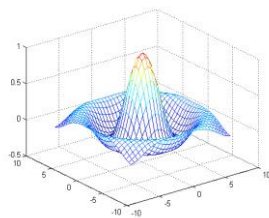


五一数学建模竞赛



题 目： _____ 让标枪飞 _____

关键词：拟合曲线函数 多元线性回归 牛顿-欧拉动力学方程 龙格-库塔算法

摘 要：

本文通过最小二乘法曲线拟合、多元线性回归、牛顿-欧拉动力学方程、龙格-库塔算法、线性规划等方法，算出了题目所给标枪的表面积、中轴线剖面面积等，建立了标枪运动模型，并分析了各个影响标枪运动因素的重要性。

对于问题一，首先将标枪横放，建立以标枪底为原点的标准坐标系，用附录中“某型标枪测量尺寸表”中长轴值和直径值的二分之一来表示标枪的散点轮廓，将整支标枪分成把手后、把手、把手前、枪尖四段，再用最小二乘法的思想来拟合曲线，通过 MATLAB 的绘图和拟合功能，得到标枪四段的单边外形轮廓和单边轮廓函数，见表 1。再用积分的方法，求出标枪的表面积、中轴线剖面面积和标枪形心的位置。

对于问题二，本文首先从数据出发，通过观察数和对比分析，初步定性得到了标枪飞行轨迹的规律；再从影响因素出发，建立出手速度、初始攻角、出手速度与标枪飞行距离的多元线性回归模型，使用最小二乘法对其参数进行估计，得到回归系数，见表 3。然后用 SPSS 对模型进行显著性检验，得出初始攻角与投掷距离相关性不显著的结论。在此基础上，本文建立了不考虑初始攻角的标枪飞行运动模型，最后通过微分方程的形式，定量得到了标枪飞行过程中的运动规律。

对于问题三，本文首先建立了以运动员为原点建立 xyz 坐标系，以标枪重心为原点建立 $\xi\eta\zeta$ 坐标系，分别研究质点的运动和标枪的转动；然后对标枪进行受力分析，建立牛顿-欧拉运动方程，用龙格-库塔算法，通过 MATLAB 编程作数值求解。对于第（1）问，代入以至已知量即可求得投掷距离约为 85.34505m。对于第（2）问，用线性规划的方法，将出手角、初始攻角作为决策变量，飞行距离作为目标函数，编写 MATLAB 程序，求得当出手速度为 30m/s 时，最佳出手角为 42° 、最佳初始攻角为 -8° 、最大投掷距离为 91.794m。

对于问题四，该问题与问题三相似，仅在问题三的基础上加上了风力、初始俯仰角速度对标枪飞行的影响，而坐标系的建立、值求解方法等均与问题三一致。最终计算结果见表 6。

对于问题五，首先在问题四的基础上，将决策变量数据进行微变，得到相应的投掷距离，作为数据基础。然后以 MATLAB 为工具，分别对各影响因素与投掷距离关系进行拟合。通过比较曲线斜率变化率，得到决策变量相对重要性由大到小依次为：出手速度>出手角>初始攻角>风速>初始俯仰角速度。

一、 问题重述

标枪运动员如何取得最佳成绩依赖于运动员水平、标枪技术参数和比赛环境等因素。为了探讨标枪出手时初始条件的良好组合规律，需要解决如下问题。

问题一：根据标枪示意图与某型标枪测量尺寸数据，估算该型标枪沿标枪中轴线剖面面积、标枪表面积和标枪形心（剖面几何中心）的位置。

问题二：根据 24 名运动员标枪投掷的实测数据，建立数学模型模拟标枪在空中的飞行轨迹，找到标枪飞行过程中的运动规律。

问题三：对标枪投掷出手瞬间、出手后受力及运动情况进行分析，建立标枪飞行的数学模型。应用建立的模型估算某运动员（投掷出手速度为 29.70m/s、出手角为 36.6° 、初始攻角为 -0.9° ）所投掷标枪距离。再利用模型，给出运动员在出手速度为 30 m/s 情况下使得投掷距离最大的出手角和初始攻角，并估算最大投掷距离。

问题四：考虑风力的影响，建立数学模型。在出手速度为 31.7 m/s 的条件下，应用模型，求解风向分别为顺风 and 逆风，风速分别为 3 m/s、6m/s 和 9m/s 条件下，使得投掷距离最大的出手角、初始攻角和初始俯仰角速度，并估算最大投掷距离。

问题五：建立数学模型分析出手速度、出手角、初始攻角、初始俯仰角速度、风向及风速等要素对标枪投掷距离影响的相对重要性。

二、 问题假设

假设 1：顺风是从运动员的正后方吹正前方，逆风是从运动员正前方吹向正后方；

假设 2：标枪的轮廓是连续光滑的曲线；

假设 3：运动员的实际投掷可以达到理论上的初始条件；

假设 4：标枪在飞行运动中做不发生形变；

假设 5：不考虑温度、湿度等复杂因素对标枪性能的影响。

三、 符号说明

序号	符号	意义
1	α	初始攻角
2	β	出手角度
3	γ	持枪角
4	ω_0	初始俯仰角速度
5	c_x	空气阻力系数
6	C	重心
7	Q	形心
8	v_w	风速

序号	符号	意义
9	W	重力
10	F	摩擦阻力
11	L	压差升力

四、 问题分析

4.1 问题一的分析

问题一要求计算标枪的表面积、中轴线剖面面积、形心的位置，因此需要将标枪放在坐标系中并拟合出标枪轮廓的函数，定性分析。考虑到标枪轮廓变化并非连续，需要对标枪分段进行拟合曲线函数。然后用积分的方法，根据面积、表面积、形心公式估算剖面面积、标枪表面积和标枪形心的位置。

4.2 问题二的分析

问题二要求找出标枪飞行过程中的运动规律，因此需要对标枪进行受力分析，进而构建标枪飞行运动的物理模型；为了简化分析，在建立模型之前，本文从观察对比、多元线性回归的角度分析，去除影响较小的因素。

然后在简化的标枪飞行过程中，对标枪受力分析，写出动力学，定量得出标枪飞行规律。

4.3 问题三的分析

为了建立标枪飞行的模型，需要基于运动学、动力学、空气动力学对标枪飞行过程进行分析。由于在整个标枪飞行过程中，标枪上升和下降过程受力有差异，所以需要分段考虑。而在飞行过程中，标枪受到环境的影响较大、质心与形心不在同一位置，所以其不仅做平动而且自身会绕重心转动，所以需要对标枪质心运动及标枪运动进行分析。因此本文建立牛顿-欧拉动力学模型，对标枪运动综合分析。

4.4 问题四的分析

问题四要求在问题三的基础上，考虑风和初始俯仰角速度对标枪飞行的影响，并找到相应风速下的最佳初始条件及最大投掷距离。因此，问题四与问题三的模型大体一致，求解方法也基本相同。

4.5 问题五的分析

问题五要求决策变量对标枪投掷距离的相对重要性，需要一定的数据基础。因此，本文基于问题四的模型，控制单一变量，输入各个单一变量的数据，得到多组相应的投掷距离。从而分析影响因素与投掷距离的相关性。为了比较相对重要性，需要以 MATLAB 为工具，分别对各影响因素与投掷距离关系进行拟合。通过比较曲线斜率变化率，得到的相对重要性。

五、 问题一模型的建立与求解

5.1 模型准备

5.1.1 坐标系建立

建立如图 1 所示直角坐标系，如下：

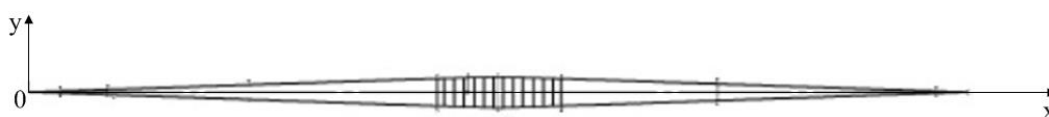


图 1 标枪几何坐标

5.1.2 最小二乘法拟合曲线

考虑到标枪轮廓变化并非连续，本文将标枪分为把手后、把手、把手前及枪尖四部分，如图 2。

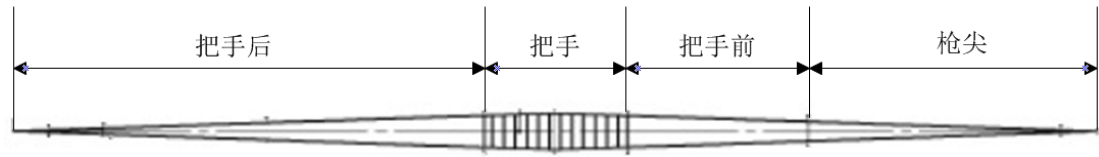


图 2 标枪示意图

根据附录“某型标枪的测量尺寸表”中的数据，将“长轴”值作为横坐标，“直径”值的二分之一作为纵坐标，用最小二乘法的方法来拟合标枪上半部分曲线，通过编写 MATLAB 程序（见附件 1）拟合得到二次曲线函数：

$$y = ax^2 + bx + c \tag{5-1-1}$$

标枪各段二次曲线函数系数如表 1

部分 \ 系数	把手后	把手	把手前	枪尖	标枪
a	-7.796e-06	-2.743e-19	-4.303e-06	-0.0001311	-7.515e-06
b	0.019	8.232e-16	0.01261	0.6396	0.02012
c	3.58	18.22	6.044	-773.2	3.094

表 1 拟合方程系数

标枪各段上部分的拟合曲线如下：（图中数据均是标枪真实宽度的 5 倍）

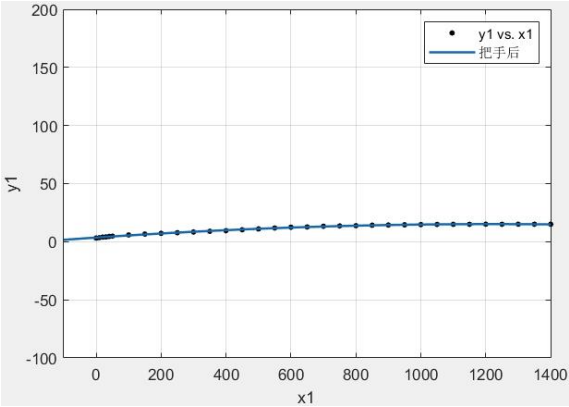


图 3 把手后拟合曲线

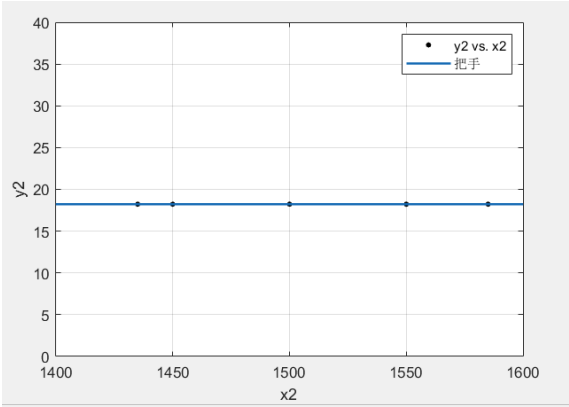


图 4 把手拟合曲线

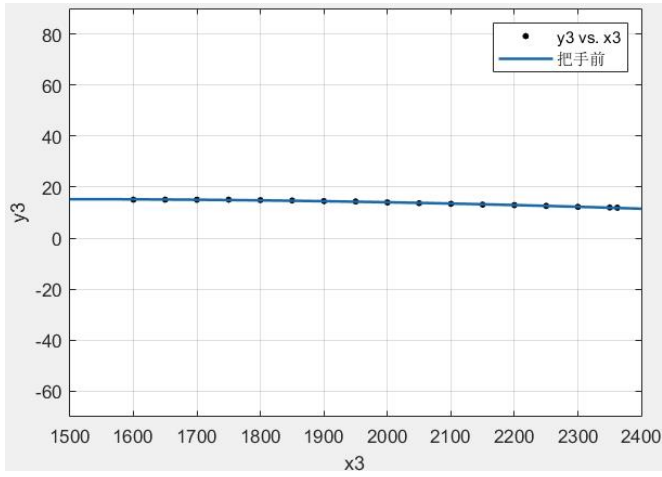


图 5 把手前拟合曲线

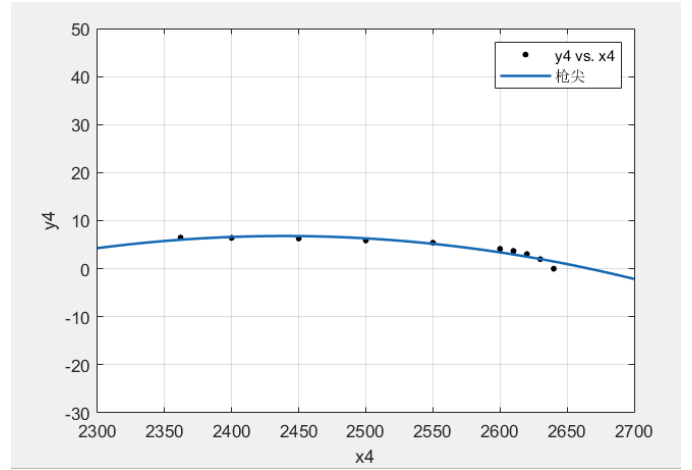


图 6 枪尖拟合曲线

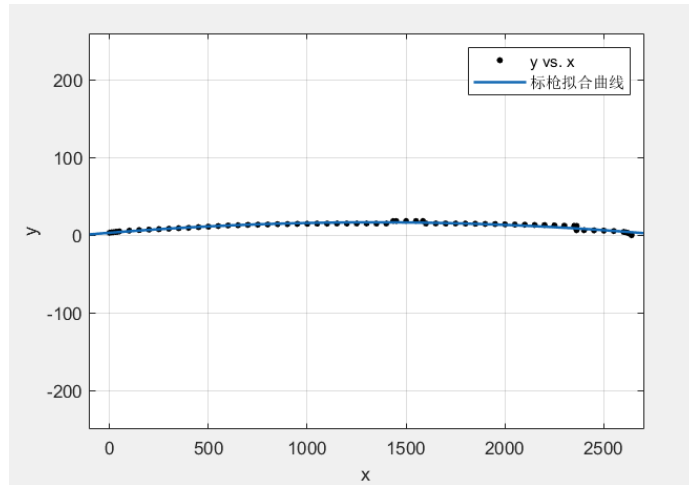


图 7 标枪拟合曲线

5.2 模型建立

由几何知识，剖面面积计算公式为：

$$S_0 = \int 2y dx \quad (5-2-1)$$

表面积计算公式为：

$$S = \int 2\pi y dx \quad (5-2-2)$$

形心位置计算公式为：

$$x = \frac{\pi \int x \cdot y^2 dx}{\pi \int y^2 dx} \quad (5-2-3)$$

5.3 模型求解

根据计算公式，在相应的区域对各段函数求定积分，通过编写 MATLAB 程序（见附件 2）计算可求得标枪各段中轴线剖面面积、表面积；对四部分面积分别求和可得到标枪沿标枪中轴线剖面面积，标枪表面积，见表 2

部分 面积	把手后	把手	把手前	枪尖	标枪
剖面面积 $S_0 (mm^2)$	33024	5467.6	21224	2990	62706
表面积 S (mm^2)	103750	17177	66677	9393.4	197000

表 2 标枪剖面面积及标枪表面积

通过编写 MATLAB 程序（见附录）计算得标枪形心位置 $Xp = 1537.3mm$ 。

六、 问题二模型的建立与求解

6.1 模型准备

对于该问题，根据运动学与力学原理进行受力分析及运动过程分析，进而找出标枪飞行过程中的运动规律。

首先，本文对标枪进行受力分析。在标枪飞行过程中一共受到重力 mg 、空气阻力 f 两个力，如图 8 所示。

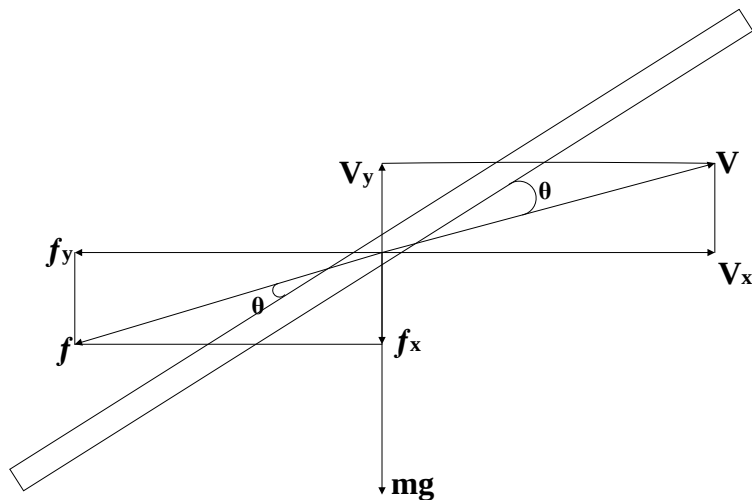


图 8 标枪飞行过程中受力分析

其中 f 表示标枪飞行过程中所受空气阻力大小，方向与标枪速度方向相反。空气阻力与飞行速度的关系为：

$$f = \frac{1}{2} c_x \rho s v^2 \quad (6-1-1)$$

式中， c_x 为空气阻力系数； ρ 为空气密度； s 为标枪迎风面积。

为了方便分析标枪的运动情况，本文建立了 xoy 直角坐标系，质心的运动轨迹如图 9 所示：

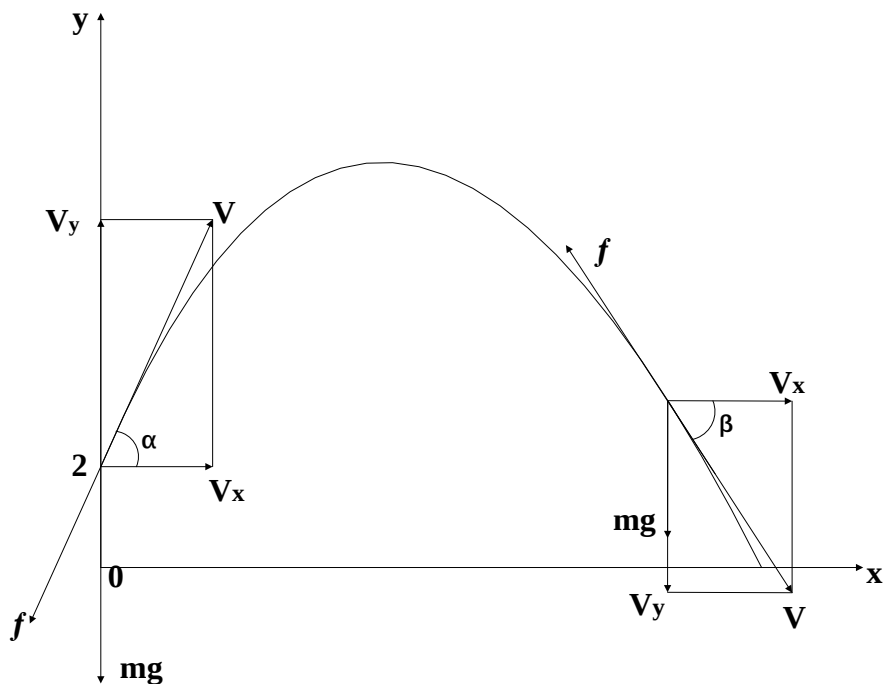


图9 标枪质心运动轨迹

6.2 模型建立

6.2.1 多元线性回归模型的建立

多元线性回归模型的一般形式为：

$$\begin{cases} y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 + \varepsilon \\ \varepsilon \sim N(0, \sigma^2) \end{cases} \quad (6-2-1)$$

其中， x_1 为出手速度， x_2 为出手角， x_3 为初始攻角。

为了方便，将 $n=24$ 位运动员的实测数据引入矩阵记号：

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} \varphi_1(\mu_1) & \varphi_2(\mu_1) & \varphi_3(\mu_1) \\ \varphi_1(\mu_2) & \varphi_2(\mu_2) & \varphi_3(\mu_2) \\ \vdots & \vdots & \vdots \\ \varphi_1(\mu_n) & \varphi_2(\mu_n) & \varphi_3(\mu_n) \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

6.2.2 多元线性回归模型的建立参数估计

选取 β 的一个估计值，记为 $\hat{\beta}$ ，使随机误差 ε 的平方和达到最小：

$$\min_{\beta} \varepsilon^T \cdot \varepsilon = \min_{\beta} (Y - X\beta)^T \cdot (Y - X\beta) = (Y - X\hat{\beta})^T \cdot (Y - X\hat{\beta}) \stackrel{\text{def}}{=} Q(\hat{\beta})$$

由最小二乘法的要求，由多元函数取得极值的必要条件可求解回归参数的标准方程如下：

$$\begin{cases} \left. \frac{\partial Q}{\partial \beta_0} \right|_{\beta_0=\beta_0} = 0 \\ \left. \frac{\partial Q}{\partial \beta_i} \right|_{\beta_i=\beta_i} = 0 (i=1,2,3) \end{cases} \quad (6-2-2)$$

可以证明，给定任意的 X, Y ，正规方程组总有解。虽然当 X 不满秩时，其解不唯一，但对任意一组解 $\hat{\beta}$ 都能使残差平方和最小，即 $Q(\hat{\beta}) = \min_{\beta} Q(\beta)$ 。

特别地，当 X 满秩时，即 $r(X) = r(X^T X) = p$ ，则正规方程组的解为 $\hat{\beta} = (X^T X)^{-1} X^T Y$ ，即为回归系数的估计值。

6.2.3 显著性差异检验

对于假设检验，其检验统计量的异常取值有2个方向，一般情况下，概率分布函数曲线两侧尾端的小概率事件都要考虑（即双尾检验）。

由于双侧检验不问差距的正负，所以给定的显著性水平 α ，需按正态对称分布的原理平均分配到左右两侧，每方各为 $\frac{\alpha}{2}$ ，相应得到下临界值为 $-\frac{Z\alpha}{2}$ ，上临界值为 $\frac{Z\alpha}{2}$ 。如图 10：

界值为 $\frac{Z\alpha}{2}$ 。如图 10：

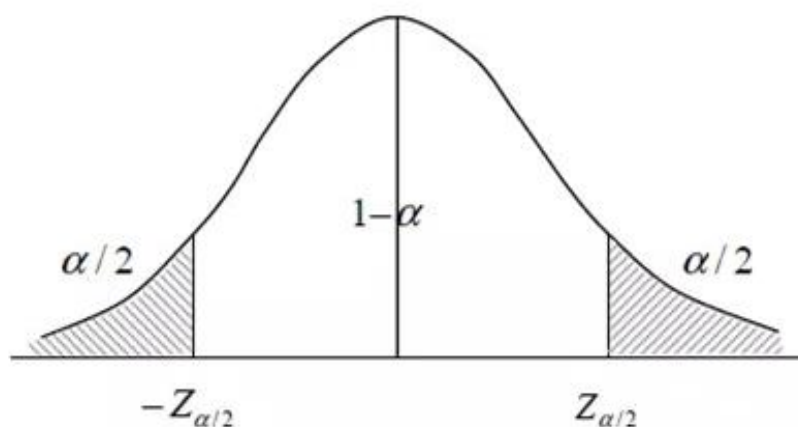


图 10 显著性水平分配

“显著性（双尾）”的值相关双变量除以二后，若小于 α ，则拒绝零假设，接受备择假设；若大于 α ，则接受零假设。

6.3 模型求解

6.3.1 多元线性回归模型的求解

基于 24 组实测数据，根据多元线性回归模型，利用软件 SPSS，求解的到回归方程： $L = \beta_1 v_0 + \beta_2 \alpha + \beta_3 \beta + \beta_4 + \varepsilon$ ，方程系数如表 3：

系数	β_1	β_2	β_3	β_4
值	4.309	0.477	-0.044	-60.542

表 3 回归系数

6.3.2 显著性差异检验的求解

提出假设：零假设 H_0 ：变量对投掷距离无显著影响；备择假设 H' ：变量对投掷距离显著影响。选定显著性水平： $\alpha=0.05$ 。利用软件 SPSS，根据显著性差异检验模型求得各变量与投掷距离差异显著性水平如表 4：

显著性系数	γ_1	γ_2	γ_3
值	0.000	0.015	0.646

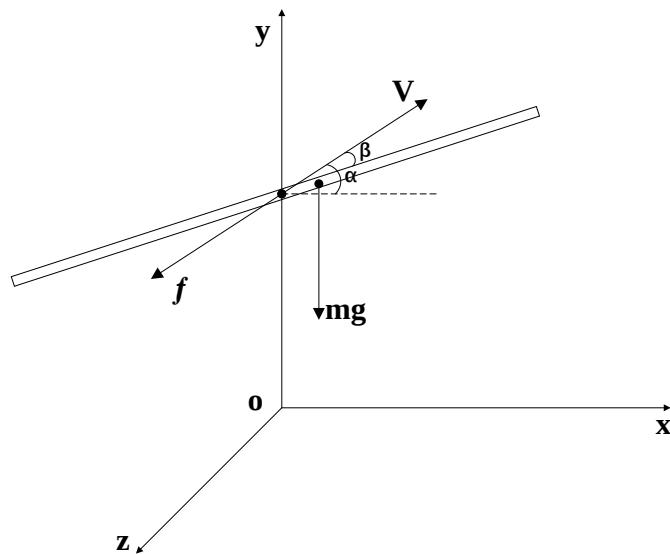
表 4 显著性系数

由表结果可知，初始攻角对投掷距离显著性水平大于 0.05，故接受原假设，即初始攻角对投掷距离影响不显著。

6.3.3 运动学与力学分析

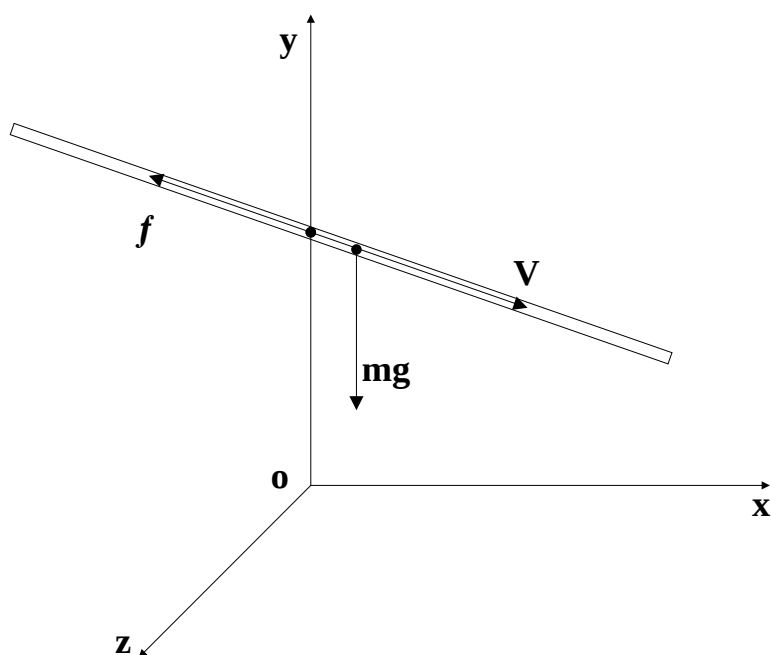
根据显著性差异检验可知初始攻角对投掷距离影响不显著，因此假设标枪长轴与标枪速度夹角为 0，即不考虑初始攻角。

对于标枪上升阶段，根据受力分析得到动力学方程组：



$$\left\{ \begin{array}{l} f = \frac{1}{2} c_x \rho s v^2 \\ a_{y_1} = \frac{mg + f_y}{m} \\ v_y = v \sin \alpha - a_{y_1} t \\ \frac{v_y}{a_{y_1}} = t_1 \\ h_1 = \frac{v_y^2}{2a_{y_1}} \\ H = 2 + h_1 \end{array} \right. \quad (6-3-1)$$

下降阶段，根据受力分析可得到动力学方程组：



$$\left\{ \begin{array}{l} a_{y_2} = \frac{mg - f_y}{m} \\ x = v \cos \alpha t - \frac{1}{2} \frac{f_x}{m} t^2 \\ h_1 = v \sin \alpha t - \frac{1}{2} a_{y_1} t^2 + 2 \\ h_2 = H - \frac{1}{2} \frac{mg - f \sin \beta}{m} t^2 \\ v_{y_1} = v \sin \alpha - \frac{mg + f \sin \beta}{m} t \\ v_{y_2} = \frac{mg - f \sin \beta}{m} t \end{array} \right. \quad (6-3-2)$$

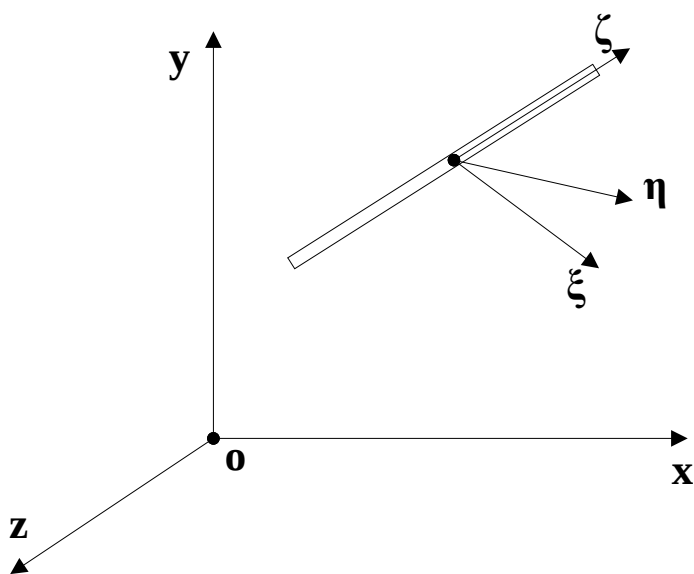
根据上述标枪上升阶段和阶段的动力学方程组，只需要给出初始速度和出手角度，可求得任意时刻标枪的任意位置，即定量反映了标枪的运动规律。

七、 问题三模型的建立与求解

7.1 模型准备

7.1.1 坐标系的建立

考虑到标枪质量较轻，受风面积较大，因此标枪在飞行运动过程中受到环境因素的影响较大。在运动过程中，标枪不仅做平动运动，自身还会绕重心转动，所以本文建立如图所示坐标系。 (x, y, z) 是固定在投掷处地面上的惯性坐标系， x 轴指向运动员投掷的正前方， y 轴垂直向上， z 轴指向其右方； (ξ, η, ζ) 是固定在标枪上的随体坐标系，原点在质心处， ζ 轴沿标枪纵轴向前指向 z 正方向， $\xi\eta\zeta$ 系不仅随标枪平动，也与标枪一起转动，它的各轴都将不断改变自己的方向。[2]



7.1.2 标枪的几何与物理参数

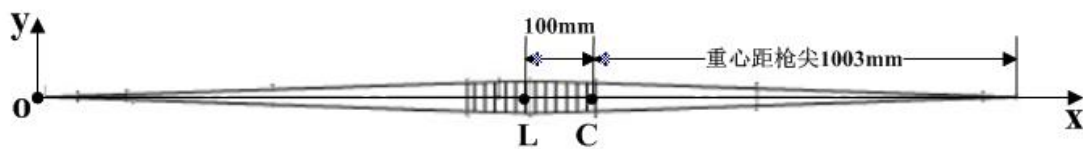


图 12 标枪几何示意图

标枪重量: $w = 0.8\text{kg}$

重心距枪尖: $X_c = 1.003\text{m}$

标枪绕通过重心的横轴转动: 转动惯量 $I_2 = 0.438(\text{kg} \cdot \text{m})$

标枪绕纵轴转动: 转动惯量 $I_1 = 2.692(\text{kg} \cdot \text{m})$

空气常数的选择:

空气密度: $\rho = 1.184 \times 10^{-3} \text{g} \cdot \text{cm}^{-3}$

切向摩擦系数 $C_\tau = 0.003$

法向阻力系数 $C_n = 1.2[3]$

风压中心位置: $X_p = \frac{S_1}{S_0} = 1368.99\text{mm}$

7.1.3 标枪受力分析

7.1.3.1 静系

标枪投掷出手瞬间, 手对标枪的作用力与标枪受到的阻力和重力抵消, 标枪所受合外力为 0, 即出手瞬间标枪只获得出手速度 V , 如图所示。

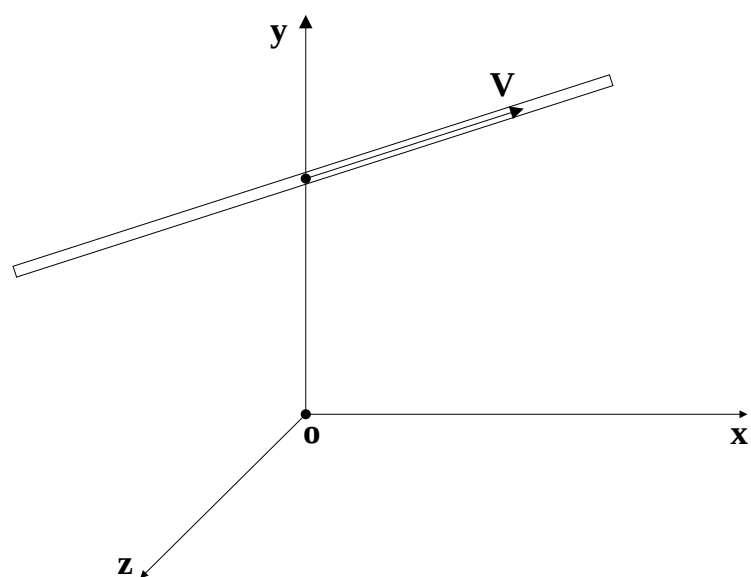


图 13 静系受力示意图

7.1.3.2 动系

标枪出手后，在飞行过程中受到重力、空气阻力及风作用力的作用。本文将空气摩擦力分解成沿标枪横轴方向的阻力与垂直于横轴方向的升力。升力 L 和阻力 F 并不作用在质心处，他们作用在标枪的风压中心 P 点。这样，升力 L 就会产生一个俯仰力矩 M ，它使标枪的方向不断变化，枪体向前翻转，枪头的方向由出手时指向前上方逐渐转成落地时指向前下方。受力情况如图所示。

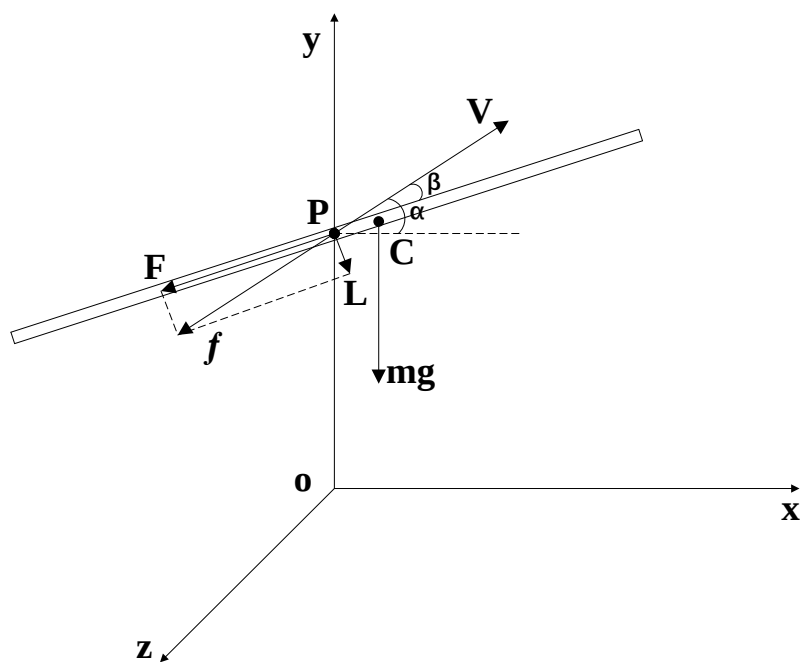


图 1411 动系受力示意图

<1>阻力与升力

将空气阻力可以分解成沿标枪切线方向的阻力 F 和法线方向升力 L ，其中：

1) 阻力 F 与空气粘性有关， 可以用下面的公式来描述：

$$F = C_r \cdot \frac{1}{2} \rho v^2 S \quad (7-1-1)$$

2) 升力与标枪受风面积和空气流速、密度有关可以用下面的公式来描述：

$$L = C_n \cdot \frac{1}{2} \rho v^2 S \quad (7-1-2)$$

〈2〉俯仰力矩

由于标枪在空气中运动时受到升力和阻力的共同作用，并且它们的合力作用点并不通过该物体的重心 C ，而是通过其压力中心 P ，这样就使得相对于重心的外力的合力矩不为零。从而在物体重心上产生一个使物体在垂直平面内上下俯仰的力矩，使得物体在飞行过程中产生抬头或低头现象。俯仰力矩可用下面的公式来描述：

$$M = C_m \frac{1}{2} \rho v^2 S \quad (7-1-3)$$

上述式 (7-1-1)、(7-1-2)、(7-1-3) 中， ρ 为空气的密度， v 为物体和气流的相对速度， C_r 为阻力系数， C_n 为升力系数， C_m 是俯仰力矩系数， S 是飞行物的特征面积。

7.2 模型建立

7.2.1 动力学及运动学分析

标枪在重力、阻力及升力三个力的共同作用下，标枪质心将遵循以下运动方程。

$$\begin{cases} L_x + F_x = ma_x \\ L_z + F_z = ma_z \\ L_y + F_y - W = ma \end{cases} \quad (7-2-1)$$

式中 m 是标枪的质量， a 是标枪的加速度；下标 x 、 y 、 z 表示力或加速度在该轴上的分量。由于重力 W 沿铅直方向所以它只有 y 轴方向的分量。

升力 L 和阻力 F 并不作用在质心 C 处，他们作用在标枪的压力中心 P 点，这样升力 L 就会产生一个俯仰力矩 M ，它使标枪的方向不断变化，枪体向前翻转，枪头的方向由出手时指向前上方逐渐转成落地时指向前下方。标枪方向用 ξ 、 η 、 ζ 坐标轴，它将遵循以下的方程组运动[2]：

$$\begin{cases} I\dot{\omega}_\xi + (I_0 - I)\omega_\tau \cdot \omega_\zeta = M_\xi \\ I\dot{\omega}_\tau + (I_0 - I)\omega_\xi \cdot \omega_\zeta = M_\eta \\ I_0\dot{\omega}_\zeta = M_\zeta \end{cases} \quad (7-2-2)$$

其中 I 和 I_0 分别是标枪对过质心的横轴和纵轴的转动惯量； ω 是标枪的角速度；下标 ξ 、 η 、 ζ 分别表示力矩或角速度在该轴上的分量，角速度在 $\xi \eta \zeta$ 系中的表达式是：

$$\begin{cases} \omega_{\xi} = \dot{\theta} \sin \psi - \dot{\phi} \sin \theta \cos \psi \\ \omega_{\eta} = \dot{\theta} \cos \psi + \dot{\phi} \sin \theta \sin \psi \\ \omega_{\zeta} = \dot{\phi} \cos \theta + \dot{\psi} \end{cases} \quad (7-2-3)$$

7.2.2 龙格-库塔算法

龙格-库塔 (Runge-Kutta) 算法是一种在工程上应用广泛的高精度单步算法。其理论基础来源于泰勒公式和使用斜率近似表达微分，它在积分区间多预计算出几个点的斜率，然后进行加权平均，用做下一点的依据，从而构造出了精度更高的数值积分计算方法。

龙格库塔算法一般形式如下：

$$\begin{cases} y_{n+1} = y_n + c_1 k_1 + c_2 k_2 + c_3 k_3 \dots c_n k_n \\ k_1 = hf(x_n, y_n) \\ k_2 = hf(x_n + a_2 h, y_n + \beta_{21} k_1) \\ k_3 = hf(x_n + a_3 h, y_n + \beta_{31} k_1 + \beta_{32} k_2) \\ \vdots \\ k_n = hf(x_n + a_n h, y_n + \beta_{n1} k_1 + \beta_{n2} k_2 \dots \beta_{nn-1} k_{n-1}) \end{cases} \quad (7-2-4)$$

本文采用龙格-库塔算法对牛顿-欧拉动力学方程进行值计算。

7.2.3 线性规划

实际中所研究的优化问题，一般都是要求使问题的某一项指标“最优”的方案，这类问题统称为最优化问题，解决这类问题的最常用方法就是线性规划方法。

针对问题三第(2)问，设出手速度、出手角及初始攻角分别为决策变量 x_1 、 x_2 、 x_3 ，最终投掷距离为 z ，则要解决的问题目标是使得（投掷距离）函数 z 取值最大。满足：

$$\begin{aligned} \max z &= f(x_1, x_2, x_3) \\ \text{s.t.} \begin{cases} 20^\circ \leq x_2 \leq 50^\circ \\ -15^\circ \leq x_3 \leq 15^\circ \end{cases} \end{aligned} \quad (7-2-5)$$

基于运动学及动力学分析，利用龙格-库塔算法，对投掷距离进行值计算。在已知标枪的特征参数及出手速度的前提下，通过线性规划模型进行数值计，计算出最佳决策变量。

7.3 模型求解

7.3.1 问题(1)的求解

已知投掷出手速度为29.70m/s、出手角为36.6°、初始攻角为-0.9°，利用龙格-库塔算法进行值计算，编写 MATLAB 程序（见附件3），可求得相应投掷距离为85.34505m。

7.3.2 问题(2)的求解

结合实际情况，优秀运动员在投掷标枪比赛中，攻角一般在 $[-10^\circ, 10^\circ]$ 范围内，出手角一般在 $[25^\circ, 45^\circ]$ 范围内。为消除其他情况干扰，本文将初始攻角限制于

