某商超蔬菜类商品动态定价与补货决策研究

摘 要

随着生鲜市场规模的持续扩大，蔬菜零售行业的竞争也愈加激烈。为帮助某商超 改善经营模式，本文基于题目所给数据信息，建立数学模型进行分析，从而制定合理 的蔬菜类商品动态定价与补货决策，提高商超收益。

针对问题一，第一问首先分别计算出蔬菜六大品类和蔬菜单品销售量的总值、平均值、最大值与最小值、标准差、偏度系数、峰度系数等描述统计量。然后进行数据可视化处理，分析各蔬菜品类销售量分布规律。结果可得各蔬菜品类的日销售量**都呈现出不同程度的季节性波动**，其中花叶类和辣椒类蔬菜日销售量的波动性较大。第二问引入 Spearman 相关系数，以各蔬菜品类及单品蔬菜的日销售量为指标，进行相关性分析，求解得出**除茄类外，其它五种品类蔬菜之间都呈现出显著的正相关关系。**然后以单品的总销售量、每日最大销售量和日均销售量为指标，通过 **K-means聚类算法**将单品划分为热销、平销和滞销三大类，进行相关性分析，结果可得**热销单品较其他单品呈现出高总销售量，高每日最大销售量和高日均销售量特点。**

针对问题二，第一问首先计算出各蔬菜品类每日成本加成定价，然后通过**Pearson相关系数**检验出各蔬菜品类销售总量与成本加成定价**呈负线性相关关系，**并且进行检验，在置信度为95%下均**检验显著**。第二问首先针对不同品类蔬菜建立不相同的**时间序列ARIMA**预测模型，考虑数据的时效性以前半年的日销售量为原始数据预测出未来一周各蔬菜品类的日销售量，再根据预测结果确立决策变量，以利润最大化为目标函数，以蔬菜的补货量、定价、损耗为约束条件建立**优化模型**，求解出**未来一周最大收益为 10091.37 元**，最后引入**蒙特卡洛模拟模型**求出最优的蔬菜品类的补货和定价策略，为实际操作提供了更为灵活和风险可控的决策方案。

针对问题三，首先根据题目要求筛选出可用数据范围，并求出其损失率，以最大化收益为目标函数，以单品种类数量、补货量和定价为约束条件建立新的优化模型求解，筛选出 **28 种**可售单品蔬菜，并估计 **2023 年 7 月 1 日最大收益为 49914.1665 元**。考虑到都是相似的优化模型，便基于问题二的蒙特卡洛模拟模型求解出最优的单品补货量和定价策略。

针对问题四，考虑到商超经营的复杂性及外部竞争环境，本文通过文献检索和市场调研，引入了客流量数据、竞争对手的价格和策略、消费者反馈等关键数据，为制定更为精准的补货和定价策略提供了必要的支持。这些数据的引入，极大地增强了模型的实用性和预测的准确性。

最后，本文对建立的模型进行了全面的讨论和评价，分析了模型在实际应用中的优势与局限性，并提出了未来可能的改进方向。

关键词：K-means聚类算法 成本加成定价 蒙特卡洛模拟 文献检索法

# 问题重述

## 问题背景

在我国经济持续增长和居民消费水平不断提升的背景下，生鲜市场，尤其是蔬菜市场，呈现出显著的增长潜力。蔬菜作为日常饮食的基本组成部分，对保障公众健康和生活质量起到了关键作用。蔬菜的消费不仅频繁，而且具有必需性，使其成为不可忽视的重要市场领域。

然而，蔬菜市场的运营具有其特有的挑战。首先，蔬菜的生物特性决定了其高度易腐败性和保鲜难度，这对商家的库存管理和物流配送提出了高要求。其次，蔬菜种类繁多，且受季节和气候的影响大，导致供应量和价格波动频繁。此外，不同地区的蔬菜成本差异也对定价策略提出了挑战。

为了在这样一个复杂多变的市场中获得竞争优势，商家需要依靠科学的数据分析和市场研究来制定定价和补货策略。通过深入分析消费者购买行为、市场供需状况以及竞争环境，商家可以更精确地预测市场趋势，制定符合市场需求的动态定价体系和优化的库存管理计划。此外，商家还需要考虑到促销活动的时机和组合，以应对市场的季节性变化和消费高峰，从而最大化销售额并提高利润率。有效的市场策略和运营调整将使商家能够不仅满足消费者需求，还能提升市场份额和业绩表现。

## 问题重述

问题一：依据附件 1 中 6 个蔬菜品类的商品信息以及附件 2 中的销售流水明细，合并统计相关数据，分别分析蔬菜各品类、蔬菜单品销售量的分布规律、相互关系，判断不同品类或不同单品之间是否存在一定的关联关系。

问题二：结合相关文献以及附件 1、2、3、4，计算出各蔬菜品类的成本加成定价，再分析其与各蔬菜品类销售总量之间的关系。同时以利润最大化为目标，制定出各蔬菜品类未来一周（2023 年 7 月 1 日-7 日）的日补货总量和定价策略。

问题三：从附件 2 中找出 2023 年 6 月 24-30 日的可售品种，按照单品订购量（销售总量）不少于 2.5 千克等要求，筛选出 27-33 个可售单品，并据此制定 7 月1 日的单品补货量和定价策略，尽量满足市场需求，获得最大收益。

问题四：整合分析 4 个附件中的数据，从季节、节日、销售形式等角度观察数据特征，整理出对于制定蔬菜商品的补货和定价决策有较大影响的相关数据，并给出选择理由，提出合理意见。

# 问题分析

## 问题一的分析

针对问题一的第一小问，首先对涉及到的数据进行数据预处理，剔除附件二的退货数据。为了更好地展示蔬菜品类与单品的分布规律，对品类与单品分别进行描述性统计分析，引入总值，平均值，最大值，最小值，标准差，偏度与峰度。除此之外，还对品类数据进行了季节性的时间序列分析并可视化。对于第二小问，计算蔬菜六大品类之间的斯皮尔曼相关系数以研究之间的相互关系；对单品则是根据其总销售量，采用Kmeans聚类模型将单品聚成热销，平销和滞销三大类。

## 问题二的分析

针对问题二，第一小问先根据成本加成定价公式计算得出各蔬菜品类三年来每日

定价情况，然后引入 Pearson 相关系数判定各蔬菜品类的销售总量与成本加成定价的线性关系。第二小问针对不同品类蔬菜建立合适的时间序列预测模型，考虑到数据的时效性，本文首先以前半年日销售量为原始数据预测未来一周各蔬菜品类的日销售量，然后根据预测结果建立优化模型。最后引入蒙特卡洛模拟模型求出最优的蔬菜品类的补货和定价策略。

## 问题三的分析

针对问题三，题目要求根据 2023 年 6 月 24-30 日的可售品种，按照单品订购量（销售总量）不少于 2.5 千克等要求，筛选出 27-33 个可售单品。以最大收益为目标，制定 7 月 1 日的单品补货量和定价策略。本题仍然为优化问题中的线性规划问题，由此本题基于问题二构建的蒙特卡洛模拟模型。首先根据题意筛选出可用的数据范围，确定目标函数及其约束条件，求出最优的单品补货量和定价策略

## 问题四的分析

针对问题四，通过文献检索法综合引入客流量数据、竞争对手的价格和策略数据、消费者反馈和评价数据等 10 种相关数据，据此制定合适的蔬菜补货和定价策略。

# 模型假设

1．假设蔬菜的历史销售数据能够代表未来的销售趋势，本文使用历史销售数据来预测未来的销售。

2．假设蔬菜的损耗是固定的，本文使用损耗率来计算蔬菜的单位成本。

3．商超的销售空间是固定的，我们假设商超的销售空间在一段时间内是恒定的。

# 符号说明

|  |  |
| --- | --- |
| **符号** | **说明** |
|  | 第种商品在第周的销售量 |
|  | 第种商品在第周的销售价格 |
|  | 第种商品的成本 |
|  | 第种商品在第周的补货量 |
|  | 第种商品的损耗率 |
|  | 第种商品的基础成本 |
|  | 第种商品的价格加成比率 |

