

目 录 基于动态规划的蔬菜类商品自动定价与补货决策分析

摘 要:

生鲜超市中,蔬菜商品因保鲜期短,交易时信息未知等多种原因,使得补货量与定价决策尤为关键。为保障生鲜商超能获得最大收益,本文针对生鲜商超补货量和定价策略的制定问题,通过分析蔬菜销售明细、损耗率、批发价格的历史相关信息,基于收益最大的目标建立有条件约束的动态规划模型。

本题数据量较大,首先对蔬菜各品类及单品销售量的数据进行预处理,删除异常值,补充缺失值,对清洗后的数据集成处理。

针对问题一,对蔬菜品类和单品,分别绘制月销售量折线图,箱线图等,通过建立小波分析模型得出各品类的月销售量波动性**周期为 7**,直观化蔬菜单品数据分布总体情况。再进行关联分析:针对蔬菜品类,以 **KS 检验**确定数据符合正态分布,计算两两之间的皮尔逊相关系数,得**最强相关系数为 0.76**;针对单品,进行聚类分析,将不同单品类蔬菜月销售量划分为热销、畅销、滞销类蔬菜。

针对问题二:首先根据商品信息将有关数据分类,建立**多元线性拟合模型**构建各品类蔬菜销售总量和成本加成定价的**线性关系**,结果详见附录 2.4。“收益最大”方案的制定属于**有条件的规划问题**。本文提取批发价的变化特征,由历史数据构建**时间序列预测模型**分析其变化规律,进而得到预测值。在此基础上,以商超收益最大为目标函数,各品类蔬菜日补货总量和定价为决策变量,综合考虑批发价的季节性变化,以进货量,批发价,损耗率,定价与销售量的关为约束条件,建立**动态规划模型**。求解该模型,最终得到不同品类蔬菜的日补货总量和定价,结果详见附录 2.5。

针对问题三:对单品的补货计划,首先筛选出 49 个可售品种,随后对其建立**单品利润排序模型**,选取其中利润贡献率最大的前 33 个蔬菜单品,结果详见附录 3.4。对所选出的单品,通过对历史批发价数据**季节性分解**,预测出 7 月 1 日的批发价信息。在问题二模型的基础上,仍以商超收益最大为目标函数,增添单品最小陈列量等约束条件,并综合考虑单品间销量关联结果,进一步细化决策变量为所选取的 33 个蔬菜单品的日补货量和定价,来建立规模更大的动态规划模型,求解得 7 月 1 日蔬菜单品补货计划,结果详见附录 3.6。

针对问题四:对于这类开放探索性题目,结合题目背景、文献参考和问题假设条件可以综合总结出最大化商超利益的更多信息需求。定价决策方面:做好周期性市场调研活动把握市场消费潮流变动,对消费群体进行用户画像提高其消费粘性,周期性评价自身运营情况以提升服务水平;补货策略方面,主要以商品本身为中心进行发散性信息收集:补充收集库存量、供应商相关信息(含交货质量、交货准点率等)、及时采集外部数据诸如气候条件变化、市场消费数据、销售渠道效果等数据,以及时调整进货策略。

关键词:自动定价 补货决策 多元线性回归 动态规划

目 录

目 录	2
一、问题重述	4
1.1 问题背景	4
1.2 问题的数据条件	4
1.3 问题的提出	4
二、问题分析	4
2.1 问题总分析	4
2.2 具体问题分析	5
2.2.1 问题一的分析	5
2.2.2 问题二的分析	5
2.2.3 问题三的分析	5
2.2.4 问题四的分析	5
三、模型假设	6
四、符号说明	6
五、数据预处理	7
5.1 数据预处理	7
六、问题一的模型建立与求解	9
6.1 不同品类销售量相互关系的探索	9
6.1.1 建模准备——数据的相关性检验分析	9
6.1.2 基于 Pearson 系数判断相关性模型的建立	10
6.2 不同品类月销售量分布规律的探索	13
6.2.1 建模准备	13
6.2.2 基于小波变换讨论每个蔬菜品类不同月份销售量的周期性变化规律	14
6.3 各蔬菜品类内不同单品月销量的分布规律	16
七、问题二求解	19
7.1 各蔬菜品类的销售总量与成本加成定价关系分析	19
7.1.1 数据处理与可视化	19
7.1.2 指标设计	22
7.1.3 模型思路与分析	23
7.2 各蔬菜品类未来一周的日补货总量和定价策略分析	23
7.2.1 未来一周各品类蔬菜进货价的预测	23
7.2.2 基于动态规划模型的利益最大化条件下品类的定价与补货策略	26
八、问题三的求解	27
8.1 单品筛选	27
8.2 基于动态规划的单品补货与定价决策	28
8.2.1 指标选择	28
8.2.2 模型建立	29
九、问题四的求解	29
9.1 定价决策	29
9.2 补货策略	30
十、模型评价、改进	31

10.1 优点	31
10.2 缺点	31
十一、参考文献	32
十二. 附录	33

一. 问题重述

1.1 问题背景

现代生活中，经济的发展使得人们对所购蔬菜质量愈发重视。在蔬菜销售市场中，由于蔬菜质量随销售时间的增长而下降，其销售量也会随之进行变化从而影响商家的销售策略。而在实际销售市场中，由于蔬菜进货及交易时间较早，商家须在具体单品和进货价格未知的情况下做出定价和补货决策。为应对此种现状，根据获取的不同蔬菜品类及单品的历史销售情况、成本和损耗率等信息以及销售空间限制等条件，商家可以制定出当日合理的补货、定价和销售策略。

1.2 问题的数据条件

附件 1 是某商超经营的六大蔬菜品类（花叶类、花菜类、水生根茎类、茄类、辣椒类、食用菌类）及其下辖的各单品种类的基本信息；附件 2 是该商超自 2020 年 7 月 1 日至 2023 年 6 月 30 日以来各商品销售数据（单价、销量、销售时间）明细；附件 3 是每天不同单品的批发价格信息；附件 4 是近期以来蔬菜的损耗率。

1.3 问题的提出

(1) 分析蔬菜各品类与销售量的分布规律以及相互关系，分析蔬菜各单品与销售量的分布规律以及分别二者之间的相互关系。

(2) 分析不同蔬菜品类销售总量与成本加成定价的关系，以品类为单位，制定该商超 2023 年 7 月 1-7 日未来一周的日补货总量和定价策略。

(3) 由于销售空间的限制，在销售单品总数约束在 27~33 个且各单品订购量满足最小陈列量 2.5KG 的条件下，在满足市场需求前提下，依据 2023 年 6 月 24-30 日的可销售品种制定出 7 月 1 日的单品补货量和定价策略方案使得收益最大化。

二、问题分析

2.1 问题总分析

本文的总体分析流程图如下：

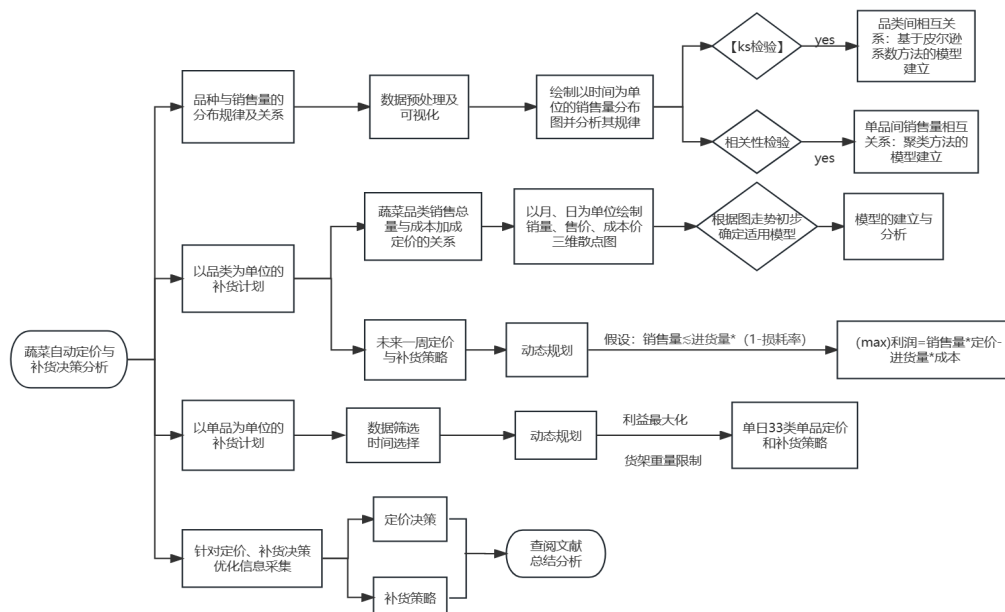


图 1-问题总分析流程图

2.2 具体问题分析

2.2.1 问题一的分析

该小题主要是根据附件 1 和附件 2 给出的商品及其流水信息对该商超各品类及单品的总体销售情况进行整体评价分析并探讨其内部的相互关系，思路流程如下：

首先进行数据的预处理，通过对附件 1、2 的整合，得到了各品类及单品的各月、日销售总量数据情况，具体数据处理信息见附件 1。

其次，针对销售量的分布规律问题，根据题目背景条件，可以得出蔬菜销售量随时间波动的结论，故可以分别绘制出各个品类以日、月为单位的销售量记录图。

然后根据图表结果初步判断销售量分布规律；针对各品类销售量相互关系，再进行相关性检验，根据检验结果选择合适的模型方法进行分析；针对各单品销售量间相互关系，由于单品数量多达 246 种，故采用聚类方法进行分析，在此之前相应的也要先进行数据相关性分析再选择合适的聚类方法类型探索其相互关系。

2.2.2 问题二的分析

问题二首先要求以蔬菜的不同品类为单位，分析其销售总量与成本加成定价的相互联系。针对此问，探究关系题目的背后，首先需要从原始数据的处理开始，逐步实现数据变化趋势的可视化、数据各维度变量的关联图，然后，在建立可视化的基础上，首先对变量间相互关系作出一定分析与猜测并选择适当的模型，当然，数据不是每次都与模型完美适配，故需要不断修正模型的参数及约束条件，逐步找出适配模型进行求解；

针对第二小问，题目要求基于第一小问得出各蔬菜品类销售总量与成本加成的基础上，对未来一周的日补货总量和定价策略做出规划以达到利益最大化的目标。针对此题，明显属于规划类题目，那么应该需要首先明确该题中的目标函数、决策变量以及约束条件，然后对各个条件和变量间的关系进行分析，并且要结合之前已经得出的结论简化变量关系，同时，如若有复杂的实际条件及限制，应做出适时的假设以简化模型的建立。

