

数理统计与应用概率

Mathematical Statistics and Applied Probability

1988

第三卷 第三期

学术期刊出版社

目 录

应用成果

细菌性痢疾的计量辅助诊断·····	申鼎煊 吴炳凤	(253)
城市用水量增长模型及其应用效果分析·····	金云程	(260)
用时间序列分析法研究齿轮系(箱)故障诊断·····	戴悟僧 安邦健	(268)

理论与方法

单向关闭的两部件串联可修系统的可靠性分析·····	何道杰	(274)
关于三维表的两个非谱系对数线性模型·····	安希忠 夏立显	(280)
均方误差的无偏估计与回归系数的有偏估计·····	察可文	(293)
两个变量的线性函数关系模型参数的一种新的估计方法——聚类估计····· ·····	吴可法 董天信	(301)
具有随机回归系数的线性试验的比较·····	陈建宝	(310)
非参数回归函数的混合型核估计的强收敛性·····	杜雪樵	(322)
每个需求是随机量且定货量依赖于状态的(s, S)库存系统·····	孟玉珂	(328)
非齐次马尔柯夫过程的可逆性·····	来向荣 程维虎	(337)
两类非时齐Poisson过程模型的参数估计·····	刘鸿翔	(341)
利用EM算法分析时间序列数据的影响·····	钱 卫	(348)
指数分布场合下恒定应力加速寿命试验的Bayes估计·····	陈 平	(353)

探索与争鸣

SFC(空间填充曲线)变换和它在判别分析中的应用·····	陈忠珽 杨振海	(360)
-------------------------------	---------	---------

综述

模型选择技巧在多元分析中的一些应用·····	赵林城	(372)
------------------------	-----	---------

CONTENTS

Applications

- Quantity Aid Diagnosis of Bacillary Dysentery.....
.....Sheng Dingxuan Wu Binfeng (253)
- The Model of Water Supply Growth in the city and the Analysis
of Its Effectiveness..... Jin Yun cheng (260)
- Using Time Series Analysis Method to Diagnose the Malfunction
of Gear Box.....Dai Wusan An Bangjian (268)

Theory and Method

- Reliability Analysis of Two-unit Series System with One
Direction Shout-off Rule.....He Daojie (274)
- On the Two Non-hierarchical Log linear Models in Three
Dimensions.....An Xizhong Xia Lixian (280)
- The Mean Squared Error Unbiased Estimate and Regression
Coefficient Baised Estimaate..... Cha Kewen (293)
- A New Method of Parameter Estimation for Two Variables Linear
Functional Relationship Model—Clustr Estimation.....
..... Wu Kefa Dong Tianxin (301)
- Comparison of Linear Experiments with Stochastic Regression
Coefficients.....Chen Jianbao (310)
- The strong Convergence of A Mixed-Type Kernel Estimate of
Nonparametric Regression Function..... Du Xueqiao (322)
- Inventory Systems With Stochastic Demand Quantity and
Order Quantity Dependent States..... Meng Yuke (328)
- The Reversiblity of Denumerable and Nonhomogeneous
Markov Processes..... Lai Xiang rong Chen Weihu (337)
- The Parameter Estimate of Two Kinds of Nonhomogeneous
Poisson Process Model.....Liu Hongxiang (341)
- Diagnostics for Time Serits VIA EM Algorithm.....Qian Wei (348)
- Bayes Estimation in Accelerated Life Testing—Constant strees Models
Vnder the Exponentiand Distribution case Chen Ping (353)

Discussion and Exploration

- SFC Transform and Its Application to Discriminant Analysis
..... Chen, Zhong lian, Yang, Zhenhai (360)

Summary

- Some Applications of the Model Selection Techniques to
Multivariate Statistics.....Zhao Lincheng (372)

细菌性痢疾的计量辅助诊断

申鼎煊 吳炳凤

(华中理工大学)

摘 要

本文用两类线性判别分析,逐步判别分析和聚类分析等多元分析方法,对细菌性痢疾作计量辅助诊断。所得结果与临床诊断基本相符。

细菌性痢疾的诊断,从来都是医生根据病人的主要临床表现,如畏寒、发热、腹痛、腹泻、脓血便和里急后重等症状,医生的知识水平和经验相结合作出决定的。本文应用现代医学发展的一个重要方向——计量医学(参阅文献[5])的观点,用多元分析的理论和方法,结合微机计算,建立数学模型,对细菌性痢疾作辅助诊断,效果很好。本文的全部计算在IBM—PC机上进行,计算机程序参阅[1]。

