大型商超的定价与补货策略研究

摘要

生鲜商超中的蔬菜保质期较短，对大部分的蔬菜，倘若当天没有把蔬菜卖出，隔日就无法再进行售卖。从供给侧角度看，商超负责人在每日补货之前对今日蔬菜的品相和成本价格一无所知。蔬菜的定价采用“成本加成定价”的方式，因此在定价时需要综合考虑各方面因素。对于商超负责人来说，基于以往数据进行可靠的市场分析并制定合理的补货决策和定价决策是非常重要的事情。蔬菜商品的销售量往往与时间有关系。

问题一要求我们分析蔬菜品类及单品销售量的分布规律及相互关系，我们首先分析了蔬菜的品类销售量，发现其销售量随时间呈周期性摆动变化，并利用AR时间模型预测出未来一周的销量情况。随后我们分析了品类之间的相关关系，依据相关系数，我们发现品类间存在较强的相关关系（茄子是个例外，它与其他每一个品类的相关性都比较低）。对于单品，由于单品数量较多，我们先采用聚类分析将单品分类，接着从分类中选取代表性的单品，依据其销售量进行分布分析。对于单品间的销量相关关系，我们依然采用相关性分析，并依据相关性强弱进行了统计性分析。

问题二要求我们以品类为基本单位，分析品类成本加成定价与品类销售总量之间的关系，随后给出一周的补货策略和定价策略使商超获得最大的收益。首先，我们确定了蔬菜品类的成本加成定价和蔬菜品类的表法方式，随后通过线性回归模型，建立起蔬菜品类成本加成定价和蔬菜销售总量之间的线性回归模型，并通过了统计性检验。制定补货和定价策略时，我们首先在蔬菜品类成本和蔬菜销售总量之间建立起线性回归模型并通过假设过去三年蔬菜品类的销售量与补货量大体一致简化了问题，从而预测出未来一周蔬菜品类的成本。随后计算了各品类商品以往的成本加成系数，最后依据问题一中预测的销售量和以往成本加成系数对未来一周的补货量和定价进行约束，从而建立了关于补货量和定价的线性规划模型，继而求得了相应的定价策略和补货策略。

问题三要求我们根据6月24日-6月30日的蔬菜单品销售情况。制定7月1日的定价策略和补货策略。首先，我们利用GM(2,1)灰色预测模型，依据6月24日-6月30日在售蔬菜单品的小样本销售量数据预测出了7月1日各单品的销售量。接着，根据要求，通过单品选择数量、最小陈列量、补货量、成本加成系数的相关要求给出了约束条件，继而建立了关于补货量和定价的线性规划模型，继而求得了相应的定价策略和补货策略。

问题四为开放性问题，需要我们从多方面、多维度(如商超所在位置、蔬菜商品季节特性等)考虑收集数据的类型。经过综合比较，最终选择出应收集的数据为：供应链数据、季节因素数据、客流量数据、库存容量数据、顾客反馈数据、市场需求数据、竞争对手数据等。

关键词：成本加成定价、AR时间序列、相关性分析、线性回归、线性规划、灰色预测

一、问题重述

1.1 问题背景

蔬菜类商品属于生鲜产品，具有一些明显的特点：一是蔬菜产品的生长与销售具有典型的周期性和季节性，容易腐烂变质；二是蔬菜产品形态脆弱，运输或加工处理过程不当均会导致损耗；三是蔬菜消费存在刚需、高频的消费特点，在生活中几乎不可或缺。因此，大型商超基于各商品的历史销售和需求情况制定蔬菜类商品的自动定价与补货策略具有重要现实意义。

1.2 要解决的具体问题

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，为此商家需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。在此条件下的各蔬菜品类的补货决策和定价策略需要满足营业需求。可靠的市场需求分析，对补货决策和定价决策尤为重要。建立数学模型解决以下问题：

问题一：附件一和附件二给出了各品类蔬菜信息和蔬菜单品的销售数据，通过对大量数据进行处理和分析，探究不同品类或不同单品之间可能存在的关联关系。分析蔬菜各品类及单品三年来的销售量的分布规律及相互关系。

问题二：蔬菜的定价一般采用“成本加成定价”方法。现在该商超以品类为单位做补货计划，请根据附件1、2、3的商品批发与销售信息，分析各蔬菜品类的销售总量与成本加成定价的关系，对未来进行合理预测，提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策略，使得商超收益最大。

问题三：因蔬菜类商品的销售空间有限，商超希望进一步制定单品的补货计划，要求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，在尽量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。

问题四：为了更加合理、精准的制定蔬菜商品的补货和定价决策，商超还需要采集哪些相关数据，这些数据对解决上述问题有何帮助。请给出建议并说明理由。

二、问题分析

2.1 对问题一的分析

针对问题一，第一小问要求探究蔬菜不同品类或单品的销售量之间可能存在的关联关系，处理此类问题通常进行相关性分析。分析之前要先对数据进行预处理，对附件中数据进行编号，将无效数据和异常值点剔除等。数据处理完成后分别对各品类和单品进行相关性分析，计算得到不同品类及不同单品之间的相关系数，结合市场规律分析其中体现的相互关系的合理性。第二小问要求探究各品类和单品销售量的分布规律，我们推测不同品类的蔬菜销售量可能存在周期性、季节性的规律，且有些品类之间具有强相关性，故我们选出几个有代表性的品类讨论其分布规律即可。以天为单位统计出各品类蔬菜每日的销售量并对其构建时间序列，在此之前先对各品类的时间序列进行Daniel平稳性检验，对非平稳时间序列进行差分运算并采用最小二乘法计算出预测模型，构建出模型后对它随时间的分布规律进行描述，同时通过该模型预测未来时间的销售量；为了更简便地分析251种蔬菜单品的分布规律，首先对蔬菜单品的销售量进行聚类分析，并从每一种分类中选出具有代表性的蔬菜单品并绘制出其随时间的分布规律。

2.2 对问题二的分析

对于问题二第一小问，要探究各蔬菜品类的销售总量与成本加成定价的关系，需建立销售总量

与各品类蔬菜定价的回归方程并进行假设检验，为了建立线性规划模型求目标函数的最优解还需要将蔬菜品类的成本与销售量建立关系，在假设销售量等于补货量的前提下我们建立起回归方程并进行检验。最后建立线性规划模型以求出收益的最大值和对应的定价、补货策略。

2.3 对问题三的分析

分析题目过后将问题三拆分成两个子问题，一是在前一个星期内所售的所有单品种选择27-33种进行补货销售。考虑到应满足总收益最大的目标，我们首先将前一星期售卖过的所有蔬菜单品名单整合，分别计算它们的利润率按降序进行排列，选出利润贡献率最大的33个蔬菜单品进行销售。该问题属于小样本短期预测问题，即通过前一个星期的短期数据来预测第八天的销售情况，综合考虑到小样本预测模型的准确性，选择GM(2,1)灰色预测模型进行预测。第二个子问题仍然是在保证收益最大的前提下制定补货和定价策略，所以依然采用线性规划模型对目标函数进行求解，并且在求解的同时要让蔬菜单品的数目在27-33间进行循环，最终比较单品数目不同的情况下总收益的大小，从而选择最适合的蔬菜单品数目，此时的收益也相应达到最大值。

三、数据预处理

3.1 数据编号

数据预处理是数据分析的关键步骤。首先对蔬菜品类、蔬菜单品和日期进行编号，花叶类、花菜类、水生根茎类、茄类、辣椒类、食用菌类等六大类蔬菜商品的编号依次为类别1，2，3，4，5，6；单品按照附件1顺序编号依次为1，2，3……251；2020年7月1日到2023年6月30日的日期编号依次为1，2，3，4……1095。这些处理为我们统计单品销售量和品类销售量提供了更加便捷的思路。

3.2 无效数据剔除

完成编号后，统计单品和品类销售量，导入附件2数据后发现存在极少量数据销售类型为“退货”，此时销售量显示为负，存在异常值，因此我们要对数据进行清洗，将此类异常值点剔除，得到Singlesales和Typesales的新数据文件。

商品销售量和时间所建立的散点图在分析商品定价与补货决策中具有重要地位，因为其能够清晰地展示商品销售量随时间的变化趋势，揭示商品销售量的季节性变化，帮助识别销售量的异常波动，同时对商超管理人员进行未来销售量的预测、制定销售计划、库存管理和生产规划具有重要意义。 因此，我们对各种品类销售量数据进行描述性分析，以天为单位绘制出蔬菜各品类销量的散点图，发现第225，579，780，874，934天蔬菜的销售量异常高，经过分析推理得出这些天数恰好对应春节等节假日前夕，人们在此时会提前大量采购蔬菜商品，导致销售量激增，这属于合理的市场需求，因此这些值作为正常值使用即可。同时发现有少部分日期所有品类销售量为零，经过分析推断这些天数商超可能处于未开业状态，因此这些数据属于异常值，需要剔除。综上，经过对原始

