

# 基于灰预测GM(1,1)模型的书号优化配置建模研究

董宁, 余波

(湖南工业大学 理学院, 湖南 株洲 412008)

**摘要:** 对不完全信息市场下的出版社书号优化配置问题进行了数学建模研究。首先确定了强势产品, 运用 DPS2000 对 2006 年书号的销售量进行了灰预测, 并且运用预测出的数据在 Lingo 中建立了经验配置模型和综合市场模型, 得到了最优分配方案。对所建立的模型用 2001~2005 年的销售额进行了 Kendall $\tau$  系数检验, 结果表明销售额具有二次增长趋势。

**关键词:** 灰预测 GM(1,1) 模型; 整数规划模型; Kendall $\tau$  系数检验

中图分类号: O29

文献标识码: A

文章编号: 1673-9833(2008)06-0024-04

## Modeling Research of Deploying Book Resources Based on GM(1,1)

Dong Ning, Yu Bo

(Science of College, Hunan University of Technology, Zhuzhou Hunan 412008, China)

**Abstract:** This paper aims at obtaining the best scheme of deploying book resources by modeling research. Firstly, we give the rule of allocation based on data of last five years. We forecast the distribution of every book's number by GM(1,1) model. Then we give two different models of integer programming in Lingo according to predicted sales number of 2006. From the outcomes of Kendall-coefficient test, the sales based on prediction of GM(1,1) increases quadratically.

**Key words:** GM(1,1); integer programming; Kendall-coefficient test

## 0 引言

出版社的资源是有限的, 要使有限的资源发挥出最大的经济效益, 就必须对资源进行科学合理的分配, 其中一种关键的资源配置就是书号的分配。2006 年全国大学生数学建模 A 题就提出了这样一个问题:

某个以教材类出版物为主的出版社, 总社领导每年需要针对各分社提交的书号(总量远大于总社的书号总量)、人力资源情况以及市场信息分析, 将总量一定的书号数合理地分配给各个分社, 在支持强势产品的原则下使出版的教材产生最好的经济效益。书号分配完成后, 各个分社(分社以学科划分)根据分配到的书号数量, 再重新对学科所属每个课程作出出版计划, 付诸实施。请依据本题给出的该出版社所掌握的一些数据资料(在 <http://mcm.edu.cn/> 下载), 利用数学

建模的方法, 在信息不足的情况下, 提出以量化分析为基础的资源(书号)配置方法, 给出一个明确的分配方案, 向出版社提供有益的建议。

试题中提出的问题是依据前 5 年的数据, 如何分配 2006 年的书号。一种解题思路是若能预测出 2006 年每个书号的销售量, 这样在书号的配置过程中就可以把预测出来的销售量当成已知量, 以产生最好的经济效益为目标, 在 Lingo 软件中可以建立整数规划模型来求得书号分配方案。

## 1 数据的处理

由于题中给出了 1 个调查问卷和 4 个数据附件, 并且数据量非常大, 因此如何处理数据及充分挖掘数据包含的信息是解题的基础。通过对试题附件中数据的

收稿日期: 2008-06-18

作者简介: 董宁(1979-), 女, 山东菏泽人, 湖南工业大学助教, 主要研究方向为应用数学;

余波(1979-), 男, 湖南株洲人, 湖南工业大学讲师, 博士生, 主要研究方向为计算数学。

分析, 得出以下结论:

1) 统计附件四中每年出版社书号量的数据, 可以得出总社每年拥有的书号资源量为 500 个。

2) 对于题中附件三给出 2001~2005 年的各课程的销售量, 由附件四已知这 5 年各课程对应的书号数目。两者相除, 则可求出 5 年内各课程每个书号平均所对应的销售量, 这些数据可以在灰预测 GM(1,1)模型中预测出 2006 年各课程每个书号的平均销售量。

3) 原题是以支持强势产品为原则的, 这就涉及到 72 门课程中哪些课程的教材为强势产品, 因此本文须对 72 门课程进行分类。运用统计软件 DPS2000 进行聚类分析, 若依据附件四各分社 2001~2005 年 72 门课程所分得的书号数目进行聚类, 则结果为: 高等数学 (23)、大学英语 (31)、概率统计 (26)、数学分析 (22)、复变函数 (34) (括号里数据为课程代码) 等分得书号较多的应为强势产品。实际并非完全这样, 因为有些课程书号可能少, 但是每个书号的销量非常大, 若出版社从利益的角度出发, 出版这类书利润很可观。为此本文对 2001~2005 年 72 门课程的每个书号销售量 (由 2) 已得) 进行聚类分析<sup>[1]</sup>, 结果为: 毛泽东思想概论 (45)、高等数学 (23)、思想道德修养 (46)、泛读 (44) 等单个书号销量较多的应为强势产品。所以在制定强势原则下利益最大化应该综合上述两个方面的考虑。