一、资 料

菌痢患者,由于病菌不同和人体免疫情况有异,临床症状表现的轻重缓急及病情的演变很不一致。除急性菌痢外,急性非典型菌痢和慢性菌痢的诊断有时比较困难。因为这时患者的体征变化表现不明显。医生确诊有赖于实验室的进一步检查。而这些检查的结果往往是数量化的或可数量化的,它们必然存在数量的统计规律性。为了建立辅助诊断数学模型,收集了我院医院1985年——1986年75例菌痢住院病人病历记录,整理出临床诊断和实验检查有关数据资料如表1形式:

表1 数据资料

样号	体温℃ (x_1)	WBC/ mm ³ (x_2)	脓球(x_3)	红细胞 (x_4)	样号	体温℃ (x_1)	WBC/ mm ³ (x_2)	脓球(x_3)	红细胞 (x_4)
1	38.5	12000	++++	++	9	39.5	13700	+++	++++
2	38.5	7700	+++	+	10	38.5	20100	++	+++
3	39	10500	++++	++++	11	38.1	9400	++	++++
4	39	13300	++	+	12	37	6200	++++	++++
5	39.5	13400	+++	++++	13	39.6	7000	++++	+
6	39.5	16200	++++	++++	14	38.5	8700	++	++
7	39.5	7000	++++	++++	15	39	7600	+++	+
8	39.3	9700	++	+++	16	37.6	10100	++++	++++

续上表

17	39.5	9000	+++	+++	47	36.8	3800	+	-
18	37.2	13000	++	+++	48	36.4	4200	+	-
19	37	8600	++++	++++	49	36.2	5100	-	-
20	38	9100	++	+++	50	37.3	8000	+	-
21	38.3	7600	+++	++++	51	36	6900	-	+
22	39	11000	+++	++	52	36.5	7000	-	-
23	38.5	9300	++	+++	53	36.2	6900	-	-
24	39.2	8200	++++	++++	54	36.5	5300	+	-
25	37.7	9800	++	++++	55	36.3	6100	+	-
26	38.3	7900	++++	+++	56	36	5000	-	+
27	38	8900	+++	+++	57	36.1	4800	+	-
28	37.5	10500	+++	++++	58	36.6	5200	-	-
29	39.2	6900	++++	+++	59	36.2	5100	-	-
30	37	8500	++	+	60	36.3	5000	-	+
31	38.5	9400	+++	+++	61	36.7	4700	+	-
32	36.2	8650	++	-	62	36.5	4300	-	-
33	38.3	7430	+	++	63	36	5000	+	+
34	39	9000	++++	++++	64	36.9	5300	+	-
35	37.7	6700	-	+	65	36.3	6000	-	-
36	36.2	6800	-	-	66	36.1	4900	-	-
37	36	3900	-	-	67	36	6000	+	-
38	36.1	4500	-	-	68	36.3	6940	-	-
39	36.2	5800	-	-	69	37	4600	-	+
40	36.5	6000	-	-	70	36.6	6100	-	+
41	36.3	5100	+	-	71	37.3	4100	++++	++++
42	36	4000	-	-	72	38.4	7000	+	++
43	36.6	4000	+	-	73	38.5	10800	+	+
44	36.2	5900	-	-	74	37	7200	+++	++++
45	36.1	9700	-	-	75	36.9	7100	+	+
46	36.8	5000	-	-					

表 1 中样号为病人代号，其中 1—35 号称为 A 组（或 A 类），即医生诊断为菌痢患者；36—70 号称为 B 组（或 B 类），即病愈将出院前夕的患者；71—75 是待判组，他们是刚入院或正在住院的患者。 x_1, x_2, x_3, x_4 各列的值，对于 A 组是刚入院时的检查记录，B 组则是病愈患者出院前一、二天内的记录。为了便于数量化处理，计算时将表 1 中的一，+，++，+++，++++ 分别赋值为 0，1，2，3，4。

对于各项观测指标，在假定它们是具有方差相同的正态总体样本的条件下，我们首先对

此为试读，需要完整PDF请访问：www.ertongbook.com

A 组总体与 B 组总体均值是否有显著差异进行了统计检验。即对各项指标, 检验假设 $H_0: u_j^{(A)} = u_j^{(B)}$ 。

设 $x_1, \dots, x_{n_1}$ 和 $y_1, \dots, y_{n_2}$ 分别是来自 A, B 两总体的样本, 则检验假设 H_0 的统计量为:

$$T = \frac{DT}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

其中 $DT = (\bar{x} - \bar{y}) / \sqrt{\frac{(n_1 - 1)S_x^2 + (n_2 - 1)S_y^2}{n_1 + n_2 - 2}}$, $S_x^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_i - \bar{x})^2$,

$$S_y^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (y_i - \bar{y})^2, \quad \bar{x} = \frac{1}{n_1} \sum_{i=1}^{n_1} x_i, \quad \bar{y} = \frac{1}{n_2} \sum_{i=1}^{n_2} y_i.$$

