蔬菜商品定价与销售的关联与规划模型

摘要

在生鲜商超中，蔬菜类商品具有保鲜期短的特点，随着销售时间的增加，品相也会逐渐变差。此外，前一天未售出的蔬菜商品在隔日通常无法再售出。考虑到蔬菜种类繁多且产地各异，商家通常在凌晨进行进货，但无法准确知道每个单品的具体进货价格。因此，根究各蔬菜商品的历史销售和需求情况做出定价与补货决策具有长远意义。

本题的数据较多，在建立模型前首先对数据进行预处理，定义了两类异常数据。类型 1 为蔬菜单品每天的销售单价与批发价格的比值大于 5, 类型 2 为蔬菜单品每天的销售单价与批发价格的比值小于 1 且当天该蔬菜单品未打折销售。在此基础上，进一步使用四分位法筛选蔬菜单品销售单价的异常数据。经过数据预处理，总共剔除了 4.63% 的异常数据。

针对问题一，以一个季度为时间周期，对蔬菜各品类的销量均值、方差等统计学指标进行可视化，并从中发现部分品类蔬菜具有周期性或互补性。采用灰色关联模型对各类蔬菜进行定量分析，最终求解出花叶类与花菜类、水生根茎类与食用菌、水生根茎类与茄类、食用菌和茄类四对品类具有高相关性，灰色关联度分别为 0.86、0.81、0.83、0.78。对各单品销量占所属品类总销量比例进行可视化，发现各类别中各单品蔬菜销售量差异较大。采用斯皮尔曼系数对各蔬菜单品对进行相关性分析，求解出菜心与西兰花、黄白菜与菠菜等七对品类显著性大于 0.99，黄白菜与菜心、奶白菜与菠菜等七对品类显著性大于 0.95。

针对问题二，对蔬菜品类的成本加成定价、蔬菜品类的批发价格等指标进行定义。依据历史数据，采用最小二乘拟合法对蔬菜品类销量与成本加成定价进行拟合，最终六类蔬菜拟合优度均小于 0.1。为提高曲线拟合的准确率，将各类蔬菜相同定价下的平均销量代替销量并进行拟合处理，最终六类蔬菜拟合优度均值为 0.67。考虑到商品销量具有周期性变化规律，故与最后一个月销售总量均值进行线性加权处理。采用 LSTM 模型预测出未来一周蔬菜的批发价格，结合约束条件建立目标规划模型，使用粒子群优化算法求解出商超未来一周收益总值为 7233.95 元。

针对问题三，首先统计出各品类蔬菜中共有 48 个可售单品，并对可售单品的成本加成定价、批发价格等指标进行定义。在问题二的基础上，为尽量满足市场对各蔬菜品类的需求，以历史平均销量上下波动 30% 为新增约束条件，建立基于 0-1 规划模型，使用基于 0-1 变量的粒子群优化算法求解出商超当天最大收益值为 715.20 元。

针对问题四，商家应该收集各蔬菜单品的新鲜度随时间变化关系函数 $\theta(t)$, 通过与微观经济学中常见需求函数 $D(p) = a - bx$ 的类比，建立蔬菜商品的需求函数 $D(y,t) = a\theta(t) - by(t)$，而商家可以通过定时对蔬菜单品进行外观、嗅觉、化学试剂检验等方式确定 $\theta(t)$。此外，还应该通过问卷调查的方式对竞争商超的单品销量和定价信息进行收集，从而调整策略以提高利润。

关键字： **灰色关联；时间序列预测；最小二乘法；基于 0-1 变量规划的粒子群算法**

一、 问题重述

1.1 问题背景

在生鲜商超中，由于一般蔬菜类商品具有保鲜期短、品相随销售时间的增加而变差和前一天剩余商品隔日就无法售出的特点。因此，研究各商品的历史销售和需求情况每天进行补货具有十分重要的意义。

由于商超销售的蔬菜品种多、产地大多不同，而蔬菜的进货交易时间通常在凌晨，于是商家在不确定具体单品和进货价格的情况下就会做出补货决策。蔬菜的定价一般采用“成本加成定价”方法，商超通常对品相较差的商品进行打折销售。可靠的市场需求分析，对补货决策和定价决策尤为重要。

1.2 问题提出

现已知某商超销售的蔬菜类相关数据（附件 1 的蔬菜品类商品信息，附件 2 的该商超 2020 年 7 月 1 日至 2023 年 6 月 30 日各商品的销售流水明细，附件 3 的 2020 年 7 月 1 日至 2023 年 6 月 30 日各商品的批发价格，附件 4 的各商品近期的损耗率数据），蔬菜的供应品种在 4 月至 10 月较为丰富。

基于上述条件，需要我们建立合适的数学模型解决以下问题：

（1）对蔬菜各品类及单品销售量的分布规律进行分析，并探究它们之间可能存在的关联关系。

（2）假设商超以品类为单位做补货计划，分析各蔬菜品类的销售总量与成本加成定价的关系，并以商超收益最大为目标，计算出各蔬菜品类未来一周（2023 年 7 月 1-7 日）的日补货总量和定价决策。

（3）由于蔬菜类商品的有限销售空间有限，商超打算制定单品补货计划。计划要求可售单品总数在 27-33 个之间，并且每个单品的订购量需要达到最小陈列量 2.5 千克。根据 2023 年 6 月 24-30 日的可售品种，给出 7 月 1 日的单品补货量和定价策略。

（4）为了使制定的蔬菜商品补货和定价决策更加完善合理，给出商超还需要采集哪些相关数据的建议。

二、 问题分析

2.1 问题一的分析

问题一要求分析蔬菜各品类及单品销售量的分布规律及相互关系。因此，我们首先以季度为时间周期统计出蔬菜各品类和蔬菜各单品在各季度的销售量。其次，对于蔬菜品类，我们衡量了蔬菜各品类在各季度的销售占比并计算了部分统计学指标以描述其分布规律。使用散点图定性的确定了蔬菜各品类的可能具有的关系，后采用灰色关联模型进一步定量的确定了相关程度。对于蔬菜单品，我们整理并筛选了数据，剔除了在某种

蔬菜品类销售量占比权重较小的蔬菜单品，后计算了余下蔬菜单品的统计学指标描述其分布规律。结合蔬菜品类可能具有的关系，使用斯皮尔曼相关系数衡量各蔬菜单品的相关关系。

2.2 问题二的分析

问题二要求以品类为单位做补货计划，分析蔬菜各品类的销售总量与成本加成定价的关系并给出蔬菜各品类的最优日补货总量和定价策略。为衡量蔬菜各品类的销售总量与成本加成的关系，采用最小二乘拟合方法进行分析，经多次尝试确定最优拟合函数。为确定最优日补货量和定价策略，我们以日收益最大为目标，建立一般规划模型，为求得约束条件，使用 LSTM 进行未来一周各蔬菜品类销量预测。

2.3 问题三的分析

问题三要求挑选出在最后一周具有销售记录的单品，在所购单品尽可能满足各类蔬菜商品的需求，针对满足要求的可售单品做补货计划，给出可售单品的最优日补货总量与定价策略。我们在问题二所建立的各类蔬菜销售总量与成本加成定价的关联模型基础上，计算出可售单品销售总量与成本加成定价的关联函数，以当日收益最大为目标，以各品类可售单品销售总量处于该类蔬菜销售总量均值波动范围区间内作为约束条件，建立单目标规划模型进行求解。

2.4 问题四的分析

问题四要求商超收集更多蔬菜商品的相关信息，使得制定的商品补货量和成本加成定价决策更加完善。通过查阅相关文献，发现蔬菜商品的新鲜程度对顾客购买率存在较大影响，需要采集各蔬菜单品新鲜度与时间的关联信息，我们在此基础上建立出蔬菜商品的需求函数，并提供部分商超采集新鲜度信息的方法；此外，商超还应考虑到与其他超市之间的竞争关系，即收集其他超市各蔬菜商品的销售量和定价，通过调整商超各单品的成本加成率，以提高超市间的产品竞争力，从而提高商品收益，对此，我们提供了以设计调查问卷的形式对其他超市商品信息进行收集的解决思路。

