其

研究方法和黑箱方法一样还是从对象外部进行。“灰色系统”不同于“灰箱”，它突破了箱的约束，主张从事物内部去研究系统，充分利用那部分白色信息。

灰色系统理论的重要贡献之一，在于它提出了“灰色系统”的概念。它打破了习惯上将系统简单地划分为“黑色系统”和“白色系统”的习惯作法，为系统理论研究开辟了一个新领域。同时，由于社会经济系统中纯粹的“黑色系统”和“白色系统”十分鲜见，而较多存在的却是灰色系统。因此，“灰色系统”之于社会经济领域的实际应用具有更为广泛的意义。系统“白化”的思路也体现了人类不断认识世界的过程，而“灰色系统控制”则是传统的“黑箱原理”和“反馈控制”理论的深入。

灰色系统理论的重要贡献之二是，它提出了“灰数”、“灰平面”、“灰色统计”等一系列全新的概念，为不确定问题的研究提供了新思路。实际上，在社会经济领域中，很难找到一个确定的“经济数”，在灰色系统理论出现以前，对经济系统中的不确定数，往往采取两种处理方法，一是统计处理，得出所谓“统计数”；二是采取模糊数学方法，得出所谓“模糊数”。对统计处理方法，需要大量的历史数据样本；而采用模糊数学处理方法，则仍不免带有主观性。而灰色系统理论所提出的以区间以及区间运算为代表的灰数处理，则是一种简便实用的方法。另外，灰色决策与西蒙的满意决策一脉相承。根据灰色系统理论的本意，灰色决策的决策结果应是一个灰数，或者说是一个区间，这与传统决策方法的“精确解”是有区别的，因此，灰色决策的结果实际上应是满意解。

灰色系统理论的重要贡献之三是，灰色系统理论提供了认识系统的中介环节。它认为系统的发展是由一无所知，到知之较多，由知之较多到认识其变化规律，直至有可能揭示事物发展变化的长期规律。从变化规律中获取预测数据，通过预测再进行决策等都体现了灰色系统灰度逐渐减少、白度逐渐增加的思想，我们把这一认识事物的过程称为灰色系统的淡化或白化。

1.3 灰色系统理论的研究内容

灰色系统是控制论的观点和方法延伸到社会、经济系统的产物；是自动控制科学与运筹学的数学方法相结合的初步尝试。这就是说，灰色系统是综合性的横断学科，它是由多学科交叉形成的一门崭新的学科。灰色系统作为系统学科一个新分支，正处于发展时期。到目前为止，灰色系统理论的主要研究内容是：

- (1) 灰色系统的建模思想、理论和方法；
- (2) 灰色因素的关联分析、灰色关联序理论；
- (3) 灰色预测理论和方法；
- (4) 灰色决策理论和方法；
- (5) 灰色系统可观性、可控性和稳定性分析；
- (6) 灰色系统的控制；
- (7) 灰色系统优化。

灰色系统理论用崭新的科学方法对系统信息进行处理、分析和建模。它既不回避灰信息的存在，也不是简单地用白数去取代灰数，而用灰度来度量灰信息的不确定度。灰

色建模方法是通过处理灰信息来揭示事物内部的物征和规律。它利用系统信息，使抽象概念量化，量化的概念模型化，最后进行模型优化。它不但考虑通过输出信息去同构系统模型，同时十分重视关联分析。关联分析是系统各因素间或各系统行为间发展势态的相关程度。它开创了灰信息之间联系程度之度量研究，从而充分利用信息。关联分析已用于经济发展分析、多变量的相关分析及综合规划方案的优劣分析。

灰色系统理论提供了对数据的“生成处理”方法，通常通过累加生成

$$x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i)$$

使杂乱无章的无序数据转化为适合于微分方程建模的有序数列。

系统分析主要包括灰色因素的关联分析、灰色统计和灰色聚类等方面的内容；灰色预测则是在关联分析、灰色建模的基础上，通过灰色预测模块实现的。灰色预测的基本思想是将已知的与时间有关的数据按某种规则加以组合，构成白色模块。最后按某种准则提高灰色模块的白化度。灰色预测按照其功能和特征可分成数列预测、灾变预测、季节灾变预测、拓扑预测和系统预测五类；灰色决策则包括灰靶决策、灰关联决策、灰色统计、灰色聚类决策、局势决策、灰色层次决策和灰色规划等。