4) 附件五对人力资源数据的处理体现在建立整数规划模型的约束条件中。

## 2 模型的建立与求解

本文目标是合理地分配 2006 年书号使得出版社的利润最大化, 因此在已知每种书的均价下, 目标函数应是第  $i$  类课第  $j$  门课程分得的书号和每个书号平均销售量的函数, 但是这 2 个量都是未知变量, 所以建立

的模型为非线性模型。根据上述思路的运作, 本文由 2001~2005 年的数据预测出 2006 年的每个书号的平均销售量, 这样目标函数就可以线性化, 在满足一定的线性约束下, 模型就是整数规划模型。所以本文的关键是如何合理地预测每个书号平均销售量。

### 2.1 对每个书号平均销售量的灰预测

由前面的数据处理过程中 2), 本文得到了 2001~2005 年各课程每个书号平均对应的销售量, 但每年的数据量较少 (5 个), 故本文采用灰预测 GM(1,1)模型来预测 2006 年的各课程书号平均对应的销售量。模型最大的优点在于需要的信息量较少, 有 4 个以上的数据就可以进行预测<sup>[2-4]</sup>。本文书号优化配置建模研究的目的是要预测 2006 年各课程 (课程 1~72) 各书号的销售情况。以课程 1 为例, 数据如表 1 所示。

表 1 2001~2005 年课程 1 的书号的销售量  
Tab. 1 Sales of books for course 1 from 2001 to 2005

年份	2001	2002	2003	2004	2005
销售量 / 册	124	113	154	240	224

在 DPS2000 软件<sup>[5, 6]</sup>中对 2006 年第一门课程书号销售量进行预测, 拟合和预测出的结果如表 2 所示。

表 2 预测出 2006 年课程 1 的书号销售量  
Tab. 2 Predicted sales of books for course 1 of 2006

年份	观察值 / 册	拟合值 / 册	误差	误差 / %
2002	113	108.603 6	4.396 4	3.890 61
2003	154	164.925 34	-10.925 34	-7.094 37
2004	240	231.133 22	8.866 78	3.694 49
2005	224	226.309 89	-2.309 89	-1.031 2
2006		303 册 (预测值)		

运用以上同样的方法在 DPS 中预测出 2006 年 72 门课程平均每个书号销售量如表 3 (精确到整数):

表 3 预测出 2006 年 72 门课程的书号销售量  
Tab. 3 Predicted sales of books for 72 courses of 2006

课程	预测值 / 册	课程	预测值 / 册	课程	预测值 / 册	课程	预测值 / 册	课程	预测值 / 册
1	303	16	667	31	795	46	8 605	61	233
2	417	17	805	32	526	47	4 426	62	632
3	289	18	601	33	883	48	1 339	63	620
4	293	19	3 616	34	330	49	743	64	793
5	190	20	1 403	35	548	50	349	65	293
6	325	21	706	36	274	51	467	66	1 466
7	514	22	944	37	119	52	901	67	609
8	71	23	9 444	38	847	53	1 560	68	484
9	430	24	1 386	39	414	54	1 262	69	597
10	470	25	1 832	40	407	55	3 822	70	464
11	1 692	26	2 887	41	6 720	56	1 068	71	615
12	2 256	27	720	42	8 179	57	521	72	1 182
13	513	28	651	43	9 414	58	208		
14	1 184	29	496	44	4 991	59	980		
15	471	30	3 044	45	22 467	60	485		

### 2.2 对书号分配的建模与求解

利用灰预测方法得到了2006年每个书号平均销售量数据,通过预测的数据可以建立整数规划模型使得出版社利益最大化。

#### 模型1: 经验配置模型

$$\max x = \sum_{i=1}^n \sum_{j=1}^{13} x_{ij} c_{ij} x_{sj_j} \quad (\text{总社的利润最大化}),$$

$$\text{s.t.} \begin{cases} \sum_j x_{ij} \geq sg_i / 2, & (1) \\ \sum_i \sum_j x_{ij} = 500, & (2) \\ \sum_i x_{ij} \leq rz_i, & (3) \\ x_{ij} \leq ksg_{ij}, & (4) \\ xx_{ij} \leq x_{ij} \leq sx_{ij}, & (5) \end{cases}$$