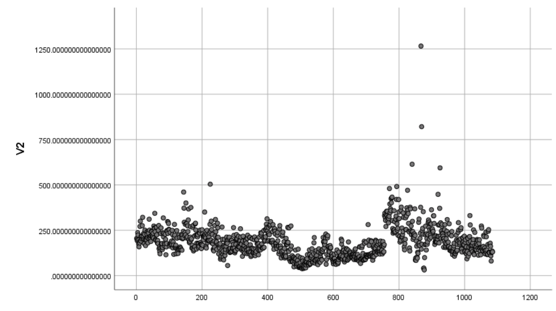


图 1: 品类*1*散点图

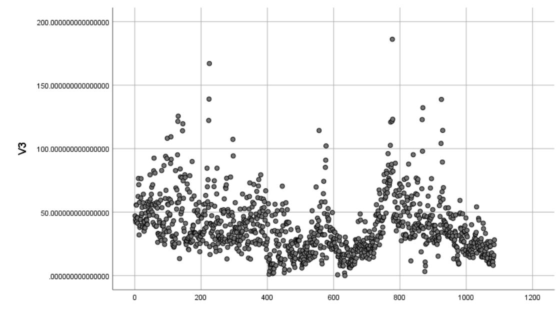


图 2: 品类*2*散点图

数据进行预处理，我们得到了一组更加合理、准确、切合实际的新数据。

四、符号说明

表 1: 符号说明

符号	含义
$S_{x,t}$	第x类商品第t天销售额
$b_{x,t}$	平稳处理后的第x类商品时间序列
e_i	第i类蔬菜品销售量
$W_{x,t}$	第x类第t天的销售额
$Q_{x,t}$	第x类蔬菜商品第t天的成本
C_{ij}	第i类蔬菜商品成本在第j天成本价
s_i	第i类蔬菜品销售额
X_{ij}	第i类蔬菜商品在第j天补货量
Y_{ij}	第i类蔬菜商品在第j天的定价
P_{ij}	预测出来第i类在第j天的销售量
per_i	蔬菜商品的利率
x_i	利润贡献率排名第i位的单品的补货量
y_i	利润贡献率排名第i位的单品的定价
m_i	利润贡献率排名第i位的单品的成本
l_i	为利润贡献率排名第i位的单品的损耗率
p_i	利润贡献率排名第i位的单品在2023.7.1当天预测出来的销售量

五、模型的建立与求解

5.1 问题一的分析与求解

5.1.1 相关性分析

首先考虑到人们在购买蔬菜时通常是一次性购买多个种类进行烹饪使营养达到均衡，因此蔬菜各品类以及单品销售量之间存在着一定关联关系，我们先对蔬菜的六个品类之间的每日销售量进行相关性检验，引入皮尔逊相关系数，即

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \tag{1}$$

式中,$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$

通过计算六种品类销售量之间的相关性系数，构建一个6*6的相关性系数矩阵以反映六种品类两两之间的相关性系数，从该矩阵可以分析出不同品类之间的相互关系，帮助商超更好地调整库存、制定销售策略，并提高销售效率。将六个品类之间的相关系数矩阵绘制成热图：

我们认为相关系数的绝对值在0.5以上的两者具有强相关性，在0.3-0.5之间的两者具有中度相关性，在0.1-0.3之间的两者具有弱相关性。分析该矩阵发现：1. 茄类与其他五种品类呈现弱的正相

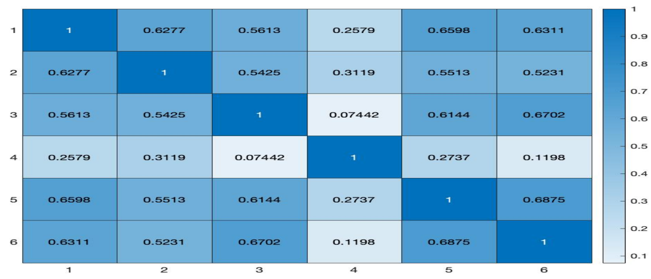


图 3: 相关性系数热图

关性。弱正相关性可能部分源于季节性因素，茄类在夏季销售量增加，而其他品类可能在不同季节有销售高峰，这种季节性差异会导致相关性较弱。2.除茄类外其他品类之间具有强的正相关性，这意味着这些蔬菜销售量的增加或减少具有相似的趋势，存在两种蔬菜在某些消费者的购物清单中常常一起出现的可能，例如：类别5和类别6的相关性系数为0.6938，很大程度是因为它们在烹饪中经常一同使用。基于此，我们可以在销售空间的限制下，优化商品的销售组合。

接着我们再分析251种单品之间的相互关系，我们对这些单品两两进行相关性分析，带入相关系数的计算公式中得到的相关系数矩阵，进一步分析得到其中有5289对弱相关对，1233对中等相关对以及405对强相关对。

5.1.2 商品销售量分布规律

继续观察六种品类销售量的时间序列曲线图，发现各个品类之间的销售量有所差异，并且发现销售量分布曲线呈现周期性波动，显示出明显的季节性变化，即销售量在特定时间段内有规律地增加或减少，时间序列模型可以用来捕捉这种季节性趋势。因此，我们采用时间序列模型预测销售量与日期之间的关系。品类3、4的时间曲线和六个品类汇总的时间曲线图如下所示。

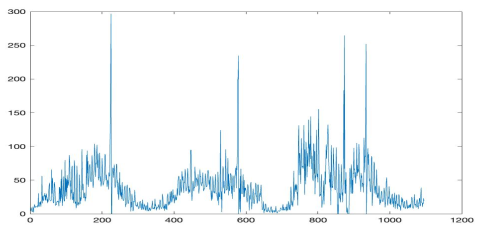


图 4: 品类3曲线图

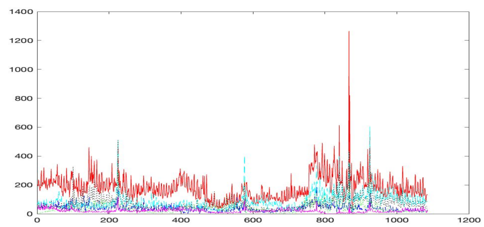


图 5: 六大品类曲线图

接着我们继续探究单品销售量的分布规律，由于单品的种类较多，不便将每个单品的分布规律一一呈现，因此我们对251种蔬菜单品的销售量进行聚类分析，并对251种蔬菜单品的销售量做统计性分析得到表2所示各项指标。

表 2: 销售量统计数据

平均值	1877.592984
中位数	212.1710000
标准偏差	4168.970768
方差	17380317.26
最小值	0.000000000
最大值	28181.74100

探究蔬菜单品随时间的分布规律时整个时间周期的长度为三年，相当于对每一个蔬菜单品而言都有1095个指标，所以应先对每一个单品的指标进行聚类，对指标的数据进行标准化处理，变量间相近性度量采用相关系数，类间相似性度量的计算选用类平均法。聚类树形图如图6、图7所示。

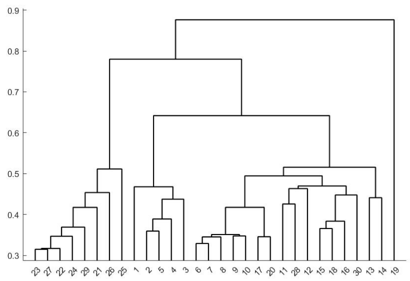


图 6: 日期树形聚类图

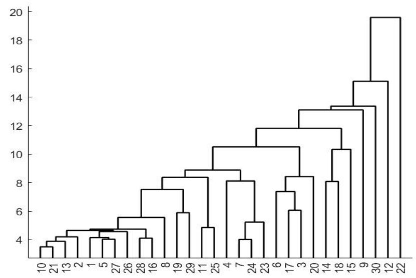


图 7: 单品树形聚类图

从日期树形聚类图可以看出数据被分成了12类，即最终对一个蔬菜单品而言保留下了12个指标。接着再根据这12个指标对251个单品进行聚类，对每个变量的数据分别进行标准化处理，样本间相似性采用欧几里得距离度量，类间距离的计算选用类平均法。

从单品树形聚类图可知251个蔬菜单品被分成5类，我们选取每一类中具有代表性的单品对它的分布规律进行分析，这个蔬菜单品便近似代表了这一类蔬菜单品的分布规律。其五种分类的代表性蔬菜单品的分布规律如下文蔬菜分布规律图所示。

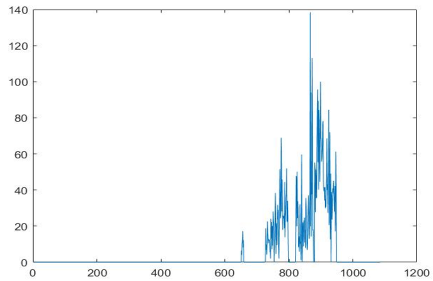


图 8: 第一类蔬菜分布规律图

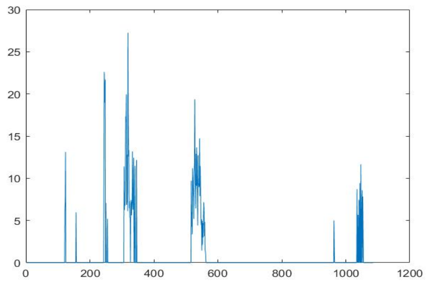
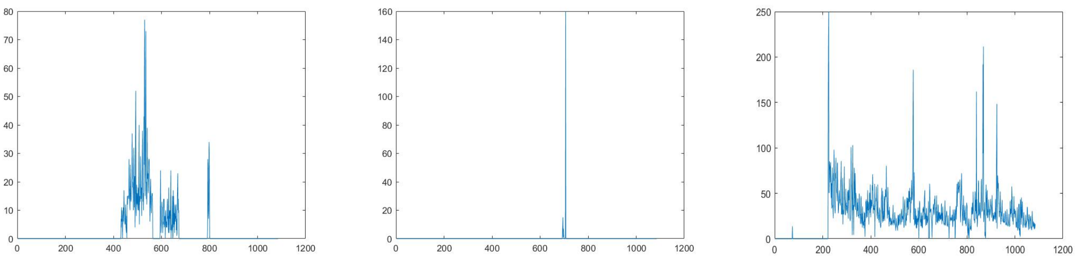