$[-15^\circ, 15^\circ]$ 范围内, 将出手角限制于 $[20^\circ, 50^\circ]$ 范围内, 在决策变量范围内进行搜索, 找出使得标枪投掷最大的初始条件。

在运动员投掷出手速度为 30m/s 时, 求解得到最佳初始条件及投掷距离, 见下表。

最佳出手角	最佳初始攻角	最大投掷距离
42°	-8°	91.794m

表 5 最佳初始条件及投掷距离

八、 问题四模型的建立与求解

8.1 模型准备

动系受力分析: 标枪出手后, 物体获得初速度, 在飞行过程中受到重力、空气摩擦力及风力的作用。将空气摩擦力及风力分解成沿标枪横轴方向的阻力与垂直于横轴方向的升力。受力情况如图所示。

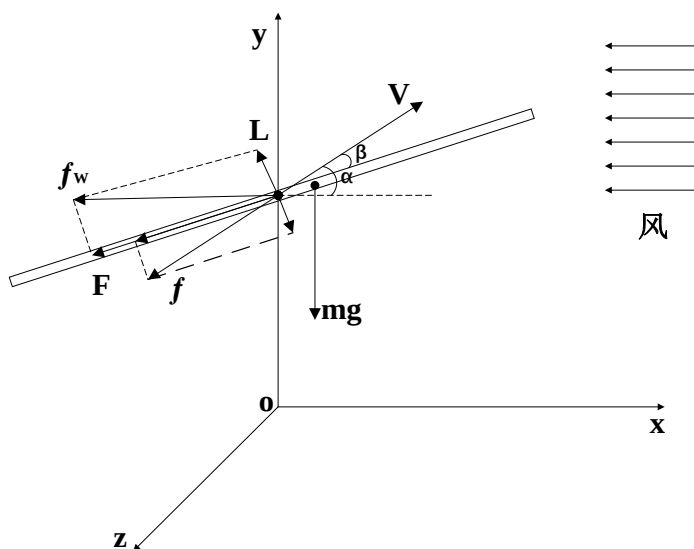


图 15 动系受力示意图

8.2 模型建立

该问题模型与问题三相似, 标枪所受阻力与升力由空气摩擦力分力及风对标枪作用力分力构成。基于问题三, 此处仅指出该问题与问题三不同处, 即线性规划中决策变量与目标函数的不同。

设出手速度、出手角、初始攻角、初始俯仰角速度及风速分别为决策变量 x_1 、 x_2 、 x_3 、 x_4 、 x_5 , 最终投掷距离为 z 。在已知 x_1 、 x_5 的前提下, 使得 (投掷距离) 函数 z 取值最大, 满足:

$$\begin{aligned} \max z &= f(x_1, x_2, x_3, x_4, x_5) \\ \text{s.t.} \quad &\begin{cases} 20^\circ \leq x_2 \leq 50^\circ \\ -15^\circ \leq x_3 \leq 15^\circ \\ -10 \leq x_4 \leq 10 \end{cases} \end{aligned} \quad (8-2-1)$$

利用线性规划模型对最佳出手条件进行求解。

8.3 模型求解

结合实际情况，优秀运动员在投掷标枪比赛中，攻角一般在 $[-10^{\circ},10^{\circ}]$ 范围内，出手角一般在 $[25^{\circ},45^{\circ}]$ 范围内。为消除其他情况干扰，本文将初始攻角限制于 $[-15^{\circ},15^{\circ}]$ 范围内，将出手角限制于 $[20^{\circ},50^{\circ}]$ 范围内，在决策变量范围内进行搜索，找出使得标枪投掷最大的初始条件。求解得到：

	顺风			逆风		
	3m/s	6m/s	9m/s	3m/s	6m/s	9m/s
最佳出手角（度）	40.91	41.21	41.74	42.13	42.27	42.42
最佳初始攻角（度）	-8.3	-8.9	-9.6	-7.7	-8.6	-9.2
最佳初始俯仰角速度（度/秒）	-0.22	-0.17	-0.11	-0.32	-0.24	-0.17
最大投掷距离（米）	101.360	101.879	102.523	101.217	101.248	101.137

表 6 不同风速下最佳出手条件及其最大投掷距离

九、 问题五模型的建立与求解

9.1 模型准备

9.1.1 投掷距离影响因素分析

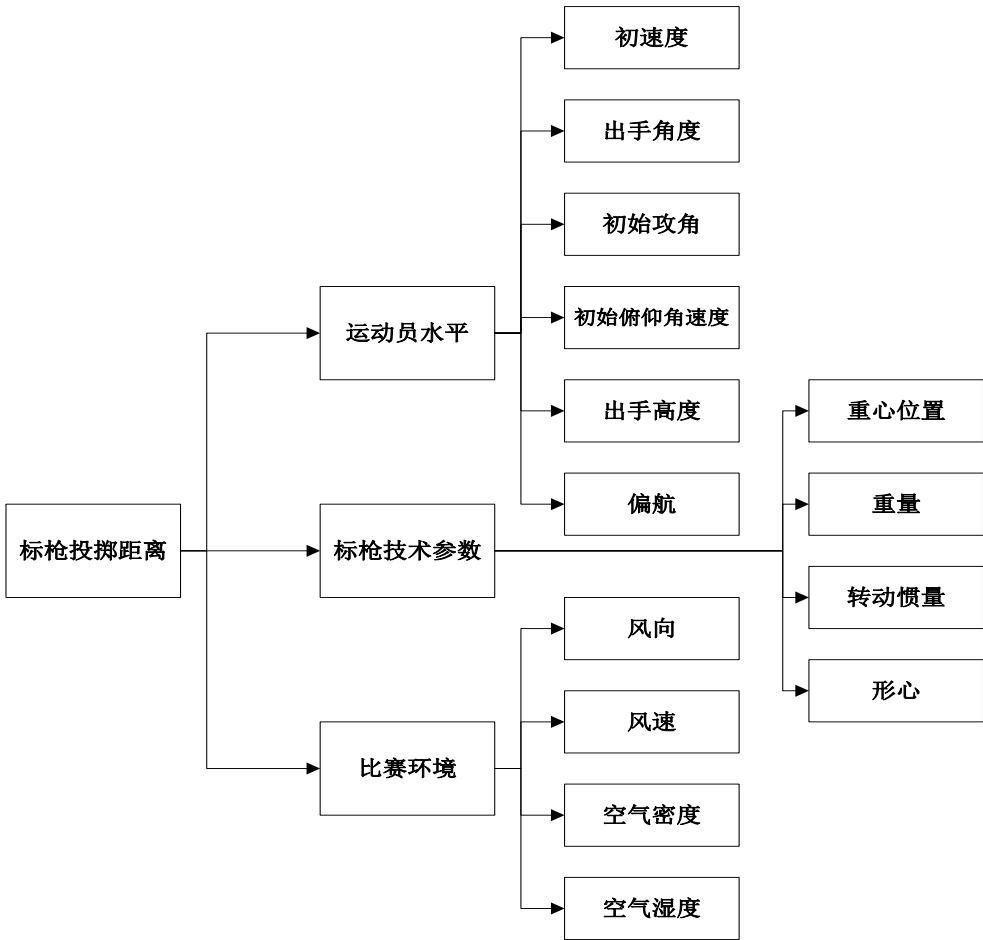


图 126 投掷距离影响因素

9.1.2 数据准备

基于问题四的标枪飞行模型，通过控制变量，对标枪投掷距离的影响因素进行数据微变，得到相应的投掷距离，如下表。

出手速度	26	27	28	29	30	31	32
投掷距离	70.55	75.62	80.90	86.30	91.99	97.78	103.08

表 7 出手速度对距离的影响

出手角	30	32	34	36	38	40	42
投掷距离	86.05	87.91	89.42	90.68	91.45	91.89	91.99

表 8 出手角对距离的影响

初始攻角	-15	-12	-9	-6	-3	0
投掷距离	90.45	91.41	91.56	91.12	90.44	89.42

初始攻角	3	6	9	12	15
投掷距离	88.54	87.91	87.63	87.67	87.67

表 9 初始攻角对距离的影响

初始俯仰角 速度	-20	-15	-10	-6	-3	0
投掷距离	72.83	74.83	79.16	80.76	80.95	80.57

初始俯仰角速度	3	6	10	15	20
投掷距离	80.04	79.78	78.16	76.69	75.09

表 10 初始俯仰角速度对距离的影响

风速	-6	-4	-2	-1	0
投掷距离	67.02	75.25	83.47	87.58	91.7

表 11 风速对距离的影响

9.2 模型建立

用最小二乘法的思想，按偏差平方和最小的原则并采取二项式方程拟合曲线函数。

6.3 模型求解

基于准备数据，以 MATLAB 为工具，分别对各影响因素与投掷距离关系进行拟合。

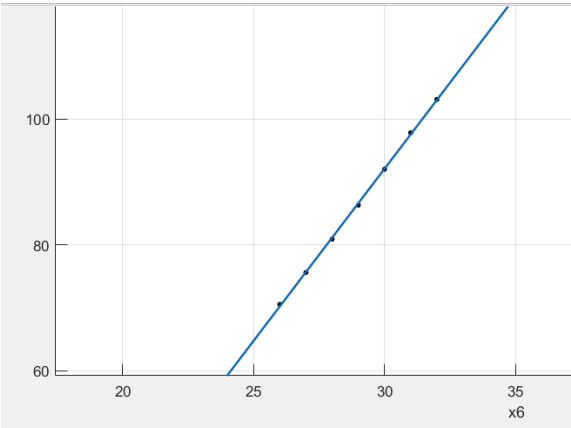


图 17 初始速度对投掷距离的影响拟合曲线

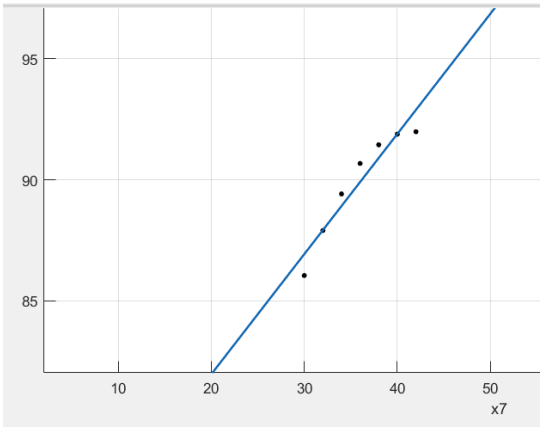


图 138 出手角对投掷距离的影响的拟合曲线

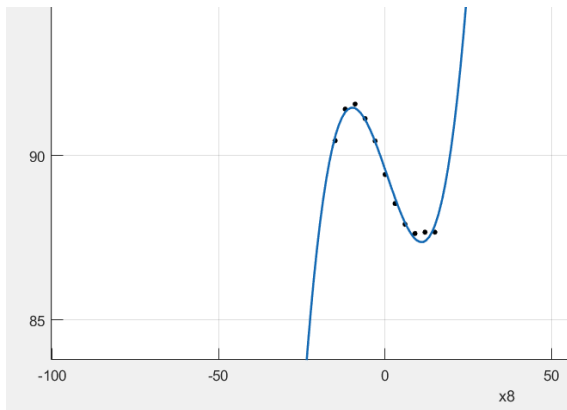


图 15 初始攻角对投掷距离的影响的拟合曲线

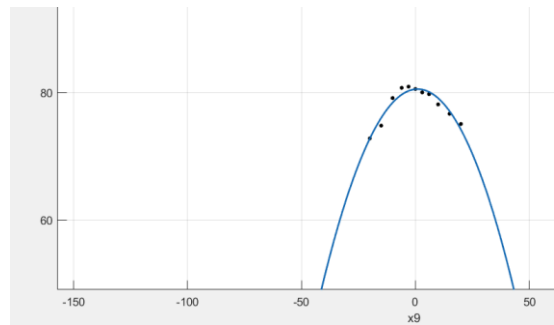


图 14 初始俯仰角速度对投掷距离的影响的拟合曲线

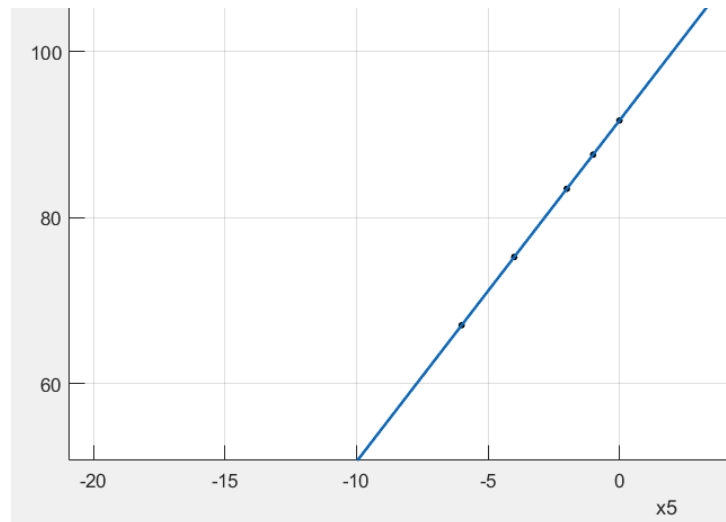


图 16 风速对投掷距离的影响的拟合曲线

若影响因素数据微变，投掷距离变化较大，即变化速度较快，则说明该影响因素对标枪投掷距离影响显著。通过对各曲线斜率的对比分析，得到决策变量相对重要性由大到小依次为：出手速度>出手角>初始攻角>风速>初始俯仰角速度。

十、 模型评价与推广

10.1 模型的优点

问题一立足于标枪散点数据的拟合曲线函数，结合标枪的轮廓特性，本文采用分段、二次多项式来拟合，观察拟合曲线图像和散点的相对位置可以发现：拟合所得曲线与真实数据十分逼近，说明本文的拟合曲线方法可信度较高、反映正是水平程度高。

在问题二上，本文从不同角度定性、定量分析标枪的飞行运动规律。不同方法之间既有相互辅助作用（如多元线性回归模型简化了物理运动模型），也有相互检

验作用，使得最后理论具有较强的说服力。

问题三和问题四的模型实质上是同一个模型，针对这个模型，本文根据牛顿-欧拉动力学理论较为全面的描述了标枪的运动，然后用高精度的龙格-库塔算法对模型进行值求解。再根据线性规划理论寻找问题中的最优解，具有较高的精度。

10.2 模型的改进

在问题二的物理模型中，为了简化模型，本文忽略了对标枪运动影响较小的攻角，但实际上，攻角的不同确实不可避免得会对标枪的运动产生一些影响。在简化模型的同时牺牲了模型的严谨性。

10.3 模型的推广

本文建立的标枪运动模型具有一定的推广价值，在已知运动员运投掷标枪初始条件的情况下，可以直接根据本模型，模拟出较为准确的投掷距离；此外，运动员可以借助本模型，固定某一个或某一些投掷初始条件，在剩余初始条件的自选范围内找出使得投掷距离最远的条件数据。本模型在标枪运动员训练具有良好的实用性。

十一、 参考文献

- [1][国家标准]GB/T 22765-2008-标枪
- [2]卢兆铭. 杨怀英. 杨娟华. 85/86 型标枪飞行特性的综合研究[J]. 天津体育学院学报, 1994, 9(2):36-42.
- [3]王其宁, 申毛毛, 陈耀庆, et al. 标枪气动力分析与初始条件的优化组合[J]. 四川体育科学, 1997(Z1):39-43.
- [4]刘晓军, 史振东, 马晋超, et al. 标枪气动力数值模拟[J]. 科学技术与工程, 2009, 9(11):3032-3035.
- [5]廖红, 张绪树, 陈维毅, et al. 考虑空气阻力时标枪最佳出手角度研究[J]. 中国体育科技, 2007, 43(1):98-101.
- [6]陈红英. 人体一标枪系统对标枪飞行初始条件的影响及标枪飞行的计算机模拟[J]. 太原理工大学, 2007.
- [8]黄国龙. 对决定标枪飞行远度相关因素的探析[J]. 河北农业大学学报(农林教育版), 2008, 8(2):184-186.
- [9]王倩. 影响标枪飞行远度的系统因素及投掷技术的综合分析[D]. 北京体育大学, 2000.
- [10]王倩. 标枪飞行轨迹的计算机仿真及实际应用[J]. 体育科学, 2001(1):73-78.

附录：

附录一：曲线拟合

$y1 = \text{polyfit}(x, y, N)$ ，这里函数 polyfit 第一个参数传递的是拟合数据的自变量，第二个参数是因变量，第三个参数是拟合多项式的阶数，这个由我们给定。

阶数确定方法：

这种方法方法可以取代用 cftool 进行判断的过程，具体实现代码如下：

```
for i=1:5
```

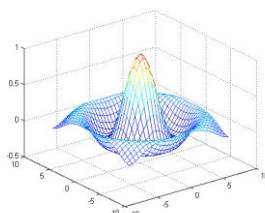
```
    y2=polyfit(x,y,i);
```

```
    Y=polyval(y2,x);%计算拟合函数在 x 处的值。
```

```
    if sum((Y-y).^2)<0.1
```

```
        c=i
```

五一数学建模竞赛



题目 标枪投掷的优化研究

关键词：标枪投掷距离 曲线拟合模型 多元回归模型 粒子群算法 动力学模型 模糊综合评判模型

摘要：

标枪投掷距离的远近受到运动员水平、标枪的技术参数、比赛环境三方面因素的影响，为了得到更远的投掷距离，研究出手速度、出手角、初始攻角、初始俯仰角速度、风向及风速等要素至关重要。因此本文对标枪投掷运动体系进行研究和优化。

针对问题一，要求算出所给型号标枪沿标枪中轴线剖面面积、标枪表面积和标枪形心的位置。本文建立以最小二乘法为基础的曲线拟合模型，对标枪外形曲线进行枪身、枪头、把手三者的分段拟合。首先画出所给数据散点图，然后建立曲线拟合方程，画出拟合曲线，最终得出结果：标枪剖面面积为 $6.352 \times 10^4 \text{mm}^2$ 、标枪表面积为 $2.098 \times 10^5 \text{mm}^2$ 、枪尾到形心距离为 1585mm 、形心到枪头为 1055mm 。

针对问题二，要求根据所给数据找出标枪飞行过程中的运动规律。本文建立多元回归模型分析投掷距离和出手速度、出手角、初始攻角三者之间的关系，得到回归方程为 $y = -60.5415 + 4.3088x_1 + 0.4769x_2 - 0.0443x_3$ ，投掷距离主要受出手角度和出手速度的影响。

针对问题三，要求建立标枪飞行的数学模型并计算相应条件下的投掷距离。本文在研究空气对运动物体的作用力、标枪飞行的受力情况的基础上，对标枪出手瞬间进行动力学分析及转动惯量的计算。对第一问进行抛体运动的研究，得出标枪的投掷距离为 90.4847m 。对第二问采用粒子群算法得出最佳的出手角为 37° ，初始攻角为 -0.17° ，最远的投掷距离为 90.3545m 。

针对问题四，要求在不同风向和风速的条件下建立标枪飞行的数学模型。本文先探讨空气动力学理论，为后文实际问题解决做铺垫；然后建立动力学模型，研究动量定理、动量矩定理、动能定理等，得出刚体平面运动微分方程；最后结合连续性方程和动量方程建立动网格守恒方程，从而得出问题结果，结果在表格中展示。

针对问题五，要求分析运动员出手速度、出手角、初始攻角、初始俯仰角速度、风向及风速等要素对标枪投掷距离影响的相对重要性。本文建立模糊综合评判模型得到模糊综合评价结果向量，再根据最大隶属度原则判断各影响因素与标枪投掷距离的关系，从而得到其相对关系：出手速度与投掷距离呈线性关系；标枪的最佳出手角为 42° ；最佳初始俯仰角速度在 $-0.17 \text{rad/s} - 0 \text{rad/s}$ 范围内；在顺风或逆风的情况下，风速在 $-6 \text{m/s} - 6 \text{m/s}$ ，投掷距离与风速成正比关系。

一 问题重述

1.1 问题背景

标枪的投掷是一项历史悠久的田径比赛项目。标枪投掷距离的远近受到运动员水平（出手速度、出手角、初始攻角、出手高度、出手时标枪的初始俯仰角速度等），标枪的技术参数（标枪的长度、重量、几何形状、重心的位置、形心的位置等）和比赛环境（空气的密度与粘度、风力、风向等）三方面因素的影响。因此，对标枪的构造、标枪的飞行过程，以及出手速度、出手角、初始攻角、初始俯仰角速度、风向及风速等要素对标枪投掷距离影响的分析至关重要。

1.2 问题提出

（1）问题一要求根据现有某型标枪的测量尺寸数据（见附件），估算出该型标枪沿标枪中轴线剖面面积、标枪表面积和标枪形心的位置。

（2）问题二要求根据现有某标枪比赛中 24 名运动员使用同型号标枪投掷的实测数据（见附件）。对这些数据进行分析，建立合适的数学模型，找出标枪飞行过程中的运动规律。

（3）问题三要求，假设某型标枪的几何参数同问题一，重心位于形心前 10cm （重心在枪尖与形心之间），出手时标枪的初始俯仰角速度为 0 。在无风的前提下，对标枪投掷出手瞬间、出手后的受力及运动情况进行分析，建立标枪飞行的数学模型并求解如下问题：

①假设某运动员投掷出手速度为 29.70m/s 、出手角为 36.6° 、初始攻角为 -0.9° ，请估算出标枪的投掷距离。

②假设某运动员投掷出手速度为 30m/s ，请给出最佳的出手角和初始攻角使得投掷距离最大并估算出标枪的投掷距离。

（4）问题四要求，假设标枪技术参数同问题 3，风向分别为顺风 and 逆风，风速分别为 3m/s ， 6m/s 和 9m/s ，运动员投掷出手速度为 31.70m/s 。建立标枪飞行的数学模型，给出最佳出手角、最佳初始攻角、最佳初始俯仰角速度使得投掷距离最大，并填写表。

（5）问题五要求，标枪技术参数同问题 3，分析运动员出手速度、出手角、初始攻角、初始俯仰角速度、风向及风速等要素对标枪投掷距离影响的相对重要性。

二 问题分析

标枪投掷距离的远近受到运动员水平（出手速度、出手角、初始攻角、出手高度、出手时标枪的初始俯仰角速度等），标枪的技术参数（标枪的长度、重量、几何形状、重心的位置、形心的位置等）和比赛环境（空气的密度与粘度、风力、风向等）三方面因素的影响。为了得到更远的投掷距离，研究出手速度、出手角、初始攻角、初始俯仰角速度、风向及风速等要素至关重要。

针对问题一，要求算出所给型号标枪沿标枪中轴线剖面面积、标枪表面积和标枪形心的位置。为了更加精确地反映标枪外形曲线，本文对这些数据进行插值与拟合处理；建立以最小二乘法为基础的曲线拟合模型，对标枪外形曲线进行枪身、枪头、把手三者的分段拟合，这样让拟合效果更加精准。首先画出所给数据散点图，然后建立曲线拟合方程，画出拟合曲线，因为所呈现的出来的曲线形状，大致为多项式函数曲线一部分；

再结合实际标枪外形大致轮廓,本文选择多项式曲线对枪身轮廓进行拟合。最后得出结果:标枪割面积为 $6.352 \times 10^4 \text{mm}^2$ 、标枪表面积为 $2.098 \times 10^5 \text{mm}^2$ 、枪尾到形心距离为1585mm、形心到枪头为1055mm。

针对问题二,要求根据所给数据找出标枪飞行过程中的运动规律。本题是对数据进行分析,要找出投掷距离和出手速度、出手角、初始攻角三者之间的关系,因此建立多元回归模型并且进行检验。最终得到回归方程为 $y = -60.5415 + 4.3088x_1 + 0.4769x_2 - 0.0443x_3$,投掷距离主要受出手角度和出手速度的影响。

针对问题三,要求在无风的前提下对标枪投掷出手瞬间、出手后的受力及运动情况进行分析,建立标枪飞行的数学模型,并计算相应条件下的投掷距离。本文首先研究空气对运动物体的作用力、标枪飞行的受力情况;然后对标枪出手瞬间进行动力学分析,及采用扭摆法进行转动惯量的计算。对于第一问,因标枪运动轨迹类似于抛体运动,因此采用运动学的斜抛运动规律解决问题,得出标枪的投掷距离为11.4847m。对于第二问,要求得到最佳的出手角和初始攻角,并且估算出标枪的投掷距离。因此采用粒子群算法,得出最佳的出手角为 42° ,初始攻角为 -0.2° ,最远的投掷距离为11.3545m。

针对问题四,要求在不同风向和风速的条件下建立标枪飞行的数学模型。因为要考虑空气阻力对标枪的作用,所以需要结合空气动力学理论,研究流体运动,为后文实际问题解决做铺垫;然后建立动力学模型,研究动量定理、动量矩定理、动能定理等,得出刚体平面运动微分方程;最后结合连续性方程和动量方程建立动网格守恒方程,从而得出问题结果,结果在后文表格中展示。

针对问题五,要求分析运动员出手速度、出手角、初始攻角、初始俯仰角速度、风向及风速等要素对标枪投掷距离影响的相对重要性。本文建立模糊综合评判模型得到模糊综合评价结果向量,再根据最大隶属度原则判断各影响因素与标枪投掷距离的关系,从而得出结论如下:

(1) 出手速度与标枪的投掷距离基本上呈线性关系。

(2) 标枪的出手角在 $0^\circ - 42^\circ$ 时,随着出手角的增加,标枪的投掷距离线性增加,超过 42° 以后,其射程略有减小。因此,标枪的最佳出手角为 42° 。

(3) 初始俯仰角速度小于 0 rad/s 时标枪飞行的远度要更远一些。最佳初始俯仰角速度大约在 $-0.17 \text{ rad/s} - 0 \text{ rad/s}$ 范围内。

(4) 风向及风速对标枪的投掷距离的影响显著。在顺风或逆风的情况下,风速在 $-6 \text{ m/s} - 6 \text{ m/s}$,投掷距离与风速成正比关系。

三 问题假设

(1) 投射角度 α 与投射初速度 v 是两个相互独立的量。

(2) 忽略标枪在飞行过程中的进动影响。

(3) 重力加速度 g 取值为 9.8 m/s^2 ,并在投掷的任意点都相等。

(4) 标枪运动轨迹在同一平面内,且地面处处水平。

四 符号说明

符号	说明
β	标枪初始攻角
α	标枪出手角度
γ	持枪角
v_0	标枪初速度
h	标枪出手高度
V_m	风速
S	标枪飞行距离
F_d	空气阻力
M	俯仰力矩
C_d	阻力系数
$S_{\text{表面积}}$	标枪表面积
$S_{\text{割面积}}$	标枪割面积
$X_h Y_h$	约束反力

五 模型的建立与求解

5.1 问题一的求解

由于某型号标枪测量尺寸表有较多的离散的测量数据，因此，需要对这些数据进行插值与拟合处理。其目的是为了从这些数据中寻找它们反映出来的规律，即通过已知的数据去确定某一类已知函数的参数或找某个近似函数，使所得到的近似的数对已知数据有较高的拟合度，更加精确地反映标枪外形曲线。标枪外形曲线如下图所示：

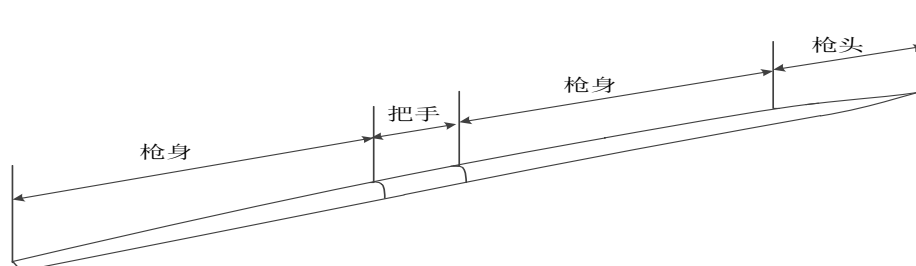


图 1 标枪外形图

5.1.1 曲线拟合模型

根据现有某型标枪的测量尺寸数据，估算出该型标枪沿标枪中轴线剖面面积、标枪表面积和标枪形心的位置。为了更清楚认识枪身、枪头、把手三者对应直径的关系，本文建立曲线拟合模型，利用 MATLAB 的拟合工具箱，画出拟合曲线。

(1) 插值

在某模型标枪给定 n 个离散数据点（称为节点）， (x_{ik}, y_k) ， $k = 1, 2, \dots, n$ 。对于 x_i 且

$x_i \neq x_{ik}$, 求 x_i 所对应的 y 的值称为内插。

$f(x)$ 为定义在给定区间上的函数, $x_{11}, x_{12}, x_{13}, \dots, x_{in}$ 为区间上 n 个互不相同的点, 若函数 $g(x)$ 满足:

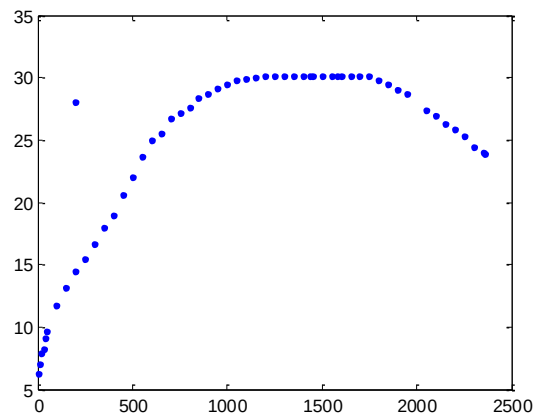
$$g(x_{ik}) = f(x_{ik})$$

则称 $g(x_{ik})$ 为 $f(x_{ik})$ 关于节点在区间上的插值函数。

(2) 拟合

根据观察已有数据, 以及标枪外形基本特征可知, 标枪大致分为三部分: 枪身、枪头、把手。枪身部分直径逐步增大, 枪头部分直径逐步减少, 把手部分直径不变。为了使拟合效果最佳, 本文按标枪外形分段拟合, 采用最小二乘法^[1]在曲线中的应用建立曲线拟合模型。

① 根据已有数据绘制枪身与对应直径的散点图:



观察枪身数据及所呈现的出来的曲线形状, 大致为多项式函数曲线一部分; 再结合实际标枪外形大致轮廓, 本文选择多项式曲线对枪身轮廓进行拟合。

② 拟合曲线方程

多项式函数表达式的建立:

$$f(x) = p_1x^3 + p_2x^2 + p_3x + p_4$$

根据最小二乘法原理有

$$M = \sum_{i=1}^n (f(x) - p_1x^3 + p_2x^2 + p_3x + p_4)$$

令 p_i 合理取值, 使得 M 取值最小, 即 $\frac{\partial M}{\partial p_i} = 0$ 时 M 可以取到最小值。

$$\frac{\partial M}{\partial p_i} = \sum_{i=1}^n (f(x) - 3p_1x^2 + 2p_2x + p_3 + p_4)$$

令上式为零, 解关于 p_i 的方程组, 运用 MATLAB 工具箱解以上正弦函数表达式的系数, 进而得出拟合曲线方程:

$$f(x) = 1.039x^3 - 6.074x^2 + 3.625x + 30.16$$

拟合曲线如下图图 2 所示:

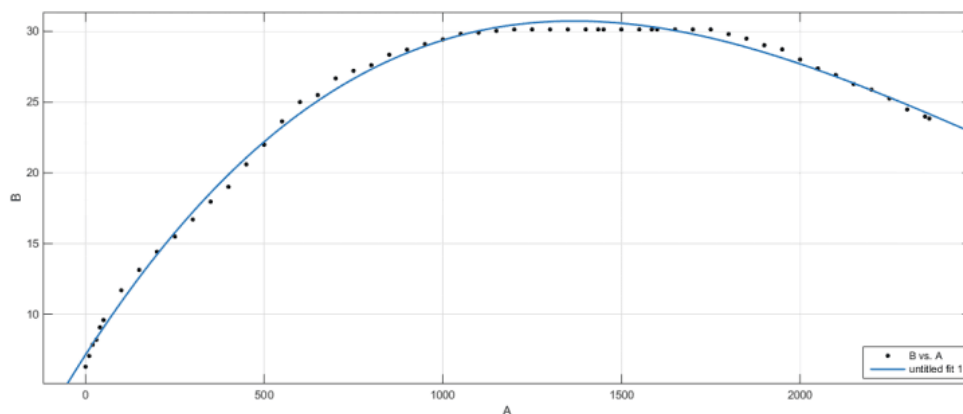
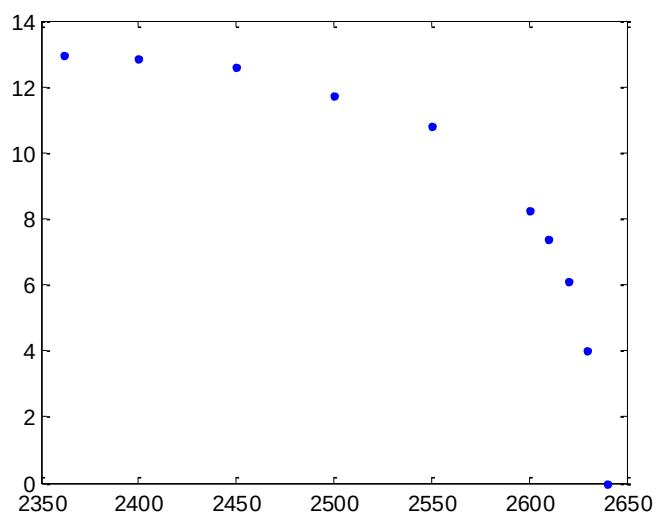


图 2

拟合度为 97.84%，拟合效果较好。

③ 同理，根据枪头与对应的直径绘制散点图：



建立多项式函数表达式

$$f(x) = p_1x^4 + p_2x^3 + p_3x^2 + p_4x + p_5$$

解得拟合函数为

$$f(x) = -2.12x^4 - 5.12x^3 - 2.436x^2 - 0.5645x + 11.46$$

拟合曲线如下图图 3 所示：

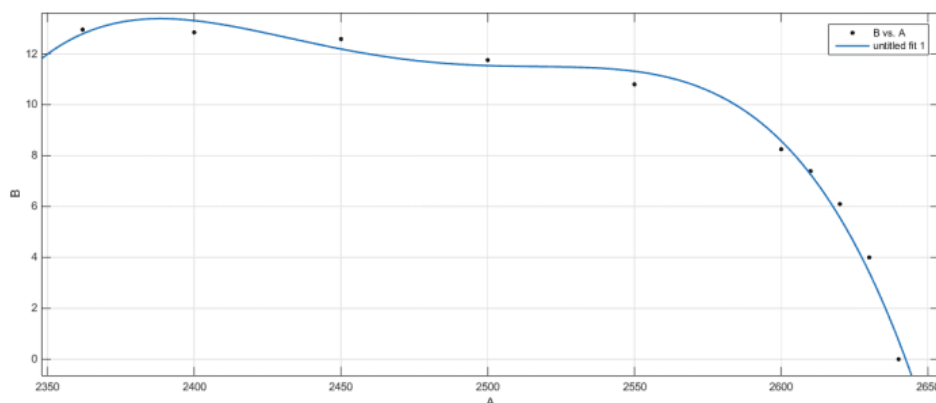


图 3

拟合度为 98.82%，拟合效果较好。

④ 由于把手部分直径不变割面积可直接求出，因此不需要进行拟合曲线。

5.1.2 割面积，表面积，形心的求解

由以上对已有数据的插值拟合，近似描述标枪外形曲线，得到的拟合方程，再利用辛普森梯形数值积分和 MATLAB 软件积分语句，求解各面积。

$$S_{\text{割面积}} = \int_0^L D(x) dx$$

$$S_{\text{表面积}} = \int_0^L \pi D(x) \sqrt{1 + \left(\frac{D'(x)}{2}\right)^2} dx$$

$$L_{\text{形心}} = \int_0^L D(x) x dx / S_{\text{割面积}}$$

经计算得出结果：

标枪割面积、标枪表面积分别为：

$$6.352 \times 10^4 \text{mm}^2, 2.098 \times 10^5 \text{mm}^2$$

枪尾到形心：

$$l = 1585 \text{mm}$$

形心到枪头：

$$l_0 = 1055 \text{mm}$$

5.2 问题二的求解

问题二是对某标枪比赛中 24 名运动员使用同型号标枪投掷的实测数据进行分析，要求建立合适的数学模型，找出标枪飞行过程中的运动规律。

附件中所给项目有选手的出手速度、出手角、初始攻角以及投掷距离，为找出其运动规律，我们需要根据这些数据，找出投掷距离和出手速度、出手角、初始攻角三者之间的关系。因此，本文对数据进行线性回归分析^[2]，建立多元回归模型解决问题。

5.2.1 多元回归模型

(1) 定义

多元回归模型是用来进行回归分析的数学模型(含相关假设), 称为回归模型。只含有一个回归变量的回归模型称为一元回归模型, 否则称为多元回归模型。

设因变量为 y , k 个自变量分别为 $x_1, x_2, \dots, x_k$, 描述因变量 y 如何依赖于自变量 $x_1, x_2, \dots, x_k$ 和误差项 ε 的方程称为多元回归模型。其一般形式可表示为: $y = B_0 + B_1x_1 + B_2x_2 + \dots + B_kx_k + E$ 。式中, $B_0, B_1, B_2, \dots, B_k$ 是模型的参数; ε 为误差项。

多元回归分析是指分析因变量和自变量之间关系, 回归分析的基本思想是: 虽然自变量和因变量之间没有严格的、确定性的函数关系, 但可以设法找出最能代表它们之间关系的数学表达形式。其分析步骤如下:

a. 确定几个特定的变量之间是否存在相关关系, 如果存在的话, 找出它们之间合适的数学表达式;

b. 根据一个或几个变量的值, 预测或控制另一个变量的取值, 并且可以知道这种预测或控制能达到什么样的精确度;

c. 进行因素分析。例如在对于共同影响一个变量的许多变量(因素)之间, 找出哪些是重要因素, 哪些是次要因素, 这些因素之间又有什么关系等等。

(2) 模型建立

本文通过分析出手速度、出手角、初始攻角(自变量)和投掷距离(因变量)之间的相关性, 来建立出手速度、出手角、初始攻角和投掷距离之间关系的数学模型。其多元回归模型建立如下:

$$Q = \sum_{i=1}^n [y_i - (b_0 + b_1x_{1i} + \dots + b_kx_{ki})]^2$$

Q 为实测值与 y_i 线性函数 $\hat{y} = b_0 + b_1x_{1i} + \dots + b_kx_{ki}$ 的残差平方和, 使 $Q = (\widehat{b_0}, \widehat{b_1}, \widehat{b_2}, \dots, \widehat{b_k}) = \min(b_0, b_1, \dots, b_k)$

求 Q 关于 $b_0, b_1, \dots, b_k$ 的偏导数, 令其为0。即 $\frac{\partial Q}{\partial b_i} = 0, i = 1, 2, \dots, k$ 。经整理为:

$$\begin{aligned} nb_0 + b_1 \sum_{i=1}^n x_{1i} + \dots + b_k \sum_{i=1}^n x_{ki} &= \sum_{i=1}^n y_i \\ b_0 \sum_{i=1}^n x_{1i} + b_1 \sum_{i=1}^n x_{1i}^2 + \dots + b_k \sum_{i=1}^n x_{1i}x_{ki} &= \sum_{i=1}^n x_{1i}y_i \\ &\dots \\ b_0 \sum_{i=1}^n x_{ki} + b_1 \sum_{i=1}^n x_{1i}x_{ki} + \dots + b_k \sum_{i=1}^n x_{ki}^2 &= \sum_{i=1}^n x_{ki}y_i \end{aligned}$$

上式称为多元线性回归的正规方程组, 令

$$X_{n(k+1)} = \begin{bmatrix} 1 & \dots & x_{k1} \\ \vdots & \ddots & \vdots \\ 1 & \dots & x_{kn} \end{bmatrix} Y = \begin{bmatrix} y_1 \\ \dots \\ y_n \end{bmatrix} B = \begin{bmatrix} b_0 \\ \dots \\ b_k \end{bmatrix}$$

则由矩阵运算法则, 正规方程组的矩阵形式为:

$$X^T X \hat{\beta} = X^T Y$$

假设 $(X^T X)^{-1}$ 存在，上式两边乘以 $(X^T X)^{-1}$ 得：

$$\hat{\beta} = [\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k]^T = (X^T X)^{-1} X^T Y$$

最后解得k元线性方程为：

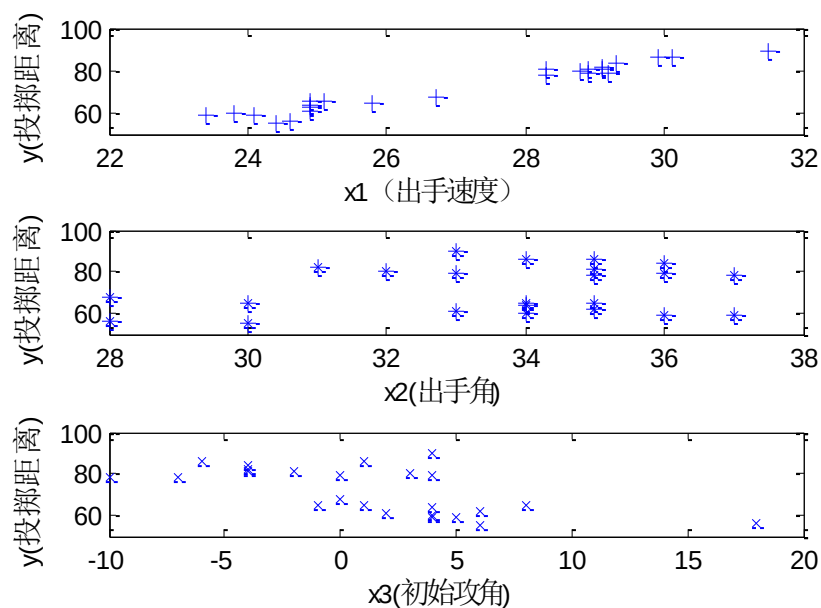
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_k x_k$$