# 问题一模型的建立与求解

## 数据预处理

附件二中存在单品的销售数据与退货数据，考虑到退货会直接影响到总销售额的计算和销售数据的准确性，且导致部分单品的销售单价为负数，剔除其中部分单品含有的退货数据。

## 第一问：品类与单品分布规律问题

### 统计性描述分析

为了清楚直观地反应出蔬菜品类与各单品的分布规律，首先引入了总值，平均值，最大值，最小值，标准差，偏度与峰度对其进行基本的描述性统计分析。

**（1）总值（总销售量）**

总值是所有数据值的总和，对于了解数据集的总量大小非常直观。

**（2）均值（日销售量平均值）**

均值是所有数据点的总和除以数据点的数量，用于描述数据集的中心位置，是最常用的中心趋势度量。

**（3）最大值与最小值（日销售量最大值与日销售量最小值）**

最大值和最小值分别是数据集中的最高点和最低点，定义了数据的范围，可以展示数据的极端情况。

**（4）标准差（总销售量标准差）**

标准差是衡量数据集中数值分散程度的统计量，是各数据偏离平均数的平均程度。一般标准差越大，表示数据的波动越大，即数据点与平均值的偏离程度较大；标准差小则表明数据较为集中。其计算公式为：

|  |  |
| --- | --- |
|  | () |

其中：是样本标准差，是样本大小，是每个观测值，是样本的平均值。

**（5）偏度Skewness（总销售量偏度）**

偏度是衡量数据分布对称性的指标。它衡量数据分布的形状，相对于正态分布的偏斜方向和程度。如果偏度值大于0，数据分布呈现右偏（长尾在右）；如果偏度值小于0，则数据分布左偏（长尾在左）。其计算公式为：

|  |  |
| --- | --- |
| Skewness | () |

其中：是样本大小，是每个观测值，是样本的平均值，是样本的标准差。

**（6）峰度Kurtosis（总销售量峰度）**

峰度是描述数据分布尖锐度的统计量，反映数据分布峰部的尖峭程度及尾部的厚度。正态分布的峰度为3，一般当峰度大于3称为尖峰态，峰度小于3称为扁平态。其计算公式为：

|  |  |
| --- | --- |
|  | () |

其中：是样本大小，是每个观测值，是样本的平均值，是样本的标准差。

基于以上，对蔬菜品类与单品分别进行描述性统计分析，所得结果如下所示。

对于**蔬菜六大品类**：

表1：蔬菜六大品类描述性统计分析表

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 指标 | 水生根茎类 | 花叶类 | 花菜类 | 茄类 | 辣椒类 | 食用菌类 |
| 总销售量 | 40607.55 | 198659.55 | 41789.78 | 22442.12 | 91645.11 | 76131.73 |
| 日销售量平均值 | 0.69 | 0.60 | 0.48 | 0.50 | 0.44 | 0.51 |
| 日销售量最大值 | 17.00 | 160.00 | 12.49 | 12.50 | 10.0140 | 25.00 |
| 日销售量最小值 | 0.05 | 0.003 | 0.022 | 0.05 | 0.004 | 0.011 |
| 标准差 | 0.55 | 0.46 | 0.20 | 0.23 | 0.31 | 0.37 |
| 偏度 | 5.91 | 77.54 | 6.03 | 5.04 | 1.42 | 3.20 |
| 峰度 | 77.54 | 42368.74 | 236.90 | 166.57 | 14.57 | 152.63 |

通过分析以上计算结果可知，蔬菜的六大品类之间存在着显著的差异。**花叶类**以198,659.55的总销售量领先，远超其他品类，显示出市场需求的强劲和消费者的高度偏好。相比之下，**水生根茎**类的总销售量仅为40,607.55，是六个品类中**最低**的，这可能反映了较低的市场需求或供应限制。在日销售量平均值方面，水生根茎类以0.69领先，尽管总量最低，表明其日常销售比其他品类更稳定。此外，花叶类的日销售最大值达到160，远高于其他品类的最大销售记录，如辣椒类的10.014和茄类的12.50，这揭示了花叶类在某些高峰期的极高销售潜力。

从统计参数来看，所有品类的销售数据都呈现出**右偏分布**，特别是花叶类的偏度达到77.54，意味着尽管平均日销售量较低，偶尔会出现极高的销售峰值。这种分布特性指示着销售数据中存在极端值。**花叶类的峰度高达42,368.74**，远超其他品类，表明其销售数据在**极端高销量日**的频率和强度都显著高于常态，这可能对库存管理和预测带来挑战。

**对于蔬菜单品：**由于单品蔬菜种类颇多，此处仅选取各类蔬菜中总销售量最大的两种单品蔬菜进行展示（其余单品见附件）。

表2：各类蔬菜中总销售量最大的两种单品蔬菜的统计分析表

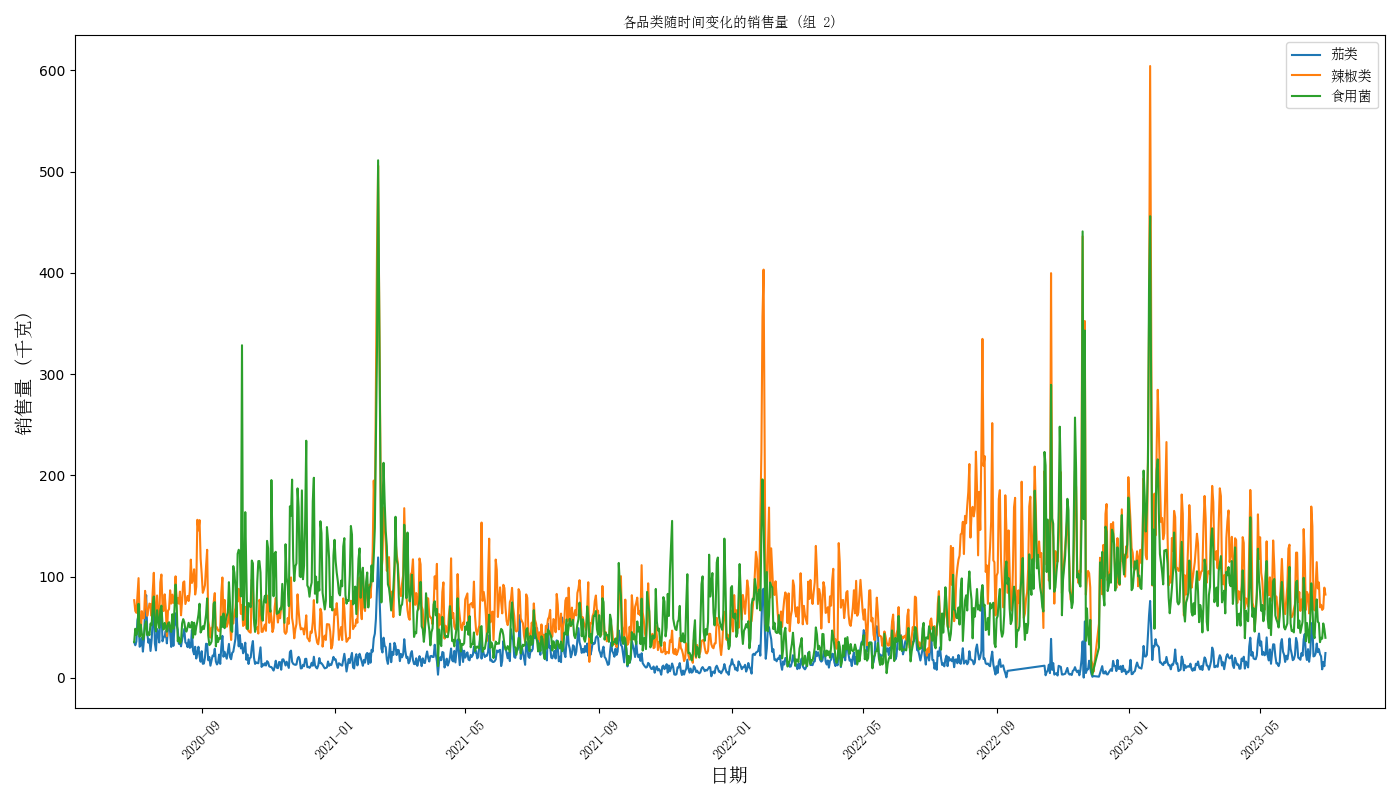
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 水生根茎类 | 花叶类 | 花菜类 | 茄类 | 辣椒类 | 食用菌类 |
| 指标 | 净藕(1) | 大白菜 | 西兰花 | 紫茄子(2) | 芜湖青椒(1) | 西峡香菇(1) |
| 总销售量 | 27166.46 | 19196.10 | 27555.95 | 13608.98 | 28181.74 | 11929.10 |
| 日销售量平均值 | 0.69 | 1.27 | 0.47 | 0.50 | 0.40 | 0.25 |
| 日销售量最大值 | 14.00 | 14.59 | 12.487 | 12.50 | 10.00 | 25.00 |
| 日销售量最小值 | 0.057 | 0.051 | 0.022 | 0.107 | 0.017 | 0.015 |
| 标准差 | 0.42 | 0.72 | 0.21 | 0.23 | 0.20 | 0.20 |
| 偏度 | 2.72 | 6.45 | 6.73 | 6.67 | 3.14 | 55.76 |
| 峰度 | 33.71 | 82.11 | 266.98 | 265.85 | 79.20 | 6278.36 |
| 指标 | 金针菇(盒) | 云南生菜 | 青梗散花 | 青茄子(1) | 小米椒(份) | 双孢菇(盒) |
| 总销售量 | 15602.00 | 15915.42 | 8394.38 | 3519.01 | 10847.00 | 4233.00 |
| 日销售量平均值 | 1.00 | 0.40 | 0.55 | 0.59 | 1.00 | 1.00 |
| 日销售量最大值 | 4.00 | 6.00 | 6.50 | 2.48 | 9.00 | 2.00 |
| 日销售量最小值 | 1.00 | 0.05 | 0.08 | 0.17 | 1.00 | 1.00 |
| 标准差 | 0.05 | 0.19 | 0.18 | 0.26 | 0.09 | 0.03 |
| 偏度 | 50.01 | 3.22 | 3.29 | 1.87 | 83.77 | 32.47 |
| 峰度 | 2730.62 | 57.58 | 84.10 | 5.55 | 7362.93 | 1052.25 |