微经机计算得 T 值如下:

体温 $t = 12.463$, 白细胞总数 $t = 13.311$,

脓球 $t = 7.972$, 红细胞 $t = 12.935$

查表得 $t_{0.01} \approx 2.65$ 。所有计算的值都大于 $t_{0.01}$, 说明在水平 $\alpha = 0.01$ 上, A, B 两组各项指标的均值都有显著差异, 即住院病人经治疗后确有显著好转。

二、两类线性判别分析

1. 两类判别及其计算结果

设有 p 个判别指标 $x_1, \dots, x_p$, 考虑建立如下形式的两类线性判别函数

$$y = c_1 x_1 + \dots + c_p x_p$$

其中 c_i 为待定判别系数。

设有 A, B 两组样本观测值:

$$x_{(\alpha)}(A) = (x_{\alpha 1}(A), \dots, x_{\alpha p}(A))', \quad \alpha = 1, \dots, n_1$$

$$x_{(\alpha)}(B) = (x_{\alpha 1}(B), \dots, x_{\alpha p}(B))', \quad \alpha = 1, \dots, n_2$$

根据Fisher准则, 可得两类判别临界值为(参阅[1]或[2])

$$y_c = \frac{n_1 \bar{y}(A) + n_2 \bar{y}(B)}{n_1 + n_2}$$

其中 $\bar{y}(A) = \sum_{i=1}^p c_i \bar{x}_i(A)$, $\bar{y}(B) = \sum_{i=1}^p c_i \bar{x}_i(B)$,

$$\bar{x}_i(A) = \frac{1}{n_1} \sum_{\alpha=1}^{n_1} x_{\alpha i}(A), \quad \bar{x}_i(B) = \frac{1}{n_2} \sum_{\alpha=1}^{n_2} x_{\alpha i}(B). \text{ 于是得判别规则为:}$$

1° 当 $\bar{y}(A) > \bar{y}(B)$ 时, 对任一样品 x , 若计算得判别函数值 $y > y_c$, 则判 $x \in A$, 反之, 判 $x \in B$ 。

2° 当 $\bar{y}(A) < \bar{y}(B)$ 时, 若计算得 $y > y_c$, 则判 $x \in B$, 反之, 判 $x \in A$ 。

p 个变量分辨两类总体的能力的显著性检验,采用统计量

$$F = \frac{n_1 n_2 (n_1 + n_2 - p - 1)}{p (n_1 + n_2) (n_1 + n_2 - 2)} D^2 \sim F(p, n_1 + n_2 - p - 1) \quad (*)$$

其中 $D^2 = (n_1 + n_2 - 2) \sum_{i=1}^p \sum_{j=1}^p d_i d_j w^{ij}$, w^{ij} 是矩阵 $W = (w_{ij})$ 的逆矩阵 $W^{-1} = (w^{ij})$

上的元素。 $w_{ij} = \sum_{\alpha=1}^{n_1} (x_{\alpha i}(A) - \bar{x}_i(A))(x_{\alpha j}(A) - \bar{x}_j(A)) + \sum_{\alpha=1}^{n_2} (x_{\alpha i}(B) - \bar{x}_i(B))$

$(x_{\alpha j}(B) - \bar{x}_j(B))$, $d_i = \bar{x}_i(A) - \bar{x}_i(B)$ 。

利用表1的数据,经微机计算得

判别函数 $y = 4.381851 \times 10^{-2} x_1 + 9.691201 \times 10^{-6} x_2 + 3.846263 \times 10^{-2} x_3 + 2.814448$
 $\times 10^{-2} x_4$

判别临界值 $y_0 = 1.817279$ (判别结果略)。

2. 判别效果的评价

1° 统计检验

经微机计算得

$$D^2 = 20.71989$$

$$F = 4.951443 > F_{0.01}(4, 65) \approx 3.64$$

故在水平 $\alpha = 0.01$ 上, A, B 两组均值向量差异显著,判别效果较好。

2° 符合率

A 组与 B 组判别符合情况如表2

表2

原分组 判别分组	A 组	B 组	小 计
A 组	32	0	32
B 组	3	35	38
小 计	35	35	70

判别分组的结果与临床诊断的结果,经计算知只在样号30, 32和35三例有差别。假阴性错误率为 $\frac{3}{35} = 8.31\%$, 符合率为 $\frac{67}{70} = 95.7\%$ 。所有的计算和检验结果都说明,用体温、白细胞总数、便检脓球量和红细胞量构成的两类线性判别函数诊断菌痢患者和治愈正常者的判别效果良好。