2.2.3 问题三的分析

问题三要求在三重限制条件（单品总数范围 27~33、单品进货总量小于 2.5KG、可售品种）下制定 7 月 1 日的单品补货量和定价策略。首先从近三年的 6 月 24~30 日中筛选出每日销售量不全为 0 的商品。其次，由于优化目标对象的时间限制在 2023 年 7 月 1 日的范围内，因此，以该时间点为中心，摘取出近三年附近时间段的可售商品单品类，并进行利润分析，再提取出利润率最高的 33 个单品类。最后可以将限制条件 2 作为约束条件通过规划分析获得最终单品的补货策略和定价决策方案。

2.2.4 问题四的分析

问题四是一类开放探索性题目，此类问题可以在结合题目条件的同时，查阅相关资料文献，同时结合个人实际生活经验来阐述。针对定价决策，商品的价格是经济学中很复杂的一门学问，受到了商品本身、商家、消费者群体特征、政府和市场的宏观调节等多重因素的影响，在分析，是可以结合这类因素从多方面综合进行阐述，同时阐明建议意见适用的假设与条件；

针对补货策略，首先，货物的质量、数量本身就对商家的利益由直接的影响，因此，需要从货物方面着手，从货物来源、货物存储、货物供应、货物销售等多方面考虑，分析他们直接对货物产生的一系列独立抑或交叉影响，评价其优缺点而进行改进。

三、模型假设

1. 仅考虑问题中的核心因素，不考虑次要因素的影响
2. 对于第二题中利益最大化要求，在模型中假设：当日售出商品量小于约等于（1-当日折损率）倍数的当日进货量
原因：在真正的现实情况中，售出量应当中包含未折损商品和已折损商品（打折商品来源的合理性），且本题中未给出每日进货量，因此当日各类商品剩余量未知（且剩余商品中按照实际情况也应当包含部分折损商品和部分未售出的完好商品，二者的价格皆未知），因此，为了简化模型的建立，假设当日售出商品量小于约等于（1-当日折损率）倍数的当日进货量，且已售出的商品基本上为未折损商品。
3. 单品名称后的编号表示不同的供应来源，在本解决方案里假设不同编号的商品类型来源归属于不同的单品类型，以简化单品类型合并难度。
4. 假设该商超所有的单品定价在历史时间范围内都是合理的。对附件 2 的数据进行分析发现，各单品的历史定价在 0.1 元/KG~119.9 元/KG 之间的范围内，经过市场调研，确定其定价在基本合理范围之内，保证了单品销售数据的完整性，

四、符号说明

符号	说明	单位
$X_{i,j}$	各蔬菜品类 i 的各月 j 的销售量	KG
Y_d	各蔬菜品类在 2020~2023 年内的 6~8 月内第 d 日均销售量	KG
X_{1d}	各品类蔬菜在 2020~2023 年内的 6~8 月内的第 d 日均定价	
X_{2d}	各品类蔬菜在 2020~2023 年内的 6~8 月内的第 d 日均进价	

五. 数据预处理

5.1 数据预处理

本题附件 1 及 2 提供了该商超近 3 年来各蔬菜单品的销售流水明细。针对较为零散且庞大的数据量，对数据进行预处理使得商品数据合并化，有利于后续问题的分析。借助 excel 软件对附件 1、2 的数据进行了整理，得到了该商超三年内各品类、各单品的月、日销售量以及相关性数据。如下表所示：

表 1-三年内各品类销售量每月合计

日期	花叶类	水生根茎类	花菜类	茄类	辣椒类	食用菌
21 年 1 月	6412.193	1922.91	1262.447	501.591	1919.11	3043.698
21 年 2 月	5741.09	1961.572	1500.92	877.194	3824.593	3691.863
21 年 3 月	5060.645	904.5	1166.262	615.974	2574.561	2246.831
21 年 4 月	4915.061	524.278	1380.398	619.437	1957.177	1379.944
21 年 5 月	5465.001	308.686	1108.386	723.611	2295.673	1164.387
...

表 2-三年内各品类销售量每日合计

分类名称	花叶类	水生根茎类	花菜类	茄类	辣椒类	食用菌
2020/7/1	205.402	4.85	46.64	35.374	76.715	35.365
2020/7/2	198.362	4.6	43.943	32.199	66.064	48.51
2020/7/3	190.779	9.572	42.076	35.896	64.253	42.442
2020/7/4	236.587	5.439	55.662	57.067	81.282	47.262
2020/7/5	223.899	4.019	55.474	61.816	98.496	73.213
...

表 3-三年内各单品销售量每月合计

单品编码	20 年 10 月	20 年 11 月	20 年 12 月	20 年 7 月	20 年 8 月	20 年 9 月
**5168	13.678	13.126	5.972	0	0	0
**5199	0	6.176	0	0	0	0
**5625	0	163.681	0.565	0	0	1.666
**5748	0	0	0	0	0	0
**5762	0	22.95	9.755	0	0	0
...

表格详细内容及其余表格见附件 1，结合以上表格内容，认定三年总销售量小于 2KG 且月销售量大于 0 的频数小于 3 的单品为不固定售出的无效数据，从单品数据集删除，其编号分别为 102900011034231、102900005119098、102900005125815、102900011033944、102900005117353 、 102900051010790 、 102900011009277 、 102900011029275 、 106973223300667。

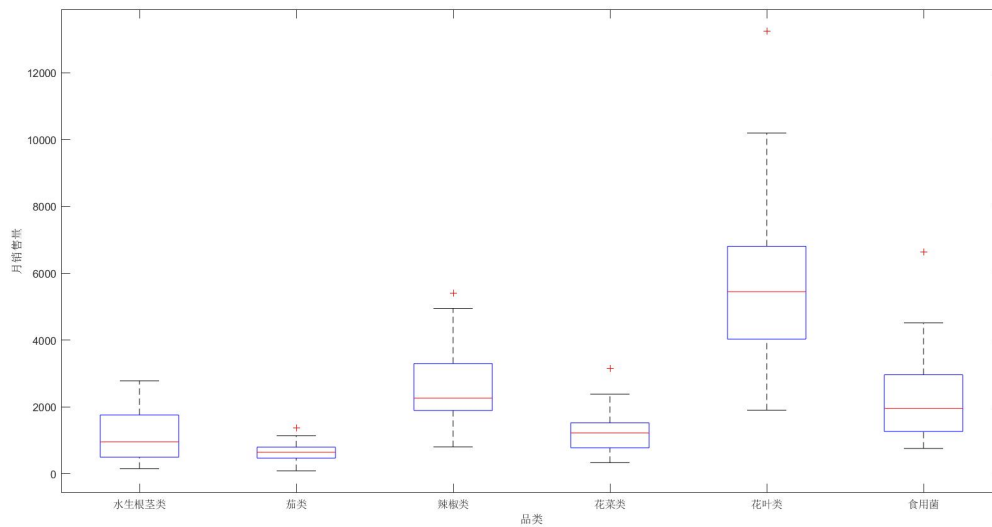


图 2-六大蔬菜品类历史月销售量箱线图

从中可以获取三年内月销量的中位数、最大值、最小值及平均数和异常值。据此可以剔除各品类月售量异常数据。

至此，数据整理与删除完成，后续在此基础上完成模型的建立。

六. 问题一的模型建立与求解

6.1 不同品类销售量相互关系的探索

6.1.1 建模准备——数据的相关性检验分析

由图 2 可知，蔬菜品类的销售量随时间变化呈现出以月份为单位的周期性变化规律且个别品类之间存在相似的波动关系。基于此，为了探索其相互关系，对各个品类的月销售数据(表 1)进行了 Kolmogorov-Smirnov 检验(K-S 检验)。由文献资料可知，KS 检验是一种用于比较两个样本分布是否相同的非参数统计检验方法。它通过计算两个样本的累积分布函数（CDF）之间的最大差异来检验它们是否来自相同的总体分布。在此问题中，选择双样本 KS 检验方法用于比较两个独立样本是否来自相同的分布。

算法步骤如下：

Step1.数据标准化

设 $X_{i,j}$ 为第 i 类蔬菜品类在第 j 个月的销售量，然后将其进行归一化处理以提高模型的稳定性，此处采用了最小-最大归一化方法

$$x_{i,j}^* = \frac{x_{i,j} - x_j^{\min}}{x_j^{\max} - x_j^{\min}} \quad (1)$$

Step2.KS 检验

指定显著性水平 $\alpha=0.05$ ，通过 matlab 中的 kstest 函数（ $[h, p, ksstat] = \text{kstest}(\text{Data}, 'Alpha', 0.05)$ ）得出检验结果 h 和给定显著性水平下观察到结果的概率 p 。如下所示，给出了六大蔬菜品类两两之间进行 KS 相关性检验后的 h 和 p 值：

表 4-各品类月销售量 KS 检验结果 h

分类名称	花叶类	水生根茎类	花菜类	茄类	辣椒类	食用菌
花叶类	0	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00
水生根茎类	1.00E+00	0	1	0	1	1
花菜类	1.00E+00	1	0	1	1	0
茄类	1.00E+00	0	1	0	1.00E+00	1
辣椒类	1.00E+00	1	1	1.00E+00	0	0
食用菌	1.00E+00	1	0	1	0	0

表 5-各品类月销售量 KS 检验下观察结果概率 p

分类名称	花叶类	水生根茎类	花菜类	茄类	辣椒类	食用菌
花叶类	1	2.31E-06	2.31E-06	2.31E-06	2.31E-06	2.31E-06
水生根茎类	2.31E-06	1	0.004596	0.433309	0.000915	0.004596
花菜类	2.31E-06	0.004596	1	0.000151	0.000151	0.065584
茄类	2.31E-06	0.433309	0.000151	1	2.31E-06	0.000151

辣椒类	2.31E-06	0.000915	0.000151	2.31E-06	1	0.186197
食用菌	2.31E-06	0.004596	0.065584	0.000151	0.186197	1

Step3.结果分析

由表 4、5 结果可知，个别类之间的月销售量有存在相同的概率分布的可能，检验通过。故在此基础上可以对符合要求的类别进行基于皮尔逊系数方法的进一步模型建立与分析。

$$\sigma_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

$$\text{cov}(X, Y) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

6.1.2 基于 Pearson 系数判断相关性模型的建立

1.指标计算

a.平均每月各品类蔬菜销售量

基于表 1 总结的各品类蔬菜 i 在不同月份 j 的月销售量合计 $X_{i,j}$,可以计算出 36 个月内各个品类的月销售量平均值 $E(X_i)$:

$$E(X_i) = \bar{X}_i = \frac{1}{n} \sum_{j=1}^n X_{i,j} \quad (2)$$

$i=1,2,3,4,5,6$

b. 各品类蔬菜月销售量的方差

$$\sigma_{X_i}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (3)$$

$i = 1,2,3,4,5,6$

c. 各品类蔬菜月销售量

$$\text{cov}(X_i, X_j) = \frac{1}{n-1} \sum_{i,j=1}^n (X_i - \bar{X})(X_j - \bar{X}) \quad (4)$$