数据显示，**西兰花**和**芜湖青椒**在其相应的蔬菜类别中表现出较高的市场需求，总销售量分别为27,555.95和28,181.74，这表明它们在市场上的受欢迎程度较高。此外，**青茄子**和**双孢菇**的总销售量相对较低，分别仅为3,519.01和4,233.00。在日销售量的平均值方面，**大白菜**和**小米椒**的日均销售量最高，分别达到1.27和1.00，表示了它们在日常市场中的稳定需求。

在本数据集中**金针菇**的标准差最低，为0.05，显示出其销售量的极高稳定性。相反，西兰花的标准差为0.21，表明其销售数据相对波动较大。西峡香菇的偏度高达55.76，峰度更是达到6278.36，表明其销售量虽然通常较低，但会偶尔出现极高的销售峰值。此外，小米椒的偏度和峰度分别为83.77和7362.93，这两单品数据进一步确认了销售数据的极端集中和偶发的高销量事件。

### 时间序列分析

鉴于蔬菜单品的销售量数据是从 2020 年 7 月至 2023 年 6 月 30 日，时间跨度比较大，对蔬菜六大品类的数据作曲线图来观察日销售量的分布情况，得到以下的曲线图。



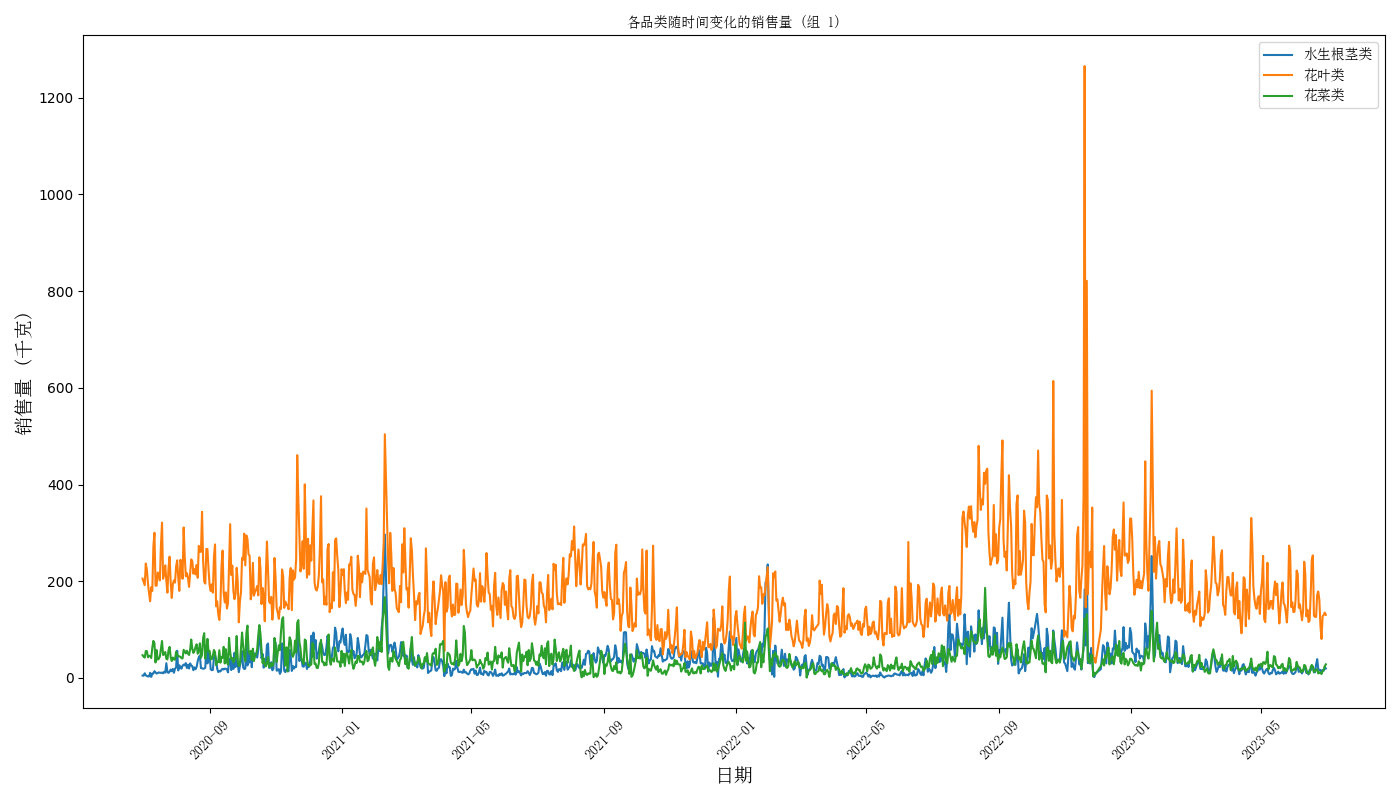


图1：蔬菜六大品类日销量分布图

从图中可以看出，各品类的销售表现具有明显的波动性和季节性特征。具体来说，每种蔬菜的销售量都有高峰和低谷。**水生根茎类**和花菜类的销售曲线较为平缓，显示其市场需求相对稳定，**食用菌类，花叶类，茄类**和**辣椒类**则展示了更加剧烈的销售波动，并且尖峰明显，表明在某些特定时期内需求显著增加。

## 蔬菜品类及单品的相互关系

### 蔬菜六大品类之间的相互关系

为了探寻蔬菜六大品类之间的相互关系，且蔬菜各品类的销售量为连续变量，因此采用 Spearman 相关系数用于分析蔬菜各品类销售量的相关关系。

斯皮尔曼相关系数（Spearman's rank correlation coefficient），是一种非参数的相关系数，用于衡量两个变量的秩次之间的统计相关性。它不假设数据服从正态分布，且不仅限于线性关系，而是可以用来度量任何单调函数关系的强度和方向。

