**基于回归模型和时间序列分析模型的蔬菜类商品的自动定价与补货决策**

**摘要**

本文主要研究基于回归模型和时间序列分析模型的蔬菜类商品的自动定价与补货决策。

针对问题1，我们不难猜到以三年总体的数据量为基础，可以用Spearman、Pearson和Kendall相关系数分析各类之间的相关性和一致性，同时可以分析最相关的两个类各自内部单品之间以及这分属于两个类的单品之间的相关性和一致性。另在探寻产品销量随时间的分布规律时，观察到各品类的销量关于季度的分布波动较大，故考虑在每个季度利用三年同时段的数据进行类似的分析，再提取单季度销售量较大的两个强相关品类之下的单品进行相关性分析，找到了若干组配对。

针对问题2，我们首先要找到成本加成定价模型的核心自变量并定义其为加成参数，再根据成本加成定价公式以及历史数据拟合出某品类在过去三年的同一时期的销售总量与该参数之间的函数关系，这里使用了logistic模型。之后分析各品类在被选购时的损坏及打折情况，并利用时间序列分析模型给出未来7天每天的各品类下单品的成本预测值，并根据历史销售占比加权得到品类的成本预测值，以确定利润与加成参数之间的函数关系。最后我们根据不同天的品类预测成本来确定不同的利润函数极值点，从而确定最佳日补货计划和定价策略。

针对问题3，在划定一个可选范围之后，我们首先要进一步确定每个单品的销售量与定价的关系，可以从模型2的品类售量与加成参数之间的关系来类比引申。再定义桥梁来由单品所在类的销售量估计出单品的日销售量来完善总销售额，最后考虑附件4里给出的单品近日损耗率来确定成本，以得到利润有关单品的函数。在加成参数于可行域变化的过程中，可以找到利润函数为正且较大的单品集合。

针对问题4，我们针对问题1、2和3的模型中的参数选取与估计以及模型本身的拟合情况，考虑哪些可以更精确化的参数以优化模型。除此之外，我们可以考虑其他的市场元素来帮助解决以上的几个问题。

**关键词：相关性分析 logistic回归模型 时间序列分析模型 优化与决策**

**一、问题重述与整体分析**

**1.1问题背景**

蔬菜类商品的保鲜期都比较短，且品相随销售时间的增加而变差，大部分品种如当日未售出，次日就无法再售。大部分商超销售的蔬菜品种众多、产地不尽相同，而蔬菜的进货交易时间通常在凌晨 3:00- 4:00。因此为追求更高的收益，商家须在不确切知道具体单品和进货价格的情况下，做出当日各蔬菜品类的补货决策。在本问题中，蔬菜的定价方式选定为“成本加成定价”方式，同时商超会对运损和品相变差的商品进行打折销售。此时，可靠的市场需求分析，把握未来可能的批发价、销售量等，对补货决策和定价决策尤为重要。

**1.2问题重述**

在该问题中，数据资料包含两百五十余种单品在内的六种蔬菜品类的各种信息。其中有近三年间各商品的销售流水明细及批发价格和各商品近期的损耗率。需要通过这些数据来求解下列问题。

问题1：蔬菜类商品不同品类或不同单品之间存在着一定的关联关系，分析蔬菜各品类及单品销售量的分布规律和相互关系。

问题 2：对于商超补货计划，根据商超近三年的销售流水明细及批发价格数据，分析各蔬菜品类的销售总量与成本加成定价的关系。并辅以近期损耗率及商品分类信息，给出各蔬菜品类 2023 年 7 月 1-7 日的日补货总量和定价策略， 使得商超收益最大。

问题 3：考虑到蔬菜类商品的销售空间有限，需要进一步制定单品的补货计划，使可售单品总数控制在27-33个，且各单品订购量满足最小陈列量 2.5 千克的要求。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，在满足市场对各品类蔬菜商品需求的前提下，使得商超收益最大。

问题 4：为更好地制定蔬菜商品的补货和定价决策，分析商超还需要采集哪些相关数据，并明确这些数据对解决上述问题有何帮助，给出意见和理由。

**1.3问题整体分析**

1.3.1问题一的分析

该问题是一个较经典的统计分析模型组合。首先由于不同的品类及单品存在一定的关联，因此通过数据处理，先找到不同品类及单品的销售规律。再注意到所给数据关于季度具有较大的波动性，故将数据按不同季度进行区分，对六个品类及其中单品的销售量进行分布规律以及相关性分析。

1.3.2问题二的分析

第二问是一个在预测分析模型基础上进行的优化问题。由于商超所卖产品拥有相对完整的历史成本信息，因而采用相对应的成本加成模型，得到定价与加成参数的关系。并采用时间序列分析，预测出不同单品在7月1日至7月7日间每天的批发价，通过计算历史数据得到每个品类的加权平均成本。此后考虑到销售量与定价的关系，提出可能的logit模型进行拟合。将此模型应用进需要优化的商超获利函数，通过对不同品类打折销售的概率进行分析，进而将食用菌与其余五项分开讨论，提出打折的概念并深化出加权平均打折率的概念，于是可按照非打折出售收入加打折出售收入减去总成本的方法，通过模拟退火算法计算得到理想的加成系数使商超获利最大。最终再检验logit模型的准确性，证明此加成系数合理。

1.3.3问题三的分析

此模型是在问题二基础上加上了规划的预测分析及优化模型。由于要求只考虑6月24日至6月30日间出售过的单品，因此首先将它们筛选出来，并按照不同品类进行聚类。此时沿用问题二的模型，并细化不同单品在其所属品类内销售的比例，进而得到在特定加成参数下某单品预测的销售量。同样的，将销售量分为打折销售与非打折销售处理，然后得到该单品的预测利润。由于问题三给了有关陈列量的约束条件，因此通过此条件，得到各个单品可行且利润最高的定价策略以及对应的补货量。此时将预测利润按从高到低排列，最终再根据单品种类数的限制分情况讨论，最终确定选取的陈列单品种类及种类数。

1.3.4问题四的分析

该问题被我们冠以“开放性影响因子挖掘题目”的称号，其一旨在分析除了题目所涉及的商超获利影响因素之外，其它未提及的因素与获利是否相关。其二旨在分析，若相关则具体是怎么样的影响。为此需要结合生活实际，思考各种可能的影响因素，分析其影响的具体表现，并得出结论。

**二、必要的符号说明**

|  |  |
| --- | --- |
|  | 变量X和Y之间的相关系数 |
|  | 变量X和Y之间的协方差 |
|  | 第j类蔬菜产品的预计成本 |
|  | 第j类蔬菜的定价/售价 |
|  | 第j类受损蔬菜的加权打折率 |
|  | 第j类蔬菜的平均损耗率 |
|  | （问题3中）某单品近日损耗率 |
|  | 某天的成本加成定价公式中的加成参数，可以是某品类加权得到的也可以是某单品直接得到的 |
|  | 针对品类的加成参数下某品类蔬菜的预测利润 |
|  | 针对单品的加成参数下某单品蔬菜的预测利润 |
|  | 由部分历史数据所得到的加权折率 |
|  | 为消除量纲影响所引进的加成折率 |
|  | 顾客购买非打折产品的概率 |
|  | 顾客购买打折产品的概率 |
|  | 问题2中的日补货量，问题3中所选7天内某单品打折销售的总销量与7天内该单品所属品类打折销售的销售总量之比 |
|  | 某品类在加成参数下的销售总量 |
|  | 问题2中某单品某天的打折销量 |
|  | 问题2中某单品某天的非打折销量 |
|  | 某单品7天的未打折销售总量 |
|  | 某单品7天的总销售量 |
|  | 某单品7天的未打折销售总量与总销售量即q与l之比 |
|  | 某单品7天的打折销售总量与总销售量，数值上等于 |
|  | 问题3中7天内某单品打折出售的总销量 |

