



统计知识(二)

张俊杰 主编

## 目 录

统计小常识 (续) .....	1
统计名言 .....	1
如何阅读《统计公报》 .....	5
什么是季节变动及调整? .....	5
什么是“贡献率” .....	6
CPI 是怎么得来的? .....	8
统计方法 .....	9
非线性回归分析技术要点 .....	9
非线性回归分析实例研究 .....	12
模型参数灵敏度分析 .....	21
周期组合模型 .....	24
模糊数学方法 .....	32
灰色系统方法 .....	58
BP 神经网络模型 .....	79
其它统计方法 .....	87
模糊聚类分析 .....	88
模糊模式识别 .....	93
模糊相似优先比方法 .....	98
模糊综合评判 .....	103
模糊关系方程求解 .....	107
综合评判的逆问题 .....	110
中华人民共和国统计法实施细则 .....	114
总则 .....	114
统计调查计划和统计制度 .....	116
统计资料的管理和公布 .....	120
统计机构和统计人员 .....	121
奖励和惩罚 .....	127

附则 .....	129
统计报告案例 .....	129
2003 年安西县社会统计分析报告 .....	129
《绍兴晚报》读者调查问卷的统计分析报告 .....	135
全省中小学校图书馆(室)建设综合统计分析报告(2001 年度) .....	139

## 统计小常识(续)

### 统计名言

至于统计，如果没有坚强的组织，尤其是 这一点在最初的章程中曾专门指出 如果没有总的领导，这项工作是无法完成的。马克思：《所谓国际内部的分裂》《马克思恩格斯全集》第 18 卷第 44 页。

一般的社会统计，特别是经济统计，最近二三十年来作出了巨大的成绩。有许多问题，而且是涉及到现代国家的经济制度和这种制度的发展的最根本的问题，过去是根据一般的估计和大致的材料加以解决的，现在如果不根据某一个一定的纲要收集并经统计专家综合的关于某一国家全国情况的浩繁材料，就无法加以比较认真地研究。尤其是争论最多的农业经济问题，更加要求根据精确的和大量的材料作出回答，况且在欧美各国所有农户进行定期调查，已经愈来愈成为一种惯例。列宁：《现代农业的资本主义制度》(1910 年)《列宁全集》第一版，第 16 卷，第 420 页。

由于分类的方法不同，同一个材料竟得出完全相反的结论。

如果根据土地面积来判断农户规模，结论是随着农户规模的扩大，农业的集约程度在下降；如果根据农户产值来判断农户规模，结论是随着农户规模的扩大，农业的集约程度在提高……。对于这个特别重要的方面，

现时的经济学家和统计学家注意得最不够。列宁：《关于农业中资本主义发展规律的新材料》(1915年)，《列宁全集》第1版，第22卷，第57—60页。

任何建设工作，任何国家工作，任何计划工作，没有正确的计算是不可想象的。而没有统计，计算是不可想象的。斯大林：《俄共(布)第十三次代表大会》(1924年5月23—31日)《斯大林全集》第6卷，第189页。

计算和监督，——这就是把共产主义社会第一阶段“调整好”，使它能正常地运转所必需的主要条件。列宁：《国家与革命》《列宁选集》第3卷，第258页。

当然，在形式上，叙述方法必须与研究方法不同。研究必须充分地占有材料，分析它的各种发展形式，探寻这些形式的内在联系。只有这项工作完成以后，现实的运动才能适当地叙述出来。这点一旦做到，材料的生命一旦观念地反映出来，呈现在我们面前的就好像是一个先验的结构了。马克思：(《资本论》第一卷第二版跋)(1873年1月24日)《马克思恩格斯选集》第三卷第217页。

统计工作不是把数字随便填到几个格格里去，而应当是用数字来说明所研究的现象在实际生活中已经充分呈现出来或正在呈现出来的各种社会类型。列宁：《莫斯科省的工作日和年工作》(1912年8月12日)《列宁全集》第一版，第18卷第254页。

在我生病期间(现在但愿病很快会痊愈)，我是无法写作的，但是，我吞下了大批统计学方面和其他方面的“材料”，对于那些肠胃不习惯于这类食物并且不能把它们迅速消化的人来说，这些材料本身就足以致病。马克思：《致路德维希·库格曼》(1868年3月6日)《马

克思恩格斯全集》第 32 卷第 526 页。

现在我们可以看到了一个可叹的事实：地方自治局的统计汇编化费了无穷无尽的劳动，收集了极丰富的、有价值的、新的(实行 11 月 9 日法会的结果!)然而没有加以结算、总结分组和复合的资料，这个汇编只有几乎等于零的一点微不足道的科学价值。列宁：《谈谈关于地方自治局统计任务的问题》(1914 年 1 月《列宁全集》第一版第 20 卷第 74 页。

为了从事实际工作，我们必须掌握数字，而中央统计局应该比谁都更早地掌握这些字。列宁：《给中央统计局的信》《列宁全集》第 33 卷第 12 页。

我很清楚知道，干巴巴的统计数字是非常不适宜作口头报告的，是会把听众吓跑的。但是为了使你们有可能来估价整个运动的真正客观基础，我还是不能不引证一些化成整数的数字。列宁：《关于 1905 年革命的报告》(1917 年 1 月 9 日)《列宁全集》第 1 版，第 23 卷，第 245 页。

我们说，现在我们的任务已经不是剥夺者，而是统计、监督、提高劳动生产率和加强纪律。列宁：《莫斯科省第七次党代表会议》(1921 年 10 月 29 日—31 日)《论新经济政策》。《列宁全集》第一版，第 33 卷，第 65—66 页。

量和质。数字是我们所知道的最纯粹的量的规定。但是它充满了质的差异。恩格斯：《自然辩证法》《马克思恩格斯选集》第 3 卷，第 69 页。

精明能干的经济学家不会去编制毫无意义的提纲，而会去细心研究事实、数字和材料，分析我们自己的实际经验，然后指出：我们在某某地方犯了错误，要如此

这般加以改正。精明能干的行政管理人員一定会根据这种研究，提出建议或自行采取措施，来调换工作人员改变月报制度，改组机构等等。在我们这里还没有看到过有人有这两种切实的态度来对待统一的经济计划。列宁：《论统一的经济计划》（1921年2月21日）《列宁全集》第二版，第40卷，第351—352页或《列宁全集》第4卷，第474页。

研究政治经济学不能随随便便，不能没有任何基础知识，不能不了解很多极重要的历史问题、统计学问题及其他问题。列宁：《书评》（亚·波格丹诺夫）《政治经济学简明教程》《列宁全集》第二版，第4卷，第5页。

我们相信中央统计局是科学的堡垒。我们认为如果没有中央统计局的数字，任何一个管理机关都不能进行计算和计划工作。我们认为中央统计局应该提供不受任何偏见影响的客观材料，因为使数字适合于某种偏见的企图是一种带有刑事性质的犯罪行为。既然如此，如果中央统计局自己都不相信自己的数字，那么人家又怎能相信它的数字呢？斯大林《联共（布）第十四次代表大会》（1925年12月18—31日）《斯大林全集》第7卷，第271—272页。

统计是社会主义建设的一项重要基础工作。我国要实现工业，农业、科学技术和国防现代化，必须实现统计工作的现代化。（《国务院关于加强统计工作的决定》1984年1月6日）。

胸中有“数”。就是说，对情况和问题一定要注意它们的数量方面，要有基本的数量分析。任何质量都表现为一定的数量，没有数量也就没有质量。我们有许

多同志至今不懂得注意事物的数量方面，不懂得注意基本的统计、主要的百分比，不懂得注意决定事物质量的数量界限，一切都是胸中无“数”，结果就不能不犯错误。毛泽东：《党委会的工作方法》《毛泽东选集》第4卷，第1380—1381页。

## 如何阅读《统计公报》

阅读《统计公报》，首先应将几个部分联系起来，不能割裂地单看某一个部分，否则就不能从总体上把握国民经济和社会发展的面貌。其次，要弄清公报中一些概念和术语含义。如什么是第一产业？什么是第二产业？国内生产总值（GDP）到底代表什么含义？等等。第三，最好是将每一年的公报连续对照起来看，从历年数量变化的轨迹中找到规律性的东西，将定量分析与定性分析结合起来，从而在更高层上观察、分析、把握社会经济形势。第四，可以对公报中有关数据进行必要的加工处理，从而了解社会经济生活的种种特色。如可从中计算出全国每天生产多少煤、电，每天完成多少基本建设项目、新建多少公路，每个人每天创造多少收入，每天每时出生多少人口等等，这不仅使得枯燥的数字可以生动化、生活化，也可增添您对生活的量化概念。

## 什么是季节变动及调整？

在一年之内，由于季节的变动，会使某些社会经济现象（一定的时间序列）产生规律性的变化，这种规律性变化通常称之为季节变动。时间序列的变化会受到社

会经济系统中各种因素的影响。从总体上看,影响时间序列变化的因素一般可以分为三类:第一类为纯粹的时间因素,如气候、日历天数和节假日等有关的季节性因素(用 S 来表示);第二类是与不同的发展阶段有关的趋势性因素(用 T 来表示);第三类是经济变量有规律的相互影响,这种相互影响随时间的延续而有规律地发展,表现在时间序列上就是有一定规律的周期波动(用 C 来表示)。另外,还有一些偶然性因素,如社会心理、政治因素等(用 I 来表示)。因此时间序列一般可以表示为:  $Y=T+C+I+S$ 。为更准确地反映客观经济现象的本质,必须事先对季节变动因素作一定消除和调整。

季节调整就是将一个时间序列分解成以上各部分。目前世界上主要有以下两种季节调整方法;X-11 方法、贝叶斯方法。我国一般采用美国商业部普查局开发的为官方使用并在世界上有着广泛影响的 X-11 季节调整程序。

## 什么是“贡献率”

在统计分析中经常使用“贡献率”,那么“贡献率”是什么含义?它是怎样计算的?

贡献率是分析经济效益的一个指标。它是指有效或有用成果数量与资源消耗及占用量之比,即产出量与投入量之比,或所得量与所费量之比。计算公式:

贡献率(%)=贡献量(产出量,所得量)/投入量(消耗量,占用量)×100%

贡献率也用于分析经济增长中各因素作用大小的程度。

计算方法是：

贡献率(%) = 某因素贡献量(增量或增长程度) / 总贡献量(总增量或增长程度) × 100%

上式实际上是指某因素的增长量(程度)占总增长量(程度)的比重。

举例说明如下：

总资产贡献率(%) = (利润总额+税金总额+利息支出) / 平均资产总额 × 100%

(1)总资产贡献率：反映企业资金占用的经济效益，说明企业运用全部资产的收益能力。

(2)社会贡献率：是衡量企业运用全部资产为社会创造或支付价值的 ability。

社会贡献率(%) = 社会贡献总额 / 平均资产总额 × 100%

社会贡献总额包括工资、劳保退休统筹及其他社会福利支出、利息支出净额、应交增值税、产品销售税金及附加、应交所得税及其他税、净利润等。为了反映企业对国家所作贡献的程度，可按上述原则计算贡献率。

企业对国家的贡献率(%) = (税金总额+上缴利润) / 社会贡献总额 × 100%

技术进步对产出增长速度的贡献率

这个指标是指在产出增长速度中，技术进步因素所占的比重，综合反映了技术进步对经济增长作用的大小。

技术进步对产出增长速度的贡献率(%) = 技术进步速度 / 产出增长速度 × 100%

上式贡献率越大则表明技术进步对经济增长的贡献和作用就越大，反之则小。

(4)各产业贡献率：

第一、二、三产业增量与国内生产总值增量之比，即为各产业的贡献率。

第三产业贡献率=第三产业当年增量/国内生产总值当年增量×100%

应该注意的是，贡献率指标比较抽象，在使用时，应说明具体含义，但也不能任意使用，要符合常规，做到标准化、规范化、通俗化。如资本收益率、资金利税率以及某些对增量因素分析的指标，已有专用名称，就没有必要改称为贡献率。另外，在计算各产业贡献率时应剔除价格变动因素，分子、分母均用可比价格的增量计算。

## CPI 是怎么得来的？

大多数国家都编制居民消费价格指数(CPI),反映城乡居民购买并用于消费的消费品及服务价格水平的变动情况，并用它来反映通货膨胀程度。

从2001年起，我国采用国际通用做法，逐月编制并公布以2000年价格水平为基期的居民消费价格定基指数，作为反映我国通货膨胀(或紧缩)程度的主要指标。经国务院批准，国家统计局城调总队负责全国居民消费价格指数的编制及相关工作，并组织、指导和管理各省区市的消费价格调查统计工作。

我国编制价格指数的商品和服务项目，根据全国城乡近11万户居民家庭消费支出构成资料和有关规定确定，目前共包括食品、烟酒及用品、衣着、家庭设备用品及服务、医疗保健及个人用品、交通和通讯、娱乐教育文化用品及服务、居住八大类，251个基本分类，约

700 个代表品种。居民消费价格指数就是在对全国 550 个样本市县近 3 万个采价点进行价格调查的基础上,根据国际规范的流程和公式算出来的。

## 统计方法

### 非线性回归分析技术要点

#### 1. 求解参数的操作过程

在进行建模分析之前,用户必须从实际问题出发,确定一个合理的模型表达式。这一点对于从事相关学科的专业人员来讲,根据实验数据、并定义数学模型是不难的。要注意的是,在 DPS 数据处理平台上正确输入相应的公式和参数,操作步骤如下:

首先,在电子表格中编辑数据和定义数据块,一行为一个样本,一列为一个变量,并将待分析的数据定义成数据块。在数据块中,第 1 列为  $x_1$ ,第 2 列为  $x_2$ ,余此类推,最后一列为  $x_p$ 。

定义数据块后,再在屏幕下部文本编辑器窗口中写入数学模型表达式,并用鼠标定义产成公式块。对需要进行模拟和参数求解的数学模型,必须将其定义成系统能够识别处理的形式,第 1 行输入待拟合的数学方程表达式,第 2 行输入待估参数的初始值和加权的变量(或表达式)。

数学方程表达式分两个部分:等号左边为因变量(或是因变量的计算表达式),等号右边是由自变量和待估参

数组组合起来的表达式。公式中待求参数用  $c_1, c_2, \dots, c_m$  表示, 它们必须从 1 开始按顺序定义。用  $x_1, x_2, \dots, x_p$  代表数据矩阵中的各列数据( $c, x$  大小写字母输入均可)。至如此, 一个基本数学模型在本系统中定义完毕。

第 2 行中, 先输入各待求参数的初始值, 各初始值之间用空格隔开。然后, 根据用户需要与否, 用 “//=” 引导, 输入在进行非线性最小二乘分析时要加权的变量或表达式。当然, 第 2 行也可只放入各待求参数的初始值, 或只放入由 “//=” 引导的加权变量或表达式, 或者什么也不输入(公式块只定义第 1 行)。

公式(包括待求参数的初值或加权变量)按上述格式编辑好之后, 拖动鼠标——把公式及其参数初值这两行拖黑。

为充分利用系统资源, 在定义公式时需要注意以下事项:

(1) 公式块中可直接引用本系统提供的全部标准函数。

(2) 表达式右边可允许使用差分变量;

(3) 等号左端只能输入表达式, 等号右端则可允许输入函数或表达式。

本系统可以分析处理 3000 个数据, 512 个样本, 求解 20 个未知参数, 一般可满足常规研究需要。

## 2. 关于参数初值的定义

如果用户定义的数学模型过于复杂, 系统在自动给出参数初值的基础上仍不能完成参数的拟合求解, 这时用户需要根据自己掌握的专业知识和研究经验及试验数据的特征, 对待拟合的参数给出较为合理的初值, 以使得收敛尽快发生, 使模型拟合能够顺利进行。对初始值

的确定，有以下几个简单而有用的原则：

(1)分析模型的性质。非线性回归模型的参数对科学家或研究人员来说通常是有意义的。这种意义可能表现在图象、生物学、物理学、化学或其它合适的形式上，因而对确定初始值可能很有用。某些参数的初始值可能从相关试验中获得，也可以根据函数图象，了解参数怎样影响函数的表现。根据模型的基本特征如拐点、极值等特征确定待模拟参数的初始值。生物学中常用的许多曲线方程，如逻辑斯蒂方程，方程本身含有参数的极限值，其大体范围是不难确定的。若是最大值，欲设定的初值可比观察值中的最大值略大一些；如是极小值，待设定初值可比观察值中的最小值略小一些。

例如在 Michaelis-Menten 酶反应模型  $f = \theta_1 x / (\theta_2 + x)$  中，参数  $\theta_1$  是酶反应的渐近速度，因而能够由试验数值的最大观察值估计；类似地， $\theta_2$  表示“半浓度”，即这样一个  $x$  值，当浓度达到该值时，速度是最大值的一半。

(2)分析一下，是否可对模型作适当转换，以获得更简单的关系，或使之线性化。将模型线性化的一般方法是取对数，如方程中有指数项，可对方程两边同时取对数，使方程线性化后，再对该线性方程进行计算以获得初始值。如前述柯布—道格拉斯方程两边取对数后可转化为线性方程。

例如，Bard(1974)在研究一个化学动力学的实例时应用了模型  $f = \exp(-\theta_1 x_1 \exp(-\theta_2/x_2))$ 。对此模型，取 2 次对数后可得到  $\ln(\ln(f)) = \ln x_1 + \ln(-\theta_1) - \theta_2/x_2$ 。该模型是一个线性模型。这时可应用线性最小二乘法来获得初始值。

(3)简化原方程。这意味着将方程从简到繁逐步求

解。有些在生物学研究中应用的模型是由一些较简单(基本)的方程演化而来的。例如从逻辑斯蒂方程衍生出来的模型有：

$$Y = \frac{K}{1 + e^{a+bx+cx^2}}$$

$$Y = \frac{K}{1 + \exp(-r(T - T_0))} (1 - \exp(-\frac{T - T_L}{\delta_L})) (1 - \exp(-\frac{T_M - T}{\delta_M}))$$

对此类模型，可先求前面的基本部分，然后再逐步增加，最后求出完整的参数。

(4)在采用以上几个方法都不能得到参数的精估计值时，可采用网格法寻求参数的初值，即对每个待估计的参数给出一个区间范围和搜索步长，进行计算，该法计算量较大，但在运算速度很快的计算机上是不难完成的。

## 非线性回归分析实例研究

### 1. 普通非线性模型

这类模型一般不含指数或复杂的数学函数，而仅仅是一些加减乘除的组合。这类模型建模时，只需要将试验数据编辑定义之后，将希望建立的数学模型(公式)进行编辑定义，一般无需给出各个参数的初值(因系统默认各个参数的初值为 0.01)就可以使用麦夸特法快速地求出模型中的各个参数。

例 1 研究“岱字棉”自播种至齐苗(以 80%出苗为准)期的天数( $Y$ )和日平均土温( $X$ , )的关系,经试验得到数据后欲建非线性经验模型(莫惠栋 1984)。根据有效积温模型,描述自播种至齐苗期天数和日平均土温相互关系

最直观的回归方程的数学表达形式为

$$Y=a/(X-b) \quad (21.16)$$

数据和公式块的定义见图 21-1 所示。图中上部为数据区，下部为公式区，数据块以灰底蓝字显示，公式块以黑底白字（反相）显示。

由于因变量  $Y$  位于数据块的第 2 列，自变量  $X$  位于数据块的第 1 列，因此令  $x_2=Y$ ,  $x_1=X$ 。因为式(21.16)中有两个待求参数  $a$  和  $b$ ，定义模型(方程)时则分别用  $c1$  和  $c2$  表示，最后定义公式块，见图 21-1 下部（黑底白字）反相显示部分。



图 21 - 1 播种至齐苗天数和日均土温相互关系模型分析

编辑定义数据和公式块后，进入菜单，选择“数学模型→单因变量参数估计→麦夸特方法”后，按回车键即刻进入分析提示界面，在屏幕的右偏下部分显示当前各参数估计、拟合误差的动态信息，并有文字提示。如果迭代结果已达到要求，或用户需要提前终止迭代计算，

可点击第一排快捷按钮最右边的那个按钮 (该按钮只有在进行计算时才显示出来)。在本例中,用麦夸特迭代模拟方法成功地给出拟合结果(图 21-1)。

如图 21-1 所示,返回的计算结果包括方差分析表、系数  $c_i$  的协方差阵(略)、系数  $c_i$  的相关阵(略)、参数拟合值及其标准误差、 $t$  测验值和相应的显著水平,最后给出各样本因变量的观察值、拟合值、拟合误差、标准残差、cook 距离和杠杆率  $H$ ,供模型诊断使用。

从以上拟合结果看, $F$  检验值为 1108.34,达极显著水平;相关系数为 0.9983,拟合度 99.64%,均优于经线性化转换后的拟合效果。将拟合结果代回原方程,有

$$T = \frac{50.7517}{X - 14.6138}$$
。如果进行区间估计,岱字棉从播种到齐苗

的发育起点温度为  $(14.61 \pm 0.19)$ ,有效积温为  $(50.75 \pm 2.36)$ 。

如果求置信概率为 95%的起点温度,需先求出自由度为 4 时的  $t_{0.05}$  临界值,可在电子表格中键入“=ttest(4,0.05)”,而后立即显示结果是 2.7764,由此求得置信概率 95%的发育起点温度为  $(14.61 \pm 2.7764 \times 0.19)$ ,有效积温为  $(50.75 \pm 2.7764 \times 2.36)$ 。

## 2 含有指数或某些函数的非线性模型

此类非线性模型因含指数或其它数学函数,往往不宜或无法转换为线性模型求解。因此对这类模型的参数估计,必须借助于非线性的迭代方法(如 Newton-Raphson 方法)求解。这对不十分熟悉或精通数理知识的生物学工作者来说是很比较困难的。但是,在我们提供的 DPS 处理平台上,用户只需将试验数据和希望建立的数学模型(公式)编辑定义成数据和公式块,便

可很快获得分析结果。因此，DPS 平台使用户在建立模型时，**所想即所见，所见即所得！**

例 2 植物病毒侵染—稀度数学模型的参数拟合。数学模型为  $y = b_1 \ln(1 + b_2 V)$ ，式中  $b_1$  和  $b_2$  为待定参数， $V$  为病毒浓度， $y$  为病毒侵染后的半叶平均枯斑数。由于上式是非线性模型，且用任何转换方法都不能直接地化为线性模型来求解参数，因此必须采用非线性迭代方法求解。Furumoto&Mickey(1967)、裴新德(1987)曾给出该模型的参数拟合方法并取得较好拟合效果。不过他们所给出的方法都是对方程两端施行对数转换，并取得了较好的最优拟合效果，但还原后的效果却并不理想。现仍利用 Furumoto&Mickey(1967)的实验数据，在此作为介绍操作步骤的例子，将其第 2 组资料的拟合过程简述如下：

分析第 2 组数据时，其因变量(半叶平均枯斑数  $y$ ) 处于第 2 列，故在分析模型中的变量名为  $x_2$ ，自变量(接种病毒浓度  $V$ ) 在第 1 列，故变量名为  $x_1$ 。定义公式时，令  $c_1 = b_1$ ， $c_2 = b_2$ ，数据和公式按图 21-2 的方式进行定义。

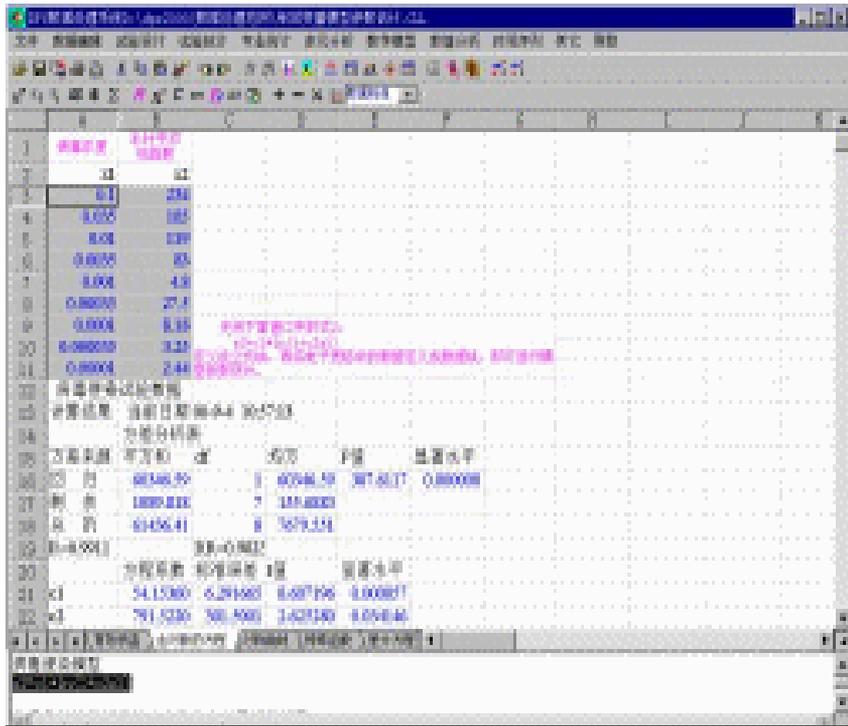


图 21-2 病毒侵染模型数据、公式的编辑与定义

编辑定义数据和公式块后，进入菜单，选择“数学模型→单因变量参数估计→麦夸特方法”后，按回车键进入分析提示界面。

分析结束时，系统自动输出方差分析表、系数  $c_i$  的协方差阵、系数  $c_i$  的相关阵、模型参数拟合值及其标准误差、 $t$  测验值和相应的显著水平，最后也给出了各样本因变量的观察值、拟合值、拟合误差、标准残差以及 cook 距离和杠杆率  $H$ ，供诊断模型时使用。

根据计算结果，将  $c_1$ 、 $c_2$  代回原方程，得第 2 组资料的植物病毒侵染—稀度的数学模型为  $y=51.1536 \ln(1+791.522I)$ 。  $F$  检验值为 1108.34,  $p < 0.0001$ , 达极显著水平。相关系数为 0.9911, 拟合度 98.23%, 拟合效果很好。

例 3 这是一个在分析过程中,需要给出各个参数初值才能进行有效拟合的例子。根据 106 个岱字棉单株纤维长度的次数分布的组值( $X$ )和次数( $Y$ )资料(莫惠栋 1984),编辑整理如图 21-3。经分析两者间关系呈对称状态,可用指数型非线性方程表为

$$Y = a \cdot e^{b(x-c)^2}$$

由于因变量  $Y$  位于数据块的第 2 列,自变量  $X$  位于数据块的第 1 列,因此令  $x_2=Y$ ,  $x_1=X$ 。又因式中有 3 个待求参数  $a$ 、 $b$  和  $c$ ,在定义模型时则分别用  $c_1$ 、 $c_2$  和  $c_3$  表示,然后就可进行公式的定义。定义公式时,如果不给初值就进行参数拟合分析,那么无论是用麦夸特法还是用加速单纯形法,拟合结果都不收敛,因此必须给出各个参数的初值。显然,本例的对称点在  $x$  为 29~30 处,故参数  $c_3$  的初值应取 30 为宜。将参数的初值  $c_1=0.01$ ,  $c_2=0.01$ ,  $c_3=30$  放在公式块的第 2 行,并按图 21-3 下部阴影处的方式定义公式块之后就可进行分析了。

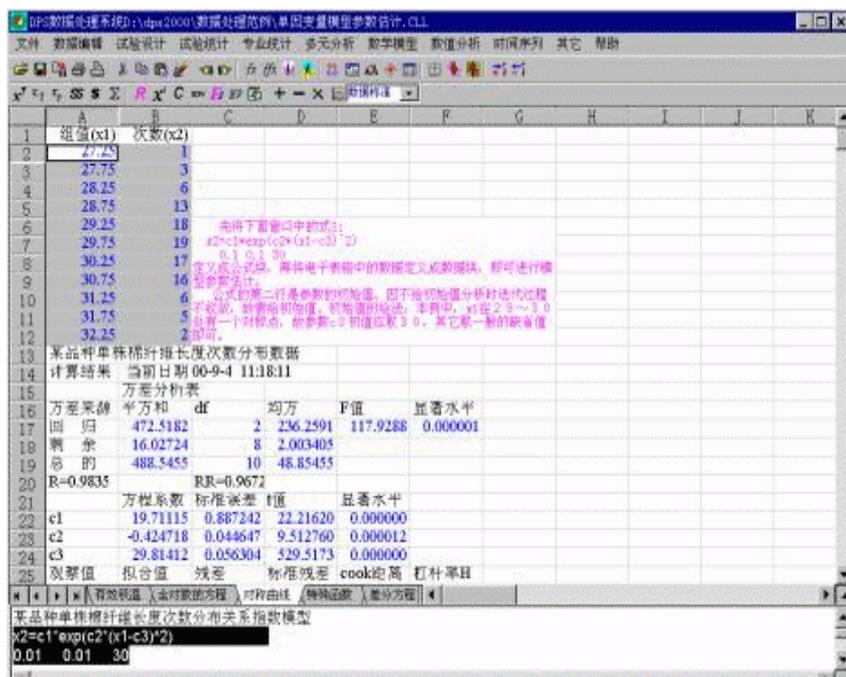


图 21-3 纤维长度次数分布组值( $X$ )和次数( $Y$ )关系建模界面

分析时,进入菜单,选择“数学模型→单因变量参数估计→麦夸特方法”后,按回车进入分析提示界面,经迭代系统拟合运算可成功地获得以下拟合结果:

其非线性经验模型为  $\hat{Y} = 19.71115e^{-0.424718X - 29.81412X^2}$ , 模型方差分析  $F$  值等于 117.9288,  $p=0.0001$ , 达极显著水平。模型确定系数  $R^2=0.9672$ , 因此拟合效果很好。

### 3 利用 DPS 提供的特殊函数建立非线性模型

#### 例 4 分段 Logistic 模型参数估计。

沈信毅(1987)提出了用分段的 Logistic 曲线描述玉米螟田间种群的季节性消长过程,该模型假定一个具有 Logistic 增长特征的  $N(t)$  在时刻  $t_c$  之前具有内禀增长率  $r_0$ , 时刻  $t_c$  之后它转变为  $r_1$ , 则  $N(t)$  在时刻  $t$  的动态模型可表示如下:

$$\frac{dN}{dt} = \begin{cases} r_0 N(1 - N/K) & t < t_c \\ r_1 N(1 - N/K) & t \geq t_c \end{cases} \quad (21.17)$$

其中  $K$  是环境容纳量,  $t_c$  为变点。设初值  $N(t_0) = N_0$ , 对上述模型积分可得到分段 Logistic 模型:

$$N(t) = \begin{cases} \frac{K}{1 + \frac{K - N(t_0)}{N(t_0)} \exp[-r_0(t - t_0)]} & t < t_c \\ \frac{K}{1 + \frac{K - N(t_c)}{N(t_c)} \exp[-r_1(t - t_c)]} & t \geq t_c \end{cases}$$

其中

$$N(t_c) = \frac{K}{1 + \frac{K - N(t_0)}{N(t_0)} \exp[-r_0(t_c - t_0)]}$$

为便于拟合模型参数, 式 21.17 可改写为

$$\frac{dN}{dt} = [r_0 + (r_1 - r_0)(t - t_c) \text{hazard}(t - t_c)] N(1 - \frac{N}{K})$$

式中 hazard 为 DPS 系统预定义的风险函数, 即该函数计算结果小于等于 0 时返回 0, 计算结果大于 0 时函数值返回 1。对上式积分, 可得

$$N = \frac{K}{1 + \frac{K - N(t_0)}{N(t_0)} \exp\{-[r_0(t - t_c) + (r_1 - r_0)(t - t_c) \text{hazard}(t - t_c)]\}}$$

若令  $\alpha = \ln \frac{K - N(t_0)}{N(t_0)} + r_0 t_0$ ,  $\beta = -r_0$ ,  $\gamma = r_1 - r_0$ , 则

$$N = \frac{K}{1 + \exp[\alpha + \beta t + \gamma(t - t_c) \text{hazard}(t - t_c)]} \quad (21.18)$$

根据该式, 即可对模型进行参数估计。

沈信毅(1987)应用约束最小二乘法给出了分段

Logistic 曲线的有关参数,但是模型中变点的选取是人为的,因此所求出的模型参数的转变点,不一定在转变点上且未指出变点是多少。李仲来(1997)应用 SAS 统计分析软件的 DUD(Doesn't used derivatives)法(Ralston and Jennrich, 1978),对该模型的参数重新进行了估计,获得了较好的结果。在这里,作者将直接根据式(21.18)拟合模型参数。

根据沈信毅(1987)资料,用 DPS 系统参数拟合技术估计分段 Logistic 模型参数,先根据公式定义,将待求参数  $K$ 、 $\alpha$ 、 $\beta$ 、 $\gamma$  和  $t_c$  分别用  $c_1$ 、 $c_2$ 、 $c_3$ 、 $c_4$  和  $c_5$  表示,并将参数  $c_1$  的初值定为 42(因  $c_1$  表示环境容纳量,故可取玉米螟密度最大值), $c_5$  的初值定位 30(因 27~30 天之后玉米螟种群数量下降,故估计转变点在 30 附近),其它参数初值置为 0.01。按图 21-4 方式定义数据和公式:

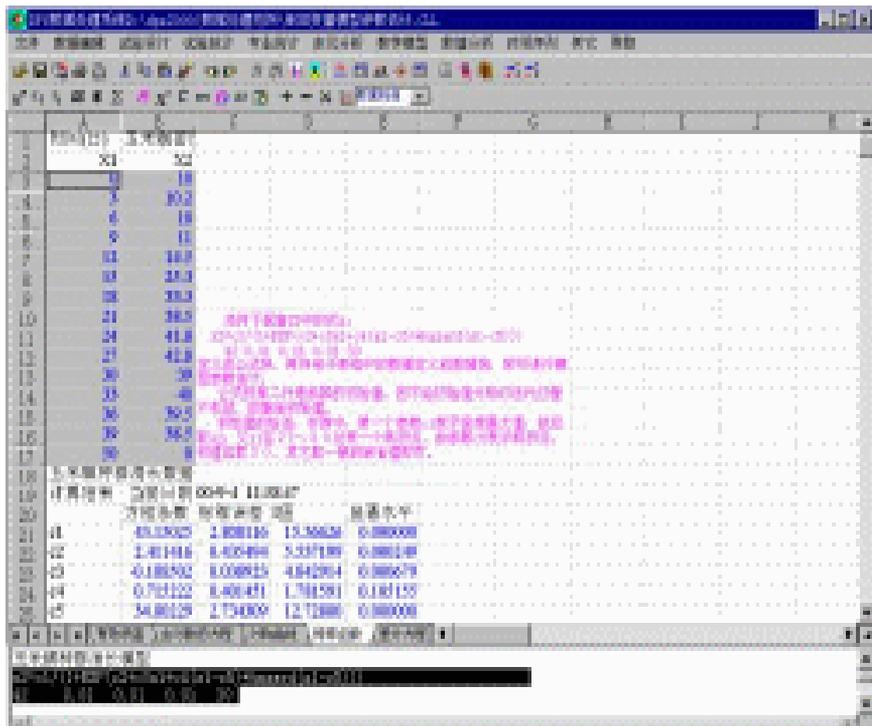


图 21-4 分段 Logistic 模型参数估计界面

然后选择“数学模型→单因变量参数估计→麦夸特方法”，按回车键，得到结果如下：

从拟合结果可以看出，所得估计值的剩余平方和为 148.55，比李仲来(1997)应用 SAS 系统中 DUD 法估计所得参数的剩余平方和 176 要小，因此应用 DPS 系统进行参数估计的效果更好。

## 模型参数灵敏度分析

建立数学模型的主要目的之一是增进我们对系统的了解，而模型参数的灵敏度分析是对数学模型的参数动态变化过程，即瞬时变化过程进行分析。因此，通过模型参数的灵敏度分析可以明确哪些参数对系统的总体输出和动态影响较大。

### 1. 方法简介

下面考虑一个经验模型，其模型的输出  $y$  可以是一种作物的产量，也可以是一头奶牛的总泌乳量等，假定对该模型已经圆满地进行了检验与评价，包括对试验数据进行了适度的拟合。模型中有些参数是生理指标，而有些是环境指标，另有一些参数，如对作物投入肥料  $x_i$ ，按其对产量的相对效应进行排序，则无论目标函数怎样，都可以得到较为客观的度量。目标函数  $y$  对参数  $x_i$  的灵敏度  $S(y, x_i)$  的定义为：

$$S(y, x_i) = \frac{\frac{\partial y}{\partial x_i}}{\frac{y}{x_i}}$$

式中  $\frac{\partial y}{\partial x_i}$  为边际函数(偏导数)， $y/x_i$  为平均的投入产出效应。 $S(y, x_i)$  表示目标函数  $y$  对输入参数  $x_i$

的灵敏度。

## 2. 灵敏度分析的计算机处理

在分析模型参数变化速率之前,先给出模型方程,并将模型及其参数按规定格式定义成公式块,按系统要求将待分析的公式放在公式块的第1行,这些待分析的公式系由变量和参数组合起来的表达式。在公式中用 $x_1, x_2, \dots, x_p$ 分别代表 $p$ 个变量(必须从1开始按顺序给出)。公式中可使用本系统的全部标准函数,最后的公式在形式上必须是合法的数学表达式,定义格式为:

方程表达式

变量1的起始值,终止值,间隔值或变量1的取值水平。

变量2的起始值,终止值,间隔值或变量2的取值水平。

.....

变量 $p$ 的起始值,终止值,间隔值或变量 $p$ 的取值水平

例如:

$21.5+4.29x_1-3.77x_2-0.059x_1^2-0.015x_2^2+0.0044x_1x_2$ , 12, 18, 0.1, 50

上述公式块中的第2行表示变量 $x_1$ 从12开始到18,每隔0.1分析一次,第3行表

示将 $x_2$ 固定为50。

例如研究产量随某种肥料用量变化的规律,求出的肥料反应经验方程为 $y = 3x + 2x^2 - 0.1x^3$ ,根据该方程进行模型的灵敏度分析,先按图22-7的方式编辑定义公式:

$3x_1+2x_1x_1-0.1x_1^3$

1160.5

图 22-7 产量函数灵敏度分析公式定义图

然后进入菜单操作,选择“模型参数变化速率分析”菜单,执行后系统即输出分析结果,包括灵敏度、导数(偏导数)、平均效应及目标函数的计算结果,详见如下输出单“

x1	灵敏度	导数(边际值)	平均效应	y/x	目标函数 y
0.0000	...	3.0000	...	0.0000	
0.5000	1.2390	4.9250	3.9750	1.9875	
1.0000	1.3673	6.7000	4.9000	4.9000	
1.5000	1.4416	8.3250	5.7750	8.6625	
2.0000	1.4848	9.8000	6.6000	13.2000	
2.5000	1.5085	11.1250	7.3750	18.4375	
3.0000	1.5185	12.3000	8.1000	24.3000	
3.5000	1.5185	13.3250	8.7750	30.7125	
4.0000	1.5106	14.2000	9.4000	37.6000	
4.5000	1.4962	14.9250	9.9750	44.8875	
5.0000	1.4762	15.5000	10.5000	52.5000	
5.5000	1.4510	15.9250	10.9750	60.3625	
6.0000	1.4211	16.2000	11.4000	68.4000	
6.5000	1.3864	16.3250	11.7750	76.5375	
7.0000	1.3471	16.3000	12.1000	84.7000	
7.5000	1.3030	16.1250	12.3750	92.8125	
8.0000	1.2540	15.8000	12.6000	100.8000	
8.5000	1.1996	15.3250	12.7750	108.5875	
9.0000	1.1395	14.7000	12.9000	116.1000	
9.5000	1.0732	13.9250	12.9750	123.2625	
10.0000	1.0000	13.0000	13.0000	130.0000	
10.5000	0.9191	11.9250	12.9750	136.2375	

11.00000.829510.700012.9000141.9000  
 11.50000.72999.325012.7750146.9125  
 12.00000.61907.800012.6000151.2000  
 12.50000.49496.125012.3750154.6875  
 13.00000.35544.300012.1000157.3000  
 13.50000.19752.325011.7750158.9625  
 14.00000.01750.200011.4000159.6000  
 14.5000-0.1891-2.075010.9750159.1375  
 15.0000-0.4286-4.500010.5000157.5000  
 15.5000-0.7093-7.07509.9750154.6125  
 16.0000-1.0426-9.80009.4000150.4000

以上结果可以作出以下解释：

第 1 阶段，灵敏度大于 1，这时的边际产量大于平均效应产量，且平均产量效应是增加的，当肥料投入量达到 10 个单位时，平均效应产量达到最高点。该点的  $x$  值约为 10。

第 2 阶段，灵敏度小于 1 但仍大于 0，目标函数在该阶段的终点达到最大值，而边际效应值下降到 0。这时的投入  $x$  约为 14。

第 3 阶段，灵敏度小于 0，目标函数趋于下降，平均效应虽为正值，但边际效应为负。

## 周期组合模型

### 1. 方法简介

第 30 章介绍的平稳时间序列分析，其基本形式是  $x_t = \mu_t + y_t$ ，通常称  $\mu_t$  为序列的确定性部分，而称  $y_t$  为零均值的平稳随机部分。ARIMA 模型是采取一些办法剔除

$\mu_t$  部分的作用, 这里所要介绍的组合模型不仅要趋势性和周期性分量分离出来, 而且要给出  $\mu_t$  的具体表达式, 因此最后建立起既有确定性又有随机性部分的组合模型, 由这两部分的组合来共同描述某些类型的非平稳过程, 常常能达到令人满意的效果。

建立组合模型的方法, 简单地说是选用最小二乘法按照某类函数拟合数据序列的确定性部分, 从低阶开始, 逐渐增加阶数, 直到模型无明显改进为止。然后对消除了确定趋势的残量序列建立适宜的 ARMA( $n, m$ ) 模型。最后, 用前述得到的两部分参数估值作为初值, 对确定性部分和 ARMA 部分的所有参数, 用非线性最小二乘法重新估计, 得出组合模型的最终估计。

我们对各种组合模型, 从简单情形到一般情形分别叙述如下:

(1) 线性趋势。数据序列含有线性趋势, 即观察数据在某一直线附近散布, 这是组合模型中最简单的一种情形。我们可以先用熟知的统计方法拟合回归直线, 然后对残差建立 ARMA( $n, m$ ) 模型, 最后把分别估计得到的两部分参数作为初值, 用非线性最小二乘法估计出组合模型的最终参数。

与此类似, 对于含有多项式趋势的序列, 我们采用适当阶数的多项式和 ARMA 模型叠加所构成组合模型来拟合。

(2) 指数趋势。数据序列有时呈现指数增长或衰减趋势, 这里所讨论的是实指数的情形。我们可用组合模

型拟合 
$$x_t = \sum_{j=1}^r A_j e^{K_j t} + y_t$$
。

其中  $K_j, A_j$  为实数,  $\{y_t\}$  为 ARMA 序列。建模的步骤

与(1)基本相同。对于  $r > 1$  即多个指数趋势混合的情形，求和阶数  $r$  的判别比较麻烦。我们要么根据实际物理背景决定应用有若干加项。要么采用使  $r=1, 2, 3, \dots$  由低到高的拟合办法，使  $r$  增加到组合模型的残差平方和无显著改进为止。

(3) 周期趋势。如果随机序列不仅含有指数衰减或增长趋势，而且还呈现出某种规律的周期性起伏，则可用如下形式的组合模型去拟合

$$x_u = \sum_{j=1}^L R_j e^{\gamma_j t} + \sum_{j=1}^K B_j e^{b_j t} \sin(j\omega t + \varphi_j) + y_t \quad (31.1)$$

其中  $\{y_t\}$  是  $ARMA(n, m)$  序列， $L$  为序列所含有的指数趋势项的项数， $K$  是周期趋势项的项数， $\varphi$  为基频，由数据的物理性质决定，单位为弧度/秒， $B_j$  和  $r_j$  分别表示周期趋势的振幅和相位， $e^{b_j t}$  控制周期振荡的增长或衰减趋势。由于在式(31.1)中，描述周期趋势所用的正弦函数是对称的，但在生物、社会、经济领域，周期并不是对称的，如1年或1天中的气温旬变化趋势，上升比下降要缓慢；生物种群数量的波动，先是较缓慢地上升，到达高峰后，较快（甚至是急剧）下降，等等。针对这种比较普遍的时间动态趋势，作者在 DPS 数据处理系统中，对式(31.1)进行了改进，即将式(31.1)改写为如下形式：

$$x_u = \sum_{j=1}^L R_j e^{\gamma_j t} + \sum_{j=1}^K B_j e^{b_j t} \sin(2\pi \left\| \frac{t}{s} \right\|^d + \varphi_j) + y_t \quad (31.2)$$

式(31.2)中的  $\left\| \cdot \right\|$  表示时间  $t$  除以周期长度  $s$  的小数部分， $0 \leq \left\| t/s \right\| \leq 1$ ， $d$  为指数。显然，当  $d$  等于1时，式(31.2)和式(31.1)是等价的；当  $d$  小于1时，式(31.2)表达的周期趋势为上升较下降快；但当  $d$  大于1

时,式(31.2)表达的周期趋势为上升较下降慢。

组合模型建模的具体步骤为:首先拟合线性或指数趋势,然后逐个地增添周期趋势项,最后对残差序列 $\{y_t\}$ 建立 ARMA 模型。各个分模型的参数都估计出来,再以它们作为初值,对整个组合模型(31.2)统一进行参数估计。在估计过程中需反复使用非线性最小二乘法。尽管模型结构比较复杂,待估计参数个数较多,建模过程还是比较麻烦,但在 DPS 数据处理系统的支持下,实现起来还是较容易的。

## 2. DPS 数据处理系统操作示例

在此,我们以时间序列分析中较经典的某航空公司 1949 年 1 月至 1960 年 12 月每月客票数(千人)时间序列资料为例,建立季节周期 - ARMA 组合模型。分析前,先将数据定义成图(31.1)所示的数据块。

在 DPS 数据处理系统中,执行“时间序列”中的“周期-ARMA 指数组合模型”功能。系统会自动检验时间序列是否具有上升、下降趋势。如果有这种趋势,系统会自动建立时间序列趋势方程、给出相关系数、并在屏幕上显示绿色的趋势曲线;给出提取趋势项后的残差序列数据曲线(红色),以及残差序列数据的正态分布检验的直方图、自相关图和偏自相关图(与 30 章平稳时间序列分析的工作界面类似)。本例显示操作界面如图 31.2 所示。

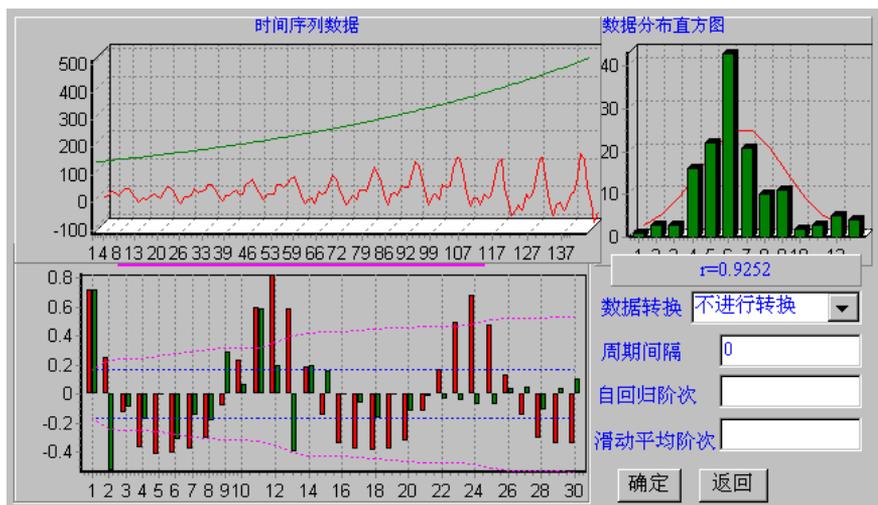


图 31.2 提取趋势项后的时间序列特征

从图 31.2 中可以看出,系统检验出该时间序列有上升趋势,并计算出趋势曲线方程的相关系数为 0.9252。从图 31.2 中残差序列的自相关函数可以看出,自相关函数有明显的周期性(在 12、24 处出现高峰),因此有理由提取以一年 12 个月为长度的季节周期。据此,在“周期间隔”的对话框中输入“12”,然后用鼠标点击“确定”按钮。这时,系统会显示提取周期后的时间序列特性如图 31.3。

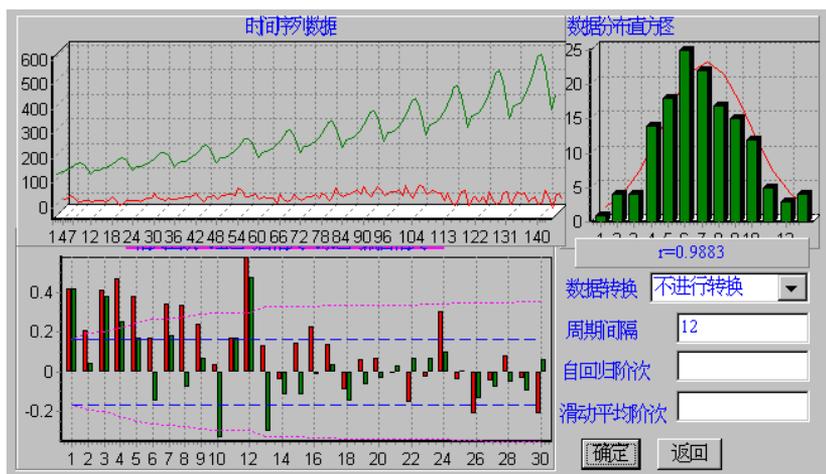


图 31.3 提取趋势项和 1 个周期项后时间序列特征

从图 31.3 中可以看出,系统将上升趋势项和以一年 12 个月为周期的周期项相加,得到复合的趋势 - 周期曲线为一上升的波浪形曲线(绿色),其复合方程的相关系数为 0.9883。此时,从图中的直方图可以看出残差序列数据呈正态分布。分析图 31.3 中残差序列的自相关函数不难发现,自相关函数在 4、8、12、16 和 24 处呈现峰值,即可能包含有以 4 个月长度的周期。因此可进一步提取以 4 个月为长度的周期,在“周期间隔”的对话框中输入“12”,然后加一空格,输入“4”,再用鼠标点击“确定”按钮。这时,系统会显示提取 2 个周期后的时间序列特性如图 31.4。

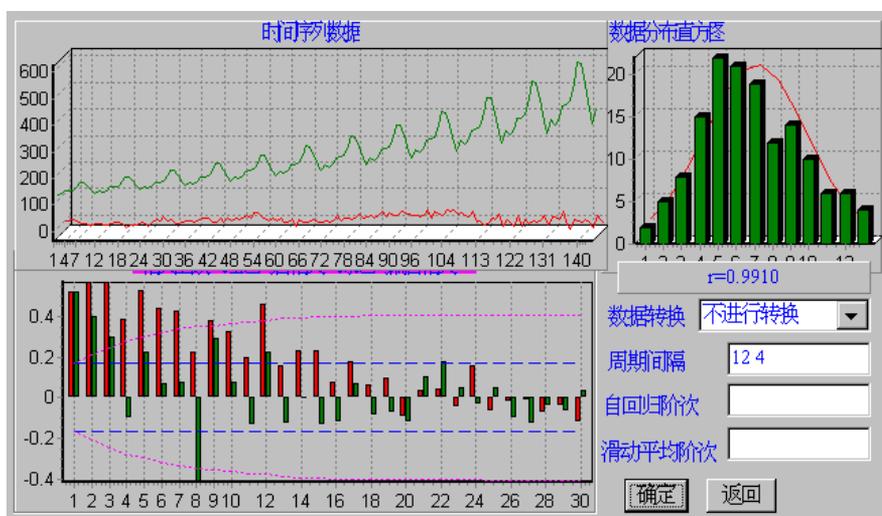


图 31.4 提取趋势项和 2 个周期项后的时间序列特征

从图 31.4 可以看出,这时的趋势 - 周期组合方程的相关系数达 0.9910,残差序列自相关已无明显的周期性,相关函数表明残差序列数据是平稳的。这时,数据序列的随机部分拟合,我们可拟合 ARMA(3,3)模型(自回归阶次取 3 为宜,滑动平均取前 3 个自相关函数较大的阶次)。

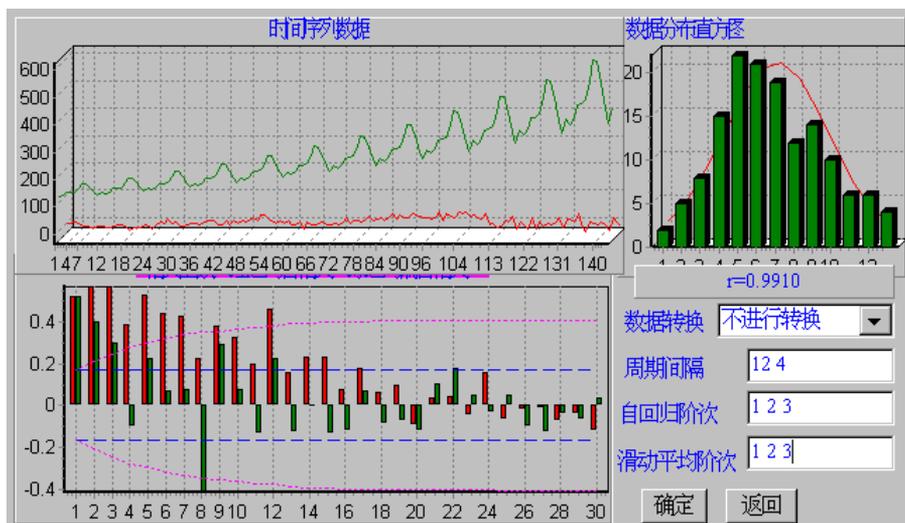


图 31.5 残差序列数据特性及 ARMA 建模参数的选取  
 如图 31.5 所示,在“自回归阶次”和“滑动平均阶次”均输入“123”(数字间用空空格开),然后点击“确定”按钮。最后,得到模型参数及参数检验结果如下

参数参数值标准误差 T 值

c1130.49655.94341021.95650

c20.0091600.00027034.08580

c321.884992.3771609.206300

c40.0121300.00075016.06910

c52.4128300.11235021.47600

c65.7114400.08480067.35460

c74.5883902.0068202.286400

c80.0096700.0041502.330100

c91.0109100.1688405.987600

c103.0702200.3077509.976300

c110.2116000.0752502.812000

c12-0.3223300.0697904.618500

c130.1547200.0670602.307300

c14-0.1110500.1269800.874600

c15-0.7343700.06880010.67470

c16-0.0411700.1284100.320600

c17-1.32036014.732910.089600

并得到组合方程  $y=x_{0t}+x_{1t}+x_{2t}+x_t$ , 其中:

时 间 序 列 趋 势 方 程

$$x_{0t}=130.49648 \cdot \text{EXP}(0.00916 t)$$

周期方程:

时间序列的第 1 个周期方程

$$x_{1t}=21.88499 \cdot \text{EXP}(0.01213 t) \cdot \sin\{2 \cdot 3.14159 \cdot | | t / 12.0 | |^{2.41283} + 5.711436\}$$

时间序列的第 2 个周期方程

$$x_{2t}=4.58839 \cdot \text{EXP}(0.00967 t) \cdot \sin\{2 \cdot 3.14159 \cdot | | t / 4.0 | |^{1.01091} + 3.070223\}$$

序列的自回归滑动平均模型为 ARMA(3,3)

$$x_t = -1.3204 + 0.2116x_{t-1} - 0.32233x_{t-2} + 0.154719x_{t-3} - 0.11105e_{t-1} - 0.73437e_{t-2} - 0.04117e_{t-3} + e_t$$

模型检验参数及预测

残差平方和  $SS=19781.4585$ , 残差标准差=12.4804,  $AIC=37.4794$ , 相关系数  $R=0.9952$ , 拟合度  $C=99.04\%$ , 并给出未来 12 个月的预测值。

将项静恬等(1986)按式(31.1)提取周期项, 建立组合模型过程与按式(31.2)提取周期项相比, 后者较为合理。例如在都提取 1 个指数趋势项的基础上, 如按式(31.2)提取 1 个周期后, 残差平方和等于 47779, 而按式(31.1)提取 1 个周期后, 残差平方和还高达 95783; 按式(31.2)提取 2 个周期后残差平方和下降到 36761, 而按式(31.1)提取 2 个周期后, 残差平方和还有 44753。实

实际上按式(31.2)提取 2 个周期后的残差平方和比按式(31.1)提取 3 个周期后的残差平方和 36897 还小。

## 模糊数学方法

在自然科学或社会科学研究中,存在着许多定义不很严格或者说具有模糊性的概念。这里所谓的模糊性,主要是指客观事物的差异在中间过渡中的不分明性,如某一生态条件对某种害虫、某种作物的存活或适应性可以评价为“有利、比较有利、不那么有利、不利”;灾害性霜冻气候对农业产量的影响程度为“较重、严重、很严重”,等等。这些通常是本来就属于模糊的概念,为处理分析这些“模糊”概念的数据,便产生了模糊集合论。

根据集合论的要求,一个对象对应于一个集合,要么属于,要么不属于,二者必居其一,且仅居其一。这样的集合论本身并无法处理具体的模糊概念。为处理这些模糊概念而进行的种种努力,催生了模糊数学。模糊数学的理论基础是模糊集。模糊集的理论是 1965 年美国自动控制专家查德(L.A.Zadeh)教授首先提出来的,近 10 多年来发展很快。

模糊集合论的提出虽然较晚,但目前在各个领域的应用十分广泛。实践证明,模糊数学在农业中主要用于病虫测报、种植区划、品种选育等方面,在图像识别、天气预报、地质地震、交通运输、医疗诊断、信息控制、人工智能等诸多领域的应用也已初见成效。从该学科的发展趋势来看,它具有极其强大的生命力和渗透力。