假设有两个变量和，它们的观测值分别为和。首先，将每个变量的数据按照大小进行排序，分配秩次。如果数据中有相同的值，则分配平均秩次。斯皮尔曼秩相关系数的计算公式为:

|  |  |
| --- | --- |
|  | () |

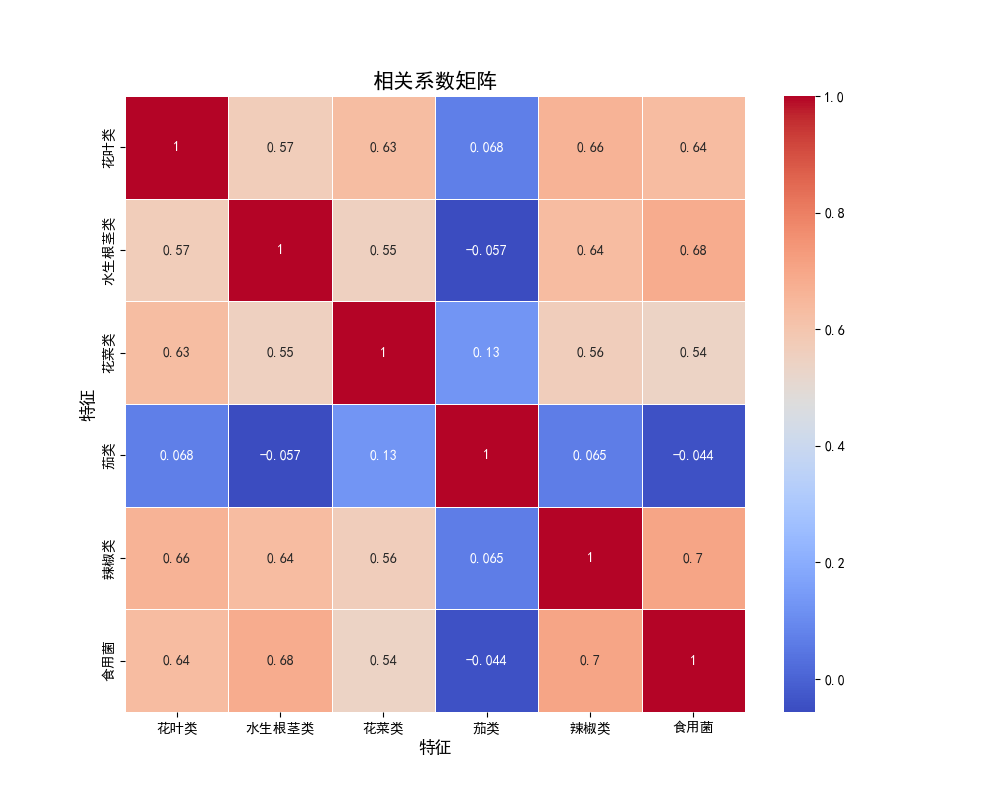
其中，是每对观测值的秩次差 (即  是观测值的数量。

利用 Python 求解蔬菜六大品类的 Spearman 相关系数如下表所示：

表3：蔬菜六大品类的 Spearman 相关系数表

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 花叶类 | 水生根茎类 | 花菜类 | 茄类 | 辣椒类 | 食用菌类 |
| 花叶类 | 1.00 | 0.58 | 0.63 | 0.08 | 0.66 | 0.64 |
| 水生根茎类 | 0.57 | 1.00 | 0.55 | -0.06 | 0.64 | 0.68 |
| 花菜类 | 0.63 | 0.55 | 1.00 | 0.13 | 0.56 | 0.54 |
| 茄类 | 0.07 | -0.06 | 0.13 | 1.00 | 0.06 | -0.04 |
| 辣椒类 | 0.66 | 0.64 | 0.56 | 0.06 | 1.00 | 0.70 |
| 食用菌类 | 0.64 | 0.68 | 0.54 | -0.04 | 0.70 | 1.00 |

绘制的六大品类蔬菜之间相关性热图如下所示：



图X:蔬菜六大品类相关系数热图

分析以上结果，我们可出结论：

1. 花叶类与水生根茎类，花菜类，辣椒类和食用菌类相关系数较高，呈正相关，但与茄类的相关系数只有0.08，几乎无相关。
2. 水生根茎类与辣椒类和食用菌类的相关系数分别为0.64与0.68，具有较强的正相关。与花菜类具有中等程度的正相关，但与茄类的的相关系数为-0.06，呈现轻微负相关。
3. 茄类除了与花菜类的低正相关外，与其他品类的相关性非常低或接近无相关，尤其是与食用菌类的轻微负相关（-0.04）。
4. 辣椒类与食用菌类之间的相关系数为0.70，是所有品类对中最高的相关系数之一。
5. 辣椒类和食用菌类与花菜类的相关系数分别为0.56和0.54，显示中等正相关。

### 蔬菜单品之间的相互关系

为了研究蔬菜单品间的相互关系，且考虑到单品数据特性，本次建模采用Kmeans聚类算法，将蔬菜单品聚为X类，分别为。

K-Means 聚类算法是一种广泛使用的无监督学习方法，适用于将数据分为预定数量的簇或组。其核心思想是将 n 个数据点分配到 k 个簇中，使得每个数据点属于离它最近的中心（质心）所代表的簇，从而以此准则最小化簇内的平方误差。该算法的计算步骤如下：

**步骤1初始化：**随机选择 k 个数据点作为初始质心。

**步骤2分配：**每个数据点被分配到最近的质心所代表的簇。计算数据点到每个质心的距离，并选择最小的那个。距离通常使用欧氏距离公式计算：

|  |  |
| --- | --- |
|  | () |

其中， 是数据点，是质心。

**步骤3更新质点：**在确定了簇的成员后，下一步是更新每个簇的质心，以便质心是簇中所有点的均值。更新的质心计算公式是：

|  |  |
| --- | --- |
|  | () |

其中，是第个簇，是簇中的点数，是簇中所有点的和。

**步骤4迭代：**重复步骤 2 和步骤 3 直到满足停止条件。停止条件通常是质心的变化小于某个阈值，或者达到了预设的最大迭代次数。这一步通常使用簇内误差平方和SSE来检查收敛，其计算公式为：

|  |  |
| --- | --- |
|  | () |

其中，是点到其质心的欧氏距离。

以单品的总销售量、每日最大销售量和日均销售量为指标，使用聚类算法进行分类，将单品分为热销、平销和滞销四大类。

表4：聚类中心

|  |  |  |  |
| --- | --- | --- | --- |
|  | 总销量(千克) | 日均销售量(千克) | 日最大销售量(千克) |
| 聚类中心1 | 552.1566863 | 3.26541862 | 14.06373039 |
| 聚类中心2 | 5965.412485 | 15.51047435 | 98.1600303 |
| 聚类中心3 | 17941.92689 | 30.45054507 | 221.8542222 |

随后求解各单品蔬菜的聚类结果如下图所示：

|  |  |  |
| --- | --- | --- |
|  |  |  |

图2：单品蔬菜聚类散点图

由于篇幅问题，这里仅仅展示部分单品蔬菜的聚类结果，如下表所示：

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 序号 | 名称 | 类别 | 序号 | 名称 | 类别 | 序号 | 名称 | 类别 |
| 1 | 牛首生菜 | 滞销 | 2 | 四川红香椿 | 滞销 | 3 | 西峡花菇(1) | 平销 |
| 4 | 白菜苔 | 滞销 | 5 | 苋菜 | 滞销 | 6 | 云南生菜 | 平销 |
| 7 | 竹叶菜 | 平销 | 8 | 小白菜 | 滞销 | 9 | 净藕(1) | 热销 |
| 10 | 上海青 | 平销 | 11 | 娃娃菜 | 热销 | 12 | 牛首油菜 | 平销 |
| 13 | 茼蒿 | 滞销 | 14 | 蔡甸藜蒿 | 滞销 | 15 | 菜心 | 滞销 |
| 16 | 木耳菜 | 滞销 | 17 | 大白菜 | 热销 | 18 | 泡泡椒(精品) | 热销 |
| 19 | 云南油麦菜 | 平销 | 20 | 红尖椒 | 滞销 | 21 | 青尖椒 | 滞销 |
| 22 | 红椒(1) | 平销 | 23 | 紫茄子(2) | 平销 | 24 | 青茄子(1) | 滞销 |
| 25 | 芜湖青椒(1) | 热销 | 26 | 金针菇(1) | 平销 | 27 | 小米椒(份) | 热销 |
| 28 | 西兰花 | 热销 | 29 | 金针菇(盒) | 热销 | 30 | 黑油菜 | 滞销 |