图 9: 第二类蔬菜分布规律图



(a) 第三类

(b) 第四类

(c) 第五类

图 10: 第三、四、五类蔬菜分布规律图

第一类蔬菜单品的分布规律是非周期性的，且销售量都集中在某段时间内；第二类蔬菜单品的分布更加离散，销售量变化的幅度较大并且持续时间短，推测该单品蔬菜的食用期较短且有季节性；第三类蔬菜单品的分布规律既有集中分布的情况又有类似尖峰脉冲的分布，考虑到可能是特殊时期或突发情况导致的销售量激增，并不具有周期性；第四类蔬菜单品的分布规律最直观，从图中可以看到只有某个特定的时间才有销售量，并且销售量很大，推测可能是为代替其他单品蔬菜的紧急缺货而大量进货；第五类蔬菜单品的分布规律具有周期性，并且销售量较大，且一直存在销售量，考虑可能是热销蔬菜菜品。

最后还可以以秒为时间单位分析一天中购买各个品类和单品的蔬菜分布规律，通过描述性分析将离散数据进行整合，同时绘制出散点图后我们得到如下的分布规律：顾客多在上午购买花菜类和花叶类蔬菜，同时对菠菜这种单品蔬菜分析可以看出购买时间也多集中在上午，这种现象符合市场规律，这类青菜往往在上午更加新鲜，因此也会有更多的顾客选择在上午时段购买。

5.1.3 时间序列AR模型

在构建时间序列模型之前，要先检验序列的平稳性，在此我们引入Danil检验方法, 该方法建立在Spearman相关系数的基础上。

Spearman相关系数是一种秩相关系数，定义为这两组秩统计量的相关系数，即Spearman相关系数是

$$q_{XY} = \frac{\sum_{i=1}^{n}(R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^{n}(R_i - \bar{R})^2}\sqrt{\sum_{i=1}^{n}(S_i - \bar{S})^2}} \tag{2}$$

式中：$\bar{R} = \frac{1}{\mathbf{n}}\sum_{i=1}^{\mathbf{n}} R_i; \bar{S} = \frac{1}{n}\sum_{i=1}^{n} S_i$。经过运算，可以证明

$$q_{XY} = 1 - \frac{6}{n(n^2-1)}\sum_{i=1}^{n} d_i^2 \tag{3}$$

式中：$d_i = R_i - S_i, i = 1, 2, \cdots, n$。

对于Spearman相关系数，作假设检验

$$H_0 : \rho_{XY} = 0, H_1 : \rho_{XY} \neq 0$$

式中：ρ_{XY}为总体的相关系数，可以证明，当（X,Y）是二元正态总体，且H_0成立时，统计量

$$T = \frac{q_{XY}\sqrt{n-2}}{\sqrt{1-q_{XY}^2}}$$

服从自由度为$n-2$的t分布$t(n-2)$

对于给定的显著水平α ，通过t分布表可查到统计量T的临界值$t_{\frac{\alpha}{2}}(n-2)$,当$|T| \leqslant t_{\frac{\alpha}{2}}(n-2)$时，接受$H_0$；当$|T| \geqslant t_{\frac{\alpha}{2}}(n-2)$时，拒绝$H_0$。

对于时间序列X_t的样本$a_1, a_2, \cdots, a_n$,记a_t的秩为$R_t = R(a_t)$,考虑变量对(t, R_t),$t = 1, 2, 3, \cdots, n$的Spearman相关系数q_s,有

$$q_x = 1 - \frac{6}{n(n^2-1)}\sum_{i=1}^{n}(t - R_i)^2 \tag{4}$$

构造统计量

$$T = \frac{q_s\sqrt{n-2}}{\sqrt{1-q_s^2}}$$

作下列假设检验

H_0:序列X_t平稳；

H_1:序列X_t非平稳（存在上升或下降的趋势）。

Daniel检验方法：对于显著水平α,由时间序列$a_t(t,R_t),t=1,2,3,\cdots,n$ 的Spearman秩相关系数q_s,若$|T|>t_{\frac{\alpha}{2}}(n-2)$，则拒绝$H_0$，认为序列非平稳。且当$q_s>0$时，认为序列有上升趋势；$q_s<0$时，认为序列有下降趋势。又当$|T|\leqslant t_{\frac{\alpha}{2}}(n-2)$ 时，接受H_0 ,可以认为X_t 是平稳序列

对六种品类的时间序列进行平稳性Daniel检验后，得到如下图所示的统计量T的分布表。

T	-2.5546	-8.4062	0.7671	-9.3868	13.9872	3.3764

对于给定的显著水平$\alpha=0.05$，通过t分布表查到统计量T的临界值为$t_0=1.9622$。只有当$|T|\leqslant t_0$时我们才认为该种品类的时间序列通过了置信水平为0.95的平稳性检验，比较可知只有品类3的时间序列通过了平稳性检验，其它的品类都未通过检验。计算出Spearman相关系数q_s的值如下所示。

q_s	-0.0774	-0.2475	0.0233	-0.2743	0.3912	0.1021

除了品类三是平稳时间序列外其它的序列都是非平稳的，我们可以通过比较q_s与0的大小关系判断出非平稳序列的变化趋势，由上表可知品类一、二、四的时间序列呈下降趋势；品类五、六的时间序列呈上升趋势。结合实际情况分析，花叶类、花菜类和茄类的销售量随时间呈下降趋势，其中花叶类和花菜类的相关性很强，这类蔬菜品种可能具有季节性生长周期，导致它们在不同季节的供应量和价格发生变化如果花叶类、花菜类和茄类蔬菜在某些季节供应过剩，价格可能下降，从而降低了销售量.

辣椒类和食用菌类的销售量随时间呈现增长趋势，且辣椒和食用菌这两种品类的蔬菜在上题中的相关系数很高，并且某些健康和饮食趋势可能会影响消费者的购物选择。辣椒类和食用菌类蔬菜被认为对一些人的饮食习惯和健康有益，这可能导致它们的销售量上升。也可能有越来越多的消费者对辣椒类和食用菌类的蔬菜产品表现出兴趣和需求。这可能是因为人们的口味逐渐变得更加多样化，或者因为这些食材被认为具有一定的营养价值而导致销售量成上升趋势，从实际情况和市场需求来分析这些品类蔬菜的时间序列的变化规律都是相对合理的。

之后我们分别对通过平稳性检验的品类3和并未通过检验的其余品类利用AR模型得到各品类蔬菜的时间预测序列模型。首先单独考虑品类3，在AR模型的基础上采用最小二乘法进行拟合得到如下所示的销售量$S_{3,t}$的表达式：

$$S_{3,t}=0.6383*S_{3,t-1}+0.2631*S_{3,t-2}+\varepsilon_t(t=1,2,\cdots,7) \tag{5}$$

接着为了构造平稳序列，我们对除了品类3以外的时间序列$S_{i,t}(t=1,2,4,5,6)$作一阶差分运算：$b_t=S_{i,t+1}-S_{i,t}$, 得到序列$b_t(t=1,2,4,5,6)$。从时间序列b_t散点图来看，时间序列是平稳的。可建立如下的自回归模型对b_t进行预测：

$$y_t=c_1y_{t-1}+c_2y_{t-2}+\varepsilon_t \tag{6}$$

式中：c_1,c_2为待定参数；ε_t为随机扰动项。这时对时间序列a_t的预测就变成了对时间序列bt的预测。

然后采用最小二乘法可计算出各品类b_t的预测模型，如下所示：

$$\begin{cases} b_{1,t} = 1 + 0.5535b_{1,t-1} + 0.4027b_{1,t-2} - 2 \\ b_{2,t} = 1 + 0.7599b_{2,t-1} + 0.1659b_{2,t-2} - 2 \\ b_{4,t} = 1 + 0.7064b_{4,t-1} + 0.2231b_{4,t-2} - 2 \\ b_{5,t} = 1 - 0.2599b_{5,t-1} - 0.2116b_{5,t-2} - 2 \\ b_{6,t} = 1 - 0.3458b_{6,t-1} - 0.1847b_{6,t-2} - 2 \end{cases} \tag{7}$$