结合实际 24 名运动员实测数据进行回归分析，参数如下表表 1 回归分析参数表：

表 1 回归分析参数表

x_1 出手速度(米/秒)	x_2 出手角(度)	x_3 初始攻角(度)	y投掷距离(米)
31.5	33	4	89.58
30.1	34	1	86.3
29.9	35	-6	86.08
29.3	36	-4	83.8
29.1	31	-4	82.04
29.1	35	-4	81.06
28.9	35	-2	81
28.3	32	3	80.76
28.8	33	4	79.58
29.2	36	0	79.14
28.9	35	-7	78.76
28.3	37	-10	78.16
26.7	28	0	67.56
25.1	35	-1	65.22
24.9	34	8	65.16
25.8	30	1	65.04
24.9	34	4	63.74
24.9	35	6	62.32
24.9	33	2	60.8
23.8	34	4	59.82
23.4	37	4	59.1
24.1	36	5	58.84
24.6	28	18	56.5
24.4	30	6	55.56

由 MATLAB 进行回归：



得出回归方程为：

$$y = -60.5415 + 4.3088x_1 + 0.4769x_2 - 0.0443x_3$$

由多元线性回归参数进行多元回归分析，可以看出本模型的建立是可靠的，与实际情况相符。根据回归模型的计算得出拟合值与实际值及残差关系，如表 2：

表 2

实测值	拟合值	残差
89.58	89.45	-1.165
86.3	86.7	0.9775
86.08	86	0.8321
83.8	83.5	0.7492
82.04	82.07	2.2352
81.06	82.01	-0.6522
81	80.9	0.2382
80.76	80.56	4.2357
79.58	78.8	0.4687
79.14	79.21	-3.3026
78.76	78.66	-2.2234
78.16	78.89	-1.3248
67.56	67.8	-0.2958
65.22	65.43	0.876

65.16	65.34	2.5535
65.04	65.04	0.1528
63.74	63.67	0.9563
62.32	62.33	-0.852
60.8	60.21	-1.5955
59.82	60.01	1.776
59.1	59.2	1.3489
58.84	58.32	-1.4061
56.5	56.78	-1.5094
55.56	55.35	-3.0733

相关系数分析：

	x_1	x_2	x_3
y	0.43088	0.4769	-0.0443

由上表可知 x_2 ， x_1 对y投掷距离影响较大，因此投掷距离主要受出手角度和出手速度的影响。

(3) 模型检验

总离差平方和

$$TSS = \sum (y_i - \bar{y})^2$$

回归平方和

$$ESS = \sum (\hat{y}_i - \bar{y})^2$$

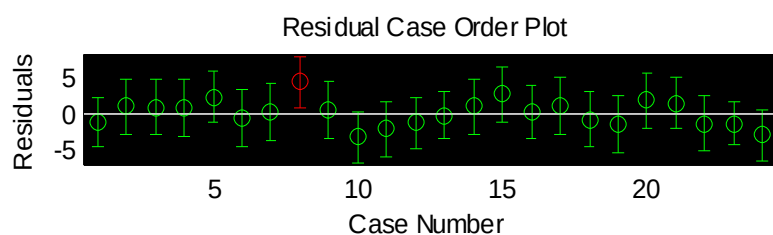
剩余平方和

$$RSS = \sum (y_i - \hat{y}_i)^2$$

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

该统计量越接近于 1，模型的拟合优度越高。

残差图：



经 MATLAB 计算得:

$$TSS = 2.8303e + 03 \quad RSS = 1.5247e + 13 \quad ESS = 1.5247e + 13$$

可决系数为 0.9735, 因此该模型的拟合优度较好。

5.3 问题三的求解

假设某型标枪的几何参数同问题 1, 重心位于形心前 10cm (重心在枪尖与形心之间), 出手时标枪的初始俯仰角速度为 0。在无风的前提下, 本文对标枪投掷出手瞬间、出手后的受力及运动情况进行分析, 建立标枪飞行的数学模型。

5.3.1 空气对运动物体的作用力

空气对在其内部运动的物体的作用, 归结为在物体表面上连续分布的法向力和切向力的作用, 这些力可简化为一个空气动力主向量 $\vec{F}$ 以及这些力对任意参考点的力矩 $\vec{M}$ 。

空气动力方向量 $\vec{F}$ 在气流坐标系 x, y, z 各轴上的投影, 分别称为阻力 X 、举力 Y 和仰力 Z 。主向量 $\vec{F}$ 在飞行器连体坐标系 x_1, y_1, z_1 各轴上的相应投影称为轴向力 X_1 、法向力 Y_1 和横向力 Z_1 。对任意参考点的空气动力方矩 $\vec{M}$ 在任一坐标系中的投影名称都是相同的, 对纵轴的分力矩称为滚转力矩, 对垂直轴的分力矩称为偏航力矩, 对横轴的分力矩称为俯仰力矩。当物体的运动速度给定时, 气动力和力矩的大小及作用方向与物体相对于飞行速度向量 $\vec{V}_\infty$ 的方位有关, 也就是说, 同攻角 α 以及偏航角 β (向量 $\vec{V}_\infty$ 和 $x_1\sigma y_1$ 平面的夹角) 有关。标枪在空气当中的受力如下图图 4 空气中的受力图所示:

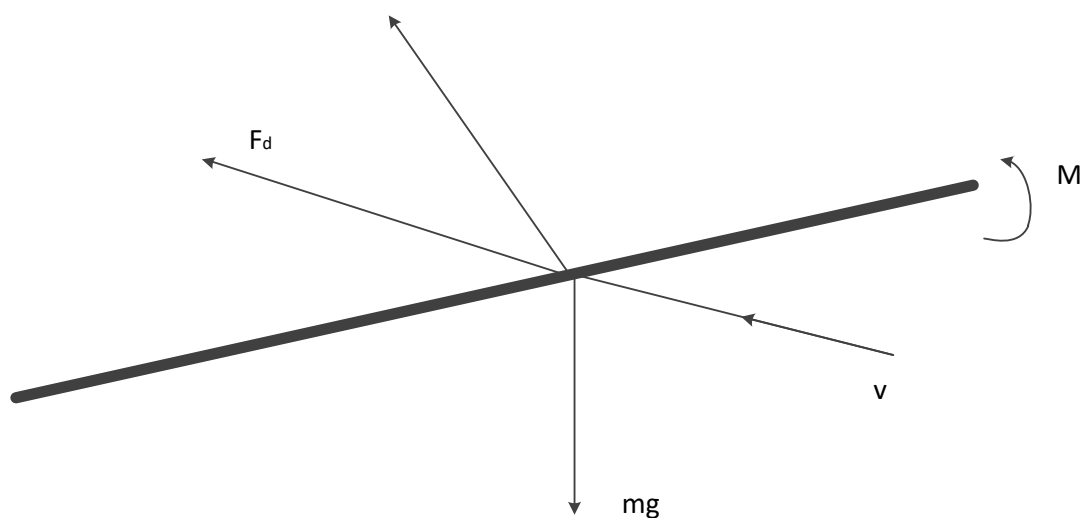


图 4 空气中的受力图

因为标枪受升力影响较小, 可以忽略不计。因此本文重点研究阻力对标枪运动的影响。

空气的阻力是由气流绕过物体流动所引起的切向应力和压强差造成的, 一般可以分为摩擦阻力、压差阻力两种。摩擦阻力与空气的粘性有关; 压差阻力与物体的形状有一定的关系。因为空气是有粘性的, 当气流流过物体表面时, 紧贴物体表面的一层空气就会粘在表面上, 这层空气的流速为零, 随着离该表面距离的增大, 相对应的空气层的流速也将增大, 直到某-距离后速度才又等于自由流速。这就使得气流在内外层之间有速度差, 使得物体在空气中运动时受到阻力的作用。我们把产生的这种阻力称为摩擦阻力。当空气流经不同形状的物体表面时, 由于物体形状和空气粘性等因素的影响, 使气流速度在物体的前后发生变化而引起压强的变化而形成了压差阻力。可以用下面的公式来描

述:

$$F_d = C_d \cdot \frac{1}{2} \rho v^2 S$$

式中， ρ 是空气的密度， v 是物体和气流的相对速度， C_d 是阻力系数， S 是飞行物的特征面积，可以取标枪的表面积或标枪纵截面积，阻力系数是个无量纲的比例系数。

5.3.2 标枪飞行的受力情况

① 标枪飞行时的气动力分析

当人体将标枪从手中投掷出去以后，由于标枪在空气中运动时受到空气的作用力，致使标枪的飞行远度受到很大的影响。以下对标枪飞行的模拟均是在考虑空气作用力的情况下进行的。下图图 5 气动力示意图是标枪飞行时的气动力示意图：

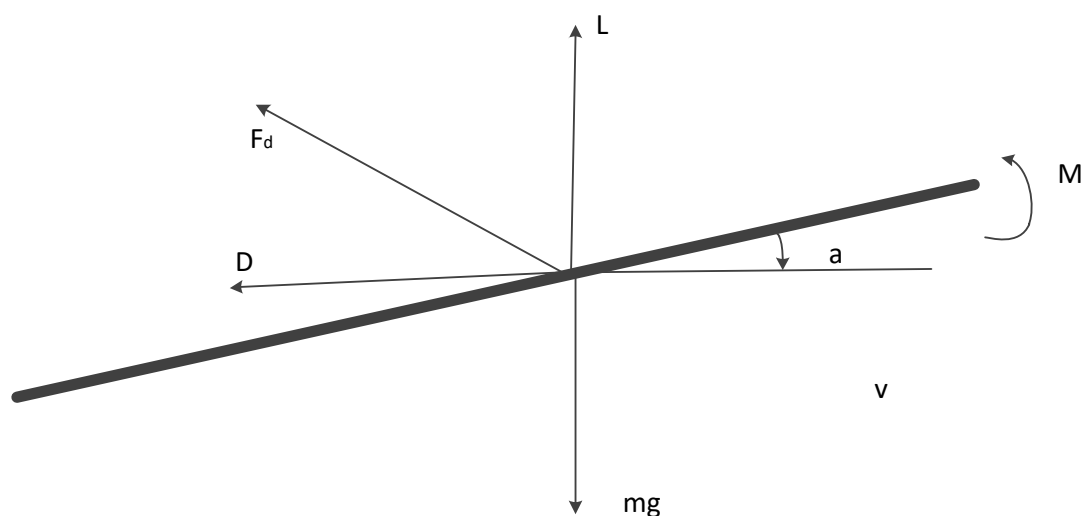


图 5 气动力示意图

其中， D 是标枪受到的空气阻力， L 是标枪受到的空气升力， R 是气动力的合力， G 是标枪的重力， v 是来流的速度。标枪飞行时，由于气动力合力的作用点(压力中心 P)偏离标枪的重心，与重心不在同一位置，因而会产生使标枪俯仰(抬头或低头)的力矩 $\vec{M}$ ，称之为俯仰力矩，并对标枪飞行的轨迹产生较大的影响。

② 出手瞬间动力学分析-多刚体系统的力学模型

为了更深入了解标枪投掷时的运动情况，理解标枪投掷的技术原理，本文将从运动学的角度对标枪投掷出手瞬间、出手后受力和标枪飞行的轨迹做分析。

从运动员起跑处至标枪投掷瞬间这一过程，选取出手瞬间进行动力学推导，如图 6 动力学分析图：

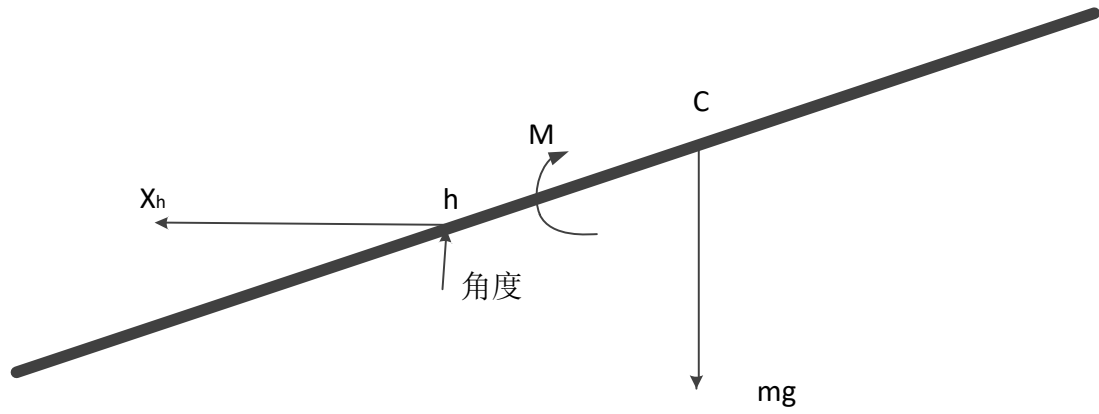


图 6 动力学分析图

在对人体投掷标枪运动的研究中，目前通常使用的人体力学模型均是由十五个刚体组成的，其中头、上下躯干、左右手、上臂、前臂、大腿、小腿、足都是由一些简单几何形状（椭球或椭圆柱或圆锥台）的均质刚体模型来代替。各刚体的连接部位(关节)为球铰或平面铰。在掷标枪运动中，将各肢体抽象为刚体，同时也将标枪假定为刚体。因此，将人体-标枪简化为多刚体系统的力学模型。在本文的研究中，根据人体-标枪运动的动作特点，将人体-标枪系统简化为由刚体组成的模型。

③转动惯量的计算是采用扭摆法，计算式如下：

$$J_y = J_z = \frac{M \times g \times a^2}{4\pi^2 h} \left(\frac{T}{n}\right)^2$$

式中 J_y 、 J_z 为标枪绕 y 轴、 z 轴的转动惯量，单位为 $\frac{kg}{m^2}$ ； M 为标枪质量，单位为 kg ； h 为挂线长度，单位为 m ； a 为悬挂点至标枪重心的水平距离，单位为 m ， $\frac{h}{a}$ 通常取 3~5； $\frac{T}{n}$ 为标枪扭摆平均周期， n 是扭摆次数，一般取 20 次， T 是扭摆 n 次的时间，单位为 s 。

5.3.3 问题（1）的求解

假设某运动员投掷出手速度为 $29.70m/s$ 、出手角为 36.6° 、初始攻角为 -0.9° 。

在出手瞬间标枪上的力有：重力 mg ，肘关节与腕关节的约束反力 X_h 、 Y_h ，关节对标枪的力矩 M 。

设标枪质心以下到把手部位长度为 l_1 ，质心以上部位长度为 l_2

$$m\ddot{x} = -X_h$$

$$m\ddot{y} = Y_h - mg$$

$$J\omega = -M - X_h l_2 \sin(\theta - 4\pi) - Y_h l_2 \cos(\theta - 4\pi) = -M - X_h l_2 \sin \theta - Y_h l_2 \cos \theta$$

出手后标枪受力有：重力 mg ，空气阻力

投掷手在出手高度为 H_0 的A点，以出手角度为 β_0 ，初速度为 V_0 投出器械，器械重心就会沿一条曲线，该曲线经过与A点同高的B点落于地面上的C点。根据运动学可知，在不考虑空气阻力的情况下，这条曲线将近似为抛物线。运动轨迹如下图图 7 运动轨迹图所示：

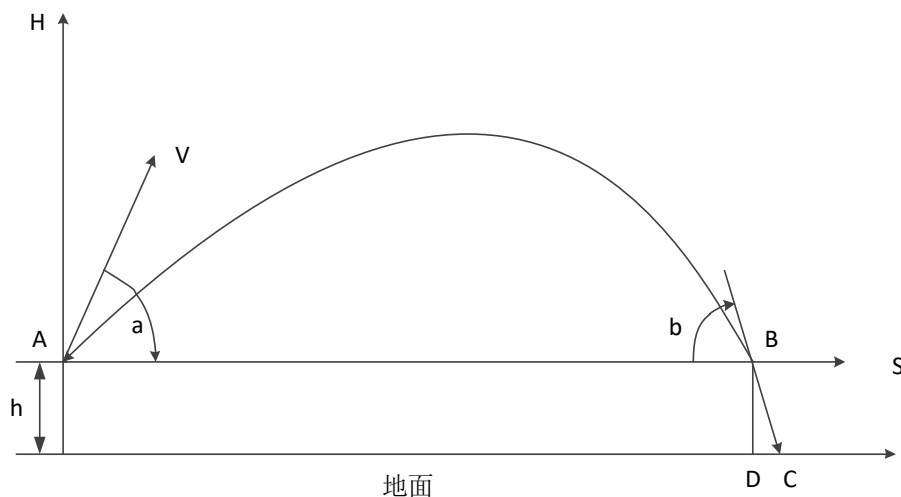


图 7 运动轨迹图

因此，为简单起见，先按抛物线轨迹来分析标枪的运动。把运动轨迹分为AB段和BC段，其中AB段是主要阶段。器械由A点投出，经过最高点而落到与A点同高的B点。根据运动学的斜抛运动可知，水平投掷距离为：

$$AB = S_1 = V_0^2 \sin(2\beta_0) / g$$

器械由B运动到C的水平距离DC，可以近似地把曲线BC看成是直线，其斜角近似等于出手角度 β_0 ，因此得：

$$DC = S_2 = H_0 / \tan(\beta_0)$$

所以，总投掷距离为：

$$S = S_1 + S_2 = \frac{V_0^2 \sin(2\beta_0)}{g} + H_0 / \tan(\beta_0)$$

由上式可知，影响投掷距离的因素有 V_0 、 β_0 和 H_0 。实际上，影响投掷距离的因素不仅有初速度、出手角度和出手高度，还有初始攻角即初速度方向与标枪长轴之间的夹角、初始俯仰角速度即出手时标枪长轴围绕通过质心横轴转动的角速度等初始条件，以及标枪本身的参数和天气因素等。

代入数据到公式，得结果如下：

标枪的投掷距离为90.4847m。

5.3.4 问题（2）的求解

假设某运动员投掷出手速度为30m/s，本文采用粒子群算法^[3]进行求解。给出使得投掷距离最大最佳的出手角和初始攻角，并且估算出标枪的投掷距离。

①算法原理：

PSO 用于解决优化问题。在 PSO 中，每个优化问题的潜在解都是搜索空间中的一只鸟，称之为粒子。所有的粒子都有一个由被优化的函数决定的适值(fitness value)，每个粒子还有一个速度决定它们飞翔的方向和距离。然后粒子们就追随当前的最优粒子在解空间中搜索。

PSO 初始化为一群随机粒子（随机解），然后通过迭代找到最优解。在每一次迭代

中，粒子通过跟踪两个极值来更新自己；第一个就是粒子本身所找到的最优解，这个解称为个体极值；另一个极值是整个种群目前找到的最优解，这个极值是全局极值。

假设在一个D维的目标搜索空间中，有N个粒子组成一个群落，其中第i个粒子表示为一个D维的向量：

$$X_i = (x_{i1}, x_{i2}, \dots, x_{iD}) \quad i = 1, 2, \dots, N$$

第i个粒子的“飞行”速度也是一个D维的向量，记为：

$$V_i = (v_{i1}, v_{i2}, \dots, v_{iD}) \quad i = 1, 2, \dots, N$$

第i个粒子迄今为止搜索到的最优位置称为个体极值，记为

$$P_{best} = (p_{i1}, p_{i2}, \dots, p_{iD}) \quad i = 1, 2, \dots, N$$

整个粒子群迄今为止搜索到的最优位置为全局极值，记为

$$g_{best} = (p_{g1}, p_{g2}, \dots, p_{gD})$$

在找到这两个最优值时，粒子根据如下的公式来更新自己的速度和位置：

$$v_{i1} = wv_{i1} + c_1r_1(p_{i1} - x_{i1}) + c_2r_2(p_{g1} - x_{i1})$$

$$x_{id} = x_{id} + v_{id}$$

其中： c_1 和 c_2 为学习因子，也称加速常数， w 为惯性因子， r_1 和 r_2 为[0, 1]范围内的均匀随机数。

上式右边由三部分组成，第一部分为“惯性”或“动量”部分，反映了粒子的运动“习惯”，代表粒子有维持自己先前速度的趋势；第二部分为“认知”部分，反映了粒子对自身历史经验的记忆或回忆，代表粒子有向自身历史最佳位置逼近的趋势；第三部分为“社会”部分，反映了粒子间协同合作与知识共享的群体历史经验，代表粒子有向群体或邻域历史最佳位置逼近的趋势， $i = 1, 2, \dots, D$, v_{iD} 是粒子的速度， $v_{iD} \in [-v_{max}, v_{max}]$, v_{max} 是常数，用来限制粒子的速度。 r_1 和 r_2 是介于[0,1]之间的随机数。^[4]

Step1 初始化粒子群，包括群体规模N，每个粒子的位置 x_i 和速度 v_i ；

Step2 计算每个粒子的适应度值 $F_{it}[i]$ ；

Step3 对每个粒子，用它的适应度值 $F_{it}[i]$ 和个体极值 $P_{best}(i)$ 比较，如果 $F_{it}[i] > P_{best}(i)$ ，则用 $F_{it}[i]$ 替换掉 $P_{best}(i)$ ；

Step4 对每个粒子，用它的适应度值和全 $F_{it}[i]$ 局极值 g_{best} 比较，如果 $F_{it}[i] > g_{best}$ 则用 $F_{it}[i]$ 替 g_{best} ；

Step5 根据公式更新粒子的速度 v_i 和位置 x_i ；

Step6 如果满足结束条件(误差足够好或到达最大循环次数)退出，否则返回 Step2。

用 MATLAB 工具求解得结果如下：

出手角是37°；初始攻角为-0.17°；标枪的投掷距离为 90.3545 m。

5.4 问题四的求解

5.4.1 空气动力学理论

前文论述中，在只研究标枪出手后在没有空气阻力的情况下，从力学的角度来看，其飞行轨迹是一条抛物线。但是，如果考虑空气阻力对标枪的作用，标枪的运动就变得

相当复杂。标枪出手以后的运动是一个自然运动的过程，不受人为因素的影响。标枪的运动完全是在空气动力的作用下进行的，受空气动力学规律^[5]的控制。

空气动力学是力学的一个分支，是在流体力学的基础上，随着航空工业和喷气推进技术的发展而成长起来的一个学科。通常所说的空气动力学研究内容是飞机，导弹等飞行器在各种飞行条件下流场中气体的速度、压力和密度等参量的变化规律，飞行器所受的举力和阻力等空气动力及其变化规律，气体介质或气体与飞行器之间所发生的物理化学变化以及传热传质规律等。从这个意义上讲，空气动力学可有两种分类法^[6]：

①根据流体运动的速度范围或飞行器的飞行速度，空气动力学可分为低速空气动力学和高速空气动力学。通常大致以 400 千米/小时这一速度 作为划分的界线。在低速空气动力学中，气体介质可视为不可压缩的，对应的流动称为不可压缩流动。对于标枪在空气中飞行的研究就属于低速空气动力学。大于这个速度的流动，须考虑气体的压缩性影响和气体热力学特性的变化。这种对应于高速空气动力学的流动称为可压缩流动。

②根据流动中是否必须考虑气体介质的粘性，空气动力学又可分为理想空气动力学(或理想气体动力学)和粘性空气动力学。

除了上述分类以外，空气动力学中还有一些边缘性的分支学科。例如稀薄气体动力学、高温气体动力学等。

5.4.2 动力学模型^[7]

风向分别为顺风 and 逆风，风速分别为 3m/s 、 6m/s 和 9m/s ，运动员投掷出手速度为 31.70m/s 。在上述三个条件下建立标枪飞行的数学模型。

标枪在运动过程中不会轻易发生形变，因此将其视为刚体。投掷过程将结合动量定理、动量矩定理、动能定理以及由此推导出来的一系列定理，建立描述标枪的运动特征与其作用量之间的关系，并从物体的机械运动进行了深入的研究。在本文中，将采用该方法建立标枪的动力学模型^[8]。

(1) 动量定理：

$$\frac{d\vec{K}}{dt} = \vec{R}^{\rightarrow(e)} = \sum \vec{F}_t^{(e)}$$

质点系的动量定理，即质点系得动量对时间的导数等于作用在质点系上的所有外力的矢量和，在标枪出手的瞬间，可视标枪为一个质点，因此动量定理可以变为：

$$\frac{d\vec{K}}{dt} = \vec{R}^{\rightarrow(e)} = \lim_{t \rightarrow 0} \sum \vec{F}_t^{(e)}$$

与动量定理相同，出手瞬间视标枪为质点，质点对任一固定点的动量矩对时间的导数，等于作用在质点系上的所有外力对同一点之矩的矢量和。

$$\frac{dL_O}{dt} = \vec{M}_O^{(e)} = \lim_{t \rightarrow 0} \sum \vec{m}_O \left(\vec{F}_t^{(e)} \right)$$

根据以上的动量矩定理，矩心或矩轴是固定点或固定点或固定轴。但是，若取质点系得质点为矩心，则其对应的质点的动量矩定理如下所示：

$$\frac{dL_{C_t}}{dt} = \vec{M}_C^{(e)} = \lim_{t \rightarrow 0} \sum \vec{m}_C \left(\vec{F}_t^{(e)} \right)$$

设平面运动刚体^[9]对轴心轴的转动惯量为 J_c ，瞬时的减速度为 $\bar{\omega}$ ，则刚体对于质心的动量矩可表示为：

$$L_{Ct} = J_c \bar{\omega}$$

而 $\frac{dL_{Ct}}{dt} = J_c \bar{\varepsilon} = J_c \ddot{\bar{\phi}}$ ，则可得：

$$m \bar{a}_c = \bar{R}^{(e)} = \sum \vec{F}_t^{(e)}$$

$$J_c \bar{\varepsilon} = \vec{M}_c^{(e)} = \sum m_c (\vec{F}_t^{(e)})$$

或

$$m \ddot{x}_c = \sum F_{tx}$$

$$m \ddot{y}_c = \sum F_{ty}$$

$$J_c \bar{\varepsilon} = \sum m_c (\vec{F}_t^{(e)})$$

上式称为刚体平面运动微分方程。可见，刚体作平面运动时，随同质心的平动决定于外力系得主矢，而绕质心的转动决定于外力系对质心的主矩。这样，就建立了外力系得主矢与主矩同刚体的瞬间关系。

(2) 在标枪出手以后，标枪将不能视为质点，结合动量方程、能量方程、连续方程建立动网格守恒方程。

①连续性方程

$$\frac{d}{dt} \int_{\Omega} \rho d\Omega + \oint_S \rho (\vec{V} - \vec{U}) \cdot \hat{n} dS = 0$$

②动量方程

$$\frac{d}{dt} \int_{\Omega} \rho d\Omega + \oint_S \left[\rho (\vec{V} - \vec{U}) \otimes \vec{V} + p \vec{I} - \vec{\tau} v \right] \cdot \hat{n} dS = \int_{\Omega} \vec{F}_b d\Omega$$

Ω 和 S 分别是移动控制体积中的体积和边界面， $\Omega = \Omega(t)$ ，而且符合几何守恒定律： $\frac{d\Omega}{dt} - \oint_S \vec{U} \cdot \hat{n} dS$ 。时间导数 $\left(\frac{d}{dt} \right)$ 表示的是时间顺着运动控制的体积的微分。所有空间梯度都是在静止参考系下计算的。

③根据上述的连续性方程和动量方程建^[10]立动网格守恒方程：

$$\frac{d}{dt} \int_V \rho \phi dV + \int_{\partial V} \rho \phi (\vec{u} - \vec{u}_g) \cdot d\vec{A} = \int_{\partial V} \Gamma \nabla \phi \cdot d\vec{A} + \int_V S_{\phi} dV$$

其中 ρ 为空气密度， $\vec{u}$ 为速度向量， $\vec{u}_g$ 移动网格的网格速度， Γ 为扩散系数， S_{ϕ} 为源项， ∂V 代表控制 V 的边界。

5.4.3 结果分析

如下表所示：

	顺风			逆风		
	3m/s	6m/s	9m/s	3m/s	6m/s	9m/s
最佳出手角（度）	39	41	40	40	39	40
最佳初始攻角（度）	-0.15	-0.17	-2.0	-0.25	-0.27	-3.0
最佳初始俯仰角速度（度/秒）	40	40	40	40	40	40
最大投掷距离（米）	90.57	104.21	118.02	79.08	91.06	105.36

5.5 问题五的求解

标枪技术参数同问题 3，本文建立模糊综合评判模型，分析运动员出手速度、出手角、初始攻角、初始俯仰角速度、风向及风速等要素对标枪投掷距离影响的相对重要性。



5.5.1 模糊综合评判模型^[11]

模糊综合评判包括 6 个主要组成要素：评判指标集、评价集、模糊评价矩阵、评判因素权重集、模糊算子和评判结果向量。

本题主要分析运动员出手速度、出手角、初始攻角、初始俯仰角速度、风向及风速等要素对标枪投掷距离影响的相对重要性。首先建立评价指标集和评价集，对被评价对象划分评价指标并建立评语集；其次根据隶属度函数法或专家评分法等方法确定隶属度矩阵模糊评价矩阵；然后计算各评价指标权重值，得到权重集；最后选择相应算法进行合成。

①建立指标体系

主要评价指标	评价内容
--------	------

出手速度	使得投掷距离最远的最佳出手速度
出手角	使得投掷距离最远的最佳出手角
初始攻角	使得投掷距离最远的最佳初始攻角
初始俯仰角速度	使得投掷距离最远的最佳初始俯仰角速度
风向及风速	使得投掷距离最远的最佳风向及风速

②决策的优劣，在优选识别过程中并不存在绝对分明的界限，具有中介过渡性，属于模糊概念。

借鉴多目标模糊优选模型，设有可选择的 n 个投掷方案，设评价指标有 m 个，如出手速度、出手角、初始攻角、初始俯仰角速度、风向及风速等，组成一个目标特征矩阵：

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix} = (x_{ij})$$

x_{ij} ——技术方案 j 目标 i 的特征值， $i = 1, 2, \dots, m; j = 1, 2, \dots, n$

③投掷方案的优劣程度依据 m 个目标特征值，按从优级到劣级的 c 个级别进行识别。可以规定优级(1级)对优的相对隶属度(简称相对优属度)为1，劣级(c 级)的相对优属度为0，1级至0级的相对优属度呈线性递减。则对于任一目标，从1级到 c 级各个级别的相对优属度标准值向量为：

$$S = \left[1 \quad \frac{c-2}{c-1} \quad \frac{c-3}{c-1} \quad \cdots \quad 0 \right] = (s_h), h = 1, 2, \dots, c$$

④投掷方案的考核目标通常分为特征值越大越优、越小越优与中间型3类。越大越优、越小越优与中间型的目标相对优属度公式分别可采用：

$$r_{ij} = \frac{x_{ij} - \min x_{ij}}{\max x_{ij} - \min x_{ij}}, \forall_j$$

$$r_{ij} = \frac{\max x_{ij} - x_{ij}}{\max x_{ij} - \min x_{ij}}, \forall_j$$

$$r_{ij} = 1 - \frac{|x_{ij} - \bar{x}_l|}{\max |x_{ij} - \bar{x}_l|}, \forall_j$$

⑤将方案 j 的 m 个目标相对优属度分别与级别矩阵 S 以进行比较后，得到投掷方案 j 级别上限值 b_j 与下限值 a_j ：

$$\begin{cases} a_j = \min a_{ij}; \\ b_j = \max a_{ij}. \end{cases}$$

⑥设 m 个目标的权重用 ω_i 表示，则投掷方案集目标权向量为：

$$W = (\omega_1, \omega_2, \dots, \omega_m), \sum_{i=1}^m \omega_i = 1.$$

⑦考虑了目标权重后，为了更完善的描述方案 j 与级别 h 之间的差异，可用加权广义

欧氏距离表示:

$$D_{hj} = \mu_{hj} d_{hj} = \mu_{hj} \left\{ \sum_{i=1}^m [\omega_{ij} (r_{ij} - s_h)]^2 \right\}^{\frac{1}{2}}$$

⑧根据目标函数构造拉格朗日函数，解得

$$\mu_j = \frac{1}{\sum_{k=a_j}^{b_j} \frac{\sum_{i=1}^m [\omega_i (r_{ij} - s_h)]^2}{\sum_{i=1}^m [\omega_i (r_{ij} - s_k)]^2}}$$

⑨应用 μ_j 可解得飞行方案集归属各个级别的最优化相对隶属矩阵:

$$U^* = \begin{bmatrix} \mu_{11}^* & \mu_{12}^* & \cdots & \mu_{1n}^* \\ \mu_{21}^* & \mu_{22}^* & \cdots & \mu_{2n}^* \\ \vdots & \vdots & & \vdots \\ \mu_{m1}^* & \mu_{m2}^* & \cdots & \mu_{mn}^* \end{bmatrix} = (\mu_{ij}^*),$$

$$j = 1, 2, \dots, n; h = 1, 2, \dots, c.$$

⑩根据级别特征值H的向量式:

$$H = (1, 2, \dots, c)(\mu_{ij}^*) = (H_1, H_2, \dots, H_n)$$

对投掷方案集进行优选，其中最小级别特征值 H_{min} ，对应的投掷方案为最远投掷距离方案。优选计算时,要求级别 $c \geq 2$, c 越大优选的精度越高,但计算量越大。本文取 $c = 4$,相当于在优级(1级)与差级(4级)之间插入良(2级)、中(3级)。即用优、良、中、差对多目标进行优选决策。

5.5.2 结果及数据分析

本文在求解问题时，考虑了空气阻力的影响。

(1) 出手速度对标枪投掷距离的影响

在决定标枪飞行远度的诸因素中，出手速度是最重要的因素，也是运动员创造优异成绩的必要条件。在本文中，研究出手速度对标枪投掷距离的影响，计算出标枪在不同出手速度下的投掷距离。

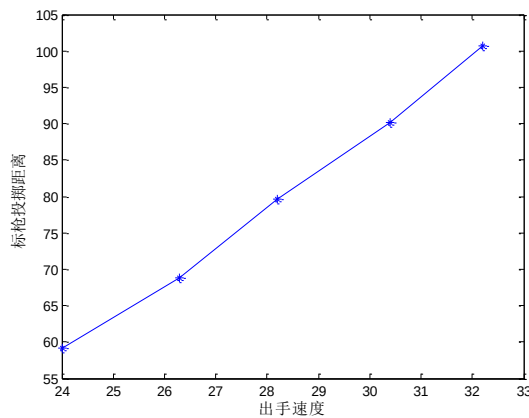


图 8 出手速度对标枪投掷距离的影响图

上图表示标枪在出手角度为 42° ，初始攻角 A_0 为零，初始俯仰角速度 W_0 也为零时，标枪在出手速度为 24 m/s - 34 m/s 的范围内标枪的飞行远度。

由图显示结果表明，标枪在出手速度为 24 m/s - 34 m/s 的范围内，随着出手速度的不断增加，标枪的投掷距离也不断增加，而且增加的幅度有加大的趋势。在本文研究的范围内，它们基本上呈线性关系。由此可知，出手速度对标枪的飞行远度的影响是十分显著的。

(2) 出手角对标枪投掷距离的影响

标枪的出手角度是标枪出手瞬间其质心的速度方向与水平面的夹角，而标枪的最佳出手角度是指在其它初始条件一定的情况下，能够使标枪的投掷距离最大的出手角度。

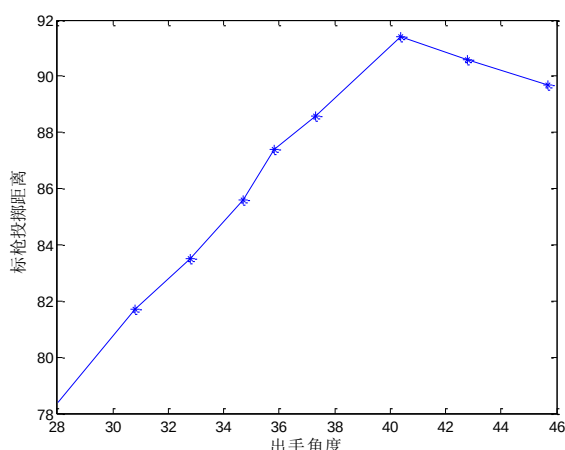


图 9 出手角对标枪投掷距离的影响图

上图是标枪在初速度 V_0 为 30 m/s ，初始攻角 A_0 为零，初始俯仰角速度 W_0 也为零，出手角度在 28° - 45° 范围内标枪的投掷距离。

由图显示结果表明，标枪在出手角度为 28° - 42° 范围内，随着出手角的增加，标枪的投掷距离基本上是线性增加的，但是达到 42° 后其投掷距离略有减小。一般认为在不考虑空气阻力的情况下，当出手角度为 45° 时，标枪的投掷距离最大。但是，在空气阻力的影响下，要使标枪飞投掷距离达到最大，出手角度并不是 45° 。当出手角度超过 42° 以后，标枪的投掷距离开始下降。由此得出，该标枪的最佳出手角度为 42° 。

(3) 初始攻角对标枪投掷距离的影响

标枪出手时的初始攻角不同，则其在飞行中的攻角也会变化，并导致作用于标枪的空气动力特性发生变化，从而影响标枪的投掷距离。本文研究标枪在初速度为 30 m/s 、出手角度为 42° 的条件下，初始攻角与初始俯仰角速度的不同组合对投掷距离的影响。

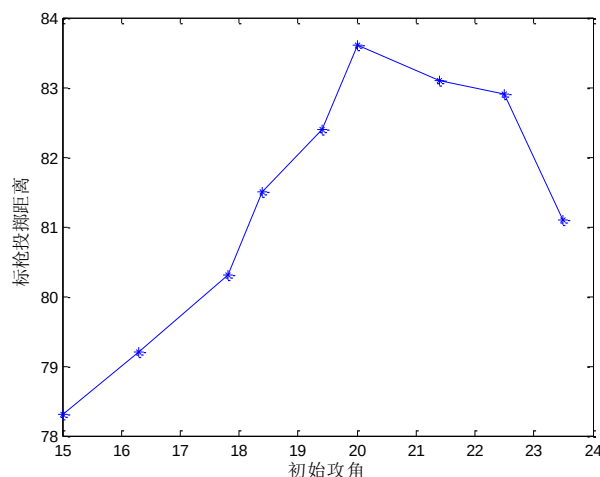


图 10 初始攻角对标枪投掷距离的影响图

由图显示结果表明,在其它初始条件一定的情况下,初始攻角的变化对标枪投掷距离的影响并不十分显著。并且,在初始攻角与初始俯仰角速度的不同组合中,其中最佳组合为 $B_0 = 42^\circ$ 、 $A_0 = -8^\circ$ 、 $W_0 = -0.09 \text{ rad/s}$ 时,标枪的投掷距离最大,达到 91.79 m。当初始条件偏离这一组合时,投掷距离有所减小。

(4) 初始俯仰角速度对标枪投掷距离的影响

当运动员投掷标枪时,由于用力方向未能通过标枪纵轴,标枪在出手时会出现一个初始俯仰角速度,使标枪出手时出现向上抬头($W_0 > 0 \text{ rad/s}$)或向下低头($W_0 < 0 \text{ rad/s}$)的现象,从而对其飞行远度产生影响。本文考虑了 $-20^\circ/\text{s} - 20^\circ/\text{s}$ 间隔为 $5^\circ/\text{s}$ 的俯仰角速度对飞行远度的影响,研究标枪的投掷距离随着初始俯仰角速度的变化情况。如下图所示:

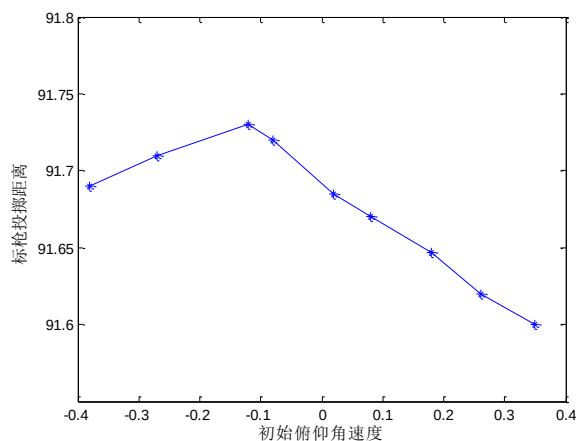


图 11 初始俯仰角速度对标枪投掷距离的影响图

由图显示结果表明,初始俯仰角速度对标枪的飞行远度并不显著,但当初始俯仰角速度 $W_0 < 0 \text{ rad/s}$ 时,标枪飞行的远度比 $W_0 > 0 \text{ rad/s}$ 远一些。标枪初速度为 30 m/s ,在初始俯仰角速度为 -0.17 rad/s 的情况下飞行远度最大。因此,认为最佳初始俯仰角速度大约在 $-0.17 \text{ rad/s} - 0 \text{ rad/s}$ 范围内。