$i, j=1,2,3,4,5,6$

d.皮尔逊相关系数

$$\rho_{X_i, X_j} = \frac{\text{cov}(X_i, X_j)}{\sigma_{X_i} \sigma_{X_j}} \quad (5)$$

$i, j=1,2,3,4,5,6$

根据公式 (2) (3) (4) (5)，可以得出最终相关系数的形式为：

$$\rho_{X_i, X_j} = \frac{E(X_i X_j) - E(X_i)E(X_j)}{\sqrt{E(X_i^2) - E^2(X_i)}\sqrt{E(X_j^2) - E^2(X_j)}} \quad (6)$$

$i, j=1, 2, 3, \dots, 4, 5, 6$

当计算出相关系数之后，可以根据其取值范围判断两个独立变量的相关强度，具体标准如下：

表 6-相关系数数值对应的相关强度

r	相关强度
0.8-1.0	极强相关
0.6-0.8	强相关
0.4-0.6	中等程度相关
0.2-0.4	弱相关
0.0-0.2	极弱相关或无相关

2.模型建立

根据表 1 和图 1 中的结果，为了避免明显异常值对相关性的影响，可以剔除明显异常的各品类各月销售量值，以其余的月销售量的平均值来代替异常值以保证模型拟合的稳定性。校正后所用的数据如下：

表 7-剔除异常值后引入模型的计算值

花叶类	水生根茎类	花菜类	茄类	辣椒类	食用菌
6577.078	332.056	1522.555	1365.551	2198.049	1617.836
7261.517	867.18	1748.658	1139.393	2822.344	1636.524
5668.902	798.931	1336.066	671.692	1992.413	1642.881
6395.312	1344.118	1644.963	771.967	1908.944	3000.532
6518.569	960.695	1943.391	432.389	1837.705	3281.028
...

确定了各品类月销售量合适数据和对应各项指标之后，构建不同品类蔬菜月销售量相关性探索模型，通过 Pearson 系数指标评价六大品类蔬菜相互之间的关系。算法求解流程如下：

Pearson 系数算法求解流程

Step1 将修改后的表格数据导入 matlab

Step2 计算均值、方差等各指标

Step3 计算各品类两两之间月销售量数组的相关系数

Step4 以热力图的形式直观得到相关性

3.求解结果

将六个蔬菜品类的各月销售量数组两两之间进行 Pearson 系数检验，得到的结果如下图所示（其中 1-花叶类，2-水生根茎类，3-花菜类，4-茄类，5-辣椒类，6-食用菌类）。

由图示结果可分析得知 花菜类和茄类商品的月销售量都与辣椒类强相关，花叶类和食用菌类强相关，花叶类与花菜类、茄类、辣椒类弱相关，花菜类与茄类、食用菌类中等相关，而水生根茎类与花叶类、花菜类、食用菌类负弱相关，且与茄类、辣椒类极弱相关。根据生活实际，辣椒作为调料类商品，其购买量与其他主菜类商品必定有紧密的相关关系，结果对这一点有所证实，而水生根茎与其他花叶花菜类作为同质类食品，他们呈现一定的负相关作用，顾客偏好一种类菜品的时候必然对另种类的购买力和购买愿望下降。

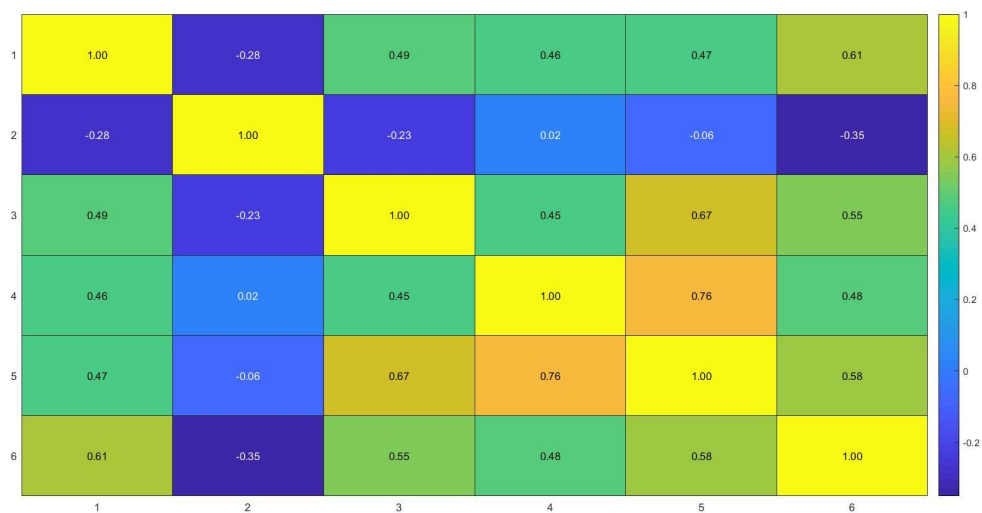


图 3-不同品类的月销售量之间 Pearson 系数热力图

6.2 不同品类月销售量分布规律的探索

6.2.1 建模准备

应用 MATLAB 可对表格数据进行可视化处理^[1]，结果如下：

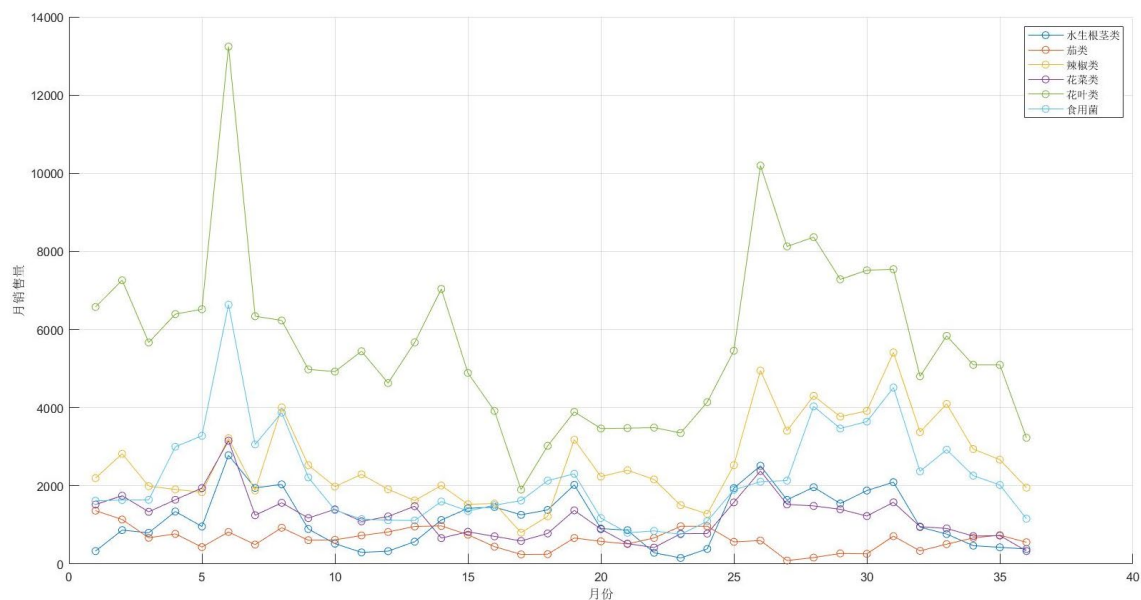


图 4-六大蔬菜品类历史销售总量变化情况

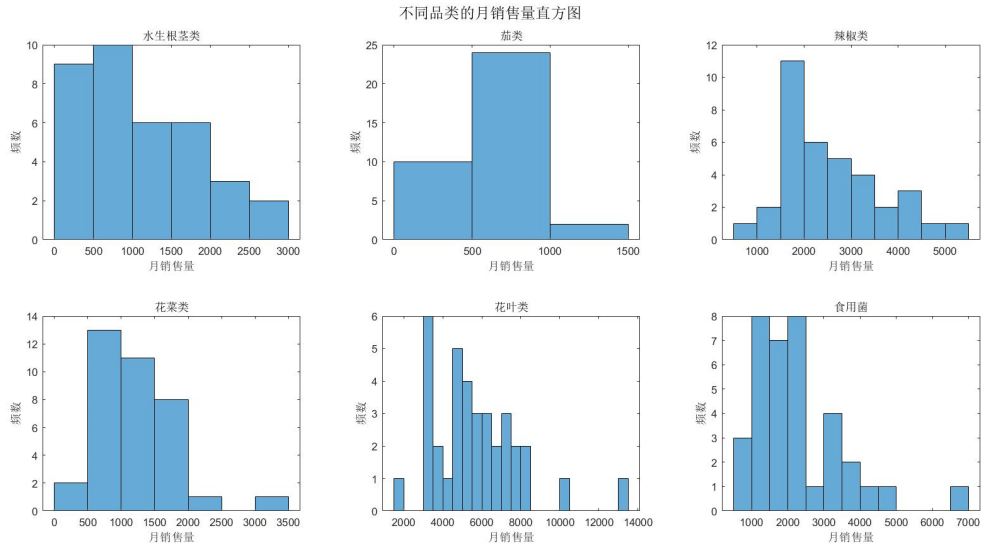


图 5-不同品类月销售量情况直方图

结合上图数据可以近似看出蔬菜品类的售量随时间变化呈现出以月份为单位的周期性变化规律。基于此，可以应用小波分析，通过逼近系数即可判断，以月份为基础单位的销售量波动规律性。根据参考资料，小波变换基本思想是：利用尺度函数表示原始信号，尺度随着尺度级别的下降而逐渐增大，其对原始信号的刻画也愈发模糊，且差异变大。那么假设引入小波函数刻画尺度函数与原始信号间差异，此时结合尺度函数与小波函数两者即可复原原信号。小波变换的优势在于，通过伸缩平移变换对信号函数多尺度细化，可以达到对时、空频率的局部化分析的目的。

基于 matlab 的算法伪代码如下：

```
import data
for m = 1 to 6 do
    Choose wavelet and decomposition level
    Perform wavelet decomposition

    Plot the approximation coefficients
    Calculate energy concentration

    Plot the energy concentration
    Calculate the main frequency component
```

6.2.2 基于小波变换讨论每个蔬菜品类不同月份销售量的周期性变化规律

1. 预备指标

a. 尺度函数^[2]

尺度函数 $\varphi(x)$ 由实数和平方可积的函数组成，将其进行 2 的伸缩变换和整数倍的平移后可以得到集合

$$\begin{aligned} & \{\varphi_{j,k}(x) \mid j, k \in \mathbb{Z}\} \\ \text{s.t. } & \varphi_{j,k}(x) = 2^{j/2} \varphi(2^j x - k) \end{aligned} \quad (7)$$