2 灰色关联度分析

2.1 灰色关联分析的一般概念

在客观世界中,有许多因素之间的关系是“灰的”,分不清哪些因素之间关系密切,哪些不密切。这样就难以找到主要矛盾和主要特性。灰因素的关联分析的目的在于定量地表征诸因素之间的关联程度,寻求系统中各因素间的主要关系,找出影响系统发展势态的重要因素,从而掌握事物的主要特征。关联分析是灰色系统理论分析和系统预测的基础。

灰色关联分析的基本思想是一种相对性的排序分析,它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的。同时它又是对一个系统发展势态进行定量描述和比较的方法。如图 2-1,由于曲线(1)与曲线(2)比较相似,我们认为曲线(1)与曲线(2)的关联度大,记关联度为 r_{12} ;曲线(1)与曲线(3)相差较大,就认为相应的关联度 r_{13} 较小;而曲线(1)与曲线(4)相差最大,则认为关联度 r_{14} 最小。如果我们将关联度按大小顺序排列起来,便组成关联序。其关联序为

$$r_{12} > r_{13} > r_{14}$$

可见,关联分析实际上是一种曲线间几何形状的分析比较,即几何形状越接近,则关联度越大,反之则小。问题在于如何去寻找一种能衡量因素间关联度大小的量化方法。

2.2 灰色关联分析方法

灰色关联分析的目的在于寻求一种能够衡量各因素间的关联度大小的量化方法,以便找出影响系统发展态势的重要因素,从而掌握事物的主要特征。系统发展变化态势的定量描述和比较方法是依据空间理论的数学基础,确定参考数列(母数列)和若干比较数列(子数列)之间的关联系数和关联度。

2.2.1 数据变换

由于系统中各因素的物理意义不同,或计量单位不同,从而导致数据的量纲不同。在

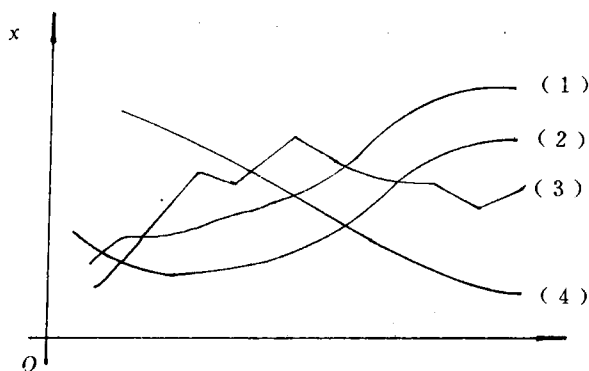


图 2-1 不同形状曲线间的关联

数据的统计过程中,有的是以人数、工时为计量单位的;有的是以元、百元、万元为计量单位的;有的则是以重量吨、万吨为计量单位等等。而且有时数值的数量级相差悬殊,如工业产值有的几万元,有的却达到上百亿元等。这样不同量纲、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果。为了便于分析就需要在各因素进行比较前对原始数据进行归一化处理。常用的数据处理方法有:初值化、均值化、中值化、区间化和归一化等。下面分别加以介绍。

2.2.1.1 初值化处理

对一个数列的所有数据均用它的第一个数去除,从而得到一个新数列的方法称为初值化处理。这个新数列表明原始数列中不同时刻的值相对于第一个时刻值的倍数。该数列有共同起点,无量纲,其数据列中的数据均大于0。