(5) 风向及风速的影响对标枪投掷距离的影响

根据风的等级情况,对 I 级(相当风速为 $\pm 1 \text{ m/s}$)、II 级(相当风速为 $\pm 2 \text{ m/s}$)、III 级

(相当风速为 $\pm 4 \text{ m/s}$)、IV级(相当风速为 $\pm 6 \text{ m/s}$)进行风向及风速的影响分析。其中, 风速为负值代表逆风, 正值代表顺风。其它初始条件一定, 即初速度 V_0 为 30m/s , 初始攻角 A_0 为零, 初始俯仰角速度 W_0 为零, 出手角度为 42° 。

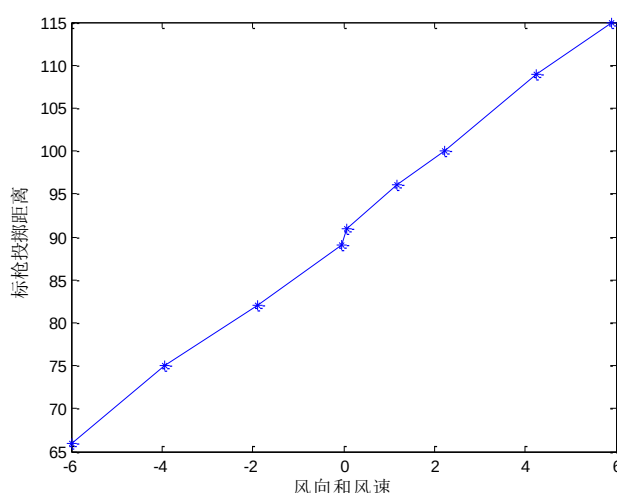


图 12 风向及风速的影响对标枪投掷距离的影响图

由图显示结果表明, 风向对标枪的投掷距离的影响比较显著。对于逆风来说, 在其它初始条件一定的情况下, 随着逆风风速的增大, 标枪的飞行远度呈现出减小的趋势; 对于顺风来说, 在其它初始条件一定的情况下, 随着顺风风速的增大, 标枪的飞行远度呈现出增大的趋势。风速在 $-6 \text{ m/s}—6 \text{ m/s}$, 标枪投掷距离与风速成正比关系。

六 模型的评价及推广

6.1 优点

(1) 本文运用了曲线拟合模型, 首先画出所给数据散点图, 然后运用图表直观表达拟合曲线, 避免了只有函数方程, 增强了本文的可读性和理解性。

(2) 本文研究了标枪的气动力性能, 对运动员投掷时初始攻角的选择具有指导意义, 而且为模拟标枪的飞行提供了前提保证。

(3) 模糊综合评判模型可以避免凭经验进行目标选择所固有的主观性, 使并购决策更加科学合理。很好地解决一般模糊综合评价模型的一些缺点, 如因素多导致各因素权重小而造成的严重失真现象或多峰值现象等。适合评价多主体对多层次多类指标评价信息的整合。

(4) 粒子群算法搜索速度快、效率高, 算法简单, 适合于实值型处理。

6.2 缺点

(1) 本文虽建立不同拟合曲线类型, 但也仅限于有限的曲线类型, 无法保证此曲线即是吻合度最高的曲线。

(2) 本文只对标枪的部分初始条件, 包括: 初速度、出手角度、初始俯仰角速度, 进行了研究分析, 而对偏航与自转对标枪飞行远度的影响并没有加以考虑。结果可能会有误差。

6.3 推广

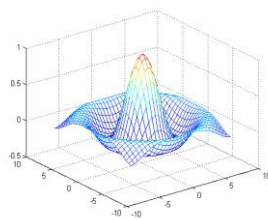
(1) 从实践意义上来说, 本文所建立的模型、得到的结果可以对掷标枪教练和运动员进行掷标枪指导和训练具有一定的指导作用, 对于运动员创造好成绩具有重要意义。

(2) 本文只是将标枪作为刚体来加以考虑, 在将来的工作中, 也可以将标枪作为弹性体来模拟其飞行情况。

七 参考文献

- [1]. 王新和, 程世洲. 曲线拟合的最小二乘法[J], 新疆职业大学学报, 2004, 12 (2): 84—86
- [2]. 谢宇. 回归分析[M]. 北京: 社会科学文献出版社, 2010
- [3]. 杨朝霞, 方建文, 李佳蓉, 等. 粒子群优化算法在多参数拟合中的作用[J]. 浙江师范大学学报, 2008, 31 (2): 173—177
- [4]. YAMAMOTO, Yoshitaka. MATSUOKA, Rutsu. ISHIDA, Yoshihisa. and SEKI, Kazuichi., Analysis of javelin throwing by high-speed photography, Proceedings of SPIE—The International Society for Optical Engineering, 1998, 799-807
- [5]. Mont Hubbard, Neville J. de Mestre and John Scott, Dependence of release variables in the shot put[J]. Journal of Biomechanics, 2001, 34(4): 449-456.
- [6]. Ferenc Mizera, Gabor Horvath. Influence of environmental factors on shot put and hammer throw range[J]. Journal of Biomechanics, 2002, 35: 785–796
- [7]. 王肇明. 标枪动力学的优化计算[J]. 力学与实践, 1983, 4: 40-42
- [8]. 王其宁, 申毛毛, 陈耀庆等. 标枪气动力分析与初始条件的优化组合[J]. 四川体育科学, 1997: 3-4, 37-41
- [9]. 蔡国华, 标枪几何参数和气动力特性测量技术报告 [R], 北京空气动力研究所 1986. 6
- [10]. 周曙, 优秀男子标枪运动员最后出手特征分析, 2005, 24(4), 485. 486
- [11]. 张健, 郑亚平, 对掷标枪最后用力技术中躯干动作特征的三维力学分析, 1997, 33(11—12), 82—84

五一数学建模竞赛



题 目：木板最优切割方案

关键词：最优化排布模型 矩形排布模型 遗传算法 整数规划模型

摘 要：

本文通过分析，确定决策变量和约束条件，建立了最优化排布模型，矩形排布模型，整数规划模型等，并用循环嵌套式算法和遗传算法等对模型进行求解，对木板最优切割方案进行了分析与研究。

针对问题一，本文首先将木板置于坐标系下，给每个木板中心坐标，分析论文实现时的决策变量和约束条件，建立最优化排布模型，再设计出循环嵌套式算法，根据流程通过 MATLAB 编程实现计算，求解出题目中规格木板能切割下的木板产品 P1 数目为 59，木板利用率为 98.3%。

针对问题二，本文先根据切割木板的实际过程，定义一系列变量表示木板产品的类型、状态和位置，再确定其约束条件。根据题意要使木板的利用率最高，确定目标函数和决策变量。目标函数为木板剩余的面积最少，决策变量为木板产品的型号和状态。然后建立了矩形排布模型，同时用约束条件对目标函数进行约束。对约束条件进行线性化改进后，通过遗传算法进行求解。通过将 P1 或 P3 木板以两条染色体表示，交叉和变异等步骤，逐步得出最佳的三种切割方案为：方案一：P1 切割 0 块，P3 切割 48 块；方案二：P1 切割 43 块，P3 切割 13 块；方案三：P1 切割 59 块，P3 切割 0 块。

针对问题三，本文先利用问题二中建立的模型，计算得出切割 P1、P3 的最优的五种方案作为基本方案，建立整数规划模型，确定目标函数和约束条件，用 LINGO 编程求解最优解，结果见表 6.2。

针对问题四，本文先在问题二基础上，改进为切割四种规格 P1、P2、P3、P4 的木板，同样利用遗传算法，先求得木板占用率较高的几种方案，再把这几种方案作为基本方案，建立整数规划模型，确定目标函数和约束条件，用 LINGO 编程求解最优解，同结果见表 7.2。

针对问题五，本文根据表 2 中的利润值，先列出目标函数，即为利润的最大值函数，根据给定 100 张 S1 木板，列出面积的约束条件，及各个方案使用次数的约束条件，最终用整数规划，LINGO 编程求解出结果，使利润最大，结果见表 9.2。

本文的特色在于可将解决大木板上切割小木板的问题方法运用于为多种相似的布局问题，该思想具有极强的适应性和可变性。解决第三、四、五问的过程中，合理地利用了前一问所获得的结论，对应用于自身上的模型进行进一步优化。

一、问题重述

在当今激烈的市场竞争中，提高材料利用率，降低生产成本，是厂商保持竞争力重要手段。本题中徐州家具厂进行家具加工的过程中需对一批 S1 木板进行切割。切割属于布局问题，是组合优化的 NP 完全问题。目前，求解该问题时多采用各种近似算法、启发式算法及智能优化算法等。此问题需要在切割大块木板的过程中设计合理切割方案，减少材料浪费，使家具厂经济生产效益达最大化。

问题一：已知完成切割所用 S1 木板的规格及需切割产品 P1 的规格。厂家需在大块木板上切割 P1 产品，通过建立数学模型，遵循节省原料的原则，给出木板利用率最高切割方案，使材料浪费达到最低，并将结果填入表中。

问题二：已知 P3 产品规格。厂家需在木板上切割 P1 和 P3 产品，通过建立数学模型，给出按照木板利用率由高到低排序的前 3 种切割方案，并将结果填入表中。

问题三：已知 P1、P3 产品规格以及生产目标。厂家在木板上切割 P1 和 P3 产品，在满足生产目标的情况下，给出木板总利用率最高的切割方案，并将结果填入表中。

问题四：在第三问的基础上，新知 P2、P4 产品规格以及它们的生产目标。厂家需在 S1 木板上切割这四种产品并完成生产目标。在满足生产目标的情况下，遵循节省原料的原则，给出木板总利用率最高的切割方案，并将结果填入表中。

问题五：已知现有 100 张 S1 木板，以及 P1、P2、P3、P4 每件产品的利润。没有四种产品的需求数量限制。厂家在 S1 木板上切割这四种木板产品，需按照给出的利润，建立数学模型，给出总利润最大的切割方案，并将结果填入表中。

二、问题假设

- 1) 根据板材切割后余料不可能再生产产品的原则，后剩余的边角料不能再利用；
- 2) 忽略实际加工时产生的误差对原设计的影响，即切割过程无损耗；
- 3) 木板厚度和割缝宽度忽略不计；

三、符号说明

序号	符号	符号说明
1	L	木板的长
2	W	木板的宽
3	x_0	坐标的原点横坐标
4	y_0	坐标的原点纵坐标
5	x	木板位置的横坐标
6	y	木板位置的纵坐标
7	x'	下一个要切割木板的横坐标
8	y'	下一个要切割木板的纵坐标
9	σ_{ik}	木板产品的类型

10	t_i	确定第 i 个木板产品是否切割下
----	-------	--------------------

四、问题分析

提高材料利用率，减少材料浪费，使经济效益达到最大化是所有厂商追求的目标。本文研究的是在一个规格为 $L \times W$ 的 S1 木板上切割不同规格小矩形木板的问题，可以将该切割木板的问题转化为在大木板中放置小木板的问题。不同规格的矩形木板在不同的条件约束下的最佳排列位置和放置方式不同。在小矩形不超出 S1 木板边界、小矩形不互相重合的情况下，要尽量多的满足各种规格小矩形摆放数量最多，S1 木板的边角余料最少，就要设计出适用于不同约束条件下的矩形排列优化方案。

4.1 问题一的分析

问题一要求在 S1 木板上切割 P1 木板，给出利用率最高的切割方案。首先，可以借鉴于循环嵌套式算法的启发，在木板最大面积内，由外而内排列 P1 木板设计能够最大限度地使用矩形木板边长的 MATLAB 程序，求解产生最优的 P1 木板长宽组合及所有 P1 木板的个数。以木板两边边长的利用率最大为优化目标，以矩形形边长的利用率最大化为优化目标，得到最终的摆放方案，并利用 EXCEL 作出示意图。

4.2 问题二的分析

问题二的分析问题二不再是单一切割 P1 矩形木板，要求在 S1 木板上切割 P1 和 P3 木板并使 S1 木板的利用率最高。由于本问含有不止一种规格的小矩形木板且需要在排入放置小矩形木板时不断进行优化，淘汰材料利用率低下的切割方案，故可以采用遗传算法求解，将 P1 或 P3 木板的排入顺序、放置状态及产品种类用两条染色体上相应的两个基因来共同表示，再进行交叉和变异，逐步得出最佳的切割方案。

4.3 问题二的分析

问题三在问题二的基础上增加了生产数量的最低要求，在第二问建立的矩形木板切割模型中，我们得出了切割 P1 和 P3 产品使木板利用率最高的前几种切割方案，因此可以在此计算结果上拓展，在已知使木板利用率最高的前几种切割方案的基础上，建立整数规划模型来求解此问题，实现对切割方案的求解。

4.4 问题四的分析

问题四在问题二的基础上增加了产品种类、四种产品生产数量的最低要求，并要求给出木板总利用率最高的切割方案。可以与第三问进行类比，同理可利用第二问中的矩形排布模型，并借助遗传算法加以改进，求出木板总利用率最高的切割方案。

4.5 问题五的分析

问题五与前四问相比出现了新的条件变化，本问中已知有 100 张 S1 木板及 P1、P2、P3、P4 每件产品的利润，与之前不同的是不再有四种产品的需求数量限制，此问要求给出总利润最大的切割方案。此问的目标函数发生变化，变为利润最大而非利润率最大，求解时可建立整数规划模型，以 100 张 S1 木板作为约束条件进行优化计算。

五、问题一的模型建立与求解

5.1 问题一的模型建立

问题一要求在一块木板上切割 P1 产品，并给出木板利用率最高的切割方案。首先将模板放入坐标系中，通过建立最优化排布模型，通过寻找约束变量和约束条件，可实现对问题的求解。

5.1.1 模型准备

在建立模型之前，需要把木板放入坐标系中，故以木板的左下角为原坐标 (x_0, y_0) ，木板的长为横坐标，宽为纵坐标，建立坐标系。

5.1.2 确定约束条件和决策变量

(1) 坐标约束

下一个即将要切割的木板的位置坐标一定在大的木板的范围内，即有以下约束：

$$x' < L, \quad y' < w$$

(2) 面积约束

如果下一次还能切割一个木板产品，那么当前木板的剩余面积一定大于一个木板产品的面积。即约束条件为：

$$LW - i l^{(1)} w^{(1)} > l^i * w^i$$

(3) 重叠约束

依次切割的矩形木块之前，不能有重叠部分，否则不符合形状规则。有如下约束条件：

$$|x_i - x_{i'}| - \left(\frac{l_i + l_{i'}}{2}\right) > 0$$

$$|y_i - y_{i'}| - \left(\frac{w_i + w_{i'}}{2}\right) > 0$$

(4) 决策变量

在确定了切割木板产品的类型和型号尺寸后，我们才能确定它们占用的面积，以求得材料利用率。因此，我们确立了如下决策变量：

i 表示第 i 块木板；

l_i 为第 i 块木板产品的长， w_i 为第 i 块木板产品的宽；

5.1.3 模型建立

至此，我们建立最优化排布模型如下：

$$\min \left\{ LW - \sum_{k=0}^i l^{(1)} w^{(1)} \right\}$$

$$s. t. \begin{cases} 0 < x' < L, 0 < y' < w \\ LW - i l^{(1)} w^{(1)} > l^i * w^i \\ |x_i - x_{i'}| - \left(\frac{l_i + l_{i'}}{2}\right) > 0 \\ |y_i - y_{i'}| - \left(\frac{w_i + w_{i'}}{2}\right) > 0 \end{cases}$$

其中 L 为大木板的长度， W 为大木板的宽度；

l_i 为第 i 块木板产品的长， w_i 为第 i 块木板产品的宽；

$l^{(k)}$ 为型号为 k 的木板产品的长， $w^{(k)}$ 为型号为 k 的木板产品的宽；

x, y 是木板位置的坐标， x', y' 是下一个木板的坐标。

5.2 模型一的求解

5.2.1 算法设计

在切割木材时，通常是从原料的边界开始切割。要在长为 L ，宽为 W 的S1木板中，切割规格为 $a \times b$ 的P1矩形产品，其等价于用若干长度为 a 和 b 的边进行组合优化，逐步提高S1木板各个边方向上的利用率，直至边利用率达到最高。

最外层排列完后，产生剩余木板，再向内部剩余木板以同样的方式进行组合排列与切割，如此逐步循环，直至剩余木板面积无法进行切割，从而达到材料利用率最大。于是基于循环嵌套式算法，通过由外至内逐步优化算法来解决木块排列问题。

算法的流程如下，并参照流程图设计出MATLAB程序：

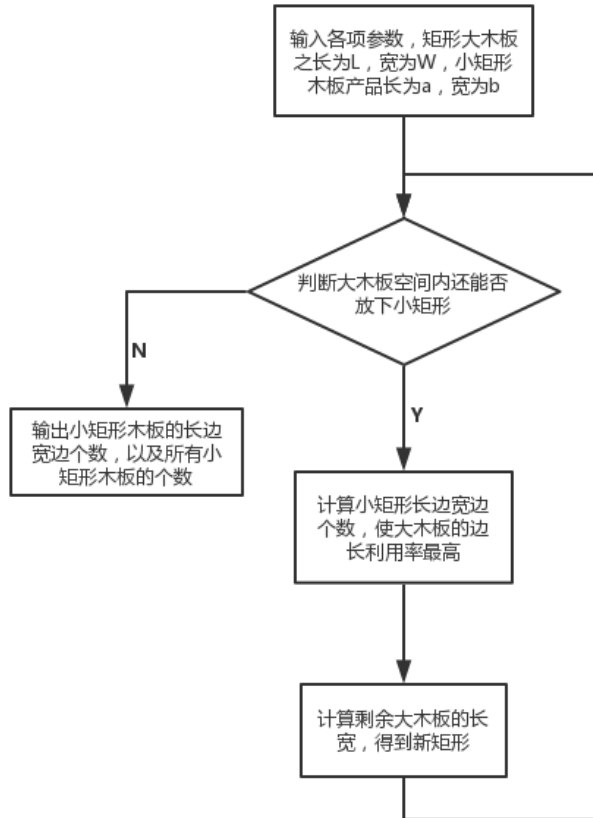


图 5.1 模型一流程图

5.2.2 结果求解

经过上述模型建立与计算过程，可以得到如下最优切割方案，见表 5.1

表 5.1 问题一的结果

P1 的数量	木板利用率
59	98.3%

用 EXCEL 作出如下 S1 模板切割示意图：

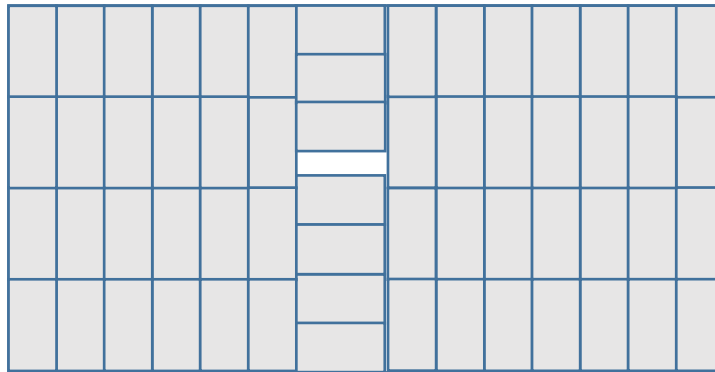


图 5.1 问题一结果示意图

5.1.3 结果分析

本问最终答案为 S1 木板上最多可切割 59 块 P1 产品，木材利用率为 98.3%

若去除长 L 宽 W 上的约束，仅以面积上 $L * W$ 来作为约束条件，则由 $(L * W)/(l_i * w_i)$ 可知最多可切割 60 块 P1，木材利用率为 99.99%。由此可看出，本问中最优化排布模型的长宽边界起到了约束作用。

六、问题二的模型建立与求解

6.1 问题二的模型建立与求解

6.1.1 确定约束条件

首先，分析了木板的切割过程，以及如何最优地安排每次切割的木板位置后。我们定义了一些在不同的情况下所需的变量作为约束变量。

(1) 二元约束变量

在这部分中，我们需要对即将切割的木板产品的类型进行分析。因此，我们引入了一个二元变量描述这些约束。

为了确定木板产品的类型，定义一个变量：

$$t_{ik} = \begin{cases} 1, & \text{第} i \text{个木板产品的类型是} k \\ 0, & \text{第} i \text{个木板产品的类型不是} k \end{cases}, k = 1, 3$$

为了确定第 i 个木板产品是否切割下，定义一个变量：

$$C_i = \begin{cases} 1, & \text{第} i \text{个木板产品已被切割下} \\ 0, & \text{第} i \text{个木板产品尚未被切割下} \end{cases}$$

(2) 切割木板产品的位置约束变量

首先，分析木板切割前后的具体情况。结合大木板的面积，我们提出了几条约束条件。定义 x_i, y_i 是第 i 个切割木板的几何中心坐标点， L 是大木板的长， W 是大木板的宽， $l^{(k)}, w^{(k)}$ 是型号为 k 的木板产品的长和宽。

则有以下的范围约束：

$$x_i \in \left[C \frac{l_i}{2}, C_i \left(L - \frac{l_i}{2} \right) \right], \quad y_i \in \left[C_i \frac{w_i}{2}, C_i \left(W - \frac{w_i}{2} \right) \right]$$

$$l_i = \sum_{k=1}^2 t_{ik} l^{(k)}$$

$$w_i = \sum_{k=1}^2 t_{ik} w^{(k)}$$

假设相继被切割的木板产品有两个，木板产品 i 尚未切割，木板产品 i' 已切割。如果 $C_i = C_{i'} = 1$ ，则 $C_{ii'} = 1$ 。于是，对这一约束描述如下：

$$C_i + C_{i'} \leq 1 + C_{ii'} \leq 2$$

一个木板产品可以在大木板的上被切割下的要求是，这个木板产品的长宽必须满足一定条件，即至少在长度、宽度的一个维度内满足约束。因此，这一约束可以表示为：

$$\max \left\{ |x_i - x_{i'}| - \left(\frac{l_i + l_{i'}}{2} \right), |y_i - y_{i'}| - \left(\frac{w_i + w_{i'}}{2} \right) \right\} \geq (1 - C_{ii'})(-M)$$

其中 $M = \max\{L, W\}$ ，引入 M 的目的是使这个不等式始终有效。

6.1.2 确定目标函数和决策变量

在给出所需约束条件后，接下来，我们将确定目标函数和决策变量。本题目标是找到一种使得木板利用率最大的方法，因此我们把目标函数表示为：

$$\min \left\{ LW - \sum_{k=0}^2 t_{ik} l^{(k)} w^{(k)} \right\}$$

在确定了切割木板产品的类型和数量后，我们才能确定它们占用的面积，以求得材料利用率。因此，我们确立了如下决策变量：

(1) 木板产品的型号变量 t_{ik} ：确定第 i 块木板的型号是否是 k ；

(2) 木板是否切割下的判断变量 C_i ：确定第 i 个木板产品是否已被切割下。

6.1.3 模型的建立

现在我们建立了一个切割矩形木块的数学模型如下：

$$\min \left\{ LW - \sum_{k=0}^2 t_{ik} l^{(k)} w^{(k)} \right\}$$

$$s. t. \begin{cases} x_i \in \left[C_i \frac{l_i}{2}, C_i \left(L - \frac{l_i}{2} \right) \right], y_i \in \left[C_i \frac{w_i}{2}, C_i \left(W - \frac{w_i}{2} \right) \right] \\ \max \left\{ |x_i - x_{i'}| - \left(\frac{l_i + l_{i'}}{2} \right), |y_i - y_{i'}| - \left(\frac{w_i + w_{i'}}{2} \right) \right\} \geq (1 - B_{ii'})(-M) \\ l_i = \sum_{k=1}^2 t_{ik} l^{(k)}, w_i = \sum_{k=1}^2 t_{ik} w^{(k)} \\ M = \max\{L, W\} \end{cases}$$

其中: C_i 为第 i 块木板是否切割下的判断变量; t_{ik} 为第 i 块木板的类型变量;

L 为大木板的长度, W 为大木板的宽度;

l_i 为第 i 块木板产品的长, w_i 为第块木板产品的宽;

$l^{(k)}$ 为型号为 k 的木板产品的长, $w^{(k)}$ 为型号为 k 的木板产品的宽。

至此, 矩形排布模型建立完毕。

6.1.4 约束条件的改进

在上式中, 木块产品的排列是非线性的。为了便于计算和求解, 我们将其线性化。在上述的对木板长宽度约束中, 木板产品至少在长度、宽度的一个维度内满足约束中。

$$\text{即} \max \left\{ |x_i - x_{i'}| - \left(\frac{l_i + l_{i'}}{2} \right), |y_i - y_{i'}| - \left(\frac{w_i + w_{i'}}{2} \right) \right\} \geq (1 - C_{ii'})(-M)$$

引入 M 的目的是使这个不等式始终有效。现对其非线性规划进行线性化:

Step 1: 将不等式的右侧更改为 0, 即移除不等式右侧的松弛变量。因此, 不等式变成了以下形式:

$$\max \left\{ |x_i - x_{i'}| - \left(\frac{l_i + l_{i'}}{2} \right), |y_i - y_{i'}| - \left(\frac{w_i + w_{i'}}{2} \right) \right\} \geq 0$$

Step 2: 将带最大函数的不等式扩展为两个不带最大函数的不等式:

$$\begin{cases} |x_i - x_{i'}| - \left(\frac{l_i + l_{i'}}{2} \right) \geq 0 \\ |y_i - y_{i'}| - \left(\frac{w_i + w_{i'}}{2} \right) \geq 0 \end{cases}$$

Step 3: 将带上述的带有绝对值的不等式扩展为下列不带绝对值的不等式:

$$\begin{cases} x_i - x_{i'} - \left(\frac{l_i + l_{i'}}{2} \right) \geq 0 \\ x_{i'} - x_i - \left(\frac{l_i + l_{i'}}{2} \right) \geq 0 \\ y_i - y_{i'} - \left(\frac{w_i + w_{i'}}{2} \right) \geq 0 \\ y_{i'} - y_i - \left(\frac{w_i + w_{i'}}{2} \right) \geq 0 \end{cases}$$

经过上述优化后，可以认为只要满足上述 Step 3 中不等式之一，Step 1 中的不等式将是有效的。因此，为了满足上述不等式系统，我们将引入松弛变量 A_k ，其中 $k = 1, 2, 3, 4$ 。此外， $C_{ii'} = \sum_{k=1}^4 A_k$ ， $M = \max\{L, W\}$ 用于松弛不等式。

得到如下不等式

$$\begin{cases} x_i - x_{i'} - \left(\frac{l_i + l_{i'}}{2}\right) \geq (1 - A_1)(-M) \\ x_{i'} - x_i - \left(\frac{l_i + l_{i'}}{2}\right) \geq (1 - A_2)(-M) \\ y_i - y_{i'} - \left(\frac{w_i + w_{i'}}{2}\right) \geq (1 - A_3)(-M) \\ y_{i'} - y_i - \left(\frac{w_i + w_{i'}}{2}\right) \geq (1 - A_4)(-M) \end{cases}$$

在上面的不等式中，当 $C_{ii'} = 1$ 时，存在一个 A_k 值为 1，而其他的 A_k 都是 0，即此不等式方程组中存在一个不等式不是松弛的，其他的不等式都是松弛的，只要一个不等式是有效的，原有不等式也是有效的。当 $C_{ii'} = 0$ 时，所有的 A_k 的值都是 0，即不等式都是松弛的，原来的不等式是有效的。

6.3 模型二的求解

根据建立的模型，本问采用遗传算法进行求解。在利用遗传算法时，首先需要建立基因编码方案。本文研究的问题中，S1 板和 P1、P3 产品都是规则的矩形，而矩形的放置存在横向放置、纵向放置两种状态：

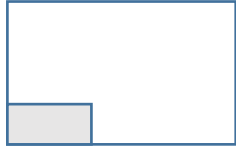


图 6.1 横向放置状态



图 6.2 纵向放置状态

如图 6.1、6.2 所示，横放状态即小矩形的长与 S1 板的长平行，纵放状态即小矩形的长与 S1 板的宽平行。将此二种状态编译为染色体的过程中，规定编号为正表示矩形放置中的横向放置状态，规定编号为负表示矩形放置中的纵向放置状态。

定义第一类染色体 GP_1 为该染色体为随机产生的第一类基因组：

$$GP_1 = (gp_{11}, gp_{12}, gp_{13}, \dots, gp_{1k}, \dots, gp_{1n})$$

$$\text{其中 } 1 \leq k \leq n, \quad -n \leq gp_{1k} \leq n \text{ 且 } gp_{1k} \neq 0, \quad |gp_{1k}| \neq |gp_{1k}|$$

此基因组内任意基因 gp_{1k} 表示在向 S1 木板中放置小矩形的过程中的第 k 个排入的小矩形。其中， $gp_{1k} > 0$ 表示此矩形横向放置， $gp_{1k} < 0$ 表示此矩形纵向放置。

定义第二类染色体 GP_2 为随机产生的第二类基因组：

$$GP_2 = (gp_{21}, gp_{22}, gp_{23}, \dots, gp_{2k}, \dots, gp_{2n})$$

$$\text{其中 } 1 \leq k \leq n, \quad gp_{2k} = 1 \text{ 或 } 3$$

那么，每一组基因 gp_1 和 gp_2 都表示矩形板的排入顺序，放置状态，产品种类。

例如， $\begin{cases} GP_1 = (-6, 4, -3, 2, 5, -1) \\ GP_2 = (1, 3, 1, 3, 3, 1) \end{cases}$ 所包含的信息可以表示为下表：

表 6.1 染该染色体组所包含的信息

排入顺序	基因编码数	产品名称	横纵放置状态
1	6	P1	纵向放置
2	4	P3	横向放置
3	3	P1	纵向放置
4	2	P3	横向放置
5	5	P3	横向放置
6	1	P1	纵向放置

确定染色体编码方案后，下面给出染色体的交叉和变异策略和适应度函数。

6.3.1 染色体的交叉

染色体的交叉是遗传算法中重要的迭代更新方法。针对矩形排样问题，本文中的遗传算子将采用单点交叉算子。基本方法如下：在选定的两个染色体 X, Y 中，在染色体的长度范围内，随机产生一个位置 s ($1 \leq s \leq n$)，染色体 X 中在 s 之前的基因段全部复制到子代染色体中 x 中，染色体 Y 中在 s 之后的基因段全部复制到子代染色体中 x 中，则子染色体 x 中的基因全部生成，子染色体 y 的生成策略相同。

6.3.2 染色体的变异

染色体的变异是遗传算法中另一种重要的迭代更新方法。变异使得群体中少数个体的一小部分染色体发生变异。对于本文建立的双层染色体编码结构，有如下策略，染色体 GP_{1i} 中发生变异时，设变异的位置为 s ($1 \leq s \leq n$)，此时染色体 GP_{1i} 中的基因 gp_{1i} 取其相反数；染色体 GP_{2i} 中发生变异时，变异策略为变异基因进行重新取值。

适应度函数：矩形排样问题主要计算占有率，各个子板材的占有率和初始板材的占有率，占有率越高，适应度越高，排版效果越好。本算法的适应度函数设置为已排入小矩形的面积总和与 $S1$ 木板的面积的比值。

6.3.3 剩余矩形

对模型进行初始化，模型产生初始放置方案即初始解的步骤如下：

Step1: 随机生成初始的染色体编译方案 $GP_j (gp_{11}, gp_{12}, gp_{13}, \dots, gp_{1j})$ ，从该染色体编译方案中可得知矩形放置的优先顺序，以及横纵放置状态。

Step2: 用一个矩形数据集合 R 来表示剩余可利用空间。本题中针对 SI 木板的初始 R 集合就是 $((0, 0), (L, W))$ 。之后每一个矩形排入前都需要与 R 进行比较，以确定是否还有空足够的间排入矩形板。排入后，更新剩余可利用空间 R 。

Step3:更新剩余可利用空间 R 后，继续进行 P1、P3 矩形板放置。此时需按照有一定先后之分的动态排样方法，放置 P1、P3 小木板。重复此种放置方法，直至 R 中无余地可放进任一 P1、P3 小矩形。

此处规定第一次更新剩余可利用空间 R 后放置 P1、P3 的过程按照如下方式进行：小矩形木板优先向下移动，在不能再移动时向左移动，在向左移的过程中如果发现存在可向下移动的空间那么接着向下移动，即下台阶规则进行放置。

遗传算法的参数设置为：交叉概率为 0.6，变异概率为 0.1，种群规模为 20，代沟为 0.3，进化代数为 15 代，40 代，100 代，对于给定的进化代数，利用 MATLAB 程序运行 5 次，取其中木板利用率最大值作为最终计算结果。经过上述模型建立与计算，可用 Excel 绘制出切割方案如图 6.3，图 6.4，图 6.5 所示：

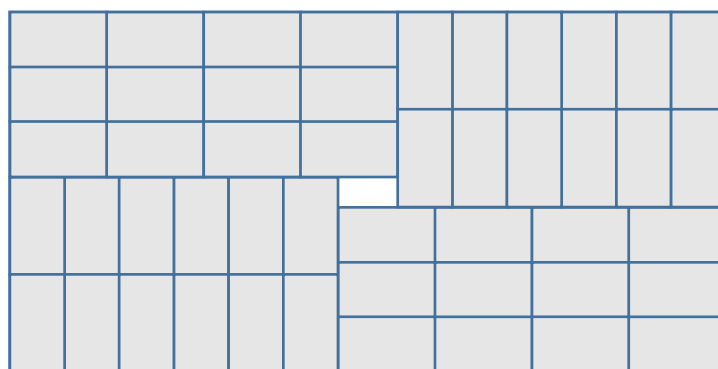


图 6.3 问题二切割方案 1

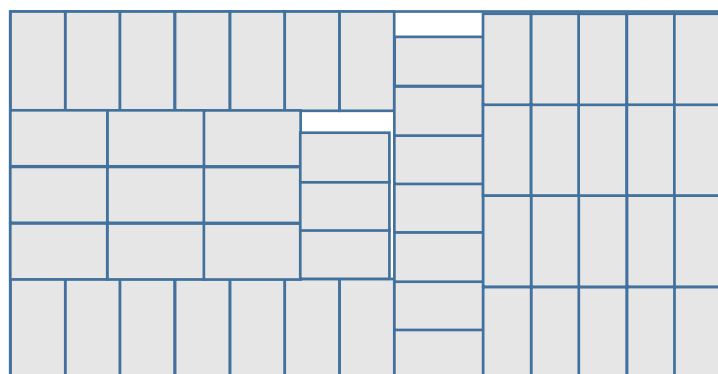


图 6.4 问题二切割方案 2

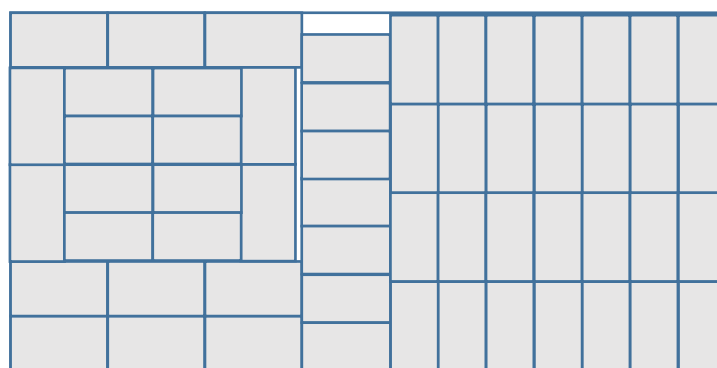


图 6.5 问题二切割方案 3

基于上述三种切割方式，可得出答案如下表所示：

表 6.2 问题的三结果

P1 的数量	P3 的数量	木板利用率
0	48	99.17%
43	13	98.50%
59	0	98.30%

6.4 结果分析

本问结果见上表。在本问的解答过程中，通过遗传算法得到上述切割方案，最优的切割方案为在一块 S1 上切割了 48 个 P3，木板利用率达到了 99.17%，但此方案的得出需要设置较大的进化代数并进行多次的计算才能得到，说明本算法对于逐步寻优还有一定的提升空间。

七、问题三的模型建立与求解

在第二问建立的矩形木板切割模型中，我们已经得出了切割 P1 和 P3 产品使木板利用率最高的前三种切割方案。我们在此计算结果上拓展，并建立整数规划模型，实现对切割方案的求解。

7.1 模型三的建立

第三问要求完成表 2 中给定数目的 P1 和 P3 产品的生产任务，并给出木板总利用率最高的切割方案。实际上，借助第二问中的遗传算法，我们可以进一步得出使木板利用率最高的前五种切割方案，如下表所示。

表 7.1 问题二中前五种利用率最高的方案

方案序号 i	P1 的数量	P3 的数量	木板利用率
1	0	48	99.2%
2	43	13	98.5%
3	59	0	98.5%
4	44	12	98.1%
5	30	23	97.5%

我们可认为前五种切割方案的利用率已达到最大，且恰好完成切割任务。故将这五种方案作为五种基本切割方案。以木板的总利用率为目标函数，五种方案的使用次数为决策变量，给定 P1、P3 的数量约束和总面积的约束，建立整数规划模型来求解此问题。

依照上表，令序号为 i 的方案的使用次数为 X_i ，并且方案中切割的 P1 个数为 m_i ，P3 的个数为 n_i ， $l^{(k)}$ 为型号为 k 的木板产品的长， $w^{(k)}$ 为型号为 k 的木板产品的宽，这

里 $k = 1, 3$ 。

则得到目标函数为：

$$\max \left\{ \frac{(\sum_{i=1}^5 m_i X_i) l^{(k)} w^{(k)} + (\sum_{i=1}^5 n_i X_i) l^{(3)} w^{(3)}}{(\sum_{i=1}^5 X_i) * 3000 * 1500} \right\}$$

约束条件：

$$s. t. \begin{cases} \sum_{i=1}^5 m_i X_i = 774 \\ \sum_{i=1}^5 n_i X_i = 1623 \\ X_i \geq 0 (i = 1, 2, 3, 4, 5) \end{cases}$$

7.2 模型三的求解

基于以上列出的目标函数和约束条件，运用 LINGO 软件编写程序求解，可得出答案如下表所示：

表 7.2 问题三的解答

木板 S1 的数量	P1 的数量	P3 的数量	木板 利用率	备注
29	0	1392	99.17%	每块木板切割方案相同
8	344	104	98.5%	每块木板切割方案相同
4	236	0	98.3%	每块木板切割方案相同
1	44	12	98.1%	每块木板切割方案相同
5	5150	115	97.5%	每块木板切割方案相同
合计数量： <u>47</u>	774	1623	木板总利用率： <u>98.78%</u>	木板总利用率 $\frac{\text{所有产品的总面积}}{\text{所有木板的总面积}}$

7.3 结果分析

本问结果见上表。在本问的解答过程中，利用第二问得到的一系列木板利用率较高的最优木材切割方案，建立起线性规划模型，以完成 P1，P3 生产任务为约束，最大化木板总利用率，得到一系列满足线性规划的解，即多重解。而在选取最优解时，通过总选取同时满足利用某切割方案次数较多和木板利用率较高的解作为最优解，可保留一些木板切割方案次数较少且木板利用率较低的解，则可以节省较多的余料，这种选择方案在实际生产过程中有一定参考价值。

八、问题四的模型建立与求解

此问基于第二三问进行改进，我们先用模型二的 xxx 模型和遗传算法算出一张木板

上的 P1、P2、P3、P4 的利用率较好的切割方案，取前 10 种最优利用率的方案为基本方案，再利用整数规划模型计算出木板总利用率最高的切割方案。

8.1 模型四的建立

第四问要求完成表 2 中给定数目的 P1、P2、P3、P4 产品的生产任务，并给出木板总利用率最高的切割方案。相比上一问增加了 P2、P4，同理我们利用第二问中的 xxx 模型，及借助遗传算法加以改进，算出在一张木板上切割 P1、P2、P3、P4 产品，使木板利用率最高的前十种切割方案，如下表所示。

表 8.1 问题四中一块前十种利用率最高的方案

方案序号 <i>i</i>	P1 的数量	P2 的数量	P3 的数量	P4 的数量	木板利用率
1	0	0	0	64	99.52%
2	0	0	48	0	99.17%
3	0	2	45	0	98.95%
4	0	33	0	0	98.64%
5	20	8	14	8	98.60%
6	0	6	39	0	98.51%
7	28	8	9	6	98.49%
8	0	18	8	18	98.32%
9	59	0	0	0	98.30%
10	0	20	8	14	98.08%

同理，我们可以认为此十种切割方案的利用率已达到最大，且恰好完成切割任务。故将这十种方案作为十种基本的切割方案。以木板的总利用率为目标函数，十种方案的使用次数为决策变量，给定 P1、P2、P3、P4 的数量约束和总面积的约束，建立整数规划模型来求解此问题。

依照上表，令序号为*i*的方案的使用次数为 X_i ，并且方案中切割的 P1 个数为 m_i ，P2 的个数为 n_i ，P3 的个数为 p_i ，P4 的个数为 q_i 。 $l^{(k)}$ 为型号为 k 的木板产品的长， $w^{(k)}$ 为型号为 k 的木板产品的宽，这里 $k = 1, 2, 3, 4$ 。

则得到目标函数为：

$$\max \left\{ \frac{(\sum_{i=1}^{10} m_i X_i) l^{(1)} w^{(1)} + (\sum_{i=1}^{10} n_i X_i) l^{(2)} w^{(2)} + (\sum_{i=1}^{10} p_i X_i) l^{(3)} w^{(3)} + (\sum_{i=1}^{10} q_i X_i) l^{(4)} w^{(4)}}{(\sum_{i=1}^{10} X_i) * 3000 * 1500} \right\}$$