在侧重于应用的模糊数学分析中,经常应用到聚类分析、模式识别和综合评判等方法。在 DPS 系统中,我

们将模糊数学的分析方法与一般常规统计方法区别开来,列专章介绍其分析原理及系统设计的有关功能模块程序的操作要领,供用户参考和使用。

## 第 1 节 模糊聚类分析

### 1. 模糊集的概念

对于一个普通的集合  $A$ , 空间中任一元素  $x$ , 要么  $x \in A$ , 要么  $x \notin A$ , 二者必居其一。这一特征可用一个函数表示为:

$$A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

$A(x)$  即为集合  $A$  的特征函数。将特征函数推广到模糊集, 在普通集合中只取 0、1 两值推广到模糊集中为  $[0, 1]$  区间。

定义 1 设  $X$  为全域, 若  $A$  为  $X$  上取值  $[0, 1]$  的一个函数, 则称  $A$  为模糊集。

如给 5 个同学的性格稳重程度打分, 按百分制给分, 再除以 100, 这样给定了一个从域  $X = \{x_1, x_2, x_3, x_4, x_5\}$  到  $[0, 1]$  闭区间的映射。

$x_1$ : 85 分, 即  $A(x_1) = 0.85$

$x_2$ : 75 分,  $A(x_2) = 0.75$

$x_3$ : 98 分,  $A(x_3) = 0.98$

$x_4$ : 30 分,  $A(x_4) = 0.30$

$x_5$ : 60 分,  $A(x_5) = 0.60$

这样确定出一个模糊子集  $A = (0.85, 0.75, 0.98, 0.30, 0.60)$ 。

定义 2 若  $A$  为  $X$  上的任一模糊集, 对任意  $0 \leq \lambda \leq 1$ , 记  $A_\lambda = \{x \mid x \in X, A(x) \geq \lambda\}$ , 称  $A_\lambda$  为  $A$  的  $\lambda$  截集。

$A_\lambda$  是普通集合而不是模糊集。由于模糊集的边界是

模糊的,如果要把模糊概念转化为数学语言,需要选取不同的置信水平  $\lambda(0 \leq \lambda \leq 1)$  来确定其隶属关系。 $\lambda$  截集就是将模糊集转化为普通集的方法。模糊集  $A$  是一个具有游移边界的集合,它随  $\lambda$  值的变小而增大,即当  $\lambda_1 < \lambda_2$  时,有  $A\lambda_1 \supseteq A\lambda_2$ 。

定义 3 模糊集运算定义。若  $A, B$  为  $X$  上两个模糊集,它们的和集、交集和  $A$  的余集都是模糊集,其隶属函数分别定义为:

$$(A \vee B)(x) = \max(A(x), B(x))$$

$$(A \wedge B)(x) = \min(A(x), B(x))$$

$$A^c(x) = 1 - A(x)$$

关于模糊集的和、交等运算,可以推广到任意多个模糊集合中去。

定义 4 若一个矩阵元素取值为  $[0, 1]$  区间内,则称该矩阵为模糊矩阵。同普通矩阵一样,有模糊单位阵,记为  $I$ ;模糊零矩阵,记为  $O$ ;元素皆为 1 的矩阵用  $J$  表示。

定义 5 若  $A$  和  $B$  是  $n \times m$  和  $m \times l$  的模糊矩阵,则它们的乘积  $C=AB$  为  $n \times l$  阵,其元素为:

$$C_{ij} = \bigvee_{k=1}^m (a_{ik} \wedge b_{kj}) \quad (i=1, 2, \dots, n; j=1, 2, \dots, l) \quad (20.1)$$

符号“ $\bigvee$ ”和“ $\bigwedge$ ”含意的定义为: $a \bigvee b = \max(a, b)$ ,  
 $a \bigwedge b = \min(a, b)$ 。

模糊矩阵乘法性质包括:1)  $(AB)C = A(BC)$ ; 2)  $AI = IA = A$ ; 3)  $AO = OA = O$ ; 4)  $AJ = JA$ ; 5) 若  $A, B$  为模糊矩阵且  $a_{ij} \leq b_{ij}$  (一切  $i, j$ ), 则  $A \leq B$ , 又若  $A \leq B$ , 则  $AC \leq BC$ ,  $CA \leq CB$ 。

## 2. 模糊分类关系

模糊聚类分析是在模糊分类关系基础上进行聚类。由集合的概念,可给出如下定义:

定义 6  $n$  个样品的全体所组成的集合  $X$  作为全域,令  $X \times Y = \{ (X, Y) \mid x \in X, y \in Y \}$ , 则称  $X \times Y$  为  $X$  的全域乘积空间。

定义 7 设  $R$  为  $X \times Y$  上的一个集合, 并且满足:

1) 反身性:  $(x_i, y_i) \in R$ , 即集合中每个元素和它自己同属一类;

2) 对称性: 若  $(x, y) \in R$ , 则  $(y, x) \in R$ , 即集合中  $(x, y)$  元素同属于类  $R$  时, 则  $(y, x)$  也同属于  $R$ ;

3) 传递性:  $(x, y) \in R, (y, z) \in R$ , 则有  $(x, z) \in R$ 。

上述三条性质称为等价关系, 满足这三条性质的集合  $R$  为一分类关系。

聚类分析的基本思想是用相似性尺度来衡量事物之间的亲疏程度, 并以此来实现分类, 模糊聚类分析的实质就则是根据研究对象本身的属性未构造模糊矩阵, 在此基础上根据一定的隶属度来确定其分类关系。

### 3. 模糊聚类

利用模糊集理论进行聚类分析的具体步骤如下:

(1) 若定义相似系数矩阵用的是定量观察资料, 在定义相似系数矩阵之前, 可先对原始数据进行变换处理, 变换的方法同系统聚类分析, 可参考第 17 章系统聚类分析一节。

(2) 计算模糊相似矩阵。设  $U$  是需要被分类对象的全体, 建立  $U$  上的相似系数  $R, R(i, j)$  表示  $i$  与  $j$  之间的相似程度, 当  $U$  为有限集时,  $R$  是一个矩阵, 称为相似系数矩阵。定义相似系数矩阵的工作, 原则上可以按系统聚类分析中的相似系数确定方法, 但也可以用主观评

定或集体打分的办法。DPS 平台，对数据集

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix} n \times m$$

提供了以下 8 种建立相似矩阵的方法：

相关系数法：

最大最小法：

算术平均最小法：

几何平均最小法：

绝对指数法：

绝对值减数法：

夹角余弦法：

欧氏距离：

(3) 聚类分析。用上述方法建立起来的相似关系  $R$ ，一般只满足反射性和对称性，不满足传递性，因而还不是模糊等价关系。为此，需要将  $R$  改造成  $R'$  后得到聚类图，在适当的阈值上进行截取，便可得到所需要的分类。将  $R$  改造成  $R'$ ，可用求传递闭包的方法。 $R$  自乘的思想是按最短距离法原则，寻求两个向量  $x_i$  与  $x_j$  的亲密度。

假设  $R^0 = (r_{ij})$ ，即  $r_{ij} = \bigvee_{k=1}^n (r_{ik} \wedge r_{kj})$ ，说明  $x_i$  与  $x_j$  是通过第三者  $K$  作为媒介而发生关系， $r_{ik} \wedge r_{kj}$  表示  $x_i$  与  $x_j$  的关系密切程度是以  $\min(r_{ik}, r_{kj})$  为准则，因  $k$  是任意的，故从一切  $r_{ik} \wedge r_{kj}$  中寻求一个使  $x_i$  和  $x_j$  关系最密切的通道。 $R^n$  随  $m$  的增加，允许连接  $x_i$  与  $x_j$  的链的边就越多。由于从  $x_i$  到  $x_j$  的一切链中，一定存在一个使最大边长达

到极小的链，这个边长就是相当于  $r_{ij}^{\infty}$ 。

在实际处理过程中， $R$  的收敛速度是比较快的。为进一步加快收敛速度，通常采取如下处理方法：

$$R \quad R^2 \quad R^4 \quad R^8 \quad \dots \quad R^{2^k}$$

即先将  $R$  自乘改造为  $R^2$ ，再自乘得  $R^4$ ，如此继续下去，直到某一步出现  $R^{2^k} = R^k = R^*$ 。此时  $R^*$  满足了传递性，于是模糊相似矩阵 ( $R$ ) 就被改造成成了一个模糊等价关系矩阵 ( $R^*$ )。

(4) 模糊聚类。对满足传递性的模糊分类关系的  $R^*$  进行聚类处理，给定不同置信水平的  $\lambda$ ，求  $R_{\lambda}^*$  阵，找出  $R^*$  的  $\lambda$  显示，得到普通的分类关系。当  $\lambda=1$  时，每个样品自成一类，随  $\lambda$  值的降低，由细到粗逐渐归并，最后得到动态聚类谱系图。

#### 4. DPS 平台操作示例

首先在编辑状态下输入编辑数据，格式是每一行为一个样本，每一列为一个变量，然后将待分析的数据定义成数据矩阵块，在菜单方式下选择“模糊数学→模糊聚类”功能项，回车执行时，系统将提示用户选择数据转换方法：

0. 不转换 1. 数据中心化 2. 对数转换 3. 数据规格化  
4. 数据标准化

作出数据转换方式的选择后，系统又将提示选择建立模糊相似关系的计算方法，共有上面所述的 8 种方法可供选择。

分析输出的结果包括各个样本的联结序号、联结水平、聚类谱系图索引及在屏幕上显示聚类谱系图(拷屏可得到谱系图硬拷贝，或按 S 将图形文件以“.BMP”格式存

放在盘上，然后可在 Windows 有关应用软件中调出)。

## 第 2 节 模糊模式识别

### 1. 方法简介

“模式”一词来源于英文 Pattern，原意是典范、式样、样品，在不同场合有其不同的含义。在此我们讲的模式是指具有一定结构的信息集合。

模式识别就是识别给定的事物以及与它相同或类似的事物，也可以理解为模式的分类，即把样品分成若干类，判断给定事物属于哪一类，这与我们前面介绍的判别分析很相似。

模式识别的方法大致可以分为两种，即根据最大隶属原则进行识别的直接法和根据择近原则进行归类的间接法，分别简介如下：

(1)若已知  $n$  个类型在被识别的全体对象  $U$  上的隶属函数，则可按隶属原则进行归类。此处介绍的是针对正态型模糊集的情形。对于正态型模糊变量  $x$ ，其隶属度为

$$A(x) = e^{\left[-\left(\frac{x-a}{b}\right)^2\right]}$$

其中  $a$  为均值， $b^2=2\sigma^2$ ， $\sigma^2$  为相应的方差。按泰勒级数展开，取近似值得

$$A(x) = \begin{cases} 1 - \left(\frac{x-a}{b}\right)^2 & x-a < b \\ 0 & x-a > b \end{cases}$$

若有  $n$  种类型  $m$  个指标的情形，则第  $i$  种类型在第  $j$  种指标上的隶属函数是

$$A_{ij}(x) = \begin{cases} 0 & x \leq a_{ij}^{(1)} - b_{ij} \\ 1 - \left( \frac{x - a_{ij}^{(1)}}{b_{ij}} \right)^2 & a_{ij}^{(1)} - b_{ij} < x < a_{ij}^{(1)} \\ 1 & a_{ij}^{(1)} \leq x \leq a_{ij}^{(2)} \\ 1 - \left( \frac{x - a_{ij}^{(2)}}{b_{ij}} \right)^2 & a_{ij}^{(2)} < x < a_{ij}^{(2)} + b_{ij} \\ 0 & a_{ij}^{(2)} + b_{ij} < x \end{cases}$$

其中  $a_{ij}^{(1)}$  和  $a_{ij}^{(2)}$  分别是第  $i$  类元素第  $j$  种指标的最小值和最大值,  $b_{ij}^2 = 2\sigma_{ij}^2$ , 而  $\sigma_{ij}^2$  是第  $i$  类元素第  $j$  种指标的方差。

(2) 若有  $n$  种类型  $(A_1, A_2, \dots, A_n)$ , 每类都有  $m$  个指标, 且均为正态型模糊变量, 相应的参数分别为

$a_{ij}^{(1)}$ ,  $a_{ij}^{(2)}$ ,  $b_{ij}$  ( $i=1, 2, \dots, n; j=1, 2, \dots, m$ )。其中,

$a_{ij}^{(1)} = \min(x_{ij})$ ,  $a_{ij}^{(2)} = \max(x_{ij})$ ,  $b_{ij}^2 = 2\sigma_{ij}^2$ , 而  $\sigma_{ij}^2$  是  $x_{ij}$  的方差。待

判别对象  $B$  的  $m$  个指标分别具有参数  $a_j, b_j$  ( $j=1, 2, \dots, m$ ), 且为正态型模糊变量, 则  $B$  与各个类型的贴近度为

$$(A_{ij}, B) = \begin{cases} 0 & a_j \leq a_{ij}^{(1)} - (b_j - b_{ij}) \\ 1 - \frac{1}{2} \left( \frac{a_j - a_{ij}^{(1)}}{b_j + b_{ij}} \right)^2 & a_{ij}^{(1)} - (b_j - b_{ij}) < a_j < a_{ij}^{(1)} \\ 1 & a_{ij}^{(1)} \leq a_j \leq a_{ij}^{(2)} \\ 1 - \frac{1}{2} \left( \frac{a_j - a_{ij}^{(2)}}{b_j + b_{ij}} \right)^2 & a_{ij}^{(2)} < a_j < a_{ij}^{(2)} + (b_j + b_{ij}) \\ 0 & a_{ij}^{(2)} + (b_j + b_{ij}) \leq a_j \end{cases}$$

记  $S_i = \min_{1 \leq j \leq m} (A_{ij}, B)$ , 又有  $S_{i0} = \max_{1 \leq j \leq n} (S_i)$ , 按贴近原则可认

为  $B$  与  $A_{j_0}$  最贴近。

根据如上介绍, DPS 系统中设计了两个功能模块: 一是根据在集合上的隶属函数, 按隶属原则识别对象, 判定样本的类别归属; 二是根据模糊集两两之间的贴近度, 按择近原则, 确定出最接近的两个模糊集。

## 2. DPS 平台的操作示例

系统规定数据输入的格式是每一行为一个样本, 每一列为一个变量。最右边的一列为样本的已知类别(如 1, 2, ...)。 (注意每一类中至少要有三个样本)。对于待判别的样本, 其分类类别用 0 表示。所有待分析数据(连同类别一起)需定义成数据块, 然后进入菜单操作, 选择“模糊数学→模糊识别”功能项, 回车执行后即可输出分析结果。输出结果包括各类参数(变量名、最小值、最大值、标准差和参数  $B$ )和各待判样本的归类结果(样本序号、对各类贴近度的最大值、最贴近的类号)。

注意事项: 系统最多可处理 20 个因子, 100 个样本。

例如, 在“有序样本最优分割”一节中, 我们将历年三化螟发生动态根据最优分割结果分成 3 类, 即将三化螟种群消长过程划分为猖獗—缓和—猖獗三个阶段, 这样的划分结果与该县历年水稻种植制度(一季中稻为主→纯双季稻→单双季混栽)的变化是相吻合的。为识别 1988 年之后三化螟发生动态, 我们也可以应用模糊识别方法进行分析。现将待识别数据和原来的历史资料按上页图 30-3 方式整理编辑和定义。

完成数据编辑定义之后, 执行选项功能“模糊识别”, 便可得到如下结果:

第 1 类

---

变量名最小值最大值标准差参数  $B$

$X(1)$ 121.000000500.000000148.979746210.689178

$X(2)$ 1497.0000004600.0000001243.9473461759.20

7208

$X(3)$ 27.00000036.0000003.3115964.683304

$X(4)$ 9.00000019.0000003.5777095.059644

$X(5)$ 8.00000012.3700001.6712082.363445

$X(6)$ 1.0600001.9100000.3411550.482466

---



---

第 2 类

---



---

变量名最小值最大值标准差参数  $B$

$X(1)$ 19.0000002100.000000587.039979830.199901

$X(2)$ 25.0000002700.000000948.2482861341.02558

7

$X(3)$ 22.00000040.0000005.0472137.137837

$X(4)$ 14.00000028.0000004.3234606.114296

$X(5)$ 0.13000010.9600002.7547763.895841

$X(6)$ 0.5500004.1700001.0928981.545592

---



---

第 3 类

---



---

变量名最小值最大值准差参数  $B$

$X(1)$  34.0000002243.000000725.4704081025.97009  
 0  
 $X(2)$  401.0000007452.0000002400.9240933395.419  
 414  
 $X(3)$  31.00000039.0000002.6692703.774917  
 $X(4)$  18.00000029.0000003.7321005.277987  
 $X(5)$  1.90000011.7900003.2501084.596346  
 $X(6)$  0.0000001.0900000.4628170.654523

---

### 各个待判样本的归类结果

---

样本序号对各类贴近度的最大值最贴近的类号

10.425643

20.897723

---

从分析结果可以看出, 1989年和1990年三化螟发生动态仍和前几年相似, 表明农业生态系统是相对稳定的。

### 第3节 模糊相似优先比方法

#### 1. 方法简介

相似优先比是模糊性度量的一种形式, 它是以成对的样本与一个固定的样本作比较, 确定哪一个与固定样本更相似, 从而选择与固定样本相似程度较大者。

假定样本  $x_i$  和  $x_j$  与固定样本  $x_k$  进行比较, 其相似优

先比  $R_{ij}$  必须满足如下要求：

(1) 若  $R_{ij}$  在  $[0.5, 1.0]$  之间，则表示  $x_i$  比  $x_j$  优先。

(2) 若  $R_{ij}$  在  $[0.0, 0.5]$  之间，则表示  $x_j$  比  $x_i$  优先。

(3) 在极值情形下有三种可能：如果  $R_{ij}=1$ ，则表示  $x_i$  比  $x_j$  显然优先；如果  $R_{ij}=0$ ，则表示  $x_j$  比  $x_i$  显然优先；如果  $R_{ij}=0.5$ ，则  $x_i$  和  $x_j$  不分伯仲，优先无法确定。

在模糊优先比分析中，一般采用海明(Harming)距离作为相似优先比中  $R_{ij}$  的测度。如对样本  $x_i$  和样本  $x_j$  与固定样本  $x_k$  之间进行比较，海明距离可定义为

$$r_{ij} = \frac{d_{ki}}{d_{ki} + d_{kj}}$$

$$R_{ji} = 1 - R_{ij}$$

式中  $d_{ki} = |x_k - x_i|$ ， $d_{kj} = |x_k - x_j|$ ，接下来，对给定的一样本集合  $X = \{x_1, x_2, \dots, x_n\}$  和固定样本  $x_k$ ，令任意  $x_i, x_j \in X$  和  $x_k$  作比较，即计算两两样本间的相似优先比，从而得到模糊相关矩阵：

$$R = (r_{ij}) \begin{cases} r_{ij} \in [0, 1] \\ i, j = 1, 2, \dots, n \end{cases}$$

建立模糊相似矩阵之后，由  $\lambda$  水平集选出相似样本，亦即在相似矩阵中，从大到小地选定  $\lambda$  值，以在  $\lambda$  值下降过程中首先到达的除主对角线元素外全行都为 1 的那一行的样本最相似，然后删除矩阵相应的行和列，并降低  $\lambda$  水平值，继续寻找。依此类推，直至截距处理完毕。

一般情形下，若每个样本有  $m$  个因素，则对每一因素都有一个模糊相似矩阵，所以，每一样本的每一因素都将产生一个反映相似程度的序号值，最后将每一样本各个因素的序号值相加，其结果便是该样本与固定样本相似程度的综合反映。

样本的序号值越小,该样本与固定样本就越相似,但严格地说,各个因素对样本的影响程度是不一样的,因此有必要给各个因素赋予一定的权重,这样得到的结果将更符合实际情况。所以当,用户在对有关因素影响的轻重程度有比较大的把握,或在分析中需突出某个因素时,可对各个因素进行加权处理以达到更好的分析效果。

## 2. DPS 平台的操作示例

数据的输入编辑格式是每一行为一个样本,每一列为一个变量,最右边的一列为已知样本的代码(用 1 表示)和待识别样本的代码(用 0 表示),并将数据和待识别样本一起定义成数据块。

在菜单下选择“模糊数学→相似优先比分析”,执行该项功能后系统将输出分析结果。结果包括待识别样本与各样本间的海明距离以及待识别样本与其它样本各个因素的模糊优先比矩阵  $R$ ,最后给出待判样品对各已知样品各变量相似程度和待判样品对各已知样品的优先比值,并按顺序排列。

例如,高素华(1981)对日本柑橘主要产地之一福冈和我国合肥、武汉、长沙、桂林、温州和成都等 7 地柑橘生长的农业气候相似程度进行了分析,选用各地年均温、年降水量、年日照时数、年极端最低气温和 1 月均温作为相似因子。现运用模糊相似优先比方法在 DPS 平台上进行分析。其数据输入、编辑整理和数据块的定义如图 30-4 所示,

地点	年均温	年降水量	年日照时数	年最低气温	1月均温	识别标识
合肥	15.7	970	2309	-20.6	1.9	1
武汉	16.3	1260	2085	-17.3	2.8	1
上海	15.7	1129	2039	-9.4	3.3	1
长沙	17.2	1422	1726	-9.5	4.6	1
桂林	18.8	1874	1709	-4.9	8.0	1
温州	17.9	1698	1846	-4.5	7.5	1
成都	16.3	976	1239	-4.6	5.6	1
日本福冈	16.2	1492	2000	-8.2	6.2	0

图 30-4 模糊优先比分析的数据编辑定义图

在执行运算时，系统会提示用户输入各个因素权重比例(注意各比例之和须等于 1)，这时如直接回车表示不考虑加权处理。本例分析结果如下。

待报样本 1 与各个样本间绝对值

$x_{10}$ .5000522.0000309.000012.40004.3000

$x_{20}$ .1000232.000085.00009.10003.4000

$x_{30}$ .5000363.000039.00001.20002.9000

$x_{41}$ .000070.0000274.00001.30001.6000

$x_{52}$ .6000382.0000291.00003.30001.8000

$x_{61}$ .7000206.0000154.00003.70001.3000

$x_{70}$ .1000516.0000761.00003.60000.6000

$x_1$  的模糊优先比矩阵  $R(1)$

1.0000.1670.5000.6670.8390.7730.167

0.8331.0000.8330.9090.9630.9440.500

0.5000.1671.0000.6670.8390.7730.167  
0.3330.0910.3331.0000.7220.6300.091  
0.1610.0370.1610.2781.0000.3950.037  
0.2270.0560.2270.3700.6051.0000.056  
0.8330.5000.8330.9090.9630.9441.000

$\lambda_2$  的模糊优先比矩阵  $R(2)$

1.0000.3080.4100.1180.4230.2830.497  
0.6921.0000.6100.2320.6220.4700.690  
0.5900.3901.0000.1620.5130.3620.587  
0.8820.7680.8381.0000.8450.7460.881  
0.5770.3780.4870.1551.0000.3500.575  
0.7170.5300.6380.2540.6501.0000.715  
0.5030.3100.4130.1190.4250.2851.000

$\lambda_3$  的模糊优先比矩阵  $R(3)$

1.0000.2160.1120.4700.4850.3330.711  
0.7841.0000.3150.7630.7740.6440.900  
0.8880.6851.0000.8750.8820.7980.951  
0.5300.2370.1251.0000.5150.3600.735  
0.5150.2260.1180.4851.0000.3460.723  
0.6670.3560.2020.6400.6541.0000.832  
0.2890.1000.0490.2650.2770.1681.000

$\lambda_4$  的模糊优先比矩阵  $R(4)$

1.0000.4230.0880.0950.2100.2300.225  
0.5771.0000.1170.1250.2660.2890.283  
0.9120.8831.0000.5200.7330.7550.750

0.9050.8750.4801.0000.7170.7400.735  
 0.7900.7340.2670.2831.0000.5290.522  
 0.7700.7110.2450.2600.4711.0000.493  
 0.7750.7170.2500.2650.4780.5071.000

$x_5$  的模糊优先比矩阵  $R(5)$

1.0000.4420.4030.2710.2950.2320.122  
 0.5581.0000.4600.3200.3460.2770.150  
 0.5970.5401.0000.3560.3830.3100.171  
 0.7290.6800.6441.0000.5290.4480.273  
 0.7050.6540.6170.4711.0000.4190.250  
 0.7680.7230.6900.5520.5811.0000.316  
 0.8780.8500.8290.7270.7500.6841.000

待判样品(1)对各个已知样品各变量相似程度  
 样本  $x_1x_2x_3x_4x_5$  相似程度

12767729.00  
 21326618.00  
 32411513.00  
 43142313.00  
 55553422.00  
 64235216.00  
 71674119.00

待判样品(1)对各已知样品优先比值

313.0000 预测样本与该样本最接近.  
 413.0000  
 616.0000  
 218.0000

719.0000

522.0000

129.0000

从分析结果来看,上海、长沙与福冈柑橘生产的农业气候条件最相似,而合肥的相似程度最小。

#### 第4节 模糊综合评判

##### 1. 方法简介

综合评判就是对受到多个因素制约的事物或对象作出一个总的评价,这是在日常生活和科研工作中经常遇到的问题,如产品质量评定、科技成果鉴定、某种作物种植适应性的评价等,都属于综合评判问题。由于从多方面对事物进行评价难免带有模糊性和主观性,采用模糊数学的方法进行综合评判将使结果尽量客观从而取得更好的实际效果。

模糊综合评判的数学模型可分为一级模型和多级模型,在此仅介绍一级模型。采用一级模型进行综合评判,一般可归纳为以下几个步骤:

(1) 建立评判对象因素集  $U = \{u_1, u_2, \dots, u_n\}$ 。因素就是对象的各种属性或性能,在不同场合,也称为参数指标或质量指标,它们能综合地反映出对象的质量,因而可由这些因素来评价对象。

(2) 建立评判集  $V = \{v_1, v_2, \dots, v_n\}$ 。如工业产品的评价,评判集是等级的集合;农作物种植区域适应性的评价,评判集是适应程度的集合。

(3) 建立单因素评判,即建立一个从  $U$  到  $F(V)$  的模糊映射

$$f: U \rightarrow F(V), \forall u_i \in U$$

$$u_i \rightarrow f(u_i) = \frac{r_{i1}}{v_1} + \frac{r_{i2}}{v_2} + \Lambda + \frac{r_{im}}{v_m}$$

$$0 \leq r_{ij} \leq 1, 1 \leq i \leq n, 1 \leq j \leq m$$

由  $f$  可以诱导出模糊关系，得到模糊矩阵

$$R = \begin{bmatrix} r_{11} & r_{12} & \Lambda & r_{1m} \\ r_{21} & r_{22} & \Lambda & r_{2m} \\ \Lambda & \Lambda & \Lambda & \Lambda \\ r_{n1} & r_{n2} & \Lambda & r_{nm} \end{bmatrix}$$

称  $R$  为单因素评判矩阵，于是  $(U, V, R)$  构成了一个综合评判模型。

(4) 综合评判。由于对  $U$  中各个因素有不同的侧重，需要对每个因素赋予不同的权重，它可表示为  $U$  上的一个模糊子集  $A = (a_1, a_2, \dots, a_n)$ ，且规定  $\sum_{i=1}^n a_i = 1$ 。

在  $R$  与  $A$  求出之后，则综合评判模型为  $B = A \circ R$ 。记  $B = (b_1, b_2, \dots, b_m)$ ，它是  $V$  上的一个模糊子集，其中

$$b_j = \bigvee_{i=1}^n (a_i \wedge r_{ij}) \quad (j=1, 2, \dots, m)。$$

如果评判结果  $\sum_{j=1}^m b_j \neq 1$ ，就对其结果进行归一化处理。

从上述模糊综合评判的 4 个步骤可以看出，建立单因素评判矩阵  $R$  和确定权重分配  $A$  是两项关键性的工作，但同时又没有统一的格式可以遵循，一般可采用统计实验或专家评分的方法求出。在 DPS 平台上，只是根据给出的评判矩阵  $R$  和确定的权重分配  $A$  进行综合评判处理。但是，用户可以在平台上利用公式计算，根据定义公式，计算其隶属函数。

## 2. DPS 平台的操作示例

根据给定的评估数据(评判矩阵  $R$ )和最终的评价结果(权重分配  $A$ )，确定出该类评估近似的权分配系数。

例如对某教师讲课质量进行综合评定。已知因素集合  $U$  为： $U=\{\text{教材熟练, 逻辑性强, 启发性强, 语言生动, 板书整齐}\}$ ，评语集合  $V=\{\text{很好, 较好, 一般, 不好}\}$ 。首先，建立评估集合(评判矩阵)，即评价矩阵数据编辑和定义。在 DPS 平台上，以行代表因素集，列代表评语集。每一行为因素集  $U$  中某一单因素的评价结果，最后一行存放最终评价值(建立权重分配)，本例的权重分配为  $A=(0.3, 0.2, 0.2, 0.2, 0.1)$  并放在数据块的最后一行。最后，评价矩阵数据块可按图 30-5 格式输入及定义：

因素\ 评语	很好	较好	一般	差	
教材熟 练	0.45	0.25	0.2	0.1	
逻辑性 强	0.50	0.40	0.1	0.0	
启发性 强	0.30	0.40	0.2	0.1	
语言生 动	0.40	0.40	0.1	0.1	
板书整 齐	0.30	0.50	0.1	0.1	
最终评 价值	0.30	0.20	0.2	0.2	0.1