注：在具体讨论某天某品类某单品是，i、j和k可能会被省略，但三者的前后顺序保持不变，具体情况文中会说明清楚。

**三、模型分析与建立**

**3.1问题1的模型**

问题1指出蔬菜类商品的不同品类和单品之间可能存在一定的关联关系，我们很容易想到直接以三年全部的数据为基础来计算六类蔬菜两两之间有关销售量的相关性，我们主要考虑Pearson、Kendall和Spearman相关系数，这三类相关系数是统计学上的三大重要相关系数，表示两个变量之间变化的趋势方向和趋势程度。Pearson相关系数的公式为 ，一般要求两个变量分别服从正态分布且标准差不为0，描述的是线性相关关系，取值为[-1, 1]。负数表示负相关，正数表示正相关，在显著性的前提下，绝对值越大，相关性越强。附件中数据的正态性较差，故我们主要考虑另外两个对正态性要求很低的相关系数。Spearman 秩相关系数的公式为，其值与两个相关变量的具体值无关，而仅仅与其值之间的大小关系有关，故它作为一种非参数的统计方法是基于秩次的，可以用于衡量两个变量之间的单调关系，即当一个变量增加时，另一个变量是否也会增加或减少。Kendall相关系数是一个用来测量两个随机变量相关性的统计值无参数假设检验，它使用计算而得的相关系数去检验两个随机变量的统计依赖性，常用于属性一致性分析。[1]

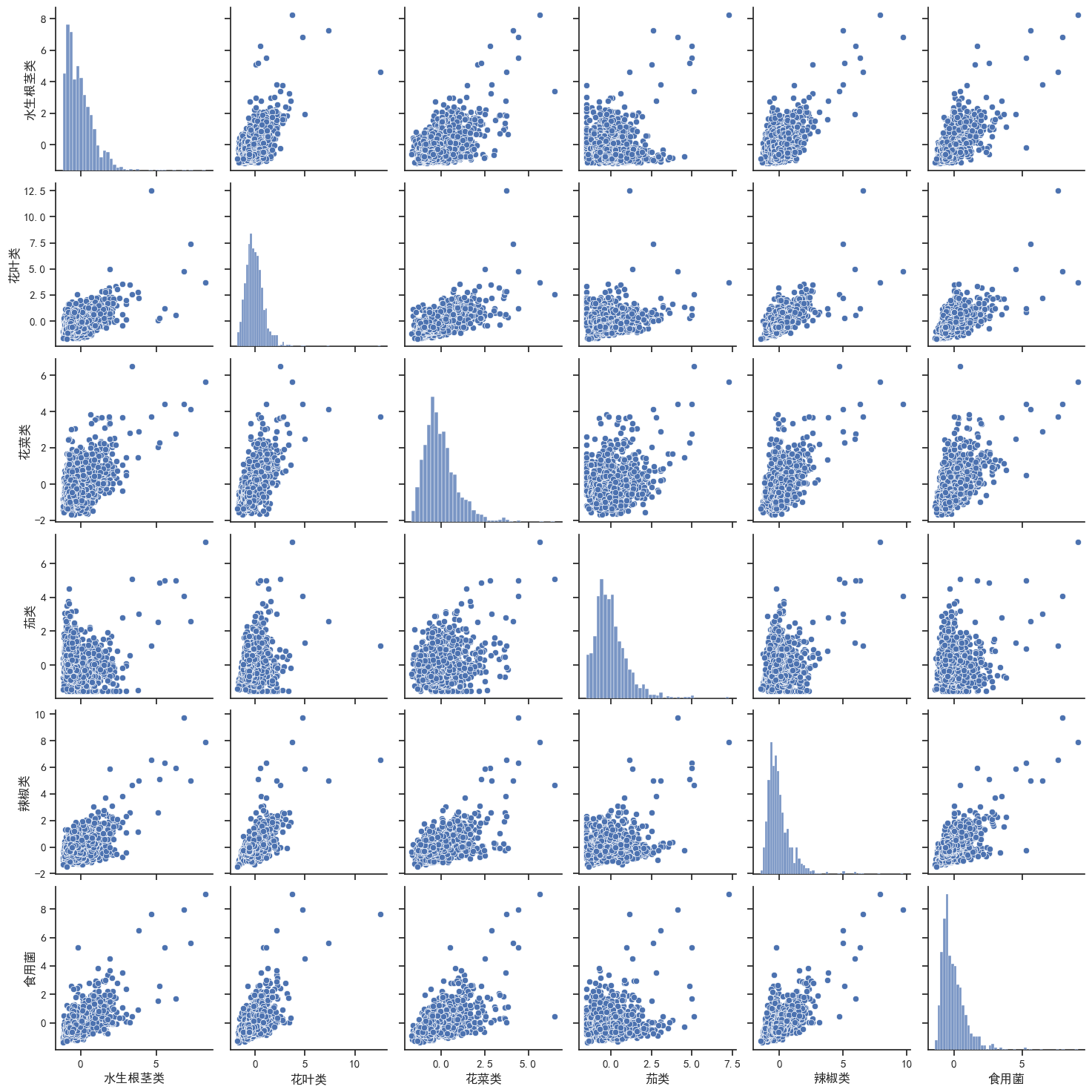
注意到附件2中也包含了退货记录，需要在处理数据之前清除。我们用python处理附件中数据并分别计算Pearson、Kendall和Spearman系数，先以天为单位计算不同类蔬菜的销售量，再将某两类蔬菜的一系列数据作相关性分析并得到三个系数。我们将结果生成的表格分别展示如下，顺序依次为Pearson、Kendall、Spearman相关系数。

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **分类名称** | **水生根茎类** | | **花叶类** | **花菜类** | **茄类** | **辣椒类** | **食用菌** |
| **水生根茎类** | 1 | 0.561287 | | 0.542507 | 0.074423 | 0.614364 | 0.670189 |
| **花叶类** | 0.561287 | 1 | | 0.62767 | 0.257879 | 0.65979 | 0.631075 |
| **花菜类** | 0.542507 | **0.62767** | | 1 | 0.311873 | 0.551288 | 0.523076 |
| **茄类** | 0.074423 | 0.257879 | | 0.311873 | 1 | 0.273673 | 0.119808 |
| **辣椒类** | 0.614364 | 0.65979 | | 0.551288 | 0.273673 | 1 | 0.687485 |
| **食用菌** | 0.670189 | 0.631075 | | 0.523076 | 0.119808 | 0.687485 | 1 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **分类名称** | **水生根茎类** | **花叶类** | **花菜类** | **茄类** | **辣椒类** | **食用菌** |
| **水生根茎类** | 1 | 0.306558 | 0.272638 | -0.13897 | 0.233193 | 0.431768 |
| **花叶类** | 0.3065585 | 1 | 0.456042 | 0.183014 | 0.425687 | 0.423923 |
| **花菜类** | 0.2726384 | **0.456042** | 1 | 0.129604 | 0.297523 | 0.318563 |
| **茄类** | -0.13897 | 0.183014 | 0.129604 | 1 | 0.076948 | -0.0747 |
| **辣椒类** | 0.2331927 | 0.425687 | 0.297523 | 0.076948 | 1 | 0.384323 |
| **食用菌** | 0.431768 | 0.423923 | 0.318563 | -0.0747 | 0.384323 | 1 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **分类名称** | **水生根茎类** | **花叶类** | **花菜类** | **茄类** | **辣椒类** | **食用菌** |
| **水生根茎类** | 1 | 0.438869 | 0.396029 | -0.20997 | 0.333866 | 0.605351 |
| **花叶类** | 0.438869 | 1 | 0.633454 | 0.252732 | 0.595045 | 0.596225 |
| **花菜类** | 0.396029 | **0.633454** | 1 | 0.193014 | 0.429713 | 0.46242 |
| **茄类** | -0.20997 | 0.252732 | 0.193014 | 1 | 0.10424 | -0.11452 |
| **辣椒类** | 0.333866 | 0.595045 | 0.429713 | 0.10424 | 1 | 0.534785 |
| **食用菌** | 0.605351 | 0.596225 | 0.46242 | -0.11452 | 0.534785 | 1 |