设有原始数列

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

对 $x^{(0)}(i)$ 作初值化处理后得 $x^{(1)}(i)$, 则

$$\begin{aligned} x^{(1)}(i) &= \left\{ \frac{x^{(0)}(1)}{x^{(0)}(1)}, \frac{x^{(0)}(2)}{x^{(0)}(1)}, \dots, \frac{x^{(0)}(n)}{x^{(0)}(1)} \right\} \\ &= \{x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)\} \end{aligned}$$

2.2.1.2 均值化处理

对一个数列的所有数据均用它的平均值去除,从而得到一个新数列的方法称为均值化处理。这个新数列表明原始数列中不同时刻的值相对于平均值的倍数。

设有原始数列

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

对 $x^{(0)}(i)$ 作均值化处理,得 $x^{(1)}(i)$, 为

$$\begin{aligned} x^{(1)}(i) &= \left\{ \frac{x^{(0)}(1)}{\bar{x}^{(0)}}, \frac{x^{(0)}(2)}{\bar{x}^{(0)}}, \dots, \frac{x^{(0)}(n)}{\bar{x}^{(0)}} \right\} \\ &= \{x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)\} \end{aligned}$$

其中

$$\bar{x}^{(0)} = \frac{1}{n} \sum_{k=1}^n x^{(0)}(k)$$

2.2.1.3 区间值化处理

对于指标数列或时间数列,当区间值的特征比较重要时,采用区间值化处理。区间值化处理可分为纵向区间值化处理和横向区间值化处理。

设有 m 个单位,每个单位用 n 个指标描述,则组成指标数列

$$x_1^{(0)}(i) = \{x_1^{(0)}(1), x_1^{(0)}(2), \dots, x_1^{(0)}(n)\}$$

$$x_2^{(0)}(i) = \{x_2^{(0)}(1), x_2^{(0)}(2), \dots, x_2^{(0)}(n)\}$$

...

$$x_m^{(0)}(i) = \{x_m^{(0)}(1), x_m^{(0)}(2), \dots, x_m^{(0)}(n)\}$$

上述指标如果按相同指标(即纵向)取值又可组成新的数列:

$$x_1^{(1)}(i) = \{x_1^{(0)}(1), x_1^{(0)}(1), \dots, x_m^{(0)}(1)\}$$

$$x_2^{(1)}(i) = \{x_1^{(0)}(2), x_2^{(0)}(2), \dots, x_m^{(0)}(2)\}$$

...

$$x_i^{(1)}(i) = \{x_i^{(0)}(n), x_i^{(0)}(n), \dots, x_m^{(0)}(n)\}$$

对数据列采用区间化处理, 得 $x_j^{(2)}(i)$, 为

$$x_1^{(2)}(i) = \{x_1^{(2)}(1), x_2^{(2)}(1), \dots, x_m^{(2)}(1)\}$$

$$x_2^{(2)}(i) = \{x_1^{(2)}(2), x_2^{(2)}(2), \dots, x_m^{(2)}(2)\}$$

... ..

$$x_n^{(2)}(i) = \{x_1^{(2)}(n), x_2^{(2)}(n), \dots, x_m^{(2)}(n)\}$$

在上述新数列中, 若作纵向区间值化处理, 则

$$x_j^{(2)}(i) = \frac{x_j^{(1)}(i) - \min_i x_i(i)}{\max_i x_j(i) - \min_i x_i(i)}$$