根据以上时间序列的预测模型我们可以预测出2023.7.1-2023.7.7的各个品类蔬菜的总销售量如下表所示：

表 3: *2023.7.1-2023.7.7*的各个品类蔬菜总预测销售量

日期 / 品类	7.1	7.2	7.3	7.4	7.5	7.6	7.7
品类1	126.6054	122.6070	118.8403	115.1454	111.5837	108.1245	104.7757
品类2	23.3850	23.9490	22.1096	21.0016	19.6763	18.4356	17.2730
品类3	18.4329	16.8757	15.6223	14.4124	13.3104	12.2885	11.3463
品类4	19.8953	19.5256	18.2306	17.2334	16.2401	15.3160	14.4116
品类5	79.7052	81.8209	81.8172	81.3704	81.4873	81.5515	81.5100
品类6	43.5975	43.8207	42.9998	43.2424	43.3102	43.2419	43.2530

为了验证已经建立的预测模型和真实的时间序列之间的拟合精度，我们做出了六种品类的真实时间序列和预测模型的曲线图如下所示，经过对比和误差分析后我们认为建立的该预测模型准确度较高，模型建立合理。

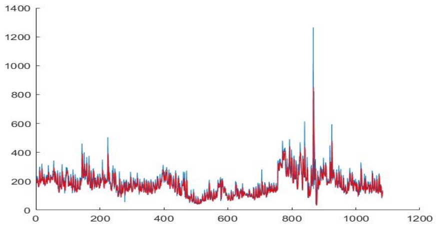


图 11: 品类1拟合曲线图

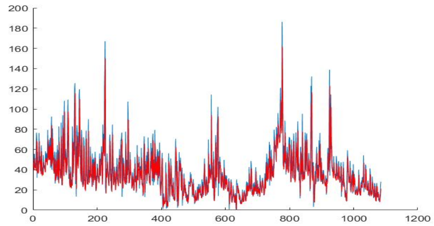


图 12: 品类2拟合曲线图

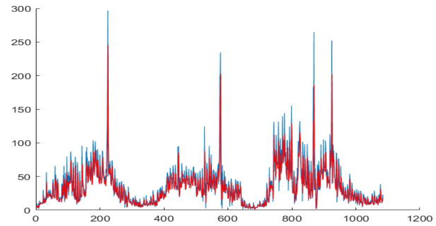


图 13: 品类3拟合曲线图

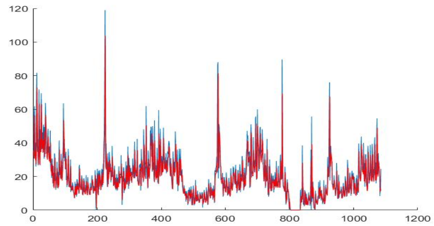


图 14: 品类4拟合曲线图

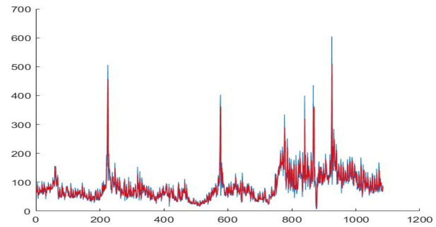


图 15: 品类5拟合曲线图

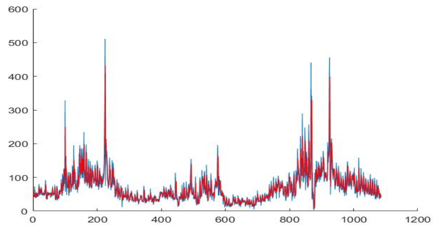


图 16: 品类6拟合曲线图

5.2 问题二的分析和求解

5.2.1

商超中蔬菜产品的定价与销售量之间存在密切的关系，这种关系通常可以总结为供需关系。价格是消费者购买决策的关键因素之一。一般来说，价格下降可能会导致销售量的增加，而价格上涨可能会导致销售量的减少。这是基本的供需关系原理，称为价格弹性。如果某种蔬菜的价格较低，消费者更有可能购买更多的数量，反之亦然。为了分析各蔬菜品类的销售总量与成本加成定价的关系，我们需要建立销量和定价之间的回归模型。

在进行回归分析之前我们要先进行数据处理，通过遍历附件二中的产品编号与附件一中的蔬菜品类信息进行对比后判断出产品编号所对应的蔬菜品类，之后利用附件三中给出的蔬菜批发价的信息和附件二中各品类蔬菜的销售量以及售价信息重新生成两个表格：TypePrice(0)和TypeCost(0), 分别为六个品类的蔬菜的每日销售额和每日的进货总成本。同时我们计算出成本加成系数（利润率），其公式如下所示：

$$p_{x,t} = \frac{W_{x,t} - Qx,t}{Q_{x,t}} \tag{8}$$

之后，我们将每日的成本加成系数数据放入新建的表格CPCType中。接着建立六个品类的蔬菜的销

售量与成本加定价间的关系。六个品类建立的多元线性回归[2]分析的模型均为

$$\begin{cases} y = \beta_0 + \beta_1 x + \varepsilon \\ \varepsilon \sim N(0, \sigma^2) \end{cases} \tag{9}$$

式中：β_0,β_1,σ^2都是与x,无关的未知参数，其中β_0,β_1称为回归系数

多元线性回归模型式中的参数$\beta_0, \beta_1, \ldots, \beta_m$用最小二乘法估计，即应选取估计值$\hat{\beta}_i$,使当$\beta_j = \hat{\beta}_i, j = 0, 1, \ldots, m$时，误差平方和

$$Q = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (b_i - \hat{b_i})^2 = \sum_{i=1}^{n} (b_i - \beta_0 - \beta_1 a_{i1} - \ldots - \beta_m a_{im})^2 \tag{10}$$

达到最小。为此，令

$$\frac{\partial Q}{\partial \beta_j}, j = 0, 1 \tag{11}$$

经过计算建立起六种品类的销售量和定价的线性回归方程如下所示

$$\begin{cases} d_1 = 228.0468 + 4.1786 * e_1 \\ d_2 = 43.9224 + 7.8461 * e_2 \\ d_3 = 12.0419 + 8.3019 * e_3 \\ d_4 = 3.3346 + 8.3582 * e_4 \\ d_5 = -52.0318 + 8.8552 * e_5 \\ d_6 = 68.4746 + 7.1604 * e_6 \end{cases} \tag{12}$$

最后对回归模型进行假设检验，经过计算，在显著性水平0.01下，查表得到上α分位数$F_\alpha = 6.6583$，并且六种品类所建立的回归模型的统计量F的值如下表所示:

$F(*10^3)$	1.6066	4.0454	5.9749	2.7029	1.0767	4.1745

通过上表可以看出我们所建立的六个线性回归方程的统计量F均大于F_α,因此全部通过了假设检验。

接下来为了制定合理的定价策略，我们不得不考虑成本的问题，即要考虑成本由什么因素来预估，因为通过第一问我们建立的时间序列AR模型可以预测出2023.7.1-2023.7.7的销售量，而且我们知道商品的补货量和成本之间存在着相关关系，同时考虑到市场需求，因此作出如下假设：补货量与销售量基本保持一致。这样以来就可以通过建立成本与销售量间的关系从而间接建立成本和补货量的回归方程。接下来分别建立六个品类成本与销售量间的多元线性回归方程，其回归的模型的形式与求解过程与上文基本一致。得到：

$$\begin{cases} C_1 = 178.6401 + 2.2837 * s_1 \\ C_2 = 42.5606 + 4.7476 * s_2 \\ C_3 = 10.2485 + 5.6925 * s_3 \\ C_4 = 3.9936 + 5.2329 * s_4 \\ C_5 = 7.8881 + 5.0873 * s_5 \\ C_6 = 42.6843 + 4.5627 * s_6 \end{cases} \tag{13}$$

$F(*10^3)$	1.1543	3.4261	7.1523	2.1946	0.7573	4.1026

最后对回归模型进行假设检验。经过计算，在显著性水平0.01下，查表得到上α分位数$F_\alpha = 6.6583$，并且六种品类所建立的回归模型的统计量F的值如上表所示

通过上表可以看出我们所建立的六个线性回归方程的统计量F均大于F_α,因此全部通过了假设检验。

建立完回归方程后我们建立线性规划模型以求出目标函数（即收益）的最大值，首先对参数

目标函数：

$$max \quad \sum_{j=1}^{7}\sum_{i=1}^{6} Y_{ij} - \sum_{j=1}^{7}\sum_{i=1}^{6} C_{ij} \tag{14}$$

约束条件为

$$\begin{cases} \mathbf{A} = \mathbf{C} + a\mathbf{B} \\ 0 \leqslant X_{ij} \leqslant P_{ij} \\ per_{imin} \leqslant \dfrac{Y_{ij} - C_{ij}}{C_{ij}} \leqslant per_{imax} \end{cases} \tag{15}$$

记矩阵数组

$$\mathbf{A} = \begin{pmatrix} C_{11} & C_{12} & \dots & C_{1j} \\ C_{21} & C_{22} & \dots & C_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ C_{i1} & C_{i2} & \dots & C_{ij} \end{pmatrix}$$

$$\mathbf{B} = \begin{pmatrix} X_{11} & X_{12} & \dots & X_{1j} \\ X_{21} & X_{22} & \dots & X_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ X_{i1} & X_{i2} & \dots & X_{ij} \end{pmatrix}$$

$$\mathbf{C} = \begin{pmatrix} D_{11} & D_{12} & \dots & D_{1j} \\ D_{21} & D_{22} & \dots & D_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ D_{i1} & D_{i2} & \dots & D_{ij} \end{pmatrix}$$