图 30-5 模糊综合评判的数据编辑定义图

完成评估数据(评判矩阵  $R$ )和最终的评价结果(权

重分配 A)数据块的编辑及定义之后,在菜单下选择“模糊数学→模糊综合评判分析”功能项,回车执行后即得分析结果。评判结果表明,对该教师的课堂教学认为“很好”的占 35.29%，“较好”的占 29.41%，“一般”的占 23.53%，“不好”的占 11.76%。根据最大隶属原则,结论是“很好”。

## 第 5 节 模糊关系方程求解

### 1. 方法简介

模糊关系方程是模糊数学的一个重要组成部分。如下所示,它其实是模糊综合评判的逆问题。

$$\begin{array}{ccc} \xrightarrow{\text{输入}A} & R & \xrightarrow{\text{输出}B=?} \\ & A.R=B & \end{array} \qquad \begin{array}{ccc} \xrightarrow{\text{输入}A=?} & R & \xrightarrow{\text{输出}B} \\ & A.R=B & \end{array}$$

这类问题具有普遍实际意义。如一些老专家、老中医或具有丰富经验的实际工作人员,在他们的头脑里,经验和技能常归结于对诸因素有一种优越的权数分配方案,这些难以言传的经验和技能,可望利用模糊数学原理,借助计算机技术进行模拟并保存下来。

在 DPS 系统中,求解模糊关系方程采用徐罗曹李方法(汪培庄 1983)。其求解过程首先考虑  $X_0R=B$  类型的模糊关系方程:

$$(x_1, x_2, \dots, x_n) \times \begin{bmatrix} r_{11} & r_{12} & \Lambda & r_{1m} \\ r_{21} & r_{22} & \Lambda & r_{2m} \\ \Lambda & \Lambda & \Lambda & \Lambda \\ r_{n1} & r_{n2} & \Lambda & r_{nm} \end{bmatrix} = (b_1, b_2, \dots, b_m)$$

按照模糊合成运算的最大—最小法则,上式可化为下面一组线性等式,称为模糊线性方程:

$$\begin{cases} (r_{11} \wedge x_1) \vee (r_{21} \wedge x_2) \vee \dots \vee (r_{n1} \wedge x_n) = b_1 \\ (r_{12} \wedge x_1) \vee (r_{22} \wedge x_2) \vee \dots \vee (r_{n2} \wedge x_n) = b_2 \\ \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \\ (r_{1m} \wedge x_1) \vee (r_{2m} \wedge x_2) \vee \dots \vee (r_{nm} \wedge x_n) = b_m \end{cases}$$

因此，求解模糊关系方程实际上就是对上述方程组的求解，其步骤简述如下：

(1) 标准化排列。将向量  $B$  写于  $A$  的上方，按  $\bar{b}_1 \geq \bar{b}_2 \geq \dots \geq \bar{b}_m$  的次序更换向量  $B$  的分量，同时矩阵  $R$  也作相应的更换，由此得到矩阵的标准化排列。

(2) 上铣。按上述规则更换后的矩阵，对每一个  $j(j=1, 2, \dots, m)$ ，用  $\bar{b}_j$  的上铣第  $j$  列，即若  $\bar{r}_{ij} > \bar{b}_j$ ，将  $\bar{r}_{ij}$  变成  $\bar{b}_j$ ；若  $\bar{r}_{ij} \leq \bar{b}_j$ ，将  $\bar{r}_{ij}$  变成空元  $\phi$ 。从而得到新的  $r_{ij}$ ，用  $r'_{ij}$  表示：

$$r'_{ij} = \begin{cases} \bar{b}_j & (\bar{r}_{ij} > \bar{b}_j) \\ \phi & (\bar{r}_{ij} \leq \bar{b}_j) \end{cases}$$

(3) 求下确界。对上铣后所得矩阵的各行求下确界，得向量  $UR=(ur_1, ur_2, \dots, ur_n)$ ，并定义空元的下确界为 1。

(4) 平铣。将上铣出来的矩阵称为平铣矩阵，即对每一个元素  $j$  分别用  $\bar{b}_j$  平铣  $\bar{R}$  的第  $j$  列：若  $\bar{r}_{ij} \geq \bar{b}_j$ ，将  $\bar{r}_{ij}$  变成  $\bar{b}_j$ ；若  $\bar{r}_{ij} < \bar{b}_j$ ，将  $\bar{r}_{ij}$  变成空元  $\phi$ 。这样得到新的  $r_{ij}$ ，以  $r''_{ij}$  表示：

$$r''_{ij} = \begin{cases} \bar{b}_j & (\bar{r}_{ij} \geq \bar{b}_j) \\ \phi & (\bar{r}_{ij} < \bar{b}_j) \end{cases}$$

(5) 划元。对平铣后的矩阵，逐行划去该行中大于上界即第  $i$  行大于  $ur_i$  的元素。

(6) 判别。若原方程有解，则上一步所得矩阵的每一列都有未被划去的非空白元素。

(7) 求解。从经过划元后所得的矩阵中的每一列，选

定一个非空白、且未被划元的元素，对这些当选元素逐行取上确界，并规定空集  $\phi$  的上确界为 0。这样得到的一组解为“拟极小解”。当被选元素不同时，可得到多个“拟极小解”。

最后得到的解之所以称为“拟极小解”，因为它们不一定是极小解，它们相互之间还可能存在着重合或优劣的模糊关系，需进一步进行筛选。实际上，当方程经判别有解时，第三步所求的上界就是方程的最大解。

应用 DPS 平台求解模糊关系方程时，当模糊方程有解时，系统将给出最大解和最小解，如模糊方程无解时则提示没有解。

## 2. DPS 平台的操作示例

在平台上求解模糊关系方程时，其数据编辑格式是：行为因素集，列为评语集，每一行为因素集  $U$  中某一单因素的评价结果，最后一行存放整个因素集  $U$  的综合评价结果，然后将数据定义成数据块。如对某单位的管理工作进行“民意测验”，得综合评价矩阵和最终评价结果，并按系统规定格式编辑整理数据和定义数据块，如图 30-6。

然后，在菜单下选择“模糊数学→模糊关系方程”功能项，回车执行后即得到归一化之后的分析结果，详见于后。

注意：本系统最多可处理 50 个因子的因子集。

	很好	较好	一般	差
政治思想工作(X1)	0.04	0.30	0.40	0.26
精神文明	0.24	0.28	0.28	0.20

建设(X2)				
行政及其 管理(X3)	0.10	0.34	0.40	0.16
职工组织 纪律(X4)	0.02	0.26	0.40	0.32
人才培训 工作(X5)	0.13	0.28	0.42	0.17
学术科研 工作(X6)	0.14	0.27	0.30	0.29
后勤管理 工作(X7)	0.03	0.14	0.42	0.41
整个因素集 U综合评价	0.15	0.29	0.34	0.22

图 30-6 求解模糊关系方程的数据编辑、定义示意图

输出结果

最小解  $x_i$  最大解归一处理后结果

0.2200<-- $x(1)$ -->0.22000.1325

0.1500<-- $x(2)$ -->0.15000.0904

0.2900<-- $x(3)$ -->0.29000.1747

0.0000<-- $x(4)$ -->0.22000.1325

0.3400<-- $x(5)$ -->0.34000.2048

0.0000<-- $x(6)$ -->0.22000.1325

0.0000<-- $x(7)$ -->0.22000.1325

## 第6节 综合评判的逆问题

### 1. 方法简介

前面的模糊综合评判和模糊关系求解是综合评判的正反两个方面。由于权重分配  $A$  的确定并无通用公式，所以它的正确与否往往只能取决于专家的判断或经验，而这些又是很难用数学公式表达出来的。与之相反，权重分配  $B$  可以通过实践的检验建立起来。由  $R$  和  $B$  反过来求  $A$ ，有利于总结专家的经验，使它得到量化。同样道理，由  $A$  和  $R$  而求  $R$  可以帮助我们检验建立起来的数学模型是否合适。从这个角度来看，综合评判的逆问题比其正问题更有意义。

上面介绍的模糊关系方程求解，即求解综合评判逆问题时，其方程可能有解，也可能无解。有解时，解也可能有多个，这需根据实际情况作出恰当选择。若无解，又应当怎么办呢？DPS 系统提供根据贴近原则解模糊关系方程的方法。

首先选择一些备择解的集合，设它们为  $J = \{A_1, A_2, \dots, A_n\}$ ，然后用贴近度原则在备择解集  $\{A_1, A_2, \dots, A_n\}$  中选取一个  $A_i$  作为模糊关系方程  $X \circ R = B$  中

的  $X$ 。将  $A_i$  与  $R$  进行合成：

$$\begin{cases} A_1 \circ R = B_1 \\ A_2 \circ R = B_2 \\ \dots \dots \dots \\ A_n \circ R = B_n \end{cases}$$

求出  $A$  之后，再用贴近度公式

$$(A_i, B) = \frac{1}{2} [A_i, B + (1 - A_i) \cdot B]$$

分别求出  $B$  与  $A_1, A_2, \dots, A_n$  的贴近度，选最大者作为近似的  $X^*$ ，将它作为方程的近似解。

## 2. DPS 平台操作示例

根据给定的评估数据(评判矩阵  $R$ 、最终的评价结果  $B$  和备择权重方案  $A$ ), 给出与预定若干个权重数分配方案的贴近程度, 从而确定出该类评估近似的权分配系数。这是一种近似求解模糊关系方程的方法, 也可用来评价某一类量对于某些标准定量的贴近程度。

(1) 评估集合。评价矩阵的数据编辑和定义格式是行为因素集, 列为评语集。每一行为因素集  $U$  中某一单因素的评价结果, 最后一行存放整个因素集  $U$  的综合评价结果, 再将数据定义成数据块。例如, 对某品牌的自行车进行评价, 已知因素集合  $U = \{\text{外型, 质量, 价格}\}$ , 评语集合  $V = \{\text{很好, 较好, 一般, 较差}\}$ , 其评价矩阵数据可按图 30-7 格式定义。

因素\评语	很好	较好	一般	差
外型	0.2	0.7	0.1	0.0
质量	0.0	0.4	0.5	0.1
价格	0.2	0.3	0.4	0.1
最终评价值	0.0	0.8	0.2	0.0

图 30-7 贴近度解模糊关系方程评语集及评价结果数据编辑定义图

(2) 可能的权分配方案集合( $A$ )的矩阵数据编辑和定义。每一行为一个权的分配方案, 每一列为某一个因素的权重。数据编辑后将数据定义成公式块。如上例根据对顾客的心理分析, 提出了四种可能的权分配方案, 其数据按图 30-8 方式编辑和定义。注意, 此处的数据块

定义时，须在按下 Ctrl 的同时拖动鼠标。

	外型	质量	价格
第一种权分配方案:	0.2	0.50	0.30
第二种权分配方案:	0.5	0.30	0.20
第三种权分配方案:	0.2	0.30	0.50
第四种权分配方案:	0.7	0.25	0.05

图 30-8 贴近度解模糊关系方程权重备择方案数据编辑定义图

完成评估数据(评判矩阵  $R$ )、最终的评价结果  $B$  以及备择权重分配  $A$  的数据的编辑、定义之后，再进入菜单操作，在菜单方式下选择“模糊数学→综合评判逆问题”分析功能项，回车后便可得到归一化之后的分析结果。

### 分析结果

#### . 贴近度分析结果

第 1 组的贴近度=0.6500

第 2 组的贴近度=0.7000

第 3 组的贴近度=0.6000

第 4 组的贴近度=0.8000

#### . 近似的权重方案

第 1 个因素的权数=0.7000

第 2 个因素的权数=0.2500

第 3 个因素的权数=0.0500

从上述分析结果可以看出,在权重备择集中应选  $A_4$  作为近似的  $\lambda^*$ , 即

$$\lambda^*=(0.7, 0.25, 0.05)$$

采用此方法,如果备择权分配方案太少,计算结果会不太理想。因此在实际应用时,为取得较满意的结果,应当尽可能多地建立一些权重分配方案。

## 灰色系统方法

### 第 1 节 关联度分析

客观世界中存在着大大小小的各类系统,都是由许多因素组成的。这些系统及系统因素之间,相互关系非常复杂。特别是表面现象变化的随机性容易混淆人们的直觉,掩盖事物的本质,使人们在认识、分析、预测和决策时得不到充分全面的信息,不容易形成明确的概念。因此,不仅不同系统之间的关系是灰的,同系统中不同因素之间的关系也是灰的。人们一时会分不清哪些因素关系密切,哪些因素关系不密切,也就是说难以找到主要矛盾,抓住主要特征与主要关系。为此,灰色系统理论提出了关联度分析的概念,其目的就是通过一定的方法理清系统中各因素间的主要关系,找出影响最大的因素,把握矛盾的主要方面。如各类产业中哪个项目的收入影响产值最明显,这种影响程度表明有关生产和销售系统之间或系统内部各因素之间的关联性。

对两个系统或两个因素之间关联性大小的量度,称为关联度。它描述系统发展过程中因素间相对变化的情况,也就是变化大小、方向及速度等指标的相对性。如果两者在系统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。可见,灰色关联度分析是对于一个系统发展变化态势的定量描述和比较。只有弄清楚系统或因素间的这种关联关系,才能对系统有比较透彻的认识,分清哪些是主导因素,哪些是潜在因素,哪些是优势而哪些又是劣势。所以,对于一个灰色系统进行分析研究时,首先要解决如何从随机的时间序列中找到关联性,计算关联度,以便为因素判别、优势分析和预测精度检验等提供依据,为系统决策打好基础。因此说,灰色因素间的关联度分析,实质上是灰色系统分析、预测、决策的基础。

灰色系统理论的关联度分析与数理统计学的相关分析是不同的,两者的区别在于第一,它们的理论基础不同。关联度分析基于灰色系统的灰色过程,而相关分析则基于概率论的随机过程;第二,分析方法不同。关联分析是进行因素间时间序列的比较,而相关分析是因素间数组的比较;第三,数据量要求不同。关联分析不要求数据太多,而相关分析则需有足够的数量;第四,研究重点不同。关联度分析主要研究动态过程,而相关分析则以静态研究为主。因此,关联度分析适应性更广,在用于社会经济系统中的应用更有其独到之处。

目前,关联度分析应用十分广泛,几乎渗透到社会和自然科学各个领域,如农业、教育、卫生、政法、环保、军事、地理、地质、石油、水文、气象、生物,等等。尤其在社会经济领域,如预测宏观经济的发展态势、

国民经济各部门投资效益、区域经济优势分析、技术经济的方案评价、产业结构的调整方向以及微观经济的因素分析等方面，都取得了较好的应用效果。

### 1. 原理与方法简介

关联度分析一般包括下列计算和步骤：(1)原始数据变换；(2)计算关联系数；(3)求关联度；(3)排关联序；(4)列关联矩阵。在应用中是否进行所有步骤，可视具体情况而定。

设有  $m$  个时间序列

$$\begin{array}{cccccc}
 t & x_1^{(t)} & x_2^{(t)} & \Lambda & x_n^{(t)} \\
 1 & x_1^{(1)} & x_2^{(1)} & \Lambda & x_n^{(1)} \\
 2 & x_1^{(2)} & x_2^{(2)} & \Lambda & x_n^{(2)} \\
 \Lambda & \Lambda & \Lambda & \Lambda & \Lambda \\
 n & x_1^{(m)} & x_2^{(m)} & \Lambda & x_n^{(m)}
 \end{array}$$

亦即

$$\{X_1^{(0)}(t)\}, \{X_2^{(0)}(t)\}, \Lambda, \{X_m^{(0)}(t)\} \quad (t=1, 2, \dots, N)$$

$N$  为各序列的长度即数据个数，这  $m$  个序列代表  $m$  个因素(变量)。另设定时间序列：

$$\{X_0^{(0)}(t)\} (t=1, 2, \dots, N)$$

该时间序列称为母序列，而上述  $m$  个时间序列称为子序列。关联度是两个序列关联性大小的度量。根据这一观点，可给关联度一个量化模型，其计算方法与步骤具体叙述如下：

#### (1) 原始数据变换

由于系统中各因素的量纲(或单位)不一定相同，如劳动力为人，产值为万元，产量为吨等，且有时数值的数量级相差悬殊，如人均收入为几百元，粮食每公顷产量为几千公斤，费用为几十万元，有些产业产值达百亿

元,有些产业才几万元,等等,这样的数据很难直接进行比较,且它们的几何曲线比例也不同。因此,对原始数据需要消除量纲(或单位),转换为可比较的数据序列。目前,原始数据的变换有以下几种常用方法:

a)均值化变换。先分别求出各个序列的平均值,再用平均值去除对应序列中的各个原始数据,得到新的数据列,即为均值化序列。其特点是量纲为一,其值大于0,并且大部分近于1,数列曲线互相相交。

b)初值化变换。分别用同一序列的第一个数据去除后面的各个原始数据,得到新的倍数数列,即为初值化数列。量纲为一,各值均大于0,且数列有共同的起点。

c)标准化变换。先分别求出各个序列的平均值和标准差,然后将各个原始数据减去平均值后再除以标准差,这样得到的新数据序列即为标准化序列。量纲为一,其均值为0,方差为1。

一般情况下,对于较稳定的社会经济系统数列作动态序列的关联度分析时,多采用初值化变换,因为这样的数列多数是增长的趋势。若对原始数列只作数值间的关联比较,可用均值化变换,譬如进行产业结构变化的关联分析,自然因素周期性变化的关联分析等。

## (2)计算关联系数

经数据变换的母数列记为 $\{X_0(t)\}$ ,子数列记为 $\{X_i(t)\}$ ,则在时刻 $t=k$ 时母序列 $\{X_0(k)\}$ 与子序列 $\{X_i(k)\}$

的关联系数 $L_{0i}(k)$ 可由下式计算
$$L_{0i}(k) = \frac{\Delta_{\min} + \Delta_{\max}}{\Delta_{0i}(k) + \rho\Delta_{\max}}$$
,式中

$\Delta_{0i}(k)$ 表示 $k$ 时刻两比较序列的绝对差,即 $\Delta_{0i}(k) = |x_0(k) - x_i(k)|$  ( $1 \leq i \leq m$ ); $\Delta_{\max}$ 和 $\Delta_{\min}$ 分别表示所有比较序列各个时刻绝对差中的最大值与最小值。因为比

较序列相交,故一般取  $\Delta_{\min} = 0$ ;  $\rho$  称为分辨系数,其意义是削弱最大绝对差数值太大引起的失真,提高关联系数之间的差异显著性,  $\rho \in (0, 1)$ , 一般情况下可取 0.1~0.5。

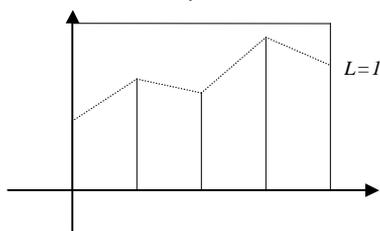
关联系数反映两个被比较序列在某一时刻的紧密(靠近)程度。如在  $\Delta_{\min}$  的时刻,  $L_{io} = 1$ , 而在  $\Delta_{\max}$  的时刻则关联系数为最小值。因此,关联系数的范围为  $0 < L \leq 1$ 。

### (3)求关联度

由以上所述可知,关联度分析实质上是对时间序列数据进行几何关系比较,若两序列在各个时刻点都重合在一起,即关联系数均等于 1,则两序列的关联度也必等于 1。另一方面,两比较序列在任何时刻也不可垂直,所以关联系数均大于 0,故关联度也都大于 0。因此,两序列的关联度便以两比较序列各个时刻的关联系数之平均值计算,即:

$$r_{0i} = \frac{1}{N} \sum_{k=1}^N L_{0i}(k)$$

式中  $r_{0i}$  为子序列  $i$  与母序列 0 的关联度,  $N$  为比较序列的长度(即数据个数)。



用几何坐标表示,即在横坐标为时间  $t$ 、纵坐标为关联系数  $L$  的坐标图中,绘出关联系数曲线(虚线)。该折线与横坐标间围成的面积,称为关联面积,记作  $S_{0i}$ ,而母序列自身的关联系数处处为 1。所以,取纵坐标  $L=1$ ,

图 22.1 关联度几何图示

作水平线与横坐标间围成的面积为重合面积, 记为  $S_{00}$ ,

则关联度的几何意义为两面积之比, 即  $r_{0i} = \frac{S_{0i}}{S_{00}}$ 。因关联

系数曲线为等时距, 且  $S_{00} = 1$ , 故  $r_{0i} = S_{0i} = \frac{1}{N} \sum_{k=1}^N L_{oi}(k)$ 。

不难看出, 关联度与下列因素有关:

- 1) 母序列  $X_0$  不同, 则关联度不同;
- 2) 子序列  $X_i$  不同, 则关联度不同;
- 3) 参考点 0 (或数据变换) 不同, 关联度不同;
- 4) 数据序列长度  $N$  不同, 关联度不同;
- 5) 分辨系数  $\rho$  不同, 关联度不同。

一般来说, 关联度也满足等价“关系”三公理, 即:

- 1) 自反性:  $r_{00} = 1$ ;
- 2) 对称性:  $r_{0i} = r_{i0}$ ;
- 3) 传递性:  $r_{0a} > r_{0b}, r_{0b} > r_{0c}$ , 则  $r_{0a} > r_{0c}$ 。

#### (4) 排关联序

将  $m$  个子序列对同一母序列的关联度按大小顺序排列起来, 便组成关联序, 记为  $\{X\}$ 。它直接反映各个子序列对于母序列的“优劣”关系。若  $r_{0a} > r_{0b}$ , 则称  $\{X_a\}$  对于相同母序列  $\{X_0\}$  有优于  $\{X_b\}$  的特点, 记为  $\{X_a|X_0\} \phi \{X_b|X_0\}$ ; 若  $r_{0a} < r_{0b}$ , 则称  $\{X_a\}$  对于母序列  $\{X_0\}$  劣于  $\{X_b\}$ , 记为  $\{X_a|X_0\} \pi \{X_b|X_0\}$ ; 若  $r_{0a} = r_{0b}$ , 则称  $\{X_a\}$  对于母序列  $\{X_0\}$  等价于 (或等于)  $\{X_b\}$ , 记为  $\{X_a|X_0\} \sim \{X_b|X_0\}$ ; 若有  $r_{0a} \geq r_{0b}$ , 称  $\{X_a\}$  对于母序列  $\{X_0\}$  优于或等于  $\{X_b\}$ , 记为  $\{X_a|X_0\} \phi \underline{\pi} \{X_b|X_0\}$ ; 若有  $r_{0a} \leq r_{0b}$ , 则称  $\{X_a\}$  对于母序列  $\{X_0\}$  劣于或等于  $\{X_b\}$ , 记为  $\{X_a|X_0\} \pi \underline{\phi} \{X_b|X_0\}$ 。

根据上述几种关系, 可定义两种有代表性的关联序,

即“有序”与“偏序”。若关联序 $\{X\}$ 为有序,那么所有元素之间必存在以下几种关系之一:“优于”( $\phi$ ),“劣于”( $\pi$ ),或“等价于”( $\sim$ )。若关联序 $\{X\}$ 为偏序,则不是所有元素都可比较的。

一般而言,各因素只要能构成关系,算出关联度,则总是“有序”的。只有在无“参考点”或无参考母序列的情况下,才可能出现“偏序”现象。

#### (5)列出关联矩阵

若有 $n$ 个母序列 $\{Y_1\}, \{Y_2\}, \dots, \{Y_n\}$  ( $n \geq 2$ )及其 $m$ 个子序列 $\{X_1\}, \{X_2\}, \dots, \{X_m\}$  ( $m \geq 1$ ),则各子序列对母序列 $\{Y_1\}$ 有关联度 $[r_{11}, r_{12}, \dots, r_{1m}]$ ,各子序列对于母序列 $\{Y_2\}$ 有关联度 $[r_{21}, r_{22}, \dots, r_{2m}]$ ,类似地,各子序列对于母序列 $\{Y_n\}$ 有关联度 $[r_{n1}, r_{n2}, \dots, r_{nm}]$ 。

将 $r_{ij}$  ( $i=1, 2, \dots, n; j=1, 2, \dots, m$ )作适当排列,可得到关联度矩阵,根据关联度矩阵,不仅可以作为优势分析的基础,而且可作为决策的依据。若关联矩阵 $R$ 中第 $i$ 列满足

$$\begin{bmatrix} r_{1i} \\ r_{2i} \\ \vdots \\ r_{mi} \end{bmatrix} > \begin{bmatrix} r_{1j} \\ r_{2j} \\ \vdots \\ r_{mj} \end{bmatrix} \quad (\forall i, j \in 1, 2, \dots, n, i \neq j)$$

则称母序列 $\{Y_i\}$ 相对于其它母序列为最优,或者说从 $Y_i$ 对于子序列 $X_j$  ( $j=1, 2, \dots, m$ )的关联度来看,序列 $\{Y_i\}$ 是系统最优序列,并记为:

$$\{Y_i\} \phi \{Y_j\} \quad (j=1, 2, \dots, n; j \neq i)$$

若有

$$\frac{1}{n} \sum_{k=1}^n r_{ki} > \frac{1}{n} \sum_{k=1}^n r_{kj} \quad (i, j=1, 2, \dots, n; i \neq j)$$

则称母序列  $\{Y_i\}$  相对于其余母序列, 或相对于子序列  $\{X_i\} (i=1, 2, \dots, m)$  的关联度是准最优的, 并记为:

$$\{Y_i\} \underset{\phi}{\phi} \{Y_j\} \quad (j \in \{1, 2, \dots, n\}, j \neq i)$$

若关联矩阵  $R$  为下三角矩阵, 即:

$$r_{11}$$

$$r_{21} \quad r_{22}$$

$$r_{31} \quad r_{32} \quad r_{33}$$

$$r_{n1} \quad r_{n2} \quad r_{n3} \dots r_{nm}$$

则称  $[Y_1]$  相对于  $[Y_i] (i \in \{2, 3, \dots, n\})$  是最优势的。

## 2. DPS 平台的操作步骤

在 DPS 平台上, 数据编辑方式一般是前面  $p$  列为子序列, 后面  $q$  列为母序列, 每一行为一个样本(或某一时刻)的数据。若无指定的母序列数据, 对有  $q$  个因子序列、 $n$  个时刻的资料输入顺序为:

因子序列

$$X_{11} X_{12} X_{13} \dots X_{1q}$$

$$\text{时 } X_{21} X_{22} X_{23} \dots X_{2q}$$

刻.....

$$X_{n1} X_{n2} X_{n3} \dots X_{nq}$$

这时系统会自动依次将各组作为母序列, 其余为子序列, 试计算关联矩阵并对因子进行排序。若有  $p$  个子序列、 $q$  个母序列、 $n$  个时刻的时间序列资料, 其数据输入、编辑、定义数据块的顺序为子序列在左, 母序列在右, 方式为:

子序列母序列

$$X_{11} X_{12} X_{13} \dots X_{1p} Y_{11} Y_{12} Y_{13} \dots Y_{1q}$$

时  $X_{21}X_{22}X_{23}\dots X_{2p}Y_{21}Y_{22}Y_{23}\dots Y_{2q}$

刻.....