$$j = 1, 2, \dots, m; i = 1, 2, \dots, n。$$

对于横向区间值化处理, 设原始数列

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

经横向区间值化处理后, 得 $x^{(1)}(i)$

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

其基本交换关系是

$$x^{(1)}(i) = \frac{x^{(1)}(i) - \min_k x(k)}{\max_k x(k) - \min_k x(k)}$$

$$k = 1, 2, \dots, n; i = 1, 2, \dots, n。$$

2.2.1.4 归一化处理

在数列中若数据的物理量不一致, 且其数值大小相差过分悬殊, 为避免造成非等权的情况, 可采用归一化处理。即对数列中的数据分别进行量级处理。

2.2.2 关联系数

系统间或因素间的关联程度是根据曲线间几何形状的相似程度来判断其联系是否紧密。因此, 曲线间关联程度的量度可作为关联程度的衡量尺度。

设母因素数列 $\{x_0(i)\}$ 和子因素数列 $\{x_j(i)\}$ 分别为

$$x_0(i) = \{x_0(1), x_0(2), \dots, x_0(n)\}$$

$$x_j(i) = \{x_j(1), x_j(2), \dots, x_j(n)\}$$

$$i = 1, 2, \dots, n; j = 1, 2, \dots, m。$$

$\{x_0(i)\}$ 与 $\{x_j(i)\}$ 的关联系数 $\xi_{0j}(i)$ 可用下述关系式表示:

$$\xi_{0j}(i) = \frac{\min_i \min_j |x_0(i) - x_j(i)| + \rho \max_j \max_i |x_0(i) - x_j(i)|}{|x_0(i) - x_j(i)| + \rho \max_j \max_i |x_0(i) - x_j(i)|} \quad (2-1)$$

式中 $\xi_{0j}(i)$ 称为 x_0 对 x_j 在 i 时刻的关联系数。

记各时刻两层次 (即两级) 的最小绝对值差为

$$\Delta_{\min} = \min_j \min_i |x_0(i) - x_j(i)|$$

记两层次 (即两级) 的最大绝对值差为

$$\Delta_{\max} = \max_j \max_i |x_0(i) - x_j(i)|$$

式(2-1)中的 ρ 为分辨系数,其作用在于提高关联系数之间的差异显著性。在一般情况下取 $\rho \in (0.1, 1)$,通常可取 $\rho=0.5$ 。

2.2.3 关联度

两个系统或者两个因素间关联性大小的度量,称为关联度。

关联度描述了系统发展过程中,因素间相对变化的情况,也就是变化大小、方向与速度等的相对性。如果两者在发展过程中,相对变化基本一致,则认为两者关联度大;反之,两者关联度就小。关联分析的实质,就是对数列曲线进行几何关系的比较。若两数列曲线重合,则关联性好,即关联系数为1。那么两数列关联度也等于1。同时,两数列曲线不可能垂直,即无关联性,所以关联系数大于0,故关联度也大于0。因为关联系数是曲线几何形状关联程度的一个度量,由公式(2-1)可见,在比较全过程中,关联系数不止一个。因此,我们取关联系数的平均值作为比较全过程的关联程度的度量。

关联度记为 r_{0j} ,其表达式为

$$r_{0j} = \frac{1}{N} \sum_{i=1}^N \xi_{0j}(i) \quad (2-2)$$

式中: r_{0j} ——子数列 j 与母数列0的关联度;

N ——数列的长度,即数据个数。

2.2.4 关联序

当比较数列有 m 个时,相对的关联度也有 m 个,若将 m 个关联度按其值大小排列起来,即为关联序。关联序直接反映各比较数列(子数列)对母数列的优劣关系。

例2-1:中国1980年~1983年间各业产值与总产值统计数据如表2-1所示,试作关联分析,以 x_0 为母因素, x_i 子因素($i=1, 2, 3, 4, 5$),计算其关联度并排出关联序。

表 2-1 中国 1980~1983 年的统计数据

年份(年)	1980	1981	1982	1983
总产值(x_0)(亿元)	8.531	9.071	9.96	11.052
农业产值(x_1)(亿元)	2.18	2.46	2.785	3.121
工业产值(x_2)(亿元)	4.897	5.12	5.506	6.088
建筑业产值(x_3)(亿元)	0.767	0.747	0.912	1.034
运输业产值(x_4)(亿元)	0.247	0.254	0.286	0.313
商业产值(x_5)(亿元)	0.44	0.49	0.474	0.496