式中$D_{(ij)} = b$

在此我们需要对约束条件进行解释：

1. 第一个约束是成本和销售量间的线性回归方程
2. 第二个约束的含义是进货量大于等于0且小于等于销售量。从市场需求考虑，由题意可知，在生鲜商超中，一般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，隔日就无法再售，因此在进货的时候在满足需求的前提下尽量不要蔬菜这类短期商品出现大量滞销的情况，故考虑补货量应该小于等于销售量；从数学角度考虑，经过对六个品类蔬菜的利润率（成本加成系数）进行计算分析后发现其集中分布在0.6附近，因此我们假设在销售量即需求量为一定值的情况下，补货量与销售量的差值也固定，例如对于销售量为3000的情况，补货量为2900或3100，当补货量超出销售量100kg时，会出现商品滞销现象，此时与供需平衡的情况进行比较（即销售量=补货量=3000时），盈利

减少100kg*(成本的金额)；当补货量比销售量低100kg时，会出现供不应求的现象，此时与供需平衡的情况相比，盈利减少100*0.6*(成本的金额)。综上所述，考虑商超收益最大的情况就应该更偏向让补货量小于等于销售量，并且我们知道当供需平衡出现时是收益最大的情况，此时就是该线性规划的极值边界。

3. 第三个约束中我们定义了一个新的变量——品类的利润率，通过绘制各品类的利润率的Q-Q图来检验正态分布，不难看出其分位数近似均匀分布在y=x这条直线上，因此可以判断各品类的利润率满足正态分布。下图是品类2的正态Q-Q图以及品类2的正态曲线图。

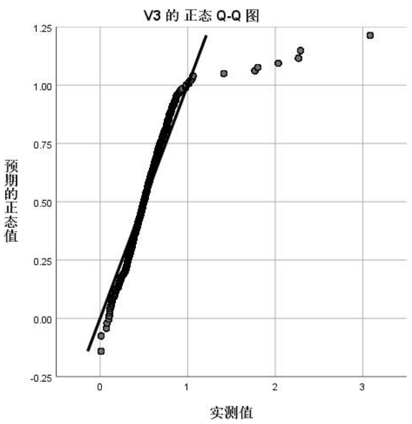


图 17: 品类2Q-Q图

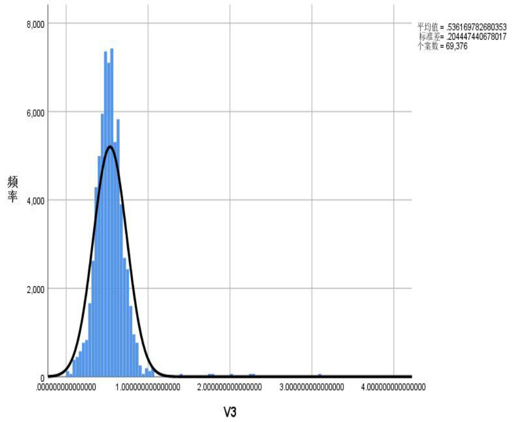


图 18: 品类2频率分布直方图

接着我们分析正态分布曲线，由于同品类蔬菜在定价时的利润率不能过高或过低，这样都不符合市场需求，且可能导致商超亏损过大，因此应该选择适当的界限作为约束条件，最终结合实际情况和市场因素最终选择了以0.8作为置信区间的边界值作为利润率的上下界，该区间涵盖了绝大多数利润率的情况，选择较为合理。

经过规划计算得到2023.7.1-2023.7.7一周时间内所有品类蔬菜的总收益为8491.30元。指定的补货策略和定价策略如下表所示(在销售量等于补货量的假设前提下，计算了品类的售价等于该品类的定价除以补货量，数据在表中所示)

表 4: 表格标题

品类	补货量(kg)	品类定价(元)	品类售价(元/kg)
7.1	126.6054	877.9029	6.9341
7.2	122.6070	860.7658	7.0205
7.3	118.8402	844.6214	7.1072
7.4	115.1454	828.7853	7.1977
7.5	11.5836	813.5196	7.2906
7.6	108.1245	798.6936	7.3867
7.7	104.7757	784.3408	7.4859

表 5:

品类	补货量(kg)	品类定价(元)	品类售价(元/kg)
7.1	25.3850	293.2474	11.5519
7.2	23.9490	280.9880	11.7327
7.3	22.4096	267.8457	11.9522
7.4	21.0016	255.8254	12.1812
7.5	19.6762	244.5111	12.4266
7.6	18.4356	233.9196	12.6884
7.7	17.2730	223.9943	12.9678

表 6:

品类	补货量(kg)	品类定价(元)	品类售价(元/kg)
7.1	18.4329	189.2729	10.2681
7.2	16.8757	174.7060	10.3524
7.3	15.6222	162.9805	10.4325
7.4	14.4124	151.6330	10.5230
7.5	13.3103	141.3536	10.6198
7.6	12.2885	131.7949	10.7250
7.7	11.3463	122.9808	10.8388

表 7:

品类	补货量(kg)	品类定价(元)	品类售价(元/kg)
7.1	19.8953	201.3101	10.1184
7.2	19.5256	197.7075	10.1255
7.3	18.2306	185.0881	10.1525
7.4	17.2333	175.3703	10.1761
7.5	16.2400	165.6909	10.2025
7.6	15.3159	156.6857	10.2302
7.7	14.4416	148.1655	10.2596

表 8:

品类	补货量(kg)	品类定价(元)	品类售价(元/kg)
7.1	79.7052	794.2104	9.9643
7.2	81.8208	814.8890	9.9594
7.3	81.8171	814.8526	9.9594
7.4	81.3703	810.4858	9.9605
7.5	81.4873	811.6285	9.9602
7.6	81.5515	812.2557	9.9600
7.7	81.5100	811.8508	9.9601

表 9:

品类	补货量(kg)	品类定价(元)	品类售价(元/kg)
7.1	43.5975	423.7794	9.7202
7.2	43.8207	425.5657	9.7115
7.3	42.9998	418.9962	9.7441
7.4	43.2424	420.9377	9.7344
7.5	43.3102	421.4801	9.7316
7.6	43.2419	420.9338	9.7344
7.7	42.2530	421.0255	9.7340

5.3 问题三的分析和求解

5.3.1 GM(2,1)模型

统计24-30号的可销售品种，统计数据包括单品成本，损耗率，利率，销售量，由于前七天的数据距离需要预测的日期较近，数据更能反映近期的真实销售情况，如果选择很长时间之前的数据进行统计可能会对预测结果产生较大误差，因此我们只统计了7月1日前七天的数据并利用这些数据进行分析，并且经过排序选出前33个利率最高的商品。

接着考虑需要用24-30号一周的数据预测7月1日的销售量，由于我们需要小样本来进行预测，且通过观察发现它不具有明显的指数单调规律，而是非单调的摆动发展序列，因此考虑采用GM(2,1)灰色预测模型对销售量进行预测以西峡花菇(1)为例进行分析。设原始序列为

$$\mathbf{x}^{(\mathbf{0})} = (x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n))$$

其1次累加生成序列$(1-AGO)x^{(1)}$ 和一次累减生成序列$(1-IAGO)\alpha^{(1)}x^{(0)}$ 分别为

$$\mathbf{x}^{(\mathbf{1})} = (x^{(1)}(1), x^{(1)}(2), \cdots, x^{(1)}(n))$$

和

$$\alpha^{(\mathbf{1})}\mathbf{x}^{(\mathbf{0})} = (\alpha^{(1)}x^{(0)}(2), \cdots, \alpha^{(1)}x^{(0)}(n))$$

其中

$$\alpha^{(1)}x^{(0)}(k) = x^{(0)}(k) - x^{(0)}(k-1), k = 2, 3, \cdots, n,$$

$x^{(1)}$的均值生成序列为

$$\mathbf{z^{(1)}} = (z^{(1)}(2), z^{(1)}(3), \cdots, z^{(1)}(n)),$$

则称

$$\alpha^{(1)}x^{(0)}(k) + a_1 x^{(0)}(k) + a_2 z^{(1)}(k) = b \tag{16}$$

为GM(2,1)模型称

$$\frac{d^2x^{(1)}(t)}{dt^2} + a_1\frac{dx^{(1)}(t)}{dt} + a_2 x^{(1)}(t) = b \tag{17}$$

为GM(2,1)模型的白化方程

且

$$\mathbf{B} = \begin{pmatrix} -x^{(0)}(2) & -z^{(1)}(2) & \dots & 1 \\ -x^{(0)}(3) & -z^{(1)}(3) & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ -x^{(0)}(n) & -z^{(1)}(n) & \dots & 1 \end{pmatrix}$$

$$\mathbf{Y} = \begin{bmatrix} \alpha^{(1)}x^{(0)}(2) \\ \alpha^{(1)}x^{(0)}(3) \\ \vdots \\ \alpha^{(1)}x^{(0)}(n) \end{bmatrix} = \begin{bmatrix} x^{(0)}(2) - x^{(0)}(1) \\ x^{(0)}(3) - x^{(0)}(2) \\ \vdots \\ x^{(0)}(n) - x^{(0)}(n-1) \end{bmatrix}$$