约束条件：

$$s. t. \begin{cases} \sum_{i=1}^{10} m_i X_i = 774 \\ \sum_{i=1}^{10} n_i X_i = 2153 \\ \sum_{i=1}^{10} p_i X_i = 1623 \\ \sum_{i=1}^{10} q_i X_i = 1614 \\ X_i \geq 0 \\ (i = 1 \sim 10) \end{cases}$$

8.2 模型四的求解

基于以上列出的目标函数和约束条件，运用 LINGO 软件编写程序求解，可得出答案如下表所示：

表 8.2 问题四的解答

木板 S1 的数量	P1 的数量	P2 的数量	P3 的数量	P4 的数量	木板 利用率	备注
6	0	0	0	384	99.52%	每块木板切割方案相同
10	0	0	480	0	99.17%	每块木板切割方案相同
7	0	14	315	0	98.95%	每块木板切割方案相同
19	0	627	0	0	98.64%	每块木板切割方案相同
3	0	18	117	0	98.51%	每块木板切割方案相同
15	420	120	135	90	98.49%	每块木板切割方案相同
33	0	594	264	594	98.32%	每块木板切割方案相同
6	354	0	0	0	98.30%	每块木板切割方案相同
39	0	780	312	546	98.08%	每块木板切割方案相同
合计数量： <u>138</u>	774	2153	1623	1614	木板总利用率： <u>98.47%</u>	木板总利用率 $\frac{\text{所有产品的总面积}}{\text{所有木板的总面积}}$

8.3 结果分析

本问结果见上表。在本问的解答过程中，通过扩展第二问的模型，得到的一系列木板利用率较高的最优木材切割方案，建立起线性模型，以完成木板产品 P_i 生产任务为约束，最大化木板总利用率，得到一系列满足线性规划的解。事实上，可以通过寻求更多木板利用率较高的最优单块木板切割方案来进行更大规模的线性规划问题，从而寻得更多的可行解，通过比较进而得到最优解。

九、问题五模型的建立与求解

9.1 模型的建立

问题五与前四问相比出现了新的条件变化,本问中已知有 100 张 S1 木板及 P1、P2、P3、P4 每件产品的利润,与之前不同的是没有四种产品的需求数量限制,要求给出总利润最大的切割方案。此问可采用整数规划:设第 i 个切割方案中 P1 个数为 m_i , P2 的个数为 n_i , P3 的个数为 p_i , P4 的个数为 q_i 按照题中给出的利润,可得到问题五的目标函数为利润的最大值:

$$\max Z = 19.9m_i + 23n_i + 21p_i + 16q_i \quad (i = 1, 2, 3, \dots)$$

由题目中木板的尺寸、产品的尺寸可以得到木板、各类产品的面积。已知 S1 木板为 100 张的限制条件,可得问题五的约束条件:

$$s.t. \begin{cases} 7.4973m_i + 13.4514n_i + 9.2974p_i + 6.9975q_i \leq 45000 \\ m_i \geq 0 \\ n_i \geq 0 \\ p_i \geq 0 \\ q_i \geq 0 \end{cases}$$

9.2 模型的求解

基于以上列出的目标函数和约束条件,运用 LINGO 软件编写程序求解,可得出答案如下表所示:

表 9.2 问题五切割方案

木板 S1 的数量	P1 的数量	P2 的数量	P3 的数量	P4 的数量	利润	木板利用率	备注
100	5900	0	0	0	117410	98.30%	每块木板切割方案相同
木板 S1 合计数 100	5900	0	0	0	总利润: <u>117410</u>	木板 总利用率: <u>98.30%</u>	木板总利用率 = $\frac{\text{所有产品的总面积}}{\text{所有木板的总面积}}$

9.3 结果分析

本问结果见上表。在本问的解答过程中,我们将 100 块 S1 木板的最大化利润切割方案的求解简化为在一块木板 S1 上求解最大化利润切割方案,从而求解出最大总利润及其对应的切割方案。由此可见,合理简化思路能极大地简化大规模的问题并快速得到最优解。

十、模型评价

10.1 模型的优点

(1) 在第一问中,我们将在大木板上切割小木板的问题转化为用多个相同类型的小矩形块由外到内去填充 S1 木板的问题,该模型可操作性强,计算简单,具有极强的适应性和可变性,该模型思想可以在市场上推广使用。

(2) 在第二问中用到了遗传算法, 并与剩余矩阵思想相结合放置小矩形, 通过生物中交叉和变异的思想, 以木板利用率最大为目标, 对方案进行筛选优化, 这是能良好适应 NP-hard 问题的解法, 解法智能、高效, 具有创新性和科学性。

(3) 在解决第三、四、五问的过程中, 合理地利用了前一问所获得的结论, 对应用于自身上的模型进行优化, 此类思想高效、巧妙。

10.2 模型的缺点

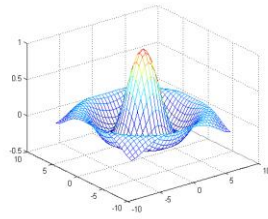
(1) 本文中运用到的遗传算法更多考虑到的是全局最优而非局部最优, 这会导致响应时间变长, 运算时间延长。

(2) 本文中模型的建立条件较为理想, 未能考虑到现实影响, 可以根据现实生活中人为失误、下刀顺序等因素带来的影响, 建立更为精确、智能的模型。

十一、参考文献

- [1] 司守奎, 孙兆亮, 数学建模算法与应用, 国防工业出版社, 2017 年
- [2] 宋雪枫, 陈江义, 张明伟, 基于遗传算法的矩形件排样问题, 煤矿机械, 第 32 卷第 05 期: 2011, 119-120
- [3] 韩喜君, 丁根宏, 矩形件优化排样问题的混合遗传算法求解, 计算机技术与发展, 第 16 卷第 06 期: 2006, 219-221
- [4] 庞剑飞, 宋丽娟. 矩形排样问题的优化设计模型, 现代制造工程, 第 02 期: 2014, 88-90
- [5] 王洪冠, 基于经验的板材切割问题, 吉林林业科技, 第 02 期: 1994 年, 59-62
- [6] 董德威, 颜云辉, 等. 矩形件优化排样的自适应遗传模拟退火算法[J]. 中国机械工程, 第 24 卷第 18 期: 2013 年, 2499-2504
- [7] 宋开胜, 基于遗传算法的二维排样研究, <https://www.docin.com/p-1398021839.html?qq-pf-to=pcqq.c2c>, 2019 年 5 月 2 日
- [8] 麻林立, 板材切割问题 LINGO 求解, <https://wenku.baidu.com/view/4ecce76fa45177232f60a2eb.html>, 2019 年 5 月 2 日

五一数学建模竞赛



题 目：_____本板最优切割方案_____

关键词：混合整数规划 最优切割 枚举法 遗传算法 Matlab 软件 LINGO 软件

摘 要：

本文主要研究了在给定条件下如何对木板最优切割问题，重点研究分析了不同情况下的切割方案，通过建立线性规划模型和混合整数规划模型，并利用 MATLAB 和 LINGO 对提出的模型进行求解，具体结果如下：

对于**问题一**：考虑到只切割 P1 产品，若想木板利用率最高，只需切割产品的总面积最大，根据两个产品不能重叠，引入了坐标的思想，建立线性规划模型，结合 MATLAB 里的 *fmincon* 函数求解，得出：在只切割 P1 产品时，一块木板上能够切割 **59** 块 P1 产品，最高木板利用率为 **98.30%**。

对于**问题二**：在问题一模型的基础上，添加变量 k_i 表示产品种类，建立新的线性规划模型，结合遗传算法求解，在两种产品数量可以为 **0** 的情况下，木板利用率由高到底排序的前三种切割方案为：

方案 1：在木板全部切割 P3 产品，能够切割 **48** 块 P3，此时木板利用率为 **99.17%**；

方案 2：在木板上切割 **47** 块 P3 产品和 **1** 块 P1 产品，此时木板利用率为 **98.77%**；

方案 3：在木板上切割 **46** 块 P3 产品和 **2** 块 P1 产品，此时木板利用率为 **98.37%**。

对于**问题三**：根据问题一、二中提出的线性规划模型，对问题 3 分析并建立新的整数规划模型，得到最优切割方案如下：用 **33** 个木板全部切割 P3，得到 **1584** 个 P3，剩余 **39** 个，利用率为 **99.17%**；用 **13** 个木板全部切割 P1，得到 **767** 个 P1，剩余 **7** 个 P1，利用率为 **98.30%**；再将剩余的 7 个 P1 和 39 个 P3 用 1 个木板切割，利用率为 **92.23%**；所需木板总数量为 **47**，木板总利用率为 **98.78%**。

对于**问题四**：首先考虑切割面积较大的产品，每次切割按面积大小依次切割，由于选取方案较多，只枚举部分切割方案，以切割的木板数量最少为目标函数，建立混合整数规划模型。运用 LINGO 求解，得到结果：需要木板 S1 的总数量为 **138**，木板总利用率为 **98.47%**，具体方案见表 6。

对于**问题五**：同问题 4 的思想相同，在分割时，首先考虑尽可能的切割单位面积利润较大的产品，即首先切割 P1 产品，建立线性规划模型，通过 LINGO 求解，得到结果：在 100 块木板上只切割 P1，总数量为 **5900**，其余产品数量均为 **0**，总利润为 **117410** 元，木板总利用率为 **98.30%**

最后，深入研究了论文中提出的模型并加以分析，对模型的优化提供了一些改进思路。

木板最优切割方案

一、问题重述

徐州某家具厂新进一批木板如表 1 所示，在家具加工的过程中，需要使用切割工具生产表 2 所示的产品。假设：木板厚度和割缝宽度忽略不计。已知产品可以纵横排放，请为该家具厂给出如下 5 个问题的木板最优切割方案。

问题 1：在一块木板上切割 P1 产品，建立数学模型，给出木板利用率最高(即剩余木板面积最小)的切割方案，计算产品的数量及木板利用率。

问题 2：在一块木板上切割 P1 和 P3 两种产品，建立数学模型，给出按照木板利用率由高到低排序的前 3 种切割方案，计算三种方案对应两种产品的数量及木板利用率。

问题 3：需要完成表 2 中 P1 和 P3 产品的生产任务，建立数学模型，给出木板总利用率最高的切割方案，并计算木板的总利用率。

问题 4：在上述问题的基础上，需要完成表 2 中 P1、P2、P3、P4 产品的生产任务，建立数学模型，给出木板总利用率最高的切割方案，计算木板 S1 的数量和总利用率。

问题 5：不考虑产品 P1，P2，P3，P4 的需求数量，给定 100 张 S1 木板，按照表 2 中给出的利润，建立数学模型，给出总利润最大的切割方案，计算在该切割下的总利润及木板的总利用率。

表 1 木板的尺寸

木板	长度(mm)	宽度(mm)
S1	3000	1500

表 2 产品尺寸及生产任务

产品名称	长度(mm)	宽度(mm)	生产任务(件)	利润(元/件)
P1	373	201	774	19.9
P2	477	282	2153	23.0
P3	406	229	1623	21.0
P4	311	225	1614	16.0

二、问题分析

此问题为板材下料问题，只切割一种产品时，切割方案比较容易，对于切割多种产品时，需要尽可能的求解板材切割方案，确定不同的方案使得木板得到更多的利用。

对于问题 1，求解出只切割 P1 产品的木板利用率最高的切割方案，若想木板利用率最高，只需切割产品的总面积最大，即在只切割同一种产品的情况下，只需切割产品的数量最多，问题可以转化为：经过对木板切割，使得获得的 P1 产品数量

最多，建立线性规划模型，结合 MATLAB 里的 fmincon 函数求解。

对于问题 2，求解出在一块木板上切割 P1 和 P3 产品，木板利用率由高到低排序前 3 的切割方案，对问题进行分析，可以将此问题转化为：在 P1 和 P3 产品个数可以为 0 的情况下，在木板上切割 P1 和 P3 产品，可将问题转化为两种产品的总面积之和最大的问题。同问题 1 一样，用线性规划模型，结合 MATLAB 里的 fmincon 函数求解。

对于问题 3，要求出木板总利用率最高的切割方案，经过对问题 3 的分析，可将问题转化为所需木板数量最少的问题，即两种木板的总面积除以所需木板的总面积最大。同问题 1、2 一样，用线性规划建立模型，结合 MATLAB 里的 fmincon 函数求解。

对于问题 4，在完成 P1, P2, P3, P4 产品的生产任务的条件下，找出木板总利用率最高的切割方案。通过建立相关模型，我们首先找出利用率较高的 12 种切割方案，再建立线性规划模型对这 12 种切割方案进行优化组合，通过 LINGO 函数进行求解。

对于问题 5，在给定 100 张 S1 木板，不考虑产品 P1, P2, P3, P4 的需求数量的条件下，找出总利润最大切割方案，经过对问题 5 的分析，可以将其转化为找出单张木板的切割利润最大的问题。类似问题 4，建立线性规划模型，通过 LINGO 函数进行求解。

三、问题假设

由于产品要求，需要对木板进行一定的打磨去除粗边的加工，切割后的木板也需要进行一定的加工处理才可投入使用。切割时需要考虑切割损失等影响因素，因此，对于本题，我们做出如下假设：

- (1) 不考虑切割时木板产生的损耗。
- (2) 木板可以沿着任意方向切割。
- (3) 木板可以任意摆放。
- (4) 用材不能重叠，不超过原材料的大小。
- (5) 原板材密度厚度均匀，形状为规则矩形。
- (6) 木板厚度和割缝宽度忽略不计。

四、符号说明

n	一块木板上切割的产品总块数
(x_i, y_i)	第 i 块产品的中心坐标
d_i	产品放置情况
$d_i = 1$	第 i 块产品的长边水平切割
$d_i = -1$	第 i 块 P1 产品短边的水平切割
k_i	产品种类
$k_i = 1$	切割产品为 P1
$k_i = -1$	切割产品为 P3

z_i	第 i 种切割方案所用的原木板的数量
L	使用的木板数
n_l	第 l 块木板上产品总数
(x_{li}, y_{li})	木板 l 上第 i 块产品的坐标

五、模型建立与问题求解

5.1 问题 1 的模型建立与求解

经过对问题 1 的分析，若想木板利用率最高，只需切割产品的总面积最大，在只切割同一种产品的情况下，只需切割产品的数量最多，因此，本题的目标是：经过对木板切割，使得获得的 P1 产品数量最多，现建立以下线性规划模型。

5.1.1 目标函数的建立

将木板如图所示摆放，四个顶点记为 O, A, B, C ，以 OA 所在直线为 x 轴， OC 所在直线为 y 轴建立平面直角坐标系。记在一块木板上切割 P1 产品的块数为 n ($1 \leq n \leq 60, n \in \mathbb{Z}$)。

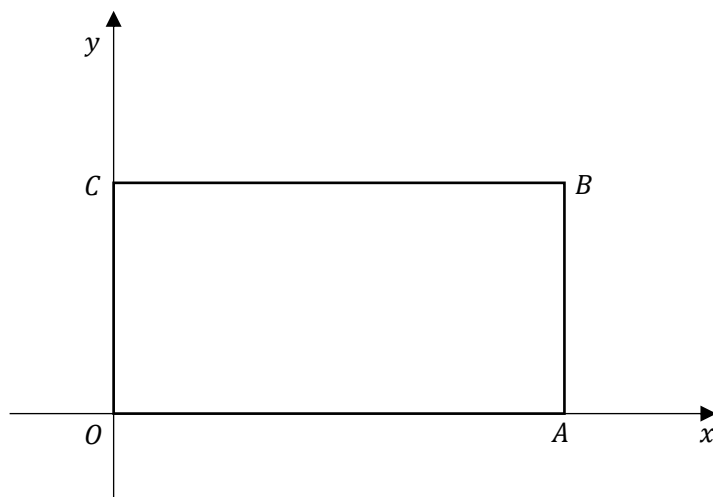


图 1 木板平面图

第 i 块 P1 产品中心的坐标与切割方向记为 $(x_i, y_i, d_i), i = 1, \dots, n, d_i = \pm 1$ ，其中 (x_i, y_i) 表示第 i 块 P1 产品的中心坐标， $d_i = 1$ 表示第 i 块 P1 产品的长边水平切割， $d_i = -1$ 表示第 i 块 P1 产品短边的水平切割。为了使木板的利用率最大，只需总的 P1 产品所占的面积最大，即 $373 \times 201 \times n$ 最大，其中 373×201 为一块 P1 产品的面积，建立目标函数 I:

$$\max f(x) = 74973n$$

5.1.2 约束条件

(1) 若要保证任意两块 P1 产品不重合，则 x_i 和 x_j 之间的距离满足：

$$|x_i - x_j| \geq \frac{d_i + 1}{4} \times 373 + \frac{d_i - 1}{-4} \times 201 + \frac{d_j + 1}{4} \times 373 + \frac{d_j - 1}{-4} \times 201$$

或 y_i 和 y_j 之间的距离满足：

$$|y_i - y_j| \geq \frac{d_i + 1}{4} \times 201 + \frac{d_i - 1}{-4} \times 373 + \frac{d_j + 1}{4} \times 201 + \frac{d_j - 1}{-4} \times 373$$

其中 $i, j = 1, \dots, n, i \neq j$.

只需满足上述条件中的其一，即可保证产品切割时不重叠；

(2) P1 产品数量不小于1，且不大于60，则有 $1 \leq n \leq 60$ ；

(3) 为了保证每块产品均落在木板之内，只需坐标满足：

$$\begin{aligned} \frac{d_i + 1}{4} \times 373 + \frac{d_i - 1}{-4} \times 201 \leq x_i \leq 3000 - \left[\frac{d_i + 1}{4} \times 373 + \frac{d_i - 1}{-4} \times 201 \right] \\ \frac{d_i + 1}{4} \times 201 + \frac{d_i - 1}{-4} \times 373 \leq y_i \leq 1500 - \left[\frac{d_i + 1}{4} \times 201 + \frac{d_i - 1}{-4} \times 373 \right] \end{aligned}$$

其中 $i = 1, \dots, n$ 。

产品间隔示意图如下所示：

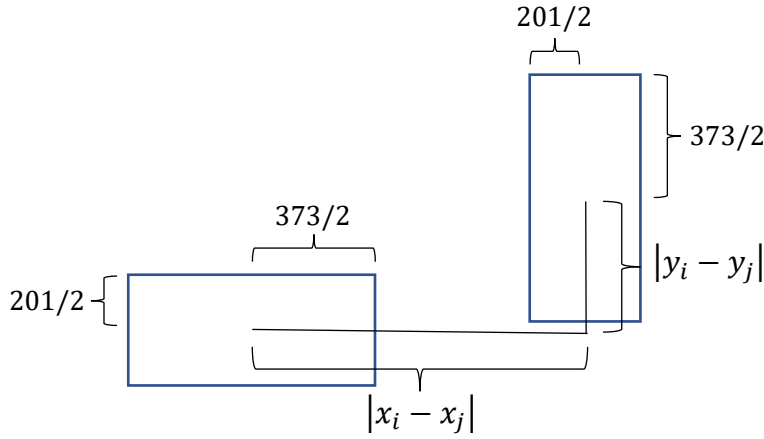


图 2 P1 产品间隔图

将上述约束条件化简，得到优化模型 I：

$$\begin{aligned} \max f(x) &= 74973n \\ \text{s.t.} \quad &\begin{cases} |x_i - x_j| \geq 43(d_i + d_j) + 287 \\ \text{or} \quad |y_i - y_j| \geq -43(d_i + d_j) + 287, i, j = 1, \dots, n, i \neq j \\ \frac{86d_i + 287}{2} \leq x_i \leq \frac{-86d_i + 5713}{2} \\ \frac{-86d_i + 287}{2} \leq y_i \leq \frac{86d_i + 287}{2}, i = 1, \dots, n \\ 1 \leq n \leq 60 \end{cases} \end{aligned}$$

根据问题一中的约束条件,利用 MATLAB 编程，得到结果：每块木板最多切割 P1 产品 **59** 块，方案之一为将木板横切 P1 产品 **7** 块，竖切 **52** 块。最终的 matlab 程序由附录 1 给出，方案如下图 3 所示：

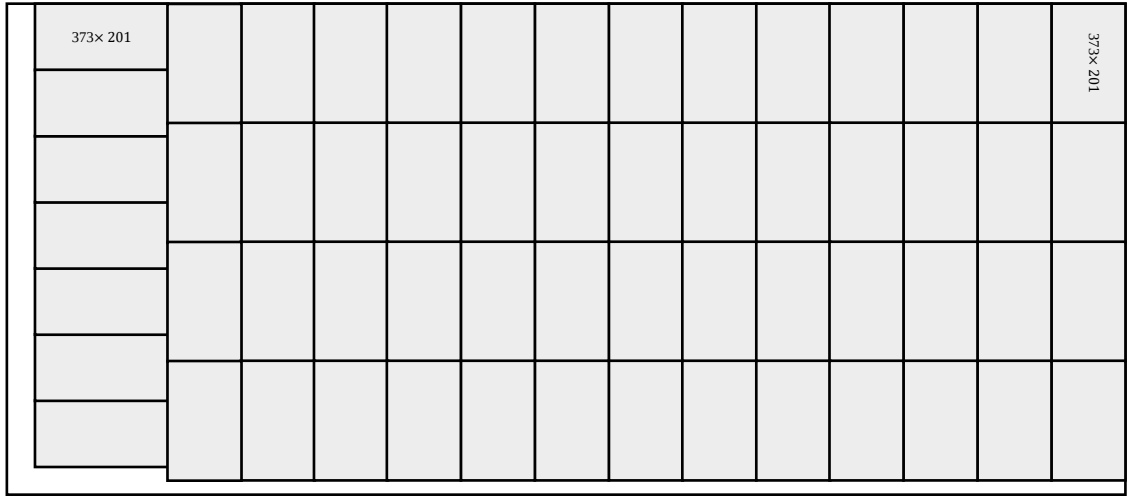


图 3 P1 产品切割图

结果见表 3:

表 3 问题 1 的结果

P1 的数量	木板利用率
59	98.30%

5.2 问题 2 的模型建立与求解

由问题 1 可得木板切割 P1 产品的最高利用率为 98.30%，这是只考虑切割一种产品的情况下。经过对问题 2 的分析，可知本题所要实现的目标是：在木板上切割 P1 和 P3 产品，找出木板利用率由高到低的前三种切割方案。可将问题转化两种产品的总面积之和最大的问题。

5.2.1 目标函数的建立

记在一块木板上切割 P1 和 P3 产品的总块数为 $n(1 \leq n \leq 60, n \in \mathbf{Z})$ 。第 i 块产品的中心坐标、种类与切割方向记为 $(x_i, y_i, d_i, k_i), i = 1, \dots, n$, $d_i = \pm 1, k_i = \pm 1$, 其中 (x_i, y_i) 为产品的中心坐标; d_i 表示产品放置情况, $d_i = 1$ 表示第 i 块产品长边水平切割, $d_i = -1$ 表示第 i 块产品的短边水平切割; k_i 表示产品种类, $k_i = 1$ 表示切割产品为 P1, $k_i = -1$ 表示切割产品为 P3。为了使木板的利用率最大, 只需两类产品的总面积最大. 产品 P1 的数量可以用 $\sum_{i=1}^n \frac{k_i+1}{2}$ 表示, 产品 P3 的数量可以用 $\sum_{i=1}^n \frac{k_i-1}{-2}$ 表示, 故两类产品的总面积可以表示为 $\sum_{i=1}^n \left(\frac{k_i+1}{2} \times 373 \times 201 + \frac{k_i-1}{-2} \times 406 \times 229 \right)$, 整理得 $\sum_{i=1}^n \frac{-18001k_i+167947}{2}$, 故以 $\sum_{i=1}^n \frac{-18001k_i+167947}{2}$ 最大为目标, 建立目标函数:

$$\max f(x) = \sum_{i=1}^n \frac{-18001k_i + 167947}{2}$$

5.2.2 约束条件

(1) 所有产品需要落在木板之内, 则中心坐标需要满足以下约束条件:

$$\frac{(d_i+1)(k_i+1)}{4} \times \frac{373}{2} + \frac{(d_i-1)(k_i+1)}{-4} \times \frac{201}{2} + \frac{(d_i+1)(k_i-1)}{-4} \times \frac{406}{2} + \frac{(d_i-1)(k_i-1)}{4} \times \frac{229}{2} \leq x_i \leq$$

$$3000 - \left[\frac{(d_i+1)(k_i+1)}{4} \times \frac{373}{2} + \frac{(d_i-1)(k_i+1)}{-4} \times \frac{201}{2} + \frac{(d_i+1)(k_i-1)}{-4} \times \frac{406}{2} + \frac{(d_i-1)(k_i-1)}{4} \times \frac{229}{2} \right],$$

$$\frac{(d_i+1)(k_i+1)}{4} \times \frac{201}{2} + \frac{(d_i-1)(k_i+1)}{-4} \times \frac{373}{2} + \frac{(d_i+1)(k_i-1)}{-4} \times \frac{229}{2} + \frac{(d_i-1)(k_i-1)}{4} \times \frac{406}{2} \leq y_i \leq$$

$$1500 - \left[\frac{(d_i+1)(k_i+1)}{4} \times \frac{201}{2} + \frac{(d_i-1)(k_i+1)}{-4} \times \frac{373}{2} + \frac{(d_i+1)(k_i-1)}{-4} \times \frac{229}{2} + \frac{(d_i-1)(k_i-1)}{4} \times \frac{406}{2} \right],$$

$$i = 1, \dots, n, \text{ 整理得 } \frac{-5k_id_i+349d_i-61k_i+1209}{8} \leq x_i \leq 3000 - \frac{-5k_id_i+349d_i-61k_i+1209}{8},$$

$$\frac{5k_id_i-349d_i-61k_i+1209}{8} \leq y_i \leq 1500 - \frac{5k_id_i-349d_i-61k_i+1209}{8}, i = 1, \dots, n.$$

(2) 产品数量之和不小于1, 且不大于60, 则有 $1 \leq n \leq 60$;

(3) 同问题一中的约束条件, 若在保证任意两块产品不重叠, 则任意两块产品中

心的横坐标距离满足约束条件: $|x_i - x_j| \geq \frac{(d_i+1)(k_i+1)}{4} \times \frac{373}{2} + \frac{(d_i-1)(k_i+1)}{-4} \times \frac{201}{2} +$

$\frac{(d_i+1)(k_i-1)}{-4} \times \frac{406}{2} + \frac{(d_i-1)(k_i-1)}{4} \times \frac{229}{2} + \frac{(d_j+1)(k_j+1)}{4} \times \frac{373}{2} + \frac{(d_j-1)(k_j+1)}{-4} \times \frac{201}{2} +$

$\frac{(d_j+1)(k_j-1)}{-4} \times \frac{406}{2} + \frac{(d_j-1)(k_j-1)}{4} \times \frac{229}{2}$, 或纵坐标距离满足约束条件: $|y_i - y_j| \geq$

$\frac{(d_i+1)(k_i+1)}{4} \times \frac{201}{2} + \frac{(d_i-1)(k_i+1)}{-4} \times \frac{373}{2} + \frac{(d_i+1)(k_i-1)}{-4} \times \frac{229}{2} + \frac{(d_i-1)(k_i-1)}{4} \times \frac{406}{2} +$

$\frac{(d_j+1)(k_j+1)}{4} \times \frac{201}{2} + \frac{(d_j-1)(k_j+1)}{-4} \times \frac{373}{2} + \frac{(d_j+1)(k_j-1)}{-4} \times \frac{229}{2} + \frac{(d_j-1)(k_j-1)}{4} \times \frac{406}{2}$, 整理

得 : $|x_i - x_j| \geq \frac{-5k_id_i+349d_i-61k_i-5k_jd_j+349d_j-61k_j+2418}{8}$, 或

$|y_i - y_j| \geq \frac{5k_id_i-349d_i-61k_i+5k_jd_j-349d_j-61k_j+2418}{8}, i, j = 1, \dots, n, i \neq j.$

综合上述约束条件, 得到优化模型 II:

$$\max f(x) = \sum_{i=1}^n \frac{-18001k_i + 167947}{2}$$

$$\text{s.t.} \begin{cases} |x_i - x_j| \geq \frac{-5k_id_i+349d_i-61k_i-5k_jd_j+349d_j-61k_j+2418}{8} \quad or \\ |y_i - y_j| \geq \frac{5k_id_i-349d_i-61k_i+5k_jd_j-349d_j-61k_j+2418}{8}, i, j = 1, \dots, n, i \neq j \\ \frac{-5k_id_i+349d_i-61k_i+1209}{8} \leq x_i \leq 3000 - \frac{-5k_id_i+349d_i-61k_i+1209}{8} \\ \frac{5k_id_i-349d_i-61k_i+1209}{8} \leq y_i \leq 1500 - \frac{5k_id_i-349d_i-61k_i+1209}{8}, i = 1, \dots, n \\ 1 \leq n \leq 60, n \in \mathbb{Z} \end{cases}$$

根据问题 2 的模型, 利用 MATLAB 编程 (程序见附录 2)。

● 情况一: 在产品 P1 和 P3 的数量可以等于 0 的情况下, 得到如下 3 种方案:

方案 1: 将木板全部切割 P3 产品, 1 块木板上能够切割 48 块 P3, 此时木板利用率 99.17%, 模型见图 4:

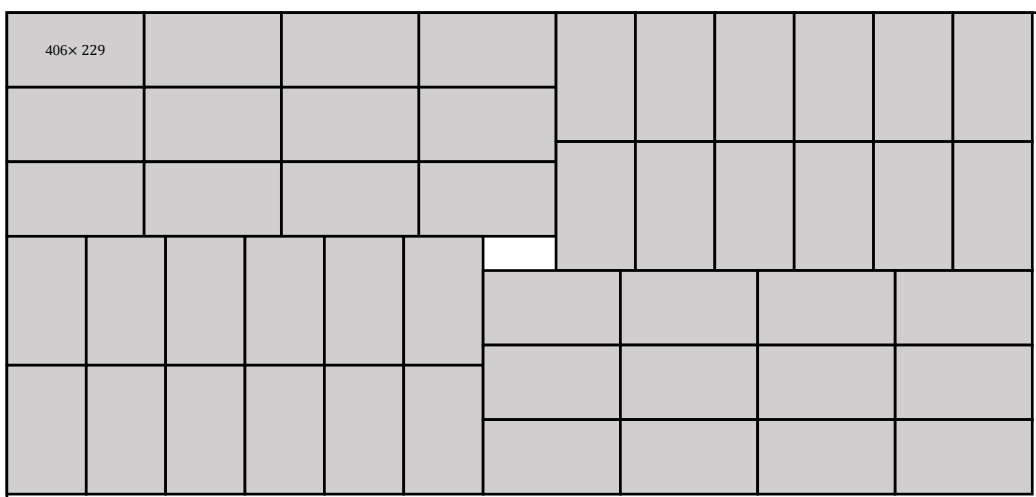


图 4 方案 1

方案 2: 将木板上切割 47 块 P3 产品和 1 块 P1 产品, 此时木板利用率为 98.77%, 模型见图 5:

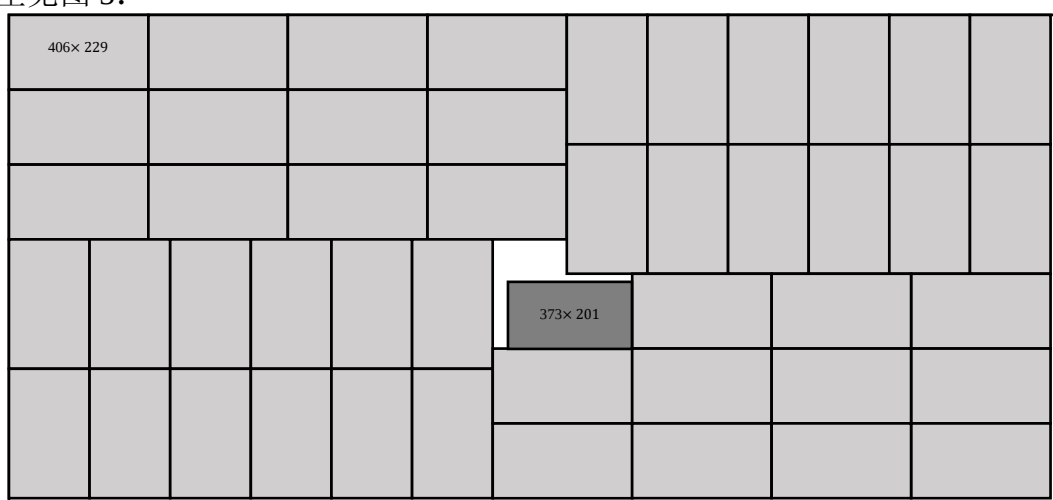


图 5 方案 2

方案 3: 将木板切割 46 块 P3 产品和 2 块 P1 产品, 此时木板利用率为 98.37%, 模型见图 6:

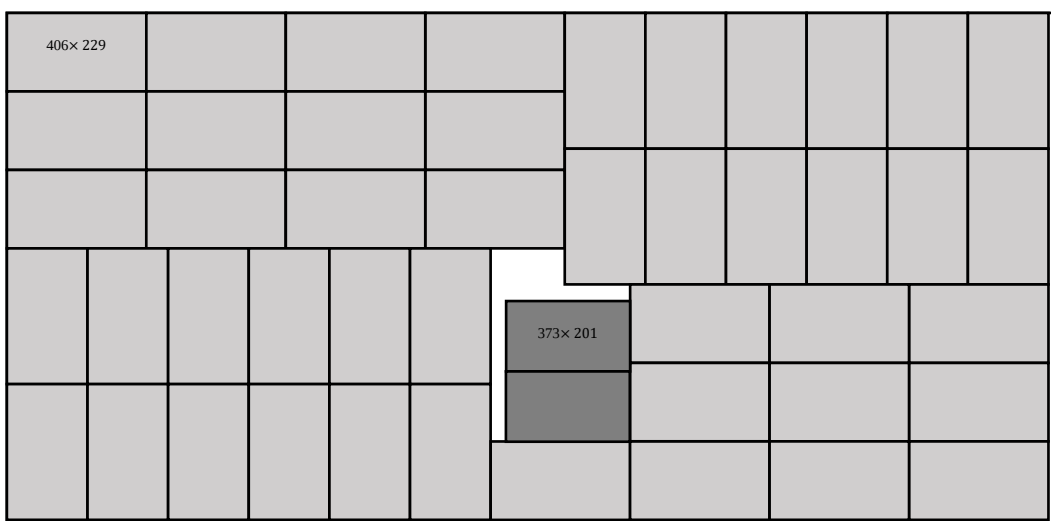


图 6 方案 3

具体方案见下表 4.1:

表 4.1 问题 2 的结果方案

方案编号	P1 的数量	P3 的数量	木板利用率
1	0	48	99.17%
2	1	47	98.77%
3	2	46	98.37%

- 情况二：在产品 P1 和 P3 的数量都不为 0 的情况下，由模型求解，得到如下 3 种方案：

表 4.2 问题 2 的结果方案

方案编号	P1 的数量	P3 的数量	木板利用率
1	1	47	98.77%
2	43	13	98.50%
3	2	46	98.37%

5.3 问题 3 的模型建立与求解

由问题 2 得出将木板全部切割 P3 产品时的木板利用率为 99.17%；将木板上切割 47 块 P3 产品和 1 块 P1 产品时的木板利用率为 98.77%；将木板切割 46 块 P3 产品和 2 块 P1 产品时的木板利用率为 98.37%。经过对问题 3 的分析，可将问题转化为所需木板数量最少的问题，即两种木板的总面积除以所需木板的总面积最大。

5.3.1 目标函数的建立

设使用木板数为 L ，考虑到每块木板得切割方案不同，设第 l 块木板上产品总数为 $n_l, l = 1, \dots, L$. 第 l 块木板上的每块产品用 $(x_{l_i}, y_{l_i}, d_{l_i}, k_{l_i}), i = 1, \dots, n_l, d_{l_i} = \pm 1, k_{l_i} = \pm 1$ 表示，其中 (x_{l_i}, y_{l_i}) 为木板 l 上第 i 块产品的坐标； d_{l_i} 表示产品的种类， $d_{l_i} = 1$ 表示第 i 块产品长边水平切割， $d_{l_i} = -1$ 表示第 i 块产品的短边水平切割； k_{l_i} 表示产品种类， $k_{l_i} = 1$ 表示切割产品为 P1， $k_{l_i} = -1$ 表示切割产品为 P3。木板 l 上产品 P1 的数量可以用 $\sum_{i=1}^{n_l} \frac{k_{l_i}+1}{2}$ 表示，产品 P3 的数量可以用 $\sum_{i=1}^{n_l} \frac{k_{l_i}-1}{-2}$ 表示，故木板 l 上产品总面积为 $\sum_{i=1}^{n_l} \left(\frac{k_{l_i}+1}{2} \times 373 \times 201 + \frac{k_{l_i}-1}{-2} \times 406 \times 229 \right)$. 故 L 块木板，切割出产品 P1 的数量为 $\sum_{l=1}^L \sum_{i=1}^{n_l} \frac{k_{l_i}+1}{2}$ ，切割出产品 P3 的数量为 $\sum_{l=1}^L \sum_{i=1}^{n_l} \frac{k_{l_i}-1}{-2}$ ，产品总面积为 $\sum_{l=1}^L \sum_{i=1}^{n_l} \left(\frac{k_{l_i}+1}{2} \times 373 \times 201 + \frac{k_{l_i}-1}{-2} \times 406 \times 229 \right)$ ，故木板的总利用率为 $\frac{\sum_{l=1}^L \sum_{i=1}^{n_l} \left(\frac{k_{l_i}+1}{2} \times 373 \times 201 + \frac{k_{l_i}-1}{-2} \times 406 \times 229 \right)}{3000 \times 1500 \times L}$ ，同模型 II，每块产品在其对应的木板内，中心坐

标需要满足对应的约束条件；除此以外，产品 P1 的数量不少于 774，即 $\sum_{n_l}^L \sum_{i=1}^{n_l} \frac{k_{l_i}+1}{2} \geq 774$ ；产品 P3 的数量不少于 1623，即 $\sum_{n_l}^L \sum_{i=1}^{n_l} \frac{k_{l_i}-1}{-2} \geq 1623$ ，整理上述约束条件，得到优化模型 III：

$$\begin{aligned} \max \quad & \frac{\sum_{n_l}^L \sum_{i=1}^{n_l} \left(\frac{k_{l_i}+1}{2} \times 373 \times 201 + \frac{k_{l_i}-1}{-2} \times 406 \times 229 \right)}{3000 \times 1500 \times L} \\ \text{s. t.} \quad & \begin{cases} |x_{l_i} - x_{l_j}| \geq \frac{-5k_{l_i}d_{l_i} + 349d_{l_i} - 61k_{l_i} - 5k_{l_j}d_{l_j} + 349d_{l_j} - 61k_{l_j} + 2418}{8} \text{ or} \\ |y_{l_i} - y_{l_j}| \geq \frac{5k_{l_i}d_{l_i} - 349d_{l_i} - 61k_{l_i} + 5k_{l_j}d_{l_j} - 349d_{l_j} - 61k_{l_j} + 2418}{8} \\ i, j = 1, \dots, n_l, i \neq j, n_l = 1, \dots, L \\ -\frac{5k_{l_i}d_{l_i} + 349d_{l_i} - 61k_{l_i} + 1209}{8} \leq x_{l_i} \leq 3000 - \frac{-5k_{l_i}d_{l_i} + 349d_{l_i} - 61k_{l_i} + 1209}{8} \\ \frac{5k_{l_i}d_{l_i} - 349d_{l_i} - 61k_{l_i} + 1209}{8} \leq y_{l_i} \leq 1500 - \frac{5k_{l_i}d_{l_i} - 349d_{l_i} - 61k_{l_i} + 1209}{8} \\ i = 1, \dots, n_l, n_l = 1, \dots, L \\ \sum_{n_l}^L \sum_{i=1}^{n_l} \frac{k_{l_i}+1}{2} \geq 774 \\ \sum_{n_l}^L \sum_{i=1}^{n_l} \frac{k_{l_i}-1}{-2} \geq 1623 \\ 1 \leq n \leq 60, n \in \mathbf{Z}, n_l \in \mathbf{Z}^* \end{cases} \end{aligned}$$

根据问题 3 的模型，用 MATLAB 编程，运行得到以下方案：

用 33 个木板全部切割 P3，那么可以切割 1584 个 P3 产品，还剩余 39 个，用 13 个木板全部切割 P1，那么可以切割 767 个 P1 产品，还剩余 7 个，再拿出一块木板切割剩余的 39 块 P3 产品和 7 块 P1 产品，因此一共需要 33+13+1=47 块木板。具体方案见下表 5：

表 5 问题 3 的结果

木板 S1 的数量	P1 的数量	P3 的数量	木板利用率	备注
33	0	1584	99.17%	每块木板切割方案相同
13	767	0	98.30%	同上
1	7	39	92.23%	同上
合计数量： 47	774	1623	木板 总利用率： 98.78%	木板总利用率 = $\frac{\text{所有产品的总面积}}{\text{所有木板的总面积}}$

5.4 问题 4 的模型建立与求解

5.4.1 问题的分析

对于问题 4，由于切割方案较多，需要确定哪些切割方案是可行的，无法进行一一列举，这里从特殊情况出发，根据各种产品的尺寸及生产任务，求解几种利用率较高的切割方案，建立线性规划模型并进行求解。

假设每个木板上只切割同一种产品，可以得到如下情况：

表 5.41 同种产品的切割情况表

产品名称	最多可切割的数量	需要块数	需要原木板 S1 的数量
P1	59	774	14
P2	33	2153	66
P3	48	1623	34
P4	64	1614	26