表5：单品蔬菜聚类

各单品的销售量相关性分析结果：

1.**热销类别中的单品通常具有高总销售量、高每日最大销售量和高日均销售量。**这意味着这些单品销售非常火爆，并且在销售指标上表现出相对一致的趋势。热销类别内的单品之间可能存在很高的相关性，即它们的销售指标相互影响较大。

2.**平销类别中的单品的销售指标在中等水平上保持稳定，没有明显的高或低**。这些单品之间可能存在较低的相关性，销售指标相对独立。

3.**滞销类别中的单品表现出较低的总销售量、每日最大销售量和日均销售量**。这些单品之间可能存在较弱的相关性，它们的销售指标相互影响较小。

# 问题二模型的建立与求解

## 成本加成定价

成本加成定价是一种简单且广泛使用的定价策略，其基本公式为：

|  |  |
| --- | --- |
|  | () |

其中成本包括直接成本和间接成本，公式表示为：

|  |  |
| --- | --- |
|  | () |

加成比率则是根据期望利润率设定的，公式表示为：

|  |  |
| --- | --- |
|  | () |

## 销售总量与成本加成定价的关系

### Pearson相关系数

根据以上公式计算出成本加成定价。为了得到蔬菜六大品类的销售总量与成本加成定价之间的关系，分别计算其Pearson相关系数。

Pearson相关系数是用于度量两个变量之间线性关系强度和方向的统计度量。核心思想是通过观测两个变量之间的共变动情况来确定它们是否存在某种线性关系，以及这种关系的强度如何。系数的大小指示关系的强度，接近 ±1 的值表示非常强的线性关系，而接近 0 的值则表示很弱的线性关系。其计算公式表示为：

|  |  |
| --- | --- |
|  | () |

其中：和是两个变量的观测值，和分别是变量X和Y的样本均值，是观测值的数量。

基于此，分别求解蔬菜六大品类销售总量与成本加成定价之间的Pearson相关系数，所得结果如下表所示：

表6：Pearson相关系数

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 品类 | 水生根茎类 | 花叶类 | 花菜类 | 茄类 | 辣椒类 | 食用菌类 |
| 相关系数 | -0.15 | -0.24 | -0.44 | -0.07 | -0.22 | -0.12 |

由上表可知，花菜类、食用菌、辣椒类、茄类、花叶类和水生根茎类的销售总量

与成本加成定价的相关系数都小于 0，并且花菜类表现明显。故各蔬菜品类的销售总量与成本加成定价呈负线性相关关系。

### 相关系数的检验

在计算出Pearson相关系数后，需要对其进行检验。检验步骤如下：

**步骤1）提出假设**

原假设𝐻0：Pearson 相关系数R ≠ 0

备择假设𝐻1：Pearson 相关系数R = 0

设定置信水平为 95%

**步骤2）计算P值**

对蔬菜六大品类的相关系数检验时计算了其P值，结果如下表所示：

表7：Pearson相关系数检验P值

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 品类 | 水生根茎类 | 花叶类 | 花菜类 | 茄类 | 辣椒类 | 食用菌类 |
| P值 | 0.032 | 0.003 | 0.038 | 0.045 | 0.01 | 0.02 |

由表 7 可知，六大蔬菜品类的销售量与其成本加成定价的 P 值均小于置信水平 0.05，则拒绝原假设，即六大蔬菜品类的销售量与其成本加成定价均呈现显著的线性相关关系。

## 优化模型制定单周日补货总量和定价策略

首先建立预测模型，使用最近半年内的数据进行预测，预测未来一周（即2023-7-1到2023-7-7日）的日销售量，根据预测结果构建目标函数，建立优化模型，最后通过蒙特卡洛求解出各蔬菜品类的最优补货和定价策略。

### 利用时间序列预测蔬菜品类未来销量

本文使用SPSS软件建立时间序列模型，在传统模型的基础上来预测各蔬菜品类未来7天的日销售量。

首先统计各品类在近半年的日均销售量，检查缺失值、异常值后统一时间格式，随后做出六大品类销售量随时间变化图。观察变化图发现，**花菜类、辣椒类、茄类、食用菌类、水生根茎类季节性明显**，呈现季节趋势，但趋势性较弱，可建立简单季节性模型；**花叶类具有明显趋势性，但季节成分不稳定**，可建立温特斯乘性模型。

根据观察分析，依据专家建模器建立模型如下：

表8：蔬菜品类时间序列模型表

|  |  |  |
| --- | --- | --- |
| 品类 | 时间序列模型类型 | 参数值 |
| 花菜类 | ARIMA（p,d,q） | p=1,d=1,q=1 |
| 辣椒类 | ARIMA（p,d,q） | p=5,d=0,q=2 |
| 茄类 | ARIMA（p,d,q） | p=1,d=1,q=7 |
| 食用菌类 | ARIMA（p,d,q） | p=0,d=1,q=9 |
| 水生根茎类 | ARIMA（p,d,q） | p=1,d=1,q=3 |
| 花叶类 | ARIMA（p,d,q） | p=1,d=0,q=9 |

### 建立优化模型

本模型的优化目标为利润最大化，其数学表达式为：

|  |  |
| --- | --- |
|  | () |

其中各符号含义表示为：：第种商品在第周的销售量，：第种商品在第周的销售价格，：第种商品的批发成本，：第种商品在第周的补货量，：第种商品的损耗率。

在确立约束条件时，主要考虑以下两个要点：

1. **销售价格约束：**

为了确保销售价格不低于成本加成定价，需要进行该约束，即：

|  |  |
| --- | --- |
| ， | () |

其中表示加成比率。

1. **补货量的约束：**

为了确保经过调整损耗率后的实际可供销售的商品数量不少于预期的销售需求，即：

|  |  |
| --- | --- |
| ， | () |

最终整合得到确定的优化模型为：

|  |  |
| --- | --- |
|  | () |

### 蒙特卡洛模拟得出最优策略

蒙特卡洛模拟是一种计算数学技术，用于理解复杂系统或数学函数的概率分布特征，通过使用随机抽样技术来进行模拟和估计各种概率事件的结果。其核心原理是利用伪随机数来模拟或近似复杂过程或数学表达式的结果。通过重复随机抽样，并计算结果的平均值或其他统计量，可以估计出一个未知量的概率分布。该算法的核心步骤如下所示：

**步骤1）定义问题**

**步骤2）构建概率模型**：基于问题定义，构建一个或多个概率模型来描述与问题相关的随机变量。

**步骤3）进行随机抽样：**从定义的概率分布中生成随机样本。

**步骤4）运行模拟**：对于每一个随机生成的样本，根据模型计算输出结果。

**步骤5）聚合结果并解释：**收集并分析所有模拟的输出结果并基于聚合的统计结果，解释或预测关于问题的解答。

### 利用以上模型进行求解

根据所建模型为近半年各蔬菜品类销售情况在2023-7-1到2023-7-7做出预测，结果如下：

表9：蔬菜品类在2023-7-1到2023-7-7的销售量预测表

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 品类 | 2023.7.1 | 2023.7.2 | 2023.7.3 | 2023.7.4 | 2023.7.5 | 2023.7.6 | 2023.7.7 |
| 花菜类 | 132.442 | 152.925 | 161.786 | 165.094 | 176.674 | 176.180 | 188.008 |
| 辣椒类 | 141.322 | 146.773 | 147.079 | 147.347 | 147.582 | 147.788 | 147.969 |
| 茄类 | 44.768 | 47.626 | 46.860 | 48.020 | 51.727 | 53.821 | 54.383 |
| 食用菌类 | 21.068 | 16.493 | 13.348 | 14.625 | 15.258 | 18.968 | 19.373 |
| 水生根茎类 | 89.638 | 92.133 | 87.910 | 88.681 | 92.920 | 98.014 | 102.990 |
| 花叶类 | 23.034 | 20.639 | 19.504 | 18.966 | 18.711 | 18.590 | 18.533 |