则GM(2,1)模型参数序列$\mathbf{u} = [a_1, a_2, b]$的最小二乘估计为

$$\hat{\mathbf{u}} = (B^TB)^{-1}B^TY \tag{18}$$

下面根据已知的$x^(0) = (6.18, 4.217, 3.582, 3.708, 3.742, 4.314, 6.572)$ 建立GM(2,1)模型

$x^{(0)}$的1-AGO序列和1-AIGO序列$\alpha^{(1)}x^{(0)}$分别为

$$\mathbf{u} = (6.18, 10.451, 14.033, 17.741, 21.483, 25.797, 32.369)$$

$$\alpha^{(\mathbf{1})}\mathbf{x^{(0)}} = (-1.909, -0.689, 0.126, 0.034, 0.572, 2.258)$$

$x^{(1)}$的均值生成序列为

$$\mathbf{z^{(1)}} = (8.3155, 12.242, 15.887, 19.612, 23.64, 29.083)$$

$$\mathbf{B} \begin{bmatrix} -x^{(0)}(2) & -z^{(1)}(2) & 1 \\ -x^{(0)}(3) & -z^{(1)}(3) & 1 \\ -x^{(0)}(4) & -z^{(1)}(4) & 1 \\ \vdots & \vdots & \vdots \\ -x^{(0)}(7) & -z^{(1)}(7) & 1 \end{bmatrix} = \begin{bmatrix} -4.271 & -8.3155 & 1 \\ -3.582 & -12.242 & 1 \\ -3.708 & -15.887 & 1 \\ -3.742 & -19.612 & 1 \\ -4.314 & -23.64 & 1 \\ -6.572 & -29.083 & 1 \end{bmatrix}$$

$$Y = [-1.909, -0.689, 0.126, 0.034, 0.572, 2.258]^T$$

$$\hat{\mathbf{u}} = \begin{bmatrix} \hat{a_1} \\ \hat{a_2} \\ \hat{b} \end{bmatrix} = (B^TB)^{-1}B^TY = \begin{bmatrix} -1.0276 \\ -0.1645 \\ -3.3591 \end{bmatrix}$$

故得GM(2,1)白化模型为

$$\frac{d^2x^{(1)}(t)}{dt^2} - 0.10276\frac{dx^{(1)}(t)}{dt} - 0.16415x^{(1)}(t) = -3.3591 \tag{19}$$

利用边界条件$x^{(1)}(1) = 6.18, x^{(1)}(7) = 32.369$，解得

$$x^{(1)}(t) = 0.867241e^{0.459772t} - 15.1516e^{0.-0.357014t} + 20.4644$$

于是GM(2,1)时间响应式为

$$x^{(1)}(k+1) = 0.867241e^{0.459772t} - 15.1516e^{0.-0.357014t} + 20.4644$$

所以

$$\hat{\mathbf{x}}^{(\mathbf{1})} = (6.18, 11.235, 15.22, 18.717, 22.287, 26.562, 32.369, 40.89)$$

做IAGO还原，有

$$x^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - x^{(1)}(k)$$

$$\hat{\mathbf{x}}^{(\mathbf{0})} = (6.18, 5.0553, 3.985, 3.4972, 3.5696, 4.2753, 5.8067, 8.5214)$$

计算结果见表

表 10: 误差检验表

序号	实际数据	预测数据	残差	相对误差
2	4.271	5.0553	-0.7843	18.3%
3	3.582	3.985	-0.40298	11.2%
4	3.708	3.4972	0.21079	5.7%
5	3.742	3.5696	0.17245	4.6%
6	4.314	4.2753	0.038711	0.89%
7	6.572	5.8067	0.76534	11.6%

计算预测结果为：

$$x^{(0)}(8) = 8.5214$$

预测结束，接着我们循环对剩余的32个单品重复进行如上操作步骤的预测模型建立并完成最终的预测结果。

预测完这些单品销售量后，我们需要建立收益最大得目标函数，进行线性规划。目标函数为：

$$max \quad \sum_{i=1}^{k}(1-l_i)y_i - \sum_{i=1}^{k}c_ix_i$$

$$\begin{cases} x_i(1-l_i) \geqslant 2.5 & i = 1, 2, \cdots, k \\ x_m(1-l_m) \leqslant p_m & \\ x_n(1-l_n) \leqslant 2.5 * (1+\alpha) & \\ per_{imin} \leqslant \dfrac{y_i - c_ix_i}{c_ix_i} \leqslant per_{imax} & i = 1, 2, \cdots, k \end{cases} \tag{20}$$

式中，$p_m > 2.5, p_n \leqslant 2.5, 1 \leqslant m, n \leqslant k$,$\alpha \in (0,1)$

下面我们对约束条件做出解释：

1. 每一个单品在这天的补货量*(1-损耗率)大于等于最小陈列量2.5kg，即实际的补货量要大于最小陈列量，以满足题目要求。
2. 首先我们将预测的销售量p_i分成两部分，分别是p_m和p_n，p_m代表某一个单品预测出的销售量大于最小陈列量2.5kg，p_n代表某一个单品预测出的销售量小于等于最小陈列量2.5kg。当预测出的销售量大于2.5kg时，为了满足收益尽可能大的标准，规定此时的补货量的上界就是该销售量，分析思路同第二问中类似；当预测出的销售量小于最小陈列量2.5kg时，考虑到蔬菜这类短期商品存在滞销情况，考虑到收益最大的目标，因此我们应该对这种情况下的补货量进行明确的上界约束，否则会导致收益受到较大影响。根据市场规律和需求情况，我们规定了最大补货量不超过最小陈列量的$(1+\alpha)$倍，α 为自定义的比例系数。最后结合相关文献，我们选定α近似等于0.2以符合题目要求。
3. 蔬菜单品的利润率界于2023.6.24-2023.6.30这七天中对应蔬菜单品的利润率最大值和最小值之间。

接着使k从27-33循环进行计算，发现当k=33且单品的补货量和定价如下表所示时收益最大，最大收益为708.5498元。定价和补货策略如下表所示(定义蔬菜单品的售价=蔬菜单品的定价/(补货量*(1-该单品的损耗率)))

表 11: 结果表

单品名称	单品补货量(kg)	单品定价(元)	单品售价(元/kg)
小米椒（份）	2.83	18.10	7.05
西兰花	16.88	232.86	15.20
西峡花菇(1)	9.55	229.88	26.91
螺丝椒	8.85	120.91	15.21
芜湖青椒(1)	5.33	33.17	6.59
紫茄子(2)	25.45	166.85	6.98
云南油麦菜(份)	29.81	135.30	5.01
长线茄	9.91	119.61	12.97
竹叶菜	14.38	56.29	4.53
娃娃菜	3.08	22.49	7.50
枝江青梗散花	3.31	48.56	16.19
姜蒜小米椒组合装(小份)	17.36	88.70	5.64
上海青	3.51	28.33	9.44
奶白菜	3.56	19.49	6.50
苋菜	4.22	16.98	4.94
螺丝椒(份)	3.31	21.86	7.29
小皱皮(份)	3.31	11.83	3.94
小青菜(1)	3.35	18.94	6.31
红薯尖	4.96	29.37	6.46
木耳菜	10.88	68.31	6.79
菱角	3.32	46.58	15.53
云南生菜(份)	37.81	226.48	6.61
高瓜(1)	4.24	70.28	23.43
海鲜菇(包)	3.00	9.02	3.01
云南生菜	6.83	63.46	10.97
七彩椒(2)	3.31	70.58	23.53
青茄子(1)	3.16	19.23	6.41
青红杭椒组合装(份)	8.69	55.21	7.01
红椒(2)	3.31	66.85	22.28
菠菜	3.68	51.66	17.22
外地茼蒿	4.06	54.69	18.23
野生粉藕	3.44	89.36	29.79
红莲藕带	3.60	35.05	11.68

5.4 问题四的分析与解答

为了更好地制定蔬菜商品的补货和定价决策，商超可以考虑采集以下数据：

1. 供应链数据：可以收集供应商交货时间、准时交货率和产品质量信息等数据。这可以帮助管理供应链，确保及时供应，减少库存和损耗。考虑与不同供应商合作，以获取更有竞争力的价格和条件。比较不同供应商的报价和质量，通过综合比对各个供应商的商品特点选择出最合理的补货策略与定价策略。此外还可以从供应链处获取采购量等数据并建立考虑供给的销量模型。这些数据可以反映供应链中的生产以及销售情况，帮助制定出更加精准的定价及补货策略。

2. 蔬菜商品的季节因素数据：由于各个品类的蔬菜生长、成熟、采集时间与季节都有着紧密的联系，某些蔬菜在某些季节可能更受欢迎，并且应季蔬菜和反季蔬菜的成本开销差异较大，因此对于每种品类的蔬菜要根据其季节性的变化规律而制定不同的补货、定价策略。这些数据有利于了解蔬菜的热销季节，例如下图是白菜苔的销售量随时间变化的散点图。