调用seaborn库里的pairplot方法生成的散点图如图3.1.1所示。

图3.1.1

观察到Spearman系数和Kendall系数最高的为花叶类和花菜类这一组，且Pearson系数也较高。考虑到原始数据依据一天为时间段排序后并不具有显著的正态性，亦不具有连续性，故Pearson系数的参考价值较低。综上所述，我们认为以全年为时间尺度上来讲，花叶类和花菜类蔬菜的销售量具有最显著的正相关性与一致性，食用菌类和水声菌类次之。

基于此，我们接着分析花菜类内的单品之间、花叶类内的单品之间和两类各自的单品之间的关系。首先分析花菜类单品与单品之间的三个系数，表格按照Pearson、Kendall和Spearman的顺序排列，由于数据量过大便不再展示且放置于附件表格之中，两类单品之间的相关性系数表格亦然做此处理。再分析花叶类的单品与单品之间的相关性，下面以同样的顺序展示相关系数。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **单品名称** | **枝江青梗散花** | **紫白菜(1)** | **紫白菜(2)** | **西兰花** | **青梗散花** |
| **枝江青梗散花** | 1 | -0.05655 | -0.02236 | 0.146813 | -0.34353 |
| **紫白菜(1)** | -0.05655 | 1 | -0.00422 | 0.00539 | 0.142286 |
| **紫白菜(2)** | -0.02236 | -0.00422 | 1 | -0.03563 | -0.01276 |
| **西兰花** | 0.146813 | 0.00539 | -0.03563 | 1 | 0.147675 |
| **青梗散花** | -0.34353 | 0.142286 | -0.01276 | 0.147675 | 1 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **单品名称** | **枝江青梗散花** | **紫白菜(1)** | **紫白菜(2)** | **西兰花** | **青梗散花** |
| **枝江青梗散花** | 1 | -0.07166 | -0.02609 | 0.008991 | -0.40661 |
| **紫白菜(1)** | -0.07166 | 1 | -0.00508 | -0.00694 | 0.068434 |
| **紫白菜(2)** | -0.02609 | -0.00508 | 1 | -0.04682 | -0.00612 |
| **西兰花** | 0.008991 | -0.00694 | -0.04682 | 1 | 0.097373 |
| **青梗散花** | -0.40661 | 0.068434 | -0.00612 | 0.097373 | 1 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **单品名称** | **枝江青梗散花** | **紫白菜(1)** | **紫白菜(2)** | **西兰花** | **青梗散花** |
| **枝江青梗散花** | 1 | -0.07776 | -0.02823 | 0.008447 | -0.48751 |
| **紫白菜(1)** | -0.07776 | 1 | -0.00509 | -0.00863 | 0.076084 |
| **紫白菜(2)** | -0.02823 | -0.00509 | 1 | -0.05735 | -0.0068 |
| **西兰花** | 0.008447 | -0.00863 | -0.05735 | 1 | 0.129205 |
| **青梗散花** | -0.48751 | 0.076084 | -0.0068 | 0.129205 | 1 |

很遗憾的是，这五个单品之间并没有较强的相关性。但是在观察未展示在论文之中的花菜类单品间的相关系数表的过程中，我们发现了一些具有较高的相关系数且销售总量较多配对组合。例如云南生菜(份)和云南油麦菜(份)有约0.84的相关系数，前者在三年的销售总量约为2700千克，后者在三年的销售总量约为1600千克，二者在销售总量方面均能排在所有单品的前列。这样的数据表明，云南生菜(份)和云南油麦菜(份)有非常好的正相关性以及一致性，我们猜测二者的产品名称相近起到了非常重要的作用，同时在所选取的时间段内，云南这一地方也曾被网络媒体带货，这亦起到了重要作用。

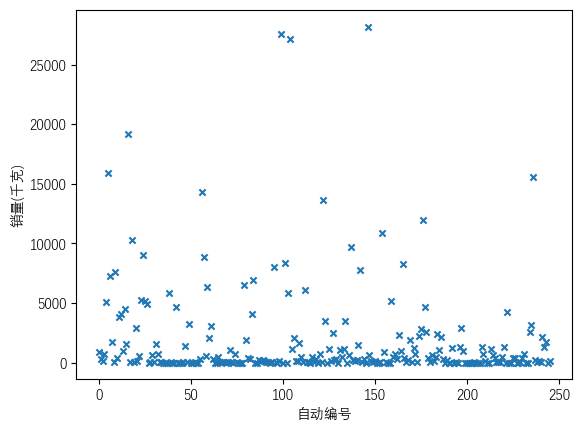
以下展示所有单品的销售总量与其在附件1中相对位置所产生的自动编号之间的散点图，如图3.1.2所示。

图3.1.2

进一步地，我们观察附件2可知某些单品的蔬菜在一年的某些季度销售较多而在其他季度销售很少甚至几乎没有售卖，所以我们不妨将2020年7月1日至2023年6月30日这12个季度的数据集分为4组，1组包括了一个季度在三年的数据集，再逐组分析各品类的分布规律以及不同品类之间的相关关系。

接下来分析分布规律，三年的时间包括了12个季度，每个季度都有3个不同年份的数据集，那么我们不妨以季度为标尺来分析。我们先计算了各个季度下各类蔬菜的销售总量，将六类蔬菜的分布规律展示于下表。

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **水生根茎类** | **花叶类** | **花菜类** | **茄类** | **辣椒类** | **食用菌** |
| **第一季度** | 12673.43 | 46345.33 | 10121.07 | 5211.662 | 28806.22 | 23153.98 |
| **第二季度** | **3481.202** | 40955.72 | **7733.269** | 6875.094 | 19955.38 | 12652.69 |
| **第三季度** | 11021.6 | 60653.48 | 12992.3 | 7155.558 | 22787.36 | 14876.48 |
| **第四季度** | 13431.32 | 50705.03 | 10943.14 | 3199.805 | 20096.15 | 25448.57 |

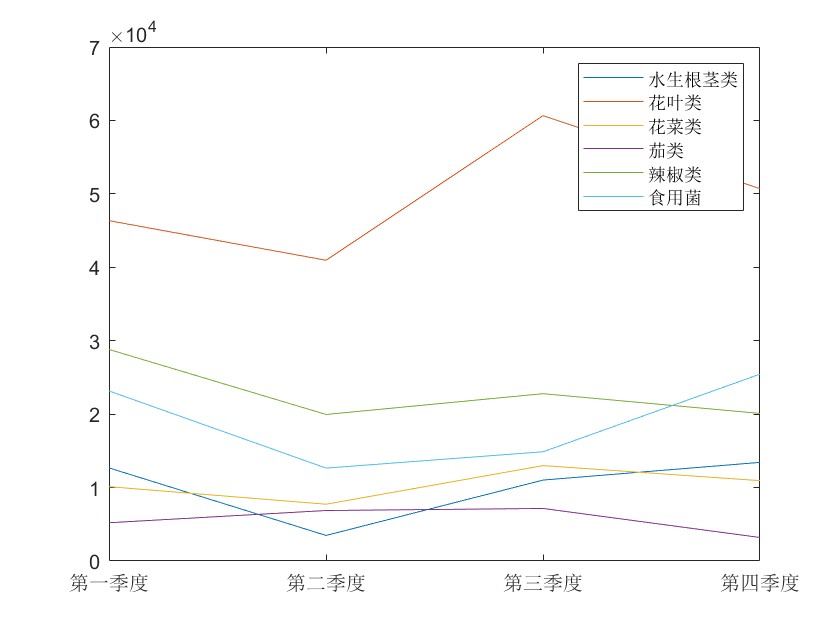
此外，我们还可以绘制折线图来反映各类蔬菜的销量随季度变化的折线图，如图3.1.3所示。

图3.1.3

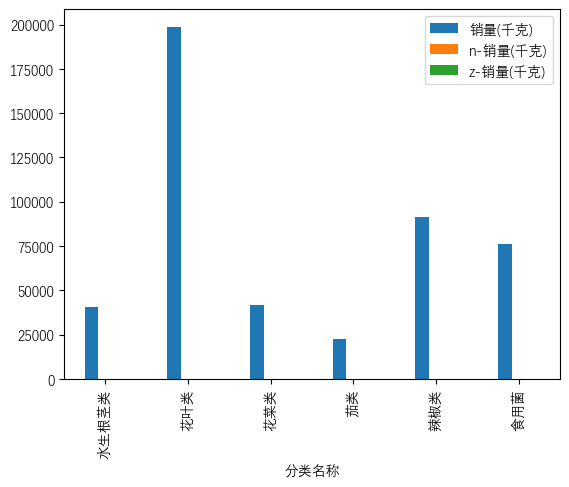
不难发现，波动性是较为显著的，比如第二季度水生根茎类的销量明显低于茄类的，但是总量上前者高于后者，如图3.1.4所示。