此处运用常见的哈尔尺度函数：

$$\varphi(x) = \begin{cases} 1 & \text{if } 0 \leq x < 1 \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

b.小波函数

给定尺度函数 $\varphi(x)$ 后，必定存在一个小波函数 $\psi(x)$ 为：

$$\psi_{j,k}(x) = 2^{\frac{j}{2}} \psi(2^j x - k) \quad \text{s.t. } (j, k \in \mathbb{Z}) \quad (9)$$

若用 W_{j_0} 表示小波函数集合 $\{\psi_{j_0,k}(x) \mid j, k \in \mathbb{Z}\}$ 所长成的空间，则可以得到

$$V_{j_0+1} = V_{j_0} \oplus W_{j_0} \quad (10)$$

且 W_{j_0} 与 V_{j_0} 的基函数正交，

$$\langle \varphi_{j_0,k}(x), \psi_{j_0,l}(x) \rangle = 0 \text{ for } k \neq l \quad (11)$$

又可变换为：

$$\psi(x) = \sum_k h_\psi(k) \sqrt{2} \varphi(2x - k) \quad (12)$$

由公式（9）（10）（11）（12）可以证明得到小波函数系数

$$h_\psi(k) = (-1)^k h_\varphi(1 - k) \quad (13)$$

c.哈尔小波函数

$$\psi(x) = \begin{cases} 1 & 0 \leq x < 0.5 \\ -1 & 0.5 \leq x < 1 \\ 0 & \text{elsewhere} \end{cases} \quad (14)$$

至此，计算各项指标之后，在 MATLAB 进行模型的建立并可视化分析。

2.模型的建立与分析

此处分别限于篇幅，以花叶类为例来观察其逐月销售量的周期性波动。其余品类的可视化小波分析详见支撑材料。

① 花叶类：

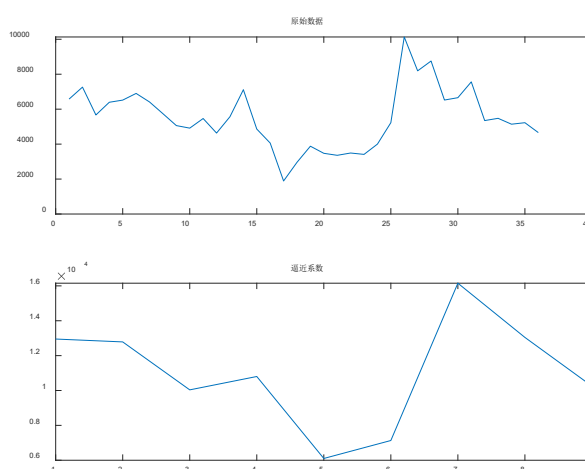


图 6-花叶类月销售量原始数据及小波变换后各分解层的逼近系数

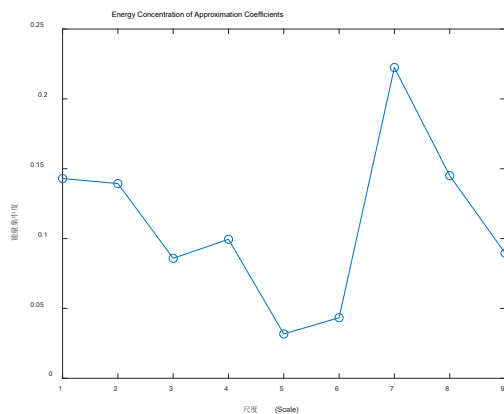


图 7-不同尺度对应的能连集中度

由图 6、7 可以得出，花叶类菜品的月销售量波动在周期为 7 时最具有规律波动性。以下是各蔬菜品类的月销售量经过小波变换后的主要频率成分及对应周期：

表 8-各蔬菜品类的主要频率成分及对应周期

蔬菜品类	主要频率成分	周期
花叶类	0.14286	7
水生根茎类	0.14286	7
花菜类	0.14286	7
茄类	1	1
辣椒类	0.125	8
食用菌	0.128	8