三、逐步判别分析

1. 逐步判别及其计算结果

为了比较和探讨,我们进一步作了Bayes 准则多类逐步判别分析。通过逐个地、“有进有出”地选取判别能力显著的变量,最终可建立判别 m 类总体的 m 个判别函数

$$y_j(x) = C_{0j} + \sum_i C_{ij} x_i, \quad j=1, \dots, m$$

分辨两两总体的能力的显著性检验统计量 $F_{.f}$ 类似于(*)式(参阅[1]或[3])。

由表1的资料,取引入和剔除变量的检验临界值 $F_1 = F_2 = 5$,经逐步计算得如下结果:

表3 各类均值与总均值向量

项 目		x_1	x_2	x_3	x_4
分类均值	1	38.42	9773.714	2.885714	2.857143
	2	36.36572	5512.572	0.3428572	0.171486
总 均 值		37.39286	7643.143	1.614286	1.514286

表4 判别系数

类号	C_0	C_1	C_2	C_3	C_4
1	-1777.371	93.96389	0	-18.95586	-0.2272116
2	-1642.519	90.54472	0	-21.17866	-2.480433

待判样品71—75号判别的结果为:

71—74号判为1类(即A组)

75号判为2类(即B组)。

2. 判别效果评价

1° 统计检验

经计算得

$$D_{12}^2 = 18.72773$$

$$F_{12} = 106.032 > F_{0.01}(3, 66) \approx 4.12$$

故在水平 $\alpha=0.01$ 上,两类总体均值差异显著。

2° 符合率与两类线性判别相同,而且都是在样号30,32和35发生误判。

由逐步判别结果可知(篇幅所限,略),判别的后验概率除样号30和71以外,其余都在0.95以上,说明错判的可能性及损失较小。从表4看出,判别系数 C_{21} C_{22} 为0,即变量 x_2 未被选入判别函数中。这时判别函数虽然只含有三个指标:体温、脓球和红细胞,但其误判情况和判别符合率却与两类线性判别一致,判别效果同样是良好的。

四、聚类分析

在撇开临床诊断分组的情况下,我们对表1前70个样号,用4个指标的观测值,按照他们的相似程度或亲疏程度,进行了样品的系统聚类分析。为了节省篇幅,有关数学原理可参阅[1]—[3]。

聚类分析情况如下:

1° 对观测数据标准化

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{S_j}, \quad i=1, \dots, n, \quad j=1, \dots, p$$

$$\text{其中 } \bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}, \quad S_j^2 = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2.$$

2° 样品相似系数采用夹角余弦

设第*i*个样品与第*j*个样品的夹角为 θ_{ij} , 则两样品的相似系数为

$$\cos \theta_{ij} = \frac{\sum_{k=1}^p x_{ik} x_{jk}}{\sqrt{\sum_{k=1}^p x_{ik}^2 \sum_{k=1}^p x_{jk}^2}}$$

3° 聚类谱系图如图 1

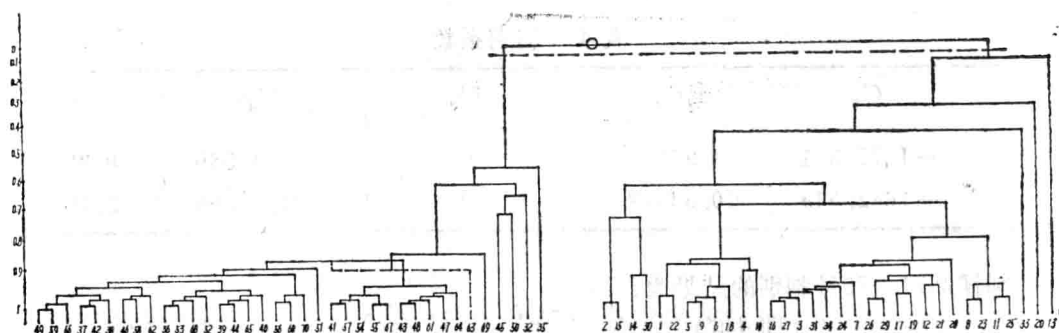


图 1

从图 1 看出, 相似系数取值在 0.1—0 之间时, 70 个样号分成两组: *B* 组除了包含临床诊断的 *B* 组全部样号外, 还包括临床诊断为 *A* 组的 32 和 35 两个样号。这个结果与判别分析的结果也几乎一致, 且更接近临床诊断结果。因此, 进一步验证了统计计量诊断的有效性及其三种统计计量诊断的一致性。

五、小 结

1. 本文介绍两类线性判别分析, 逐步判别分析和聚类分析三种多元分析方法作为菌痢的计量诊断方法。结合计算机的应用构成了计算机辅助诊断。三种统计计量方法可以单独结合计算机使用, 也可以综合运用。从方法和计算角度看, 两类线性判别较简单。逐步判别选取诊断菌痢的最重要指标作判断, 这是它的优点, 聚类分析则作了纯数量的统计计量诊断, 不依赖于经验。但无论如何, 它们还有赖于医生的最后判断。