图3.1.4

按照季度进行类与类之间相关性分析的时候，Pearson系数对于原始数据正态性的高要求使得其结果误差极大，下面展示各类在考虑第一季度三年总计的销售量之时的Pearson系数表格。

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **分类名称** | **水生根茎类** | **花叶类** | **花菜类** | **茄类** | **辣椒类** | **食用菌** |
| **水生根茎类** | 1 | 0.999261 | 0.99953 | 0.999223 | 0.999348 | 0.999694 |
| **花叶类** | 0.999261 | 1 | 0.999677 | 0.999484 | 0.999649 | 0.999676 |
| **花菜类** | 0.99953 | 0.999677 | 1 | 0.999592 | 0.99965 | 0.99966 |
| **茄类** | 0.999223 | 0.999484 | 0.999592 | 1 | 0.999808 | 0.999476 |
| **辣椒类** | 0.999348 | 0.999649 | 0.99965 | 0.999808 | 1 | 0.999629 |
| **食用菌** | 0.999694 | 0.999676 | 0.99966 | 0.999476 | 0.999629 | 1 |

于是，下文的分析仅考虑Kendall和Spearman系数，以下分别是第一至第四各个季度的Kendall和Spearman系数表，以便在不同季度分析各个类之间的相关关系。

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **分类名称** | **水生根茎类** | **花叶类** | **花菜类** | **茄类** | **辣椒类** | **食用菌** |
| **水生根茎类** | 1 | 0.577118 | 0.622561 | 0.347233 | 0.376103 | 0.730019 |
| **花叶类** | 0.577118 | 1 | 0.566426 | 0.478214 | 0.514568 | 0.603315 |
| **花菜类** | **0.622561** | 0.566426 | 1 | 0.442395 | 0.454157 | 0.589949 |
| **茄类** | 0.347233 | 0.478214 | 0.442395 | 1 | 0.602245 | 0.392676 |
| **辣椒类** | 0.376103 | 0.514568 | 0.454157 | 0.602245 | 1 | 0.416199 |
| **食用菌** | **0.730019** | 0.603315 | 0.589949 | 0.392676 | 0.416199 | 1 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **分类名称** | **水生根茎类** | **花叶类** | **花菜类** | **茄类** | **辣椒类** | **食用菌** |
| **水生根茎类** | 1 | 0.77907 | 0.813881 | 0.479095 | 0.505832 | 0.903587 |
| **花叶类** | 0.77907 | 1 | 0.773657 | 0.67101 | 0.70245 | 0.804841 |
| **花菜类** | **0.813881** | 0.773657 | 1 | 0.600714 | 0.619323 | 0.795928 |
| **茄类** | 0.479095 | 0.67101 | 0.600714 | 1 | 0.786415 | 0.554422 |
| **辣椒类** | 0.505832 | 0.70245 | 0.619323 | 0.786415 | 1 | 0.581486 |
| **食用菌** | **0.903587** | 0.804841 | 0.795928 | 0.554422 | 0.581486 | 1 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **分类名称** | **水生根茎类** | **花叶类** | **花菜类** | **茄类** | **辣椒类** | **食用菌** |
| **水生根茎类** | 1 | 0.268514 | 0.275203 | 0.103201 | 0.341615 | 0.438127 |
| **花叶类** | 0.268514 | 1 | 0.406593 | 0.343526 | 0.511228 | 0.450072 |
| **花菜类** | 0.275203 | 0.406593 | 1 | 0.260392 | 0.30387 | 0.300048 |
| **茄类** | 0.103201 | 0.343526 | 0.260392 | 1 | 0.233158 | 0.224558 |
| **辣椒类** | 0.341615 | **0.511228** | 0.30387 | 0.233158 | 1 | 0.582418 |
| **食用菌** | 0.438127 | 0.450072 | 0.300048 | 0.224558 | **0.582418** | 1 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **分类名称** | **水生根茎类** | **花叶类** | **花菜类** | **茄类** | **辣椒类** | **食用菌** |
| **水生根茎类** | 1 | 0.40197 | 0.398178 | 0.13569 | 0.465561 | 0.608603 |
| **花叶类** | 0.40197 | 1 | 0.566127 | 0.485982 | 0.717183 | 0.64624 |
| **花菜类** | 0.398178 | 0.566127 | 1 | 0.36965 | 0.458703 | 0.43412 |
| **茄类** | 0.13569 | 0.485982 | 0.36965 | 1 | 0.328884 | 0.323937 |
| **辣椒类** | 0.465561 | **0.717183** | 0.458703 | 0.328884 | 1 | 0.786446 |
| **食用菌** | 0.608603 | 0.64624 | 0.43412 | 0.323937 | **0.786446** | 1 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **分类名称** | **水生根茎类** | **花叶类** | **花菜类** | **茄类** | **辣椒类** | **食用菌** |
| **水生根茎类** | 1 | 0.45582 | 0.223469 | 0.007948 | 0.470313 | 0.386629 |
| **花叶类** | **0.45582** | 1 | 0.318841 | 0.228612 | 0.661057 | 0.43899 |
| **花菜类** | 0.223469 | 0.318841 | 1 | 0.447873 | 0.338943 | 0.301075 |
| **茄类** | 0.007948 | 0.228612 | 0.447873 | 1 | 0.165498 | 0.066854 |
| **辣椒类** | **0.470313** | **0.661057** | 0.338943 | 0.165498 | 1 | 0.423562 |
| **食用菌** | 0.386629 | 0.43899 | 0.301075 | 0.066854 | 0.423562 | 1 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **分类名称** | **水生根茎类** | **花叶类** | **花菜类** | **茄类** | **辣椒类** | **食用菌** |
| **水生根茎类** | 1 | 0.643341 | 0.331011 | -0.00112 | 0.637239 | 0.563069 |
| **花叶类** | **0.643341** | 1 | 0.457342 | 0.31803 | 0.848706 | 0.616812 |
| **花菜类** | 0.331011 | 0.457342 | 1 | 0.633285 | 0.478589 | 0.442228 |
| **茄类** | -0.00112 | 0.31803 | 0.633285 | 1 | 0.240578 | 0.103682 |
| **辣椒类** | **0.637239** | **0.848706** | 0.478589 | 0.240578 | 1 | 0.586404 |
| **食用菌** | 0.563069 | 0.616812 | 0.442228 | 0.103682 | 0.586404 | 1 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **分类名称** | **水生根茎类** | **花叶类** | **花菜类** | **茄类** | **辣椒类** | **食用菌** |
| **水生根茎类** | 1 | 0.503341 | 0.348303 | 0.395349 | 0.432237 | 0.404972 |
| **花叶类** | **0.503341** | 1 | 0.328522 | 0.497461 | 0.465918 | 0.341353 |
| **花菜类** | 0.348303 | 0.328522 | 1 | 0.367014 | 0.457364 | 0.352045 |
| **茄类** | 0.395349 | **0.497461** | 0.367014 | 1 | 0.430634 | 0.20556 |
| **辣椒类** | 0.432237 | 0.465918 | 0.457364 | 0.430634 | 1 | 0.484095 |
| **食用菌** | 0.404972 | 0.341353 | 0.352045 | 0.20556 | 0.484095 | 1 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **分类名称** | **水生根茎类** | **花叶类** | **花菜类** | **茄类** | **辣椒类** | **食用菌** |
| **水生根茎类** | 1 | 0.67502 | 0.491343 | 0.55752 | 0.595848 | 0.568637 |
| **花叶类** | **0.67502** | 1 | 0.444066 | 0.678501 | 0.633703 | 0.475104 |
| **花菜类** | 0.491343 | 0.444066 | 1 | 0.503445 | 0.630531 | 0.514817 |
| **茄类** | 0.55752 | **0.678501** | 0.503445 | 1 | 0.603813 | 0.309506 |
| **辣椒类** | 0.595848 | 0.633703 | 0.630531 | 0.603813 | 1 | 0.668714 |
| **食用菌** | 0.568637 | 0.475104 | 0.514817 | 0.309506 | 0.668714 | 1 |