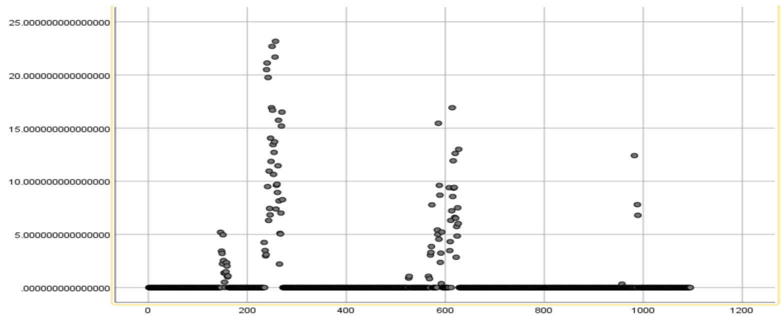


图 19: 白菜苔销量散点图

从上图可以明显看到白菜苔的热销时间成季节性，周期性变化，通过查阅资料得到白菜苔最佳食用季节是春季，恰好与时间的变化吻合，因此对白菜苔在制定补货和定价策略时应考虑到季节这种较为重要的因素，对于某些品类，考虑季节性因素，以确定何时需要加大或减少补货，这样可以使制定出的策略更加精准。

3. 突发情况数据收集：对于高需求或急需的情况，（例如除夕春节的需求量激增以及疫情、台风等突发事件的影响），为此应考虑紧急补货的问题，但要谨慎处理，避免成本飙升，这可以帮助调整补货策略和定价策略。

4. 商超库存数据：商超库存数据是指超市或零售企业的库存信息，通常包括库存周转率、库存成本、毛利率等关键指标。库存周转率反映了商品库存的流动速度。较高的周转率通常表示更高的效率和更少的库存积压。这可以帮助超市减少资金占用成本。这可以让策略制定时了解商超的商品库存情况，以便及时调整补货策略，这样可以有效防止商品的长期滞销或者长期供不应求的情况的发生，有效提高了库存空间的利用率，增加了利润。库存成本：库存成本是维护和管理库存所需的费用。了解库存成本有助于优化库存管理策略，确保库存的最佳利用。毛利率是超市销售额与成本之间的比率。通过监控毛利率，超市可以评估商品的定价策略和盈利能力。同时，合理的库存管理需要对订单信息进行合理掌控，确保货物供应充足，同时避免过度订购。此外，超市可以通过分析销售数据，了解不同商品的销售状况，以便调整库存策略。并且商超可以采取一系列措施，如订单联动库存中心、高库存预警等，以提升库存管理效率。综合以上数据和信息，商超可以制定更合理的库存策略，提高运营效率，降低成本，提供更好的客户服务，从而实现更好的经营绩效。

5. 顾客的反馈和评价数据：商超可以对顾客对各品类蔬菜的满意程度以及偏好程度的数据进行收集，还可以对顾客的购买意向进行调查，这些数据可以显示出消费者对于各品类蔬菜的偏好（例如冬季更多人购买根茎类蔬菜，夏季更多人购买叶菜类蔬菜等）以及价格敏感度等信息，商超

可以听取客户的投诉和建议，这样可以改进蔬菜的品质和服务，提高客户忠诚度，同时优化定价和补货策略，提高商品利润率。

6. 客流量数据：不同的日期和时间可能会有不同的客流量，客流量和蔬菜的销售量有一定关联，比如对于蔬菜而言，顾客会更偏向于上午去购买新鲜的青菜例如菠菜，而土豆和茄子等易存放的蔬菜可能不会呈现出在某个集中时间段进行购买的情况。这些数据有利于了解各品类蔬菜销售的高峰和低谷时期，有利于帮助商超调整陈列策略、促销活动时间以及补货、定价策略。

7. 市场需求数据：商超可以对蔬菜市场的需求侧进行调研，通过市场研究和消费者调查等手段，收集市场趋势和消费者洞察等数据，例如对当地人口、消费趋势和健康意识进行调研，这有助于了解市场需求，选择适合市场的蔬菜品种。

8. 生产、采购和运输成本数据：除了蔬菜进价的成本以外还存在着其他因素导致的成本支出，因此对各种成本数据进行调查有助于确保定价策略能够盈利，提高商品利润率，制定出更为合理、精准的定价策略。

9. 竞争对手数据：调研竞争对手的价格、促销策略和产品种类。这可以帮助商超了解市场动态，制定竞争性定价策略，以防恶性竞争现象的发生。还可以帮助商超了解市场竞争格局和竞争对手的表现，以制定差异化的补货和定价策略，提高市场份额和利润率。

10. 促销和折扣数据：商超促销折扣数据通常指的是超市或零售店在销售过程中提供的促销、折扣和优惠信息的相关数据。这些数据对于制定促销策略、吸引顾客、提高销售和了解市场反应非常重要。其中商超可以对如下有关促销的关键数据进行收集。促销类型：记录促销的具体类型，例如满减、买一送一、折扣、赠品等。促销时间：包括促销开始日期和结束日期，以及促销活动的持续时间。促销商品：列出参与促销的商品或商品类别。这些数据有助于了解哪些产品受到了促销的影响。原价和促销价：记录商品的原价和促销期间的价格，以及折扣幅度。促销渠道：指定促销活动在哪些销售渠道或门店中进行，以及是否在线上或线下促销。促销效果：分析促销活动对销售额、利润、库存周转率和客户购买行为的影响。促销成本：记录促销活动的成本，包括广告费用、降价费用和赠品成本等。促销策略：详细描述促销活动的策略，包括宣传方式、促销时机和目标受众。商超促销折扣数据有助于超市管理层更好地了解促销活动的效果，优化促销策略，提高销售额，增强市场竞争力，同时也能提供更多的购物价值给消费者。这些数据可以通过POS系统、销售记录、顾客反馈和市场研究等方式来收集和分析。这些数据有助于商超制定出更加合理、精准的补货策略和定价策略。

11. 顾客购买行为数据：商超客户购买行为数据是指记录顾客在超市或零售店购物时的行为和消费数据。这些数据对于了解顾客的购物偏好、行为模式以及制定更有效的市场策略至关重要。其中商超可以对如下有关顾客购买行为的重要数据进行收集。购买历史：记录顾客的购买历史，包括购买的产品、购买日期和购买金额。这有助于了解哪些产品更受欢迎。购买频率：记录每位顾客的购物频率，即他们多久会光顾一次商超。这有助于制定定期促销策略。购买渠道：指定顾客是在线上还是线下购买商品。这对于跟踪多渠道销售的效果很重要。购买金额：记录每次购物的总金额，以及购买的单个商品或商品类别的金额。购买时段：分析顾客购物的时间段，包括每天的高峰和低谷购物时段。购物篮分析：了解顾客在同一次购物中购买的商品组合。这有助于进行交叉销售和推荐产品。促销响应：跟踪顾客对促销活动的反应，包括折扣、特价商品和赠品等。购物地点：记录顾客购物的具体位置，以便确定哪些店铺或部门最受欢迎。支付方式：分析顾客使用的支付方式，例如现金、信用卡、移动支付等。购物偏好：了解顾客的购物偏好，包括品牌、商品类型和包装大小。退货和投诉：记录退货情况和顾客的投诉，以改进产品和服务。顾客忠诚度：评估顾客的忠诚度和重复购买率，以制定保持顾客满意度的策略。商超客户购买行为数据可以通过POS系统、会员卡数据、在线交易记录、顾客调查和市场研究等方式收集和分析。这些数据有助于超市管理层更好

地理解顾客需求，提供个性化的购物体验，增加销售额，提高客户满意度，同时也有助于制定更精确的市场定位和促销策略。此外，这些数据还有助于商超制定出更加合理、精准的补货策略和定价策略。

参考文献

[1] 黄辉. 基于消费者异质性的ATO系统生产及补货策略研究[D].重庆大学,2017.

[2] 司守奎等．数学建模算法与应用 [M]. 北京：国防工业出版社，2011.8 第一版；

[3] 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,2022.

[4] 孙雷.回归技术在超市商品销售趋势预测模型中的应用[J].江苏广播电视大学学报,2008,19(05):51-53.

[5] 杨种学.基于回归技术商品销售趋势预测模型的实现[J].保山师专学报,2009,28(05):64-67.