为了更加直观地展示出品类在2023-7-1到2023-7-7的销售量预测结果，做了相应散点图如下：

|  |  |
| --- | --- |
|  |  |

图3：花菜类（左）和花叶类（右）

|  |  |
| --- | --- |
|  |  |

图4:辣椒类（左）和茄类（右）

|  |  |
| --- | --- |
|  |  |

5X：食用菌类（左）和水生根茎类（右）

未来一周各蔬菜品类日销量预测如下表：

表10：各蔬菜品类在 2023 年 7 月 1 日-7 日的日补货总量

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 日期 | 花菜类 | 辣椒类 | 茄类 | 食用菌类 | 水生根茎类 | | 花叶类 |
| 2023-7-1 | 151.88 | 161.08 | 50.27 | 22.50 | | 98.42 | 25.01 |
| 2023-7-2 | 175.37 | 167.29 | 53.48 | 17.61 | | 101.15 | 22.41 |
| 2023-7-3 | 185.53 | 167.64 | 52.62 | 14.26 | | 96.62 | 21.18 |
| 2023-7-4 | 189.32 | 167.94 | 53.92 | 15.62 | | 97.36 | 20.59 |
| 2023-7-5 | 202.60 | 168.21 | 58.09 | 16.30 | | 102.02 | 20.31 |
| 2023-7-6 | 202.04 | 168.45 | 60.44 | 20.26 | | 107.61 | 20.18 |
| 2023-7-7 | 215.60 | 168.65 | 61.07 | 20.69 | | 113.08 | 20.12 |

根据优化模型求解得未来一周最大收益为 10091.37 元。根据优化目标，建立蒙特卡洛模拟模型，求解得到 2023 年 7 月 1 日-7 日各蔬菜品类日补货量(单位:kg)与定价策略如下：

表10：2023 年 7 月 1 日—7 日各蔬菜品类日补货量(单位:kg)与定价策略

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 日期 | 花菜类 | | 辣椒类 | | 茄类 | |
| 日补货量 | 定价 | 日补货量 | 定价 | 日补货量 | 定价 |
| 2023-7-1 | 151.88 | 11.48 | 161.08 | 7.28 | 50.27 | 9.33 |
| 2023-7-2 | 175.37 | 11.48 | 167.29 | 7.28 | 53.48 | 9.33 |
| 2023-7-3 | 185.53 | 11.48 | 167.64 | 7.28 | 52.62 | 9.33 |
| 2023-7-4 | 189.32 | 11.48 | 167.94 | 7.28 | 53.92 | 9.33 |
| 2023-7-5 | 202.60 | 11.48 | 168.21 | 7.28 | 58.09 | 9.33 |
| 2023-7-6 | 202.04 | 11.48 | 168.45 | 7.28 | 60.44 | 9.33 |
| 2023-7-7 | 215.60 | 11.48 | 168.65 | 7.28 | 61.07 | 9.33 |
| 日期 | 食用菌类 | | 水生根茎类 | | 花叶类 | |
| 日补货量 | 定价 | 日补货量 | 定价 | 日补货量 | 定价 |
| 2023-7-1 | 22.50 | 9.25 | 98.42 | 12.52 | 25.01 | 9.17 |
| 2023-7-2 | 17.61 | 9.25 | 101.15 | 12.52 | 22.41 | 9.17 |
| 2023-7-3 | 14.26 | 9.25 | 96.62 | 12.52 | 21.18 | 9.17 |
| 2023-7-4 | 15.62 | 9.25 | 97.36 | 12.52 | 20.59 | 9.17 |
| 2023-7-5 | 16.30 | 9.22 | 102.02 | 12.52 | 20.31 | 9.15 |
| 2023-7-6 | 20.26 | 9.22 | 107.61 | 12.52 | 20.18 | 9.15 |
| 2023-7-7 | 20.69 | 9.21 | 113.08 | 12.52 | 20.12 | 9.14 |

# 问题三模型的建立与求解

## 数据预处理

根据题意，筛选出2023.6.24日到2023.6.30日的可售单品作为补货选择，进一步框选补货对象，制定更加合理的补货与定价策略。同时，筛选出各个补货单品的批发价格与损耗率，求得最大收益。

共筛选出该期间单品（49件）、批发价、损耗率如下表所示（仅部分展示，其余见附录）：

表11：单品蔬菜每日批发单价与损耗率表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 日期 | 苋菜 | 菠菜 | 娃娃菜 | 青红杭椒 | 七彩椒 |
| 2023-6-24 | 2.31 | 3.9 | 4.75 | 3.53 | 11.92 |
| 2023-6-25 | 2.34 | 0 | 4.83 | 3.44 | 11.93 |
| 2023-6-26 | 2.33 | 0 | 4.84 | 3.36 | 11.92 |
| 2023-6-27 | 2.34 | 4.13 | 4.85 | 3.06 | 12.53 |
| 2023-6-28 | 2.37 | 4.19 | 4.85 | 2.96 | 12.84 |
| 2023-6-29 | 2.37 | 4.2 | 4.6 | 2.95 | 12.13 |
| 2023-6-30 | 2.21 | 4.07 | 4.4 | 3.32 | 0 |
| 损失率 | 18.52% | 9.43% | 2.48% | 9.43% | 9.43% |

说明：上表中单品蔬菜批发单价为 0，即当天为对其进行补货。

## 优化模型制定单日日补货总量和定价策略

1. **确定目标函数：**

要使2023.7.1日的总收益最大，目标函数为：

|  |  |
| --- | --- |
|  | () |

其中表示第个单品，取值为0或1，时，表示选择补货该单品，时，表示不补货该单品，其他变量同前文所述。

1. **确定约束条件：**

各单品订购量满足最小陈列量 2.5 千克，即补货量最低2.5千克，故有

|  |  |
| --- | --- |
|  | () |

要求可售单品总数控制在27-33个, 对销售单品总数做了限制, 单品总数为 , 故有:

|  |  |
| --- | --- |
|  | () |

防止补货无限制导致收益不收敛, 因此需对补货量做上限要求, 故有:

|  |  |
| --- | --- |
|  | () |

1. **总体优化模型：**

|  |  |
| --- | --- |
|  | () |

### 模型说明

在问题三的背景下，本题仍然为优化问题中的线性规划问题，所以本题仍选用蒙特卡洛模拟模型求解。

### 蒙特卡洛模拟模型求解补货及定价策略

根据优化模型筛选出 28 种可售单品，并求解得 2023 年 7 月 1 日最大收益为

4994.1665 元，根据优化目标，建立模拟退火模型， 采用 Python 软件求解。结果如下：

表12：单品蔬菜的补货量与定价策略

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 单品蔬菜名称 | 单位成本 | 日补货量 | 定价 | 收益 |
| 1 | 红莲藕带 | 5.60 | 2.50 | 9.09 | 224.11 |
| 2 | 金针菇(盒) | 1.45 | 18.02 | 1.80 | 5.49 |
| 3 | 双孢菇(盒) | 3.40 | 9.19 | 5.02 | 8.64 |
| 4 | 奶白菜 | 2.53 | 7.59 | 5.02 | 282.00 |
| 5 | 竹叶菜 | 2.32 | 12.85 | 3.68 | 388.42 |
| 6 | 青线椒(份) | 2.83 | 2.50 | 4.40 | 62.80 |
| 7 | 菠菜(份) | 4.10 | 10.08 | 5.55 | 374.89 |
| 8 | 净藕(1) | 10.75 | 5.94 | 13.67 | 336.41 |
| 9 | 菜心 | 4.61 | 2.50 | 6.07 | 154.15 |
| 10 | 海鲜菇(包) | 1.95 | 8.20 | 2.84 | 7.23 |
| 11 | 野生粉藕 | 16.06 | 2.50 | 26.01 | 484.75 |
| 12 | 七彩椒(2) | 12.21 | 2.50 | 20.05 | 268.30 |
| 13 | 木耳菜(份) | 1.57 | 2.50 | 3.44 | 32.25 |
| 14 | 枝江青梗散花 | 9.36 | 6.23 | 13.03 | 526.45 |
| 15 | 青茄子(1) | 4.05 | 3.30 | 5.72 | 61.39 |
| 16 | 云南生菜 | 5.73 | 2.67 | 8.80 | 225.21 |
| 17 | 云南油麦菜 | 3.87 | 2.50 | 7.53 | 114.91 |
| 18 | 蟹味菇与白玉菇双拼(盒) | 3.21 | 2.50 | 4.78 | 2.82 |
| 19 | 小皱皮(份) | 1.54 | 11.59 | 2.67 | 155.56 |
| 20 | 紫茄子(2) | 3.75 | 11.84 | 6.20 | 240.27 |
| 21 | 白玉菇(袋) | 3.50 | 2.50 | 5.89 | 51.55 |
| 22 | 红薯尖 | 3.20 | 4.70 | 5.56 | 115.78 |
| 23 | 鲜木耳(份) | 1.30 | 3.62 | 2.45 | 40.22 |
| 24 | 螺丝椒(份) | 3.28 | 10.89 | 4.83 | 319.74 |
| 25 | 高瓜(2) | 13.47 | 2.50 | 16.33 | 310.29 |
| 26 | 圆茄子(2) | 4.04 | 2.50 | 6.80 | 60.96 |
| 27 | 紫茄子(1) | 4.34 | 2.50 | 9.15 | 90.28 |
| 28 | 娃娃菜 | 4.73 | 8.16 | 6.53 | 81.06 |