由表 8 可知，除茄类以外，其他各类蔬菜品类呈现以周期为 7 或者 8 个月的规律性波动。

6.3 各蔬菜品类内不同单品月销量的分布规律

通过 excel 对附件 2 进行数据正路可以得到不同蔬菜品类内各单品的月销售量总和，并通过 box_plot 函数在 matlab 上绘制出对应的箱线图。

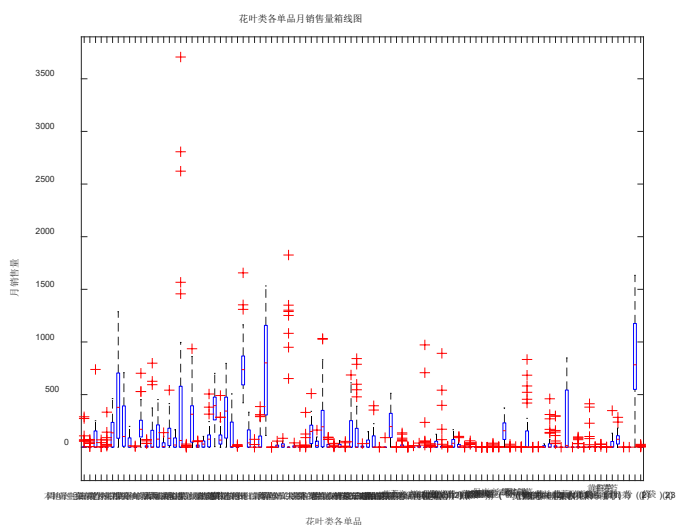


图 8-花叶类各单品月销售量箱线图

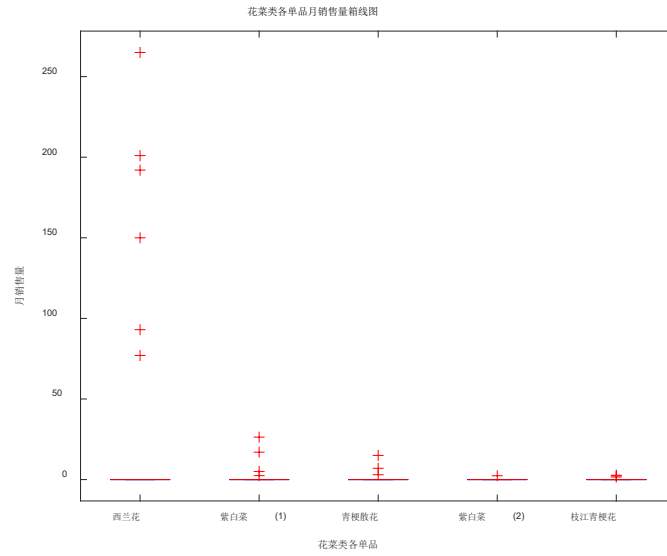


图 9-花菜类各单品月销售量箱线图

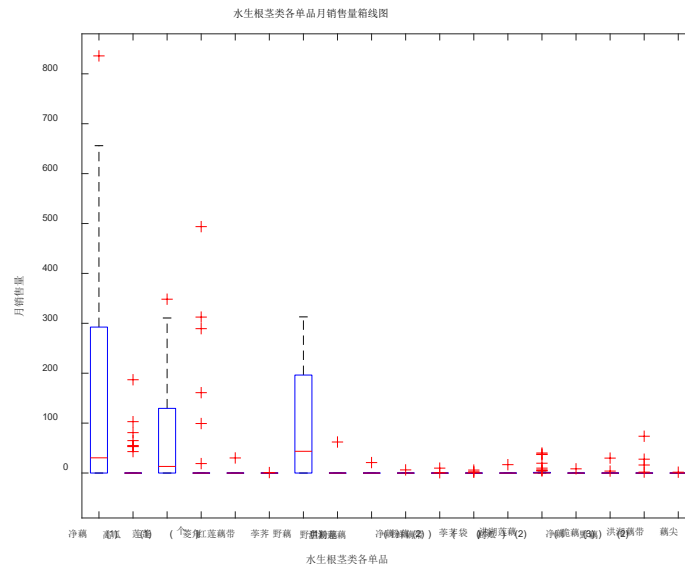


图 10-水生根茎类各单品月销售量箱线图

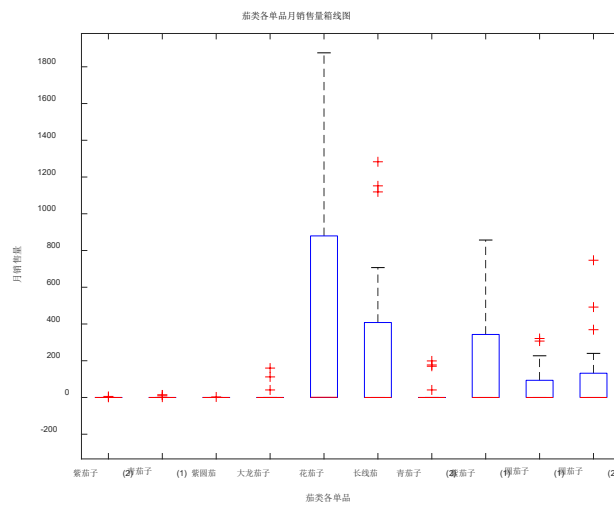


图 11-茄类各单品月销售量箱线图

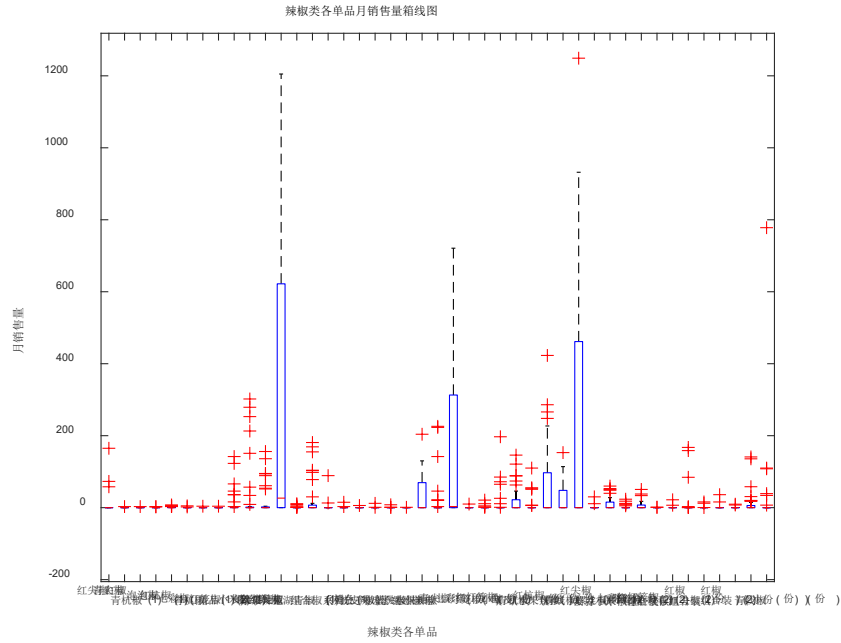


图 12-辣椒类各单品月销售量箱线图

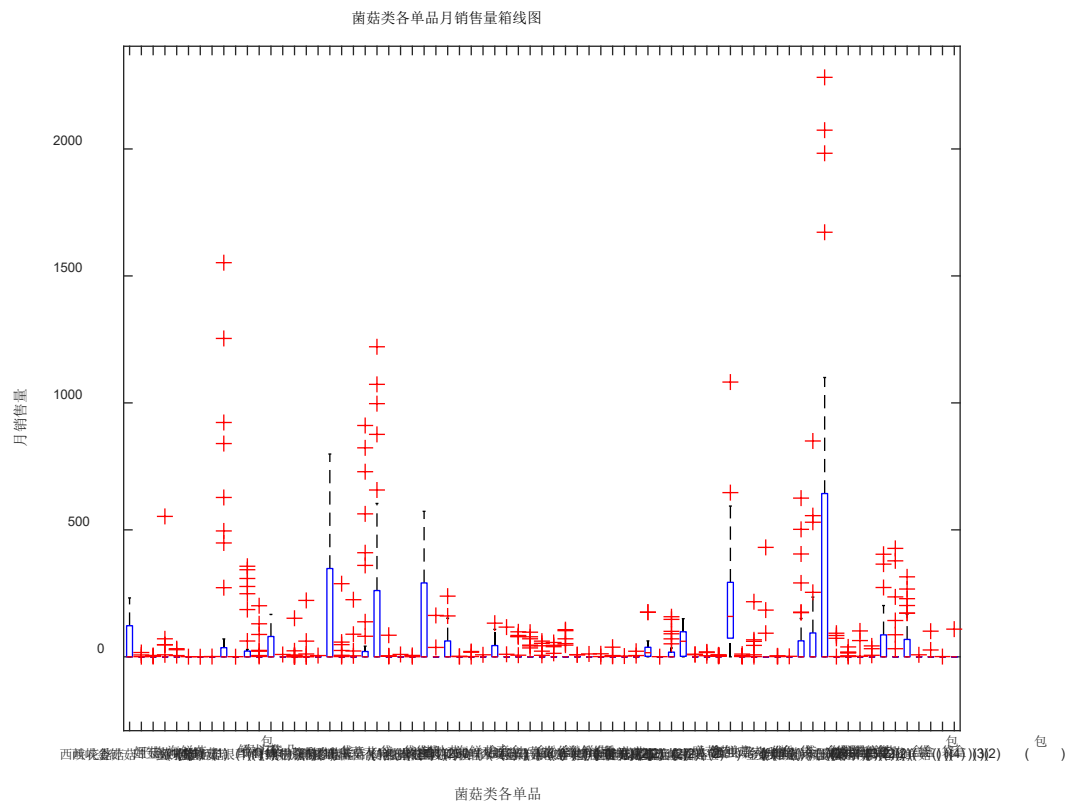


图 13-菌菇类各单品月销售量箱线图

七. 问题二求解

7.1 各蔬菜品类的销售总量与成本加成定价关系分析

7.1.1 数据处理与可视化

根据对附件 2 的数据处理，可以得到各品类商品每日平均定价随每日时间变化的关系折线图，如下所示：

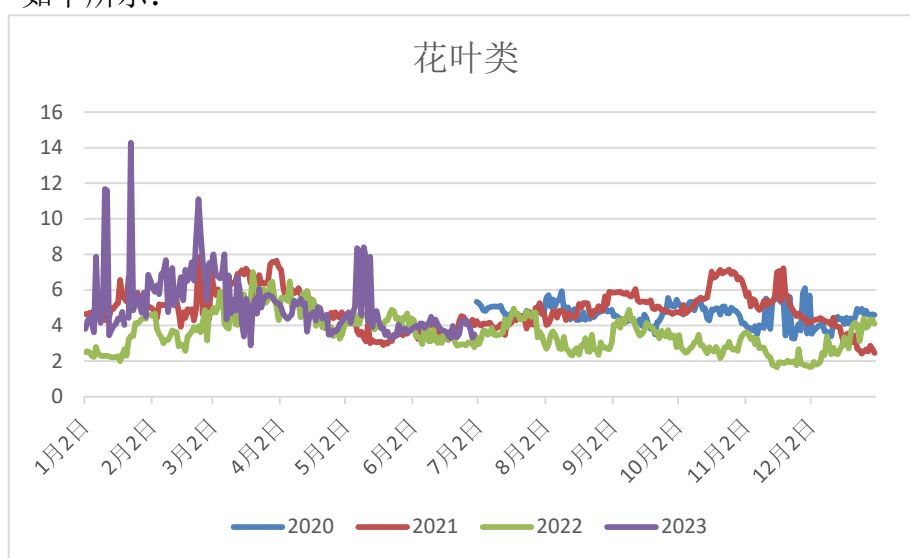


图 14-花叶类商品每日平均定价日变化趋势

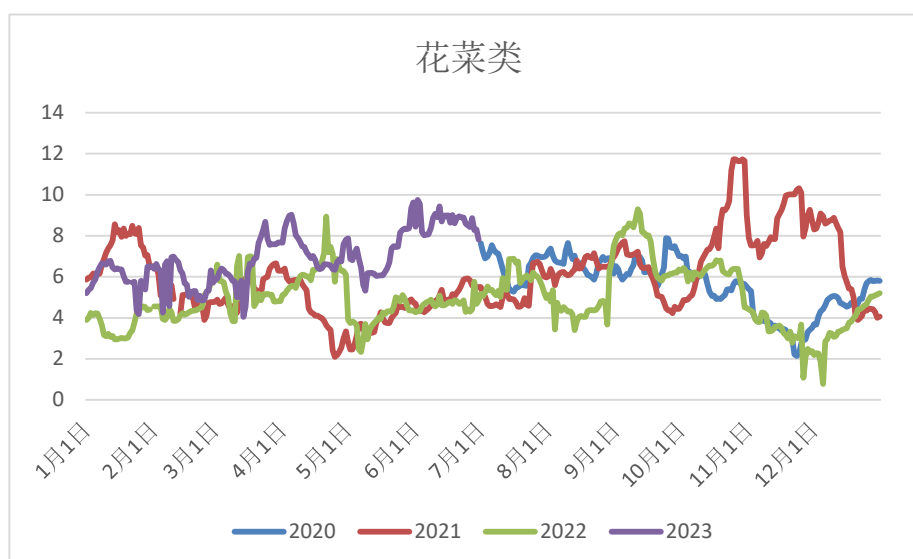


图 15-花菜类商品每日平均定价日变化趋势

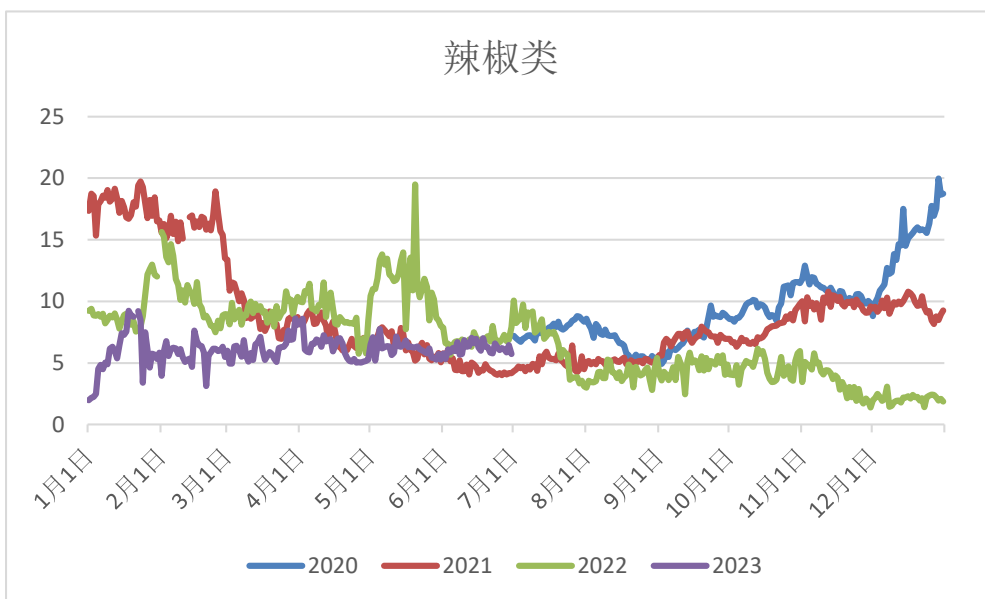


图 16-辣椒类商品每日平均定价日变化趋势

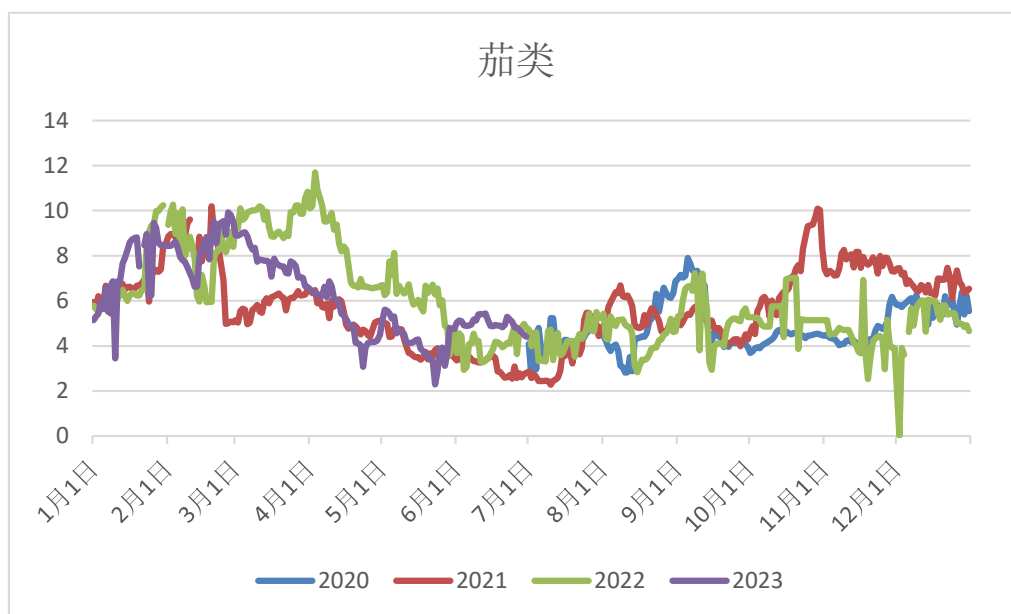


图 17-茄类商品每日平均定价日变化趋势

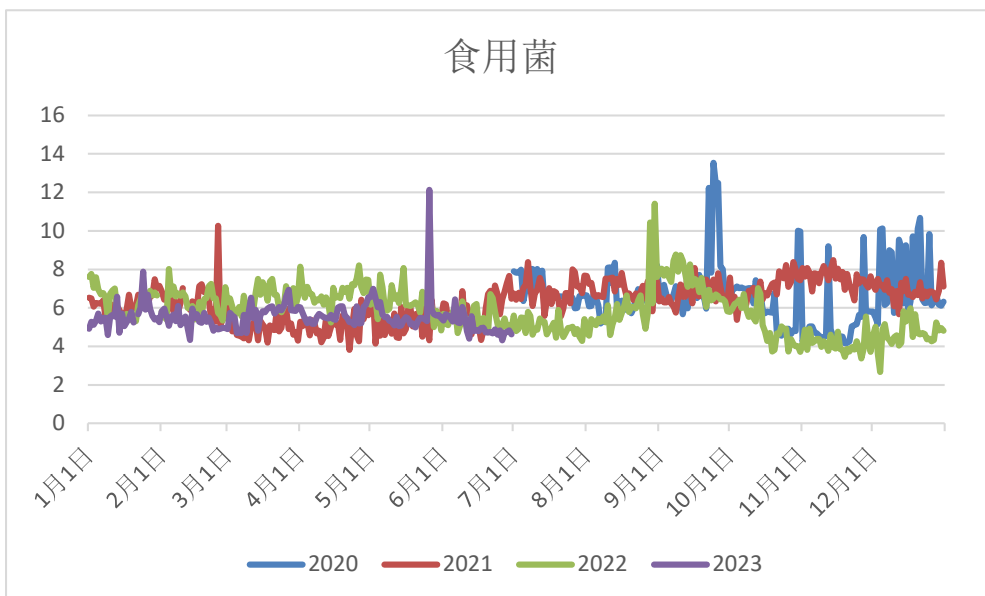


图 18-食用菌类商品每日平均定价日变化趋势

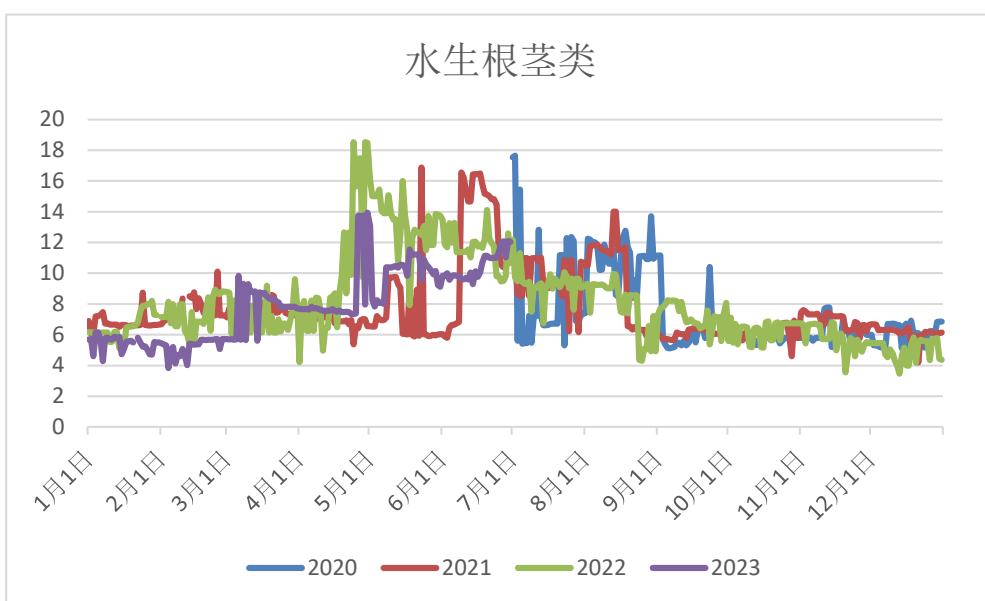
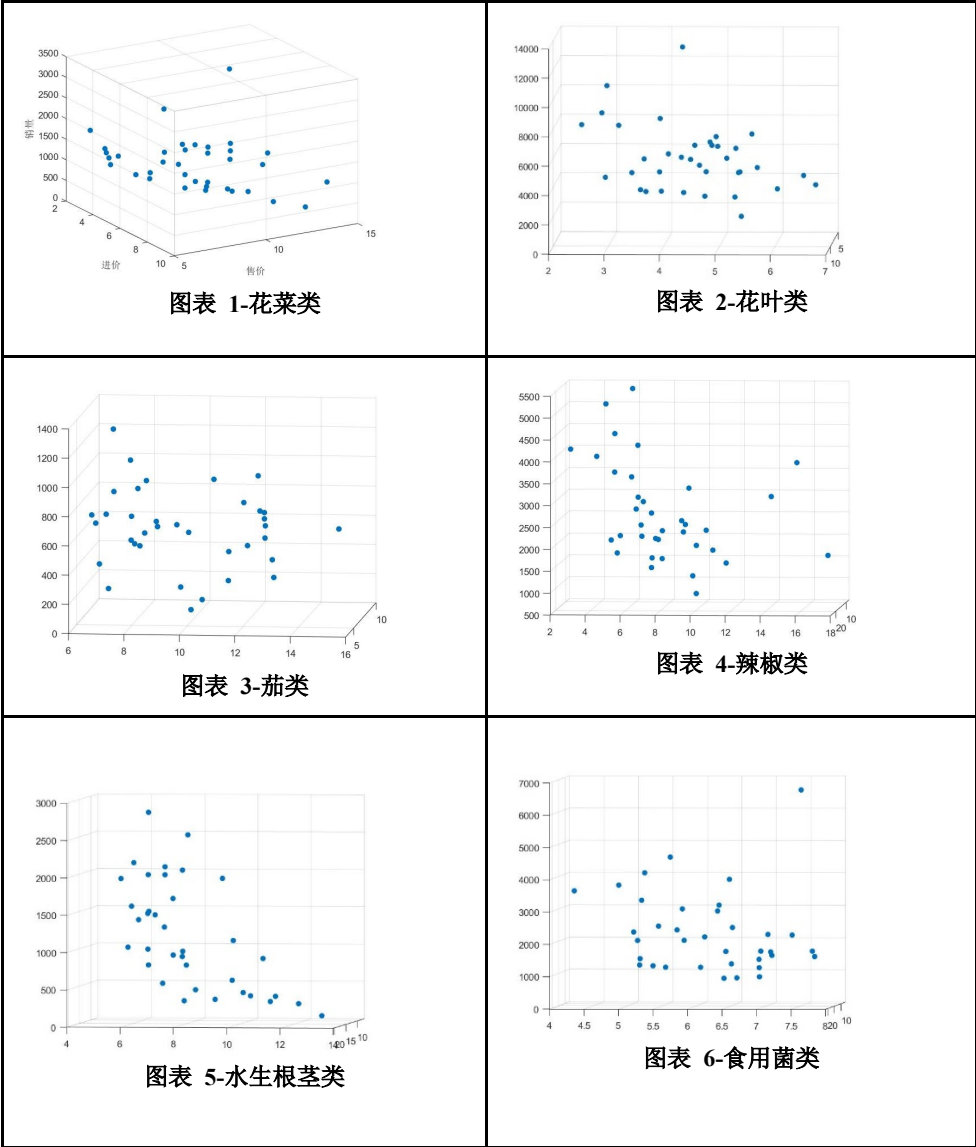


图 19-水生根茎类商品每日平均定价日变化趋势

根据上图可以分析出 2020 年、2021 年、2022 年以及 2023 年的各品类蔬菜在 6、7、8 月内的每日平均定价呈现大致相同的走势，为了简化模型减少数据量的处理且更好的拟合关系，即只采用三年分别的 6、7、8 月的品类日均定价数据。据此，可以绘制出近三年内 7/8/9 月中每天不同品类商品日售量与当日售价和进价之间的三维散点图，如下所示：