可以发现，第一季度水生根系类和食用菌类蔬菜的正相关性和一致性最强，和花菜类的次之；第二季度辣椒类与食用菌类的正相关性和一致性最强，与花叶类的次之；第三季度辣椒类与花叶类的正相关性和一致性最强，水生根茎与花叶类的、水生根茎与辣椒类的均的次之，茄类与水生根茎类的相关性极其差；第四季度花叶类与水生根茎类和茄类的正相关性和一致性难分高下，均是最强。

最后做一些必要的说明，按季度分开数据之后，我们仍然试图找寻可能的单品配对使得二者的某季度内销量较高且相关性、一致性较好，但是我们在海量数据中几乎无法找到这样的组合，原因之一是无论同类还是不同类之间的单品组合很难在某季度同时有较高的售卖量，更不要提相关性、一致性同时均较高的组合。我们只寻找了三年总销售量下的一些单品组合，目前我们认为几乎最好的例子是云南生菜(份)和云南油麦菜(份)。

**3.2问题2的模型假设——logistic模型、时间序列分析模型**

明确成本定价模型

建立销售量与加成参数的逻辑模型

根据题意提出加权平均成本及加权平均折率概念

通过时间序列分析预测7天中品类加权平均成本

建立预测利润函数

通过模拟退火算法找到w合理的最优解

确定日补货量及定价策略

问题2旨在让我们根据历史数据分析各蔬菜品类的销售总量与成本加成定价的关系，这里用到了成本加成定价模型[2]，即：

我们在以下的讨论中都只针对某一类（不妨是第j类，j∈{1，2，3，4，5，6}）蔬菜商品而言，故略去j下标，并且设该类商品共有n个单品，遍历元为k即k∈[1,n]且k为正整数。所以成本加成定价模型写为：

其中为我们规定的定价参数，并没有量纲；为由时间序列分析方法所预测出来的7月1日至7月7日的预计成本。我们的目的是分析未来一周的最佳策略，所使用的数据集有两种选择，要么使用过去三年的数据，要么使用过去三年里这七天即共计21天的数据。由于周期性，我们在估算利润时要使用的根据单品销量加权得到的折率是用过去三年里这7天的值，这是因为，我们观察到实际上某类商品三年共计的打折销售量只占总销售量的很少一部分，一些单品的折率更是如此，那么我们在只知道单品在某天的销售情况下，①如果全部都是未打折即正常出售，则折率为1，②存在打折出售，折率为打折价/标价，③未出售，折率与售量的乘积为0故其值并无影响，于是我们根据单品销量加权得到的的折率为：

这实际上是不合理的，因为对于某些单品非常多的类而言，需要处理的数据非常多，计算量过大。最重要的是，商品的销售是有周期性的，所以我们只需要考虑每年7月1日到7月7日所销售的这些单品，而对于这些单品中绝大多数单品的任何一个，它在一年内的大多数时间是并不销售的，它自身的折率其实并没有对等式做出贡献，故这样算出来的折率其实并不是严谨的。因此，我们使用过去三年里每年这七天内的历史数据，即我们认为加权折率满足：

我们使用logistic模型来解决问题。[3]同样的道理，我们在用logistic拟合销售量即的时候也应当用这21天的数据。所谓logistic模型，是一种广泛应用于社会学、经济学等领域的常见回归模型，选用此模型的原因是题中场景由于空间和时间限制而存在阻滞性，即即使商品定价足够低，由于当地人们的购买力和时间成本有限，会存在一个最大销售量和最小的销售量，且定价越低，显然销量的变化率的绝对值减小。所以设我们认为

其中，为预测的未来七天中某天的销售总额，为该类中第k个单品的当日销量，即为前文提到的定价参数，可认为是21天历史数据里单日该类商品销量的最大值,是21天历史数据里单日该类商品销量最小值。

现在我们将所预测的分类，分为非打折销售的量和打折销售的量，除此之外，我们还要考虑顾客购买非打折产品和打折产品的概率，分别记作和，之和为1，前者即21天内该类非打折的总销量与21天内该类总销量的比值大小。那么我们有：

这样进行简单处理往往是没问题的，因为顾客的消费习惯可以认为是不太可能随时间变化而变化的。这两个概率已知，分析数据可得：

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 概率 | 水生根茎类 | 花叶类 | 花菜类 | 茄类 | 辣椒类 | 食用菌 |
|  | 0.985547 | 0.978968 | 0.991700 | 0.997966 | 0.998997 | 0.885183 |
|  | 0.014453 | 0.021032 | 0.008300 | 0.002034 | 0.001003 | 0.114817 |

分析如上数据可知很少有打折的单品可以被售出，经检验，对除了食用菌类之外的其他品类，补货量中损耗的部分一定会剩余许多且进货量可以直接认为存在等式：

其中，为各个类别的平均损耗率，在这里可以用vscode软件打开附件4可以得到相关数据，如图3.2.1所示。

图3.2.1

考虑到除了食用菌以外，其他品类的均在0.975以上，已经极其逼近1，故打折销售的产品可以被忽略。

对于食用菌这一类商品来说，需要在计算利润时进行更细致的优化，在此不再展开。这样一来，如果给出较好的拟合，各个品类的就可以几乎完全由确定。

再考虑每日凌晨买入一批货，在运输过程中的损失等问题会导致一部分单品需要打折销售而另一部分正常价格销售。我们在讨论除了食用菌之外的五个品类如何定价之时只需要保证不超过后者且不超过前者即可。

里的和是可以由21天的历史数据拟合得出，进而我们可以分别拟合出进行利润的讨论。

在根据21天历史数据具体拟合求解里的和的过程中，有：

我们只需要确定当天该类产品的值。针对于该类产品的所有单品而言，若在某天不存在打折销售，根据可知，，为当天该类产品中某单品的非打折售价，则该类产品当日的=，为该单品的当日非打折销量即日总销量，分母上的为该单品所在类的当日总销量；若该类产品的若干单品当天存在打折销售，则可以做类似处理，即得该类产品当日的，其中为某单品的打折销量，这是因为当天的在上文中已经被完备定义（情形②的定义）且已知。如此一来，确定当天该类产品的值的相关问题便迎刃而解。

之后，我们需要找到历史数据中的每一天的加权折率，以便于考虑预测的打折产品对应的销售额，实际上可以完全被确定。

最后的准备工作落脚点为未来7天某品类的加权平均成本上，首先，对三年内每个单品的批发价进行时间序列分析，预测出该单品在7月1日至7日一周内每日可能的批发价，记为, 即第i类第j天第k个单品。

这里对时间序列分析方法做说明，本题应当采用时间序列预测方法中的Prophet模型。Prophet模型，是Facebook公司开源的一个专门用于大规模时间序列分析的模型，基于加性模型（Additive Model），利用年月日等的周期性再加上假期影响去拟合非线性的趋势。

题目中的数据是按天为时间单位分成了若干间隔，因此具有较强周期性并且拥有几个周期的数据，适合使用该算法。并且Prophet模型对缺失值，趋势偏移和异常值都有着较好的支持，题目中这样的数据点不在少数。Prophet的大致原理如下，它将一个时间序列看成是三部分的组合：趋势，季节和假日，即——代表趋势项，用于拟合时间序列里面的非周期项；代表周期性变化（如季节性变化）；代表了假期的影响（通常表现为对某些时间点的特殊影响）；误差项代表了所有未被模型考虑到的误差因素。同时该模型将一个预测类问题，映射为了曲线拟合类问题，这天然的就和那些用于分析数据间有依赖性的模型有所不同。[4]