综上，问题一、二、三、四可视分析流程如图 1 所示

三、 模型的假设

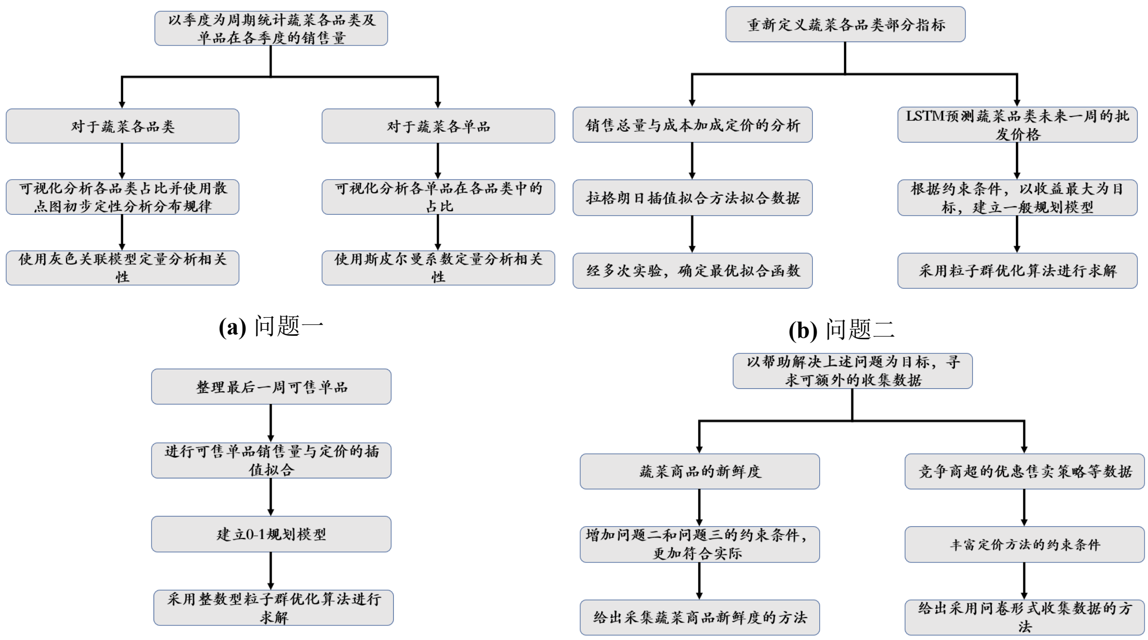
假设 1：两组蔬菜单品所属蔬菜品类之间具有强相关性的单品才可能具有相关性。

假设 2：附件所给蔬菜各品类及单品在一定时间内销售量不会发生突变，满足一定销售规律。

假设 3：附件 2 所给蔬菜各单品的售价即为成本加成定价。

假设 4：蔬菜各单品的损耗是指除正常售卖外的所有损耗。

假设 5：蔬菜各品类及单品的销售量趋于稳定，不会出现与附件所给数据相差很大的情况。



(a) 问题一　**(b)** 问题二　**(c)** 问题三　**(d)** 问题四

图 1: 问题分析总体可视图

四、 符号说明

符号	符号说明	符号	符号说明
χ	灰色关联相关分辨系数	v_i	灰色关联母序列
τ_{ij}	i、j 母序列关联度	ξ_i	灰色关联系数
Q_j	序列 X_i 的第 j 个四分位点	r_s	斯皮尔曼相关系数
I_{QR}	序列 X_i 的四分位点距	d_{ij}	i 序列与 j 序列等级差
χ	灰色关联的相关分布系数	r_i	最小二乘偏差

- 注：其他符号将在正文中给出定义。

五、 数据预处理

本题附件 2 和附件 3 分别给出了蔬菜类商品的销售明细和蔬菜类商品的批发价格。由于数据量过大，为了避免可能存在的错误数据对计算结果的影响导致误差的发生，需要采取以下预处理措施。

借助 Excel 软件对附件 2 和附件 3 进行整理，得到蔬菜各单品的编码，蔬菜各单品每天的批发价格、销售单价、销量和损耗率，蔬菜各单品每天的销售单价与批发价格的比值 η。如表 1 所示：

基于表 1，定义了以下两种异常数据类型：

类型 1：蔬菜单品每天的销售单价与批发价格的比值 $\eta > 5$。

表 1: 对附件 2 和附件 3 的统计性分析表

日期	单品编码	批发价格	销售单价	销量	损耗率	η
2023-07-01	102900005117056	3.88	7.60	0.515	7.08%	1.96
2023-07-01	102900005115779	6.72	8.00	0476	15.25%	1.19
2023-07-01	102900005115786	3.19	6.00	0.452	13.62%	1.88
…	…	…	…	…	…	…

- 注：由于篇幅有限，完整的整理数据将在附件中展示。

类型 2：蔬菜单品每天的销售单价与批发价格的比值 $\eta < 1$ 且当天该蔬菜单品未打折销售。

经过数据筛选我们剔除了 413 个类型 1 异常数据，236 个类型 2 异常数据。此外，经数据分析，我们还发现蔬菜单品的销售单价存在异常数据。为进一步筛选异常数据，我们采用了四分位法 [1] 对蔬菜单品的销售单价进行异常数据筛选，具体过程如下：

Step1. 数据排序

对蔬菜各单品在 2020 年 7 月 1 日至 2023 年 6 月 30 日每天的销售单价进行升序排列，得到 $X_i = \{x_1, x_2, \cdots, x_n\}(i = 1, 2 \cdots 251)$。

Step2. 计算四分位数和四分位距

设 X_i 前三个四分位数分别为 Q_1, Q_2 和 Q_3，若 $n = 2k(k = 1, 2, \cdots)$，则 Q_2 为中间两个数的平均值。则 Q_2 点将 X_i 分为两个子序列，分别计算两子序列的中位数 Q_2' 和 Q_2''，则有 $Q_1 = Q_2'$，$Q_3 = Q_2''$

若 $n = 4k + 3(k = 1, 2, \cdots)$，则有

$$Q_1 = \frac{3}{4}x_{k+1} + \frac{1}{4}x_{k+2} \quad Q_3 = \frac{1}{4}x_{3k+1} + \frac{3}{4}x_{3k+2} \tag{1}$$

设四分位距为 I_{QR}，则 $I_{QR} = Q_3 - Q_1$。

Step3. 确定异常值内限范围

根据上述计算结果，可以通过 $Whisker$ 上限 W_1 和下限 W_2 来剔除序列中的异常值，序列 X_i 中的异常值内限范围定义如下：

$$[W_1, W_2] = [Q_1 - 1.5I_{QR}, Q_3 + 1.5I_{QR}] \tag{2}$$

经过 Python 软件编程筛选，剔除了 1513 个异常数据。

至此完成数据的整理与剔除，剔除了 4.63% 的异常数据，后文的建模和求解均建立在此数据预处理的基础上进行。

六、 模型的建立与求解

6.1 问题一

题目要求对蔬菜类商品的销售量进行分析，探索不同品类或单品之间的销售分布规律和相互关系。因此，本文首先以蔬菜各品类及单品销售量的分布规律季为时间周期统计蔬菜各品类及单品的销售量，得出可能存在的分布规律。继而针对蔬菜各品类进行灰色关联分析。最后，我们在蔬菜各品类关系的基础上，进行各单品的关联分析。

6.1.1 蔬菜各品类及单品销售量的分布规律分析

（一）对于蔬菜各品类

我们利用 Excel 软件整理了蔬菜各品类在 12 个季度的销售总量数据，并统计出每个季度蔬菜各品类销售量所占比例，具体结果如图 2 所示。此外，我们计算出蔬菜各品类销售量的一些重要统计指标，如表 2 所示。

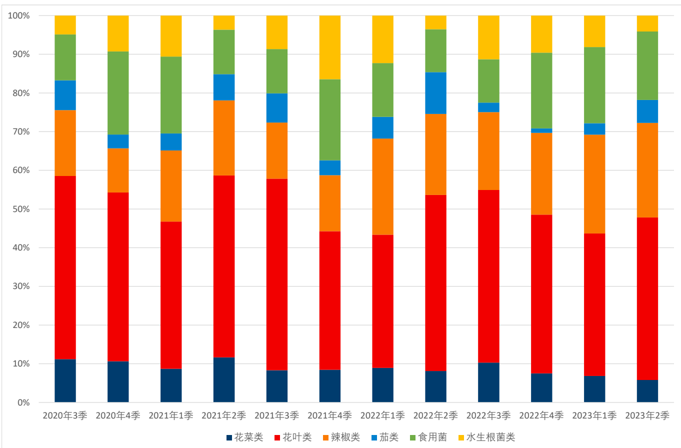


图 2: 蔬菜各品类各季度销量所占比例

表 2: 蔬菜各品类销售量的部分统计学指标

品类名称	总销量	销量均值	销量标准差	销量最大值	销量最小值
花菜类	40843.82	3403.65	1122.94	5408.766	1944.76
花叶类	192508.00	16042.34	4252.81	21672.06	8717.40
辣椒类	86653.53	7221.13	2734.93	12418.94	3594.12
茄类	22087.28	1840.60	744.21	3143.71	620.00
食用菌	74513.63	6209.47	2713.24	10427.89	2649.79
水生根茎类	39155.83	3262.97	1635.08	5861.86	745.47

• 注：上述各统计学指标的单位均为千克

结合图 2 和表 2，做出以下分析：

（1）在蔬菜各品类中，花叶类的销售占比最大，茄类的销售占比最小。

（2）大多数蔬菜品类的销售量在每年的第 3 季度达到最大值，在每年的第 1 季度达到最小值，满足蔬菜的供应品种在 4 月至 10 月较为丰富的前提条件。

（3）花叶类蔬菜每季度的销量量最稳定，茄类蔬菜每季度的销售最不稳定，即花叶类蔬菜的销售不具有季节性且茄类的销售具有季节性。

（4）辣椒类蔬菜和花叶类蔬菜的季度销量具有一定的互补性，即辣椒类蔬菜销量大的季度而花叶类蔬菜销量少，辣椒类蔬菜销量小的季度而花叶类蔬菜销量大。

（二）对于蔬菜各单品

我们使用 Excel 软件对数据进行整理，统计出每个单品销售量占其所属品类的比例。因篇幅有限，以花叶类为例，做出销售量占比图，如图 3 所示。

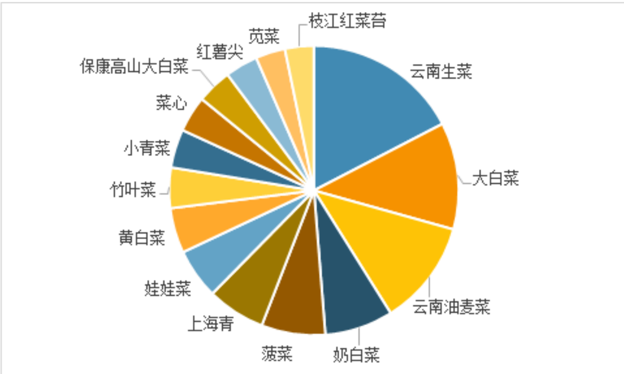


图 3: 花叶类单品销售量占比图

据图 3 可分析：在花叶类蔬菜中，云南生菜销售量所占比例最大。枝江红菜苔、苋菜和红薯尖销售量所占比例最小。同理可分析出其它蔬菜品类的销售量分布规律。

此外，由于蔬菜单品数量过多，在下文的分析中，我们只使用蔬菜各品类销售量大的单品进行分析。

6.1.2 蔬菜各品类及单品销售量的相互关系分析

（一）对于蔬菜各品类

为了确切的建立模型，我们首先对蔬菜各品类之间进行了定性分析，得到了蔬菜各品类之间的散点图，如图 4 所示。

依据图 4，可定性做以下分析：

根据之前的定性分析，我们初步确定了一些蔬菜品类之间可能存在一定的相关性。例如，花叶类和花菜类可能存在正相关关系，而茄类和水生根菌类可能存在负相关关系。接下来，我们将进行定量分析，以衡量蔬菜各品类之间的相关性程度。