$X_{n1}X_{n2}X_{n3}\dots X_{np}Y_{n1}Y_{n2}Y_{n3}\dots Y_{nq}$

数据输入并编辑好之后定义数据块，然后进入主菜单，选择“其它方法→灰色系统→关联分析”功能项。按回车执行时，系统还需用户根据分析要求作出一些选择（对每项选择系统都有默认值，即每次按回车也可得到分析结果）：

(1)在系统提示下输入母序列的个数后回车。如果不指定母序列数目，按回车键即可。(2)在系统提示下选择数据转换方式。一般为均值化。

(3)选择数据转换方式之后，系统提示用户输入分辨系数，这时如直接回车则取分辨系数的系统默认值(0.1)。

(4)最后，系统询问是否将分析参数  $\Delta_{\min}$  取值为 0。如直接回车或键入“Y”后再回车则将参数  $\Delta_{\min}$  取值为 0；如键入“N”后再回车，则参数  $\Delta_{\min}$  的取值由各个数据序列各个时刻绝对差值的比较来确定。

最后系统输出分析结果，包括数据转换结果、母序列与其它因子序列的绝对差值、最大差值  $\Delta_{\max}$ 、关联系数及其关联序等等。

### 3.应用示例

(1)无母序列情形。设有下述 4 组时间序列，若依次将各组作为母序列，其余为子序列，试计算关联矩阵并对因子进行排序。在 DPS 平台的编辑状态下，按图 33-1 方式编辑定义待分析数据块。

序号	X1	X2	X3	X4
1	39.1	45.8	3.4	6.7
2	41.6	43.4	3.3	6.8
3	43.9	42.3	3.5	5.4
4	44.9	41.9	3.5	4.7

图 33-3 关联分析的数据编辑、定义示意图

然后在菜单下执行灰色关联分析选项功能，在分析过程中对数据进行均值化变换，分辨系数取 0.1，参数  $\Delta_{\min}$  取值为 0，输出结果见后：

均值化变换结果（略）

X1 与其它因子的绝对差值

X10.00000.00000.00000.0000

X20.13380.01940.06020.0930

X30.07000.01820.01410.0377

X40.21290.17080.12070.2630

最大差值  $\max=0.26298$

关联系数  $G(1,2)=0.31589G(1,3)=0.48153G(1,4)=$

0.12828

关联序  $X_3 > X_2 > X_4$

X2 与其它因子的绝对差值

X10.13380.01940.06020.0930

X20.00000.00000.00000.0000

X30.06380.03760.04610.0553

X40.07910.15140.06050.1699

最大差值  $\max=0.16994$

关联系数

$G(2,1)=0.23842$   $G(2,3)=0.25637$   $G(2,4)=0.14699$

关联序  $\lambda_3 > \lambda_1 > \lambda_4$

$\lambda_3$  与其它因子的绝对差值

$\lambda_1$  0.07000.01820.01410.0377

$\lambda_2$  0.06380.03760.04610.0553

$\lambda_3$  0.00000.00000.00000.0000

$\lambda_4$  0.14290.18900.10660.2253

最大差值  $\max=0.22529$

关联系数

$G(3,1)=0.44647$   $G(3,2)=0.31319$   $G(3,4)=0.12700$

关联序  $\lambda_1 > \lambda_2 > \lambda_4$

$\lambda_4$  与其它因子的绝对差值

$\lambda_1$  0.21290.17080.12070.2630

$\lambda_2$  0.07910.15140.06050.1699

$\lambda_3$  0.14290.18900.10660.2253

$\lambda_4$  0.00000.00000.00000.0000

最大差值  $\max=0.26298$

关联系数

$G(4,1)=0.12828$   $G(4,2)=0.20862$   $G(4,3)=0.14497$

关联序  $\lambda_2 > \lambda_3 > \lambda_1$

-----关联矩阵-----

1.000000.315890.481530.12828

0.238421.000000.256370.14699

0.446470.313191.000000.12700

0.128280.208620.144971.00000

(2)有母序列情形。当指定某一组或多组数据为母序

列时,这时母序列数据必须放在数据矩阵右边。例如,王淑荣(1995)应用灰色关联分析大豆主要数量性状的选择,通过田间调查株高、主茎节数、单株荚数、单株粒数及室内考种测得百粒重等诸项指标。现以这些因素作为子因素,以大豆产量作为母因素进行分析。先将试验观察数据在 DPS 平台上按规定格式编辑和定义,如图 31-2 所示。

品	株高	主茎节数	单株荚数	单株粒数	百粒重	产量
系	X1	X2	X3	X4	X5	Y1
v1	78.8	14.3	28.9	74.8	17.7	1974.0
v2	60.5	13.4	33.9	89.9	16.8	1688.4
v3	74.3	13.7	26.5	70.9	17.7	1786.5
v4	69.3	14.3	34.8	83.6	19.9	1796.6
v5	77.0	12.6	27.9	75.4	20.0	1796.3
v6	76.5	12.4	21.5	55.3	18.1	1826.1
v7	67.5	11.0	24.8	70.0	20.3	1715.5
v8	78.7	12.2	30.6	81.4	18.3	1991.5
v9	76.1	13.3	38.0	94.7	18.4	1847.4
v10	75.5	13.1	23.1	62.8	17.0	1854.6

图 31-2 灰色关联分析的数据编辑与定义(子序列在左,母序列在右)

按图 31-2 方式编辑定义数据之后,在菜单下选择“其它方法→灰色系统→关联分析”功能项。执行灰色关联分析时,按系统提示输入母序列个数 1 之后再回车,数据转换方式选择标准化,分辨系数取 0.5,参数  $\Delta_{\min}$

的取值从各个数据序列各个时刻的绝对差值的比较来确定(键入“N”后再回车)。分析结束时系统输出如下结果:

### 数据标准化

0.91621.2506-0.0187-0.0900-0.57701.5056

-2.20020.36440.91631.1682-1.2981-1.4333

0.14990.6598-0.4675-0.4150-0.5770-0.4239

-0.70161.25061.08460.64331.1860-0.3199

0.6096-0.4234-0.2057-0.04001.2661-0.3230

0.5245-0.6204-1.4025-1.7149-0.2564-0.0164

-1.0081-1.9990-0.7854-0.49001.5065-1.1545

0.8991-0.81730.29920.4600-0.09621.6857

0.45640.26591.68301.5682-0.01600.2028

0.35420.0689-1.1033-1.0899-1.13790.2769

### $\gamma_1$ 与其它因子的绝对差值

$\chi_{10}$ .58940.76680.57370.38170.93270.54090.1464

0.78650.25360.0773

$\chi_{20}$ .25501.79771.08361.57060.10040.60400.8446

2.50300.06310.2080

$\chi_{31}$ .52432.34960.04361.40450.11731.38610.3691

1.38651.48011.3802

$\chi_{41}$ .59562.60160.00890.96320.28301.69850.6645

1.22571.36541.3668

$\chi_{52}$ .08250.13520.15311.50591.58910.24012.6610

1.78180.21891.4148

最小差值  $\min=0.00890$  最大差值  $\max=2.66097$

### 关联系数

$G(1,1)=0.74660$  $G(1,2)=0.67148$  $G(1,3)=0.59674$  $G($

$$1,4)=0.58433G(1,5)=0.61111$$

关联序

$$\lambda_1 > \lambda_2 > \lambda_5 > \lambda_3 > \lambda_4$$

-----关联矩阵-----

$$0.746600.671480.596740.584330.61111$$

从分析结果来看,大豆主要数量性状与产量之间的关联度第一位是株高,其次是主茎节数,再其次分别是百粒重、单株荚数和单株粒数。

第2节灰色动态(GM)建模基本原理(略)

第3节灰色数列GM(1,1)模型

灰色数列预测主要是指利用GM(1,1)模型,对时间序列数据进行数量大小的预测。例如人口预测、劳力预测、产品产量预测、各业产值预测及病虫发生趋势预测等,就是利用历年统计资料,对其未来的发展进行预测。这类预测不仅应用最广,而且方法步骤也有普遍意义。因此,下面作比较详细的介绍。

1.建立GM(1,1)模型的基本原理(略)

2.DPS平台的操作示例

在DPS平台上,先将系列原始数据资料按时间顺序从左到右、从上到下地在编辑器中输入,每行的数据个数不一定相同。编辑好数据后,将待分析的数据定义为数据矩阵块。

然后,在菜单下选择GM(1,1)建模分析功能项,执行运算时,系统会先提示输入待预测时刻(样本)的个数,再选择待识别的残差类型(一般选缺省值1)。在每次分析完毕时,屏幕上显示当前的残差值,并请用户决定是否继续进行残差序列的建模分析。

最后系统输出模型参数、观察值、拟合值、误差并

对模型作出评价,同时还给出未来的预测值。

例如,汪笃栋等(1994)对彭泽县 1960~1986 年三代棉红铃虫百株卵量,运用灰色系统理论和方法,建立了 GM(1,1)模型,进行长期灾变预测。根据该县数据,在 DPS 平台上作同样分析。

根据原作者研究,将全代卵量大于 3000 粒以上的年份作为大发生年。据此,我们以全代卵量 3000 粒为阈值,对离散函数

$$\begin{aligned} Y^{(0)} &= \{Y^{(0)}(1), Y^{(0)}(2), Y^{(0)}(3), \dots, Y^{(0)}(27)\} \\ &= \{584, 775, 2888, 1343, 4435, 2652, 1530, 3117, 954, \\ &410, 428, 2920, 464, 5870, 270, 12385, 4284, 2920, 1799, 1 \\ &527, 6148, 2242, 7295, 2378, 2533, 453, 1384\} \end{aligned}$$

作灾变映射  $\xi: \{Y^{(0)}\} = \{Y_{\xi}^{(0)}\}$ , 按  $\xi \geq 3000$  得到灾变集

$$\{Y_{\xi}^{(0)}(1), Y_{\xi}^{(0)}(2), \dots, Y_{\xi}^{(0)}(27)\} = \{12385, 4435, 3117, 4284, 7295, 5870, 6148\}。 其中$$

$\forall Y_{\xi}^{(0)}(i) \geq \xi$  且  $Y_{\xi}^{(0)}(i) \in Y^{(0)}$ ,  $i \in \{1, 2, 3, 4, 5, 6, 7\}$ , 经对  $Y^{(0)}$  编序后再作映射  $X$ , 得灾变日期集:

$$X^{(0)} = \{X^{(0)}(1), X^{(0)}(2), X^{(0)}(3), \dots, X^{(0)}(7)\} = \{6, 11, 12, 15, 17, 23, 25\}。$$

这即为我们建立灰色模型 GM(1,1)时所需的原始数据(图 22-4)。

将数据编辑定义成数据块(图 31-3)之后,再执行 GM(1,1)建模分析功能。按系统提示输入待预测的样本数(3),然后选择待识别的残差类型为累加序列与其理论值的误差,作两次残差序列的建模分析时得到如下结果:

模型参数:  $a=-0.178152$   $b=8.481744$

$x(t+1)=53.609524e^{0.178152t}-47.609524$

原始数据及其拟合观察值拟合值误差误差%

$X(2)=11.0000$   $10.4542$   $0.5457$   $54.96$

$X(3)=12.0000$   $12.4929$   $-0.4929$   $-4.11$

$X(4)=15.0000$   $14.9291$   $0.0708$   $90.47$

$X(5)=17.0000$   $17.8404$   $-0.8403$   $-4.94$

$X(6)=23.0000$   $21.3194$   $1.6806$   $07.31$

$X(7)=25.0000$   $25.4768$   $-0.4768$   $-1.91$

对当前模型的评价

$C=0.1353$  很好

$\rho=1.0000$  很好

未来 3 个时刻预测值

$X(t+1)=30.44500$

$X(t+2)=36.38200$

$X(t+3)=43.47675$

\*\*\*继续对残差序列进行建模分析\*\*\*

第 1 次残差序列分析结果

模型参数

$a=-0.214610$   $b=0.219015$

$x(t+1)=2.282920e^{0.214610t}-1.020524$

原始数据及其拟合观察值拟合值误差误差%

$$X(2)=11.000010.45420.545754.96$$

$$X(3)=12.000011.77700.223011.86$$

$$X(4)=15.000014.83690.163091.09$$

$$X(5)=17.000017.8394-0.83943-4.94$$

$$X(6)=23.000022.35980.640232.78$$

$$X(7)=25.000025.0856-0.08564-0.34$$

对当前模型的评价

$$C=0.0783 \text{ 很好}$$

$$p=1.0000 \text{ 很好}$$

未来 3 个时刻预测值

$$X(t+1)=31.32643$$

$$X(t+2)=37.64597$$

$$X(t+3)=45.21485$$

\*\*\*继续对残差序列进行建模分析\*\*\*

第 2 次残差序列分析结果

模型参数

$$a=0.057166b=0.606903$$

$$x(t+1)=-9.940109e-0.057166t+10.616449$$

原始数据及其拟合观察值拟合值误差误差%

$$X(2)=11.000010.45420.545754.96$$

$$X(3)=12.000011.77700.223011.86$$

$$X(4)=15.000014.71290.287131.91$$

$$X(5)=17.000017.5216-0.52161-3.07$$

$$X(6)=23.000022.85240.147600.64$$

$$X(7)=25.000024.91070.089330.36$$

对当前模型的评价

$C=0.0523$  很好

$\rho=1.0000$  很好

未来 3 个时刻预测值

$$X(t+1)=31.31251$$

$$X(t+2)=37.60764$$

$$X(t+3)=45.15345$$

根据计算结果,第一个原始数据序列模型与原作者所建立的模型完全相同,但是在对残差数据序列进行建模时,由于采用的残差类型不同,所得到的结果也不同。从分析可以看出,采用该系统,经过一次残差拟合,就得到了较原作者经过两次残差修正后的模型还要高的精度。

从对历史资料回测验证和 1987~1991 年预测结果验证来看,实况和预测值的吻合程度很高。

#### 第 4 节灰色数列 GM(2,1) 模型

##### 1. 建立 GM(2,1) 模型的基本原理(略)

##### 2. DPS 平台的操作示例

原始数据资料的编辑定义方式与前述 GM(1,1) 模型的数据编辑定义方式完全相同,即按时间顺序从左到右、从上到下地在编辑器中输入数据(数据放在一行或一列都可以),每行的数据个数不一定相同。将待分析的数据定义成数据块。

邓聚龙(1985)应用 GM(2,1) 模型分析某地某作物历年种子用量情况,我们引用其数据在 DPS 平台上进行分

析。将原始数据按图 31 - 4 的方式编辑定义，然后在菜单下选择执行 GM(2,1)建模分析功能项，可立即得到同样的计算结果。输出结果有模型参数、观察值、拟合值、误差等，并对模型作出评价，最后给出预测结果。

#### 模型参数

$$a_1=25.3658 \quad a_2=-0.9621 \quad U=75.41002$$

$$r_1=0.03787 \quad r_2=-25.40371$$

$$C_1=81.35986 \quad C_2=-0.00435$$

$$x_1(t+1)=81.35986e^{0.03787t}-0.00435e^{-25.40371t}-78.38051$$

$$x_0(2)=3.24574 \dots 3.27800e=0.03226q=0.9841$$

$$x_0(3)=3.26161 \dots 3.33700e=0.07539q=2.2593\%$$

$$x_0(4)=3.38750 \dots 3.39000e=0.00250q=0.0737\%$$

$$x_0(5)=3.51826 \dots 3.67900e=0.16074q=4.3692\%$$

$$C=0.2381 \text{ 很好}$$

$$p=1.0000 \text{ 很好}$$

$$\text{预测: } G_0=16.2871067$$

$$\hat{x}_1(6)=19.94116 \quad \hat{x}_0(6)=3.65406$$

$$\hat{x}_1(7)=23.73626 \quad \hat{x}_0(7)=3.79510$$

$$\hat{x}_1(8)=27.67785 \quad \hat{x}_0(8)=3.94158$$

#### 第 5 节灰色数列 GM(1, N) 模型

##### 1. 建立 GM(1, N) 模型的基本原理 (略)

##### 2. DPS 平台的操作示例

在 DPS 平台上，数据编辑方式是第一列为待分析的对象，后面各列存放有关影响因素，每一行为一个样本(或某一时刻)的数据。对有  $N$  个因子序列、 $M$  个时刻的资料，输入顺序为：

←因子→

$X_{11} X_{12} X_{13} \dots X_{1N}$

时  $X_{21} X_{22} X_{23} \dots X_{2N}$

刻 .....

$X_{M1} X_{M2} X_{M3} \dots X_{MN}$

数据编辑之后将所有待分析数据定义成数据块，在菜单下选择“其它方法→灰色系统→GM(1, M)模型分析”功能项，按回车执行后系统给出分析结果。

	山西太谷 1978-84 年农业 产值		
年份	农业总产值	农业 产值	农村 工业 产值
1978	<b>5344</b>	<b>4500</b>	<b>734</b>
1979	<b>6724</b>	<b>5279</b>	<b>1345</b>
1980	<b>5968</b>	<b>4442</b>	<b>1526</b>
1981	<b>6373</b>	<b>4697</b>	<b>1676</b>
1982	<b>7309</b>	<b>5485</b>	<b>1824</b>
1983	<b>8509</b>	<b>6019</b>	<b>2490</b>
1984	<b>11542</b>	<b>7875</b>	<b>3668</b>

图 33 - 5GM(1, M)建模分析数据编辑、定义示意图

如对山西省太谷县农业产值构成动态进行分析。现抽取 1978~1984 年农业产值资料，并按图 31 - 5 方式进行编辑、定义，然后在菜单方式下选择执行 GM(1, M)建模分析功能项，最后系统将给出矩阵  $B$  的各个参数、GM(1, M)模型参数、观察值、拟合值、误差以及各个因素对作用对象的动态环节传递函数，等等。

Matrix  $A_{ij} =$   
 -8706.000 -15052.000 -21222.500 -28063.500 -3597  
 2.500 -45998.000  
 9779.000 14221.000 18918.000 24403.000 30422.000  
 38297.000  
 2079.000 3605.000 5281.000 7105.000 9595.000 1326  
 3.000

=====系数向量=====

$$^a = 2.08515 \quad ^b = 2.13172 \quad ^c = 1.94612$$

系统动态环节及其传递函数

$$X_1 1.0223$$

=

$$X_2 1 + 0.4796S$$

$$X_1 0.9333$$

=

$$X_3 1 + 0.4796S$$

$$^x_1(t+1) = (5344.00000 - 1.02233x_2 - 0.93332x_3) \exp(-2.08515t) + 1.02233x_2 + 0.93332x_3$$

$$-----^x_1(j)-----$$

$$^x_1(1) = 5344.00000 \dots 5344.00000$$

$$^x_1(2) = 11118.23793 \dots 12068.00000$$

$$^x_1(3) = 17709.20981 \dots 18036.00000$$

$$^x_1(4) = 24233.02974 \dots 24409.00000$$

$$^x_1(5) = 31572.98063 \dots 31718.00000$$

$$\begin{aligned} \hat{x}_1(6) &= 40055.60594 \dots 40227.00000 \\ \hat{x}_1(7) &= 51530.76365 \dots 51769.00000 \\ \dots \hat{x}_0 \dots e \dots \% \dots \\ \hat{x}_0(2) &= 5774.23796724.0000 \quad e=949.7621 \quad q=14.12\% \\ \hat{x}_0(3) &= 6590.97195968.0000 \quad e=-622.9719 \quad q=-10.44\% \\ \hat{x}_0(4) &= 6523.81996373.0000 \quad e=-150.8199 \quad q=-2.37\% \\ \hat{x}_0(5) &= 7339.95097309.0000 \quad e=-30.9509 \quad q=-0.42\% \\ \hat{x}_0(6) &= 8482.62538509.0000 \quad e=26.3747 \quad q=0.31\% \\ \hat{x}_0(7) &= 11475.157711542.0000 \quad e=66.8423 \quad q=0.58\% \end{aligned}$$

从分析结果可以得到两个因素对系统影响的动态环节传递函数：

$$\frac{x_1^{(1)}}{x_2^{(1)}} = \frac{1.0223}{1+0.4796s}, \quad \frac{x_1^{(1)}}{x_3^{(1)}} = \frac{0.9333}{1+0.4796s}$$

由这两个函数看出，系统在 1.5~2 年内呈摆动状态，但 1.5~2 年后会形成稳态经济状态而逐步发展。

## BP 神经网络模型

### 第 1 节基本原理简介

近年来全球性的神经网络研究热潮的再度兴起，不仅仅是因为神经科学本身取得了巨大的进展，更主要的原因在于发展新型计算机和人工智能新途径的迫切需要。迄今为止在需要人工智能解决的许多问题中，人脑远比计算机聪明的多，要开创具有智能的新一代计算机，就必须了解人脑，研究人脑神经网络系统信息处理的机制。另一方面，基于神经科学研究成果基础上发展出来

的人工神经网络模型,反映了人脑功能的若干基本特性,开拓了神经网络用于计算机的新途径.它对传统的计算机结构和人工智能是一个有力的挑战,引起了各方面专家的极大关注.

目前,已发展了几十种神经网络,例如 Hopfield 模型, Feldmann 等的连接型网络模型, Hinton 等的玻尔茨曼机模型,以及 Rumelhart 等的多层感知机模型和 Kohonen 的自组织网络模型等等。在这众多神经网络模型中,应用最广泛的是多层感知机神经网络。多层感知机神经网络的研究始于 50 年代,但一直进展不大。直到 1985 年, Rumelhart 等人提出了误差反向传递学习算法(即 BP 算),实现了 Minsky 的多层网络设想,如图 34-1 所示。

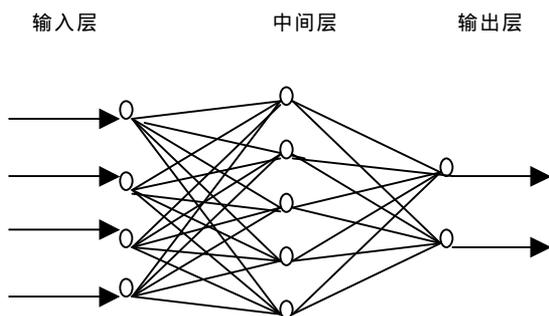


图 34-1 BP 神经网络模型

BP 算法不仅有输入层节点、输出层节点,还可能有 1 个或多个隐含层节点。对于输入信号,要先向前传播到隐含层节点,经作用函数后,再把隐节点的输出信号传播到输出节点,最后给出输出结果。节点的作用的激励函数通常选取 S 型函数,如

$$f(x) = \frac{1}{1 + e^{-x/Q}}$$

式中  $Q$  为调整激励函数形式的 Sigmoid 参数。该算法的学习过程由正向传播和反向传播组成。在正向传播过程中, 输入信息从输入层经隐含层逐层处理, 并传向输出层。每一层神经元的状态只影响下一层神经元的状态。如果输出层得不到期望的输出, 则转入反向传播, 将误差信号沿原来的连接通道返回, 通过修改各层神经元的权值, 使得误差信号最小。

社含有  $n$  个节点的任意网络, 各节点之特性为 Sigmoid 型。为简便起见, 指定网络只有一个输出  $y$ , 任一节点  $i$  的输出为  $O_i$ , 并设有  $N$  个样本  $(x_k, y_k) (k=1, 2, 3, \dots, N)$ , 对某一输入  $x_k$ , 网络输出为  $y_k$  节点  $i$  的输出为  $O_{ik}$ , 节点  $j$  的输入为

$$\text{net}_{jk} = \sum_i W_{ij} O_{ik}$$

并将误差函数定义为

$$E = \frac{1}{2} \sum_{k=1}^N (y_k - \hat{y}_k)^2$$

其中  $\hat{y}_k$  为网络实际输出, 定义

$$E_k = (y_k - \hat{y}_k)^2, \quad \delta_{jk} = \frac{\partial E_k}{\partial \text{net}_{jk}}, \text{ 且 } O_{jk} = f(\text{net}_{jk}), \text{ 于是}$$

$$\frac{\partial E_k}{\partial W_{ij}} = \frac{\partial E_k}{\partial \text{net}_{jk}} \frac{\partial \text{net}_{jk}}{\partial W_{ij}} = \frac{\partial E_k}{\partial \text{net}_{jk}} O_{ik} = \delta_{jk} O_{ik}$$

当  $j$  为输出节点时,  $O_{jk} = \hat{y}_k$

$$\delta_{jk} = \frac{\partial E_k}{\partial \hat{y}_k} \frac{\partial \hat{y}_k}{\partial \text{net}_{jk}} = -(y_k - \hat{y}_k) f'(\text{net}_{jk}) \quad (34.1)$$

若  $j$  不是输出节点, 则有

$$\begin{aligned}\delta_{jk} &= \frac{\partial E_k}{\partial \text{net}_{jk}} = \frac{\partial E_k}{\partial O_{jk}} \frac{\partial O_{jk}}{\partial \text{net}_{jk}} = \frac{\partial E_k}{\partial O_{jk}} f'(\text{net}_{jk}) \\ \frac{\partial E_k}{\partial O_{jk}} &= \sum_m \frac{\partial E_k}{\partial \text{net}_{mk}} \frac{\partial \text{net}_{mk}}{\partial O_{jk}} \\ &= \sum_m \frac{\partial E_k}{\partial \text{net}_{mk}} \frac{\partial}{\partial O_{jk}} \sum_i W_{mi} O_{ik} \\ &= \sum_m \frac{\partial E_k}{\partial \text{net}_{mk}} \sum_i W_{mj} = \sum_m \delta_{mk} W_{mj}\end{aligned}$$

因此

$$\begin{cases} \delta_{jk} = f'(\text{net}_{jk}) \sum_m \delta_{mk} W_{mj} \\ \frac{\partial E_k}{\partial W_{ij}} = \delta_{mk} O_{ik} \end{cases} \quad (34.2)$$

如果有  $M$  层，而第  $M$  层仅含输出节点，第一层为输入节点，则 BP 算法为：

第一步，选取初始权值  $W_0$ 。

第二步，重复下述过程直至收敛：

a. 对于  $k=1$  到  $N$

a). 计算  $O_{ik}$ ,  $\text{net}_{jk}$  和  $\hat{y}_k$  的值(正向过程)；

b). 对各层从  $M$  到 2 反向计算(反向过程)；

b. 对同一节点  $j$  的  $W_{ij}$ , 由式(34.1)和(34.2)计算  $\delta_{jk}$ ;

第三步, 修正权值,  $W_{ij} = W_{ij} - \mu \frac{\partial E}{\partial W_{ij}}$ ,  $\mu > 0$ , 其中

$$\frac{\partial E}{\partial W_{ij}} = \sum_k^N \frac{\partial E_k}{\partial W_{ij}}。$$

从上述 BP 算法可以看出, BP 模型把一组样本的  $1/0$  问题变为一个非线性优化问题, 它使用的是优化中最普通的梯度下降法。如果把神经网络看成输入到输出的映射, 则这个映射是一个高度非线性映射。

设计一个神经网络专家系统重点在于模型的构成和学习算法的选择。一般来说, 结构是根据所研究领域及要解决的问题确定的。通过对所研究问题的大量历史资料数据的分析及目前的神经网络理论发展水平, 建立合适的模型, 并针对所选的模型采用相应的学习算法, 在网络学习过程中, 不断地调整网络参数, 直到输出结果满足要求。

## 第 2 节 DPS 数据处理系统操作步骤

在 DPS 数据处理系统中, 数据的输入格式是一行为一个样本, 一列为一个变量, 输入节点(变量)放在数据块左边, 输出节点(因变量)放在数据块右边, 输完一个样本后再输下一个样本。对于待识别(预测)的样本, 不需要输入输出变量(因变量)。

数据输入完毕后, 定义数据块。如有待识别(预测)的样本, 可在按下 Ctrl 键时再按下并拖动鼠标, 将待预测的样本定义成第二个数据块。

在进行神经网络学习之前, 系统出现如图 34-2 所示界面, 这时需要你提供若干参数, 各个参数取值的基本

原则是：

隐含网络层数	1	数据转换方式 <input type="radio"/> 不变换 <input type="radio"/> 平方根变换 <input type="radio"/> 自然对数变换 <input checked="" type="radio"/> 标准化变换
输入层节点数	7	
最小训练速率	0.1	
动态参数	0.6	
Sigmoid 参数	0.9	
允许误差	0.00001	
最大训练次数	1000	

图 34-2 神经网络参数设置对话框

网络参数确定原则：

、网络节点网络输入层神经元节点数就是系统的特征因子(自变量)个数，输出层神经元节点数就是系统目标个数。隐层节点选按经验选取，一般设为输入层节点数的 75%。如果输入层有 7 个节点，输出层 1 个节点，那么隐含层可暂设为 5 个节点，即构成一个 7-5-1BP 神经网络模型。在系统训练时，实际还要对不同的隐层节点数 4、5、6 个分别进行比较，最后确定出最合理的网络结构。

、初始权值的确定初始权值是不应完全相等的一组值。已经证明，即便确定存在一组互不相等的使系统误差更小的权值，如果所设  $W_{ji}$  的的初始值彼此相等，它们将在学习过程中始终保持相等。故而，在程序中，我们设计了一个随机发生器程序，产生一组 - 0.5~+0.5 的随机数，作为网络的初始权值。

、最小训练速率在经典的 BP 算法中，训练速率是由经验确定，训练速率越大，权重变化越大，收敛越快；但训练速率过大，会引起系统的振荡，因此，训练速率在不导致振荡前提下，越大越好。因此，在 DPS 中，训练速率会自动调整，并尽可能取大一些的值，但用户可

规定一个最小训练速率。该值一般取 0.9。

、动态参数动态系数的选择也是经验性的，一般取 0.6~0.8。

、允许误差一般取 0.001~0.00001，当 2 次迭代结果的误差小于该值时，系统结束迭代计算，给出结果。

、迭代次数一般取 1000 次。由于神经网络计算并不能保证在各种参数配置下迭代结果收敛，当迭代结果不收敛时，允许最大的迭代次数。

、Sigmoid 参数该参数调整神经元激励函数形式，一般取 0.9~1.0 之间。

、数据转换。在 DPS 系统中，允许对输入层各个节点的数据进行转换，提供转换的方法有取对数、平方根转换和数据标准化转换。

### 第 3 节应用实例

原始数据整理：本例令影响棉铃虫发生程度的因素指标集序列由麦田 1 代幼虫量、6 月降水天数、5 月积温、6 月积温、5 月相对湿度、5 月降水天数和 6 月相对湿度等 7 个生态和生物因子构成，2 代发生程度按照全国植保站颁发的标准分级，并规定发生程度重、偏重、中、偏轻和轻分别赋值为 0.9、0.7、0.5、0.3 和 0.1。在建立 BP 神经网络模型时，取 1982~1991 年的数据作为学习、训练样本，1992 和 1993 年为试报样本。在数据分析前将数据定义成数据块(图 34-3)。

	A	B	C	D	E	F	G	H	I	J
1	年份	麦田1代 幼量	6月降水 天数	5月积温	6月积温	5月相对 湿度	5月降水 天数	6月相对 湿度	发生程 度	
2	1982	71.3	6	650	737	48	3	59	0.7	
3	1983	79.4	7	648	784	68	7	52	0.9	
4	1984	58	6	619	760	63	4	63	0.5	
5	1985	43.3	10	604	746	67	12	63	0.3	
6	1986	36.7	7	649	777	56	4	57	0.9	
7	1987	18.9	9	631	739	59	9	63	0.1	
8	1988	111	6	596	765	67	6	62	0.9	
9	1989	90	9	643	753	54	6	57	0.5	
10	1990	105	8	607	743	62	8	64	0.7	
11	1991	92.3	10	610	738	58	9	63	0.9	
12	1992	2608	8	615	770	63	9	62	0.9	
13	1993	1356	7	625	779	65	8	57	0.9	
14										