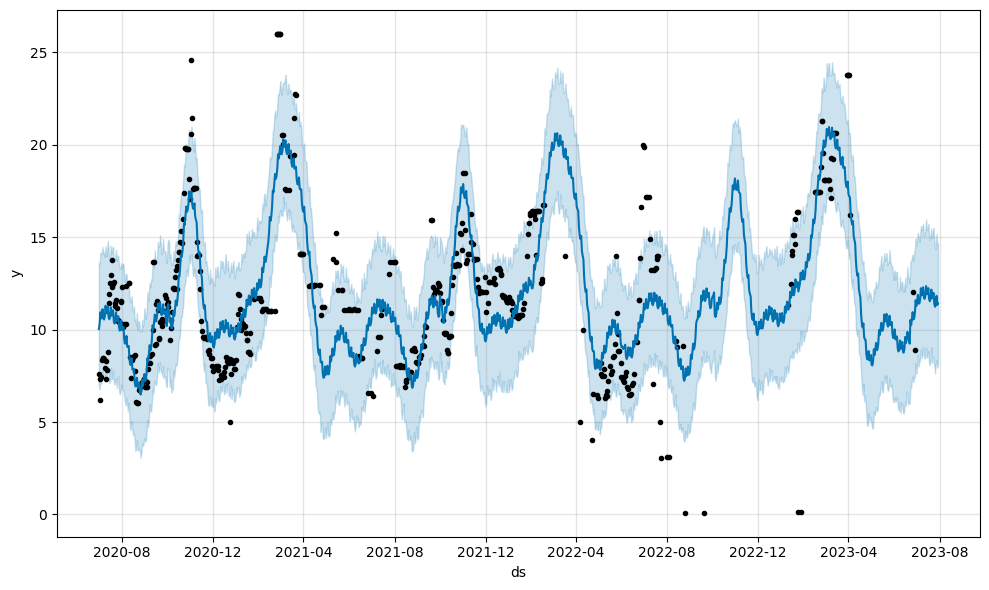
这种时间序列分析的手段可以直观的用图3.2.2来表示。

图3.2.2 七彩椒(1)进价预测模型

但在预测过程中，我们发现有各别单品在此一周内可能的批发价为负值或者预测的最小值大于前三年中得到的观测数据中最大值的三倍又或者预测的最大值小于前三年中的得到的观测数据中最小值的一半。因此，将这些单品筛选出来，观察它们在不同月份的销售情况，可以很明显地发现该些异常单品在七月份几乎没有进货，而只在其他月份有进货记录。因此我们可以认为这些单品属于季节性蔬菜，只在特定月份进货并销售。这也就说明这些异常单品在七月并不是应季的蔬菜，可以认为商超并不会在七月购进这些蔬菜。还有另外一种现象是某些单品在前三年内在七月份总共只有极少数天进货。经分析，可能是该些天内因某平台带货突然爆火导致商超只是突然进了几天货，但热度很快下去之后，商超便又不再进货等极特殊的原因，在此后七天内可认为基本不会发生。因此我们可以将这些异常单品从计算加权平均成本的计算中剔除。虽然筛去了许多数据，但经分析，由于题目所限定的时间范围只是7月1日至7月7日这较为局限的7天，因此在这段时间内进货的商品比较局限的现象也是合理的。

于是此时记该类中剩余单品总数为，＜。对于给定的某，为21天，例如7月1日分别为2020、2021、2022年历史记录中7月1日某单品的销量，分别为2020、2021、2022年7月1日该品类中所有单品的累计销量，所以2023年7月1日至7月7日的每一天的该品类的加权平均成本都可以由以下式子确定：

所有准备工作做好之后，我们可以引入利润公式。对于除了食用菌以外的所有品类，我们有：

再针对食用菌这一类蔬菜而言，其日损耗量无法覆盖掉预计的日打折销售量，但是日利润不一定因为剩余的损耗产品没有被打折售出而减小，这里我们有：

,

= ,

以上讨论的是只关于加成参数的一元函数，我们利用图像分析其极值情况以求出对于给定的种类来说每天不同下的其极值点，即解出题目中所求的商超利润最大的定价策略，同时给出未来7天每天的最佳日补货量，即之下的销售量***B***，即为所求。

对于除了食用菌以外的所有品类，考虑到我们并不知道w的具体范围，因此采用了可以随机选取初始解与最终解的模拟退火算法。我们代入所有可以计算或拟合的参数，简单划定加成参数的范围是，得到了五类蔬菜7月1日（标为0）至7月7日（标为6）这七天内的最大及对应的极值点和日总售价的预测值，表格依次展示如下：

|  |  |
| --- | --- |
|  | **水生根茎类** |
| **0** | (85.03544241841007, 0.6495554036886869, 20.492876938851026) |
| **1** | (86.62780560212954, 0.721035903757209, 18.39148434850515) |
| **2** | (59.5325655104034, 0.7094908736837592, 19.464852114514727) |
| **3** | (70.77261058749534, 0.6386230605760601, 23.096292489888263) |
| **4** | (69.94187899101692, 0.6800155831434472, 22.036839264978443) |
| **5** | (65.9069462584689, 0.684925480839549, 22.55312007153998) |
| **6** | (62.002814316830154, 0.6753089016787844, 23.723742577189416) |

|  |  |
| --- | --- |
|  | **花叶类** |
| **0** | (569.8902493636199, 1.3720413759252765, 130.15596305466877) |
| **1** | (581.8132493708387, 1.3919172333727658, 129.91879001334067) |
| **2** | (621.768795844352, 1.482026001060882, 129.39178353144487) |
| **3** | (614.5614089842734, 1.4569328786991176, 129.45578816896926) |
| **4** | (593.3370706927568, 1.4580018983085565, 129.41769741377541) |
| **5** | (608.2124057346628, 1.4564880094668884, 129.39150722357502) |
| **6** | (489.32249777171137, 1.1702143251159687, 132.67484515919742) |

|  |  |
| --- | --- |
|  | **花菜类** |
| **0** | (377.307331361222, 1.3784828694061495, 44.10845831935055) |
| **1** | (431.710799472452, 1.4998288660008396, 44.91908270604457) |
| **2** | (423.6422695612223, 1.4510251864614425, 46.44765844144574) |
| **3** | (420.91923246651936, 1.3947002593815478, 48.076921793702745) |
| **4** | (404.03833581691276, 1.3442649720360167, 49.738339695586504) |
| **5** | (393.5575727832876, 1.271288687442056, 51.54336340830964) |
| **6** | (465.1984104323397, 1.4476695502562587, 52.24248425510075) |

|  |  |
| --- | --- |
|  | **茄类** |
| **0** | (138.69933224844655, 1.4094084975849759, 29.970815510573722) |
| **1** | (129.36163311479845, 1.3151149771809847, 30.660168663344212) |
| **2** | (126.24028833742658, 1.25111954259172, 31.073568131558822) |
| **3** | (131.26385740712414, 1.4933036040016792, 28.679700738404218) |
| **4** | (136.8906402685074, 1.4762206985390496, 28.657725183947253) |
| **5** | (137.15305414975074, 1.4578302081055963, 28.64744246049227) |
| **6** | (131.85069749524988, 1.4824784994261102, 28.25433455080689) |

|  |  |
| --- | --- |
|  | **辣椒类** |
| **0** | (705.4433421279177, 1.4745400340400299, 79.90431384554626) |
| **1** | (632.0395651220194, 1.4779221264836864, 78.20915559572029) |
| **2** | (611.9303219970305, 1.4958100157793779, 76.6933998309766) |
| **3** | (652.0028959483657, 1.4783750014744999, 74.60682572214715) |
| **4** | (620.5652470715314, 1.429018914745553, 71.98141865805978) |
| **5** | (547.0679853555564, 1.4024175994698433, 69.68136388221367) |
| **6** | (556.6956875180119, 1.4433208930238952, 68.44160404580064) |