2. 三种计量诊断与临床诊断虽然有 2 或 3 例结果不一致, 但从文中所述看出, 不论从统计检验, 后验概率或回代计算的符合率角度看, 它们都是行之有效的。

3. 计量诊断不但可以区分 *A*, *B* 两组, 还可以随时观察病情的变化。如 71—75 号, 经判别计算知, 71—74 号是病情较严重的菌痢患者。75 号则属病愈阶段了。在逐步判别或聚类分析中, 增加划分的类别, 则可将病情严重程度划分成不同等级。

参 考 文 献

- [1] 林少宫等, 多元统计分析及计算程序, 华中工学院出版社 (1987)
- [2] 张尧庭等, 多元统计分析引论, 科学出版社 (1978)
- [3] 方开泰, 数学的实践与认识, 第 1、2 期 (1978)
- [4] 方开泰, 多元分析资料汇编, 中国科学院数学所概率统计室编 (1977)
- [5] 黄秉宪等, 计量医学, 上海科技出版社 (1984)

Quantity Aid Diagnosis of Bacillary Dysentery

Sheng Dingxuan Wu Binfeng

(Huazhong University of Science and Tecnology)

Abstract

By means of the method of Multivariate Analysis, Such as 2—Class Linear Discriminant Analysis, Stepwise Discriminant Analysis and Cluster Analysis, etc. This article deals with the quantity aid diagnosis. Its results basicallly accords with clinical diagnosis.

城市用水量增长模型及其应用效果分析

金云程

(华南工学院自动化系统工程研究室)

摘 要

本文选用时间序列分析中的自适应滤波调整的预测法,建立某市用水量的预测模型。得到了用水需求量发展变化规律。预测该市“七、五”期间的逐年用水量,并对预测模型进行评价。1987年的预测值与实际用水量比较,精度达99.41%。在此基础上,文章还对预测结果进行分析,为该市的自来水公司的经济发展,提出了决策。

一、引 言

城市用水水平的预测,是城市规划设计的重要内容和依据。科学地预测城市用水量,对合理规划城市水网结构及水厂的选择与布局,具有重要的作用和意义。

城市用水量的预测是要从现状的水平出发,预测未来的发展水平。它要求具有科学性。而预测的科学性,应是有反映客观规律性的科学数据,采取适应发展规律的科学方法。选用符合实际的科学参数,以达到预测的目标。

城市用水量的预测方法较多,外推法、平滑法、分解法、回归法等等。为了加强科学性、实用性本文选用了近代的时间序列分析中的自适应滤波调整法建立某市用水量的预测模型,并通过此模型预测该市“七、五”期间的逐年用水量。在此基础上,进一步分析其增长的趋势其对该市自来水公司经济发展的影响。并提出了对策分析。

关键词:自适应滤波调整法、时间序列、中短期预测、对策分析。

二、用水量预测模型的建立

1. 自适应滤波调整法原理

时间序列分析的自适应滤波调整法和移动平均法、指数平滑法和多项式拟合法一样,是根据过去观察值的加权平均数来确定预测值的一种方法。这种预测方法可表示为:

$$\hat{X}_{t+1} = \sum_{i=1}^N W_i X_{t-i+1} \quad (1)$$

式中: $\hat{X}_{t+1}$ 为 $t+1$ 时期的预测值;

W_i 为 i 期观察值的权数;

X_i 为 i 时期的观察值;

N 为用以计算 X_{t+1} 的时期(权数)的数目。

滤波调整的主要内容是,首先确立一组权数,然后多次反复地试用这些权数,以便得到均方误差为最小的预测值。

它可利用以下的简单方程式,在每次反复之后修改其权数,即可得到上述的结果。

$$W'_i = W_i + 2Ke_{t+1}X_{t-i+1} \quad (2)$$

式中: $i=1, 2, 3, \dots, N$;

$t=N+1, N+2, \dots, n$;

W'_i 为修改后第 i 期的权数;

K 为被称为调试常数中的一个常数;

e_{t+1} 为 $t+1$ 期的误差。

2. 某市用水量预测模型的建立。

自适应滤波调整法对社会经济系统的各项指标的预测是一个较好的方法。一方面,它能“自适应”地调节权值以适应不同类型的实际数据,显得比移动平均或指数平滑方法等更为合理,另一方面,它又无需进行因素分析或结构分析以构造预测对象的数学模型。特别是社会经济的系统问题(如区域经济社会发展规划、决策分析等)经常有多项数值指标需要预测,而其中绝大多数的数学模型又难以通过机理的分析来确定。所以适应性强的滤波调整预测法就有其方法简单易行且能通过实际数据的分析来揭示事物内部的规律。特别是对中短期的预测有它明显的效果。因而得到了广泛的应用。