计算蔬菜各品类之间的关联度，是根据蔬菜各品类每季度的销售量发展趋势的相似或相移程度确定的，由于部分数据过度偏离，造成传统的关联分析拟合效果不明显。因此，本文选用灰色关联分析对蔬菜各品类的关联度进行分析，并判断哪些蔬菜品类的相关性最高。

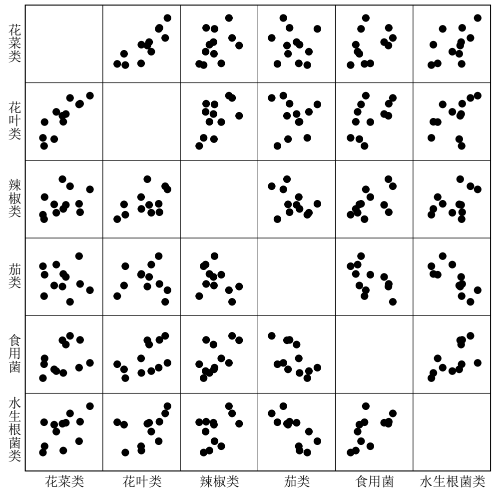


图 4: 蔬菜各品类间销量散点图

灰色关联分析 [2] 体现一种模糊关联的思想，其衡量了两个序列的整体相似性，而整体相似度又以局部相似度为基础进行讨论。现将六类蔬菜视为六个待求关联度的母序列 $v_1, v_2, v_3 \cdots v_6$，分别将这六个序列两两组合，以求出其蔬菜各品类的相关性。

设评价对象为 $m = 12$ 个，而评价指标变量为 $n = 6$ 个，凭借母序列 $v_1, v_2, v_3 \cdots v_6$ 可设：

（1）参考序列为

$$x_0 = \{x_0(k)|k = 1, 2 \cdots n\} = ((x_0(1)), (x_0(2)), \cdots (x_0(n))) \tag{3}$$

（2）比较序列可设置为

$$x_i = ((x_i(1)), (x_i(2)) \cdots (x_i(n))) \quad i = 1, 2 \cdots m \tag{4}$$

通过对参考序列与比较序列进行数据处理，可计算得到两两母序列间的灰色关联系数为：

$$\xi_i(k) = \frac{min_s min_t |x_0(t) - x_s(t)| + \chi max_s max_t |x_0(t) - x_s(t)|}{|x_0(t) - x_s(t)| + \chi max_s max_t |x_0(t) - x_s(t)|} \tag{5}$$

其中，$\chi \epsilon [0, 1]$，称为灰色关联相关分辨系数。

由灰色系统理论可知，χ 越大，则分辨率越大；χ 越小，则分辨率越小，为降低相关系数对实验仿真的影响，一般取分辨系数 $\chi = 0.5$

最后，求解关联系数的均值 $\bar{\xi}$ 满足：

$$\bar{\xi} = \frac{\sum_{i=1}^{n-1} \xi_i}{m} \tag{6}$$

则可以得到第 i 个母序列与第 j 个母序列的关联度为：

$$\tau_{ij} = |\bar{\xi}| = \left\| \frac{\sum_{i=1}^{n-1} \xi_i}{m} \right\| \tag{7}$$

根据所查文献，局部相似度的度量效果对整体序列关联度的计算起到决定性作用。则可知，灰色关联度 τ_{ij} 越大的两个序列，其关联程度也越大。现将各指标变量数据带入比较序列中，并将母序列两两组合，利用上述模型求解灰色关联度，可得到图 5。

	花菜类	花叶类	辣椒类	茄类	食用菌	水生根茎类
花菜类	1.00	0.86	0.71	0.68	0.69	0.72
花叶类	0.86	1.00	0.71	0.69	0.72	0.69
辣椒类	0.71	0.71	1.00	0.66	0.74	0.66
茄类	0.68	0.69	0.66	1.00	0.78	0.83
食用菌	0.69	0.72	0.74	0.78	1.00	0.81
水生根茎类	0.72	0.69	0.66	0.83	0.81	1.00

图 5: 蔬菜各品类灰色关联度图

根据蔬菜各品类销售量差异进行适当加权平均化处理，从而得到相关关系标准，即对于 $\tau_{ij} > 76.95\%$ 的关联度称为高相关性。

现结合图 5 与上文论述进行综合分析：依据上述标准，不难发现花叶类与花菜类、水生根茎类与食用菌、水生根茎类与茄类和食用菌与茄类四对蔬菜品类具有高相关性，其相关系数均大于相关系数标准。具有这种关系的原因很可能与这几类蔬菜品类的生活环境和收获季节相近。

（二）对于蔬菜品类各单品

依据模型假设 1，我们认为只有属于上述四对蔬菜品类的蔬菜单品才可能具有相关性，故下文只对属于该四对蔬菜品类的蔬菜单品进行讨论分析。

在数据科学领域中，正态分布被广泛应用，并成为以概率分布为核心的主要研究对象。在对蔬菜各单品进行分析时，如果数据符合正态分布，将在极大程度上简化模型。以花叶类与花菜类具有强相关关系的蔬菜品类对为例（完整的数据分析图表在附录中展示），我们对蔬菜各单品的销售数据进行正态检验，具体如图 6 所示。其中直方图是各数据指标在各值的数量，黄色曲线各数据指标代表核密度估计。

据图 6 可做如下分析：

蔬菜各单品中只有云南生菜和云南油麦菜的数据可近似为正态分布，其它数据指标无法视作正态分布进行处理。因此，一些需要指标数据满足正态分布的相关性分析方法无法使用，例如皮尔逊相关性分析方法。所以，我们采用斯皮尔曼相关系数分析方法。

设 A 和 B 为两组数据，其斯皮尔曼相关系数：

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \tag{8}$$

其中，d_i 为 A_i 和 B_i 之间的等级差。

经 SPSS 进行模型求解，得到各单品之间的斯皮尔曼系数图，如图 7 所示。

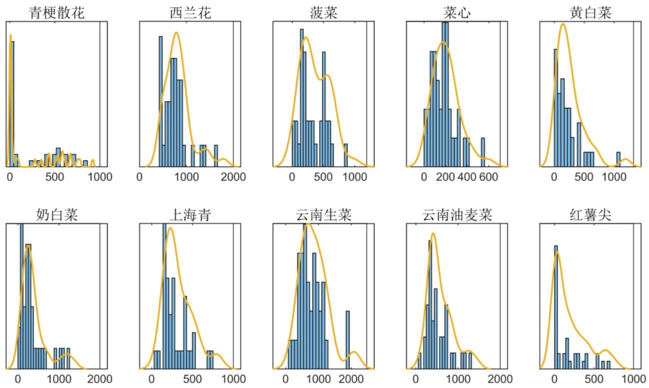


图 6: 蔬菜各单品分析图

	青梗散花	西兰花	菠菜	菜心	黄白菜	奶白菜	上海青	云南生菜	云南油麦菜	红薯尖
青梗散花	1.000	0.296	-0.585	0.254	0.430	-0.676	-0.359	-0.416	-0.134	0.092
西兰花	0.296	1.000	-0.161	**0.783****	0.462	-0.343	0.189	0.021	0.308	-0.035
菠菜	0.585	-0.161	1.000	-0.385	**0.762****	0.671*	0.510	0.420	0.126	-0.636
菜心	0.254	**0.783****	-0.385	1.000	0.615*	-0.224	0.063	0.336	0.552	0.245
黄白菜	0.430	0.462	**0.762****	0.615*	1.000	-0.469	-0.196	-0.021	0.224	**0.790****
奶白菜	0.676	-0.343	0.671*	-0.224	-0.469	1.000	**0.727****	0.566	0.126	-0.259
上海青	0.359	-0.189	0.510	0.063	-0.196	**0.727****	1.000	**0.867****	0.650*	-0.021
云南生菜	0.416	0.021	0.420	0.336	-0.021	0.566	**0.867****	1.000	**0.748****	0.126
云南油麦菜	0.134	0.308	0.126	0.552	0.224	0.126	0.650*	**0.748****	1.000	0.126
红薯尖	0.092	-0.035	-0.636	0.245	**0.790****	-0.259	-0.021	0.126	0.126	1.000

注：**.在 0.01 级别相关性显著；*.在 0.05 级别相关性显著

图 7: 各蔬菜单品间的斯皮尔曼系数图

据图 7 可做以下分析：在花叶类-花菜类中，菜心与西兰花、黄白菜与菠菜、红薯尖与黄白菜、上海青与奶白菜、云南生菜与上海青、云南油麦菜与云南生菜六对品类显著性最高，在 99% 的概率下具有相关性，其相关系数分别为 0.783、0.762、0.790、0.727、0.867、0.748。黄白菜与菜心、奶白菜与菠菜、云南油麦菜与上海青三对品类显著性较高，在 95% 的概率下具有相关性，其相关系数分别为 0.615、0.671、0.650。其余品类之间可认为无相关性。据补充 1 可做以下分析：在水生根茎类-茄类中，净藕与紫茄子一对品类显著性最高，在 99% 的概率下具有相关性，其相关系数为-0.734。净藕与青茄子一对品类显著性较高，在 95% 的概率下具有相关性，其相关系数为-0.671。其余品类之间可认为无相关性。据补充 2 可做以下分析：在水生根茎类-食用菌中，净藕与高瓜、洪湖莲藕与金针菇两对品类显著性较高，在 95% 的概率下具有相关性，其相关系数分别为-0.601、0.678。其余品类之间可认为无相关性。据补充 3 可做以下分析：在食用菌与茄类中，金针菇与青茄子一对品类显著性较高，在 95% 的概率下具有相关性，其相关系数为-0.587。其余品类之间可认为无相关性。