由产品最多可切割的数量可知，P2 产品的面积最大，我们这里考虑首先尽可能的切割面积较大的产品，即首先切割 P2 产品，其次为 P3，P1，P4。每次切割按面积大小依次切割，面积大的产品优先且尽可能多的切割，当面积大的产品切割完后，不再考虑切割此产品，由程序可得，选取方案较多，在这里我们仅考虑 12 种利用率较高的切割方案，如下表所示：

表 5.42 方案情况表

产品名称	P1	P2	P3	P4	利用率
方案 1	12	14	12	18	99.07%
方案 2	1	0	47	0	98.77%
方案 3	14	13	13	6	98.37%
方案 4	2	19	8	14	98.43%
方案 5	1	6	33	7	98.67%
方案 6	6	24	4	5	97.77%
方案 7	35	0	19	0	97.56%
方案 8	0	18	1	26	96.3%
方案 9	0	5	21	16	83.21%
方案 10	13	11	15	6	94.86%
方案 11	2	17	3	22	94.55%
方案 12	16	15	0	14	93.26%

问题可转化为在满足需要的条件下，按照哪些合理的方案，切割的原木板的数量最少，下面将对此目标进行讨论。

5.4.2 模型建立及求解

设 z_i 表示第 i 种切割方案所用的原木板的数量， z_i 均为非负整数。以切割的原材料板材的数量最少为目标函数，建立如下模型：

$$\begin{aligned}
 \text{Min } f(z) &= z_1 + z_2 + z_3 + z_4 + z_5 + z_6 + z_7 + z_8 + z_9 + z_{10} + z_{11} + z_{12} \\
 \text{s.t. } \begin{cases} 12z_1 + z_2 + 14z_3 + 2z_4 + z_5 + 6z_6 + 35z_7 + 13z_{10} + 2z_{11} + 16z_{12} \geq 774 \\ 14z_1 + 13z_3 + 19z_4 + 6z_5 + 24z_6 + 18z_8 + 5z_9 + 11z_{10} + 17z_{11} + 15z_{12} \geq 2153 \\ 12z_1 + 47z_2 + 13z_3 + 8z_4 + 33z_5 + 4z_6 + 19z_7 + z_8 + 21z_9 + 15z_{10} + 3z_{11} \geq 1623 \\ 18z_1 + 6z_3 + 14z_4 + 7z_5 + 5z_6 + 26z_8 + 16z_9 + 6z_{10} + 22z_{11} + 14z_{12} \geq 1614 \end{cases}
 \end{aligned}$$

根据问题 4 的模型,编写 LINGO 程序（附录 4），运行得到以下最优解：

$$z_1 = 62, z_2 = 0, z_3 = 0, z_4 = 53, z_5 = 14, z_6 = 0$$

$$z_7 = 0, z_8 = 8, z_9 = 1, z_{10} = 0, z_{11} = 0, z_{12} = 0$$

具体方案见下表 6：

表 6 问题 4 的结果

木板 S1 的数量	P1 的数量	P2 的数量	P3 的数量	P4 的数量	木板 利用率	备注
53	636	742	636	424	99.07%	每块木板切割方案相同
14	14	84	462	98	98.67%	同上
62	124	1178	496	868	98.43%	同上
8	0	144	8	208	96.30%	同上
1	0	5	21	16	83.21%	同上
合计数量： 138	774	2153	1623	1614	木板 总利用 率:98.47%	木板总利用率 = $\frac{\text{所有产品的总面积}}{\text{所有木板的总面积}}$

5.5 问题 5 的模型建立与求解

5.5.1 问题的分析

通过对问题的分析我们可知，此问题可以转化为在不考虑产品 P1，P2，P3，P4 的需求数量，给定 100 张 S1 木板，且单个产品的利润一定的前提下，找出总利润最大的切割方案。

表 5.51 产品利润表

产品名称	产品面积（ mm^2 ）	利润（元/件）	单位面积的利润
P1	74973	19.9	0.000265
P2	134514	23.0	0.000171
P3	92974	21.0	0.000226
P4	69975	16.0	0.000228

通过对比，可以得出单位面积的利润 $P1 > P4 > P3 > P2$ ，同问题 4 思想相同，在分割时，首先考虑尽可能的切割单位面积利润较大的产品，即首先切割 P1 产品，其次为 P4，P3，P2。记 $t_i (i=1,2,3,4)$ 为在一块木板上切割四种产品对应的数量， $t_i \in \mathbf{Z}^+$ ，可以建立如下线性规划模型：

$$\begin{aligned} \max Y &= 100 \times (19.9t_1 + 23t_2 + 21t_3 + 16t_4) \\ \text{s.t. } \begin{cases} 33 \leq t_1 + t_2 + t_3 + t_4 \leq 64 \\ 74973t_1 + 134514t_2 + 92924t_3 + 69975t_4 \leq 450000 \end{cases} \end{aligned}$$

根据问题 5 的模型,编写 LINGO 程序（附录 5），运行得到以下最优解：

$$t_1 = 59, t_2 = 0, t_3 = 0, t_4 = 0$$

具体方案见下表 7：

木板 S1 的数量	P1 的数量	P2 的数量	P3 的数量	P4 的数量	利润	木板 利用率	备注
100	5900	0	0	0	117410	98.30%	每块木板切割方案相同
木板 S1 合计数量 100	5900	0	0	0	总利润: 117410	木板 总利用率: 98.30%	木板总利用率 = $\frac{\text{所有产品的总面积}}{\text{所有木板的总面积}}$

六、模型的分析、改进及推广

6.1. 模型的分析

通过 MATLAB、LINGO 编程能够计算出大量的切割方案，本文只选取部分较优的切割方案，问题 1 中根据两个产品不能重叠，引入了坐标的思想，以此建立相关模型。问题 2 和问题 3 的模型是在问题 1 模型的基础上建立的，问题 4 和问题 5 思想相同，根据所给出的权重切割产品。

6.2. 模型的改进与推广

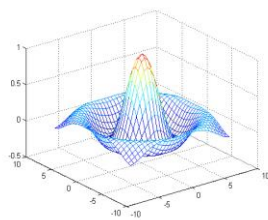
1. 由于问题 4 中的切割方案较多，我们这里只考虑了部分切割方案。
2. 对于问题 5 在实际情况中，不确定因素太多，尽可能增加这些考虑，结果会更加合理。
3. 因数据资料搜集的不完整，准确性也有待商榷，而且没有对最终方案进行更为细致的讨论研究，这些方面有待改进。

七、参考文献

参考文献：

[1] 姜启源, 谢金星, 叶俊,《数学模型(第三版)》,北京: 高等教育出版社, 2003,73-85
 [2] 谢金星, 薛毅,《优化建模与 LINDO/LINGO 软件》,北京: 清华大学出版社, 2005, 105-120
 [3] 卓金武,《MATLAB 在数学建模中的应用》, 成都, 西南交通大学出版社, 2005,73-86

五一数学建模竞赛



题 目： 基于多元线性回归模型对科创板问题的研究

关键词： 主成分分析法 灰色预测模型 多元线性回归模型 估值量化模型

摘 要：

本文针对科创板模拟上市企业估值的问题进行研究，使用主成分分析、回归分析等方法，建立灰色预测模型、多元线性回归模型以及估值量化模型。

问题一要求计算 2018 年中国 A 股市场和美国 NASDAQ 市场的估值水平以及估值溢价或折价水平建立估值算法公式，和题目附件中所给实际估值水平对比，得出两个市场都是溢价水平。

问题二要求定量分析两个市场的估值指标与基本面指标、流动性指标之间存在的关系。本题使用主成分分析的方法，根据总得分表达式中各自变量前的系数大小对指标的重要程度做出分析。总得分表达式记为：

$$F = \sum_{i=1}^6 a_i Z X_i$$

得出结论：中国 A 股市场受流动性指标影响程度较大，而美国 NASDAQ 市场受基本面指标影响程度较大。

问题三要求预测两个市场 2019 年的基本面指标、流动性指标以及估值指标。本题通过近十年各指标的平均值，先建立灰色预测模型得出两个市场基本面指标和流动性指标的预测值，再通过建立多元线性回归模型：

$$y = 0.010x_5 - 1.217 \times 10^{-11}x_6 + 1.103, y = \sum_{i=1}^6 b_i x_i + c$$

得出中、美两个市场 2019 年的估值指标分别为 4.726 和 2.335。

问题四要求预测我国首批科创板企业的估值水平。根据问题一中有关估值的概念并结合美国 NASDAQ 市场的平均市销率数据建立估值量化模型：

$$VL_{19} = \frac{\overline{Bi_{19}}}{\overline{Bi_{18}}} \times \overline{P/S_{19}}$$

进而通过问题三中中国 A 股市场流动性指标的预测值计算出我国首批科创板企业的估值水平为 5.4。

一、问题背景与重述

1.1 背景分析

科创板是习近平在中国进口博览会上宣布成立，作为新设板块，独立于现有主板市场，并在此内进行注册制试点。通过交易、发行，和引入中长期资金链配套措施，增加试点，逐步争取科创板实现市场平衡、融资平衡、利益平衡，有效促进市场形成更加良好的预期。科创板重点在新材料、新能源等新兴产业上进行发展，推动高新科技产业的融合，满足了国家的重大需求，能够符合国家战略。科创板作为一个新的板块，发行压力并不大，有更多能满足条件的中小型企业能够上市，它对于营业收入、净利润没有具体要求，会有更多的科技股诞生，中国能衍生出更多的高新科技，能够有利于国家产业结构的升级。从创业板到新三板，再到现在的科创板，都有十分重要的作用，是股票市场的新兴力量。

纳斯达克是美国全国证券交易商协会自动报价表，能有效地规范混乱的场外交易和为小企业提供融资平台。它作为美国资本市场的重要力量，培养了许多新兴产业，帮助美国转型。它是首家电子化的股票市场，比其他股票市场更加先进，有利于美国资本市场经济的发展，做到了技术密集型产业来代替传统的服务业，让美国的经济长期诱惑力。科创板有望成为中国的纳斯达克，它有中国政策的支持，参与科创板投资的公司市价会更加合理，而且所需成本较低，科创板有着较好地发展前景。但是，科创板是新设的板块，它从监管到市场都并不十分成熟，属于中国资本市场制度改革的一种实验，发展的时间比较短，表明了它存在一些不利的竞争因素。不能达到美国的纳斯达克的相同水准。但从另一方面来说，科创板的出现和发展能够让我国的各大不同的金融机构进入投行业务中，实现金融强国的国家重要战略，改变了金融体系的脆弱性，深化中国金融供给侧结构性改革。

1.2 问题重述

题目对于新设板块科创板拟上市企业进行估值。科创板补充了创新科技企业的融资短板，有希望成为中国的纳斯达克。那么本题就对科创板企业上市后进行估值，可以采用市盈率法，由于是新型板块，不考虑企业盈利能力，可以参考美国上市公司的平均市销率，但由于国内和国外企业的迎来能力及各方面的不同，造成会存在较大差异，导致中国和美国的股票市场存在估值溢价或折价。

根据题目所给的数据资料，建立数学模型解决以下 4 个问题：

- 1、根据题目所给数据，用平均市销率作为估值水平，算出两个市场的估值水平，再计算出它们的估值溢价或折价水平；
- 2、建立模型，对中国 A 股市场的 3 个指标，估值指标、基本面指标、流动性指标进行定量分析，这 3 个指标每个又包含 3 个小指标，比较其差异；
- 3、分别对两个股票市场的基本面指标，流动性指标进行预测分析，并根据结果，计算它们 2019 年的估值指标；
- 4、根据题目所给数据，参照美国市场的估值量化模型、中国 A 股的流动性指标，预测我国科创板上市后的估值水平。

二、问题分析

2.1 对于问题一的分析

问题一要求以平均市销率作为估值水平，计算 2018 年中国 A 股市场以及美国 NASDAQ 市场的估值水平，进而求出估值溢价或折价水平。由于大数据中总有错误信息的存在，首先要对数据进行预处理，删除所有干扰信息。考虑到市销率的合理范围为 (0,10)，可以先对该区间分段，并从投资的角度对每段区间进行分析，从而理解股票市场的估值水平，然后根据所列的相关公式对平均市销率即估值水平进行计算。溢价和折价指的是基金交易时的交易价格与基金价值的价差量，二者属于相对关系，而估值溢价或折价是对估值即平均市销率的变化所作的分析，故思考问题的角度不应只停留在折溢价上，而需要与估值变化量相结合。

2.2 对于问题二的分析

问题二要求建立数学模型，定量分析中国 A 股市场和美国 NASDAQ 市场的估值指标与基本面指标和流动性指标之间存在的关系，并对两个市场进行比较。首先对估值指标进行定值处理，即确定一个平均市销率，为使数据包含的信息更加准确、广泛，本文选定问题一中计算出的中、美市场平均市销率为参考值，并对 2009-2018 年每年中市销率近似等于该值的上市公司整合后，可以进行主成分分析。由于 6 个指标之间的单位不尽相同，因此需要对数据进行归一化处理。然后根据得出的综合得分表达式中各指标前系数的大小对 6 个指标的重要程度作出分析，亦可将其理解为 6 个指标与估值指标之间的关系。最后根据基本面指标和流动性指标对上述 6 个指标的分类情况综合分析各自与估值指标之间的关系，并将中国 A 股市场与美国 NASDAQ 市场得出的结论进行对比。

2.3 对于问题三的分析

问题三要求建立数学模型，预测中国 A 股市场与美国 NASDAQ 市场的 2019 年基本面指标和流动性指标，并根据预测的结果计算中、美市场 2019 年的估值指标。观察数据可知，需要对 2009-2018 年的数据进行均值处理，即利用 Excel 计算出各指标在本年度中的平均值，以便于指标的预测和估值指标的计算。

针对于基本面指标和流动性指标的预测问题，可以建立灰色预测模型对数据进行分析，根据 6 个指标 10 年的数据变化趋势，预测出中、美两个市场 2019 年的数据。针对于两个市场 2019 年估值指标的计算问题，可以使用逐步回归分析的方法，建立多元线性回归模型。逐步回归法是对全部因子按照自身对被解释变量影响程度从大到小依次引入回归方程的预测方法，通过 SPSS 软件对数据进行分析处理，利用显著性较高的自变量构建有关估值指标的多元线性回归方程。最后将使用灰色预测模型得到的数据代入方程即可求出中国 A 股市场和美国 NASDAQ 市场 2019 年的估值指标。

2.4 对于问题四的分析

问题四要求根据 93 家企业的基本面数据、美国 NASDAQ 市场的估值量化模型以及中国 A 股市场的流动性指标来预测我国首批科创板企业的估值水平。首先要假设估值水平仍然是平均市销率，然后求出 93 家企业 2016-2018 年基本面指标的平均值，并根据问题一中有关估值的概念和问题三中求出的中国 A 股市场的流动性指标对估值水平进行判断。

三、模型假设

结合本题实际，为确保模型求解准确性和合理性，本文排除一些因素干扰，提出以下几点假设：

- 1、企业在数据统计期间不受自然灾害、金融危机等因素的影响；
- 2、假设国内市场在 2018 年中的未知数据相对于前几年无大幅度变动；
- 3、假设题目中除了选取到的因素外，其他影响因素的影响效果可忽略不计。

四、符号说明

为便于问题的求解，本文给出以下符号说明：

符号	说明
SSR	回归离差平方和
Bi	主营业务收入
Msr	市销率
Tmv	总市值
Var	方差
Cov	协方差
VI	估值水平
θ	可容覆盖区间
SST	总离均差平方和
λ	原始时间序列的级比

五、模型的建立与求解

经过以上分析和准备，本文将逐步建立以下数学模型，进一步阐述模型的实际建立过程。

5.1 问题一模型的建立与求解

5.1.1 问题一数据的预处理

由附件 1 中的数据可知，其中包含了中国 A 股市场和 B 股市场（上证指数成分股）的共同数据，为方便计算 2018 年中国 A 股市场的估值水平，本文将所涉及的 B 股市场的数据全部剔除，并将所有企业按照估值数据市销率从小到大进行排序。对于估值数据市销率为 0 的企业来说，由于其流动性指标为 0，可理解为该企业在本年度未进行股券交易，故将该类型企业全部剔除。

由附件 2 中的数据可知，美国 NASDAQ 市场因为资本主义政治制度的潜在影响，企业很容易面临停产、倒闭等情况，因此可将只有企业名称无指标数据（即停产）的企业以及营业收入无数据（即倒闭）的企业全部剔除。又当对美国 NASDAQ 市场的数据进行排序时，发现估值数据市销率中存在负数，而由公式：

$$Msr = \frac{P}{S} = \frac{Tmv}{Bi}$$

其中， Msr 为市销率（下文写作 P/S ）， P 为每股价格， S 为每股销售额， Tmv 为总市值， Bi 为去年主营业务收入。由此可知市销率一定大于 0，故需要将市销率为负值的

企业数据全部剔除。又通过查阅相关资料，可知市销率大于 10 的企业数据并无意义，则对附件 1、附件 2 中该类数据全部剔除。

5.1.2 问题一模型的建立

本文选取平均市销率来作为估值水平。而在股票的投资问题中，根据专家的意见，市销率在 0-1 之间时，表明股票市场低估了该上市公司的实际能力，同时也意味着该公司的股票具有很大的潜力和较高的投资价值；市销率在 1-2 之间时，对于新型的上市公司而言属于正常现象，而对于成熟的上市公司来说，意味着本公司的经济发展状况良好；而当市销率大于或等于 4 时，表明股票市场高估了该上市公司的实际能力，股价是虚高的，具有很大的风险性和极小的投资价值。

针对于上述思想，为使估值水平更加精确，本文对中国 A 股市场和美国 NASDAQ 市场的市销率进行分段处理，利用 Excel 软件对其平均市销率进行计算，结果如下表所示：

表 1 2018 年两个市场的平均市销率

市销率区间	(0,2)	[2,4)	[4,6)	[6,8)	[8,10)
中国 A 股市场	1.03	2.82	4.91	6.83	8.94
美国 NASDAQ 市场	0.83	2.94	4.88	6.92	8.91

由上表可知，市销率在 (0,2) 之间时，中国 A 股市场的平均市销率明显大于美国 NASDAQ 市场，说明目前中国的上市公司中成长型科技创新企业数量较多，相对于美国的上市公司而言仍处于成长阶段，甚至是初级阶段；市销率在 [2,10) 之间时，中、美两国的平均市销率基本趋于一致，说明成熟的上市公司都具有完善的经营管理体制，无论是中国还是美国，对于本公司的平稳运营都能起到关键性的作用。

对整体数据的计算步骤如下：

Step1: 根据上述条件，中国 A 股市场和美国 NASDAQ 市场的平均市销率 $\overline{P/S}$ 的计算公式为：

$$\overline{P/S} = \frac{\sum_{i=1}^6 (\overline{p/s_i} \cdot n_i)}{N}$$

其中， $\overline{p/s_i}$ 为表 1 中各区间的平均市销率， n_i 为各区间的公司数量， N 为参与（包括倒闭企业）本年度股市交易的上市公司数量。

Step2: 所求出的平均市销率为实际估值水平；

Step3: 溢价（折价）是指高出（低出）市场标准价格、标准利率或者证券票面价值的部分。那么，用公式计算出一个估值水平，具体公式如下：

$$VI = \frac{\overline{Bi_{18}}}{\overline{Bi_{17}}} \times \overline{P/S_{18}}$$

其中， VI 为 2018 年的估值水平， $\overline{Bi_{17}}$ 为 2017 年平均营业收入， $\overline{P/S_{18}}$ 为 2018 年的平均市销率。

Step4: 所求的估值溢价就是将计算得到的 2018 的估值水平与实际估值水平进行对比，若大于 1，就说明股票市场为溢价水平，若在 0-1 之间，就是折价水平。

5. 1. 3 问题一模型的求解

通过上述公式，对两国股票市场的计算结果如下表：

表 2 2018 年两国股票市场的计算数据

	中国 A 股市场	美国 NASDAQ 市场
平均市销率（倍数）	4.896	4.846
2018 估值水平	4.357	3.420
平均溢/折价比例	1.123	1.417

由对上表分析可以知道，中国的 A 股市场为估值溢价水平，美国 NASDAQ 市场也为估值溢价水平，但两者相比较，两个股票市场的溢价水平仍存在差距，中国股票市场的估值溢价水平较低。

5. 2 问题二模型的建立与求解

5. 2. 1 数据的选取与模型的确立

问题二要求先对中国 A 股市场以及美国 NASDAQ 市场的估值指标（即市销率）与基本面指标和流动性指标间的关系作定量分析，再对二者进行差异比较。考虑到基本面指标和流动性指标各自都包含 3 个低级指标，本文使用主成分分析法对数据进行降维处理，建立评价模型，根据变量的系数对估值指标与 6 个低级指标间的关系进行说明。

主成分分析法要求估值指标应选取一个定值（近似即可）作为总综合得分的参考量。根据问题一的结果，本文选取中国 A 股市场和美国 NASDAQ 市场的估值指标（即平均市销率）分别为 4.89 和 4.84，以此为目标在 2009-2018 年中选出市销率最接近该值的上市公司作为研究对象，选择结果如下表所示：（完整表格见附录 I）

表 3 研究对象的选取

年份	中国 A 股市场	美国 NASDAQ 市场
2009	天宸股份	威瑞信 (VERISIGN)
2010	长春一东	微妙通讯
2011	锦江股份	MERIDIAN BIOSCIENCE
2012	洲际油气	网易
⋮	⋮	⋮
2015	内蒙一机	韩美金融
2016	康美药业	SOUTH STATE CORP
2017	广州港	布林摩尔银行
2018	熊猫金控	BANCFIRST

5. 2. 2 问题二评价模型的建立

首先假设 $X_1, X_2, \dots, X_6$ 为原始变量经处理后的指标， $F_1, \dots, F_m (m < 6)$ 为新变量指标， F 与 X 满足以下关系：

$$\begin{cases} F_1 = a_{11}X_1 + a_{12}X_2 + \dots + a_{16}X_6 \\ F_2 = a_{21}X_1 + a_{22}X_2 + \dots + a_{26}X_6 \\ \vdots \\ F_m = a_{m1}X_1 + a_{m2}X_2 + \dots + a_{m6}X_6 \end{cases}$$

其中， a, F 满足如下条件：

(1) 每个主成分的系数平方和为 1。即：

$$a_{1j}^2 + a_{2j}^2 + \cdots + a_{mj}^2 = 1$$

(2) 主成分之间相互独立，即无重叠的信息。即：

$$\text{Cov}(F_i + F_j) = 0, i \neq j, i, j = 1, 2, \cdots m.$$

(3) 主成分的方差依次递减，重要性依次递减。即：

$$\text{Var}(F_1) \geq \text{Var}(F_2) \geq \cdots \geq \text{Var}(F_m)$$

主成分分析步骤：（以中国 A 股市场的为例）

Step1：为了消除各项财务指标之间在量纲化和数量级上的差别，对指数数据进行归一化处理：

$$z_{ij} = \frac{\overline{x_{ij}} - \overline{x_j}}{s_j}, \quad \overline{x} = \frac{1}{n} \sum_{i=1}^n x_{ij}$$

表 4 归一化处理的数据

年份	ZX_1	ZX_2	ZX_3	ZX_4	ZX_5	ZX_6
2009	-0.568	-0.467	-0.287	0.226	1.035	-0.157
2010	-0.489	-0.469	0.198	-0.730	-0.061	-0.787
2011	-0.278	-0.178	0.240	-0.867	-0.975	-0.858
2012	-0.342	-0.122	1.261	0.449	-0.443	-0.252
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
2015	-0.295	-0.545	-1.103	-0.081	0.530	0.403
2016	2.665	2.767	0.665	0.993	-0.869	1.884
2017	0.655	0.190	-0.056	2.193	2.243	1.592
2018	-0.555	-0.544	-1.823	-0.906	-0.455	-0.969

（完整表格见附录 I）

Step2：求指标数据的相关系数矩阵：

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix}$$

其中， $r_{ij}(i, j = 1, 2, \cdots, p)$ 为原变量 x_i 与 x_j 的相关系数， $r_{ij} = r_{ji}$ ，其计算公式为：

$$r_{ij} = \frac{\sum_{k=1}^n (x_{ki} - \overline{x_i})(x_{kj} - \overline{x_j})}{\sqrt{\sum_{k=1}^n (x_{ki} - \overline{x_i})^2 \sum_{k=1}^n (x_{kj} - \overline{x_j})^2}}, (i = 1, \cdots, n; j = 1, \cdots, p)$$

通过 SPSS 软件，计算得到指标数据的相关系数如下表所示：

表 5 相关性系数

	ZX_1	ZX_2	ZX_3	ZX_4	ZX_5	ZX_6
ZX_1	1.000	0.978	0.256	0.603	-0.054	0.847
ZX_2	0.978	1.000	0.358	0.491	-0.214	0.748
ZX_3	0.256	0.358	1.000	0.184	-0.224	0.175
ZX_4	0.603	0.491	0.184	1.000	0.666	0.879
ZX_5	-0.054	-0.214	-0.224	0.666	1.000	0.428
ZX_6	0.847	0.748	0.175	0.879	0.428	1.000

故相关系数矩阵为：

$$R = \begin{bmatrix} 1.000 & 0.978 & 0.256 & 0.603 & -0.054 & 0.847 \\ 0.978 & 1.000 & 0.358 & 0.491 & -0.214 & 0.748 \\ 0.256 & 0.358 & 1.000 & 0.184 & -0.224 & 0.175 \\ 0.603 & 0.491 & 0.184 & 1.000 & 0.666 & 0.879 \\ -0.054 & -0.214 & -0.224 & 0.666 & 1.000 & 0.428 \\ 0.847 & 0.748 & 0.175 & 0.879 & 0.428 & 1.000 \end{bmatrix}$$

Step3：计算特征值和特征向量。解特征方程 $|\lambda I - R| = 0$ ，求出的特征值按大小顺序排列 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ ；然后分别求出特征值的向量 $a_i (i=1, 2, \dots, p)$ 且 $|a_i| = 1$ 。

Step4：计算主成分的方差贡献率及累积方差贡献率：

1、方差贡献率：

$$\frac{\lambda_i}{\sum_{k=1}^p \lambda_k} (i=1, 2, \dots, p)$$

2、累积方差贡献率：

$$\frac{\sum_{k=1}^i \lambda_k}{\sum_{k=1}^p \lambda_k} (i=1, 2, \dots, p)$$

一般取累计贡献率达到 80% 以上的特征值， $\lambda_1, \lambda_2, \dots, \lambda_m$ 对应第 1、第 2、……、第 $m (m \leq p)$ 个主成分。

表 6 总方差的解释

成分	初始特征值			提取载荷平方和		
	总计	方差百分比	累积 (%)	总计	方差百分比	累积 (%)
1	3.435	57.251	57.251	3.435	57.251	57.251
2	1.654	27.574	84.824	1.654	27.574	84.824
3	0.805	13.410	98.234			
4	0.064	1.062	99.295			
5	0.038	0.632	99.928			
6	0.004	0.072	100.000			

由表 3 可知成分 1 与成分 2 的累计贡献率之和为 84.824%，大于 80%，故成分 1 与成分 2 可分别对应主成分 F_1 与 F_2 。也可利用 SPSS 软件绘制的碎石图加以说明，如下图所示：

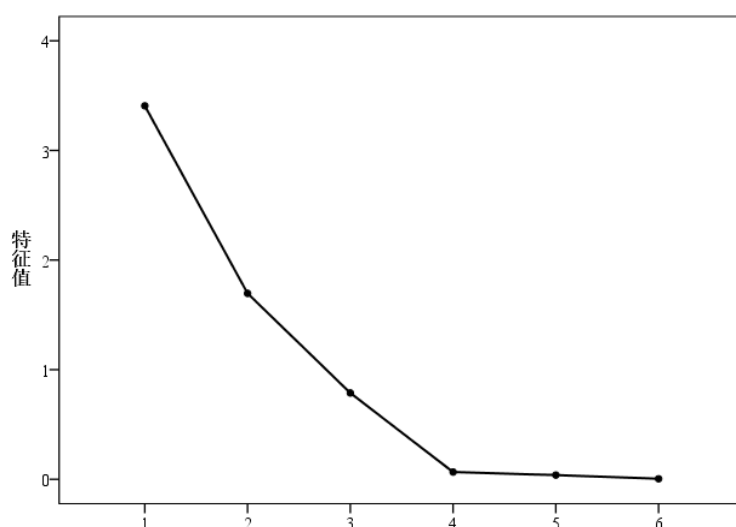


图 1 碎石图

在碎石图中，我们取特征值大于 1 的为主成分，则依照上图可清晰直观的反映出应选择前两个为主成分。

Step5：计算主成分载荷量：

$$l_{ij} = \rho(z_{ij}, x_j) = \sqrt{\lambda_i a_{ij}} \quad (i, j = 1, 2 \cdots p)$$

得到因子载荷矩阵（成份得分系数矩阵）。如下表所示：

表 7 成份得分系数矩阵

指标	成分 1	成分 2
营业收入（元）	0.268	-0.175
归母净利润（元）	0.251	-0.275
净资产收益率（%）	0.097	-0.304
年成交量（股）	0.248	0.274
年平均换手率（%）	0.080	0.563
年成交额（元）	0.281	0.115

表中的每一列表示一个主成分作为原来变量线性组合的系数。因此，我们得到两个主成分关于指标的线性组合为：

$$\begin{cases} F_1 = 0.268ZX_1 + 0.251ZX_2 + 0.097ZX_3 + 0.248ZX_4 + 0.080ZX_5 + 0.281ZX_6 \\ F_2 = -0.175ZX_1 - 0.275ZX_2 - 0.304ZX_3 + 0.274ZX_4 + 0.563ZX_5 + 0.115ZX_6 \end{cases}$$

于是综合得分表达式可表示为：

$$F = 0.123ZX_1 + 0.081ZX_2 - 0.034ZX_3 + 0.259ZX_4 + 0.234ZX_5 + 0.227ZX_6$$

同理可得到美国 NASDAQ 市场综合得分表达式可表示为：

$$F = 0.208ZX_1 + 0.264ZX_2 + 0.198ZX_3 + 0.105ZX_4 + 0.071ZX_5 + 0.158ZX_6$$

5.2.3 问题二评价模型的求解

根据中国 A 股市场的综合得分公式可知，营业收入、归母净利润、净资产收益率、

年成交量、年平均换手率以及年成交额对总得分的影响系数分别为 0.123、0.081、-0.034、0.259、0.234、0.227，而本题将总得分与估值指标（平均市销率）相联系，可得出结论：中国 A 股市场的估值指标与净资产收益率呈负相关，而与其他 5 个指标存在正相关的关系；进而求得基本面指标、流动性指标对估值指标的影响系数分别为 0.17、0.72，故中国 A 股市场的流动性指标对估值指标的影响比重较大。

根据美国 NASDAQ 市场的综合得分公式可知，营业收入、归母净利润、净资产收益率、年成交量、年平均换手率以及年成交额对总得分的影响系数分别为 0.208、0.264、0.198、0.105、0.071、0.158，而本题将总得分与估值指标（平均市销率）相联系，可得出结论：美国 NASDAQ 市场的估值指标与 6 个低级指标均存在正相关的关系；进而求得基本面指标、流动性指标对估值指标的影响系数分别为 0.67、0.334，故美国 NASDAQ 市场的基本面指标对估值指标的影响比重较大。

表 8 中美两国估值指标受影响系数

	基本面指标影响系数	流动性指标影响系数
中国 A 股市场	0.17	0.72
美国 NASDAQ 市场	0.67	0.334

根据上表比较可知，中国 A 股市场更加重视上市公司的利润率，而美国 NASDAQ 市场更加重视金融的流动量；相比较而言，美国市场较活跃，从而会导致上市公司的倒闭现象大量存在，而中国市场由于成长型科技创新企业较多，处在发展阶段，经济更趋向于稳定，故中美两国的经济状况存在较大差异。

5.3 问题三模型的建立与求解

5.3.1 问题三数据的处理与模型的选取

问题三要求预测出中美两个市场 2019 年的基本面指标和流动性指标，并据此计算两个市场本年度的估值指标。本文对中国 A 股市场和美国 NASDAQ 市场 2009-2018 年的数据进行分析处理，利用 Excel 计算出各指标本年度的平均值，方便各指标的预测和估值指标的计算。因此，本文使用灰色预测和回归分析的方法，分别建立 GM (1,1) 模型和多元线性回归模型。

5.3.2 问题三灰色预测模型的建立

灰色预测模型的建立步骤：（以中国 A 股市场总营业收入为例）

Step1：数据的检验与处理。

设有原始时间数列（从 2009 年开始）：

$$x^{(0)} = x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(10) \\ = (13412218854.7425, \dots, 29982732330.6282)$$

为了确保建模方法的可行性，需要对上述数列作必要的检验处理。计算数列的级比：

$$\lambda(k) = \frac{x^{(0)}(k-1)}{x^{(0)}(k)}, k = 2, 3, \dots, 10.$$

如果所有级比 $\lambda(k)$ 都落在可容覆盖区间 $\theta = (e^{-\frac{2}{n+1}}, e^{\frac{2}{n+2}})$ 内，则序列 $x^{(0)}$ 可作为模型 GM (1,1) 的数据进行灰色预测。否则，就要对序列 $x^{(0)}$ 做变换处理，使其能够落入可容覆盖区间内。即取适当的常数 c 做平移变换：

$$y^{(0)}(k) = x^{(0)}(k) + c, k = 1, 2, \dots, 10.$$

使得序列 $y^{(0)} = \{y^{(0)}(1), y^{(0)}(2), \dots, y^{(0)}(10)\}$ 的级比:

$$\lambda_y(k) = \frac{y^{(0)}(k-1)}{y^{(0)}(k)} \in \theta, k = 2, 3, \dots, 10.$$

则对于 $x^{(0)}$, 级比为:

$$\lambda = \{\lambda(2), \lambda(3), \dots, \lambda(10)\}$$

$$= (0.7154, 0.9039, 0.9194, 0.9063, 0.9567, 0.9713, 1.0440, 1.0637, 0.8044)$$

由于所有的 $\lambda(k) \in [0.7154, 1.0637], k = 2, 3, \dots, 10$. 且此时 θ 中的 n (即 $k_{\max}$) = 10, 所

以可容覆盖区间 $\theta = (e^{-\frac{2}{11}}, e^{\frac{2}{12}}) \approx (0.368, 1.948)$, 故可以用 $x^{(0)}$ 作为 $GM(1,1)$ 模型的数据。

Step2: 对原始数列进行累加运算:

$$x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i), k = 1, 2, 3, \dots, 10.$$

新生成的数列为:

$$x^{(1)} = (13412218854.7425, \dots, 232901341182.83)$$

Step3: 构造数据矩阵 B 及数据向量 Y :

$$B = \begin{bmatrix} -\frac{1}{2}(x^{(1)}(1) + x^{(1)}(2)) & 1 \\ -\frac{1}{2}(x^{(1)}(2) + x^{(1)}(3)) & 1 \\ \vdots & \vdots \\ -\frac{1}{2}(x^{(1)}(9) + x^{(1)}(10)) & 1 \end{bmatrix}, Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(10) \end{bmatrix}, X \begin{bmatrix} a \\ b \end{bmatrix}, Y = BX$$

Step4: 用最小二乘法估算 $\hat{X}$:

$$\hat{X} = (\hat{a}, \hat{b})^T = (B^T B)^{-1} B^T Y = \begin{bmatrix} -0.04363093311 \\ 19155397314.572 \end{bmatrix}$$

于是得到 $a = -0.04363093311, b = 19155397314.572$ 。

Step5: $x^{(1)}$ 近似服从指数规律, 即满足如下灰色预测的微分方程 $GM(1,1)$ 为:

$$\frac{dx^{(1)}}{dt} + \hat{a}x^{(1)} = \hat{b}$$

求解得:

$$\begin{aligned} \hat{x}^{(1)}(k+1) &= (x^{(0)}(1) - \frac{\hat{b}}{\hat{a}})e^{-\hat{a}k} + \frac{\hat{b}}{\hat{a}} \\ &= 454444711382e^{0.04363093311k} - 439032492527 \end{aligned}$$

Step6: 残差检验。

令残差为 $\varepsilon(k)$, 计算:

$$\varepsilon(k) = \frac{x^{(0)}(k) - \hat{x}^{(0)}(k)}{x^{(0)}(k)}, k = 1, 2, \dots, 10.$$

这里 $x^{(0)}(1) = x^{(0)}(1)$ ，如果 $|\varepsilon(k)| < 0.2$ ，则认为达到一般要求；如果 $|\varepsilon(k)| < 0.1$ ，则认为达到较高的要求。经计算，残差为 0.1520，故达到一般要求。

Step7：级比偏差值检验。

首先由参考数据 $x^{(0)}(k-1), x^{(0)}(k)$ 计算出级比 $\lambda(k)$ ，然后用发展系数 a 求出相应的级比偏差：

$$\rho(k) = 1 - \left(\frac{1 - 0.5a}{1 + 0.5a} \right) \lambda(k)$$

如果 $|\rho(k)| < 0.2$ ，则认为达到一般要求；如果 $|\rho(k)| < 0.1$ ，则认为达到较高要求。经计算， $|\rho(k)| = 0.1597 < 0.2$ ，故该序列达到一般要求。

5.3.3 问题三灰色预测模型的求解

根据上述步骤，通过 Matlab 软件，计算得到的预测结果如下表：

表 9 2019 年中、美市场各指标的预测值

	中国 A 股市场	美国 NASDAQ 市场
营业收入（元）	29969897907.85437	134654.0084303869
归母净利润（元）	2580781206.407135	11983.40029135154
净资产收益率（%）	0.550620357009539	-71.67168969307582
年成交量（股）	4898739834.687134	39770321.78732681
年平均换手率（%）	427.8173765560377	89.65135510113032
年成交额（元）	53798561769.29260	373479.9392504115

5.3.4 问题三多元线性回归模型的建立

Step1: 自变量的确定。

根据问题二可知，估值指标与基本面指标、流动性指标均存在一定关系，故选取估值指标作为因变量 y_i ，营业收入、归母净利润、净资产收益率、年成交量、年平均换手率以及年成交额分别作为自变量 $x_1, x_2, x_3, x_4, x_5, x_6$ 。

Step2: 建立多元线性回归方程：

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{m1} & x_{m2} & \cdots & x_{mk} \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{pmatrix}$$

多元线性回归模型假设：解释变量 x_i 为确定性变量，而不是随机变量；解释变量之间不存在相关性；随机误差项 ε 不满足序列相关，与解释变量之间不存在相关性；随机误差项的平均值为 0，满足同方差的正态分布。则利用 SPSS 软件，绘制的相关图形如下：

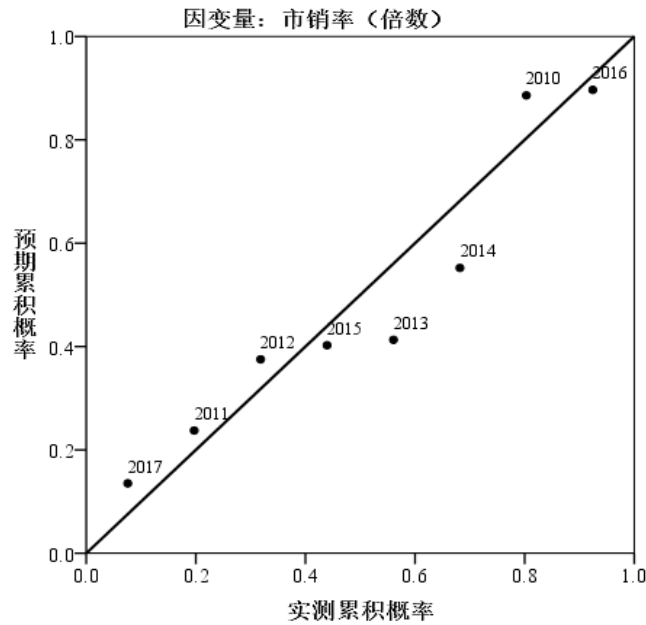


图2 回归标准化残差的正态 $P-P$ 图

由上图可知，标准化残差围绕直线上下波动幅度较小，故所选取的数据符合正态分布，部分假设成立。

Step3: 回归方程的检验。

多元线性回归方程应满足显著性检验、残差检验。显著性检验分为：回归方程的显著性检验（即 F 检验）和回归系数的显著性检验（即 t 检验）。

回归方程的显著性检验是检验被解释变量与所有解释变量之间的线性关系是否显著。对于多元线性回归方程，检验统计量为：

$$F = \frac{SSR / p}{SSE / (n - p - 1)} = \frac{\sum (y - \bar{y})^2 / p}{\sum (y - \hat{y})^2 / (n - 2)} \sim F(p, n - p - 1)$$

若大于则代表显著。

回归系数的显著性检验是检验回归方程中的被解释变量与每一个解释变量之间的线性关系是否显著。对于多元线性回归方程，检验统计量为：

$$t = \frac{\beta_1}{\frac{\sigma}{\sqrt{\sum (x_i - \bar{x})^2}}} \sim t(n - p - 1), \text{ 其中 } \sigma = S_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n - p - 1}}$$