# 问题四模型的建立与求解

为了更好地制定蔬菜商品的补货和定价决策，商超需要采集以下相关数据，并基于这些数据进行分析和决策：

**1. 客流量数据**

采集内容：记录不同时间段、日期的客流量变化，特别是高峰和低谷时段的详细数据。

分析与决策：客流量的变化直接影响到蔬菜的销售量，通过掌握客流量的高低峰，商超可以灵活调整商品陈列策略、促销活动的时间点以及进货量。这样可以有效提升顾客满意度，最大化销售机会，并避免商品的积压和损耗。

**2. 竞争对手的价格和销售策略信息**

采集内容：收集竞争对手的定价信息、促销活动、商品组合、市场定位等详细数据。

分析与决策：在特定商圈内，竞争对手的数量和实力会显著影响商超的定价策略。通过对比分析竞争对手的定价模式和促销策略，商超可以优化自身的价格定位和促销方案，确保在市场上保持竞争力。同时，深入研究竞争对手的市场策略，还可以帮助商超识别潜在的市场机会，调整自身经营模式，以吸引更多顾客。

**3. 消费者反馈和评价数据**

采集内容：系统收集顾客对蔬菜新鲜度、品质、价格等方面的反馈和评价。

分析与决策：通过分析顾客的反馈，商超可以深入了解顾客的需求和偏好，并据此做出针对性的销售调整。首先，顾客对新鲜度的评价可以帮助商超改进商品的储存和运输方式，确保商品以最佳状态售出。其次，顾客对商品品质的反馈有助于商超优化进货渠道。最后，价格反馈则有助于调整定价策略，使其更加符合顾客的期望和市场需求。

**4. 顾客需求数据**

采集内容：通过顾客调查、购买记录以及反馈，收集顾客对蔬菜种类、包装规格、购买频率等方面的需求数据。

分析与决策：了解顾客的需求和偏好，商超可以优化产品的陈列组合，增加受欢迎的蔬菜种类和规格，确保供应多样化的商品。根据顾客对品质和有机认证等需求的关注，商超还可以选择优质的供应商，并合理制定定价策略，以提升商品的吸引力和竞争力。此外，通过分析购买频率和顾客偏好，商超可以设计更具针对性的促销活动，提升顾客的忠诚度和购买欲望。

**5. 天气数据**

采集内容：收集未来的天气预报和节假日等相关信息。

分析与决策：天气变化对顾客的购物行为影响显著，商超可以基于天气数据，提前调整库存和促销策略。例如，暴雨天气下顾客可能更倾向于在线购物，商超可以提前准备，确保线上渠道的库存充足。此外，在节假日期间，商超可以推出针对性的促销活动，吸引更多消费者。

**6. 供应链和物流数据**

采集内容：记录供应商的交货时间、产地信息、产量、运输时间以及商品质量等相关数据。

分析与决策：通过评估供应商的交货能力和准时性，商超可以更好地安排库存和补货计划，避免因供货延迟或不足导致的缺货问题。监控供应商提供的商品质量，结合顾客反馈，商超可以优化供应链合作伙伴的选择，确保商品的整体质量，提升顾客满意度。

**7. 库存状况**

采集内容：详细记录蔬菜库存量、库存周转率以及潜在的库存积压情况。

分析与决策：通过实时监控库存状况，商超可以预测和避免库存过剩或短缺。当某种蔬菜库存过多且需求不足时，商超可以采取打折销售或促销活动来降低库存积压风险，确保商品的新鲜度和销售率。而对于库存紧张的蔬菜种类，商超可以考虑适时提高价格或加大补货力度，以维持稳定的供应链和合理的利润。

**8. 财务数据**

采集内容：分析各蔬菜商品的成本、利润率和销售额等关键财务数据。

分析与决策：财务数据是制定定价策略的重要依据。通过了解商品的成本和利润率，商超可以合理设定售价，确保盈利。此外，基于销售数据和市场趋势，商超可以调整商品组合，引入更多高利润率或受欢迎的蔬菜品种，从而优化整体盈利能力。

**9. 退货数据**

采集内容：收集并分析退货原因、退货商品类别及相关统计数据。

分析与决策：退货数据可以揭示商品质量和顾客满意度的问题。通过分析退货原因，商超可以识别并解决潜在的质量问题，提升商品质量。此外，退货数据还可以帮助商超优化商品组合，淘汰不受欢迎的商品，增加更受顾客喜爱的商品，减少退货率，提高顾客满意度。

通过整合这些数据，商超可以更好地制定补货和定价决策。这些数据不仅提供了对市场动态和顾客需求的全面洞察，还为优化供应链、提升商品质量、制定促销策略提供了坚实的基础，最终帮助商超在竞争激烈的市场环境中取得更好的业绩表现。

# 模型的评价、改进与推广

## 模型的优点

1.时间序列预测，对不同品类单独分析，分别建立模型求解，充分利用了各品类近三年的销量信息，对未来预测使用了更据时效性的数据，更加准确有效。

2.蒙特卡洛模拟求解优化模型，对于具有复杂数据，多维度优化问题时表现优异，减少了维度带来的计算复杂度，不会陷入局部最优解，更有机会找到全局最优解。

3.K-means聚类分析数据，相对简单，对于大数据量，能够迅速完成聚类任务，对于均匀分布的数据集，聚类结果更加稳定，适合本题中的多单品的大数据集。

## 模型的缺点

1.K-means容易受到噪声与异常值的干扰，影响聚类结果。

2.蒙特卡洛方法是基于随机性的，使得结果的稳定性和可解释性较差。

## 模型的推广

K-means可以进一步优化为K-means++，不仅限于传统的数据分析任务，它可以应用于图像颜色空间量化、客户以及产品群体分类、社会网络分析、金融领域的风险分级等等。

# 参考文献

[1] 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].2021.

[2] 魏泽娥,陈刚,丁胜春,耿军霞.大规模定制产品的成本加成定价方法研究[J].黑龙江

对外经贸,2007,No.152(02):84-85.

[3] 刘保政,刘德宝,高立群.供不应求季节性商品的价格控制和生产销售决策模型[J].

东北大学学报,2005(11):23-26.

[4] 王艳.中小型连锁超市定价的影响因素及其定价策略探析[J].兰州工业学院学

报,2014,21(03):99-102.

[5] 朱洪文.应用统计[M].北京:高等教育出版社.2004,7,1:205-235.

[6] Estrada G,Elizabeth,Villaseñor, et al. Shapiro-Wilk test for multivariate

skew-normality[J]. Computational Statistics,2022(prepublish).

[7] 薛毅.统计建模与 R 软件（第 1 版）[M].北京:清华大学出版社.2007,4:402-418

[8] Lunn D, Jackson Christopher, et al. A Practical Introduction to Bayesian

Analysis[M].CRC Press,2013.

[9]易丹辉. 数据分析与 EViews 应用[M].中国统计出版社,2002.10.166-179.