相应的，各品类每日的最佳日补货量也是可以根据当日的极值点对应的预计销售量计算得到的，即使用以下等式计算可得到：

这里不再赘述。至于食用菌品类，我们也可以近似地用这个利润函数处理，得到的表格如下：

|  |  |
| --- | --- |
|  | **食用菌** |
| **0** | (460.15336872341896, 1.3777687040993318, 61.713743579188254) |
| **1** | (509.5507487540344, 1.4090432601480207, 61.723143457239885) |
| **2** | (523.4996562432052, 1.360259970917236, 61.703645766599685) |
| **3** | (528.4918865248176, 1.4454684062191425, 61.730842603078194) |
| **4** | (520.4334248440664, 1.4978698927311491, 61.739612944951276) |
| **5** | (514.378252372089, 1.3396289002430704, 61.68727316091006) |
| **6** | (555.8079126014651, 1.4951117608992037, 61.73828570038329) |

经过检验，起始使用的针对食用菌的较为精确的利润函数,= ,计算出来的值与使用另外五个品类对应的利润函数计算出来的值相差微乎其微，故亦不再赘述。

**3.3问题3——基于问题2模型的进一步优化**

在第二问模型的基础上，根据题意从6月24日至6月30日中找出可售单品共计***G***个，记***i***为其中第***i***种。此时，单品成本及对应定价的关系为：

再沿用问题2中与之间的关系，我们有：

这里为单品的预测销量。我们也需要一些准备工作，记：

其中为6月24-6月30日这七天内某单品的打折出售的销量，而是其中某天的销量。类似的，我们记该单品7天的总销售量为，未打折销售总量为，之后我们定义：

故我们有某单品的利润期望公式：

c

其中，为天内该单品未打折销售的总销量与7天内该单品所属品类未打折销售的销售总量之比，为7天内该单品打折销售的总销量与7天内该单品所属品类打折销售的销售总量之比。根据题意，需要满足最小陈列量的要求，也就是说：

即：

然后按照每个单品可获利润的大小由高到低排列，在为正的条件下选入个单品，则可以确定以下三种情形：

①若≥33，则选最大的33个即可。

②若7≤≤33，我们选这个即可。

③若＜27，则重复模拟退火过程，最坏情况下，选此个加上亏损最少的(27-m)个即可。

整体思路已经在上述内容中讨论了，下面代入数据之后，将利润按从高到低的顺序排列如下，三列数据分别是的值：

|  |  |
| --- | --- |
|  | **0** |
| **七彩椒(2)** | (673.1127232208081, 1.3764544291694831, 92.97260296129917) |
| **红椒(2)** | (539.8583971603435, 1.3358737418924693, 92.74785546026553) |
| **螺丝椒** | (453.05533340019804, 1.2933243833283956, 93.28274175333665) |
| **西峡花菇(1)** | (446.92082346009926, 1.443784703960977, 60.72830119334889) |
| **芜湖青椒(1)** | (445.2437994273788, 1.4567456332942663, 89.71676366038692) |
| **外地茼蒿** | (392.68974631917047, 1.3765842025279023, 110.00128718925191) |
| **菠菜** | (343.0566668864039, 1.0677521091537021, 106.74124729636787) |
| **菜心** | (295.22779638625525, 1.4211480245051644, 93.89171876828344) |
| **云南生菜** | (287.0835094740437, 1.3493185505881056, 96.02447591074535) |
| **上海青** | (271.5402916201964, 1.3775499100087012, 94.91642001101519) |
| **红薯尖** | (266.3979921591461, 1.4581910233388917, 88.34647390748104) |
| **云南油麦菜** | (263.7930309201696, 1.3588017223204543, 93.27103280013581) |
| **螺丝椒(份)** | (231.3368842424091, 1.3294102481612642, 92.71189050293178) |
| **娃娃菜** | (202.1432310127197, 1.4525177613478457, 82.98209983463532) |
| **青线椒(份)** | (198.08702976760017, 1.4894959251342756, 93.5884158857457) |
| **竹叶菜** | (197.9811130215381, 1.4912335146979563, 93.56848886926397) |
| **小青菜(1)** | (176.93004620510473, 1.378778587017708, 90.56954152179792) |
| **奶白菜** | (168.05416246730502, 1.244353161477371, 97.68751254560883) |
| **青红杭椒组合装(份)** | (161.93010923719837, 1.2285467170059137, 92.1450492590311) |
| **小米椒(份)** | (158.86626657363877, 1.49344308262252, 93.60963040740535) |
| **苋菜** | (141.3547983214569, 1.3581985656243556, 99.81169337494643) |
| **西兰花** | (140.35177881656483, 1.3518919409305492, 15.503417549910793) |
| **长线茄** | (130.5800089513238, 1.2861006623004674, 17.961654810608938) |
| **姜蒜小米椒组合装(小份)** | (125.40406717486745, 1.3062356464518505, 92.58257091982374) |
| **云南生菜(份)** | (96.93631371993663, 1.214678392271322, 91.43511293133423) |
| **紫茄子(1)** | (91.76629609365621, 1.2791028572660814, 18.49136328947295) |
| **紫茄子(2)** | (82.68310730695643, 1.3306294184013239, 17.631927414300065) |
| **小皱皮(份)** | (81.50964680720573, 1.3811792083523777, 92.9986490963087) |
| **圆茄子(2)** | (77.19668557462622, 1.2703153924729904, 17.98634685275068) |
| **枝江青梗散花** | (60.36456430111679, 1.2547636422802613, 15.701205039028638) |
| **双孢菇(盒)** | (58.972943581690814, 1.4097564073391382, 54.27436869524987) |
| **云南油麦菜(份)** | (55.3577807027165, 1.114028533886642, 94.08460656251988) |
| **青茄子(1)** | (55.15504300238296, 1.4460480304342958, 17.00184163220531) |
| **虫草花(份)** | (53.8264554021261, 1.297260936143932, 59.77690522735789) |
| **野生粉藕** | (38.71204096528369, 1.290987066765081, 12.51579147292112) |
| **金针菇(盒)** | (34.23102177360124, 1.4240717234440088, 54.41243223513795) |
| **菱角** | (32.58041193789386, 1.4751228869203163, 11.836018262893237) |
| **海鲜菇(包)** | (31.843169861499415, 1.3651496419652223, 54.15859189623394) |
| **木耳菜** | (22.048739703040237, 0.8133110694960966, 119.13267022967322) |
| **蟹味菇与白玉菇双拼(盒)** | (15.90320268816366, 1.4904985647998013, 54.632184890950484) |
| **红莲藕带** | (12.2764573808103, 1.3385102205072565, 13.022039493894608) |
| **净藕(1)** | (7.727282976085768, 1.443208734039458, 11.359403287384351) |
| **白玉菇(袋)** | (5.726700654297119, 1.24075150720779, 57.917760308424995) |
| **菠菜(份)** | (0.1426421737838997, 1.3236423043951862, 90.05032167863749) |