拟合优度指的是回归直线对观测值的拟合程度。度量拟合优度的统计量为可决系数（ R^2 ），其最大值为 1，且越接近 1 说明拟合程度越好。其计算公式为：

$$R^2 = \frac{SSR}{SST}$$

其中， SSR 表示回归离差平方和， SST 表示总离均差平方和。

5.3.5 问题三多元线性回归模型的求解

本文通过 SPSS 软件，选用逐步回归分析的方法，对所选取的数据进行回归分析处理。逐步回归法是对全部因子按照自身对被解释变量影响程度从大到小依次引入回归方程的科学预测方法，其随时对回归方程所包含的全部变量进行检验，观察是否仍然显著，若不显著则将其剔除。然后从剩余的因子中，选出对被解释变量影响最大的因子，并对其进行显著性检验，一直到没有因子能被引入，逐步回归结束。详细流程图如下图所示：

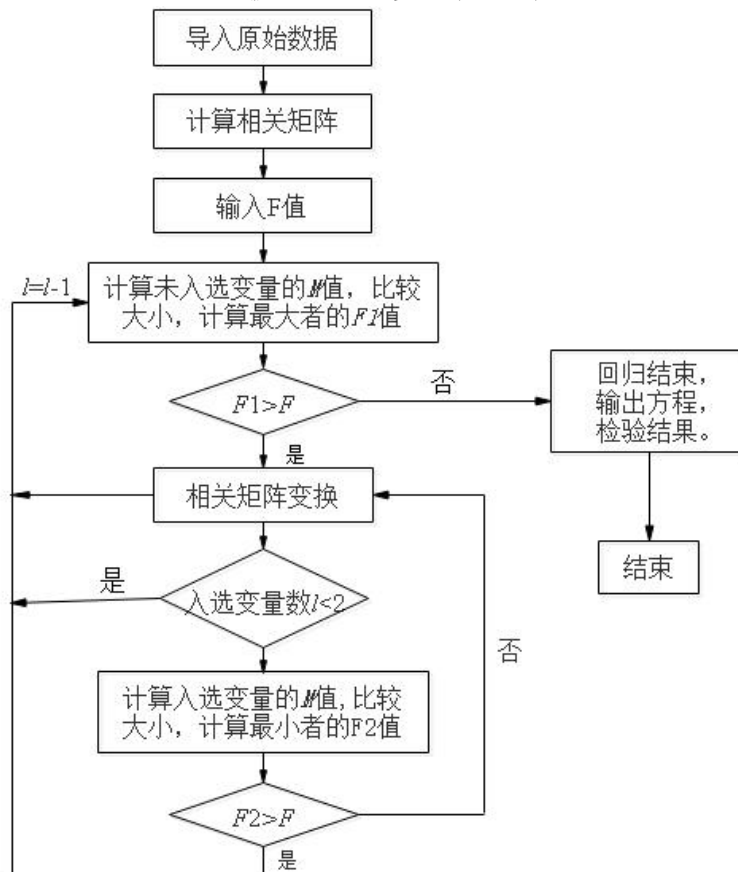


图 3 流程图

根据上图所画的流程图，建立多元线性回归。

将中国 A 股市场 2009-2018 年各项指标的平均值整合代入 SPSS 软件，通过上述操作得到的系数值如下表：

表 9 各项指标平均值的系数

模型	变量	非标准化系数		标准系数		Sig.
		B	标准误差	Beta	t	
1	(常量)	1.712	0.407		4.208	0.006
	年平均换手率 (%)	0.006	0.002	0.826	3.586	0.012
2	(常量)	1.103	0.484		2.279	0.072
	年平均换手率 (%)	0.010	0.003	1.501	3.553	0.016
	年成交额 (元)	-1.217×10^{-11}	0.000	-0.762	-1.805	0.131

由上表可知，中国 A 股市场的对于估值指标的多元回归线性方程为：

$$y = 0.010x_5 - 1.217 \times 10^{-11}x_6 + 1.103$$

该公式的拟合优度 $R^2 = 0.807$ ，由此说明所选的变量对因变量的解释程度优良。统计量 $F = 10.477 > F\alpha(\alpha = 0.05)$ ，说明回归方程总体上存在着线性关系。

同理可算出美国 NASDAQ 市场的对于估值指标的多元线性方程为：

$$y = 1.292 \times 10^{-5} x_1 + 1.358 \times 10^{-5} x_2 + 0.008 x_3 - 4.127 \times 10^{-9} x_4 + 0.005 x_5 - 1.306 \times 10^{-6} x_6 + 1.210$$

根据表 8 中的数据，结合上述公式，得到两个市场 2019 年的估值指标（即平均市销率）分别为 4.726 和 2.335。

5.4 问题四模型的建立与求解

5.4.1 数据的处理

问题四要求根据 93 家企业的基本面数据、美国 NASDAQ 市场的估值量化模型以及中国 A 股市场的流动性指标来预测我国首批科创板企业的估值水平。首先要结合问题一中有关估值的概念建立估值量化模型，然后要求出 93 家企业 2016-2018 年基本面指标的平均值，并根据问题三中预测出的中国 A 股市场的流动性指标，对估值水平进行判断。本文假设估值水平仍是平均市销率。

5.4.2 问题四估值量化模型的建立

根据问题一中对估值定义的公式以及美国 NASDAQ 市场近年来的平均市销率变化情况，我国首批科技创新板企业在上市后的估值水平可以通过如下公式进行测算：

$$VI_{19} = \frac{\overline{Bi}_{19}}{\overline{Bi}_{18}} \times \overline{P/S}_{19}$$

其中， VI_{19} 为 93 家企业 2019 年的预测估值水平， $\overline{Bi}_{18}$ 为 2018 年平均营业收入， $\overline{P/S}_{19}$ 为 2019 年的平均市销率。

5.4.3 问题四估值量化模型的求解

根据问题三中的灰色预测模型对 93 家企业 2019 年的基本面指标进行预测，并与中国 A 股市场 2019 年的流动性指标预测值一同记入下表：

表 10 指标预测结果

基本面指标	营业总收入（元）	归母总利润（元）	净资产收益率（%）
93 家上市公司	1661587705.5196	169833410.72639	21.57088
流动性指标	年成交量（股）	年平均换手率（%）	年成交额（元）
中国 A 股市场	4898739834.687	427.81738	53798561769.293

根据上表数据，并结合问题三中国 A 股市场的估值指标多元回归线性方程得出 2019 年的平均市销率为 4.726。则根据估值量化模型的公式可得 93 家企业的预测估值水平为 5.4。

六、模型的检验

6.1 问题三的检验

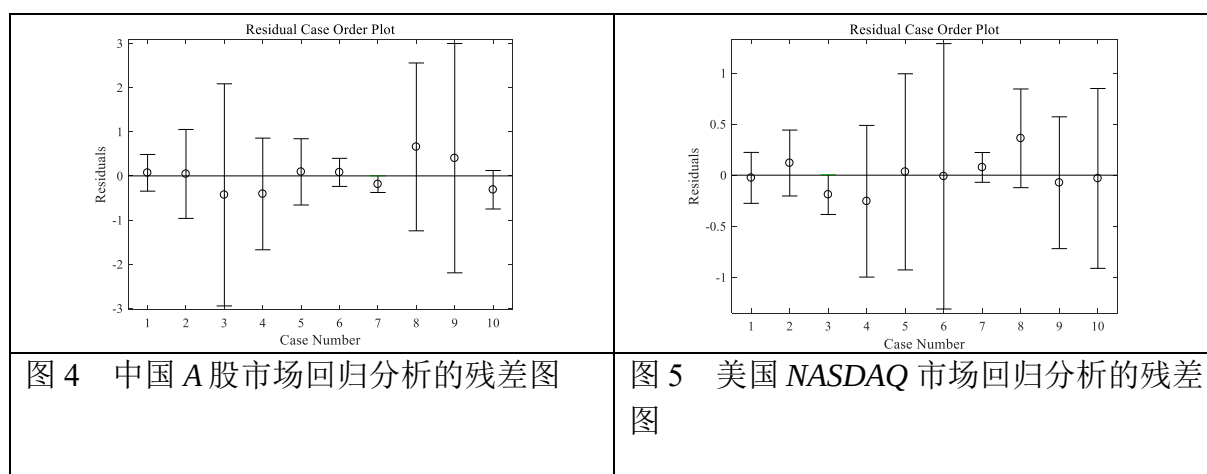
残差指的是由回归方程计算得出的预测值与实际样本值之间的差距，定义为：

$$\varepsilon_i = y_i - \hat{y}_i = y_i - (\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)$$

对于线性回归分析而言，若方程能较好的反映出被解释变量的特征和规律性，则残

差序列中不应含有明显的规律性。残差分析包括以下内容：残差应服从正态分布，其平均值要等于 0；残差不存在自相关；残差的取值与 x 的取值无关；残差的方差相等。

针对问题三的多元线性回归模型，本文对其进行了残差检验，具体如下图所示：



观察命令 $rcoplot(r, rint)$ 所画的残差分布，可知所有残差的置信区间均包含零点，无异常点存在，故该多元线性回归模型的效果较好。

由图可知，美国 NASDAQ 市场的回归效果同中国 A 股市场的对比无过大差别，故该多元线性回归模型的效果也很好。

七、模型的评价及改进

7.1 模型的评价

7.1.1 模型的优点

- 1、回归分析法在分析多因素模型是，更加地方便，简洁；
- 2、回归分析能准确地计算各个因素相关程度的高低，提高预测的精准度；
- 3、灰色预测模型不需要大量的样本，定量分析结果与定性分析结果不会不一致，计算工作量小；
- 4、主成分分析模型能较好地消除评价指标之间的相关影响，能一定程度地减少指标选择的工作量，使结果更加直观、清晰。

7.1.2 模型的缺点

- 1、在回归分析中，选用的因子在某些情况下受限；
- 2、在主成分分析中，要保证所提取的前几个累计方差贡献率达到一个较高的水平，被提取的成分要符合实际背景和意义。

7.2 模型的改进

问题二中的主成分分析法实际可以使用问题三中的回归分析法代替，建立多元线性回归模型，从而问题三可直接使用回归模型的计算结果进行预测。

对于传统的 $GM(1,1)$ 模型而言，其在预测时经常会产生较大的误差，对此本文将对

$GM(1,1)$ 模型进行改进。 $GM(1,1)$ 模型实质是一种外推法预测模型,其是用指数曲线 $X^{(1)}$ 来拟合序列 $X^{(1)}$,拟合的方法是最小二乘法。拟合得到的曲线 $X^{(1)}$ 在坐标平面 $(k, X^{(1)}(k))$ 上必经过点 $(1, X^{(1)}(1))$ 。然而由最小二乘法的原理可知,拟合的曲线并不一定通过第一个数据点,本文将 $X^{(1)}(1) = X^{(1)}(1)$ 作为已知条件的理论依据并不成立。又 $X^{(1)}(1)$ 实际上为最旧的数据,且不能通过累加生成得到。因此本文对 $X^{(1)}(1) = X^{(1)}(1)$ 的条件进行排除,而对于其他数据如 $X^{(1)}(n) = X^{(1)}(n), n = 2, 3, \dots, n$ 仍然可作为已知条件使用,从而得到改进的预测公式为:

$$X^{(1)}(k+1) = \left(X^{(1)}(n) + \frac{u}{a} \right) e^{-ak+a(n-1)} + \frac{u}{a}$$

这里的 n 可以由实际情况选择 $1, 2, \dots, n$ 中的任意整数,通过分析,选取能够使测算误差达到最小的 n 值,从而建立最佳的预测模型。

八、模型的推广

主成分分析法是由 K.皮尔森 (Karl Pearson) 对非随机变量引入的,然后 H.霍特林将此方法推广到随机向量的情形。主成分分析 (PCA) 是一种统计方法,它通常用于解决数据降维、算法加速和数据可视化等问题。主成分分析是设法将原来众多具有一定相关性 (比如 P 个指标) 的指标,重新组合成一组新的互相没有关系的综合指标来代替原来的相关性指标。能够尽可能地减少信息损失,而且各个主成分之间互不相关,更好地简化系统结构,使问题简单化。主成分分析作为一种基础的数学分析法,它的实际应用十分广泛,例如人口统计学、数学建模、数理分析等学科中均有应用,它是一种普遍适用的多变量分析方法。它在数理统计中主要用于指标综合评价,也能应用到系统分析、证券投资、经济评价、财务管理与分析、医院事物管理等众多领域。

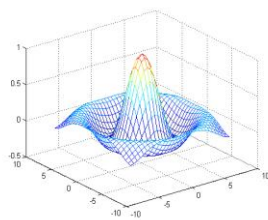
线性回归是利用回归分析,以此来确定两种或两种以上变量之间的定量关系,它在实际应用中十分广泛。线性回归有多元线性回归和一元线性回归,一元线性回归分析能在工程技术经济领域中进行应用,能够合理地预测或确定工程造价,是用料最节省;也可以用在一定养护时间内地聚合物的抗压强度进行测试,能够得到新兴材料的抗压强度。多元线性回归分析可以在岩溶矿坑排水中应用,由于实际生活中排水不及时,排水设施能力不足,导致矿坑设备被淹,多元线性回归能分析降雨量和矿坑涌水量的变化规律,减少损失;多元线性回归在分析城乡居民收入差距因素中十分适用。线性回归在生活中的应用十分有用。

灰色预测模型是通过一些少量的、不完全的信息,建立数学模型并做出预测的一种预测方法。主要运用运筹学的思想方法解决实际问题,制定可执行的发展战略和政策。在进行一系列重要问题的决策时,都必须对未来进行科学有效的预测。常见的灰色预测有五种:数列预测、灾变与异常值预测、季节灾变、波形预测和系统预测等几种类型。在工程技术、经济、农业、生态、环境、等各种系统中会遇到信息不完全的情况,灰色预测能利用时间序列确定微分方程的参数,再建立模型作出预报。实际生活中的道路交通事故,可以建立预测模型,不考虑影响交通事故的各种因素,从道路交通事故数据自身的时间序列中寻求数据间的内在联系,解决实际问题。灰色预测适合“数据少”、波动不大的中短期预测。灰色系统可以解决小样本系统的分析、建模、优化、决策等问题,对我国经济发展具有重要意义。

九、参考文献

- [1]孙晓春 孔伟军, 基于股票估值方法的上市公司投资价值分析, 中文核心期刊要目总览, 1002-5812(2018)16-0042-03, 42-44, 2018;
- [2]叶佩娣, 市盈率估值法和市销率估值法在我国 A 股市场的应用分析, 山东 财政学院学报, 250014, 42-45, 2008;
- [3]冯科 蔡杭坚 窦尔翔等, 相对估值指标在中国股票市场的有效性研究, 北京大学, 102600, 1002-4034 (2015) 02-0057-11, 57-67, 2015;
- [4]陈祥, 市销率在股票投资中的应用, 合作经济与科技, 62-63, 2011;
- [5]李彦芙, 基于多元线性回归模型的江苏省 GDP 增长影响因素研究, 新纪元学校, 山东 潍坊 261108, 84-88, 2019;
- [6]张曦, 一元线性回归分析在工程技术经济领域中的应用, 福州市规划设计研究院, 福建, 350003, 1002-851X (2009) s1-07-0005-04, 2009;
- [7]胡亚楠 王铁铮 巩晓欣等, 基于线性回归的电力物资配套采购数量的预测分析, 国网北京市电力公司物资分公司, 北京 100053, 农村经济与科技第 30 卷第 02 期, 138-139, 2019;
- [8]汪兴瑞 乔克林, 基于层次分析法的股票价值评估模型的研究, 延安大学数学与计算机科学学院, 716000, 陕西, 2018;
- [9]孔雪 王丽 冯益华等, 灰色预测 GM(1,1)模型应用现状与展望, 齐鲁工业大学(山东省科学 机械与汽车工程学院), 济南 250353, 2018;
- [10]陈文彬, 基于中国股市特点的股票估值研究, 华北电力大学管理硕士专业学位论文, 2010;
- [11]桂浩明, 科创板: 金融供给侧改革“试点”, 申万宏源证券研究所首席市场专家, 2019;
- [12]邓聚龙著, 灰色系统基本方法[M]. 华中工学院出版社, 1987;
- [13]刘思峰著, 灰色系统理论及其应用[M], 科学出版社, 1999。

五一数学建模竞赛



题目： 科创板拟上市企业估值

关键词： 股市预测估值；模糊层次分析法；强灵敏 SOM 神经网络；鱼群算法；加权灰色预测

摘要：

在首届中国国际进口博览会上宣布设立的科创板弥补了我国科技型创新企业的融资短板。由于科创板与美国 NASDAQ 市场具有一定的共性，科创板的估值可参照美国 NASDAQ 市场的平均市销率。然而，国内市场与美国市场在企业盈利能力与投资者结构方面的不同，会造成科创板的估值产生严重的估值溢价或折价。为了准确评估科创板上市企业，本文做了以下工作：

针对问题 1，考虑到单只股票对市场的影响，对市场内的每一个企业进行分析，通过对市场内各个企业的市销率、营业收入、年成交额、年成交量和年平均换手率进行计算，获得了 A 股市场与美股市场各自的当前实际估值，同时由数据中给出的市销率计算获得了两市场的预期估值，通过这两种估值的比较，获得了两市场的溢价或折价水平。

针对问题 2，要在基本面指标、流动性指标以及估值指标之间建立定量数学模型，应当考虑到股价的波动情况和换手率情况对流动性指标的影响、市场企业的成长性与价值性对基本面指标的影响。首先，建立各指标的量化模型，然后将股价波动情况和换手率作为影响流动性指标的因素；对各项基本面指标采取了模糊层次分析法，最终获得了综合的基本面评价指标。在获得了各指标的量化模型后，通过绘制三个指标的空间散点图观察三者间的函数关系，分析各自市场的公司分布状况。经过观察后，首先采用强灵敏 SOM 聚类分析方法，对企业类别进行划分，理清市场的内部结构关系，进而在每个类别中，进一步建立市场基本面指标、流动性指标以及估值指标定量关系插值数学模型。

针对问题 3，通过分析 2009-2018 年 10 年间的数据规律，采用灰色预测模型对 2019 年中国 A 股市场与美国 NASDAQ 市场的基本面指标、流动性指标分别进行预测；采用鱼群算法优化的方法对加权灰色预测模型的参数进行优化，进一步体现背景值在建模中的作用，并增加了模型的约束条件，对于不满足约束条件的指标，采用多项式进行拟合预测。根据问题 2 所得的指标定量关系，进一步计算出两个市场的市销率指标。

针对问题 4，根据 2019 年科创板基本面指标、流动性指标的预测值，结合问题 2 中获得的美国 NASDAQ 市场的估值量化模型，对科创板上市后的估值进行计算。

§1 问题重述

为适应中国高科技产业自主化趋势，改变以间接融资为主的融资结构，增强资本市场对实体经济的包容性和服务性以及深化资本市场改革，科创板应运而生。2018年11月5日国家主席习近平在首届中国国际进口博览会上正式宣布设立科创板。科创板的成立弥补了我国科技型创新企业的融资短板。

由于科创板对企业的盈利不做要求，且与美国 NASDAQ 市场具有一定的共同性，采用 NASDAQ 上市企业的平均市销率作为科创板的估值水平。但在企业盈利能力与投资者结构方面国内市场与美国市场仍有不同，基本面与流动性均存在较大差异。为了增强美国 NASDAQ 市场对科创板的可参照性，我们建立了模型解决以下问题：

（1）根据要求将中国 A 股市场作为衡量国内市场与美国 NASDAQ 市场差异的基准，通过平均市销率，分别测算出中国 A 股市场与美国 NASDAQ 市场的估值水平。选取相关指标，分别求解两个市场的估值溢价或折价水平。

（2）要求根据中国 A 股市场与美国 NASDAQ 市场的估值指标、基本面指标以及流动性指标间，研究其中的内在关系，同时对其内在关系寻求定量分析，并针对定量分析结果，给出两个市场的差异比较。

（3）通过建立数学模型，分别预测 2019 年中国 A 股市场与美国 NASDAQ 市场基本面指标、流动性指标，并依据之前求得的定量关系求解两个市场的估值指标，即市销率。

（4）依据前面得到的美国 NASDAQ 市场估值量化模型以及中国 A 股市场流动性指标，以已受理在申报科创板上市的 93 家企业在 2016-2018 年之间的基本面数据为基础，对我国科创板企业上市后的估值水平进行估测。

§2 问题分析

问题（1）的分析：

对于两个市场层面的数据分析，首先要对市场内的每一个企业进行分析，通过对市场内各个企业的市销率、营业收入、年成交额、年成交量和年平均换手率进行计算，得出中国 A 股市场与美国 NASDAQ 市场各自的平均市销率，并进一步求解两个市场的估值溢价或折价水平。

问题（2）的分析：

为了能够获得较为准确的中国 A 股市场与美国 NASDAQ 市场的基本面指标、流动性指标以及估值指标间的关系，首先应当建立基本面指标、流动性指标以及估值指标的定量数学模型。然后通过查阅资料等方式，分别查找影响这三个指标的相关因素，建立各指标的量化模型。将股价波动情况和换手率作为影响流动性指标的因素；对于基本面指标而言，应当同时关注市场企业的成长性与价值性。对于估值指标而言，利用市销率对该指标进行度量。在获得了各指标的量化模型后，应当对这些指标绘制空间散点图，观察三者间的函数关系。由于空间散点图具有集聚性的特点，普通的拟合方法不能适用，因此本文采用了强灵敏 SOM 聚类分析的方法，大体划分了企业类别，理清了市场的内部结构关系。之后再在每个类别中，寻求市场所具有的基本面指标、流动性指标以及估值指标定量关系。

问题（3）的分析：

通过分析 2009-2018 年 10 年间的数据规律，对 2019 年中国 A 股市场与美国 NASDAQ 市场的基本面指标、流动性指标分别进行预测，由于整个市场的整体数据量有限，采用灰色预测模型对基本面指标和流动性指标进行预测，同时为了体现背景值在建模中的作用，采用鱼群算法优化的方法对加权灰色预测模型的参数进行优化，并为其增加了约束条件，对于不满足约束条件的指标，则利用多项式进行拟合预测。再根据问题（2）所得的指标定量关系，进一步计算出两个市场的市销率指标。

问题（4）的分析：

为了分析科创板 2019 年的估值水平，首先应当对 2019 年的科创板基本面指标、流动性指标进行预测，根据预测值，结合问题（2）中获得的美国 NASDAQ 市场的估值量化模型，对科创板上市后的估值进行计算。

§3 名词解释与符号说明

符号	符号含义	符号	符号含义
PS	市销率	P	市值估值
S	营业收入	SL	营业收入增长率
G	归母净利润	GL	归母净利率
J	净资产收益率	E	年成交额
C	年成交量（股）	P_t	实际市值
H	年平均换手率	V	估值溢价或折价水平
A	股价	L	表示流通股数
LD	流动性评价指标	R	神经元数量
a	交易价格波动幅度	Q	训练样本个数
α	权值的学习速率	β	阈值的学习速率
T	最大迭代次数	d	距离
q	已知类别的公司个数	B	累加矩阵

§4 问题假设

1、不考虑政策导向和操盘以及投资者交易情绪等人为因素对中国 A 股市场与美国 NASDAQ 市场的影响。

2、中国 A 股市场与美国 NASDAQ 市场之间不存在相互影响关系，且各市场内的各企业为相互独立的个体，不存在企业间的相互影响

3、市场中企业的估值指标、基本面指标以及流动性指标与市场指标的关系具有一致性。

§5 问题 1 市场的估值溢价或折价水平求解

计算传统股市企业上市后的估值通常采用的是市盈率法，而科创板为 2018 年 11 月 5 日由习近平主席在首届中国国际进口博览会上新设立的板块，独立于现有主板市场。科创板对企业盈利不做要求，采用市销率法更为合理。由于科创板的估值参照美国同类型市场的平均市销率，通过采用平均市销率对国内现有板块和美国市场进行测算找出两个市场之间的估值溢价或折价水平，有利于更合理的参照美国市场对科创板进行估值。

选取平均市销率作为估值水平，市销率(Price-to-sales, PS)的定义为：

$$PS = \frac{P}{S} \quad (5.1)$$

其中， P 表示市值估值， S 表示营业收入。市销率 PS 的值越低，说明该企业股票目前的投资价值越大。

两个市场的平均市销率 $\overline{PS}$ 可以通过市值总估值除以总营业收入进行求解。总营业收入可以根据附件 1 和 2 中 2018 年各企业的营业收入进行求和计算，估计市值可以根据公式 (5.1)，利用每一个企业 2018 年的市销率与主营业务收入关系先求出每一个企业的估计市值，再进行求和运算就得到了市值总估值 $P_{\text{总}}$ ，具体的求解数值见附录 2，将数据整理如下表 5.1 所示：

表 5.1 中国 A 股市场和美国 NASDAQ 市场的平均市销率

中国 A 股市场		
总营业收入 $S_{\text{总}}$ (元)	市值总估值 $P_{\text{总}}$ (元)	平均市销率 $\overline{PS}$ (倍数)
32,545,270,536,213.70	4.76907×10^{13}	1.46536576
美国 NASDAQ 市场		
总营业收入 $S_{\text{总}}$ (万美元)	市值总估值 $P_{\text{总}}$ (万美元)	平均市销率 $\overline{PS}$ (倍数)
434,155,336.94	1,696,451,287	3.90747537

从表中可以看出，中国 A 股市场的平均市销率明显低于美国 NASDAQ 市场，说明从目前的市场状况来说，中国市场总体上的投资价值要更大。

要求得两个市场的估值溢价或折价水平，则需要知道每一个企业的单只股票实际市值 P_r 与市值估值 P ，实际市值 P_r 可以通过股市的数据进行计算，具体步骤如下，其中 $\bar{A}$ 表示股票均价， L 表示流通股数。

$$H = \frac{C}{L} \quad (5.2)$$

$$\bar{A} = \frac{E}{C} \quad (5.3)$$

$$P_r = L\bar{A} = E/H \quad (5.4)$$

市值估值在上文已经求得，估值溢价或折价水平的求解公式见式 (5.5)：

$$V = \frac{\sum P_r - P}{\sum P} \quad (5.5)$$

由此分别计算出两个市场的股票的市值和估计市值，求得两个市场的总市值和总市值估值，进而求得总估值溢价或折价水平 V ，其中正值表示溢价，负值表示折价，主要数据见下表 5.2:

表 5.2 中国 A 股市场和美国 NASDAQ 市场的估值溢价或折价水平

中国 A 股市场		
实际总市值 $P_{r总}$ (元)	市值总估值 $P_{总}$ (元)	估值溢价或折价水平 V (%)
8,242,113,760	4.76907×10^{13}	-99.9827176
美国 NASDAQ 市场		
实际总市值 $P_{r总}$ (万美元)	市值总估值 $P_{总}$ (万美元)	估值溢价或折价水平 V (%)
3,218,636,065	1,696,451,287	89.7275854

可以看出，中国 A 股市场的总市值估值要明显高于实际总市值，股市的折价率较高，在估值方面的优势较大，具备较高的长期投资价值；而美国 NASDAQ 市场的总市值估值却明显低于实际总市值，股市的溢价率较高，从估值角度来看，美国 NASDAQ 市场的公司普遍较贵，长期投资价值不明显。这与市销率的所反映的市场情形相一致，说明市销率作为一种估值指标，具备一定的合理性。

通过上述分析可知，当参照美国 NASDAQ 市场的平均市销率对科创板进行估值时，会降低科创板的实际估值水平，并在一定程度上加大了科创板的折价率或有估值溢价出现的可能，因此在参照美国市场对科创板进行估值时，应当适当对市值的估计进行修正，从而减少估值溢价的出现。

§6 问题 2 估值指标与基本面指标、流动性指标之间的关系

6.1、流动性指标的建立

为了估值指标与基本面指标、流动性指标之间的关系，首先对于单只股票三个指标之间的关系进行研究。流动性是指投资者以最小成本、最小价格影响和最快速度完成交易大宗股票的容易程度。通过上述定义，我们可以知道流动性指标必须同时体现股票成本、股票价格波动情况以及股票价格波动时间。以下针对单只股票建立单只股票的流动性指标。本文采用的股票流动性评价指标 $G(\Delta t)$ 定义如下：

$$G(\Delta t) = \frac{H(\Delta t)}{a(\Delta t)\Delta t} \quad (6.1)$$

其中， $H(\Delta t)$ 为单只股票在给定时间的总换手率， Δt 为股票所研究的时间区间， $a(\Delta t)$ 为给定时间区间的交易价格波动幅度，可通过下式进行定义。

$$a(\Delta t) = \frac{A_{\max}(\Delta t) - A_{\min}(\Delta t)}{A_{\min}(\Delta t)} \quad (6.2)$$

为了保证结果有意义，对于最高股价与最低股价相同的情况，本文做出如下补充定义。

$$a(\Delta t) = \frac{h}{A_{\min}(\Delta t)} \quad (6.3)$$

其中 h 为当前所计算的市场的的市场的最小价格变动单位。在 A 股市场，该值取 0.01。综上所述， $a(\Delta t)$ 可用下列函数进行表示。

$$a(\Delta t) = \begin{cases} [A_{\max}(\Delta t) - A_{\min}(\Delta t)] / A_{\min}(\Delta t) A_{\max} \neq A_{\min} \\ h / A_{\min}(\Delta t) & A_{\max} = A_{\min} \end{cases} \quad (6.4)$$

同时，由经济学公式易知给定时间内的交易均价可用下式进行表示。

$$A(\Delta t) = \frac{E(\Delta t)}{C(\Delta t)} \quad (6.5)$$

其中， $E(\Delta t)$ 为给定时间内的年成交总额， $C(\Delta t)$ 为给定时间内的成交量。由上可以获得用于最终计算的流动性评价指标定义式如下

$$LD(\Delta t) = \begin{cases} \frac{H(\Delta t) \left(\frac{E(\Delta t)}{C(\Delta t)} \right)_{\min}}{\left[\left(\frac{E(\Delta t)}{C(\Delta t)} \right)_{\max} - \left(\frac{E(\Delta t)}{C(\Delta t)} \right)_{\min} \right] \Delta t} & \left(\frac{E(\Delta t)}{C(\Delta t)} \right)_{\max} \neq \left(\frac{E(\Delta t)}{C(\Delta t)} \right)_{\min} \\ \frac{H(\Delta t) \left(\frac{E(\Delta t)}{C(\Delta t)} \right)_{\min}}{h \Delta t} & \left(\frac{E(\Delta t)}{C(\Delta t)} \right)_{\max} = \left(\frac{E(\Delta t)}{C(\Delta t)} \right)_{\min} \end{cases} \quad (6.6)$$

通过上式从 2009 年至 2018 年每隔十年作为统计周期，可以计算全体 A 股及 NASDAQ 市场股票的流动性指标。其中部分流动性指标计算结果如下表 6.1 所示。其中，由于部分股票存在退市等原因，导致流动性指标异常，影响 A 股市场流动性的整体分析，因此处理数据时已将这部分数据删去。

表 6.1 中美市场流动性指标

中国 A 股市场		美国 NASDAQ 市场	
公司代码	流动性指标	公司代码	流动性指标
600616.SH	0.1516	ACLS.O	0.2486
600617.SH	0.0695	ABCB.O	0.7991
600618.SH	0.1248	ACTG.O	0.8411
600619.SH	0.1822	ADBE.O	1.0649
600620.SH	0.0618	ADI.O	3.4820
600621.SH	0.1470	ADMP.O	0.4132
600622.SH	0.1206	ADP.O	1.9620
600623.SH	0.0817	ADSK.O	2.1862
600624.SH	0.2228	ADTN.O	6.4017
600381.SH	0.0650	ADUS.O	0.5241

6.2、基本面指标的建立

上市公司的基本面包括财务状况、盈利状况、市场占有率、经营管理体制、人才构成等各个方面。本文从附件 1 和 2 中所提供的公司财务数据出发，主要考虑上市公司的价值性与成长性。考虑到净资产收益率与营业收入与归母净利润的相关性，本文选取营业收入指标与归母净利润指标作为价值性评价指标；选取营业收入同比增长率指标、归母净利润同比增长率作为成长性评价指标。

6.2.1 基本面指标影响因素的定义

由于上市公司普遍存在某公司的考察参数严重偏离市场平均值的情况，为了保证指标归一化后的结果质量，需要对考察参数进行筛选，因此本文定义了用于存放不同类别的考察数据集合如下：

$$A_t = \{X_t \mid X_t \leq Q_{t1} - 1.5IQR_t, t = 1, 2, 3, 4\} \quad (6.7)$$

$$B_t = \{X_t | Q_{t3} + 1.5IQR_t \geq X_t \geq Q_{t1} - 1.5IQR_t, t = 1, 2, 3, 4\} \quad (6.8)$$

$$C_t = \{X_t | X_t \geq Q_{t3} + 1.5IQR_t, t = 1, 2, 3, 4\} \quad (6.9)$$

其中, Q_{t1} 、 Q_{t3} 、 IQR_t 分别为第 t 个指标的下四分位数, 上四分位数以及四分位距。 X_t 表示各项指标所对应的参数值。其中 X_1 表示营业收入; X_2 表示归母净利润; X_3 表示营业收入同比增长率、 X_4 表示归母净利润同比增长率。

为了方便统计与运算, 可将营业收入指标 x_1 、归母净利润指标 x_2 、营业收入同比增长率指标 x_3 、归母净利润同比增长率指标 x_4 利用下式统一进行计算, 其表达式 6.10 如下:

$$(x_t)_i = \begin{cases} 0 & X_t \in A \\ \frac{\sum_{j=1}^{p_t} (X_t)_{ij} - [(X_t)_{ij}]_{\min}}{[(X_t)_{ij}]_{\max} - [(X_t)_{ij}]_{\min}} & X_t \in B_t, \text{其中 } t=1, 2, 3, 4 \\ p_t & \\ 1 & X_t \in C_t \end{cases} \quad (6.10)$$

其中, $(x_t)_i$ 为第 i 个公司的第 t 项指标, $(X_t)_{ij}$ 为第 i 个公司第 j 年的第 t 项指标所对应的参数值。 p_t 为第 t 个指标实际统计的年数。

6.2.2 基于模糊层次分析法的基本面指标模型建立

对于评价分类的问题, 可以采取神经网络方法^[1]、模糊方法^[2]、层次分析法^[3]等方法。神经网络 (ANN) 在样本数量较多且符合一定规律条件下, 效果较好, 但是考虑到本问题中的训练样本差异较大, 所以不宜采用。模糊方法 (Fuzzy Logic) 是通过建立模糊集以及隶属度函数, 进而构建模糊数学模型进行评价。层次分析法 (AHP: the Analytic Hierarchy Process) 是 20 世纪 70 年代美国运筹学 T.L. Saaty 教授提出的一种定性定量相结合的系统分析方法。该方法对于量化评价指标, 选择最优方案提供了依据。

层次分析法 (AHP) 能将人的主观判断过程数学化、思维化, 使决策依据易于被人接受, 因此, 更能适合复杂的社会科学领域的情况。但是, 其思维一致性很难保证。检验判断矩阵是否一致非常困难。在这种情况下, 将模糊法与层次分析法的优势结合起来形成的模糊层次分析法 (FAHP), 将能很好地解决这一问题。模糊层次分析法的基本思想和步骤与 AHP 的步骤基本一致, 但仍有以下两方面的不同点:

(1) 建立的判断矩阵不同: 在 AHP 中是通过元素的两两比较建立判断一致矩阵; 而在 FAHP 中通过元素两两比较建立模糊一致判断矩阵

(2) 求矩阵中各元素的相对重要性的权重的方法不同

模糊层次分析法 (FAHP) 改进了传统层次分析法存在的问题, 提高了决策可靠性。

6.2.2.1 模糊层次分析法解决问题的步骤

模糊层次分析法的基本思想是根据多目标评价问题的性质和总目标, 把问题本身按层次进行分解, 构成一个由下而上的梯阶层次结构。因此在运用 FAHP 决策时, 大体上可以分为以下四个步骤。

(1) 分析问题, 确定系统中各因素之间的因果关系, 对决策问题的各种要素建立多级 (多层次) 递阶结构模型。

(2) 对同一层次 (等级) 的要素以上一级的要素为准则进行两两比较, 并根据评定尺度确定其相对重要程度, 最后据此建立模糊判断矩阵。

(3) 通过一定计算, 确定各要素的相对重要度。

(4) 通过综合重要度的计算, 对所有的替代方案进行优先排序, 从而为决策人

选择最优方案提供科学的决策依据

下面进行建立模糊互补判断矩阵、模糊互补判断矩阵权重的计算方法以及模糊互补判断矩阵的一致性判断。

6.2.2.2 模糊互补判断矩阵的建立

根据附件 1 和 2 所整理出的关于基本面指标的信息，对评价体系进行分层。构架如下：

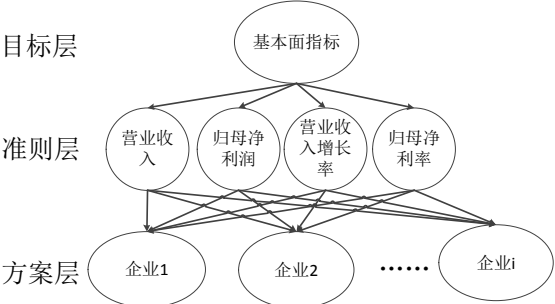


图 6.1 评价指标体系构架图

在模糊层次分析中，各因素间的两两比较判断时，采用一个因素比另一个因素的重要程度定量表示，则得到的模糊判断矩阵 $A=(a_{ij})n \times n$ ，如果其具有如下性质：

- (1) $a_{ij} = 0.5, i = 1, 2, \dots, n$
- (2) $a_{ij} + a_{ji} = 1, i = 1, 2, \dots, n$

则这样的判断矩阵称为模糊互补判断矩阵。为了使任意两个方案关于某准则的相对重要程度得到定量描述，通常采用如表 1 的 0.1~0.9 标度法给予数量标度。

表 6.2 尺度的含义

标度	定义	说明
0.5	两元素相比较	同等重要两元素相比较
0.6	两元素相比较	一元素比另一元素稍微重要
0.7	两元素相比较	一元素比另一元素明显重要
0.8	两元素相比较	一元素比另一元素重要得多
0.9	两元素相比较	一元素比另一元素极端重要
0.1, 0.2, 0.3, 0.4	反比较	若元素 a_i 与元素 a_j 相比较得到判断 r_{ij} ，则元素 a_j 与元素 a_i 相比较得到的判断为 $r_{ji}=1-r_{ij}$

$a_{ii} = 0.5$ 表示因素与自己相比同样重要；若 $a_{ij} \in [0.1, 0.5)$ ，则表示因素 x_j 比 x_i 重要；若 $a_{ij} \in [0.5, 0.9)$ ，则表示因素 x_i 比 x_j 重要。

依据上面的数字标度，列出各个因素，则得到如下模糊互补判断矩阵表：

表 6.3 模糊判断矩阵表

序号	营业收入	归母净利润	营业收入增长率	归母净利润率
营业收入	0.5	0.9	0.9	0.9
归母净利润	0.1	0.5	0.3	0.4
营业收入增长率	0.1	0.7	0.5	0.7
归母净利润率	0.1	0.6	0.3	0.5

应用排序向量公式

$$P_i = \frac{1}{n} - \frac{1}{2a} + \frac{1}{na} \sum_{k=1}^n r_{ik} \tag{6.11}$$

其中 $a=(n-1)/2$ ，进行层次单排序，得：

$$P=(0.542 \quad 0.208 \quad 0.325 \quad 0.242)^T$$

归一化处理后得到所求特征向量：

$$P=(0.413 \quad 0.158 \quad 0.247 \quad 0.184)^T$$

对模糊判断矩阵进行一致性检验，因为 R 矩阵中任意两行（列）对应元素之差为常数，所以 R 是完全一致矩阵。

表 6.4 权重汇总表

营业收入	归母净利润	营业收入增长率	归母净利润率
0.413	0.158	0.247	0.184

由式（6.11）得到的权重值是否合理，还应该进行比较判断的一致性检验。当偏移一致性过大时,表明此时将权向量的计算结果作为决策依据是不可靠的。推导出用模糊判断矩阵的相容性来检验其一致性原则的方法。主要包括检验 m 个判断矩阵 A_k 的满意一致性以及检验判断矩阵间的满意相容性。

$$I=(A_k,P^{(k)})\leq A,k=1,2,\dots,m \tag{6.12}$$

$$I(A_k,A_l)\leq A,k\neq l,l=1,2,\dots,m \tag{6.13}$$

可以证明在模糊互补判断矩阵 $A_k(k=1,2,\dots,m)$ 是一致可接受的情况下，它们的综合判断矩阵也是一致可接受的。权重向量表达式：

$$P=(P_1,P_2,\dots,P_n) \tag{6.14}$$

式中， $P_i=\frac{1}{n}\sum_{k=1}^n P_i^{(k)} (i=1,2,\dots,m)$

只要当上述两个条件同时满足时， m 个权重值的均值作为因素集 x 的权重分配向量就是合理的。

代入 $n=4$ ，得 $I=0.046<0.1$ ，故矩阵具有满意的一致性，可以作为评价因子的权向量。

6.3、综合评价模型的建立

根据上述方法，可以获得全体上市公司的基本面指标、流动性指标以及市销率估值指标，其部分数据如下表所示。

表 6.5 中美市场部分公司基本面、流动性以及市销率指标情况

中国 A 股市场				美国 NASDAQ 市场			
代码	市销率指标	流动性指标	基本面指标	代码	市销率指标	流动性指标	基本面指标
600608.SH	0.2765	0.0872	0.1752	ACIW.O	2.0549	2.0055	0.0967
600613.SH	4.4746	0.0951	0.1765	ACLS.O	1.0885	0.2487	0.0942
600655.SH	71.4120	0.0721	0.1778	AIRT.O	0.2595	0.6292	0.0935
600621.SH	0.6623	0.1257	0.1761	ADBE.O	7.6887	1.0650	0.1103
601901.SH	1.5397	0.3155	0.1760	ADI.O	5.2005	3.4821	0.1055
600642.SH	4.1312	0.0226	0.1789	ADP.O	3.3040	1.9621	0.1294
900911.SH	0.6759	0.0839	0.1752	ADSK.O	6.2726	2.1862	0.1000
600668.SH	40.4583	0.0708	0.1742	ARNA.O	30.5847	0.9159	0.0942
600687.SH	2.2808	0.036	0.1807	AROW.O	3.0313	2.9304	0.0938
600696.SH	1.0339	0.1376	0.1790	ARRY.O	8.5511	2.4822	0.0938