表 9-各品类日销量与定价、进价三维散点关系图



由图可以看出三者大致呈现线性相关性，则可以采用多元线性回归分析模型建立不同品类商品日销售量与当日售价、进价之间线性关系。^[3]

7.1.2 指标设计

a.目标函数

各品类在 2020~2023 年内的 6~8 月内日均销售量为目标函数：
假设为 Y_d ($d \in [1,92]$ ，第 d 日)

b.决策变量

各品类蔬菜在 2020~2023 年内的 6~8 月内的日均定价为 X_{1d} ，
各品类蔬菜在 2020~2023 年内的 6~8 月内的日均进价为 X_{2d}
此时，将上述三类指标融合为一个三维矩阵 $X=[X_{1d}, X_{2d}, Y_d]$

c. 马氏距离

$$D_M(X) = \sqrt{(X - \mu)^T \Sigma^{-1} (X - \mu)}$$
 (s.t. Σ 为多维随机变量协方差矩阵， μ 样本均值)

7.1.3 模型思路与分析

算法步骤:

步骤	算法
Step1	导入预处理后表格数据
Step2	读取三个维度变量并储存在三维矩阵中
Step3	计算每个数据点的马氏距离
Step4	设置离散群阈值
Step5	标记离群点并获得其索引并剔除得到无离群点的数据样本
Step6	对无离群点数据样本进行多元线性回归拟合

其中 step4 中的离散群阈值通过卡方逆累积分布函数来计算

卡方分布是一个单参数曲线族。参数 ν 是自由度。卡方分布的 icdf 为:

$$x = F^{-1}(p | \nu) = \{x: F(x | \nu) = p\}$$

$$\text{其中, } p = F(x | \nu) = \int_0^x \frac{t^{(\nu-2)/2} e^{-t/2}}{2^{\nu/2} \Gamma(\nu/2)} dt$$

(ν 是自由度, $\Gamma(\cdot)$ 是 gamma 函数。结果 p 是来自自由度为 ν 的卡方分布的单个观测值落在区间 $[0, x]$ 内的概率。)

求解结果:

表 10-Estimated Coefficients:

	Estimate	SE	tStat	pValue
Tntercept	65.5785865715634	2.8245	23.217	4.5232e-97
X1d	-2.69350017422215	0.26417	-10.196	2.264e-23
X2d	-0.0292929631304963	0.02035	-1.4395	0.15031