[10]袁新生、邵大宏、郁时炼. LINGO 和 Excel 在数学建模中的应用[M].科学出版

社,2007.32

附录

|  |
| --- |
| 附录1 |
| 介绍：第二问蒙特卡洛模拟代码 |
| import pandas as pd import numpy as np # 加载Excel文件 file\_path = r"D:\WeChat Files\wxid\_pph1gnrllef322\FileStorage\File\2024-08\补货与定价.xlsx" data = pd.read\_excel(file\_path) # 预测销售量 predicted\_sales = np.array([  [132.442, 152.925, 161.786, 165.094, 176.674, 176.180, 188.008], # 花菜类  [141.322, 146.773, 147.079, 147.347, 147.582, 147.788, 147.969], # 辣椒类  [44.768, 47.626, 46.860, 48.020, 51.727, 53.821, 54.383], # 茄类  [21.068, 16.493, 13.348, 14.625, 15.258, 18.968, 19.373], # 食用菌类  [89.638, 92.133, 87.910, 88.681, 92.920, 98.014, 102.990], # 水生根茎类  [23.034, 20.639, 19.504, 18.966, 18.711, 18.590, 18.533] # 花叶类 ])  # 计算每类的平均值 grouped\_data = data.groupby('分类名称').mean()  # 从分组数据中提取需要的字段 wholesale\_costs = grouped\_data['平均批发价格(元/千克)'].values loss\_rates = grouped\_data['平均损耗率'].values / 100 markup\_rates = grouped\_data['平均加成比率'].values sales\_prices = grouped\_data['平均销售单价(元/千克)'].values  num\_simulations = 1000 num\_days = 7  # 初始化记录最佳价格和补货量 best\_prices = np.zeros((6, num\_days)) best\_restocks = np.zeros((6, num\_days)) weekly\_best\_profit = 0  # 执行模拟 for product in range(6):  for day in range(num\_days):  day\_best\_profit = -np.inf  for \_ in range(num\_simulations):  price = sales\_prices[product]  restock = predicted\_sales[product, day] / (1 - loss\_rates[product]) # 考虑损耗率  profit = predicted\_sales[product, day] \* (price - wholesale\_costs[product]) - restock \* wholesale\_costs[product] \* loss\_rates[product]   if profit > day\_best\_profit:  day\_best\_profit = profit  best\_prices[product, day] = price  best\_restocks[product, day] = restock   weekly\_best\_profit += day\_best\_profit # 累加最佳利润  # 输出最优结果 print("Best Prices per day per product:") print(best\_prices) print("\nBest Restocks per day per product:") print(best\_restocks) print(f"\nTotal Weekly Best Profit: {weekly\_best\_profit:.2f}") |

|  |
| --- |
| 附录2 |
| 介绍：第三问蒙特卡洛模拟代码 |
| import pandas as pd import numpy as np import random  # 加载数据 data = pd.read\_excel(r"C:\Users\always\Desktop\merged\_output.xlsx") # 请替换为你的数据文件路径  # 确保日期格式正确解析 data['日期'] = pd.to\_datetime(data['日期'])  # 选择特定日期范围内的数据 date\_start = '2023-06-24' date\_end = '2023-06-30' data\_range = data[(data['日期'] >= date\_start) & (data['日期'] <= date\_end)]  # 计算平均值 average\_data = data\_range.groupby('单品名称').mean().reset\_index()  # 模拟参数 num\_simulations = 10000 num\_products = average\_data.shape[0] best\_profit = -np.inf best\_strategy = None  for \_ in range(num\_simulations):  selected\_indices = np.random.choice(average\_data.index, random.randint(27, 33), replace=False)  selected\_data = average\_data.loc[selected\_indices].reset\_index(drop=True)   # 模拟补货量和定价  restocks = np.maximum(2.5, np.random.normal(selected\_data['平均销量(千克)'], selected\_data['平均销量(千克)'] \* 0.1))  price\_fluctuations = np.random.uniform(-0.05, 0.05, size=len(selected\_indices))  prices = selected\_data['平均销售单价(元/千克)'] \* (1 + price\_fluctuations)   # 计算利润  profits = (prices - selected\_data['平均批发价格(元/千克)']) \* restocks - selected\_data['平均批发价格(元/千克)'] \* \  selected\_data['平均损耗率'] \* restocks  total\_profit = profits.sum()   # 更新最佳策略  if total\_profit > best\_profit:  best\_profit = total\_profit  best\_strategy = {  'Selected Product Names': selected\_data['单品名称'].tolist(),  'Unit Costs': selected\_data['平均批发价格(元/千克)'].tolist(),  'Restocks': restocks,  'Prices': prices,  'Profits': profits # 直接存储每个产品的利润  }  # 输出最佳策略的详细信息 print("Best Profit:", best\_profit) print("Best Strategy Details:") for i, product in enumerate(best\_strategy['Selected Product Names']):  print(f"{i + 1}. Product Name: {product}")  print(f" Unit Cost: {best\_strategy['Unit Costs'][i]:.2f} RMB/kg")  print(f" Daily Restock Amount: {best\_strategy['Restocks'][i]:.2f} kg")  print(f" Price: {best\_strategy['Prices'][i]:.2f} RMB/kg")  print(f" Expected Revenue: {best\_strategy['Profits'][i]:.2f} RMB") |

|  |
| --- |
| 附录3 |
| 介绍：用Python语言画相关系数热图 |
| import pandas as pd  import seaborn as sns  import matplotlib.pyplot as plt  # 读取Excel文件中的数据  file\_path = r"D:\WeChat Files\wxid\_pph1gnrllef322\FileStorage\File\2024-08\data\_filled.xlsx"  df = pd.read\_excel(file\_path)  # 删除日期列，假设日期列名为'销售日期'，请根据实际情况调整列名  df = df.drop('销售日期', axis=1)  # 计算相关系数矩阵  correlation\_matrix = df.corr()  # 设置中文显示字体和解决负号无法显示的问题  plt.rcParams['font.sans-serif'] = ['SimHei'] # 中文标签支持  plt.rcParams['axes.unicode\_minus'] = False # 负号显示支持  # 绘制相关系数矩阵的热图  plt.figure(figsize=(10, 8))  heatmap = sns.heatmap(correlation\_matrix, annot=True, cmap='coolwarm', linewidths=0.5)  # 设置中文标签  heatmap.set\_xlabel('特征', fontsize=12) # 设置x轴标签  heatmap.set\_ylabel('特征', fontsize=12) # 设置y轴标签  heatmap.set\_title('相关系数矩阵', fontsize=15) # 设置标题  plt.show() |

|  |
| --- |
| 附录4 |
| 介绍：用Python计算描述性统计量 |
| import pandas as pd  from scipy.stats import skew, kurtosis  import numpy as np  # 读取1.xlsx和2.xlsx文件  file1\_path = 'C:/Users/24361/Desktop/C题/附件1.xlsx'  file2\_path = 'C:/Users/24361/Desktop/C题/附件2.xlsx'  df1 = pd.read\_excel(file1\_path)  df2 = pd.read\_excel(file2\_path)  # 将1.xlsx中的信息合并到2.xlsx中  merged\_df = pd.merge(df2, df1[['单品编码', '分类名称']], on='单品编码', how='left')  # 将合并后的数据保存到新的xlsx文件中  output\_path = 'C:/Users/24361/Desktop/C题/分类后数据1.xlsx' # 替换为你的文件路径  merged\_df.to\_excel(output\_path, index=False)  # 对不同品类的销售量进行描述性统计  grouped = merged\_df.groupby('分类名称')['销量(千克)'].describe().reset\_index()  # 计算每个分类的偏度和峰度  def compute\_skew\_kurt(series):  if series.std() == 0:  return pd.Series([np.nan, np.nan], index=['skew', 'kurtosis'])  return pd.Series([skew(series), kurtosis(series)], index=['skew', 'kurtosis'])  # 计算偏度和峰度  skew\_kurt = merged\_df.groupby('分类名称')['销量(千克)'].apply(compute\_skew\_kurt).reset\_index()  # 将偏度和峰度的结果与描述性统计数据合并  result = pd.merge(grouped, skew\_kurt, on='分类名称')  # 将结果保存到新的xlsx文件中  output\_stats\_path = 'C:/Users/24361/Desktop/C题/分类销售统计1.xlsx' # 替换为你的文件路径  result.to\_excel(output\_stats\_path, index=False)  print("数据合并和统计已完成，并保存到新的xlsx文件中。") |

|  |
| --- |
| 附录4 |
| 介绍：用Python计算描述性统计量 |
|  |

|  |
| --- |
| 附录4 |
| 介绍：用Python计算描述性统计量 |
|  |

|  |
| --- |
| 附录4 |
| 介绍：用Python计算描述性统计量 |
|  |