显然满足条件①的情况，因此选最大的33个即可，如下表所示。

|  |  |
| --- | --- |
|  | **0** |
| **七彩椒(2)** | (673.1127232208081, 1.3764544291694831, 92.97260296129917) |
| **红椒(2)** | (539.8583971603435, 1.3358737418924693, 92.74785546026553) |
| **螺丝椒** | (453.05533340019804, 1.2933243833283956, 93.28274175333665) |
| **西峡花菇(1)** | (446.92082346009926, 1.443784703960977, 60.72830119334889) |
| **芜湖青椒(1)** | (445.2437994273788, 1.4567456332942663, 89.71676366038692) |
| **外地茼蒿** | (392.68974631917047, 1.3765842025279023, 110.00128718925191) |
| **菠菜** | (343.0566668864039, 1.0677521091537021, 106.74124729636787) |
| **菜心** | (295.22779638625525, 1.4211480245051644, 93.89171876828344) |
| **云南生菜** | (287.0835094740437, 1.3493185505881056, 96.02447591074535) |
| **上海青** | (271.5402916201964, 1.3775499100087012, 94.91642001101519) |
| **红薯尖** | (266.3979921591461, 1.4581910233388917, 88.34647390748104) |
| **云南油麦菜** | (263.7930309201696, 1.3588017223204543, 93.27103280013581) |
| **螺丝椒(份)** | (231.3368842424091, 1.3294102481612642, 92.71189050293178) |
| **娃娃菜** | (202.1432310127197, 1.4525177613478457, 82.98209983463532) |
| **青线椒(份)** | (198.08702976760017, 1.4894959251342756, 93.5884158857457) |
| **竹叶菜** | (197.9811130215381, 1.4912335146979563, 93.56848886926397) |
| **小青菜(1)** | (176.93004620510473, 1.378778587017708, 90.56954152179792) |
| **奶白菜** | (168.05416246730502, 1.244353161477371, 97.68751254560883) |
| **青红杭椒组合装(份)** | (161.93010923719837, 1.2285467170059137, 92.1450492590311) |
| **小米椒(份)** | (158.86626657363877, 1.49344308262252, 93.60963040740535) |
| **苋菜** | (141.3547983214569, 1.3581985656243556, 99.81169337494643) |
| **西兰花** | (140.35177881656483, 1.3518919409305492, 15.503417549910793) |
| **长线茄** | (130.5800089513238, 1.2861006623004674, 17.961654810608938) |
| **姜蒜小米椒组合装(小份)** | (125.40406717486745, 1.3062356464518505, 92.58257091982374) |
| **云南生菜(份)** | (96.93631371993663, 1.214678392271322, 91.43511293133423) |
| **紫茄子(1)** | (91.76629609365621, 1.2791028572660814, 18.49136328947295) |
| **紫茄子(2)** | (82.68310730695643, 1.3306294184013239, 17.631927414300065) |
| **小皱皮(份)** | (81.50964680720573, 1.3811792083523777, 92.9986490963087) |
| **圆茄子(2)** | (77.19668557462622, 1.2703153924729904, 17.98634685275068) |
| **枝江青梗散花** | (60.36456430111679, 1.2547636422802613, 15.701205039028638) |
| **双孢菇(盒)** | (58.972943581690814, 1.4097564073391382, 54.27436869524987) |
| **云南油麦菜(份)** | (55.3577807027165, 1.114028533886642, 94.08460656251988) |
| **青茄子(1)** | (55.15504300238296, 1.4460480304342958, 17.00184163220531) |

问题3到此结束。

**3.3问题4——开放性的影响因子挖掘**

其一，商超还需要了解全年内每天的客流量。很典型的例子是某些节日前夕及当天某种单品的销量有上升趋势，这时需要增大补货量来为升高的客流量做好充足的准备。例如在春节前夕，人们要大量购买各种蔬菜原材料为过节做准备，显然会产生更高的需求量。如果有充足的全年每天的客流量数据，我们就可以引入客流量的影响函数，这个函数是独立于加成参数而影响利润的，并且提醒我们在客流量较高的时候适当增大整体补货量。

其二，获取一天内的每个时间段内的客流量。一天内的不同时间段，消费者的消费欲望也是不同的，注意到每天的所有单品鲜有打折销售的记录，所以我们可以针对一天内客流量较大的时间段集中投放打折产品在商超内较为显眼的位置，也有一定的宣传效果。进一步地，非常有创造性的想法是我们可以在一天内的若干时间段将同一个正常价销售的单品做定价差异化处理，类似的做法也可以用在打折率的处理上，可以进一步精细化定价策略。

其三，消费者的反馈和意见也十分重要。我们认为商超应该定期听取消费者的声音，这可以帮助商家了解到哪些品类或者单品更收到顾客的欢迎，以及定价策略应该如何调整。

其四，在问题2的解决过程中，总销售量与加成参数之间的回归函数关系至关重要，而这个函数是有确界的，上下确界分别为和。题中场景由于空间限制、时间限制和人力资源限制等等原因而存在阻滞性，即使商品极其便宜，由于当地人们的购买力和时间成本有限，会存在一个最大销售量和最小的销售量。同时我们注意到定价过了一个临界值之后越低，销量的变化率的绝对值减小，这也是合乎常理的，所以商家要根据未来的销售数据挖掘出比问题2模型中的处理情况更精确的上下确界，并通过提升店内环境、资源管理配置水平等等手段来整体上增大值域。

其五，在问题3的解决过程中，我们面临着数据量很小的问题，这样的问题本质上是因为同类别的数据太少。我们建议该商超与周边的其他商超积极交流，例如为天内该单品未打折销售的总销量与7天内该单品所属品类未打折销售的销售总量之比，为7天内该单品打折销售的总销量与7天内该单品所属品类打折销售的销售总量之比，这两个数据量在模型的产出起到作用，我们希望得到的是有更多数据支撑之下的更精确的值。所以商超需要在交流之中采集更多可用的相关数据，这样就有更多的数据点，我们更能准确估计和的值。

其六，及时确定未来的天气预报数据。常见的台风、暴雨、地震、极端高温等等异常天气会极大削减消费者的消费热情，这时候就不需要在极端天气持续的几天制定激进的补货策略；极端天气前夕亦和节假日前夕的需求侧情况十分类似，大家倾向于囤货，因此客流量与销售量上升，商超应当对此做出准备。

**四、模型评价**

**4.1优点**

对于问题1，首先我们采用了三年全部的数据，有足够多的数据来分析六类蔬菜两两之间的相关性，提升了分析的可靠性。同时又在此基础上添加了按季度

由于附件中的数据经分析后，正态性较差，因此选择了Kendall和Spearman这两个对正态性要求很低的经典相关系数进行相关性分析，而不是选择Pearson相关系数。使结果分析更加可靠。

对于问题2，我们所采用的成本定价模型适用于对于产品有较完整的历史成本信息的情况，与此问题相符。

在寻找销售量与定价关系时，使用了logit模型进行拟合。该模型计算概率时有着更高的准确性，且其概率严格的落入1-0之间，比较合理。且logit模型有较好的直观解释，可认为它是反映在最大值及最小值间最有可能所处的位置。

对于问题2，我们在时间序列分析中选择了Prophet模型，其优势在于灵活性，可以很容易地适应季节性与多个时期，对趋势做出不同的假设。对具有趋势性、周期性等明显内在规律的数据敏感。这是相对于其他时间序列处理办法的优势[5]。且与ARIMA模型相比，其测量不需要有规律的间隔，不需要插入丢失的值，显得更为方便。

在问题2的最后使用了模拟退火算法，其优势在于计算过程简单，适用于并行处理,可用于求解复杂的非线性优化问题.且由于初始解与最终解都是随机选取的，因而具有一定抵御外界不稳定因素的能力。

对于问题3，我们采用化归思想，利用单品与所属品类间关系系数A、B，将利润公式与问题2中所解模型进行关联，在具有合理性的前提下，简化求解过程，且最后得到理想实验结果。

4.2缺点

问题2中使用的模拟退火算法具有收敛速度慢,执行时间长,算法性能与初始值有关及参数敏感等缺点.如果降温过程足够缓慢,得到的解性能会比较好,但与此相对的是收敛速度太慢;如果降温过程过快,很可能得不到全局最优解，因此还存在改进的空间。另外，对于二元logit回归模型来说，常用的拟合优度指标R^2没有特别明显的意义。

**五、参考文献**

[1] Pearson, Spearman, Kendall 三大相关系数简单介绍，[https://zhuanlan.zhihu.com/p/60059869](https://zhuanlan.zhihu.com/p/60059869/)，

2020年5月9日.

[2] 成本加成定价法评介款，韩俊华，干胜道，<https://x.cnki.net/xmlRead/xml.html?pageType=web&fileName=CKYK201222033&tableName=CJFDTOTAL&dbCode=CJFD&topic=&fileSourceType=1&taskId=&from=&groupId=&appId=CRSP_BASIC_PSMC&act=&customReading=>

[3] 姜启源，薛金星，叶俊.数学模型[M].北京：高等教育出版社，2018.

[4] 时间序列预测（五）—— Prophet模型，<https://blog.csdn.net/kewei168/article/details/90375897>，

2019年5月20日.

[5] Prophet时序模型在短期水质溶氧预测中的应用，沈时宇，陈明，<https://x.cnki.net/xmlRead/xml.html?pageType=web&fileName=HDXY202003005&tableName=CJFDTOTAL&dbCode=CJFD&topic=&fileSourceType=1&taskId=&from=&groupId=&appId=CRSP_BASIC_PSMC&act=&customReading=>