6.2 问题二

6.3 建模前的准备

为分析蔬菜各品类的销售总量与成本加成定价的关系。我们首先统计了六类蔬菜品类的销售总量，其次，依据假设 3，附件 2 中蔬菜各单品的销售单价即为该单品的成本加成定价，则蔬菜某品类的成本加成定价为：

$$\text{蔬菜某品类的成本加成定价} = \sum_{\text{所有单品}} \frac{\text{单品成本加成定价 * 单品销售量}}{\text{该蔬菜品类的总销量}} \tag{9}$$

为分析蔬菜品类的日补货总量和定价策略使得商超收益最大，必然需要计算商超每天得进货蔬菜各单品的进货量，依据假设 4，蔬菜各单品的损耗是指除正常售卖外的所有损耗，则可计算蔬菜各单品的进货量：

$$\text{蔬菜某单品的进货量} = \frac{\text{该蔬菜单品的售货量}}{\text{1-该蔬菜单品的损耗率}} \tag{10}$$

此外，我们定义蔬菜各品类的批发价格如下：

$$\text{蔬菜某品类的批发价格} = \sum_{\text{所有单品}} \frac{\text{单品批发价格 * 单品销售量}}{\text{该蔬菜品类总销售量}} \tag{11}$$

蔬菜各品类的损耗率如下：

$$\text{蔬菜某品类的损耗率} = \sum_{\text{所有单品}} \frac{\text{单品损耗率 * 单品销售量}}{\text{该蔬菜品类总销售量}} \tag{12}$$

6.3.1 对蔬菜各品类的销售总量与成本加成定价的分析

为定量地分析蔬菜各品类的销售总量与成本加成定价，我们使用了蔬菜各品类三年的数据进行最小二乘拟合方法 [3] 进行分析拟合。

最小二乘法的主要思想是使得最小二乘标准偏差最小，最小二乘偏差的定义如下：

$$r_i = f(x_i; \alpha_1, \alpha_2, \cdots, \alpha_n) - y_i, i = 1, 2, \cdots, N \tag{13}$$

使用 MATLAB 软件进行求解，发现拟合效果不佳。后经分析发现，部分蔬菜品类在成本加成定价相同时，销售总量却不同，这对拟合效果产生了较大的影响。为此，我们近一步进行数据处理，对于成本加成定价相同时，销售总量却不同的蔬菜品类的销售量数据，采用平均销售量替代销售量，得到了较好的拟合效果。如图 8 所示。

其中，各子图的的横坐标为该蔬菜品类的成本加成定价，纵坐标为销售总量。由图 6 可知，花菜类和食用菌的拟合曲线具有相同的趋势，花叶类、辣椒类和茄类的拟合曲线具有相同的趋势，水生根茎类的拟合曲线最为不同。

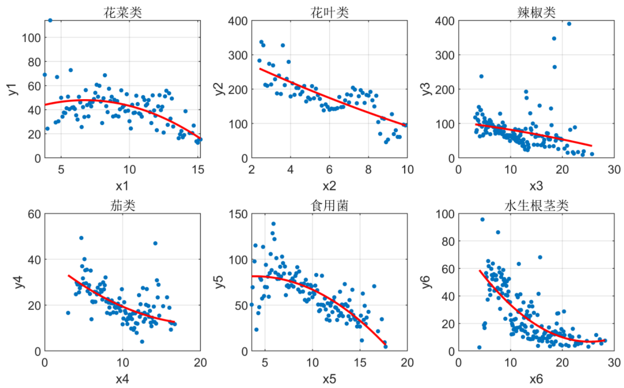


图 8: 蔬菜各品类拟合曲线图

为进一步求取拟合曲线的表达式，经多次尝试各种函数进行拟合，我们发现形如 $y = ax^2 + bx + c$ 的二次函数拟合程度最高。蔬菜各品类的拟合曲线系数及拟合优度如下表所示。

表 3: 蔬菜各品类的拟合曲线系数表

品类名称	a	b	c	拟合优度
花菜类	-0.4468	6.0297	27.5476	0.6549
花叶类	0.4420	-27.4844	323.2484	0.8146
辣椒类	-0.0310	-1.8916	103.8160	0.7156
茄类	0.0694	-2.8557	40.9736	0.5189
食用菌	-0.3949	3.1004	75.3746	0.7849
水生根茎类	0.1151	-5.8167	80.2147	0.5461

6.3.2 日补货量与定价策略的确定

为了给出蔬菜各品类未来一周 (2023 年 7 月 1-7 日) 的日补货总量和定价策略，我们使用 LSTM 时间序列预测模型预测未来一周的蔬菜各品类的定价，后结合蔬菜各品类的销售总量与成本加成定价的关系建立规划模型，以商超最大收益为目标进行求解。

（一）LSTM 模型预测未来一周的蔬菜各品类的定价

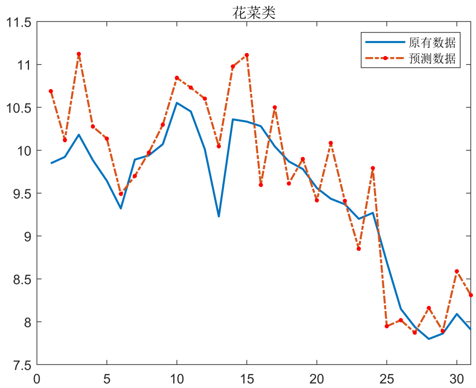
LSTM 长短期记忆神经网络是一种可以用于对时间序列进行分析和预测的神经网络，它很好地解决了一般递归神经网络的长期依赖问题。LSTM 长短期记忆神经网络具有十分强大的学习能力，十分适合用于解决间隔和延迟相对较长的时间序列.

我们使用蔬菜品类前 35 个月的批发价格数据作为模型的训练集，以蔬菜品类 2023 年最后一个月的批发价格数据作为模型的测试集。模型部分参数如下表所示。

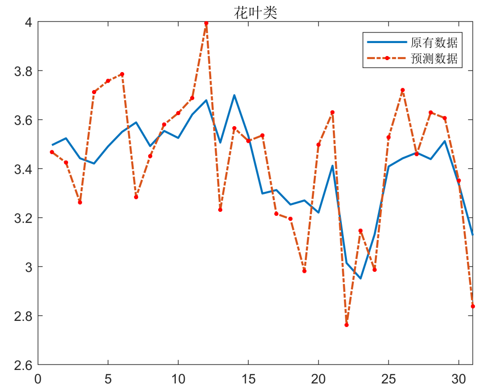
表 4: LSTM 网络部分参数

参数名	参数值
最大训练回合数	500
训练梯度阈值	1
初始学习率	0.01
学习率系数因子	0.15

经 MATLAB 编程求解，为验证模型性能，我们可视化地对比了最后一个月模型预测的批发单价与实际批发单价的差异，如图 9 所示（注：由于篇幅有限，完整的对比图将在附录中展示）



(a) 花菜类



(b) 花叶类

图 9: 部分蔬菜品类测试对比图

据图 9 可做分析：模型对蔬菜品类批发价格进行预测的结果与实际的批发价格较为接近，且二者整体趋势相似，表明模型的效果较好。使用已完成训练的模型对蔬菜各品类未来一周的批发价格进行预测，预测结果如表 5 所示。

表 5: 蔬菜各品类未来一周的批发价格预测值

日期	花菜类	花叶类	辣椒类	茄类	食用菌	水生根茎类
7 月 1 日	9.20	3.60	3.44	4.95	3.96	12.41
7 月 2 日	7.73	3.26	2.95	3.98	3.60	12.31
7 月 3 日	8.15	3.55	3.27	4.14	3.73	13.34
7 月 4 日	7.74	3.41	3.42	4.74	3.50	13.10
7 月 5 日	7.82	3.49	3.34	4.79	3.08	12.25
7 月 6 日	7.38	3.04	3.31	4.15	2.90	11.18
7 月 7 日	7.70	3.05	3.61	4.57	4.39	11.73

• 注：上述数据的单位均为元/千克

（二）补货策略与定价的确定

为了便于确定补货策略与定价，我们定义以下变量，具体如下表所示：

符号	符号说明
m_{it}	第 i 类蔬菜品类在第 t 天的补货量
n_{it}	第 i 类蔬菜品类在第 t 天的销售量
v_i	第 i 类蔬菜品类在历史数据中每天销售量的平均值
x_{it}	第 i 类蔬菜品类在第 t 天的成本加成定价
x_{imax}	第 i 类蔬菜品类历史成本加成定价的最大值
x_{imin}	第 i 类蔬菜品类历史成本加成定价的最小值
z_{it}	第 i 类蔬菜品类在第 t 天的批发价格
s_{it}	第 i 类蔬菜品类在第 t 天的损耗率

其中，$i\epsilon\{1,2\cdots6\}$，$t\epsilon\{1,2\cdots7\}$，视各蔬菜品类而定。依据上文论述 z_{it} 已使用 LSTM 预测出。

且依据假设 4，蔬菜各品类的销售量与成本加成定价有如下关系：

$$m_{it} \geq \frac{n_{it}}{1-s_{it}} \tag{14}$$

由上文可得蔬菜各品类销售量与成本加成定价有 $n_{it}=a_ix_{it}^2+b_ix_{it}+c_i$ 的关系，但考虑到 n_{it} 与 x_{it} 的拟合效果有偏差，且蔬菜各品类销售总量具有周期性规律，故我们对每天的销售总量进行线性加权处理并将新变量定义为 d_{it}

$$d_{it} = 0.5n_{it} + 0.5v_i \tag{15}$$

依据假设 5，蔬菜各品类每天的销售量有如下关系：

$$x_{imin} \le x_{it} \le x_{imax} \tag{16}$$

依据上几式可建立以商超收益 W 最大为目标的一般规划模型：

$$\text{目标函数：} Max\ W = \sum_{i=1}^{6} x_{it}d_{it} - z_{it}m_{it}$$

$$s.t. \begin{cases} d_{it} = 0.5n_{it} + 0.5v_i \\ n_{it} = a_i x_{it}^2 + b_i x_{it} + c_i \\ m_{it} \ge \dfrac{n_{it}}{1 - s_{it}} \\ x_{imin} \le x_{it} \le x_{imax} \\ i\epsilon\{1, 2\cdots 6\}, t\epsilon\{1, 2\cdots 7\} \end{cases} \tag{17}$$