$x_{ij}, sg_i, rz_i, ksg_{ij}, xx_{ij}, sx_{ij}$  均为整数。

其中,目标函数中 $x_{ij}$ 为要求的第*i*类课第*j*门课程所分得的书号数,为决策变量; $c_{ij}$ 为第*i*类课第*j*门课程的均价; $x_{sj_j}$ 为第*i*类课第*j*门课程的销售量(已由前述两类方法中预测出)。约束条件中,(1)表示第*i*类课程分得的书号数应不小于第*i*分社欲申请数目 $sg_i$ 的一半;(2)表示总社所有的书号数为500;(3)表示第*i*分社的书号量不大于其人力资源 $rz_i$ 所能处理的上限;(4)表示每门课程分得的书号不大于其申请量 $ksg_{ij}$ ;(5)表示每门课程分得的书号有其上、下限,其中上限 $sx_{ij} = \min\{\text{书号历年均值加上极差,欲申请书号数目}\}$ ,下限 $xx_{ij} = \max\{\text{历年均值减去极差,欲申请书号数目的一半}\}$ 。原则是按照以往的经验给出每门课程分配的书号不至于太大(超过申请的数目)和太小(以利润最大化为目标可能使得某些课程分配到书号量为零)。这也是经验配置模型名称的由来。

采用灰预测 $GM(1,1)$ 预测 $x_{sj_j}$ 时,对上述模型1在Lingo 8.0中编程<sup>[7]</sup>,计算的结果如表4。

表4 由模型1得到的2006年各课程分得的书号数

Tab. 4 Book numbers for all courses of 2006 from model 1

课程类别	各课程分得的书号数 / 个										合计 / 个
计算机类	9	9	1	2	2	9	10	2	9	5	58
经管类	8	4	4	4	2	5	6	8	6	4	51
数学类	3	23	36	3	12	18	5	2	13	5	120
英语类	40	2	2	13	7	6	2	6	5	6	89
两课类	4	8	8	9	6	8	9	13			65
机械类	15	0	0	15	4	8					42
化工类	0	4	4	1	4	9					22
地质类	1	6	6	7	1	4					25
环境类	6	4	7	4	4	3					28

由以上的分配方案可以算出出版社总的销售额(即目标函数)为2 622万元。

#### 模型2: 综合市场模型

模型1虽然在其约束条件下给出了一个最优解,

但是从解的结果看,由于最后一个约束条件是以加减书号极差作为上下限的,因此解的跨度比较大,以至于解中出现了一些零,说明出版社不会对此种教材分配书号,这对于一些专业基础课是不利的。例如:化学化工类教材《化学与现代文明》在众多出版社中仅有P115和P357两家出版,虽然销量不大,但是作为少数专业基础课是应该扶持的,否则会出现教材紧缺的情形;另一方面,模型1没有考虑市场需求和课程教材市场占有率等方面的因素,因此模型带有一定的主观经验性。以下综合市场与专业扶持等因素,从市场占有率的角度考虑建立模型,由此称模型2为综合市场模型。

首先在Excel中对2001~2004年72门课程的市场占有率作散点图,由此把出版的72门课程教材分成了3类:1)占有率80%以上(高点);2)占有率40%~80%(中点);3)占有率40%以下(低点)。分别对高点、中点、低点类的课程所应分配的书号数进行一定的限制,具体来说是通过前4年求出均值,再对占有率为高、中、低的课程分别设待定正数*a*和*b*使得:

$a \times \text{前4年的均值} \leq 90\%$ 以上05年的点 $\leq b \times \text{前4年的均值}$ ,从而确定了占有率的系数*a*和*b*。最后把得到的*a*和*b*乘上2001~2005年的书号均值得到2006年应分得书号的上下限,结果如下:

1) 占有率低的课程书号分配区间为 $(0.7 \times 5 \text{年}(2001\sim 2005) \text{书号分的数目均值}, 1.3 \times 5 \text{年}(2001\sim 2005) \text{书号分的数目均值})$ ;

2) 占有率中的课程书号分配区间 $(0.75 \times 5 \text{年}(2001\sim 2005) \text{书号分的数目均值}, 1.5 \times 5 \text{年}(2001\sim 2005) \text{书号分的数目均值})$ ;

3) 占有率高的课程书号分配应在 $(0.85 \times 5 \text{年}(2001\sim 2005) \text{书号分的数目均值}, 06 \text{年欲申报的书号数目})$ 。