Root Mean Squared Error: 21.6

R-squared: 0.0906, Adjusted R-Squared: 0.0889

F-statistic vs. constant model: 53.8, p-value = 5.11e-23

最终得到蔬菜品类日销售量关于售价和进货价的多元线性回归方程:

$$Y = 65.5786 + -2.6935 * X1 + -0.0293 * X2 \quad (15)$$

7.2 各蔬菜品类未来一周的日补货总量和定价策略分析

7.2.1 未来一周各品类蔬菜进货价的预测

基于时间序列预测模型 ARMA 的各品类未来一周进价预测

(1) 算法步骤

步骤	算法
Step1	导入预处理后表格数据为二阶矩阵
Step2	创建测试和训练数据
Step3	使用 Augmented Dickey-Fuller 检查数据是否平稳, 若否则进行差分处理
Step4	通过自相关图和偏自相关图确定 arma 模型阶数

- Step5 若 step4 无法准确得到阶数，则使用 AIC 准则
 Step6 绘制阶数热力图，可视化不同模型 AIC 值以确定最佳模型
 Step7 基于所选阶数，使用函数进行拟合
 Step8 预测
 Step9 模型验证：对预测数据进行反差分以还原原始数据
 Step10 可视化预测数据，并计算相对误差
-

(2) 模型建立：

①数据指标：

a.自回归模型 AR (p)

如果时间序列 x 满足：

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \varepsilon_t \quad (16)$$

x_t 为时间序列 x 在时刻的值， $\phi_1, \phi_2, \dots, \phi_p$ 为自回归系数， ε_t 为独立同分布的随机变量满足：

$$E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) > 0 \quad (17)$$

则称时间序列 x 服从 p 阶自回归模型 AR (p)

b.移动平均模型 MA (q)

若时间序列 x 满足：

$$x_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} \cdots - \theta_q \varepsilon_{t-q} \quad (18)$$

$\theta_1, \theta_2, \dots, \theta_q$ 为移动平均系数，成时间序列 x 服从 q 阶移动平均模型 MA (q)

c.自回归移动平均模型 ARMA(p, q)

若时间序列 x 满足：

$$x_t - \phi_1 x_{t-1} \cdots - \phi_p x_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} \cdots - \theta_q \varepsilon_{t-q} \quad (19)$$

②模型求解

以花菜类为例，在以其日均进价为时间序列创建 ARMA 模型可视化过程如下：

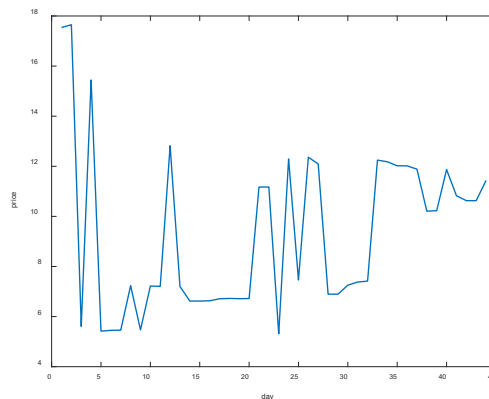


图 20-花菜类日均成本价原始时间序列

检验结果发现不平稳，则进行了差分处理，结果如下：

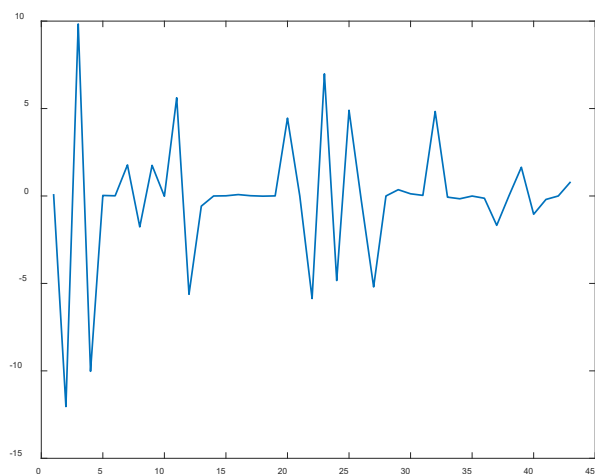


图 21-差分化处理原始序列

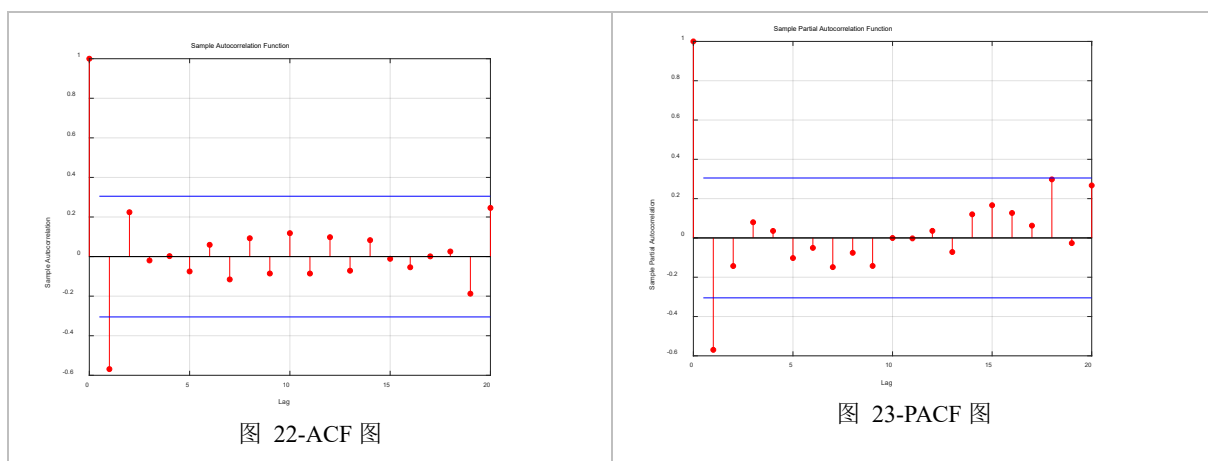


图 22-ACF 图

图 23-PACF 图

依据图 21、22，根据通用规则：如果 ACF 图在滞后 p 处截尾，而 PACF 图在滞后 q 处截尾，则可以初步估计模型为 $ARMA(p, q)$ ；如果 ACF 图在滞后 p 处截尾，而 PACF 图在滞后 q 处截尾，但 $q=0$ ，则可以初步估计模型为 $AR(p)$ ；如果 ACF 图在滞后 q 处截尾，而 PACF 图在滞后 p 处截尾，但 $p=0$ ，则可以初步估计模型为 $MA(q)$ 。最终确定 $p=4, q=10$ 。

据此，将此参数代入 `arima` 函数可以获取未来七天的预测结果。数据如下：

表 11-各蔬菜品类 7 月未来一周的日均进价预估

品类	1 日	2 日	3 日	4 日	5 日	6 日	7 日
花菜类	8.59	8.5	8.43	8.86	8.28	8.31	7.81
花叶类	4.24	4.19	4.33	4.1	3.84	3.31	3.5
辣椒类	6.28	6.06	6.21	6.1	5.86	6.44	5.73
茄类	5.06	4.83	4.81	4.7	4.61	4.48	4.4
菌菇类	4.6	4.84	4.31	4.71	4.75	4.84	4.62
水生根茎类	11	11.4	11.93	12.09	11.03	12.1	12.04

7.2.2 基于动态规划模型的利益最大化条件下品类的定价与补货策略

1.前提:

根据假设 2，假设各品类日均销售量小于约等于（1-损耗率）倍的进货量，并且约定：

$$\text{日利润} = \text{品类日均售价} \times \text{日均销售量} - \text{日均进货量} \times \text{进货价}$$

2.指标处理

a.决策变量:

各品类 i 日均定价 P_i 、各品类日均进货量 I_i

b.约束条件:

各品类日均销售量小于约等于（1-损耗率）倍的进货量，即

$$S_i \leq (1-\eta)I_i \quad (20)$$

(其中 S_i 表示各品类日均销售量， η 为当天该品类日均损耗率)

c.目标函数

各品类日均利润 R_i 为优化的目标函数，则

$$R_i = P_i \times S_i - C_i \times I_i \quad 21$$

(其中 C_i 为各品类日均进价)

3.算法步骤

步骤	算法
Step1	定义参数（定价、日进货量、初始利润）
Step2	以六大品类为外层循环
Step3	内层循环：在考虑损耗率的情况下，以销售额大于成本条件进行利润的计算
Step4	以利润为判断标准，更迭定价与进货量
Step5	规划结果输出

4.模型求解

联立公式 15、20、21，可以得知各品类日均销售量可有公式 15 的回归分析方程中的售价和成本加成得到具体数值，而各品类日均成本价可以通过 7.2.1 中的 arma 预测模型得到，再在约束条件的限制下（其中损耗率的取值来源于附件 4 中各品类中各单品损耗率的均值），最终求得各品类未来一周的补货总量和定价决策，如下表所示：

（假设 P 代表最佳售价， I 代表最佳日进货量， R 代表最大日利润）

表 12-各品类未来一周的补货总量和定价决策

品类		1 日	2 日	3 日	4 日	5 日	6 日	7 日
花菜类	P	18.00	18.00	17.00	17.00	17.00	17.00	17.00
	I	17.00	17.00	20.00	20.00	20.00	20.00	20.00
	R	136.71 6	138.46	139.93	139.93	143.35	143.35	154.06

花叶类	P	26.00	26.00	26.00	26.00	26.00	26.00	26.00
	I	98.00	98.00	98.00	98.00	98.00	98.00	98.00
	R	2057.91	2063.31	2063.31	2073.01	2101.04	2158.17	2158.17
辣椒类	P	22.00	22.00	26.00	22.00	22.00	22.00	22.00
	I	44.00	44.00	44.00	44.00	44.00	44.00	44.00
	R	665.97	676.43	676.43	676.43	685.93	685.93	692.11
茄类	P	18.00	18.00	18.00	18.00	18.00	18.00	18.00
	I	10.00	10.00	10.00	10.00	10.00	10.00	10.00
	R	123.67	126.13	126.34	127.52	128.48	129.87	130.73
菌菇类	P	11.00	11.00	10.00	10.00	10.00	10.00	10.00
	I	37.00	37.00	43.00	43.00	43.00	43.00	43.00
	R	216.64	216.64	228.27	228.27	228.27	228.27	228.27
水生 根茎类	P	17.00	17.00	17.00	17.00	17.00	17.00	17.00
	I	15.00	15.00	15.00	15.00	15.00	15.00	15.00
	R	62.11	64.46	64.60	64.57	64.58	64.47	64.46

八. 问题三的求解

8.1 单品筛选

1.针对“根据 2023 年 6 月 24-30 日可售品种”这一要求，从附件 4 总结整理出了 2023 年的 6 月 24~30 日这个时间段内，通过 excel 筛选出这十天的每日售出量不全为 0 的单品类别，合计共 49 种单品。

2.针对“要求出售单品总数数量范围在 27~33 内”这一条件，那么要进一步筛选单品数量，而总目标在于使得商超利益最大化，故需要从这 49 种单品中筛选出利润率最高的 33 种单品。

筛选标准和过程如下：

步骤	方法
Step1	明确筛选标准：利润=（定价-进货价）*销量
Step2	明确时间段：由要求 1，选择三年内最接近 6 月 24~30 日这一时间段的数据
Step3	明确数据来源:销量（预测回归拟合），进货价（附件 4 整理并计算均值）
Step4	计算：49 种单品利润结果
Step5	排序择取

筛选结果如下表 13 所示：

表 13-筛选后的 33 种指定时间段内利润率最高的单品

编号	名称
102900005115250	西峡花菇(1)
102900005115762	茼蒿
102900005115779	云南生菜
102900005115823	上海青
102900005115908	菜心
102900005115946	木耳菜
102900005115984	云南油麦菜
102900005115786	竹叶菜
...	...