城市用水量是受到各种各样的因素影响的。它与自然条件(如气温、雨量、气压和湿度等)有关。也与城市的供水系统本身供水能力、管网设施、服务系统的管理等有关。更与城市人口数量、人民生活水平以及对外开放、旅游业的发展都有很大的关系。情况是错综复杂的。要全面地取得各有关因素的资料以及分析各因素对它的影响是十分困难的。因此用时间序列分析法的自适应滤波调整法来预测某市用水量是比较合适的。

由表 I 给出某市从1966年至1985年各年年末的用水量(单位万吨)。共升期。考虑用水量与国民经济计划发展有关,故选 $N=5$ 。选择合适的调试常数 K ,经过多次权重优化计算直到预测均方误差最小或稳定为止。由表 I 给出。

表 I: 用水量预测模型计算表

年 份	序号	实际售水量 X_t	第一轮计算		第二轮计算		第三轮计算	
			$\hat{X}_t$	e_t	$\hat{X}_t$	e_t	$\hat{X}_t$	$\hat{e}_t$
1966	1	10035.81						
1967	2	10808.94						
1968	3	10769.65						
1969	4	12451.68						
1970	5	13554.68						
1971	6	15076.56	11524.15	3552.414	15095.76	19.20117	15095.69	-19.12793

续上表

1972	7	16669.51	12991.22	3678.289	16423.91	245.5996	16422.72	246.7871
1973	8	18940.87	14775.24	4165.627	18008.48	932.3868	18006.69	934.1816
1974	9	20919.02	17327.25	3591.766	20333.43	585.586	20331.87	587.1465
1975	10	23305.84	20088.79	3217.047	22720.21	585.6309	22718.46	587.3848
1976	11	25107.47	23328.34	1779.127	25492.3	-384.8281	25490.75	-383.2774
1977	12	27441.06	26424.99	1016.066	28036.93	-595.8711	28036.32	-575.2559
1978	13	30452.36	29569.71	882.6524	30641.16	-188.8027	30641.63	-189.2734
1979	14	34217.58	32977.84	1239.738	33590.3	627.2813	33590.97	626.6133
1980	15	37651.59	37231.13	420.4649	37514.55	137.043	37514.36	137.2344
1981	16	41033.97	41372.52	338.5508	41462.1	-428.1328	41460.98	-427.0117
1982	17	43294.72	45298.23	-2003.508	45311.6	-2016.879	45311.43	-2016.707
1983	18	46533.96	47175.81	-641.8555	47176.79	-642.832	47178.5	-644.543
1984	19	49998.14	50347.35	-349.2031	50349.29	-351.1485	50351.04	-352.8945
1985	20	55775.08	53704.86	2070.223	53705.27	2069.813	53705.78	2069.305

经过计算可得到某市用水量预测模型为

$$\begin{aligned}\hat{X}_{t+1} = & 0.2729933X_t + 0.2656945X_{t-1} + 0.2594909X_{t-2} + 0.2559704X_{t-3} + \\ & + 0.2516611X_{t-4}\end{aligned}\quad (3)$$

三、预测模型的评价

从预测模型的精确度来分析模型的预测能力。利用模型(3)通过GF20/IIC 汉字电脑的计算, 得出某市1971年至1985年理论用水量 $\hat{X}_{t+1}$ 与实际用水量的比较, 由表Ⅱ给出。

表Ⅱ: 实际用水量与理论用水量比较

年 份	序 号	实际值 (万吨)	理论值 (万吨)	相对误差
1971	6	15076.56	15095.6	$-1.262761E-03$
1972	7	16669.51	16423.98	$1.472936E-02$
1973	8	18940.87	17969.29	$-5.129554E-02$
1974	9	20919.02	20113.01	.0385299

续上表

1975	10	23305.84	22336.39	$4.159692E-02$
1976	11	25107.47	24893.28	$8.53075E-03$
1977	12	27441.06	27515.09	$-2.697686E-03$
1978	13	30452.36	30328.72	$4.060197E-03$
1979	14	34217.58	33347.25	$2.543512E-02$
1980	15	37651.59	36841.86	$2.150598E-02$
1981	16	41033.97	40610.49	$1.032024E-02$
1982	17	43294.72	44780.92	-0.0343275
1983	18	46533.96	48910.75	$-5.107644E-02$
1984	19	49998.14	53101.87	$-6.207692E-02$
1985	20	55775.08	57226.12	$-2.601598E-02$

从上表可知，其预测理论值准确度都达到93%以上。平均在97.38%。因此可认为模型(3)是适合预测某市用水量发展趋势的。

利用模型(3)对1986年的用水量进行预测，并与实际用水量进行比较如下：(单位万吨)