根据此标准,本文模型2的约束条件中对每个书号分配的上下限和作了上述改进,在Lingo 8.0中建模,获得的结果如表5。

表5 由模型2得到的2006年各课程分得的书号数

Tab. 5 Book numbers for all courses of 2006 from model 2

课程类别	各课程分得的书号数 / 个										合计 / 个
计算机类	9	9	1	2	2	9	12	2	9	3	58
经管类	8	4	2	4	2	5	6	8	6	4	49
数学类	3	23	37	5	1	18	8	3	13	8	120
英语类	30	2	2	16	5	10	2	4	5	3	79
两课类	4	8	8	9	6	8	9	11			63
机械类	18	1	3	15	4	8					49
化工类	2	3	4	1	4	9					23
地质类	4	6	6	7	1	4					28
环境类	6	6	6	6	4	3					31

从以上计算统计结果可以看出,由Lingo软件算出出版社总的销售额(即目标函数)为2559.8万元。另

外结果显示模型2中没有出现分配书号为零的课程, 即对某些专业进行了扶持; 同时销量较好的书分配了较多的书号, 符合前面提出的强势原则。

### 3 模型的检验

进行销售额差分的 Kendall  $\tau$ 系数的有关趋势的检验<sup>[3]</sup>, 其目的在于衡量销售额具有怎样的上升趋势。由数据分析方法知若原序列呈现直线状趋势, 则它的一次差分 $\Delta y_t = y_t - y_{t-1}$ 构成序列将没有趋势; 若原序列呈现二次型趋势, 则二次差分序列 $\Delta^2 y_t = \Delta y_t - \Delta y_{t-1}$ 有可能减弱趋势, 这种推广可以继续下去。

为此在教材均价保持不变的情况下可以推算2001~2006年的销售额、一次差分 and 二次差分如表6。

表6 2001~2006年销售额的差分

Tab. 6 Difference of sales from 2001 to 2006

年份	销售额 / 万元	一次差分	二次差分
2001	860.6		
2002	932	0.071	
2003	1 114	0.182	0.111
2004	1 436	0.322	0.140
2005	1 789	0.353	0.031
2006	2 559	0.770	0.417

从表上可以看出, 2001~2006年销售额数据呈现上升趋势, Kendall  $\tau$ 系数检验可以检验出数据上升趋势是否为线性趋势。首先对差分序列观察值的  $n(n-1)/2$  种可能配对进行检验, 若在一对观测值中, 时间  $t$  大一些的观察值大于另一个观测值, 则记这一对为“向上对”; 反之记为“向下对”。令  $N_u$  = 向上对的总数,  $N_d$  = 向下对的总数, 则有: 若序列无趋势, 则  $N_u - N_d$  几乎

为零; 若序列具有向上趋势, 则  $N_u - N_d > 0$ ; 若序列具有向下趋势, 则  $N_u - N_d < 0$ ; 对于假设  $H_0$ : 一次差分序列无趋势;  $H_1$ : 一次差分序列有趋势, 当  $n \leq 10$  时, 则检验统计量  $\tau = (N_u - N_d) / \frac{n(n-1)}{2}$ ; 拒绝域为  $|\tau| \geq \tau_{\alpha, n}$ 。现取显著水平, 由2001~2005年和2006年灰预测数据计算, 查表得, 结果显示拒绝, 即数据不是线性趋势上升。实际上, 本文对二次差分序列同样做上面的检验, 结果显示不拒绝原假设, 即数据上升趋势为二次函数上升。

#### 参考文献:

- [1] 施锡铨. 数据分析方法[M]. 上海: 上海财经大学出版社, 1997.
- [2] 邓聚龙. 灰预测与灰决策[M]. 武汉: 华中科技大学出版社, 2002.
- [3] 程晓鑫, 周渝慧. 基于灰色改进模型的电价预测[J]. 华北电力大学学报, 2006, 33(1): 47-50.
- [4] 王霜, 殷国富, 罗中先, 等. 基于GM(1,1)模型的用电量预测方法[J]. 西华大学学报: 自然科学版, 2007, 26(1): 77-78.
- [5] 唐启义, 冯明光. 实用统计分析及其DPS数据处理系统[M]. 北京: 科学出版社, 2002.
- [6] 吴建国. 数学建模案例精编[M]. 北京: 中国水利水电出版社, 2005.
- [7] 谢金星, 薛毅. 优化建模与Lingo软件[M]. 北京: 清华大学出版社, 2005.

(责任编辑: 罗立宇)