解:第一步,将原始数据作初值化处理得:

$$x_0 = \{1, 1.063, 1.168, 1.296\}$$

$$x_1 = \{1, 1.128, 1.278, 1.432\}$$

$$x_2 = \{1, 1.046, 1.124, 1.243\}$$

$$x_3 = \{1, 0.974, 1.189, 1.348\}$$

$$x_4 = \{1, 1.028, 1.158, 1.267\}$$

$$x_5 = \{1, 1.114, 1.077, 1.127\}$$

第二步, 求绝对差

即

$$\begin{aligned}\Delta_{0j} &= |x_0(i) - x_j(i)| (j = 1, 2, 3, 4, 5; i = 1, 2, 3, 4) \\ \Delta_{01} &= \{\Delta_{01}(1), \Delta_{01}(2), \Delta_{01}(3), \Delta_{01}(4), \Delta_{01}(5)\} \\ &= \{0, 0.065, 0.11, 0.136\} \\ \Delta_{02} &= \{\Delta_{02}(1), \Delta_{02}(2), \Delta_{02}(3), \Delta_{02}(4), \Delta_{02}(5)\} \\ &= \{0, 0.017, 0.044, 0.053\} \\ \Delta_{03} &= \{\Delta_{03}(1), \Delta_{03}(2), \Delta_{03}(3), \Delta_{03}(4), \Delta_{03}(5)\} \\ &= \{0, 0.089, 0.021, 0.053\} \\ \Delta_{04} &= \{\Delta_{04}(1), \Delta_{04}(2), \Delta_{04}(3), \Delta_{04}(4), \Delta_{04}(5)\} \\ &= \{0, 0.035, 0.01, 0.029\} \\ \Delta_{05} &= \{\Delta_{05}(1), \Delta_{05}(2), \Delta_{05}(3), \Delta_{05}(4), \Delta_{05}(5)\} \\ &= \{0, 0.051, 0.091, 0.169\}\end{aligned}$$

第三步, 求两级最大差与最小差:

由第二步可得最大差与最小差为

$$\begin{aligned}\Delta_{\max} &= \max_j \max_i |x_0(i) - x_j(i)| = 0.169 \\ \Delta_{\min} &= \min_j \min_i |x_0(i) - x_j(i)| = 0\end{aligned}$$

第四步, 求关联系数:

$$\xi_{0j}(i) = \frac{\Delta_{\min} + \rho \cdot \Delta_{\max}}{\Delta_{0j}(i) + \rho \cdot \Delta_{\max}}$$

若取 $\rho=0.5$, 可得关联系数 $\xi_{0j}(i)$:

$$\begin{aligned}\xi_{01}(1) &= \frac{0 + 0.5 \times 0.169}{0 + 0.5 \times 0.169} = 1 \\ \xi_{01}(2) &= \frac{0 + 0.5 \times 0.169}{0.065 + 0.5 \times 0.169} = 0.564 \\ \xi_{01}(3) &= \frac{0 + 0.5 \times 0.169}{0.11 + 0.5 \times 0.169} = 0.434 \\ \xi_{01}(4) &= \frac{0 + 0.5 \times 0.169}{0.136 + 0.5 \times 0.169} = 0.382\end{aligned}$$

同理可得

$$\begin{aligned}\xi_{02}(1) &= 1, \xi_{02}(2) = 0.825, \xi_{02}(3) = 0.383, \xi_{02}(4) = 0.617 \\ \xi_{03}(1) &= 1, \xi_{03}(2) = 0.485, \xi_{03}(3) = 0.799, \xi_{03}(4) = 0.615 \\ \xi_{04}(1) &= 1, \xi_{04}(2) = 0.706, \xi_{04}(3) = 0.894, \xi_{04}(4) = 0.748 \\ \xi_{05}(1) &= 1, \xi_{05}(2) = 0.626, \xi_{05}(3) = 0.481, \xi_{05}(4) = 0.333\end{aligned}$$

第五步, 求关联度:

由

$$r_{0j} = \frac{1}{N} \sum_{i=1}^N \xi_{0j}(i)$$

得

$$r_{01} = \frac{1}{4} \sum_{i=1}^4 \xi_{01}(i) = \frac{1}{4} [\xi_{01}(1) + \xi_{01}(2) + \xi_{01}(3) + \xi_{01}(4)]$$