为求解上述一般规划模型，我们使用粒子群优化算法 [4] 进行求解，粒子群优化算法的核心思想是利用群体中的个体对信息的共享使整个群体的运动在问题求解空间中产生从无序到有序的演化过程，从而获得问题的可行解，粒子群优化算法流程如图 10 所示。

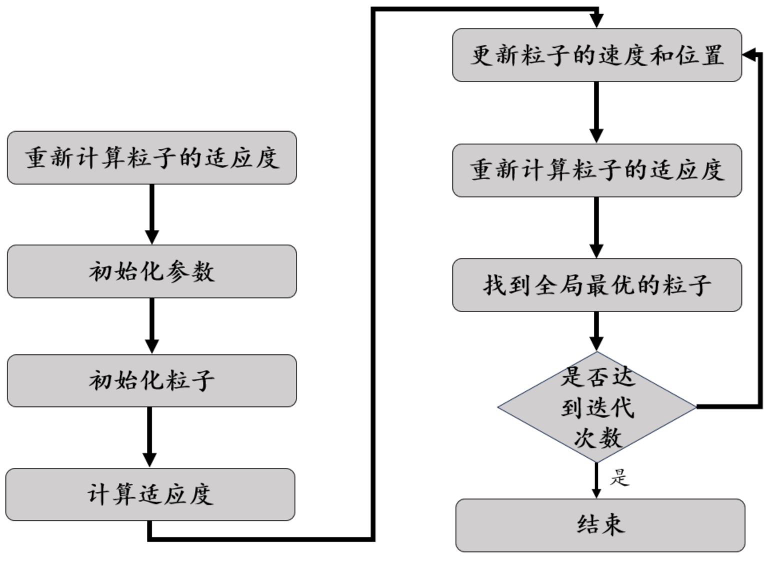


图 10: 粒子群算法实现步骤

经 MATLAB 编程进行多次实验，得到了粒子群优化算法较好的结果。

可得商超未来一周的最优日补货总量和成本加成定价分布如表 6 和表 7 所示。相应的日收益最大值因篇幅有限，将在附录中展示。

表 6: 蔬菜各品类未来一周的最优补货量

日期	花菜类	花叶类	辣椒类	茄类	食用菌	水生根茎类
7 月 1 日	12.27	171.6	99.07	30.10	57.57	17.05
7 月 2 日	11.39	116.14	74.29	24.87	37.61	17.59
7 月 3 日	8.43	84.06	75.35	22.15	41.31	10.84
7 月 4 日	14.90	133.40	70.35	8.81	40.57	15.74
7 月 5 日	16.10	131.01	68.97	15.75	54.08	15.04
7 月 6 日	24.40	135.71	89.52	11.56	48.53	22.28
7 月 7 日	29.70	138.10	87.10	25.96	41.89	20.31

- 注：上述数据的单位均为千克

表 7: 蔬菜各品类未来一周的最优定价

日期	花菜类	花叶类	辣椒类	茄类	食用菌	水生根茎类
7 月 1 日	15.41	5.66	6.64	8.13	6.44	19.09
7 月 2 日	14.20	5.20	6.21	6.74	5.96	19.34
7 月 3 日	15.70	6.60	7.20	7.10	6.63	21.19
7 月 4 日	12.82	5.75	7.58	8.45	5.85	20.73
7 月 5 日	14.68	6.21	8.26	9.07	5.58	21.9531
7 月 6 日	13.10	5.42	7.53	8.00	5.19	19.57
7 月 7 日	12.57	5.50	7.28	8.59	6.80	19.23

- 注：上述数据的单位均为元/千克

据表 6 和表 7 可做分析：为使商超未来一周日收益最大，花叶类蔬菜未来一周的补货量最多，花菜类未来一周的补货量最少。蔬菜各品类未来一周的定价与历史数据相比不会发生太大的变化。

6.4 问题三

6.4.1 建模前的准备

我们使用 Excel 软件对 2023 年 6 月 24 日至 2023 年 6 月 30 日的可售品种进行筛选，统计出蔬菜各品类中共有 48 个可售单品，因篇幅有限，完整信息放在附件中。

对可售单品按“花菜类，花叶类，辣椒类，茄类，食用菌，水生根茎类”顺序进行排列编号。其中，编号为 1-2 的单品为花菜类记为 i_1，3-19 的单品为花叶类记为 i_2，

20-29 的单品为辣椒类记为 i_3，30-34 的单品为茄类记为 i_4，35-42 的单品为食用菌记为 i_5，43-48 的单品为水生根茎类记为 i_6。

为了便于描述定义以下变量，如下表所示：

符号	符号说明
m_l	各可售单品在 2023 年 7 月 1 日的补货量
n_l	各可售单品在 2023 年 7 月 1 日的销售量
v_l	各可售单品在 2023 年最后 1 个月的平均值
x_l	各可售单品在 2023 年 7 月 1 日的成本加成定价
z_l	各可售单品在 2023 年 7 月 1 日的批发价格
s_l	各可售单品在 2023 年 7 月 1 日的损耗率
$\bar{u}_i^j$	蔬菜各品类历史数据销售量的平均值
q_l	各可售单品在 2023 年 7 月 1 日是否补货的二值变量

其中

$$q_l = \begin{cases} 1, m_l \neq 0 \\ 0, m_l = 0 \end{cases} \tag{18}$$

对于 $\bar{u}_i^j$ 我们计算的数值如下表所示。

表 8: 蔬菜各品类历史销售平均值 ($\bar{u}_i^j$)

名称	花菜类	花叶类	辣椒类	茄类	食用菌	水生根茎类
$\bar{u}_i^j$	34.39	160.44	60.03	29.99	43.02	15.22

6.4.2 问题三模型的建立与求解

通过对 48 个可售单品的销售总量与定价关系的分析（同问题二），得到部分蔬菜单品拟合效果图，如图 11 所示。采用形如 $y = ax^2 + bx + c$ 的二次函数拟合，可得拟合系数，如下表 9 所示。

考虑到 n_l 与 x_l 的拟合具有偏差（同问题二），且各单品销售总量具有周期性规律，故我们对销售总量进行线性加权处理并将新变量定义为 d_l

$$d_l = 0.5n_l + 0.5v_l \tag{19}$$

依据假设 4，可得各可售单品的补货量与销售量有如下关系：

表 9: 蔬菜各单品的拟合曲线系数表

单品名称	a	b	c
西兰花	-0.4468	6.0297	27.5476
苋菜	0.4420	-27.4844	323.2484
竹叶菜	-0.0310	-1.8916	103.8160

- 注：篇幅有限，完整的拟合系数数据将在附件中展示

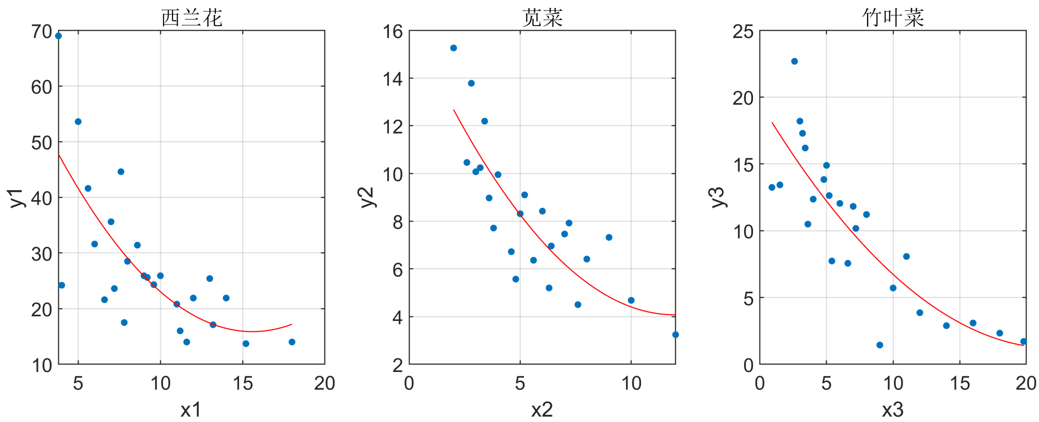


图 11: 蔬菜部分单品拟合图

$$m_l \geq \frac{n_l}{1-s_l} \tag{20}$$

由于题目规定各单品订购量需要满足最小陈列量 2.5 千克，可得：

$$m_l \geq 2.5 \tag{21}$$

因蔬菜类商品的销售空间有限，要求可售单品总数控制在 27-33 个，可得：

$$27 \leq \sum_{l=1}^{48} q_l \leq 33 \tag{22}$$

为满足市场对各类蔬菜商品的需求，且考虑到销售量会发生波动，我们定义波动率为 30%, 可得：

$$0.7\bar{u}_i^j \leq \sum_{l\epsilon i_j} n_l \leq 1.3\bar{u}_i^j \tag{23}$$

建立以商超收益 W 最大为目标的 0-1 规划模型：

$$\text{目标函数：} Max\ W = \sum_{l=1}^{48} q_l x_l d_l - z_l m_l$$

$$
s.t.\begin{cases}
d_l = 0.5n_l + 0.5v_l \\
m_l \geq \dfrac{n_l}{1-s_l} \\
m_l \geq 2.5 \\
27 \leq \sum\limits_{l=1}^{48} q_l \\
0.7\bar{u}_i^j \leq \sum\limits_{l\epsilon i_j} n_l \leq 1.3\bar{u}_i^j \\
27 \leq \sum\limits_{l=1}^{48} q_l \leq 33 \\
l\epsilon\{1,2\cdots 48\}, j\epsilon\{1,2\cdots 6\}
\end{cases} \tag{24}
$$