图 34-3BP 神经网络数据编辑定义示意图

进入 BP 神经网络训练时,系统会显示如图 34-3 所示界面。这时我们可按网络的结构确定网络的参数,这里输入层节点数为 7,隐含层 1 层,最小训练速率取 0.1,动态参数 0.7, Sigmoid 参数为 0.9,允许误差 0.00001,最大迭代次数 1000。并对输入节点的数值进行标准化转换。

点击“确定”按钮后,设置隐层的神经元个数(这里取 5),运行 1000 次后,样本误差等于 0.0001427。输出各个神经元(节点)的权值如下:

第 1 隐含层各个结点的权重矩阵

1.6327102.4498203.0897100.2127105.392370

1.6274202.6001101.9875505.2404103.146180

1.7438302.0563005.2384800.5505900.380470

1.6308303.1634002.5134803.6583000.965040

1.6292901.0916000.6772300.7448802.091490

1.5456001.6525400.8966701.1618102.270320

1.6110404.5644601.9452300.7099802.607020

输出层各个结点的权重矩阵

-1.488610

-4.286470

-5.201220

5.492000

4.719190

学习样本的拟合值和实际观察值,以及根据 BP 神经网络对 1992、1993 年 2 代棉铃虫发生程度进行预测的结果与实际值的比较列于表 34-1。结果表明,应用 BP 神经网络进行二代棉铃虫发生程度预测,不仅历史资料的拟合率高,而且 2 年的试报结果与实际完全符合。

## 其它统计方法

在自然科学或社会科学研究中,存在着许多定义不很严格或者说具有模糊性的概念。这里所谓的模糊性,主要是指客观事物的差异在中间过渡中的不分明性,如某一生态条件对某种害虫、某种作物的存活或适应性可以评价为“有利、比较有利、不那么有利、不利”;灾害性霜冻气候对农业产量的影响程度为“较重、严重、很严重”,等等。这些通常是本来就属于模糊的概念,为处理分析这些“模糊”概念的数据,便产生了模糊集合论。

根据集合论的要求,一个对象对应于一个集合,要么属于,要么不属于,二者必居其一,且仅居其一。这样的集合论本身并无法处理具体的模糊概念。为处理这些模糊概念而进行的种种努力,催生了模糊数学。模糊数学的理论基础是模糊集。模糊集的理论是 1965 年美国自动控制专家查德(L.A.Zadeh)教授首先提出来的,近 10 多年来发展很快。

模糊集合论的提出虽然较晚,但目前在各个领域的应用十分广泛。实践证明,模糊数学在农业中主要用于病虫害测报、种植区划、品种选育等方面,在图像识别、天气预报、地质地震、交通运输、医疗诊断、信息控制、人工智能等诸多领域的应用也已初见成效。从该学科的发展趋势来看,它具有极其强大的生命力和渗透力。

在侧重于应用的模糊数学分析中,经常应用到聚类分析、模式识别和综合评判等方法。在 DPS 系统中,我们将模糊数学的分析方法与一般常规统计方法区别开来,列专章介绍其分析原理及系统设计的有关功能模块程序的操作要领,供用户参考和使用。

## 模糊聚类分析

### 1. 模糊集的概念

对于一个普通的集合  $A$ , 空间中任一元素  $x$ , 要么  $x \in A$ , 要么  $x \notin A$ , 二者必居其一。这一特征可用一个函数表示为:

$$A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

$A(x)$ 即为集合  $A$  的特征函数。将特征函数推广到模糊集, 在普通集合中只取 0、1 两值推广到模糊集中为  $[0,1]$  区间。

定义 1 设  $X$  为全域, 若  $A$  为  $X$  上取值  $[0,1]$  的一个函数, 则称  $A$  为模糊集。

如给 5 个同学的性格稳重程度打分, 按百分制给分, 再除以 100, 这样给定了一个从域  $X = \{x_1, x_2, x_3, x_4, x_5\}$  到  $[0,1]$  闭区间的映射。

$x_1$  : 85 分 , 即  $A(x_1)=0.85$

$x_2$  : 75 分 ,  $A(x_2)=0.75$

$x_3$  : 98 分 ,  $A(x_3)=0.98$

$x_4$  : 30 分 ,  $A(x_4)=0.30$

$x_5$  : 60 分 ,  $A(x_5)=0.60$

这样确定出一个模糊子集  $A=(0.85,0.75,0.98,0.30,0.60)$ 。

定义 2 若  $A$  为  $X$  上的任一模糊集 , 对任意  $0 \leq \lambda \leq 1$  , 记  $A_\lambda = \{x \mid x \in X, A(x) \geq \lambda\}$  , 称  $A_\lambda$  为  $A$  的  $\lambda$  截集。

$A_\lambda$  是普通集合而不是模糊集。由于模糊集的边界是模糊的, 如果要把模糊概念转化为数学语言, 需要选取不同的置信水平  $\lambda(0 \leq \lambda \leq 1)$  来确定其隶属关系。 $\lambda$  截集就是将模糊集转化为普通集的方法。模糊集  $A$  是一个具有游移边界的集合, 它随  $\lambda$  值的变小而增大, 即当  $\lambda_1 < \lambda_2$  时, 有  $A_{\lambda_1} \supseteq A_{\lambda_2}$ 。

定义 3 模糊集运算定义。若  $A, B$  为  $X$  上两个模糊集, 它们的和集、交集和  $A$  的余集都是模糊集, 其隶属函数分别定义为 :

$$(A \vee B)(x) = \max(A(x), B(x))$$

$$(A \wedge B)(x) = \min(A(x), B(x))$$

$$A^C(x) = 1 - A(x)$$

关于模糊集的和、交等运算, 可以推广到任意多个模糊集合中去。

定义 4 若一个矩阵元素取值为  $[0, 1]$  区间内, 则称该矩阵为模糊矩阵。同普通矩阵一样, 有模糊单位阵, 记为  $I$ ; 模糊零矩阵, 记为  $0$ ; 元素皆为 1 的矩阵用  $J$  表示。

定义 5 若  $A$  和  $B$  是  $n \times m$  和  $m \times l$  的模糊矩阵, 则它们的乘积  $C=AB$  为  $n \times l$  阵, 其元素为 :

$$C_{ij} = \bigvee_{k=1}^m (a_{ik} \wedge b_{kj}) \quad (i=1,2,\dots,n; j=1,2,\dots,l) \quad (20.1)$$

符号“ $\bigvee$ ”和“ $\bigwedge$ ”含意的定义为： $a \bigvee b = \max(a, b)$ ， $a \bigwedge b = \min(a, b)$ 。

模糊矩阵乘法性质包括：1)  $(AB)C = A(BC)$ ；2)  $AI = IA = A$ ；3)  $A0 = 0A = 0$ ；4)  $AJ = JA$ ；5) 若  $A, B$  为模糊矩阵且  $a_{ij} \leq b_{ij}$  (一切  $i, j$ )，则  $A \leq B$ ，又若  $A \leq B$ ，则  $AC \leq BC$ ， $CA \leq CB$ 。

## 2. 模糊分类关系

模糊聚类分析是在模糊分类关系基础上进行聚类。由集合的概念，可给出如下定义：

定义 6  $n$  个样品的全体所组成的集合  $X$  作为全域，令  $X \times Y = \{ (X, Y) \mid x \in X, y \in Y \}$ ，则称  $X \times Y$  为  $X$  的全域乘积空间。

定义 7 设  $R$  为  $X \times Y$  上的一个集合，并且满足：

1) 反身性： $(x_i, y_i) \in R$ ，即集合中每个元素和它自己同属一类；

2) 对称性：若  $(x, y) \in R$ ，则  $(y, x) \in R$ ，即集合中  $(x, y)$  元素同属于类  $R$  时，则  $(y, x)$  也同属于  $R$ ；

3) 传递性： $(x, y) \in R$ ， $(y, z) \in R$ ，则有  $(x, z) \in R$ 。

上述三条性质称为等价关系，满足这三条性质的集合  $R$  为一分类关系。

聚类分析的基本思想是用相似性尺度来衡量事物之间的亲疏程度，并以此来实现分类，模糊聚类分析的实质就是根据研究对象本身的属性未构造模糊矩阵，在此基础上根据一定的隶属度来确定其分类关系。

## 3. 模糊聚类

利用模糊集理论进行聚类分析的具体步骤如下：

(1)若定义相似系数矩阵用的是定量观察资料,在定义相似系数矩阵之前,可先对原始数据进行变换处理,变换的方法同系统聚类分析,可参考第17章系统聚类分析一节。

(2)计算模糊相似矩阵。设 $U$ 是需要被分类对象的全体,建立 $U$ 上的相似系数 $R$ , $R(i,j)$ 表示 $i$ 与 $j$ 之间的相似程度,当 $U$ 为有限集时, $R$ 是一个矩阵,称为相似系数矩阵。定义相似系数矩阵的工作,原则上可以按系统聚类分析中的相似系数确定方法,但也可以用主观评定或集体打分的办法。DPS平台,对数据集

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{1m} \\ x_{21} & x_{22} & x_{2m} \\ \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & x_{nm} \end{bmatrix} n \times m$$

提供了以下8种建立相似矩阵的方法:

相关系数法:

最大最小法:

算术平均最小法:

几何平均最小法:

绝对指数法:

绝对值减数法:

夹角余弦法:

欧氏距离:

(3)聚类分析。用上述方法建立起来的相似关系 $R$ ,一般只满足反射性和对称性,不满足传递性,因而还不是模糊等价关系。为此,需要将 $R$ 改造成 $R^*$ 后得到聚类图,在适当的阈值上进行截取,便可得到所需要的分类。将 $R$ 改造成 $R^*$ ,可用求传递闭包的方法。。 $R$ 自乘的思想是按

最短距离法原则, 寻求两个向量  $x_i$  与  $x_j$  的亲密度。

假设  $R^2=(r_{ij})$ , 即  $r_{ij}=\bigvee_{k=1}^n(r_{ik} \quad r_{kj})$ , 说明  $x_i$  与  $x_j$  是通过第三者  $K$  作为媒介而发生关系,  $r_{ik} \quad r_{kj}$  表示  $x_i$  与  $x_j$  的关系密切程度是以  $\min(r_{ik}, r_{kj})$  为准则, 因  $k$  是任意的, 故从一切  $r_{ik} \quad r_{kj}$  中寻求一个使  $x_i$  和  $x_j$  关系最密切的通道。 $R^m$  随  $m$  的增加, 允许连接  $x_i$  与  $x_j$  的链的边就越多。由于从  $x_i$  到  $x_j$  的一切链中, 一定存在一个使最大边长达到极小的链, 这个边长就是相当于  $r_{ij}^\infty$ 。

在实际处理过程中,  $R$  的收敛速度是比较快的。为进一步加快收敛速度, 通常采取如下处理方法:

$$R \quad R^2 \quad R^4 \quad R^8 \quad \dots \quad R^{2^k}$$

即先将  $R$  自乘改造为  $R^2$ , 再自乘得  $R^4$ , 如此继续下去, 直到某一步出现  $R^{2^k}=R^k=R^*$ 。此时  $R^*$  满足了传递性, 于是模糊相似矩阵( $R$ )就被改造成成了一个模糊等价关系矩阵( $R^*$ )。

(4)模糊聚类。对满足传递性的模糊分类关系的  $R^*$  进行聚类处理, 给定不同置信水平的  $\lambda$ , 求  $R_\lambda^*$  阵, 找出  $R^*$  的  $\lambda$  显示, 得到普通的分类关系。当  $\lambda=1$  时, 每个样品自成一类, 随  $\lambda$  值的降低, 由细到粗逐渐归并, 最后得到动态聚类谱系图。

#### 4.DPS 平台操作示例

首先在编辑状态下输入编辑数据, 格式是每一行为一个样本, 每一列为一个变量, 然后将待分析的数据定义成数据矩阵块, 在菜单方式下选择“模糊数学→模糊聚类”功能项, 回车执行时, 系统将提示用户选择数据转换方法:

0.不转换 1.数据中心化 2.对数转换 3.数据规格化 4.数据标准化

作出数据转换方式的选择后,系统又将提示选择建立模糊相似关系的计算方法,共有上面所述的8种方法可供选择。

分析输出的结果包括各个样本的联结序号、联结水平、聚类谱系图索引及在屏幕上显示聚类谱系图(拷屏可得到谱系图硬拷贝,或按 S 将图形文件以“.BMP”格式存放在盘上,然后可在 Windows 有关应用软件中调出)。

## 模糊模式识别

### 1.方法简介

“模式”一词来源于英文 Pattern,原意是典范、式样、样品,在不同场合有其不同的含义。在此我们讲的模式是指具有一定结构的信息集合。

模式识别就是识别给定的事物以及与它相同或类似的事物,也可以理解为模式的分类,即把样品分成若干类,判断给定事物属于哪一类,这与我们前面介绍的判别分析很相似。

模式识别的方法大致可以分为两种,即根据最大隶属原则进行识别的直接法和根据择近原则进行归类的间接法,分别简介如下:

(1)若已知  $n$  个类型在被识别的全体对象  $U$  上的隶属函数,则可按隶属原则进行归类。此处介绍的是针对正态型模糊集的情形。对于正态型模糊变量  $x$ ,其隶属度为

$$A(x) = e^{-\left(\frac{x-a}{b}\right)^2}$$

其中  $a$  为均值,  $b^2=2\sigma^2$ ,  $\sigma^2$  为相应的方差。按泰勒级数展开, 取近似值得

$$A(x) = \begin{cases} 1 - \left(\frac{x-a}{b}\right)^2 & x-a < b \\ 0 & x-a > b \end{cases}$$

若有  $n$  种类型  $m$  个指标的情形, 则第  $i$  种类型在第  $j$  种指标上的隶属函数是

$$A_{ij}(x) = \begin{cases} 0 & x \leq a_{ij}^{(1)} - b_{ij} \\ 1 - \left(\frac{x - a_{ij}^{(1)}}{b_{ij}}\right)^2 & a_{ij}^{(1)} - b_{ij} < x < a_{ij}^{(1)} \\ 1 & a_{ij}^{(1)} \leq x \leq a_{ij}^{(2)} \\ 1 - \left(\frac{x - a_{ij}^{(2)}}{b_{ij}}\right)^2 & a_{ij}^{(2)} < x < a_{ij}^{(2)} + b_{ij} \\ 0 & a_{ij}^{(2)} + b_{ij} < x \end{cases}$$

其中  $a_{ij}^{(1)}$  和  $a_{ij}^{(2)}$  分别是第  $i$  类元素第  $j$  种指标的最小值和最大值,  $b_{ij}^2 = 2\sigma_{ij}^2$ , 而  $\sigma_{ij}^2$  是第  $i$  类元素第  $j$  种指标的方差。

(2) 若有  $n$  种类型  $(A_1, A_2, \dots, A_N)$ , 每类都有  $m$  个指标, 且均为正态型模糊变量, 相应的参数分别为

$a_{ij}^{(1)}$ ,  $a_{ij}^{(2)}$ ,  $b_{ij}$  ( $i=1, 2, \dots, n; j=1, 2, \dots, m$ )。其中,

$a_{ij}^{(1)} = \min(x_{ij})$ ,  $a_{ij}^{(2)} = \max(x_{ij})$ ,  $b_{ij}^2 = 2\sigma_{ij}^2$ , 而  $\sigma_{ij}^2$  是  $x_{ij}$  的方差。待判别

对象  $B$  的  $m$  个指标分别具有参数  $a_j, b_j (j=1, 2, \dots, m)$ , 且为正态型模糊变量, 则  $B$  与各个类型的贴近度为

$$(A_{ij}, B) = \begin{cases} 0 & a_j \leq a_{ij}^{(1)} - (b_j - b_{ij}) \\ 1 - \frac{1}{2} \left( \frac{a_j - a_{ij}^{(1)}}{b_j + b_{ij}} \right)^2 & a_{ij}^{(1)} - (b_j - b_{ij}) < a_j < a_{ij}^{(1)} \\ 1 & a_{ij}^{(1)} \leq a_j \leq a_{ij}^{(2)} \\ 1 - \frac{1}{2} \left( \frac{a_j - a_{ij}^{(2)}}{b_j + b_{ij}} \right)^2 & a_{ij}^{(2)} < a_j < a_{ij}^{(2)} + (b_j + b_{ij}) \\ 0 & a_{ij}^{(2)} + (b_j + b_{ij}) \leq a_j \end{cases}$$

记  $S_i = \min_{1 \leq j \leq m} (A_{ij}, B)$  , 又有  $S_{i0} = \max_{1 \leq j \leq n} (S_i)$  , 按贴近原则可认

为 B 与  $A_{i0}$  最贴近。

序号	年份	幼虫发生量		发生期		增殖系数		类别
		第二代	第三代	第二代	第三代	二至三代	三至四代	
1	1962	344	3333	29	9	9.69	1.91	1
2	1963	121	1497	27	19	12.37	1.34	1
3	1964	187	1813	32	18	9.70	1.06	1
:	:	:	:	:	:	:	:	:
27	1988	162	2817	34	21	2.64	0.00	3
28	1989	760	877	39	33	1.15	0.00	0
29	1990	458	199	35	27	0.43	0.00	0

图 30-3 模糊识别分析的数据编辑定义图

根据如上介绍, DPS 系统中设计了两个功能模块: 一是根据在集合上的隶属函数, 按隶属原则识别对象, 判定样本的类别归属; 二是根据模糊集两两之间的贴近度, 按择近原则, 确定出最接近的两个模糊集。

## 2.DPS 平台的操作示例

系统规定数据输入的格式是每一行为一个样本,每一列为一个变量。最右边的一列为样本的已知类别(如 1,2,...)。(注意每一类中至少要有三个样本)。对于待判别的样本,其分类类别用 0 表示。所有待分析数据(连同类别一起)需定义成数据块,然后进入菜单操作,选择“模糊数学→模糊识别”功能项,回车执行后即可输出分析结果。输出结果包括各类参数(变量名、最小值、最大值、标准差和参数  $B$ )和各待判样本的归类结果(样本序号、对各类贴近度的最大值、最贴近的类号)。

注意事项:系统最多可处理 20 个因子,100 个样本。

例如,在“有序样本最优分割”一节中,我们将历年三化螟发生动态根据最优分割结果分成 3 类,即将三化螟种群消长过程划分为猖獗—缓和—猖獗三个阶段,这样的划分结果与该县历年水稻种植制度(一季中稻为主→纯双季稻→单双季混栽)的变化是相吻合的。为识别 1988 年之后三化螟发生动态,我们也可以应用模糊识别方法进行分析。现将待识别数据和原来的历史资料按上页图 30-3 方式整理编辑和定义。

完成数据编辑定义之后,执行选项功能“模糊识别”,便可得到如下结果:

### 第 1 类

---

变量名 最小值 最大值 标准差 参数  $B$

X(1) 121.000000 500.000000 148.979746 210.689178

X(2) 1497.000000 4600.000000 1243.947346 1759.207208

---

X(3)27.00000036.0000003.3115964.683304

X(4)9.00000019.0000003.5777095.059644

X(5)8.00000012.3700001.6712082.363445

X(6)1.0600001.9100000.3411550.482466

---

---

## 第 2 类

---

变量名最小值最大值标准差参数  $B$

X(1)19.0000002100.000000587.039979830.199901

X(2)25.0000002700.000000948.2482861341.025587

X(3)22.00000040.0000005.0472137.137837

X(4)14.00000028.0000004.3234606.114296

X(5)0.13000010.9600002.7547763.895841

X(6)0.5500004.1700001.0928981.545592

---

---

## 第 3 类

---

变量名最小值最大值准差参数  $B$

X(1)34.0000002243.000000725.4704081025.970090

X(2)401.0000007452.0000002400.9240933395.419414

X(3)31.00000039.0000002.6692703.774917

X(4)18.00000029.0000003.7321005.277987

X(5)1.90000011.7900003.2501084.596346

X(6)0.0000001.0900000.4628170.654523

---

各个待判样本的归类结果

---

样本序号对各类贴近度的最大值最贴近的类号

10.425643

20.897723

---

从分析结果可以看出,1989年和1990年三化螟发生动态仍和前几年相似,表明农业生态系统是相对稳定的。

## 模糊相似优先比方法

### 1.方法简介

相似优先比是模糊性度量的一种形式,它是以成对的样本与一个固定的样本作比较,确定哪一个与固定样本更相似,从而选择与固定样本相似程度较大者。

假定样本  $x_i$  和  $x_j$  与固定样本  $x_k$  进行比较,其相似优先比  $R_{ij}$  必须满足如下要求:

(1)若  $R_{ij}$  在  $[0.5,1.0]$  之间,则表示  $x_i$  比  $x_j$  优先。

(2)若  $R_{ij}$  在  $[0.0,0.5]$  之间,则表示  $x_j$  比  $x_i$  优先。

(3)在极值情形下有三种可能:如果  $R_{ij}=1$ ,则表示  $x_i$  比  $x_j$  显然优先;如果  $R_{ij}=0$ ,则表示  $x_j$  比  $x_i$  显然优先;如果  $R_{ij}=0.5$ ,则  $x_i$  和  $x_j$  不分伯仲,优先无法确定。

在模糊优先比分析中,一般采用海明(Harming)距离

作为相似优先比中  $R_{ij}$  的测度。如对样本  $x_i$  和样本  $x_j$  与固定样本  $x_k$  之间进行比较,海明距离可定义为

$$r_{ij} = \frac{d_{ki}}{d_{ki} + d_{kj}}$$

$$R_{ji} = 1 - R_{ij}$$

式中  $d_{ki} = |x_k - x_i|$ ,  $d_{kj} = |x_k - x_j|$ , 接下来,对给定的一样本集合  $X = \{x_1, x_2, \dots, x_n\}$  和固定样本  $x_k$ , 令任意  $x_i, x_j \in X$  和  $x_k$  作比较,即计算两两样本间的相似优先比,从而得到模糊相关矩阵:

$$R = (r_{ij}) \begin{cases} r_{ij} \in [0, 1] \\ i, j = 1, 2, \dots, n \end{cases}$$

建立模糊相似矩阵之后,由  $\lambda$  水平集选出相似样本,亦即在相似矩阵中,从大到小地选定  $\lambda$  值,以在  $\lambda$  值下降过程中首先到达的除主对角线元素外全行都为 1 的那一行的样本最相似,然后删除矩阵相应的行和列,并降低  $\lambda$  水平值,继续寻找。依此类推,直至截距处理完毕。

一般情形下,若每个样本有  $m$  个因素,则对每一因素都有一个模糊相似矩阵,所以,每一样本的每一因素都将产生一个反映相似程度的序号值,最后将每一样本各个因素的序号值相加,其结果便是该样本与固定样本相似程度的综合反映。

样本的序号值越小,该样本与固定样本就越相似,但严格地说,各个因素对样本的影响程度是不一样的,因此有必要给各个因素赋予一定的权重,这样得到的结果将更符合实际情况。所以当,用户在对有关因素影响的轻重程度有比较大的把握,或在分析中需突出某个因素时,可对各个因素进行加权处理以达到更好的分析效果。

## 2.DPS 平台的操作示例

数据的输入编辑格式是每一行为一个样本,每一列为一个变量,最右边的一列为已知样本的代码(用 1 表示)和待识别样本的代码(用 0 表示),并将数据和待识别样本一起定义成数据块。

在菜单下选择“模糊数学→相似优先比分析”,执行该项功能后系统将输出分析结果。结果包括待识别样本与各样本间的海明距离以及待识别样本与其它样本各个因素的模糊优先比矩阵  $R$ ,最后给出待判样品对各已知样品各变量相似程度和待判样品对各已知样品的优先比值,并按顺序排列。

例如,高素华(1981)对日本柑橘主要产地之一福冈和我国合肥、武汉、长沙、桂林、温州和成都等 7 地柑橘生长的农业气候相似程度进行了分析,选用各地年均温、年降水量、年日照时数、年极端最低气温和 1 月均温作为相似因子。现运用模糊相似优先比方法在 DPS 平台上进行分析。其数据输入、编辑整理和数据块的定义如图 30-4 所示,

地点	年均温	年降水量	年日照时数	年最低气温	1 月均温	识别标识
合肥	15.7	970	2309	-20.6	1.9	1
武汉	16.3	1260	2085	-17.3	2.8	1
上海	15.7	1129	2039	-9.4	3.3	1
长沙	17.2	1422	1726	-9.5	4.6	1
桂林	18.8	1874	1709	-4.9	8.0	1
温州	17.9	1698	1846	-4.5	7.5	1
成都	16.3	976	1239	-4.6	5.6	1
日本	16.2	1492	2000	-8.2	6.2	0

福冈						
----	--	--	--	--	--	--

图 30-4 模糊优先比分析的数据编辑定义图

在执行运算时,系统会提示用户输入各个因素权重比例(注意各比例之和须等于 1),这时如直接回车表示不考虑加权处理。本例分析结果如下。

待报样本 1 与各个样本间绝对值

$x_{10}$  5.000522.0000309.000012.40004.3000

$x_{20}$  1.000232.000085.00009.10003.4000

$x_{30}$  5.000363.000039.00001.20002.9000

$x_{41}$  0.000070.0000274.00001.30001.6000

$x_{52}$  6.000382.0000291.00003.30001.8000

$x_{61}$  7.000206.0000154.00003.70001.3000

$x_{70}$  1.000516.0000761.00003.60000.6000

$X_1$  的模糊优先比矩阵  $R(1)$

1.0000.1670.5000.6670.8390.7730.167

0.8331.0000.8330.9090.9630.9440.500

0.5000.1671.0000.6670.8390.7730.167

0.3330.0910.3331.0000.7220.6300.091

0.1610.0370.1610.2781.0000.3950.037

0.2270.0560.2270.3700.6051.0000.056

0.8330.5000.8330.9090.9630.9441.000

$X_2$  的模糊优先比矩阵  $R(2)$

1.0000.3080.4100.1180.4230.2830.497

0.6921.0000.6100.2320.6220.4700.690

0.5900.3901.0000.1620.5130.3620.587

0.8820.7680.8381.0000.8450.7460.881  
0.5770.3780.4870.1551.0000.3500.575  
0.7170.5300.6380.2540.6501.0000.715  
0.5030.3100.4130.1190.4250.2851.000

### X3 的模糊优先比矩阵 $R(3)$

1.0000.2160.1120.4700.4850.3330.711  
0.7841.0000.3150.7630.7740.6440.900  
0.8880.6851.0000.8750.8820.7980.951  
0.5300.2370.1251.0000.5150.3600.735  
0.5150.2260.1180.4851.0000.3460.723  
0.6670.3560.2020.6400.6541.0000.832  
0.2890.1000.0490.2650.2770.1681.000

### X4 的模糊优先比矩阵 $R(4)$

1.0000.4230.0880.0950.2100.2300.225  
0.5771.0000.1170.1250.2660.2890.283  
0.9120.8831.0000.5200.7330.7550.750  
0.9050.8750.4801.0000.7170.7400.735  
0.7900.7340.2670.2831.0000.5290.522  
0.7700.7110.2450.2600.4711.0000.493  
0.7750.7170.2500.2650.4780.5071.000

### X5 的模糊优先比矩阵 $R(5)$

1.0000.4420.4030.2710.2950.2320.122  
0.5581.0000.4600.3200.3460.2770.150  
0.5970.5401.0000.3560.3830.3100.171  
0.7290.6800.6441.0000.5290.4480.273

0.7050.6540.6170.4711.0000.4190.250

0.7680.7230.6900.5520.5811.0000.316

0.8780.8500.8290.7270.7500.6841.000

待判样品(1)对各个已知样品各变量相似程度

样本  $x_1x_2x_3x_4x_5$  相似程度

12767729.00

21326618.00

32411513.00

43142313.00

55553422.00

64235216.00

71674119.00

待判样品(1)对各已知样品优先比值

313.0000 预测样本与该样本最接近.