年 份	序 号	实 际 值	理 论 值	相 对 误 差
1986	21	61844.61	61994.41	0.0024

其准确度可达99.76%。从而证实了对近期来说，此预测模型是合理的、可靠的。(注)

分析模型的稳定性。由表Ⅲ给出了从1971年为起始数据建立了预测模型的实际值与理论值的比较，并得知此模型其参数和预测结果都不会受较大的影响，则表明此模型具有抗干扰的能力较强，可预测的时间也较长，也具有较高的稳定性。

表Ⅲ：实际值(从1976年算起)与理论值比较

年 份	序 号	实 际 值 (万 吨)	理 论 值 (万 吨)	相 对 误 差
1976	6	25107.47	24893.63	$8.516826E-03$
1977	7	27441.06	27515.91	$2.727721E-03$
1978	8	30452.36	30329.54	$4.033324E-03$
1979	9	34217.58	33347.44	$2.542964E-02$
1980	10	37651.59	36841.26	$2.152196E-02$
1981	11	41033.97	40610.52	$1.031939E-02$
1982	12	43294.72	44782.03	$3.435321E-02$
1983	13	46533.96	48912.87	$5.112203E-02$
1984	14	49998.14	53103.69	$6.211325E-02$
1985	15	55775.08	57226.77	0.0260276

注：1987年用水量为67085.20万吨，准确度达99.41%。

四、用水量的预测对某市自来水公司建设对策分析

利用模型(3)对“七、五”期间某市的用水量进行预测,并由表Ⅳ给出:

表Ⅳ:“七、五”期间用水量预测值(单位万吨)

年 份	序 号	实 际 值	理 论 值	增长率(%)
1986	21	61844.61	61994.41	10.88
1987	22		67483.27	9.12
1988	23		73836.21	9.41
1989	24		80994.19	9.69
1990	25		89106.85	10.01

某市在“七、五”期间预测用水量将达到373265.13万吨,比“六、五”期间的236635.87万吨预测平均年递增长率约达9.84%。1985年平均日用水量达152.81万吨,而至1990年平均日用水量将达到244.13万吨。若要适应城市发展的需要,必须投资相当四个的建设期3至4年,供水能力为20万吨/日,投资4800万元的水厂。加上管网供水设施及改善水源设施工程,若以1:2:0.4投资比例估算,共需要投资资金达73440万元。如何解决这样巨大的建设资金是该市自来水公司所面临着一个非常突出的问题。而在“七、五”期间所预测的总利润将达到15936.95万元,除留15%后再缴能源基金外,能用于“以水养水”建设投资的资金仅有13211.24万元。其资金缺口高达约6亿元。这说明城市的自来水建设资金仅来源于自来水利润的“以水养水”的资金是永远不能满足当前国家建设发展的需要。为解决此矛盾,提出以下几个对策:

1. 要解决好自来水建设资金短缺,就要确定合理的城市建设投资中的自来水建设投资比例。改变过去对自来水投资比例偏低的情况。(据有关资料统计六十年代占1.4%,七十年代占1.11%，“六、五”期间占0.867%，而1985年仅占0.46%)部分专家认为一般比例要达到2.5%为宜。所以对自来水建设中大的重点项目国家应给予相应投资,解决资金缺口。这样才使得该市自来水建设在“七、五”期间能得到合理的发展。不然将可能会产生更为突出的供需矛盾。

2. 现行该市自来水价格背离价值,对城市基础设施的建设是不利的。水价偏低对城市节水也不利。合理调整水售价,使之适应经济价值的规律。根据上述预测,若从一九八八年开始,按现行工商用水与生活用水的价格比5:3,以及工商用水与生活用水按目前约7:13计算,利用盈亏平衡分析法推算,工商用水售价提高0.145元/吨。(即水价为0.345元/吨)而生活用水售价提高0.087元/吨。(即水价为0.207元/吨)这样可以通过理顺水价,增收“以水养水”的资金。以确保水建设资金收支基本达到综合平衡。

3. 由于预测“七、五”期间每年平均要递增9.84%过快的供水增长速度,必然带来了基础工作跟不上等的矛盾。要解决此问题的另一对策是控制这个增长速度,以适应供需平衡。

这样就要对模型(3)进行修正。由于“六、五”期间,平均用水量增长为 $\alpha_0=0.0752$,又令“七、五”期末的用水量增长率为 α ,则模型(3)可修改为

$$\hat{X}_{t+1} = \left(\frac{1+\alpha}{1+\alpha_0} \right)^{t-19} (0.2729933X_t + 0.2656945X_{t-1} + 0.2594909X_{t-2} + 0.2559704X_{t-3} + 0.2516611X_{t-4}) \quad (4)$$