表 10: 各蔬菜单品未来一周的最优补货量与定价

名称	补货量	定价	名称	补货量	定价
西兰花	15.16	11.95	奶白菜	10.67	4.42
枝江青梗散花	10.18	13.42	小青菜（1）	6.5	5.2
苋菜	9.36	3.72	云南生菜（份）	35.64	4.60
竹叶菜	14.83	5.20	云南油麦菜（份）	19.90	4.39
木耳菜	5.84	6.22	菠菜（份）	15.23	5.31
娃娃菜	6.60	9.68	上海青	3.80	7.85
长线茄	5.50	12.00	螺丝椒	7.34	10.17
芜湖青椒（1）	12.57	5.50	小米椒（份）	14.62	4.87
小皱皮（份）	5.14	2.71	螺丝椒（份）	11.25	4.51
姜蒜小米椒组合	7.67	4.58	紫茄子（2）	14.46	6.00
红薯尖	5.75	6.42	西峡花菇（1）	5.69	23.33
双孢菇（盒）	14.30	5.50	金针菇（盒）	19.20	2.00
海鲜菇（包）	12.05	2.90	净藕	6.13	15.30
高瓜（1）	3.09	15.57	高瓜（2）	2.5	17.28
洪湖藕带	5.06	25.25	菱角	2.50	14.24
青茄子（1）	2.50	5.80			

- 注：上述数据的单位均为元/千克

经 MATLAB 采用粒子群算法求解（同问题二），可售单品数为 31 个满足题意。各可售单品具体的补货量和定价如表 10 所示。求得的蔬菜各品类满足率分别为：0.7366

（花菜类），0.8359（花叶类），0.9539（辣椒类），0.7490（茄类），1.1908（食用菌），1.2664（水生根茎类）。

6.5 问题四

为了更好地制定蔬菜商品的补货和定价决策，商超还需要采集以下相关数据：

（一）各蔬菜单品 t 时刻的新鲜度 $\theta(t)$

大量研究表明消费者在购买生鲜食品时，新鲜度越高，购买的意愿就越强烈，生鲜商品的需求量就越高 [5]。设 $y(t)$ 为蔬菜商品单位周期内的售卖价格，设 $D(y,t)$ 为 t 时刻蔬菜商品的需求量，由此可得蔬菜商品的需求函数为：

$$D(y,t) = a\theta(t) - by(t) \tag{25}$$

其 a 为市场基本需求，b 为价格敏感系数，即单位价格的变动产生的需求变动。

倘若商超能采集各蔬菜单品 t 时刻的新鲜度 $\theta(t)$，将对于我们解决上述问题具有很大帮助。对于问题一，我们可以分析不同蔬菜各品类及单品的 t 时刻的新鲜度差异，进一步分析蔬菜各品类及单品的相互关系。对于问题二和问题三，可以进一步完善进货策略，在原问题的基础上，进一步分析销售总量与需求量的关系。以下给出商超采集蔬菜单品新鲜度的部分建议：

（1）定时检查对蔬菜单品进行外观检查和嗅觉检查，例如颜色、外表质地、表面光泽和蔬菜单品的气味。

（2）定时使用 PH 值检测方法，生鲜食品的 PH 值会随着时间而变化。变化较大的 PH 值可能表明食品已不新鲜。

（3）定时使用专业的化学试剂检测蔬菜单品的化学成分变化，例如，测试氨基酸含量和糖含量。

（二）竞争商超蔬菜各单品销售量、销售单价、销售策略等信息

了解竞争对手的蔬菜商品定价和促销策略。可以帮助商超制定竞争性策略，并更好地了解市场格局，以便制定补货策略，对我们解决上述问题有很大帮助。对于问题二和问题三，可以进一步完善进货策略，在原问题的基础上，进一步分析销售总量与需求量的关系。对于此种信息的采集，问卷调查可以很好的适合这种方式，以下给出商超设计问卷的思路及部分问题的方向。

（1）设计问题方向：蔬菜新鲜度；蔬菜需求量；蔬菜价格；营销策略；优惠政策等。

（2）问卷设计思路：设计的问卷应分为两部分：第一部分为参与调查的顾客基本信息收集，该部分的信息主要用于了解参与调查客户的个人基本信息如性别、年龄阶段、收入水平、学历和工作状态，用于统计分析顾客群像特征。第二部分问题评价部分，由参与的顾客更加自身购物经历进行评价。为方便后期统计数据，第二部分问卷设计可采用李克特量表，数据调查为封闭形式，参与问卷调查的客户按照自己的意愿从已经拟定的几个答案当中勾选一个认为最符合的答案出来。

6.5.1 模型的优点

（1）问题一对蔬菜品类进行分析时，考虑到销售品类仅有六类，样本量较少，且通过品类间散点图初步定性分析时未发现明显规律，而采用的灰色关联分析方法仅是通过曲线形状的相似度来判断变量之间的关系，避免了传统数理统计方法需要大量数据，且服从典型的概率分布的缺点，也不会出现定量分析结果与定性分析结果不一致的问题。

（2）问题二对各品类蔬菜分析其销售总量与定价的关系时，初始拟合优度均小于 0.1，通过对各类蔬菜在相同定价下的平均销量代替销量进行再拟合，拟合优度均值达到 0.67，并考虑到销量的周期性变化规律，与最后一个月的销售量均值进行线性加权处理，从而进一步提升模型的准确性。

（3）问题三将各蔬菜品类历史数据销售量均值的波动范围（0.7，1.3）作为 2023 年 7 月 1 日的商品需求区间，通过新增各品类中单品补货量总值处于商品需求区间作为约束条件，从而将看似为双目标规划问题进行化简。

（4）问题四通过查找文献，收集各蔬菜单品新鲜度与时间变化关系 θt，并类比于微观经济学中的需求函数，建立蔬菜商品需求函数，从而使蔬菜商品补货与定价决策更加准确。

6.5.2 模型的缺点

问题三中“在尽量满足市场对各品类蔬菜商品需求的前提下”所产生的约束条件很难定量描述。我们只有 27-33 种单品要满足共有 251 种单品蔬菜商品的需求。若过分的满足要求可能会出现约束过强而无法得到最优解，且会降低商超的收益，而削弱要求又恐不符合题议。本模型综合考虑实际情况中商超对收益更为看重，削弱对于“在尽量满足市场对各品类蔬菜商品需求的前提下”的要求。

6.5.3 模型的推广

本模型不仅可以很好的适合蔬菜类商品的自动定价与补货决策，对于类似于定价和补货的相关问题都可以处理，譬如快递运输系统中运费与运货量的相关问题等。

参考文献

[1] 毛燕. 四分位法和迭代法对数据分散的能力验证检测数据统计分析结果的比较 [J]. 冶金分析,2016,36(05):76-81.DOI:10.13228/j.boyuan.issn1000-7571.009841.

[2] 易德生, 郭萍. 灰色理论与方法—提要・题解・程序・应用 [M]. 北京: 石油工业出版社,1992.

[3] 司守奎. 数学模型（第三版）[M], 北京: 高等教育出版社,2003.

[4] 毛开富, 包广清, 徐驰. 基于非对称学习因子调节的粒子群优化算法 [J]. 计算机工程,2010,36(19):182-184.

[5] 曾敏敏. 基于时间情境 A 生鲜社区超市的动态定价策略研究 [D]. 西南财经大学,2021.DOI:10.27412/d.cnki.gxncu.2021.002858.

附录 A 支撑文件列表

```
1 data1.mat
2 data2.mat
3 dj.mat
4 fun.m
5 fun_2.m
6 hc_hy.mat
7 hc_hy_2.mat
8 last.month.mat
9 pfjg.mat
10 Q1_hemidu.m
11 Q1_huiseguanlian.m
12 Q1_第一问各品类销售总量.py
13 Q1_剔除销售单价得异常值.py
14 Q2_data.mat
15 Q2_data_new.mat
16 Q2_guihua.m
17 Q2_hc.m
18 Q2_hstm.m
19 Q2_nihe.m
20 Q3_data.m
21 Q3_guihua.m
22 Q3_nihe.m
23 Q3_sjcl.m
24 shl_data.m
25 xishu.mat
```

附录 B 问题二补充图

附录 C 代码

```
fun.m
介绍：问题二粒子群算法适应度函数

function y=fun(x,xs,shl_temp,pfjg_temp,mean_each)
   y=0;
   for i=1:6
      n=xs(1)*x(i).*x(i)+xs(2)*x(i)+xs(3);
      n=0.5*mean_each+0.5*n;
      m=n/(1-shl_temp);
      y=y+x(i)*n-pfjg_temp(i)*m;
   end
```

	洪湖莲藕	高瓜	净藕	青茄子	紫茄子	圆茄子	长线茄
洪湖莲藕	1.000	-0.147	0.189	-0.462	-0.154	-0.280	0.007
高瓜	-0.147	1.000	**-0.601***	0.434	0.098	0.084	0.197
净藕	0.189	**-0.601***	1.000	**-0.671***	**-0.734****	-0.182	-0.331
青茄子	-0.462	0.434	**-0.671***	1.000	0.406	0.140	0.000
紫茄子	-0.154	0.098	**-0.734****	0.406	1.000	0.427	0.338
圆茄子	-0.280	0.084	-0.182	0.140	0.427	1.000	-0.141
长线茄	0.007	0.197	-0.331	0.000	0.338	-0.141	1.000

注：**.在 0.01 级别相关性显著；*.在 0.05 级别相关性显著

(a) 补充 1

	洪湖莲藕	高瓜	净藕	白玉菇	海鲜菇	姬菇	金针菇
洪湖莲藕	1.000	-0.147	0.189	0.427	0.497	0.014	**0.678***
高瓜	-0.147	1.000	**-0.601***	-0.056	-0.483	0.155	-0.189
净藕	0.189	**-0.601***	1.000	0.091	0.559	0.359	0.378
白玉菇	0.427	-0.056	0.091	1.000	0.552	0.380	0.503
海鲜菇	0.497	-0.483	0.559	0.552	1.000	0.472	0.280
姬菇	0.014	0.155	0.359	0.380	0.472	1.000	0.070
金针菇	**0.678***	-0.189	0.378	0.503	0.280	0.070	1.000