413.0000

616.0000

218.0000

719.0000

522.0000

129.0000

从分析结果来看,上海、长沙与福冈柑橘生产的农业气候条件最相似,而合肥的相似程度最小。

## 模糊综合评判

### 1.方法简介

综合评判就是对受到多个因素制约的事物或对象作出一个总的评价,这是在日常生活和科研工作中经常遇到的问题,如产品质量评定、科技成果鉴定、某种作物种植适应性的评价等,都属于综合评判问题。由于从多方面对事物进行评价难免带有模糊性和主观性,采用模糊数学的方法进行综合评判将使结果尽量客观从而取得更好的实际效果。

模糊综合评判的数学模型可分为一级模型和多级模型,在此仅介绍一级模型。采用一级模型进行综合评判,一般可归纳为以下几个步骤:

(1)建立评判对象因素集  $U=\{u_1, u_2, \dots, u_n\}$ 。因素就是对象的各种属性或性能,在不同场合,也称为参数指标或质量指标,它们能综合地反映出对象的质量,因而可由这些因素来评价对象。

(2)建立评判集  $V=\{v_1, v_2, \dots, v_m\}$ 。如工业产品的评价,评判集是等级的集合;农作物种植区域适应性的评价,评判集是适应程度的集合。

(3)建立单因素评判,即建立一个从  $U$  到  $F(V)$  的模糊映射

$$f: U \rightarrow F(V), \forall u_i \in U$$

$$u_i \mapsto f(u_i) = \frac{r_{i1}}{v_1} + \frac{r_{i2}}{v_2} + \Lambda + \frac{r_{im}}{v_m}$$

$$0 \leq r_{ij} \leq 1, 1 \leq i \leq n, 1 \leq j \leq m$$

由  $f$  可以诱导出模糊关系,得到模糊矩阵

$$R = \begin{bmatrix} r_{11} & r_{12} & \Lambda & r_{1m} \\ r_{21} & r_{22} & \Lambda & r_{2m} \\ \Lambda & \Lambda & \Lambda & \Lambda \\ r_{n1} & r_{n2} & \Lambda & r_{nm} \end{bmatrix}$$

称  $R$  为单因素评判矩阵, 于是  $(U, V, R)$  构成了一个综合评判模型。

(4) 综合评判。由于对  $U$  中各个因素有不同的侧重, 需要对每个因素赋予不同的权重, 它可表示为  $U$  上的一个模糊子集  $A=(a_1, a_2, \dots, a_n)$ , 且规定  $\sum_{i=1}^n a_i = 1$ 。

在  $R$  与  $A$  求出之后, 则综合评判模型为  $B=A \circ R$ 。记  $B=(b_1, b_2, \dots, b_m)$ , 它是  $V$  上的一个模糊子集, 其中

$$b_j = \bigvee_{i=1}^n (a_i \wedge r_{ij}) \quad (j=1, 2, \dots, m)。$$

如果评判结果  $\sum_{j=1}^m b_j \neq 1$ , 就对其结果进行归一化处理。

从上述模糊综合评判的 4 个步骤可以看出, 建立单因素评判矩阵  $R$  和确定权重分配  $A$  是两项关键性的工作, 但同时又没有统一的格式可以遵循, 一般可采用统计实验或专家评分的方法求出。在 DPS 平台上, 只是根据给出的评判矩阵  $R$  和确定的权重分配  $A$  进行综合评判处理。但是, 用户可以在平台上利用公式计算, 根据定义公式, 计算其隶属函数。

## 2. DPS 平台的操作示例

根据给定的评估数据(评判矩阵  $R$ )和最终的评价结果(权重分配  $A$ ), 确定出该类评估近似的权分配系数。

例如对某教师讲课质量进行综合评定。已知因素集合  $U$  为:  $U=\{\text{教材熟练, 逻辑性强, 启发性强, 语言生动, 板书整齐}\}$ , 评语集合  $V=\{\text{很好, 较好, 一般, 不好}\}$ 。首先, 建立评估集合(评判矩阵), 即评价矩阵数据编辑和定义。在 DPS 平台上, 以行代表因素集, 列代表评语集。每一行为因素集  $U$  中某一单因素的评价结果, 最后一行存放最

终评价值(建立权重分配),本例的权重分配为 $A=(0.3,0.2,0.2,0.2,0.1)$ 并放在数据块的最后一行。最后,评价矩阵数据块可按图 30-5 格式输入及定义:

因素 \ 评语	很好	较好	一般	差	
教材熟练	0.45	0.25	0.2	0.1	
逻辑性强	0.50	0.40	0.1	0.0	
启发性强	0.30	0.40	0.2	0.1	
语言生动	0.40	0.40	0.1	0.1	
板书整齐	0.30	0.50	0.1	0.1	
最终评价值	0.30	0.20	0.2	0.2	0.1

图 30-5 模糊综合评判的数据编辑定义图

完成评估数据(评判矩阵  $R$ )和最终的评价结果(权重分配  $A$ )数据块的编辑及定义之后,在菜单下选择“模糊数学→模糊综合评判分析”功能项,回车执行后即得分析结果。评判结果表明,对该教师的课堂教学认为“很好”的占 35.29%，“较好”的占 29.41%，“一般”的占 23.53%，“不好”的占 11.76%。根据最大隶属原则,结论是“很好”。

## 模糊关系方程求解

### 1. 方法简介

模糊关系方程是模糊数学的一个重要组成部分。如下所示，它其实是模糊综合评判的逆问题。

$$\begin{array}{ccc} \xrightarrow{\text{输入}A} & R & \xrightarrow{\text{输出}B=?} \\ & A.R=B & \end{array} \qquad \begin{array}{ccc} \xrightarrow{\text{输入}A=?} & R & \xrightarrow{\text{输出}B} \\ & A.R=B & \end{array}$$

这类问题具有普遍实际意义。如一些老专家、老中医或具有丰富经验的实际工作人员，在他们的头脑里，经验和技能常归结于对诸因素有一种优越的权数分配方案，这些难以言传的经验和技能，可望利用模糊数学原理，借助计算机技术进行模拟并保存下来。

在 DPS 系统中，求解模糊关系方程采用徐罗曹李方法(汪培庄 1983)。其求解过程首先考虑  $X \circ R = B$  类型的模糊关系方程：

$$(x_1, x_2, \dots, x_n) \times \begin{bmatrix} r_{11} & r_{12} & \Lambda & r_{1m} \\ r_{21} & r_{22} & \Lambda & r_{2m} \\ \Lambda & \Lambda & \Lambda & \Lambda \\ r_{n1} & r_{n2} & \Lambda & r_{nm} \end{bmatrix} = (b_1, b_2, \dots, b_m)$$

按照模糊合成运算的最大—最小法则，上式可化为下面一组线性等式，称为模糊线性方程：

$$\begin{cases} (r_{11} \wedge x_1) \vee (r_{21} \wedge x_2) \vee \dots \vee (r_{n1} \wedge x_n) = b_1 \\ (r_{12} \wedge x_1) \vee (r_{22} \wedge x_2) \vee \dots \vee (r_{n2} \wedge x_n) = b_2 \\ \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \\ (r_{1m} \wedge x_1) \vee (r_{2m} \wedge x_2) \vee \dots \vee (r_{nm} \wedge x_n) = b_m \end{cases}$$

因此，求解模糊关系方程实际上就是对上述方程组的求解，其步骤简述如下：

(1) 标准化排列。将向量  $B$  写于  $A$  的上方，按  $\bar{b}_1 \geq \bar{b}_2 \geq \dots \geq \bar{b}_m$  的次序更换向量  $B$  的分量，同时矩阵  $R$  也作

相应的更换，由此得到矩阵的标准化排列。

(2)上铣。按上述规则更换后的矩阵，对每一个  $j(j=1,2,\dots,m)$ ，用  $\bar{b}_j$  的上铣第  $j$  列，即若  $\bar{r}_{ij} > \bar{b}_j$ ，将  $\bar{r}_{ij}$  变成  $\bar{b}_j$ ；

若  $\bar{r}_{ij} \leq \bar{b}_j$ ，将  $\bar{r}_{ij}$  变成空元  $\phi$ 。从而得到新的  $r_{ij}$ ，用  $r'_{ij}$  表示：

$$r'_{ij} = \begin{cases} \bar{b}_j & (\bar{r}_{ij} > \bar{b}_j) \\ \phi & (\bar{r}_{ij} \leq \bar{b}_j) \end{cases}$$

(3)求下确界。对上铣后所得矩阵的各行求下确界，得向量  $UR=(ur_1,ur_2,\dots,ur_n)$ ，并定义空元的下确界为 1。

(4)平铣。将上铣出来的矩阵称为平铣矩阵，即对每一个元素  $j$  分别用  $\bar{b}_j$  平铣  $R$  的第  $j$  列：若  $\bar{r}_{ij} \geq \bar{b}_j$ ，将  $\bar{r}_{ij}$  变成  $\bar{b}_j$ ；

若  $\bar{r}_{ij} < \bar{b}_j$ ，将  $\bar{r}_{ij}$  变成空元  $\phi$ 。这样得到新的  $r_{ij}$ ，以  $r''_{ij}$  表示：

$$r''_{ij} = \begin{cases} \bar{b}_j & (\bar{r}_{ij} \geq \bar{b}_j) \\ \phi & (\bar{r}_{ij} < \bar{b}_j) \end{cases}$$

(5)划元。对平铣后的矩阵，逐行划去该行中大于上界即第  $i$  行大于  $uR_i$  的元素。

(6)判别。若原方程有解，则上一步所得矩阵的每一列都有未被划去的非空白元素。

(7)求解。从经过划元后所得的矩阵中的每一列，选定一个非空白、且未被划元的元素，对这些当选元素逐行取上确界，并规定空集  $\phi$  的上确界为 0。这样得到的一组解为“拟极小解”。当被选元素不同时，可得到多个“拟极小解”。

最后得到的解之所以称为“拟极小解”，因为它们不

一定都是极小解,它们相互之间还可能存在着重合或优劣的模糊关系,需进一步进行筛选。实际上,当方程经判别有解时,第三步所求的上界就是方程的最大解。

应用 DPS 平台求解模糊关系方程时,当模糊方程有解时,系统将给出最大解和最小解,如模糊方程无解时则提示没有解。

## 2.DPS 平台的操作示例

在平台上求解模糊关系方程时,其数据编辑格式是:行为因素集,列为评语集,每一行为因素集  $U$  中某一单因素的评价结果,最后一行存放整个因素集  $U$  的综合评价结果,然后将数据定义成数据块。如对某单位的管理工作进行“民意测验”,得综合评价矩阵和最终评价结果,并按系统规定格式编辑整理数据和定义数据块,如图 30-6。

然后,在菜单下选择“模糊数学→模糊关系方程”功能项,回车执行后即得到归一化之后的分析结果,详见于后。

注意:本系统最多可处理 50 个因子的因子集。

	很好	较好	一般	差	
政治思想工作(X1)	0.04	0.30	0.40	0.26	评价矩阵
精神文明建设(X2)	0.24	0.28	0.28	0.20	
行政及其管理(X3)	0.10	0.34	0.40	0.16	
职工组织纪律(X4)	0.02	0.26	0.40	0.32	

人才培训 工作(X5)	0.13	0.28	0.42	0.17	
学术科研 工作(X6)	0.14	0.27	0.30	0.29	
后勤管理 工作(X7)	0.03	0.14	0.42	0.41	
整个因素集 U综合评价	0.15	0.29	0.34	0.22	

图 30-6 求解模糊关系方程的数据编辑、定义示意图

输出结果

最小解  $X_i$  最大解归一处理后结果

0.2200<--x(1)-->0.22000.1325

0.1500<--x(2)-->0.15000.0904

0.2900<--x(3)-->0.29000.1747

0.0000<--x(4)-->0.22000.1325

0.3400<--x(5)-->0.34000.2048

0.0000<--x(6)-->0.22000.1325

0.0000<--x(7)-->0.22000.1325

## 综合评判的逆问题

### 1.方法简介

前面的模糊综合评判和模糊关系求解是综合评判的

正反两个方面。由于权重分配  $A$  的确定并无通用公式, 所以它的正确与否往往只能取决于专家的判断或经验, 而这些又是很难用数学公式表达出来的。与之相反, 权重分配  $B$  可以通过实践的检验建立起来。由  $R$  和  $B$  反过来求  $A$ , 有利于总结专家的经验, 使它得到量化。同样道理, 由  $A$  和  $R$  而求  $R$  可以帮助我们检验建立起来的数学模型是否合适。从这个角度来看, 综合评判的逆问题比其正问题更有意义。

上面介绍的模糊关系方程求解, 即求解综合评判逆问题时, 其方程可能有解, 也可能无解。有解时, 解也可能有多个, 这需根据实际情况作出恰当选择。若无解, 又应当怎么办呢? DPS 系统提供根据贴近原则解模糊关系方程的方法。

首先选择一些备择解的集合, 设它们为  $J=\{A_1, A_2, \dots, A_n\}$ , 然后用贴近度原则在备择解集  $\{A_1, A_2, \dots, A_n\}$  中选取一个  $A_i$  作为模糊关系方程  $X \circ R = B$  中

的  $X$ 。将  $A_i$  与  $R$  进行合成: 
$$\begin{cases} A_1 \circ R = B_1 \\ A_2 \circ R = B_2 \\ \dots \dots \dots \\ A_n \circ R = B_n \end{cases}$$

求出  $A$  之后, 再用贴近度公式

$$(A_i, B) = \frac{1}{2} [A_i, B + (1 - A_i) B]$$

分别求出  $B$  与  $A_1, A_2, \dots, A_n$  的贴近度, 选最大者作为近似的  $X^*$ , 将它作为方程的近似解。

## 2. DPS 平台操作示例

根据给定的评估数据(评判矩阵  $R$ 、最终的评价结果  $B$  和备择权重方案  $A$ , 给出与预定若干个权重数分配方案的

贴近程度,从而确定出该类评估近似的权分配系数。这是一种近似求解模糊关系方程的方法,也可用来评价某一类量对于某些标准定量的贴近程度。

(1)评估集合。评价矩阵的数据编辑和定义格式是行为因素集,列为评语集。每一行为因素集  $U$  中某一单因素的评价结果,最后一行存放整个因素集  $U$  的综合评价结果,再将数据定义成数据块。例如,对某品牌的自行车进行评价,已知因素集合  $U=\{\text{外型,质量,价格}\}$ ,评语集合  $V=\{\text{很好,较好,一般,较差}\}$ ,其评价矩阵数据可按图 30-7 格式定义。

因素\评语	很好	较好	一般	差
外型	0.2	0.7	0.1	0.0
质量	0.0	0.4	0.5	0.1
价格	0.2	0.3	0.4	0.1
最终评价值	0.0	0.8	0.2	0.0

图 30-7 贴近度解模糊关系方程评语集及评价结果数据编辑定义图

(2)可能的权分配方案集合(A)的矩阵数据编辑和定义。每一行为一个权的分配方案,每一列为某一个因素的权重。数据编辑后将数据定义成公式块。如上例根据对顾客的心理分析,提出了四种可能的权分配方案,其数据按图 30-8 方式编辑和定义。注意,此处的数据块定义时,须在按下 Ctrl 的同时拖动鼠标。

	外型	质量	价格
--	----	----	----

第一种权分 配方案:	0.2	0.50	0.30
第二种权分 配方案:	0.5	0.30	0.20
第三种权分 配方案:	0.2	0.30	0.50
第四种权分 配方案:	0.7	0.25	0.05

图 30-8 贴近度解模糊关系方程权重备择方案数据编辑定义图

完成评估数据(评判矩阵  $R$ )、最终的评价结果  $B$  以及备择权重分配  $A$  的数据的编辑、定义之后,再进入菜单操作,在菜单方式下选择“模糊数学→综合评判逆问题”分析功能项,回车后便可得到归一化之后的分析结果。

### 分析结果

#### .贴近度分析结果

第 1 组的贴近度=0.6500

第 2 组的贴近度=0.7000

第 3 组的贴近度=0.6000

第 4 组的贴近度=0.8000

#### .近似的权重方案

第 1 个因素的权数=0.7000

第 2 个因素的权数=0.2500

第3个因素的权数=0.0500

从上述分析结果可以看出,在权重备择集中应选  $A_4$  作为近似的  $X^*$ , 即

$$X^*=(0.7,0.25,0.05)$$

采用此方法,如果备择权分配方案太少,计算结果会不太理想。因此在实际应用时,为取得较满意的结果,应当尽可能多地建立一些权重分配方案。

## 中华人民共和国统计法实施细则

(1987年1月19日经国务院批准,1987年2月15日国家统计局发布;2000年6月2日国务院批准修订,2000年6月15日国家统计局发布)

### 总则

#### 第一条

根据《中华人民共和国统计法》(以下简称《统计法》)的规定,制定本细则。

#### 第二条

《统计法》所指的统计,是指运用各种统计方法对国民经济和社会发展情况进行统计调查、统计分析,提供统计资料和统计咨询意见,实行统计监督等活动的总称。国民经济和社会发展的统计项目分类,由国家统计局规定、调整。

#### 第三条

国家有计划地用现代信息技术装备各级人民政府统计机构，建立健全国家统计信息自动化系统。国务院各部门根据工作需要，有计划地用现代信息技术装备本部门及其管辖系统的统计机构。

县级以上各级人民政府应当将国家统计信息工程建设列入发展计划。国家统计信息工程建设，由国家统计局统一领导，县级以上地方各级人民政府统计机构分级负责。

#### 第四条

统计机构和统计人员实行工作责任制，实行考核和奖惩制度，不断提高工作质量和工作效率。

统计机构和统计人员依法独立行使下列职权：

(一)统计调查权 调查、搜集有关资料，召开有关调查会议，检查与统计资料有关的原始记录和凭证。统计调查对象应当依照《统计法》和国家有关规定，如实提供统计资料 and 情况，不得虚报、瞒报、拒报、迟报，不得伪造、篡改。

(二)统计报告权 将统计调查取得的统计资料 and 情况加以整理、分析，向上级领导机关和有关部门提出统计报告。任何单位或者个人不得阻挠和扣压统计报告，不得篡改统计资料。

(三)统计监督权 根据统计调查和统计分析，对国民经济和社会发展情况进行统计监督，检查国家政策和计划的实施，考核经济效益、社会效益和工作成绩，检查和揭露存在的问题，检查虚报、瞒报、伪造、篡改统计资料的行为，提出改进工作的建议。有关部门和单位对统计机构、统计人员反映、揭露的问题和提出的建议，应当及时处理，作出答复。

## 第五条

县级以上地方各级人民政府、各部门、各企业事业组织，应当根据国家统计任务和本地区、本部门、本单位的需要，在下列方面加强对统计工作的领导和监督：

(一)领导和支持统计机构、统计人员和其他有关人员执行统计法规和统计制度，准确、及时地完成统计工作任务，加强统计工作现代化建设；

(二)吸收和组织统计人员参加讨论有关政策和计划、研究经济和社会发展问题的会议，发挥统计的服务和监督作用；

(三)根据国家统一部署，组织实施重大的国情国力普查；

(四)按照规定审批统计调查计划，切实解决经批准的统计调查需要的人员和经费。

## 第六条

国家统计局和县级以上地方各级人民政府统计局是国家执行统计法规和统计制度的机关，负责监督检查统计法规和统计制度的实施，为同级机构和统计人员的职权，同违反统计法规和统计制度的行为作斗争。

## 统计调查计划和统计制度

### 第七条

县级以上各级人民政府统计机构和有关部门按照下列三类情况，分别建立统计制度，编制统计调查计划，按照规定经审查机关批准后实施：

(一)国家统计调查，是指全国性基本情况的统计调查，包括国家统计局单独拟订的和国家统计局与国务院

有关部门共同拟订的统计调查项目。国家统计调查计划中新的、重大的统计调查项目，由国家统计局报国务院审批；经常性的、一般性的统计调查项目，由国家统计局审批。

各地方、各部门、各单位必须严格按照国家统计调查方案实施国家统计调查。

(二)部门统计调查，是指各部门的专业性统计调查。部门统计调查计划和统计调查方案，由该部门的统计机构组织本部门各有关职能机构编制。其中，统计调查对象属于本部门管辖系统内的，由本部门领导人审批，报国家统计局或者本级地方人民政府统计机构备案；统计调查对象超出本部门管辖系统的，报国家统计局或者本级地方人民政府统计机构审批，其中重要的，报国务院或者本级地方人民政府审批。各部门统计调查管辖系统的划分办法，由国家统计局会同国务院有关部门提出，报国务院批准后实施。

(三)地方统计调查，是指地方人民政府需要的地方性的统计调查。地方统计调查计划和统计调查方案的报批办法，由省、自治区、直辖市人民政府统计机构规定，报国家统计局备案。

#### 第八条

部门统计调查和地方统计调查不得与国家统计调查重复、矛盾。

国家统计调查与部门统计调查、地方统计调查的分工，由国家统计局会同国务院有关部门和省、自治区、直辖市人民政府统计机构具体商定。

#### 第九条

县级以上各级人民政府的综合协调部门需要的统计

资料，应当从本级人民政府统计机构和有关部门搜集；确实需要直接进行统计调查的，应当编制统计调查计划和统计调查方案，依照《统计法》和本细则的有关规定，经批准后实施。

县级以上各级人民政府有关部门组织实施的部门统计调查，应当及时向本级人民政府统计机构报送基本统计资料或者综合统计资料。

国家统计局和县级以上地方各级人民政府统计机构，应当定期、无偿地向本级人民政府部门提供有关综合统计资料。

#### 第十条

统计调查计划按照统计调查项目编制。统计调查项目，是指一定时期内为实现特定统计调查目的而组织实施的统计调查。统计调查项目的计划应当列明：项目名称、调查机关、调查目的、调查范围、调查对象、调查方式、调查时间、调查的主要内容。

编制统计调查计划，必须同时编制统计调查方案。统计调查方案应当包括下列内容：

- (一) 供统计调查对象填报用的统计调查表和说明书；
- (二) 供整理上报用的统计综合表和说明书；
- (三) 统计调查需要的人员和经费及其来源。

#### 第十一条

县级以上各级人民政府统计机构、各部门统计机构对送审的统计调查计划和统计调查方案的必要性、可行性、科学性应当进行严格审查；对不符合本细则规定的，应当退回修改或者不予批准。编制和审查统计调查方案，应当遵循下列原则：

(一)在已经批准实施的各种统计调查中能够搜集到资料的,不得重复调查;

(二)抽样调查、重点调查或者行政记录可以满足需要的,不得制发全面统计调查表;一次性统计调查可以满足需要的,不得进行经常性统计调查;按年统计调查可以满足需要的,不得按季统计调查;按季统计调查可以满足需要的,不得按月统计调查;月以下的进度统计调查必须从严控制;

(三)编制新的统计调查方案,必须事先试点或者征求有关地方、部门和基层单位的意见,进行可行性论证,保证切实可行,注重调查效益;

(四)统计调查需要的人员和经费应当有保证。

#### 第十二条

国家建立周期性的普查制度。周期性普查由国务院和地方人民政府统一领导,组织统计机构和有关部门共同实施,所需要的经费由中央和地方财政共同负担。

进行经常性抽样调查,应当通过基本统计单位普查和行政记录的方式,查明基本统计单位及其分布情况,建立科学的抽样框,按照随机原则在调查总体中选取足以代表总体的样本单位,减少抽样误差。

#### 第十三条

按照规定程序批准的统计调查表,必须在右上角标明表号、制表机关、批准或者备案机关、批准或者备案文号、有效期限。

对未标明前款所列内容或者超过有效期限的统计调查表,有关统计调查对象有权拒绝填报,统计机构有权废止。

#### 第十四条

统计调查方案所规定的指标涵义、调查范围、计算方法、分类目录、调查表式、统计编码等,未经批准该统计调查方案的机关同意,任何单位或者个人不得修改。

## 统计资料的管理和公布

### 第十五条

各地方、各部门、各单位应当健全统计资料的审核制度,保障统计资料的准确性和及时性。

各部门、各企业事业组织提供的统计资料,由本部门、本单位领导人或者统计负责人审核、签署或者盖章后上报。有关财务统计资料由财务会计机构或者会计人员提供,并经财务会计负责人审核、签署或者盖章。县级以上各级人民政府统计机构和乡、镇统计员提供的统计资料,由本级人民政府统计机构负责人或者乡、镇统计员审核、签署或者盖章后上报。

### 第十六条

各级领导机关制定政策、计划,检查政策、计划执行情况,考核经济效益、社会效益和工作成绩,进行奖励和惩罚等,需要使用统计资料的,必须依照《统计法》第十三条的规定,以统计机构或者统计负责人签署或者盖章的统计资料为准。

### 第十七条

县级以上各级人民政府统计机构必须做好统计信息咨询服务工作,充分利用可以公开的社会经济信息为社会公众服务。

符合国家有关规定,在《统计法》和统计制度规定之外提供统计信息咨询,实行有偿服务。具体办法由国

国家统计局会同国务院价格主管部门制定。

### 第十八条

各地方、各部门、各单位必须执行国家有关统计资料保密管理的规定，加强对统计资料的保密管理。

### 第十九条

各地方、各部门、各单位必须建立统计资料档案制度。统计资料档案的保管、调用和移交，应当遵守国家有关档案管理的规定。

### 第二十条

国家建立健全统计资料定期公布制度。

国家统计局统计调查取得的统计数据，由国家统计局公布。

国务院有关部门统计调查取得的统计数据，由国务院有关部门公布；其中，与国家统计局统计调查取得的统计数据有重复、交叉的，应当同国家统计局协商后，由国务院有关部门公布。国务院有关部门公布统计数据，应当自公布之日起 10 日内报国家统计局备案。

县级以上地方各级人民政府统计机构和有关部门公布其统计调查取得的地方统计数据，比照前款规定执行。

### 第二十一条

国家建立健全统计数据质量监控和评估的制度，加强对各省、自治区、直辖市国内生产总值等重要统计数据的监控和评估。

## 统计机构和统计人员

### 第二十二条

国家统计局履行下列职责：

(一)根据有关法律、行政法规和国家有关政策和计划,制定统计工作规章,制订统计工作现代化规划和国家统计调查计划,组织领导和协调全国统计工作,监督检查统计法规和统计制度的实施;

(二)健全国民经济核算制度和统计指标体系,制定全国统一的基本统计报表制度;制定或者与有关部门共同制定国家统计标准,审定部门统计标准;

(三)在国务院领导下,会同有关部门组织重大的国情国力普查,组织、协调全国城市和农村社会经济抽样调查;

(四)根据国家制定政策、计划和进行管理的需要,搜集、整理、提供全国性的基本统计资料,对国民经济和社会发展情况进行统计分析、统计预测和统计监督;

(五)审查国务院各部门编制的统计调查计划和统计调查方案,管理国务院各部门制发的统计调查表;

(六)检查、审定、管理、公布、出版全国性的基本统计资料,定期发布全国国民经济和社会发展情况的统计公报;

(七)统一领导和管理全国城市和农村社会经济抽样调查队;

(八)组织指导全国统计科学研究、统计教育、统计干部培训和统计书刊出版工作;

(九)开展统计工作和统计科学的国际交流。

第二十三条县级以上地方各级人民政府统计机构履行下列职责:

(一)完成国家统计调查任务,执行国家统计标准,执行全国统一的基本统计报表制度;

(二)制订本行政区域内的统计工作现代化规划、统

计调查计划和统计调查方案，统一领导和协调本行政区域内包括中央和地方单位的统计工作，监督检查统计法规和统计制度的实施；

(三)根据本行政区域内制定计划和进行管理的需要，搜集、整理、提供基本统计资料，对本行政区域内国民经济和社会发展情况进行统计分析、统计预测和统计监督；

(四)审查本行政区域内各部门的统计调查计划和统计调查方案，管理本行政区域内各部门制发的统计调查表；

(五)按照国家有关规定，检查、审定、管理、公布、出版本行政区域内的基本统计资料；省、自治区、直辖市人民政府统计机构定期发布本行政区域内国民经济和社会发展情况的统计公报；自治州、县、自治县、市、市辖区人民政府统计机构按照本级人民政府的决定，发布本行政区域内国民经济和社会发展情况的统计公报；

(六)统一管理本行政区域内城市和农村社会经济抽样调查队；

(七)组织指导本行政区域内各部门、各单位加强统计基础工作建设，加强统计教育、统计干部培训和统计科学研究工作；对本行政区域内人民政府统计机构干部和乡、镇统计员进行考核和奖励。

县级以上地方各级人民政府统计机构受本级人民政府和上级人民政府统计机构的双重领导，在统计业务上以上级人民政府机构的领导为主。

#### 第二十四条

乡、镇统计员执行乡、镇综合统计的职能，履行下列职责：

(一)完成国家统计局调查和地方统计调查任务,执行国家统计标准,执行全国统一的基本统计报表制度,执行统计法规和统计制度,监督检查统计法规和统计制度的实施;

(二)按照国家有关规定,搜集、整理、分析、提供和管理本乡、镇的基本统计资料;

(三)组织指导本乡、镇各有关单位、人员加强农村统计基础工作建设,健全本乡、镇的统计台账制度和统计档案制度,组织乡、镇以下的统计业务工作。

乡、镇人民政府应当根据《统计法》等有关规定和统计工作的需要,设置专职的或者兼职的统计员,建立健全乡、镇统计信息网络。乡、镇统计员和乡、镇统计信息网络在统计业务上受县级人民政府统计机构的领导。

村的统计工作,由村民委员会指定专人负责,其在统计业务上受乡、镇统计员的领导。

第二十五条县级以上各级人民政府有关部门的统计机构或者统计负责人执行本部门综合统计的职能,履行下列职责:

(一)组织指导、综合协调本部门各职能机构(包括生产、供销、基建、劳动人事、财务会计等机构)的统计工作,共同完成国家统计局调查、部门统计调查和地方统计调查任务,执行统计法规和统计制度,监督检查统计法规和统计制度的实施;

(二)制订本部门的统计工作现代化规划、统计调查计划和统计调查方案,组织指导本部门及其管辖系统内企业事业组织的统计工作,加强统计队伍和统计基础工作建设;

(三)按照国家有关规定,向上级领导机关和本级人民政府统计机构报送和提供本部门的基本统计资料,会同计划和其他有关职能机构对本部门执行政策、计划和经营管理效益的情况,进行统计分析、统计预测和统计监督;

(四)管理本部门制发的统计调查表和基本统计资料;

(五)会同本部门的人事教育机构,组织指导本部门的统计教育和统计干部培训;对本部门统计人员进行考核和奖励;加强本部门统计科学研究工作。

县级以上各级人民政府有关部门统计机构的设置,应当根据实际需要,本着精简、效能的原则,依照《统计法》的规定执行。

#### 第二十六条

企业事业组织的统计机构或者统计负责人执行本单位综合统计的职能,履行下列职责:

(一)组织指导、综合协调本单位各职能机构和下属机构的统计工作,共同完成国家统计调查、部门统计调查和地方统计调查任务,制订、实施本单位的统计工作计划和统计制度,执行统计法规和统计制度,监督检查统计法规和统计制度的实施;

(二)按照国家有关规定,向主管部门和所在地人民政府统计机构或者乡、镇统计员报送和提供统计资料,对本单位计划的执行情况和经营管理的效益,进行统计分析和统计监督;

(三)管理本单位的统计调查表和基本统计资料;

(四)会同本单位有关职能机构完善计量、检测制度,建立健全原始记录、统计台账和核算制度。

企业事业组织的统计机构或者统计负责人在统计业务上，受所在地人民政府统计机构或者乡、镇统计员的指导。

中小型企业事业组织不单设统计人员的，可以指定人员专门负责统计工作。

### 第二十七条

统计负责人，是指代表本部门或者本单位履行《统计法》规定职责的主要责任人员。不设统计机构的，一般应当由具备相当统计专业技术职务条件的人员担任统计负责人。

### 第二十八条

各地方、各部门、各单位应当根据国家有关规定和工作需要，设置统计专业技术职务。

### 第二十九条

具有统计专业技术职务的人员的调动，应当分别征求本地区、本部门、本单位统计机构或者统计负责人的意见；其中，具有中级以上统计专业技术职务的人员的调动，应当征得上级统计机构的同意。

县级以上地方各级人民政府统计机构主要负责人的调动，应当征得上一级人民政府统计机构的同意。乡镇统计员的调动，应当征得县级人民政府统计机构的同意。

各部门和企业事业组织统计负责人的调动，应当征求上级主管部门和所在地人民政府统计机构的意见。

### 第三十条

国家统计局和县级以上地方各级人民政府统计机构，应当有计划地对统计人员进行培训，加强对统计人员的职业道德教育，提高统计人员的业务素质。

国家统计局和县级以上地方各级人民政府统计机构

增加和补充统计人员，应当从具备统计专业知识的人员中选调。

## 奖励和惩罚

### 第三十一条

县级以上各级人民政府统计机构、各部门、各企业事业组织，应当依照国家或者企业事业组织的规定，对有下列表现之一的统计人员或者集体，定期评比，给予奖励：

(一)在改革和完善统计制度、统计方法等方面，做出重要贡献的；

(二)在完成规定的统计调查任务，保障统计资料的准确性、及时性方面，做出显著成绩的；

(三)在进行统计分析、统计预测和统计监督方面，有所创新，取得重要成绩的；

(四)在运用和推广现代信息技术方面，取得显著效果的；

(五)在改进统计教育和统计专业培训，进行统计科学研究，提高统计科学水平方面，做出重要贡献的；

(六)坚持实事求是，依法办事，同违反统计法规和统计制度的行为作斗争，表现突出的；

(七)揭发、检举统计违法行为有功的。

奖励分为：通令嘉奖、记功、记大功、晋级、升职、授予荣誉称号，并可以发给奖品、奖金。奖金按照国家或者企业事业组织的规定在有关经费中开支。

### 第三十二条

下列行为，属于《统计法》第二十七条第一款所称

情节较重的违法行为：

(一)虚报、瞒报、伪造、篡改统计资料数额较大或者占应报数额的份额较多的；

(二)虚报、瞒报、伪造、篡改或者拒报统计资料，二年内再次发生的；

(三)虚报、瞒报、伪造、篡改、拒报或者屡次迟报统计资料，被责令改正而拒不改正的；

(四)虚报、瞒报、伪造、篡改、拒报或者屡次迟报统计资料，造成严重后果或者恶劣影响的；

(五)在接受统计检查时，拒绝提供情况、提供虚假情况或者转移、隐匿、毁弃原始统计记录、统计台账、统计报表以及与统计有关的其他资料的；

(六)使用暴力或者威胁的方法阻挠、抗拒统计检查的；

(七)国家统计局依法认定的其他行为。

### 第三十三条

企业事业组织有《统计法》第二十七条第一款所列违法行为之一的，由县级以上人民政府统计机构予以警告，并可以处5万元以下的罚款。

个体工商户有《统计法》第二十七条第一款所列违法行为之一的，由县级以上人民政府统计机构予以警告，并可以处1万元以下的罚款。

### 第三十四条

任何单位或者个人有《统计法》第二十九条第二款所列违法行为的，由县级以上人民政府统计机构责令改正，没收违法所得，并可以处违法所得1倍以上3倍以下的罚款；没有违法所得的，可以处3万元以下的罚款。

## 附则

### 第三十五条

中华人民共和国境外的组织、个人需要在中华人民共和国境内进行统计调查活动的，应当委托中华人民共和国境内具有涉外统计调查资格的机构进行。

统计调查范围限于省、自治区、直辖市行政区域内的，应当持有关证明文件和统计调查方案，向省、自治区、直辖市人民政府统计机构提出申请，由省、自治区、直辖市人民政府统计机构审批；统计调查范围跨省、自治区、直辖市行政区域的，应当持有关证明文件和统计调查方案，向国家统计局提出申请，由国家统计局审批。

### 第三十六条

本细则自发布之日起施行。

## 统计报告案例

### 2003 年安西县社会统计分析报告

2003 年，我县经济和社会各项事业在县委、县政府的正确领导下，高举邓小平理论伟大旗帜，深入贯彻落实党的十六大精神，努力实践“三个代表”重要思想，紧紧依靠全县人民，采取强有力措施，努力克服“非典”疫情及自然灾害带来的不利影响。以全面建设小康社会统揽全局，坚持发展不动摇，咬住目标不放松，团结奋进，顽强拼搏，崇尚实干，狠抓落实，圆满完成了县第

十五届一次人代会确定的目标任务，改革开放和经济社会发展取得了有目共睹的成绩，实现了本届政府各项工作的良好起步和精彩开局。全县生产总值达到 8.97 亿元，比上年增长 15.5%（可比价增长 9.68%），增速达到了近几年以来的最高水平，其中：第一产业增加值 3.49 亿元，增长 19.8%（可比价增长 8.41%）；第二产业增加值 2.02 亿元，增长 23.5%（可比价增长 19.74%）；第三产业增加值 3.46 亿元，增长 7.6%（可比价增长 6.05%）。一、二、三次产业之间的比例关系由上年的 37.5 : 21.1 : 41.4。调整到 2003 年的 38.9 : 22.5 : 38.6。人均生产总值接近万元大关，达到 9645 元。全县城镇居民人均可支配收入达 6390 元，比上年增加 480 元，增长 8.1%。人均消费性支出 5516.81 元，下降 1.5%。城镇居民家庭恩格尔系数为 30%。农民人均纯收入 4046 元，比上年增加 330 元，增长 8.9%。人均生活消费支出 2439 元，增长 14.1%。农村居民家庭恩格尔系数 27%，下降 4.8 个百分点。全县综合经济实力有了进一步增强。

### 一、教育和科技

教育事业稳步发展。中小学入学率达到了 100%，年巩固率达到了 98.4%；高中阶段普及率达到了 90% 以上；正规国民教育上线率达到了 82.5%，提高了 1.45 个百分点。完成了安西中学教学大楼修建和 5 所农村中小学 3000 平方米的危房改造工程。农村教师工资上划县财政统一发放。欠发工资的问题逐步得到解决。县中学后勤管理社会化改革迈出了成功的一步，教育后勤管理体制得到创新。职教、普教实行了合并办学，实现了教育资源的优化配置。

科技队伍基本稳定，科技投入明显增加，科技成果

转化成效明显。年末全县共有各类专业技术人员 1475 人。工业企业从事科研活动人员 132 人,其中工程师 28 人。全县各级财政投入科技三项经费 20 万元,增长 100%。全年共取得应用技术成果 15 项,受理专利申请 1 件。

## 二、文化、广播、电影、电视和体育事业

文化艺术、广播电视、新闻出版等各项事业健康发展。年末全县共有艺术表演团体 3 个,文化馆 1 个,公共图书馆 1 个,博物馆 1 个。广播发射台和转播台 2 座,广播综合人口覆盖率达到 100%;电视发射台和转播台 2 座,每千人广播电视播出时间 21 小时/周,电视综合人口覆盖率达到 78%。全县广播电视光纤传输网络达到 112.8 公里,有线电视用户 1.16 万户,有线电视入户率达到 68.9%。体育事业蓬勃发展,全民健身运动广泛开展。全年共举办各种体育竞赛活动 78 场次,参加人员达 3.5 万人次,全县共有 90%的在校学生达到《国家体育锻炼标准》。

## 三、卫生事业

抗击“非典”斗争取得阶段性胜利。认真落实预防“非典”责任制,全面开展群防群控工作,实现了“非典”疫情的零输入目标。进一步完善了县乡疾病预防控制网络,增强了全民疾病预防控制和应对突发性公共卫生事件的能力。卫生事业规范运行。年末全县共有卫生机构 38 个,其中医院 1 所。卫生机构床位数 305 张,其中医院 120 张。全县卫生机构人员 479 人,其中卫生技术人员 352 人,注册护士 115 人。卫生防疫防治机构 1 个,卫生技术人员 13 人。妇幼卫生机构 1 个,卫生技术人员 10 人。全县共有乡镇卫生院 12 所,床位数 145 张,卫生技术人员 197 人。农村有医疗点的村占总村数的比

重达 79.5%，有乡村医生和卫生员 51 人。新型农村合作医疗制度已初步建立，农户参保率达到 63.3%。

#### 四、政法工作

政法工作力度进一步加大。全县各级以维护社会稳定为首要任务。自 2000 年，开通“148”法律咨询电话专线后，为全县经济发展和社会稳定发挥了保驾护航作用。公、检、法部门继续以抓黄、赌、毒为重点，严厉打击各种刑事犯罪。2003 年各类刑事案件立案数 280 起，破案 187 起，破案率达 66.79%；发生各类治安案件 287 起，查处 249 起，治安案件查处率达 86.76%。2003 年发生火灾 35 起，造成直接经济损失 203 万元，交通事故 146 起，造成直接经济损失 23.31 万元。

#### 五、社会保障

社会保障体系进一步完善。年末全县参加基本养老保险职工 2209 人，综合覆盖率 70%，社会化发放率 100%；有 2393 名职工参加失业保险，综合覆盖率 80%；有 3452 名职工参加基本医疗保险。“两个确保”基本实现，全年分别发放企业离退休人员基本养老金 344 万元、失业保险金 68.2 万元，支付医疗费 235 万元。全县得到最低生活保障救济人数 1714 人，比上年增长 19.5%；农村社会养老保险人数 2030 人，下降 32.3%。救济工作成效显著，共发放救灾款 31 万元，救济灾民 3123 人。重视关心孤寡残疾人员、贫困人口等弱势群体的生活，贫困人口生产生活条件得到改善。全县除个别新的移民安置点外，全部实现了通路、通电、通电话、通广播电视、通自来水。人民居住条件明显改善，生活质量明显提高。

#### 六、社会团体、干部情况

2003 年全县中共党员 5448 人，比上年增加 85 人，

增长 1.58%；共青团员达 2650 人，比上年减少 276 人，减少 9.43%。工会基层组织有 170 个，建立基层工会的单位职工人数达到 9603 人，比上年同期减少 42 人，减 0.44%。全县人大代表 136 人，其中女代表 29 人。人大代表议案数 2 件，人大落实议案数 2 件。人大建议批评建议数 41 件，人大落实批评建议数 41 件。政协委员 79 人，其中女委员 17 人。政协委员提案数 42 件，政协落实提案数 42 件。政协常委会建议案 1 件。2003 年我县机关单位干部总数 562 人，比上年增加 18 人，增长 3.31%。其中女干部 113 人，比上年增加 11 人，增长 10.78%。事业单位干部总数 1349 人。其中女干部 453 人。

### 七、环境保护

环境保护事业取得新进展。年末全县有自然保护区 2 个，其中国家级自然保护区 1 个，省级自然保护区 1 个，自然保护区总面积 112.42 万公顷，占全县国土总面积的 46.5%。生态示范点及生态示范工程进展顺利，共建成生态示范点 4 个，示范工程项目 3 个。环保投入进一步加大，人居环境继续改善。全县环保投资 1423.25 万元，占 GDP 比重达 1.65%。城市绿化覆盖率达到 21.01%，饮用水达标率 100%，垃圾定点堆放、集中处理率达到 100%。

### 八、问题与对策

根据以上对我县社会、经济与科技等发展情况分析，我认为我县目前社会发展领域的现状是既面临挑战，又存在历史机遇。目前我县经济和社会生活中还存在一些突出的困难和问题。我县社会发展的总体水平还不高，特别是在社会转型时期，一方面历史形成的一些困难和

问题还没有得到彻底解决，同时旧体制下长期隐藏的诸多社会问题又呈现显性化的趋势，而且随着经济体制改革的深化，经济体制改革过程中的各种矛盾，也必然会在社会领域突出表现出来。如社会保障体系还不够健全完善，就业压力增大、隐性失业问题的显性化；产业结构性矛盾和财政收支矛盾比较突出；资源制约性较大，工农业主导产品不够明显，相对优势减弱；企业技术创新能力及适应市场能力还不高，市场竞争能力不强；总体经济运行质量不高，效益回升缓慢；社会治安问题，人口素质偏低与经济社会发展的要求不相适应的问题，以及科技成果转化率低、社会保障体系尚不完善等等，这些问题都是今后我县经济建设和社会发展过程中急需解决的主要问题。

为更好地促进我县社会经济与科技事业的全面发展，我县应做好以下几个方面工作：

- 调整产业结构，提高国民经济整体的效益水平和产品的市场竞争力；
- 确立就业优先的战略，千方百计解决下岗职工就业问题；
- 强化措施，提高人口素质；
- 创造良好的教育、科技环境，加大教育、科技的投入力度，确保财政对科教投入的增长速度高于财政收入的增长速度；
- 继续强化九年制义务教育，全面推行素质教育，培养创造性人才；
- 积极扶持科技风险投资，促进科技成果的转化；
- 建立健全多层次、全方位的社会保障制度；
- 开展消费信贷，促进消费方式的改变，引导城乡居民实现消费升级；
- 加大环保投入，保护生态环境；
- 加强精神文明建设，净化社会环境。

2004年机遇和挑战并存，希望和困难同在，我们一定要正视困难和问题，充满信心，充分调动和发挥全县人民的积极性和创造性，认真研究，加以解

决。2004 年全县经济社会发展的任务依然非常艰巨，我们要高举改革发展的旗帜，全面贯彻“三个代表”重要思想，按照“马上办，盯住干”抓落实的总要求，在县委的正确领导下，把广大干部群众的思想 and 行动统一到既定目标任务上来，把全县上下各方面的力量凝聚到全面建设小康社会的伟大事业上来，与时俱进，开拓创新，同心同德，崇尚实干，为把安西的现代化建设推向新阶段而努力奋斗。

安西县统计局舒芙蓉

二 00 四年五月十三日

## 《绍兴晚报》读者调查问卷的统计分析报告

### A . 样本分析

性别：本次统计中男性有 8 9 5 人，女性有 6 1 3 人，男女比例接近 6 : 4，基本符合统计分析需要。

年龄：统计显示，年龄段分布比较平均，22 周岁以上 60 周岁以下的读者超过 60%，而 60 周岁以上的读者不到 20%。有老龄化倾向，但不明显。

职业：统计显示，读者职业分布总体比较分散，符合统计需要。其中公务员占 4%，事业单位工作人员占 16%，企业管理人员占 11%，普通职员占 23%，商务人员占 8%，学生占 10%，离退休和其他占 28%。

### B . 要素分析

读者文化程度：本次统计中，文化程度分布主要集中在高中和中专、大专文化，占一半以上(56%)，而大学及以上学历仅占 10%。和我市市民平均文化程度

对比看,这样的读者文化程度说明晚报的读者群比较广,而且说明晚报是一张适合普通市民阅读并深入市民的报纸。

读者报纸来源:本次统计表明,1608人中有984人是个人自费订户,占61%,还有12%的读者在报摊购买。主动读者、自费读者的比例非常大。另外,单位订阅的为20%左右,另有10%的读者来源为“借阅或赠阅”。

### C. 比较分析

#### 版面数量和阅读时间

在调查中,有42%的读者认为现在平均每天16版的规模很好,更少数的读者认为要有比32版更多的版面。

不少读者在来信中表达了对报纸版面多少的看法,大多数读者都认为版面多少并不是评判报纸好坏的主要标准。他们关心的不是扩版与否,而是信息量是否能满足他们的需要。来信中将近一半的读者认为像杭城一些厚报版面过多,其中广告和垃圾信息又占据不少版面,阅读很累。可见,版面多并不是读者看重的,新闻质量、报纸的信息量更为重要。

对于阅读时间,有24%的读者阅读一份晚报的时间在1个小时以上,而阅读时间超过半个小时的更是占三分之二强(68%)。作为绍兴本地最强势的平面媒体,其受关注程度由此可见一斑。

关于看报时间,我们得到的数字是:三分之二(68%)的读者在晚饭前后阅读,反映出晚报在大多数读者心目中并不是一份信息“快餐”,而是一份每天都得品尝的信息“主食”。面对厚报的竞争,晚报在质量上

任重而道远。

#### 收入状况和报价定位

在本次参与调查的人群中，读者的平均月收入在5000~10000元、10000~15000元以及15000~20000元这三档的居多，总共占将近三分之二；仅59人（4%）填写30000元以上。从这个读者收入状况看，晚报缺少高收入阶层的读者，这多少是对晚报品位的质疑。

不过，不同收入状况的读者对报纸价格的看法并没有分歧。调查中，近三分之二（65%）的读者对晚报0.5元/份的价格表示“满意”或“较满意”，感到无所谓的占26%，感到不满意的不到10%。

#### D. 重点分析

##### 最先关注和非常喜欢

调查显示，读者翻开报纸最先关注的四类版面（按排名先后顺序）：社会新闻、国内要闻、深度报道、本地要闻。

读者非常喜欢的四类版面也已“出炉”，它们分别是（按排名先后顺序）：社会新闻、纪实和亲历、本地要闻和国内要闻。

这个选择在读者的来信中有更翔实的解释。比如：本地社会新闻贴近性强，有的就发生在身边，影响较大；看本地要闻是想了解绍兴的一些重要政策、动向，对老百姓生活也很重要；记者亲历的报道优势也很明显，可读性强。

值得一提的是，尽管“消费生活”和“生活前沿”尚未进入读者非常喜欢的版面行列，但已经有将近一半的读者选择了它们，体现出读者对财经新闻和时尚信息

的关注度也不低。

### 最重要和最吸引

本次调查中的读者“最重要”和“最吸引”的新闻观，体现了我们读者的“实用主义”的一面，某种程度上也指明了我们报纸的发展方向。

参与调查的读者认为，最重要的新闻报道是报道的信息量和报道的深度（这两项比例相差甚微）。在信息爆炸时代，信息量是报道有价值的基础，而深度则体现报纸对信息的理解力，也能张扬一张报纸的质量和特色。

参与调查的读者中有65%认为，“有发生在身边的新闻”是一份报纸最吸引人的要素，这是读者贴近性的体现。事实上，身边的新闻即“本埠新闻”正是晚报要不断做大做强的，做大做强本埠新闻也完全符合晚报目前的定位。另外，读者对信息量和时效性也比较关注，相比之下，版面的冲击力和图片的质量并不是读者重视的。

### 社会新闻最吸引眼球

社会新闻是晚报的传统强项，如今注重贴近、时效、独家、深度的社会新闻版依然深受读者喜爱。“通常重点关注哪些类别的内容”、“最喜欢哪些版面”和“希望增加哪些内容的报道力度”这三大调查问题中，“社会新闻”这一选项全部脱颖而出，统计数字高居榜首。

### 四大标准均得认可

调查数据显示，晚报的“报道及时性”、“客观公正性”、“权威性”和“舆论监督力度”的读者满意度均超过90%。这四个评判标准得到高度认可，无疑是报纸质量和品位的体现。

## 全省中小学校图书馆(室)建设综合统计分析报告

### (2001年度)

随着素质教育的不断深入,我省中小学校图书馆(室)建设有了进一步的发展。根据我厅《关于填报中小学校图书馆(室)建设统计表的通知》(粤教装备[2001]10号)文件精神,我们对全省中小学校图书馆建设工作进行了一次全面统计。现简要分析报告如下。

#### 一:基本情况

1:从报表提供的数据得到,目前我省中小学校图书馆(室)建筑面积合计 1277709.1m<sup>2</sup>,平均面积为 47.73 (m<sup>2</sup>/所)。全省拥有计算机管理的学校图书馆有 825 所,全省藏书室建筑面积 874488m<sup>2</sup>(其中教师阅览室面积 285820m<sup>2</sup>,学生阅览室面积 905164m<sup>2</sup>)。全省学校数与全省教师阅览室和学生阅览室的比例约为 4:1:2 (26772:7252:16393)。

2:全省中小学校的藏书量合计 18809.3 万册(其中含工具书类、教学参考书、学生用书、通用图书、报刊、非书资料等)。全省中小學生均图书数为 13.95(册/生)。其中,省一级学校生均图书数为 30.40(册/生),高级完中生均图书数为 20.47(册/生),初级中学生均图书数为 16.31(册/生),小学生均图书数为 11.04(册/生)。

3:全省中小学校现有图书价值 97575.07 万元。当年图书购置经费合计 15304.95 万元,其中由上级拨款购置资金 3529.7 万元,自筹购置资金 11775.25 万元。全省尚需图书购置资金 88012.77 万元,图书尚需数量 17456.76 万册。

4:全省中小学校图书馆管理人员 24093 人,其中专职人员 4465 人,兼职人员 19628。从职称情况看,拥有高级职称有 596 人,中级职称 4163 人,初级职称 10547 人。从学历情况看,具有大专以上学历的管理人员有 4985 人,中专学历的有 12345 人,其他的有 6071 人。

## 二、综合分析结果

依据以上提供的统计数据,经过综合分析我们可以得出目前我省图书馆(室)建设的一些基本情况。

### 一、取得的主要成绩:

1:我省中小学校图书馆(室)建设规模逐步发展:全省中小学校有 26772 所,图书馆建设面积为 1277709.1m<sup>2</sup>,

实行计算机管理的学校有 825 所(其中省一级学校 177 所,高级完中 188 所,初级中学 226,小学 234 所)比去年同期增加 160 所,增长率达 24%。全省中小学校藏书 18809.3 万册,比去年增加 3382.3 万册,增长 22%。阅览室(含教师阅览室和学生阅览室)全省合计 23645 个,比去年增加 828 个,增长率 12%。全省生均图书数 13.95(册/生),比去年增加 2.11(册/生),增长 17%(其中省一级学校增加 5.38(册/生)高级中学增加 1.81(册/生),初级中学增加 2.62(册/生),小学增加 1.64(册/生)。

### 2:图书馆的经费投入逐年增加:

全省中小学校现有图书价值合计 97575.07 万元,比去年增加 10478.07 万元,增长了 12%。当年图书购置经费达 15304.95 万元,比去年同期增加 3381.95 万元(其中自筹资金 11775.25 万元,比去年增加 3551.25 万元)。

3:图书馆馆藏结构有所改善,馆藏质量不断提高。

全省中小学校藏书室有 20877 个(其中省一级学校有 537 个,高级完中有 1000 个,初级中学有 2591 个,小学 16749 个)。藏书量有 18809.3 万册,(其中工具书类有 1189 万册,教学参考书类有 3001 万册,学生用书类有 6032 万册,通用图书类有 6096 万册,报刊类有 600 万册,非书资料类 1891.3 万册)。六类图书占藏书量的比例分别是 6% : 16% : 32% : 32% : 3% : 10%。

4 :图书馆管理人员的总体水平不断提高,队伍建设不断加强。全省中小学校现有管理人员 24093 人,比去年增加 692 人,其中专职管理人员 4465 人,增加 304 人,兼职管理人员 19628 人,增加 388 人。从现有职称人员的情况来看,具有高级职称 596 人,中级职称 4163 人,初级职称 10547 人,比去年分别增加 131 人,503 人和 794 人。从学历情况看,具有大专以上学历的管理人员有 4985 人,中专 12345 人,其他的 6071 人。从以上统计数字得到,我省中小学校图书馆(室)管理人员的总体水平在不断提高。管理人员的队伍建设不断得到加强和充实。

### 三、存在问题

1 :图书馆建设规模偏小,藏书室数量不足。

全省图书馆建设面积 1277709.1m<sup>2</sup>,平均每所学校的图书馆建

设面积为 47.72m<sup>2</sup>(其中省一级学校为 612.82m<sup>2</sup>,高级中学为 290.71m<sup>2</sup>),初级中学为 87.17m<sup>2</sup>,小学为 25.29m<sup>2</sup>),现有建筑面积较小,规模有待扩大。全省藏书室 20887 个,面积 874488m<sup>2</sup>,平均每所学校藏书室有 0.78 个,藏书室的数量不足。

全省拥有计算机管理的中小学校有 825 所,省一级

学校为 177 所,高级中学为 188 所,初级中学为 226 所,小学为 234 所。

2:藏书量不足,图书缺口数量大。

全省中小学校现有藏书 18809.3 万册,生均图书数为 13.95(册/生),其中省一级学校 30.40(册/生),高级完中 20.47(册/生),初级中学 16.31(册/生),小学 11.04(册/生)。依据《广东省中小学图书馆(室)建设规程》(粤教装备[2000]11 号)第十三条的规定,各类学校的生均图书数分别是省一级学校 40—60(册/生),高级完中 40(册/生),初级中学 30(册/生),小学 20(册/生)。全省合计相差 26.05(册/生)。其中省一级学校相差 19.6(册/生),高级完中相差 19.53(册/生),初级中学相差 13.69(册/生),小学相差 8.96(册/生)。

全省尚需购置图书数量 17456.75 万册(其中省一级学校 833.49 万册,高级中学 1481.8 万册,初级中学 2685.72 万册,小学 12455.75 万册)。

3:图书经费投入不足。

全省图书购置经费尚需资金较多,全省合计共需资金 88012.77 万元(其中省一级学校尚需 5922.34 万元,高级中学尚需 12779.85 万元,初级中学尚需 20862.19 万元,小学尚需 48449.39 万元)。

4:图书馆现有专业管理人员短缺,培训工作有待加强。

(1) 据统计我省中小学校现有管理人员 24093 人,其中专职管理人员 4465 人。兼职管理人员 19628 人,专职人员与兼职人员的比例约为 1:4,专职人员少,兼职人员多。

(2) 管理人员专业管理水平和学历不高,具有大专

以上学历的管理人员有 4985 人,中专学历的有 12345 人。全省现有图书馆管理人员中,具有中级以上职称的管理人员占总数人员的比例不高,全省比例约为 1:5(其中重点中学 1:3,高级完中 1:4,初级中学 1:4,小学 1:5)。

#### 四、问题的原因及今后努力方向。

从综合统计的数据分析,了解到目前我省中小学校图书馆(室)建设工作中存在的一些问题及各方因素。

##### 1: 关系没有理顺,管理工作没有明确。

目前,我省还有一些地方没有明确管理职能部门,图书馆管理工作尚未归口到装备部门。到目前为止,还有一些地方如:广州市、中山市、东莞市、江门市图书馆建设工作仍未按规定实行归口管理。造成管理上的混乱,使管理工作难以到位,影响了事业的向前发展。

##### 2: 认识不足,投入不够。

目前,还有许多地方部门没有充分认识到图书馆在学校教育教学中的重要作用,把图书馆作为一个可有可无的机构,因而个别地方领导没有把图书馆建设管理工作提到日常工作日程上,对图书馆建设工作不重视,对图书馆的投入非常有限,从而出现图书馆建设规模较小,藏书量不足,专职管理人员缺少等等问题。

##### 3: 管理队伍人员短缺,总体素质不高。

长期以来,图书馆管理人员的职称,待遇问题得不到重视,没有把它归入教师职称系列的评聘,因此,出现了管理人员专职的少,兼职多的现象。另对现有管理人员的专业业务知识培训和再教育工作跟不上,从而造成管理人员总体素质不高的现象。

鉴于以上的问题,我们应采取一些有效的措施。

1: 理顺关系、提高认识,按照我厅《规程》把图书馆管理职能归口到装备部门,实行统一管理。充分认识图书馆工作在现代教育教学中的重要作用。在中小学教育中,学校的图书馆将逐步发展成为教育事业的基础,成为教育事业的一个重要系统。

2: 加大对图书馆的投入,中小学校图书馆建设作为基础教育的一个重要组成部分,各教育行政部门、学校应加大对图书馆的经费投入,从每年的教育经费预算中安排图书馆专项经费,并逐年有所增加。学校在学校教育事业费中应不小于3%的资金用于购置图书。

3: 培养一支高素质的中小学图书馆管理队伍。要对现有管理人员组织多种形式的培训。各级教育行政部门要贯彻落实好有关政策,解决好图书馆人员的编制、待遇问题,在各方面应与一线教师同等待遇,加强对管理人员的培养工作,提高管理人员的总体综合素养。

广东省教育装备中心

二〇〇二年八月日