为了清楚表明各个公司的市销率与流动性、基本面间的关系，本文分别做出了流动性、基本面与市销率的散点图，其散点图 6.2 如下所示。

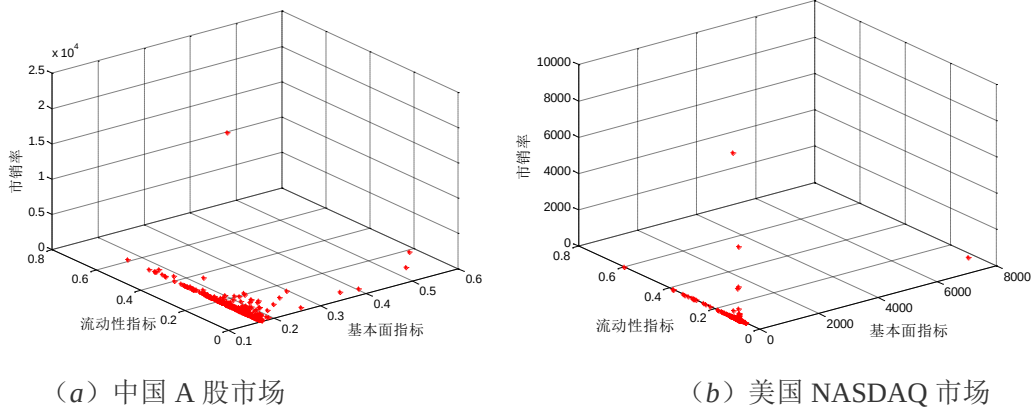


图 6.2 流动性与市销率、基本面与市销率的散点图

从散点图中，我们可以看出这些数据具有集群的特点，在某些区间存在聚集效应，而在其余区间的数据则较为稀疏。因此，利用普通的回归分析基本不可能得到较为符合的拟合函数，本文为了更为清晰的掌握数据间的集群效应，将具有类似流动性、价值型以及估值的数据进行了聚类处理，从而获得较为直观的流动性指标、价值性指标以及市销率估值间的内在关系。

对于聚类算法的实施，最关键的一步是要计算出两个数据点的“相似性”，常用的聚类算法有：层次聚类、K-means 聚类、谱聚类、Chameleon 聚类、PCCA、SOM(self-organizing map)^[4,5]、Affinity Propagation 等。

SOM 聚类是 Kohonen 提出的竞争网络降维算法，它可以将高维的数据投影到节点平面上，实现高维数据可视化，然后也可以继续根据降维之后的数据再进行聚类。它是基于无导师的竞争神经网络结构的一种算法，其与真实人脑中的神经网络类似，可以通过不断地观察分析与比较，自动揭示样本中的内在规律和本质，从而可以对具有近似特征（属性）的样本进行准确地分类和识别。这与股市中诸多企业的分类结果可学习知识较少，并且总共要分多少类别都没有相关的知识储备的情况比较相似，故采用 SOM 神经网络聚类方法进行归类处理。

6.3.1 SOM 神经网络设计及分析

1、竞争神经网络结构

(1) 竞争神经网络学习算法

竞争神经网络在学习的过程中，仅用到了样本的输入信息。输入层由 R 个神经元构成，竞争层由 S^1 个神经元构成。设训练样本的输入矩阵为：

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1Q} \\ p_{21} & p_{22} & \cdots & p_{2Q} \\ \cdots & \cdots & \cdots & \cdots \\ p_{R1} & p_{R2} & \cdots & p_{RQ} \end{bmatrix}_{R \times Q} \quad (6.15)$$

其中， Q 为训练样本的个数； p_{ij} 表示第 j 个训练样本的第 i 个输入变量，并记 $p_i = [p_{i1}, p_{i2}, \cdots, p_{iQ}]$ ， $i=1, 2, \dots, R$ ，则网络的初始连接权值为：

$$IW^{1,1} = [\omega_1, \omega_2, \cdots, \omega_R]_{S^1 \times R} \quad (6.16)$$

网络的初始阈值为：

$$b^1 = \left[e^{1-\log(\frac{1}{S^1})}, e^{1-\log(\frac{1}{S^1})}, \dots, e^{1-\log(\frac{1}{S^1})} \right]_{S^1 \times 1} \quad (6.17)$$

同时，在学习之前需要初始化相关参数。设权值的学习速率为 α ，阈值的学习速率为 β ，最大迭代次数为 T ，迭代次数初始值为 $N=1$ ；

(2) 计算获胜神经元

随机选取一个训练样本 p 根据

$$n_i^1 = \sqrt{\sum_{j=1}^R (p_j - IW_{ij}^{1,1})^2} + b_i^1 \quad i=1,2,\dots,S^1 \quad (6.18)$$

计算竞争层神经元的输入。其中， n_i^1 表示竞争层第 i 个神经元的输出； p^i 表示样本 p 第 j 个输入变量的值； $IW_{ij}^{1,1}$ 表示竞争层第 i 个神经元与输入层第 j 个神经元的连接权值； b_i^1 表示竞争层第 t 个神经元的阈值。

设竞争层第 h 个神经元为获胜神经元，则应满足下式的要求。

$$n_k^1 = \max(n_i^1), i=1,2,\dots,S^1 \quad k \in [1, S^1] \quad (6.19)$$

(3) 权值、阈值更新

获胜神经元 h 对应的权值和阈值分别按照：

$$IW_k^{1,1} = IW_k^{1,1} + \alpha(p - IW_k^{1,1}) \quad (6.20)$$

$$b^1 = e^{1-\log((1-\beta)e^{1-\log(b^1)} + \beta \times \alpha^1)} \quad (6.21)$$

进行修正，其余神经元的权值和阈值保持不变。其 $IW_k^{1,1}$ 为 IW 的第 k 行，即表示与获胜神经元 k 对应的权值； a^1 为竞争层神经元的输出，即

$$a^1 = [a_1^1, a_2^1, \dots, a_{S^1}^1] \quad a_i^1 = \begin{cases} 1 & i = k \\ 0 & i \neq k \end{cases} \quad (6.22)$$

(4) 迭代结束判断

若样本没有学习完，则再另外随机抽取一个样本，返回步骤(2)。否则，若 $N < T$ ，令 $N = N+1$ ，返回步骤(2)；否则，迭代结束。

(5) 竞争神经网络的 MATLAB 工具箱函数

函数 newc 用于创建一个竞争神经网络，其调用格式为：

`net = Newcastle (PR,S,KIR,CLR)`

其中，PR 为输入变量的最小及最大值构成的矩阵均为竞争层神经元个数；K LR 为权值 的学习速率（默认为 0.01），CLR 为阈值的学习速率（默认为 0.001）；net 为创建好的竞争网络。

6.3.2 SOM 神经网络强灵敏分析建模

SOM神经网络能够识别向量空间中最有意义的特征等，因而特别适用于高维数据的无指导聚类分析，但它仍然存在一些缺陷，具体可以总结如下：

(1)收敛速度慢，为了获得网络全局极小点，学习速度需取较小的值，只能缓慢地逼近全局极小点。

(2)输入顺序敏感性。SOM算法在训练的过程中，是通过具有最小失真的训练矢量进行调整，这样使算法对输入数据的训练序列的依赖性较大，也就是输入数据敏感性问题。

而造成 SOM 缺陷的原因有以下几个方面：

(1)学习速率的影响。在 SOM神经网络的自组织的过程中，如果学习速率一直

过大甚至接近于1，则会使权重向量的更新过程来回振荡，学习稳定性较差；反之，如果学习速率过小甚至趋向于0，虽然可以增强学习稳定性，但是会使收敛时间过长。因此，SOM算法不仅要调整神经元的权值，还要对神经元邻域内的所有神经元进行权值修正，导致它收敛速度慢，这是SOM算法应用的主要瓶颈。

(2)迭代次数的影响。在SOM神经网络的训练过程中，对于那些复杂的网络和样本，若给定较小的迭代次数，那么会导致迭代往往还没有收敛时学习就结束了。若给定太大的迭代次数，往往又会造成训练速度缓慢，导致极大的浪费。因此，确定一个准则来判断迭代是否已经收敛十分必要。

(3)输入顺序的影响。对于SOM模型，每次进行连接权修正后，各输出神经元对应于输入模式空间的最邻近区都在不断地变化，不同的输入模式对应着不同的变化，因此使网络最终的收敛结果与输入顺序有关。

为了加快网络的收敛速度，本文对自适应SOM算法进行改进，提出了强灵敏的SOM算法。SOM神经网络在训练过程中权值修正公式一般如式(6.23)所示。

$$\Delta w_{ij}(t+1) = \alpha(t) \times h_{ci}(t) \times \Delta w_{ij}(t) \quad (6.23)$$

$$h_{ci}(t) = \tau + \alpha(t-1) \times (1 - 0.01 \times t) \quad (6.24)$$

式中， $\Delta w_{ij}(t+1)$ 和 $\Delta w_{ij}(t)$ 表示权值增量， $\alpha(t)$ 表示学习速率， $h_{ci}(t)$ 表示邻域函数如式(6.24)所示。

在SOM算法的训练过程中，网络中每个输出节点都可以成为获胜节点，但是每个节点成为获胜节点的概率都不一样。有的节点响应次数多，其对应的权值与邻域内权值调整的机会就大；而有的节点响应次数少，其对应的权值与邻域内权值调整的机会就小，这就影响了一部分权值的利用效率。

因此，为了避免一部分权值的重复调整，而另一部分权值不被调整或者很少被调整，引入敏感参数 b_i 。 b_i 的定义如下。

给每个输出节点都设置一个敏感参数 b_i ，对应的节点每响应一次， b_i 则加1，也就是说 b_i 代表输出节点的响应次数。

从前面的分析可以看出，导致SOM算法收敛速度慢的主要原因是不仅要调整神经元的权值，还要对神经元邻域内的所有神经元进行权值修正。因此，此时权值修正公式改进为式(6.25)。

$$W_{ij}(t+1) = W_{ij}(t) + f(b_i)[x_j(t) - W_{ij}(t)] \quad (6.25)$$

式中， $f(b_i)$ 的计算公式如(6.26)所示， a 为正常数

$$f(b_i) = e^{\frac{b_i}{a}} \quad (6.26)$$

6.3.3 聚类结果分析

经过数据的预处理，数据集向量 $u_{ij} = [G_j, T_j, P_j]$ ，其中 G_j, T_j, P_j 分别为流动性指标列向量、基本面指标列向量以及估值列向量。

以及SOM神经网络聚类之后，得到了5类主要的股市评估类别。其聚类中心的结果参数如表6.6、表6.7所示。

表 6.6 中国 A 股市场聚类中心结果

指标	聚类类别				
	1	2	3	4	5
估值	43.522818	102.650770	225.480171	4.162478	172.536030
基本面	0.177159	0.177031	0.181723	0.180993	0.187743
流动性	0.123989	0.102691	0.112583	0.128383	0.162714

表 6.7 美国 NASDAQ 市场聚类中心结果

指标	聚类类别				
	1	2	3	4	5
估值	2.7588	268.4671	346.6942	82.5134	170.7393
基本面	0.1011	0.0926	0.1008	0.0992	0.0949
流动性	2.2660	2.1581	2.1508	3.0119	3.6488

由聚类结果可知，美国市场较中国市场具有较好的流动性，主要原因在于中国市场属于不成熟市场，换手率偏高，且股价波动较大。同时，中国市场的基本面指标总体偏高，说明 A 股市场中的大多数公司的基本面水平相对偏高；从估值指标来看，美股的估值较高，而中国市场相对较低。

6.4、流动性指标、基本面指标、估值指标的定量关系确定

上文已经对数据进行了大致分类，为了进一步明确流动性指标、基本面指标、估值指标的定量关系，以下给出给定某公司流动性指标 G_0 与基本面指标 T_0 后，对该公司 Co 的估值进行估算的方法。设 5 类对象分别对应集合 O_1 、 O_2 、 O_3 、 O_4 、 O_5 。

(1) 取聚类中心参数的流动性指标 G 、基本面指标 T ，按照下式计算该公司与各类中心的距离 d_i 。^[6]

$$d_i = \sqrt{(G - G_i^*)^2 + (T - T_i^*)^2} \quad (6.27)$$

(2) 依据下式确定该公司所属类。

$$Co \in O_i \text{ 其中 } \{i | d_i = \min\{d_1, d_2, d_3, d_4, d_5\}\}$$

(3) 于是公司的估值可用下列函数进行确定

$$P = \begin{cases} P_i^* & q < 20 \\ P_k & q \geq 20 \end{cases} \quad (6.28)$$

其中， q 为已知类别中的公司个数。 P_i^* 为某类中心值参数的估值指标， P_k 中 k 满足 $\{k | d_k = \min\{d_{ik}\}\}$ 。

§7 问题 3 基于鱼群算法优化的加权灰色模型预测估值指标

7.1、市场各指标历年数据

根据附件 1、附件 2 数据整理出的附录 4 数据中，可以得出市场历年的基本面指标、流动性指标数据，其中基本面指标包括营业收入指标、归母净利润指标、净资产收益率指标、营业收入同比增长率指标、归母净利润同比增长率以及净资产收益率同比增长率指标。

本部分以中国 A 股市场和美国 NASDAQ 市场为研究对象，根据 2009-2018 的数据作为依据，利用鱼群算法优化的加权 GM(1, 1)模型预测确定 2019 年的中国 A 股市场和美国 NASDAQ 市场指标，并且得出历年市场各项指标的模拟回归值；

7.2、基于现行市场基本面指标与流动性指标的预测

7.2.1 灰色预测模型及其改进

本文采用鱼群算法优化的加权灰色预测模型 GM(1,1)，利用附录 4 的数据，对

2019 年市场的进行预测。

当观测数据序列较长时，各种数学建模方法均可获得满意的预报结果；但对于短数据序列，由于信息量少，规律性不强，使得传统统计预测法预测不准确。在这方面，灰色预测理论显示了一定的优越性，灰色预测通过鉴别系统因素之间发展趋势的相异程度，即进行关联分析，并对原始数据进行生成处理来寻找系统变动的规律，生成有较强规律性的数据序列，然后建立相应的微分方程模型，从而预测事物未来发展趋势的状况。

但是，目前存在的问题是，以往的 GM(1,1) 预测模型都是以固定的背景值（1-AGO 的紧邻均值）为基础进行建模的，这样并不足以反映背景值在建模中的作用，而事实上应考虑所建模型进行预测计算的平均模拟相对误差达到最小来确定预测模型，这样所建立的模型其预测效果会更优。

即，灰色微分方程中的背景值 $x(t+1)$ 以及 $x(t)$ 不再按照传统的方法去平均，而是从变形体实际的变形规律出发，引入一种新的背景值计算方法，使得原始值与模型预测之差的平均模拟相对误差达到最小。

$$\bar{\Delta} = \frac{1}{n} \sum_{k=1}^n \left| \frac{\varepsilon(k)}{x_{(k)}^{(0)}} \right| = \min \quad (7.1)$$

这样背景值的确定不再是取其平均值，而是依据 $x(t+1)$ 、 $x(t)$ 的影响大小用加权的方式来确定，即

$$z(t+1) = px(t+1) = (1-p)x(t) \quad (7.2)$$

式中， p 为背景值最佳生成系数，其取值范围为 $[0, 1]$ 。

为了选取恰当的背景值最佳生成系数 p ，采用鱼群算法对其进行优化。人工鱼群算法^[7]是由李晓磊等人在 2002 年提出的一类基于鱼群行为的群体智能优化算法。是一种自下而上的寻优算法^[8]，具有全局收敛性^[9]。该算法通过模拟鱼群的觅食、聚群、追尾和随机四项基本行为在搜索域中进行寻优求解，是集群体智能思想的具体应用。人工鱼群算法的主要流程包括以下步骤：

(1) 鱼群初始化

鱼群中每条人工鱼为给定范围内随机产生的一组实数，例如，鱼群的大小为 N ，待优化的参数分别为 x 、 y ，取值范围分别为 $[x_1, x_2]$ 和 $[y_1, y_2]$ ，则需要产生一个 2 行 N 列的初始鱼群，每列数值表示一条人工鱼的两个参数。

(2) 觅食行为

假设人工鱼群的当前状态为 X_i ，随机选取感知范围内的一个状态 X_j ，在求解极小值问题中如果第 i 条人工鱼所在位置的食物浓度 $Y_i < Y_j$ ，则进行下一步骤：

$$X_{\text{next}} = X_i + \text{rand}() \cdot \text{Step} \cdot \frac{X_j - X_i}{\|X_j - X_i\|} \quad (7.3)$$

式中， X_{next} 表示鱼群向前进一步的位置， $\text{rand}()$ 表示取值范围内的随机数， Step 表示鱼群移动的最大步长。反之，则重新选择随机状态 X_j ，

$$X_{\text{next}} = X_i + \text{rand}() \cdot \text{Step} \quad (7.4)$$

(3) 聚群行为

假设人工鱼的当前状态 X_i ，探索当前领域内满足 $\|X_j - X_i\| < \text{Visual}$ 的伙伴数目 n_f 和中心位置 X_c ，假如 $Y_c/n_f > Y_i$ (δ 表示拥挤度)，则表明在当前领域内，伙伴中心有较多食物且不太拥挤，则朝伙伴中心位置的方向进一步，反之，再次进行觅食行为。

(4) 追尾行为

假设人工鱼当前状态 X_i ，探索当前领域内满足 $\|X_j - X_i\| < \text{Visual}$ 的伙伴数目 n_f 和这

些伙伴中 Y_j 为最大的伙伴 X_j ，若满足 $Y_j / n_f > \delta Y_i$ ，则表明 X_j 的状态有较高的食物浓度且周围环境不太拥挤，则朝 X_j 的方向进一步，反之，重复执行觅食行为。

(5) 随机行为

随机行为的实现相对简单，在视野范围内随机选取一个状态，然后朝着该方向移动，此过程也可认为是觅食过程的一个缺省行为，即：

$$X_{i|next} = X_i + rand() \cdot Visual \quad (7.5)$$

7.2.2 优化目标的选取

为了所优化的背景值最佳生成系数 p 可以使得灰度模型的预测更为接近实际数据，优化的目标函数取预测数据与实际数据的方差和最小，即

$$\sigma^2 = \sum_{i=2009}^{2018} (x_i - r_i)^2 \quad (7.6)$$

式中， x_i 、 r_i 分别为 i 年的预测数据和实际数据。

7.2.3 优化约束条件

为使得预测结果的方差限制在一定范围内，对优化进行一定的约束。这里因为需要优化的指标较多，且量纲不一致，所以采取归一化方法设置约束条件。

$$\sigma_i = \sum_{i=2009}^{2018} \frac{|G-E|}{\bar{x}} \quad (7.7)$$

$$\sigma < 5 \quad (7.8)$$

人工鱼群算法的整体流程如图 7.1 所示：

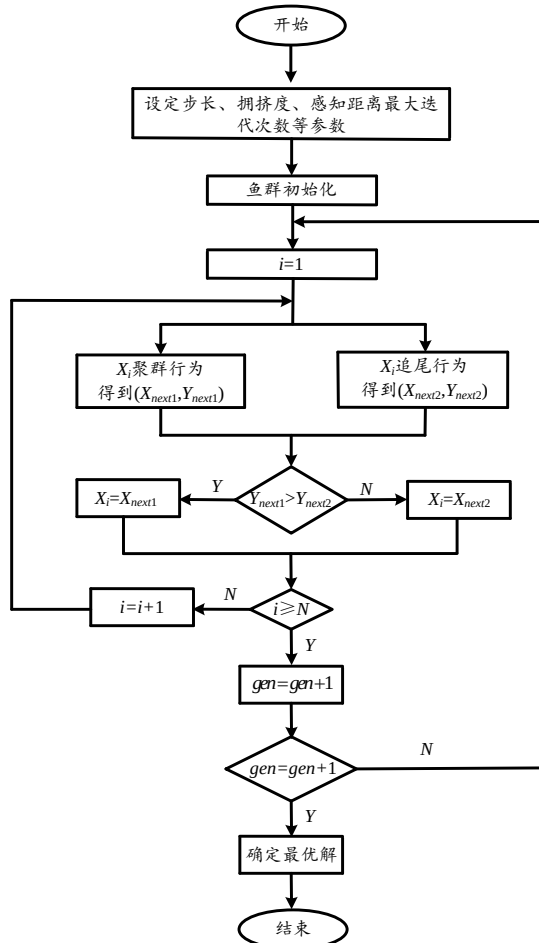


图 7.1 人工鱼群算法流程图

7.3、基本面指标与流动性指标的预测

鱼群算法所选取的参数：人工鱼数量为 50 只，最大迭代次数为 50 次，觅食最大试探次数为 100 次，感知距离为 1，拥挤度因子取 0.618，移动步长为 0.1。经过多次优化后，选择表 7.1 的各数值为待优化参数的优化结果。

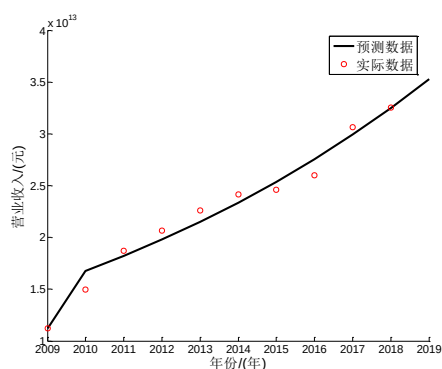
表 7.1 背景值最佳生成系数优化结果

市场类别	参数	数值
中国 A 股市场	p_1	0.4632
	p_2	0.3587
	p_3	0.6645
	p_4	0.4785
	p_5	0.5852
	p_6	/
	p_7	0.2589
美国 NASDAQ 市场	p_1'	0.4258
	p_2'	0.2965
	p_3'	/
	p_4'	/
	p_5'	0.5122
	p_6'	0.3578
	p_7'	0.3145

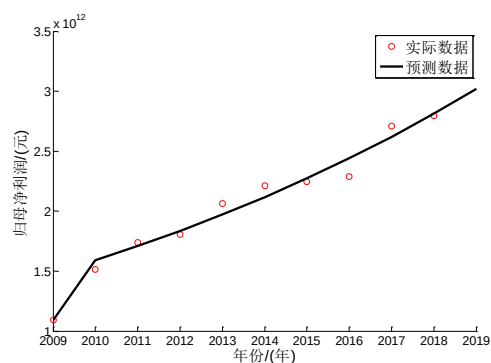
基于鱼群算法优化的加权灰色预测的程序求解步骤如下：

- (1) 对原始数据进行累加；
- (2) 构造累加矩阵 B 与常数向量；
- (3) 利用鱼群算法优化背景值最佳生成系数；
- (4) 求解灰参数；
- (5) 将参数带入预测模型进行数据预测。

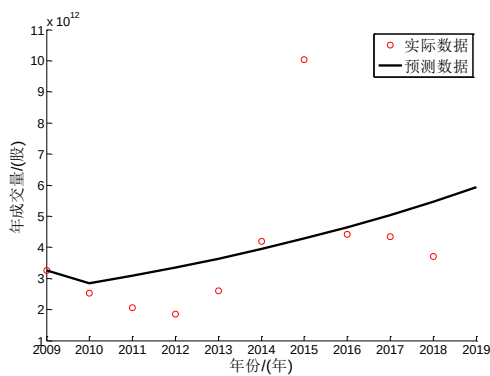
其预测结果如图 7.2，即 2019 年的中国 A 股市场的基本面指标与流动性指标。



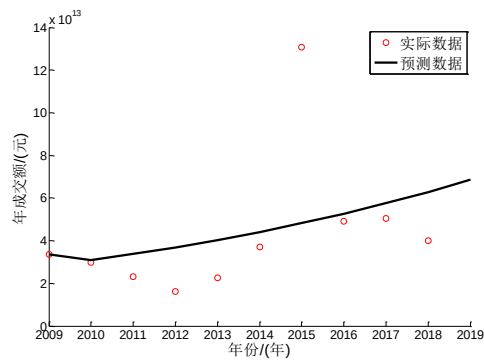
(a) 营业收入



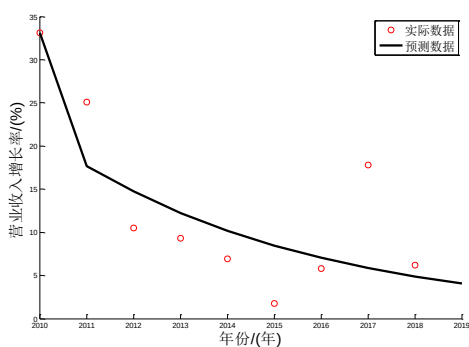
(b) 归母净利润



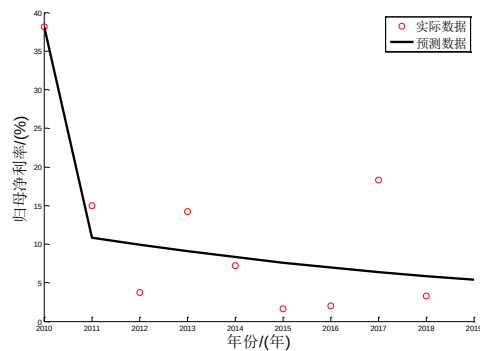
(c) 年成交量



(d) 年成交额



(e) 营业收入增长率



(f) 归母净利润率

图 7.2 中国 A 股市场基于鱼群算法优化的加权灰色预测结果

从图 7.2 中可以看出，营业收入、归母净利润、年成交额、营业收入增长率、归母净利润率的预测较为准确，而年均换手率则由于不满足约束条件，无法完成预测。从实际数据，可以看出，年均换手率呈较为明显的波动，灰色预测无法对其进行精确的建模预测，故考虑采取多项式拟合对其进行预测，其结果见图 7.3。

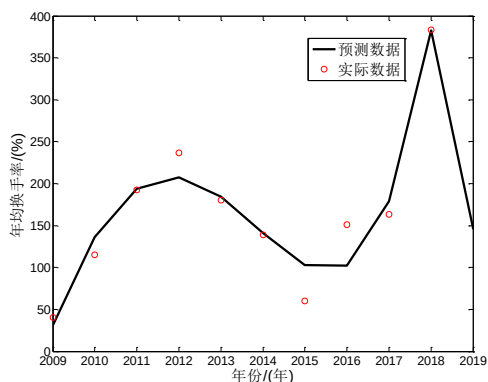


图 7.3 中国 A 股市场基于多项式拟合预测的年均换手率

最终，得到的中国 A 股市场基本面指标、流动性指标的预测见表 7.2:

表 7.2 中国 A 股市场预测结果

指标	指标名称	预测数据
基本面指标	营业收入 S (元)	3.529×10^{13}
	归母净利润 G (元)	3.019×10^{12}
	营业收入增长率 SL (%)	4.071
	归母净利率 GL (%)	5.367
流动性指标	年成交量 C (股)	5.927×10^{12}
	年均换手率 H (%)	145.46
	年成交额 E (元)	6.873×10^{13}

对于美国 NASDAQ 市场，基于鱼群算法优化的加权灰色模型预测结果如下图所示：

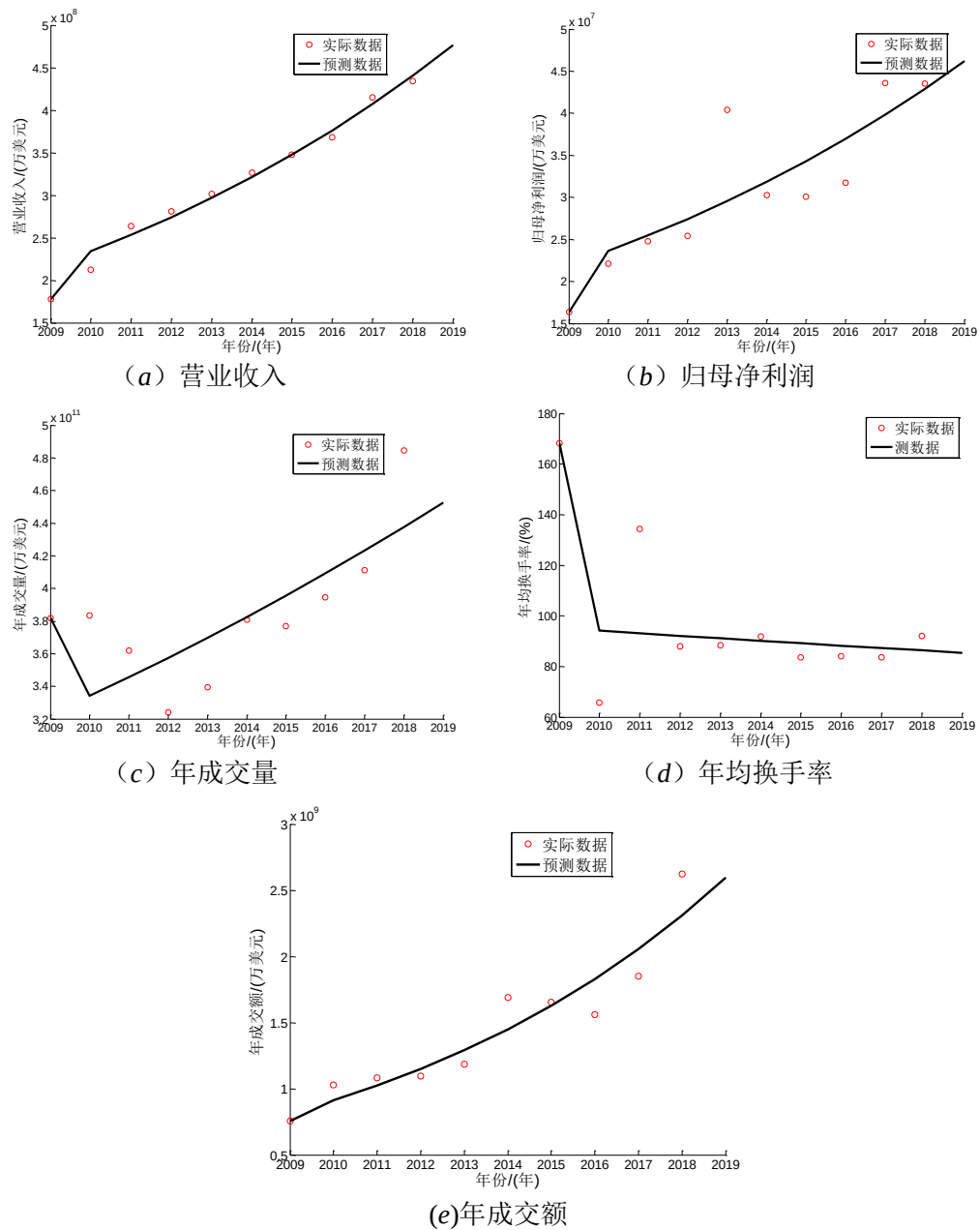


图 7.4 美国 NASDAQ 市场基于鱼群算法优化的加权灰色预测预测结果
同样，对于不满足优化约束条件的营业收入增长率和归母净利率，采用多项式

预测，结果见图 7.5。

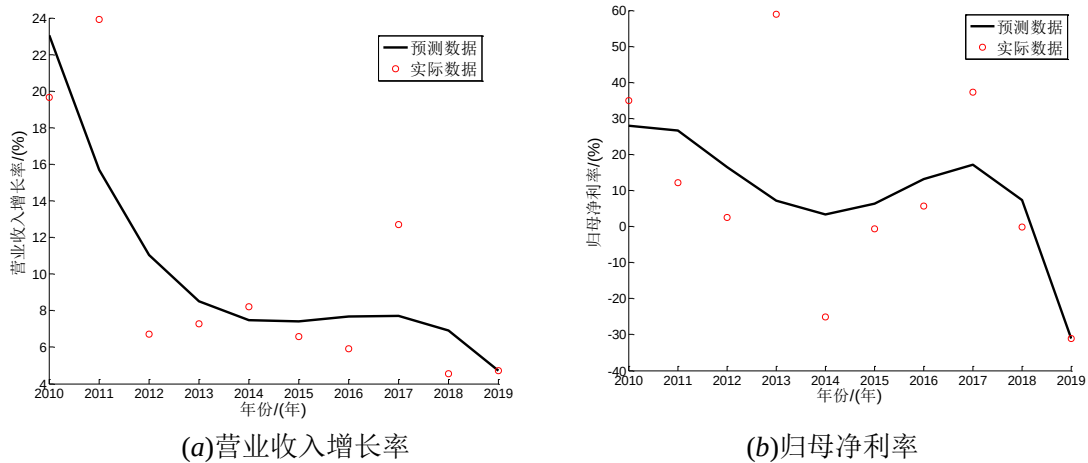


图 7.5 美国 NASDAQ 市场基于多项式拟合预测的年均换手率

最终，得到的美国 NASDAQ 市场基本面指标、流动性指标的预测见表 7.3:

表 7.3 美国 NASDAQ 市场预测结果

指标	指标名称	预测数据
基本面指标	营业收入 S (万美元)	4.767×10^8
	归母净利润 G (万美元)	4.621×10^7
	营业收入增长率 SL (%)	4.7
	归母净利润率 GL (%)	-31.09
流动性指标	年成交量 C (股)	4.525×10^{11}
	年均换手率 H (%)	85.44
	年成交额 E (万美元)	2.594×10^9

7.4、市场估值指标计算

由预测结果及问题 2 中所建立的市场估值指标与基本面指标、流动性指标之间的关系,可以进一步计算出中国 A 股市场和美国 NASDAQ 市场 2019 年的估值指标。

通过比较中国市场 2018 年的估值水平,可得到 A 股市场营业收入指标、归母净利润率指标、营业收入增长率指标、归母净利润率指标分别为 0.0081、0.0068、0.2760、0.5042,进而可得到基本面指标为 0.1653,流动性指标为 0.116,与 A 股中三友化工较为接近,因此对应中国 A 股市场的市销率为 3.9 倍,中国 A 股市场的估值约为 1.38×10^{14} 元。

通过比较美国 NASDAQ 市场 2018 年的估值水平,可得到美国 NASDAQ 市场营业收入指标 0.0032、归母净利润率指标 0.0890、营业收入增长率指标 0.014、归母净利润率指标分别为 0.476,进而可得到基本面指标为 0.1064,流动性指标为 1.74,通过查找美股公司可知该指标与美股 SUNESIS PHARMACEUTICALS 较为相近,因此可推算对应美股市场的市销率为 20.58,美国 NASDAQ 市场的估值约为 9.8×10^9 万美元。

§8 问题 4 科创板估值水平预测

根据附件 3 中数据，结合问题二中所建立的美国 NASDAQ 市场的估值量化模型，以及 A 股市场的流动性指标，可以得出 2019 年科创板上市后的估值水平。图 8.1 为采用问题 3 中的基于鱼群算法优化的加权灰色预测模型求解的科创板 2019 年营业收入与归母净利润的预测数值。营业收入增长率与归母净利率由于只有两年的数据，预测误差较大，所以直接采用了 2016-2018 年所提供数据结合所预测的 2019 年营业收入与归母净利润的预测数值求解计算得出。

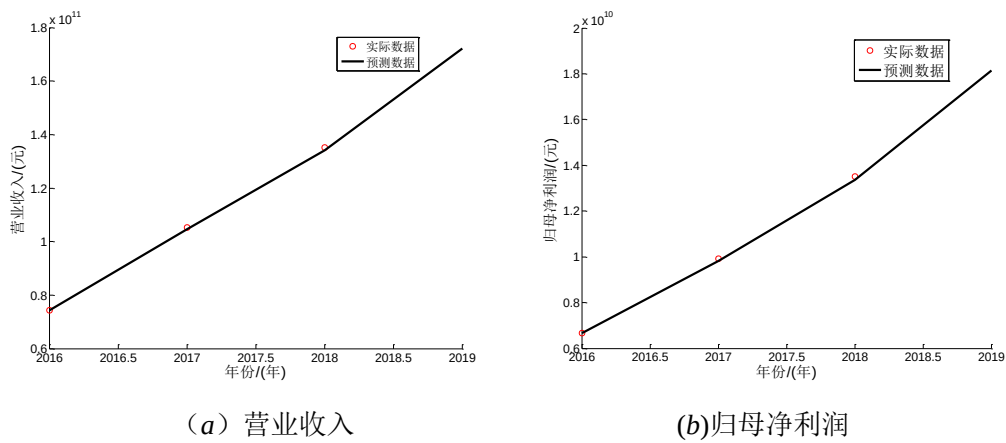


图 8.1 2019 年科创板基本面指标预测

而对于流动性指标，主要由投资者的投资情绪以及对于市场的估值预期所决定，因此，对于该指标而言，在短期内一般具有一致性特征，依据 2019 年中国 A 股市场的流动性指标近似代替科创板平均换手率。

由此，得到的中国创业板市场基本面指标、流动性指标的预测见表 8.1：

表 8.1 中国创业板市场预测

指标	指标名称	预测数据
基本面指标	营业收入 S (元)	1.722×10^{11}
	归母净利润 G (元)	1.815×10^{10}
	营业收入增长率 SL (%)	27.414
	归母净利率 GL (%)	34.299
流动性指标	年成交量 C (股)	3.67×10^{11}
	年均换手率 H (%)	145.46
	年成交额 E (元)	4.2×10^{12}

为了衡量科创板市场的公司平均基本面状况，以 2018 年美股市场的其余股票基本面作为参考，可获得科创板公司的营业收入指标、归母净利润指标、营业收入增长率指标以及归母净利润指标分别为 0.006、0.003、0.066、0.57，由此可进一步获得科创板公司的平均基本面指数为 0.124，对应流动性指标为 2.9。根据第二问中所给出的市销率估算方法可知，其与美股中的 PFG.0 信安金融具有相似的流动性指标与基本面指标，因此可估算出科创板的市销率约为 1.21 倍，对应市值约为 2.135×10^{11} 元。

§9 模型优缺点分析

9.1 模型优点

(1) 模型中尽可能多的考虑了股市企业的有效数据，同时对市场中的无效数据进行了剔除，力求能够尽量真实的反映股市实际状况。

(2) 建立市场的基本面指标分析模型时，对于基本面的各项指标综合过程中采用模糊层次分析法，权重选取更为客观，可以使得评价更为科学。后期的验证也不断证明了所建立模型的科学性。

(3) 采用强灵敏 SOM 变量聚类的方法，将市场分为 5 种主要的股市评估类别，能够将股市中的企业归类于其中，并利用企业数据距离 5 个主要的股市评估类别中心数据的空间距离进行插值，得到了市场估值指标、基本面指标和流动性指标之间的定量关系。

(4) 采用基于鱼群算法优化的加权灰色模型结合多项式拟合对 2019 年的股市各项指标进行了预测。改进的灰色模型可以更准确的预测数据，并在置信度较低时停止预测，而多项式拟合预测此时则起到了替补作用。

9.2 模型缺点

(1) 本模型所基于的数据仅为 2009-2018 年的统计数据，总体数据量还是较少，对于股市来说，时间跨度不算长，且未考虑政策引导，社会经济情况灯等一些外在因素对于股市的影响。

(2) 由于附录中所给出的参数有限，导致在预测基本面指标与流动性指标时只能仅仅依据时间作为自变量进行预测，虽然时间与这两种因素具有相关性，但是这种相关性属于间接相关性，如需要进一步提高指标的预测精度，应当尝试利用直接相关性数据进行拟合与预测。

参考文献

- [1] 焦李成, 杨淑媛, 刘芳,等. 神经网络七十年:回顾与展望[J]. 计算机学报, 2016, 39(8):1697-1716.
- [2] 袁晓, 肖瑾. 应用模糊目标规划方法求解多目标线性分式规划问题[J]. 数学的实践与认识, 2016, 46(6):158-164.
- [3] 李炳军, 刘思峰, 任盈盈. 层次灰色准优势分析模型及应用(英文)[J]. Transactions of Nanjing University of Aeronautics and Astronautics, 2005, 22(1):78-83.
- [4] 基于 SOM 聚类的数据挖掘方法及其应用研究[D]. 浙江大学, 2006.
- [5] Wang C H, Chen C Y, Hung K N. Toward a new task assignment and path evolution (TAPE) for missile defense system (MDS) using intelligent adaptive SOM with recurrent neural networks (RNNs)[J]. IEEE Transactions on Cybernetics, 2017, 45(6):1134-1145.
- [6] Klauberg C, Hudak A T, Bright B C, et al. Use of ordinary kriging and Gaussian conditional simulation to interpolate airborne fire radiative energy density estimates[J]. International Journal of Wildland Fire, 2018, 27(4):228-238.
- [7] 李晓磊. 一种新型的智能优化算法—人工鱼群算法[D]. 杭州: 浙江大学, 2003.
- [8] 李晓磊, 邵之江, 钱积新. 一种基于动物自治体的寻优模式: 鱼群算法[J]. 系统工程理论与实践, 2002, 11: 32-38.
- [9] 黄光球, 刘嘉飞, 姚玉霞. 人工鱼群算法的全局收敛性证明[J]. 计算机工程, 2012, 38(2): 204-206.