注：*.在 0.05 级别相关性显著

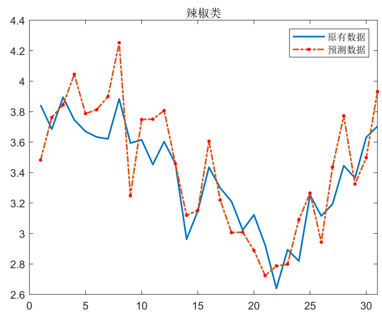
(b) 补充 2

	青茄子	紫茄子	圆茄子	长线茄	白玉菇	海鲜菇	姬菇	金针菇
青茄子	1.000	0.406	0.140	0.000	0.042	-0.287	0.211	-0.587*
紫茄子	0.406	1.000	0.427	0.338	0.021	-0.350	-0.387	-0.308
圆茄子	0.140	0.427	1.000	-0.141	0.371	-0.315	0.056	0.196
长线茄	0.000	0.338	-0.141	1.000	-0.380	-0.303	-0.447	-0.380
白玉菇	0.042	0.021	0.371	-0.380	1.000	0.552	0.380	0.503
海鲜菇	-0.287	-0.350	-0.315	-0.303	0.552	1.000	0.472	0.280
姬菇	0.211	-0.387	0.056	-0.447	0.380	0.472	1.000	0.007
金针菇	-0.587*	-0.308	0.196	-0.380	0.503	0.280	0.070	1.000

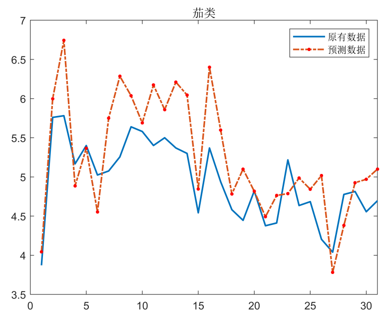
注：*.在 0.05 级别相关性显著

(c) 补充 3

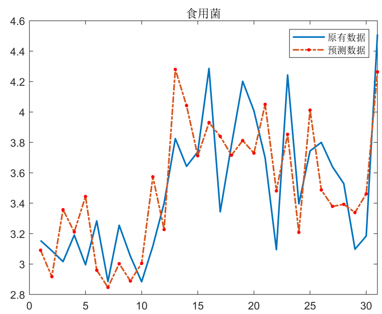
各蔬菜单品斯皮尔曼系数补充图



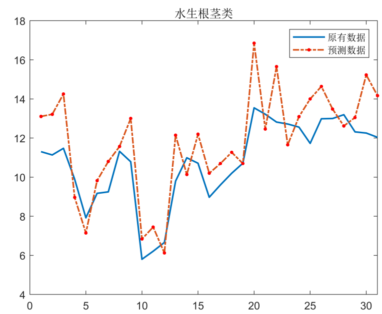
(a) 辣椒类



(b) 茄类



(c) 食用菌



(d) 水生根茎类

各蔬菜单品斯皮尔曼系数补充图

蔬菜各品类未来一周的日最高收益

日期	花菜类	花叶类	辣椒类	茄类	食用菌	水生根茎类	总收益
7 月 **1** 日	76.21	351.76	317.13	95.80	142.89	113.92	1097.73
7 月 **2** 日	73.76	225.48	241.98	68.67	88.79	123.69	822.40
7 月 **3** 日	63.65	255.78	296.04	65.50	119.60	85.07	885.65
7 月 **4** 日	75.79	311.91	293.02	32.68	95.32	120.06	928.78
7 月 **5** 日	110.93	356.10	338.91	67.32	135.30	145.81	1154.40
7 月 **6** 日	141.90	322.34	377.35	44.53	110.67	187.20	1184.00
7 月 **7** 日	144.71	338.68	320.02	104.31	100.92	152.29	1160.95