(1985年对应 $t=20$)

假设“七、五”期末的用水量增长率分别控制为 $\alpha=0.07, 0.06, 0.05$ 和 0.04 的四种情况。利用模型(4)可预测用水量。由表V给出了不同增长速度对水公司经济发展的影响。

表V：不同增长率对水公司经济发展的影响比较

指 标 值	$\alpha=0.07$	$\alpha=0.06$	$\alpha=0.05$	$\alpha=0.04$
用水量(万吨)	364038.55	347120.64	331076.5	316288.16
供水能力(万吨/日)	239.37	228.24	223.2	207.97
需投资总额(万元)	15640	14960	12240	9520
总利润(万元)	15542.99	14820.68	14135.66	13504.26
能用于投资资金(万元)	12884.66	12285.89	11718.03	11194.62
资金缺口(万元)	2755.34	2674.11	521.97	-1674.62

从上表做出决策分析是最佳控制用水量增长速度在 $\alpha=0.05$ 。投资资金缺口最少,又能保持一定的增长速度。但要能实现这一决策,相应的许多措施得跟上,而从另一侧面也说明大力加速发展自来水建设事业的紧迫感。

五、几点注记

第一,自适应滤波调整法误差控制问题

原算法的每次迭代都得到一次预测误差。用 $E_{x,1}$ 表示

$$E_{x,1} = \frac{1}{n-N} \sum_{t=N}^n e_{t+1}^2 = \frac{1}{n-N} \sum_{t=N}^n \left[X_{t+1} - \sum_{i=1}^N W_i^{(t-1)} \cdot X_{t-i+1} \right]^2$$

其中 $W_i^{(t-1)}$ 随 t 从 N 变化到 n 不断改变。

这是本次“锻炼”过程的平均(假误差)而并不是代表本次“锻炼”过程的最后所确定的权 $W_i^{(n-N)}$ 的预测平均误差(真误差)。用 E_x 表示

$$E_x = \frac{1}{n-N} \sum_{t=N}^n \left[X_{t+1} - \sum_{i=1}^N W_i^{(n-N)} X_{t-i+1} \right]^2$$

这样采用真误差 E_x 作为判决误差的标准是合理的。用假误差作为判决误差的好处是计算简单。

为了考虑 W_i 的稳定性。取相邻二列预测平均误差

$$E_{\hat{X}} = \frac{1}{n-N} \sum_{t=N}^n \left[\hat{X}_{t+1,j} - \hat{X}_{t+1,j+1} \right]^2$$

其中 $\hat{X}_{t+1,j}$ 表示第 j 轮的预测值。同样也可以采用 $E_{\hat{X}}$ 作为判决误差的一个来判断 W_i 的稳定性依据。

当既考虑预测平均误差, 又考虑 W_i 的稳定性, 可由综合二者的误差控制效应

$$E = E_x \cap E_{\hat{X}}$$

表示判决误差标准。

第二, 在滤波调整中的一个重要参数 K 。它起的作用是调整权数值的快慢, 从而决定出每次调整巡回中误差减少量的大小。对不同的 K 值会得出完全不同的结果。因此研究不同的 K 值的效果是很有意义的。 K 值较小时, 每次巡回中误差减少得较小, 这就要求更多的巡回次数, 才能达到误差给定的水平。 K 值较大时, 可能使其结果较差。这是因为权数值调整得过快, 这样就会对历史观察值中的随机波动有影响。可能在调整过程中不会收敛于一组最优权数值上。因此审慎地选择一组权数的初始值也是很重要的。在没有其他信息可供使用的情况下, 可选用 $1/N$ 作为自然的起点。

第三, 最优权数 N 的确定。

原则上说, 权数 N 可取从 2 到 $n-2$ 的任意值。但考虑到自适应滤波调整本质在利用已知实际数据修正权值。若 N 取得较大, 可供“锻炼”权的数据个数 $n-N$ 就少, 迭代结果将集中依赖于少数几个数据, 可靠性降低了。因此一般可限制权数 N 不超过 $n-N$, 即有 $N \leq \frac{n}{2}$ 。至于有些问题是通过实际经验与具体地对问题的分析来作选择的。本文的最优权数 $N=5$ 就来通过对问题的分析以及经过多级比较中选优的。

参 考 文 献

- [1] Steven C. Wheelwright, Spyros Makridakis. 《Forecasting Methods For Management》JOHN WILEY & SOHS 1980 THIRD EDITION
- [2] 哈利·琼尼, 人里安·特惠斯, 《用于计划决策的技术预测》, 复旦大学出版社(陆廷纲, 桑赓陶等译)
- [3] 项静恬、杜金观、史久恩, 《动态数据处理——时间序列分析》, 气象出版社, 1986。