8.2 基于动态规划的单品补货与定价决策

8.2.1 指标选择

a.目标函数:

单品日利润 = 单品日销量*单品定价 - 单品日进货量*单品进货价

(假设分别用 R_d , S_d , P_{sd} , I_d , P_{id} 来表示单品利润 (daily profit)、单品日销量 (daily sale)、单品定价 (daily sale price)、单品日进货量 (daily input)、单品进货价 (daily input price)) 那么上述文字公式可以表述为:

$$R_d = S_d \times P_{sd} - I_d \times P_{id} \quad 22$$

b.决策变量:

选择单品定价 P_{sd} 和单品日进货量 I_d 作为决策变量。

c.其他非决策性变量

对于其他变量, 单品进货价可以基于三年内每天的进货价时间序列预测得出 2023 年 7 月 1 日的进货价;

单品日销量是基于 2020 年 7 月 1 日~7 月 1 日、2021 年 6 月 15 日~7 月 1 日、2022 年 6 月 15 日~7 月 1 日、2023 年 6 月 15 日~6 月 30 日这四个时间段, 通过各单品日销量与单品定价进行线性回归, 得到此二者之间的线性关系。拟合结果如下:

表 14-33 类单品日销量与定价的线性回归关系

单品	拟合关系
单品 1	$s = -0.8510 * p + 27.9318$
单品 2	$s = -0.6952 * p + 12.4512$
单品 3	$s = -0.9434 * p + 27.5004$
单品 4	$s = -0.9513 * p + 18.3402$
单品 5	$s = -0.2074 * p + 6.7451$
...	...

(其中单品序号顺序与表 13 保持一致)

d.约束条件

各单品订购量满足最小陈列量 2.5 千克，即： $\sum_{i=1}^{33} I_{d_i} \leq 2.5$

8.2.2 模型建立

本题的模型采用与 7.2.2 相近的动态规划模型进行求解。基本算法与其一致，得到 33 类单品 7 月 1 日的补货量和定价策略如下：

表 15-2023-07-01 单品补货计划和补货策略			
单品名称	销售定价	补货量	预计利润
西峡花菇(1)	26	7	41.75
茼菜	11	6	38. 14
云南生菜	17	14	123.32
竹叶菜	11	10	61.43
上海青	17	4	38.25
菜心	13	5	31.54
...

九、问题四的求解

9.1 定价决策

考虑到价格弹性在商品定价中的作用^[4]，从商家、消费者、政府、市场变化、商品自身出发来分析该商超的定价决策。该商超的蔬菜类产品，其质量受到贮存环境、天气、运输条件、人为活动等诸多因素的影响，存在保存期限短、价格易浮动等特点，故需从多方面衡量其价格预定及其后续调整。

其一，从商品自身来看，蔬菜品类繁多，贵贱不一，某些商品之间具有可替代性，某商品的可替代商品越多，该商品的价格需求弹性随着其可替代品的相似程度的提高而增大。从商品在消费者生活中的必需性来看，一些基础的应季蔬菜以及调料性蔬菜，作为人体维生素和膳食纤维的重要来源，是广大消费者的生活必需品，其需求价格弹性较小，而一些高端珍稀蔬菜和非应季商品，它们的价格需求弹性较大，消费者会随着其定价的提高而明显减少对其的购买量；

其二，从商家自身来看，该商超所处地理位置、周边繁华程度等因素直接的影响了该商超的商品定价。比如说，若该商超地处一线城市繁华地段，由于租金、人力资源成本、运营成本等，相应的就需要提高部分商品的价格来保证成本维护的合理性，特别是针对那些运输成本高、保存条件苛刻的蔬菜类产品。同时，要避免价格溢出，需在市场范围内做好市场调研与价格流动监控，避免个别失误定价决策而导致消费者群体流失。

其三，从消费者群体特征来看，结合第二点来说，店面的地理位置间接的决定了消费者人群特征，消费者的性别比例、年龄特征、经济状况等因素都会影响消费者的购买力能力和对商品定价的认可度。比如，一些年轻消费者和部分男性群体，他们对日常生活类商品，特别是食物等缺乏一定的价格认知，往往会盲目的消费此类产品；而与前者

相反的是，对于一些老年人群体和部分中年妇女群体，他们对价格的敏感度往往比较熟稔，故在此类群体居多的情况下，根据实际条件调控价格，比如说，适时的打折促销活动，可以提高此类消费者人群的用户黏性和活跃度；

其三，宏观因素的影响。季节性因素，很多蔬菜类产品基本都是具有时节性产出特征的，当其处于应季时间段内时，其数量往往会梯度式上升，此时其价格会相应降低，反之，反季节性蔬菜由于培育、贮存条件成本较高，数量较少，往往定价偏高；政府和市场的宏观调控因素，部分蔬菜类产品，作为生活必需品，其价格的稳定程度与社会的稳定有一定关联，政府为了社会的福祉与安定，往往会对此类必需品的价格做出相应的限制。比如说，新冠疫情期间，“买菜难”一度成为群众热点问题，部分地区甚至出现“一菜难求”“抢菜浪潮”等现象，不少商家趁机坐地起价，干扰市场消费环境，政府也是对违规涨价的商家做出了一定处罚以稳定市场秩序。

根据以上分析，商户可以采集以下信息：**周期性的周边市场调研活动：**收集当前市场数据迎合潮流；**用户画像：**通过走访、调查问卷等形式采集该商超消费群体特征并作出适时调整以迎合大众的消费需求和特点，实现利益最大化；**对商家自身运营优劣的精准分析：**假如说该商超体量较大，拥有多家分店，那么商家要及时对下辖商店运营的优劣做出量化性分析，比如环境整洁条件、货物新鲜度、人员服务态度等，结合相关指标对自身运营进行优良和改进，以提升社会评价、提高消费者群体黏性最终达到利益最大化。

9.2 补货策略

根据题目条件，该商超已经收集了近三年蔬菜类商品的销售流水以及衡量了每个蔬菜类单品的损耗率，有助于商家分析出过往历史商品交易的特点，有助于商超进行补货决策，但是还缺乏一定的量化数据。

①**库存量：**记录每天每个蔬菜类单品及品类的库存水平（包括进货量、库存量、进货日期、进货成本等），有助于量化蔬菜类产品的销售质量，以及预估商品销售趋势。

②**供应商信息：**记录包括供应商名称、联系方式、交货准时度、交货质量等因素在内的与供应商的历史合作情况，这有助于商户衡量众多商品源的优劣程度，供其选择或者淘汰其中的供养商，有利于提升商品质量、降低商品成本，同时可以根据其准时度合适选择进货时间，及时补充货源。

③**外部数据（季节性环境变化数据、市场消费数据及竞争对手相关信息）：**蔬菜类商品的销售量受诸多因素（气温、降雨量、气候条件等）影响，通过提前收集该类信息有助于商家合理规划进货量；同时，周期性关注、调查市场消费趋势，诸如节日、节气等特殊时间段对特定商品销量的具有较大的影响以及个别特殊时间段或者地区存在的部分的消费潮流；针对周边地区出售同类商品的竞争对手，可以收集竞争对手的定价信息、促销活动情况和市场份额等信息，及时调整补货策略和价格策略以了解市场竞争情况并制定竞争策略。

④**销售渠道效果数据：**假设说该商超规模属于大型类，那么商户还可以对不同销售渠道（实体店铺销售、网络购买、社区团购等）的销售效果进行分析，以确定各个销售渠道的利润率并调整进货策略。

十. 模型评价、改进

10.1 优点

- (1) 本文模型是对大量数据认真分析和处理后所得到建立的，相比小样本模型，更具有普适性，可靠性，可推广使用。
- (2) 本文在分析单品销售量相互关系时，采用聚类分析模型，有利于计算更为科学准确的关联系数，得到适于组合销售的单品蔬菜。
- (3) 本文对单日内蔬菜销售价格和是否折扣与时间段的关系进行定性分析，有利于科学决策折扣时间。
- (4) 本文在进行品类批发价预测时，采用 ARIMA 模型，更加简单和易于实现，通过截取变化趋势稳定的时间段，使得批发价的预测更加准确，
- (5) 本文在进行单品批发价预测时，采用季节性分解方法，更加适用中短期预测，使得预测结果更接近真实水平。
- (6) 本文制作了大量可视化图表，使得数据的分析、沟通和决策制定更直观和便捷，有利于更好地理解 and 探索数据变化规律。
- (7) 本文第三问的模型以第二问的模型为基础，对其修改和增加相应的约束条件，可见模型具有复用性。

10.2 缺点

- (1) 本文建立的商超收益为最大目标的动态规划所考虑到的复杂约束条件较少。
- (2) 本文所建立的动态规划模型，由于其内部递归函数的使用。将会占用较大内存。
- (3) 本文缺少对是否打折以及打折力度的综合考虑，结果的真实性和准确性有所降低。

十一、参考文献

- [1] 崔汝麦,司守奎.数据可视化处理的数学模型[J].海军航空工程学院学报,2010,25(01):105-108+112.
- [2] 白丽,王孟伟,周悦等.我国辛辣类蔬菜价格波动特征研究[J/OL].北方园艺:1-14[2023-09-09].<http://kns.cnki.net/kcms/detail/23.1247.S.20230906.1135.002.html>
- [3] 王鹏飞,李涛,于春亮等.基于多元线性逐步回归和 BP 神经网络建立鸭梨盛花期预测模型[J].山东农业科学,2023,55(07):159-166.DOI:10.14083/j.issn.1001-4942.2023.07.022.
- [4] 伍希.需求价格弹性在商品定价中的应用研究[J].商场现代化,2011(23):13-14.

十二. 附录