• 注：上述数据的单位均为元

```
    y=-y;%转化为求最小值
end
```

```
fun_2.m
介绍：问题三粒子群算法适应度函数

function y=fun_2(x,xs,shl_temp,pfjg_temp,mean_each,mzl)
    y=0;
    temp=zeros(1,6);%创建一个矩阵，记录各品类销量
    %计算x中0的个数
    count_zeros = sum(x== 0);
    %计算品类满足率
    temp(1)=sum(x(1:2));
    temp(2)=sum(x(3:19));
    temp(3)=sum(x(20:29));
    temp(4)=sum(x(30:34));
    temp(5)=sum(x(35:42));
    temp(6)=sum(x(43:48));
    %计算出上线波动30%的范围
    mzl_upper=mzl*1.3;
    mzl_upper=mzl*0.7;
    for i=1:48
        n=xs(i,1)*x(i).*x(i)+xs(i,2)*x(i)+xs(i,3);
        n=0.5*mean_each(i)+0.5*n;
        m=n/(1-shl_temp);
        y=y+x(i)*n-pfjg_temp(i)*m;
    end
    %增加两个惩罚项
    if count_zeros<=15 || count_zeros>=21%当单品数不满足27-33时，增加一个惩罚项
            y=y+100000000;
    end
    index=(temp>=mzl_upper)&(temp<=mzl_upper);
```

```
    if all(index)
        y=y;
    else
        y=y+100000000;
    end
    y=-y;%转化为求最小值
end
```

```
Q1_hemidu.m
介绍：问题一绘制核密度函数图像

clc;clear;
load hc_hy_2.mat
% 假设 data 是你的数据向量
close all
rows=size(data,1);%行数
cols=size(data,2);%列数
label={'青梗散花','西兰花','菠菜','菜心','黄白菜','奶白菜','上海青','云南生菜','云南油麦菜','
红薯尖'};
figure('Units','normalized','Position',[0.2,0.2,0.6,0.5])
for k = 1:cols
    subplot(2,5,k)
    % 绘制直方图
    h1 = histogram(data(:,k), 20, 'Normalization', 'pdf', 'FaceAlpha', 0.5);
    % 计算核密度估计
    [f, xi] = ksdensity(data(:,k));
    % 创建第二轴
    ax1 = gca;
    ax2 = axes('Position', get(ax1, 'Position'), 'XAxisLocation', 'top', 'YAxisLocation',
        'right', 'Color', 'none', 'XColor', 'k', 'YColor', 'k');
    % 在第二轴上绘制核密度估计
    line(xi, f, 'Parent', ax2, 'Color', [0.9290 0.6940 0.1250],'LineWidth', 1.2);
    set(ax2, 'XTick', [],"YTick",[]);
    set(ax1, "YTick",[]);
    linkaxes([ax1 ax2], 'x');
    ax2.XLim = ax1.XLim;
    box on
    title(label{k}, 'FontSize', 10);
end
```

```
Q1_huiseguanlian.m
介绍：问题一计算灰色关联度函数

%灰色关联用于求各品类之间的关联度
clear;clc;
load data2.mat%导入各品类的数据
%数据预处理部分
rows=size(data,1); %求行数
data_average = mean(data);%求列向量平均值
std_data = data ./ repmat(data_average,rows,1);%进行标准化处理
```

```
%计算灰色关联系数
mo_seq = std_data(:,6); %将第六列确定为母序列(每次运行需要自己更换不同列作为母列)
ch_seq = std_data(:,1:5); % 将第一列到第五列品类信息确定为子序列
cols=size(ch_seq,2);%求列数
im_matrix = abs(ch_seq - repmat(mo_seq,1,cols));%得|X0(k)-Xi(k)|的整个矩阵
rat_num = 0.5; %将res_rat取默认值0.5
max_data = max(max(im_matrix));
min_data = min(min(im_matrix));
gl_data = (min_data+rat_num*max_data) ./ (im_matrix + rat_num*max_data); % 代入公式计算
result=mean(gl_data)%平均求和得出灰色关联度
```

```
Q1_第一问各品类销售总量.py
介绍：问题一计算各品类销售总量得函数

import pandas as pd
#读取文件
df = pd.read_excel(r"C:\Users\第二题第1小题.xlsx",sheet_name='合并损耗率')
# 根据'单品名称'分组，然后对销量进行求和，对销售日期使用nunique计算出现天数
agg_df = df.groupby(['单品名称', '分类名称']).agg({
    '销量': 'sum',
    '销售日期': 'nunique'
}).reset_index()
agg_df.rename(columns={'销售日期': '出现天数'}, inplace=True)
# 将计算后的结果写入文件
agg_df.to_excel('各品类销售总量.xlsx', index=False, engine='openpyxl')
```

```
Q1_剔除销售单价的异常值.py
介绍：数据预处理使用四分位法剔除销售单价异常值的函数

import pandas as pd
# 读取Excel文件
excel_data = pd.read_excel('out.xlsx')
# 计算上四分位数
Q1 = excel_data.groupby('dpbm')['xsdj'].quantile(0.25)
#计算下四分位数
Q3 = excel_data.groupby('dpbm')['xsdj'].quantile(0.75)
#由I四分位数计算IQR
IQR = Q3 - Q1
# 定义一个函数来标记异常值
def remove_func(row):
    IQR_value = IQR.get(row['dpbm'])
    if IQR_value:
        return (row['xsdj'] < (Q1[row['dpbm']] - 1.5 * IQR_value)) | (row['xsdj'] >
            (Q3[row['dpbm']] + 1.5 * IQR_value))
    else:
        return False
# 调用函数标记异常值
excel_data['is_outlier'] = excel_data.apply(remove_func, axis=1)
# 统计被标记的异常值的个数
outlier_count = excel_data['is_outlier'].sum()
```

```
print('总共有 {} 个异常值'.format(outlier_count))
# 剔除异常值
excel_data = excel_data[~excel_data['is_outlier']]
# 将剔除了异常值之后的数据存储到Excel中
excel_data.to_excel('cleaned.xlsx', index=False)
```

```
Q2_guihua.m
介绍：问题二使用粒子群算法求解目标函数

clc;clear;
load xishu.mat
load dj.mat
load pfjg.mat
load last_month
shl=[10.75154275,12.5249767,8.015658326,6.350473961,6.795504752,8.097707213];%各品类损耗率
labels=["花菜类","花叶类","辣椒类","茄类","食用菌","水生根茎类"];
%求出各品类的成本加成定价的最大值最小值
min_dj=min(data);
max_dj=max(data);
%利用最后一个月求销量均值
mean_each=mean(last_month);
%%设置粒子群算法的参数
n = 30; % 粒子数量
var_num = 6; % 变量个数
a = 2.05;
b = 2.05;
C = a+b;
%定义收缩因子
fai = 2/abs((2-C-sqrt(C^2-4*C)));
w = 0.9;
iter_num = 3000; % 迭代的次数
v_max =2; %定义粒子的最大速度
%定义上下界
x_lower =min_dj(1)*ones(1,var_num);
x_upper =max_dj(1)*ones(1,var_num);
%将粒子的位置随机初始化
x = zeros(n,var_num);
for i = 1: var_num
   x(:,i) = x_lower(i) + (x_upper(i)-x_lower(i))*rand(n,1);
end
%随机初始化粒子的速度
v = -v_max + 2*v_max .* rand(n,var_num);
 %将粒子的适应度全部初始化为0
fitness = zeros(n,1);
%循环，调用fun函数计算粒子的适应度
for i = 1:n
   fitness(i) = fun(x(i,:),xishu(1,:),shl(1),pfjg(:,1),mean_each(1));
end
%初始化记录最佳位置为初始值x
best_pos = x;
%找到最优适应度的位置
```

```
min_index = find(fitness == min(fitness), 1);
%记录全局最优的位置
all_best = x(min_index,:);
% 初始化每次迭代得到的最佳的适应度
best_fit = ones(iter_num,1);
for d = 1:iter_num
   for i = 1:n
      v(i,:) = fai * (w*v(i,:) + a*rand(1)*(best_pos(i,:) - x(i,:)) + b*rand(1)*(all_best -
          x(i,:))); % 更新第i个粒子的速度
      x(i,:) = x(i,:) + v(i,:);
      %判断粒子的位置是否越界,如果越界则将粒子速度设置为边界
      for j = 1: var_num
         if x(i,j) < x_lower(j)
            x(i,j) = x_lower(j);
         elseif x(i,j) > x_upper(j)
            x(i,j) = x_upper(j);
         end
      end
      fitness(i) = fun(x(i,:),xishu(1,:),shl(1),pfjg(:,1),mean_each(1));
      %判断新的粒子是否有适应度优于当前最优的,如果有则替换当前最优的记录
      if fitness(i) < fun(best_pos(i,:),xishu(1,:),shl(1),pfjg(:,1),mean_each(1))
        best_pos(i,:) = x(i,:);
      end
      %判断新的粒子是否有适应度优于历史最优的,如果有则替换历史最优的记录
      if  fitness(i) < fun(all_best,xishu(1,:),shl(1),pfjg(:,1),mean_each(1))
         all_best = best_pos(i,:);
      end
   end
end
disp('最优解为: '); disp(all_best)
```

```
Q2_lstm.m
介绍: 问题二使用lstm预测批发价格

%%做一个最后一个月的对比图
clc;clear;
load Q2_hc.mat%%需要手动更改不同品类的数据,当前为花菜类数据
pred_result=[];
n=length(data);
train_data = data(1:n-29); %用前35个月做训练集
Validate_data = data(n-30:n); %最后一个月的数据用来作为验证集
dataTrain=data(:,1);
%数据预处理
data_mean = mean(dataTrain); %均值
data_std = std(dataTrain); %标准差
standardized_train_data = (dataTrain - data_mean) / data_std;%标准化
% 训练数据
train_x = standardized_train_data(1:end-1)';
train_y = standardized_train_data(2:end)';
%创建LSTM回归网络,并设置较优参数
features_num = 1;
```

```
response_num = 1;
hidden_unit_num = 300;
layers = [sequenceInputLayer(features_num)
            lstmLayer(hidden_unit_num)
            fullyConnectedLayer(response_num)
            regressionLayer];
  % 设置网络参数
options = trainingOptions('adam', ...
       'MaxEpochs', 500, ...
       'GradientThreshold', 1, ...
       'InitialLearnRate', 0.01, ...
       'LearnRateSchedule', 'piecewise', ...
       'LearnRateDropPeriod', 200, ...
       'LearnRateDropFactor', 0.15, ...
       'Verbose', 0, ...
       'Plots', 'training-progress');
% 训练
net = trainNetwork(train_x,train_y,layers,options);
net = predictAndUpdateState(net, train_x);
%进行预测
[net, predict_y] = predictAndUpdateState(net, train_y(end));
%用训练的数据去预测最后一个月的数据以及7.1-7.7的数据
pred_num= 38; %预测多少天的
for i = 2:pred_num
   [net, predict_y(:, i)] = predictAndUpdateState(net, predict_y(:, i-1),
       'ExecutionEnvironment', 'gpu');
end
% 去标准化
predict_y =data_std*predict_y + data_mean;
figure(2)
plot(data(end-30:end),'LineWidth',1.5,'MarkerSize',3);
hold on
plot(1:38,predict_y,"LineWidth",1.5);
```

```
Q2_nihe.m
介绍：问题二使用最小二乘拟合函数

clc;clear;
load Q2_data_new.mat
labels={'花菜类','花叶类','辣椒类','茄类','食用菌','水生根茎类'};
% 创建2x3的子图
figure;
% fit_data=fit_data(end-271:end,:);
fit_data=data;
xishu=zeros(6,3);
for i = 1:6
   %获取拟合的数据
   fit_x = fit_data(:, 2*i-1);
   fit_y = fit_data(:, 2*i);
   %对数据进行多项式拟合
   p_fit = polyfit(fit_x, fit_y, 2);
```

```
    xishu(i,:)=p_fit;
    %计算拟合出来的y值
    x_fit = linspace(min(fit_x), max(fit_x), 1000);
    y_fit = polyval(p_fit, x_fit);
    %在子图上绘制拟合的效果
    subplot(2, 3, i);
    scatter(fit_x, fit_y,10,'filled');
    hold on;
    plot(x_fit, y_fit, 'r-'); % 拟合曲线用红色线表示
    box on
    title(labels(i));
    xlabel(['x', num2str(i)]);
    ylabel(['y', num2str(i)]);
    grid on;
end
```

```
Q3_guihua.m
介绍：问题三使用基于0-1整数规划的粒子群算法求解目标函数

clc;clear;
load xishu.mat
load shl_data.mat
load pfjg.mat
load last_month.mat
%求出各品类的成本加成定价的最大值最小值
min_dj=min(data);
max_dj=max(data);
%利用最后一个月求销量均值
mean_each=mean(last_month);
%%设置粒子群算法的参数
n = 50; % 粒子数量
var_num = 48; %对48个单品进行0-1整数规划的粒子群算法求解
matrix = randi([0, 1], 1, 48);%创建一个1行48列的0-1变量矩阵
a = 1.05;
b = 1.05;
C = a+b;
%定义收缩因子
fai = 2/abs((2-C-sqrt(C^2-4*C)));
w = 0.9;
iter_num = 3000; % 迭代的次数
v_max =6; %定义粒子的最大速度
%定义上下界
x_lower =min_dj.*ones(1,var_num);
x_upper =max_dj.*ones(1,var_num);
%将粒子的位置随机初始化
x = zeros(n,var_num);
for i = 1: var_num
    matrix = randi([0, 1], 1, 48);%创建一个1行48列的0-1变量矩阵
    x(:,i) = x_lower(i) + (x_upper(i)-x_lower(i))*rand(n,1);
    x(:,i)=x(:,i).*matrix;%对变量乘0-1变量矩阵
end
```

```
%随机初始化粒子的速度
v = -v_max + 2*v_max .* rand(n,var_num);
 %将粒子的适应度全部初始化为0
fitness = zeros(n,1);
%循环，调用fun函数计算粒子的适应度
for i = 1:n
   fitness(i) = fun_2(x(i,:),xishu(1,:),shl(1),pfjg(:,1),mean_each(1),mzl);
end
%初始化记录最佳位置为初始值x
best_pos = x;
%找到最优适应度的位置
min_index = find(fitness == min(fitness), 1);
%记录全局最优的位置
all_best = x(min_index,:);
% 初始化每次迭代得到的最佳的适应度
best_fit = ones(iter_num,1);
for d = 1:iter_num
   for i = 1:n
      v(i,:) = fai * (w*v(i,:) + a*rand(1)*(best_pos(i,:) - x(i,:)) + b*rand(1)*(all_best -
          x(i,:))); % 更新第i个粒子的速度
      x(i,:) = x(i,:) + v(i,:);
      %判断粒子的位置是否越界，如果越界则将粒子速度设置为边界
      for j = 1: var_num
         if x(i,j) < x_lower(j)
            x(i,j) = x_lower(j);
         elseif x(i,j) > x_upper(j)
            x(i,j) = x_upper(j);
         end
      end
      fitness(i) = fun_2(x(i,:),xishu(1,:),shl(1),pfjg(:,1),mean_each(1),mzl);
      %判断新的粒子是否有适应度优于当前最优的，如果有则替换当前最优的记录
      if fitness(i) < fun_2(best_pos(i,:),xishu(1,:),shl(1),pfjg(:,1),mean_each(1),mzl)
        best_pos(i,:) = x(i,:);
      end
      %判断新的粒子是否有适应度优于历史最优的，如果有则替换历史最优的记录
      if  fitness(i) < fun_2(all_best,xishu(1,:),shl(1),pfjg(:,1),mean_each(1),mzl)
         all_best = best_pos(i,:);
      end
   end
end
disp('最优解为：'); disp(all_best)
```

```
Q3_sjcl.m
介绍：问题三数据处理

%对相同的x对应的y取平均值
data=data(:,1:2);
%获取数据的列数和行数
[row_num, col_num] = size(data);
%使用一个for循环处理每两列数据
result = []; % 用于存储处理后的结果
```

```
for col = 1:2:col_num
    x = data(:,col);
    y = data(:,col+1);
    % 对相同的x值求y的均值
    x_unique = unique(x);
    y_mean = zeros(length(x_unique), 1);
    for i = 1:length(x_unique)
        y_mean(i) = mean(y(x == x_unique(i)));
    end
    % 将结果合并到result数组
    result = [result, x_unique, y_mean];
end
```