论文附录:

第一部分: 本部分包含用到的所有可运行的源程序代码,运行源程序时,请顺次执行下列程序,需要预置的文件包括:
“a1.xlsx(附件 1 的 A 列)” ,“a2.xlsx(附件 1 的 C 列)”, “a3.xlsx(附件 2 的 C 列)”, “a4.xlsx(附件 2 的 D 列)”, “a5.txt(附件 2 的 A 列)”, “a6.xlsx(附件 2 的 E 列)”, “c1.xlsx(附件 3 的 A 列)”, “c3.xlsx(附件 3 的 B 列)”, “c4.xlsx(附件 3 的 C 列)”, “d1.xlsx(附件 4 的 A 列)”, “d2.xlsx(附件 4 的 C 列)”, “change.txt(附件 2 的 F 列)”, “discount.txt(附件 2 的 G 列)”, “SingleName.txt(附件 1 的 B 列)”

Problem1_1.m:

```
clc,clear
format compact
format short
%SIC:单品编码数据
%L1:SIC 的长度
%CC:分类编码
%corrType:相应位置对应的分类编码
%编号对应关系:花叶类-1011010101-1;花菜类-1011010201-2;水生根茎类-
1011010402-3;
%茄类-1011010501-4;辣椒类-1011010504-5;食用菌-1011010801-6
%TypeSales:6 种单品在 2020 年 7 月 1 日至 2023 年 6 月 30 日的销量,下标对应分类
%a2STC:销量表中的单品编码数据
%L2:a2STC 的长度
%SISales:销量表中的单品销量
%SISalesnum:统计销量表中的单品销量里的无效数据个数
%date:导入日期数据
%Sdate:经过处理的 date 数据,显示了每个日期跟 2020 年 6 月 30 日的差距天数.
%L3:Sdate 中的最大数据,显示天数
%SingleSales:L1 种单品在 2020 年 7 月 1 日至 2023 年 6 月 30 日的销量,每行是一种单
品所有天数的销售额
%TypeSales:6 种品类在 2020 年 7 月 1 日至 2023 年 6 月 30 日的销量,每行是一种品类所
有天数的销售额
[numData1,textData1,rawData1] = xlsread('a1.xlsx');
SIC=textData1;
[numData2,textData2,rawData2] = xlsread('a2.xlsx');
CC=textData2;
L1=length(SIC);
corrType=zeros(L1,1);
search1='1011010101';
match1=ismember(CC,search1);
matchIndices1=find(match1);
corrType(matchIndices1)=1;
search2='1011010201';
```

```
match2=ismember(CC,search2);
matchIndices2=find(match2);
corrType(matchIndices2)=2;
search3='1011010402';
match3=ismember(CC,search3);
matchIndices3=find(match3);
corrType(matchIndices3)=3;
search4='1011010501';
match4=ismember(CC,search4);
matchIndices4=find(match4);
corrType(matchIndices4)=4;
search5='1011010504';
match5=ismember(CC,search5);
matchIndices5=find(match5);
corrType(matchIndices5)=5;
search6='1011010801';
match6=ismember(CC,search6);
matchIndices6=find(match6);
corrType(matchIndices6)=6;
[numData3,textData3,rawData3] = xlsread('a3.xlsx');
a2SIC=textData3;
L2=length(a2SIC);
SISales=readmatrix('a4.xlsx');
SISalesnum=sum(isnan(SISales));
date = importdata('a5.txt');
baseDate = datetime('2020-06-30', 'Format', 'yyyy-MM-dd');
Sdate= zeros(L2,1);
for i = 1:L2
    currentDate=datetime(date{i}, 'Format', 'yyyy-MM-dd');
    Sdate(i)=days(currentDate-baseDate);
end
L3=max(Sdate);
SingleSales=zeros(L1,L3);
TypeSales=zeros(6,L3);
change=importdata('change.txt');
search='销售';
for i=1:L2
    if strcmp(change(i),search)
        day=Sdate(i);         %先确定这条数据的天数
        temp1=a2SIC(i);       %确定这条数据的单品编码
        temp2=find(ismember(SIC,temp1));
        SingleSales(temp2,day)=SingleSales(temp2,day)+SISales(i);
```

```
TypeSales(corrType(temp2),day)=TypeSales(corrType(temp2),day)+SISale
s(i);
    end
end
for i=1:6
    figure
    plot([1:1095],TypeSales(i,:));
end
index=[];
for i=1:1095
    if TypeSales(:,i)==0
        index=[index;i];
    end
end
writematrix(TypeSales,'TypeSales0.xlsx');
writematrix(SingleSales,'SingleSales0.xlsx');
TypeSales(:,index)=[];
SingleSales(:,index)=[];
writematrix(TypeSales,'TypeSales.xlsx');
writematrix(SingleSales,'SingleSales.xlsx');
writematrix(Sdate,'Sdate.xlsx');
writematrix(corrType,'corrType.xlsx');
```

Problem1_2.m:

```
clc,clear
format compact
%formattedDate1:销量很高的点的日期
%formattedDate2:销量为 0 的点的日期
%c1:6 种品类的相关系数矩阵
%c2:251 种单品的相关系数矩阵
format short
L=1085;
SingleSales=readmatrix('SingleSales.xlsx');
TypeSales=readmatrix('TypeSales.xlsx');
for i=1:6
    figure
    plot([1:L],TypeSales(i,:));
end
c1=corrcoef(TypeSales')
c2=corrcoef(SingleSales');
figure
plot([1:L],TypeSales(1,:),'r-');
hold on
```

```
plot([1:L],TypeSales(2,:),'b--');
plot([1:L],TypeSales(3,:),'g:');
plot([1:L],TypeSales(4,:),'m-');
plot([1:L],TypeSales(5,:),'c--');
plot([1:L],TypeSales(6,:),'k:')
hold off
writematrix(c2,'SingleCoref.xlsx');
% 使用 heatmap 创建热图
figure
h = heatmap(c1);
colorbar;
h.FontSize = 10;
h.XDisplayLabels = cellstr(h.XDisplayLabels); % 将 X 轴刻度标签转换为字符串数组
h.YDisplayLabels = cellstr(h.YDisplayLabels); % 将 Y 轴刻度标签转换为字符串数组
```

Problem1_3.m:

```
clc,clear
format compact
format short
%c2:251 种单品的上三角相关系数矩阵
%row1,col1:弱相关性商品对坐标
%row2,col2:中相关性商品对坐标
%row3,col3:强相关性商品对坐标
%L:弱、中、强相关性商品对的数量
c2=readmatrix('SingleCoref.xlsx');
for i=1:251
    for j=1:i
        c2(i,j)=0;
    end
end
[row1,col1]=find(abs(c2)>0.1&abs(c2)<=0.3);
[row2,col2]=find(abs(c2)>0.3&abs(c2)<=0.5);
[row3,col3]=find(abs(c2)>0.5);
L=[length(row1),length(row2),length(row3)]
writematrix([row1,col1],'weak.xlsx');
writematrix([row2,col2],'middle.xlsx');
writematrix([row3,col3],'strong.xlsx');
```

Problem1_4.m:

```
clc,clear
format compact
SingleSales=readmatrix('SingleSales.xlsx');
```

```
%本文件用于对单品进行聚类分析
a=SingleSales;
b=zscore(a);                        %数据标准化
r=corrcoef(b);                      %计算相关系数矩阵
%d=tril(1-r);d=nonzeros(d)';        %另一种计算距离方法
z=linkage(b','average','correlation');        %按类平均法聚类
%correlation 表示类之间的距离用 1-相关系数
h=dendrogram(z);
set(h,'Color','k','LineWidth',1.3)
T=cluster(z,'maxclust',12);         %把变量分成 12 类
for i=1:12
   tm=find(T==i);
   fprintf('第%d 类的有%s\n',i,int2str(tm'));
end
T=[1081,619,878,756,529,461,1,75,223,873,580,706];
T=sort(T);
figure
a=a(:,T);
b=zscore(a);
z=linkage(b,'average');
h=dendrogram(z);
set(h,'Color','k','LineWidth',1.3)
for k=5
   fprintf('划分为%d 类的结果如下:\n',k)
   T=cluster(z,'maxclust',k);
   for i=1:k
      tm=find(T==i);
      fprintf('第%d 类的有:%s\n',i,int2str(tm'));  %显示分类结果
   end
   fprintf('********************************************\n')
end
A=[1 81 89 149 241];
for i=1:5
   figure
   plot(SingleSales(A(i),:));
end
```

Problem1_5.m:

```
clc,clear
format compact
format short
%本文件使用 AR 时间序列法来分析销售量和时间之间的关系,并预测未来一周的销售量
%QS:记录 6 个时间序列模型的 qs 值
%T:记录 6 个时间序列的统计量的值
```

```
%Pre:6*7矩阵,用来承接6种品类2023年7月1日-7月7日的预测销售量
TypeSales=readmatrix('TypeSales.xlsx');
for i=1:6
    a=TypeSales(i,:);
    Rt=tiedrank(a);                         %求原始时间序列的秩
    n=length(a);t=1:n;
    t_0=tinv(0.975,n-2);
    Qs(i)=1-6/(n*(n^2-1))*sum((t-Rt).^2);   %计算Qs的值
    T(i)=Qs(i)*sqrt(n-2)/sqrt(1-(Qs(i))^2);        %计算T统计量的值
    if T(i)>t_0
        b=diff(a);                          %求原始时间序列的一阶差分
        m=ar(b,2,'ls')                      %利用最小二乘法估计模型的参数
        bhat=predict(m,b');                  %求预测值，第二个参数必须为列向量
        bhat(end+1)=forecast(m,b',1);         %计算1个预测值，第二个参数必须
为列向量
        ahat=[a(1),a+bhat'];                  %求原始数据的预测值，并计算t=15
的预测值
        k=forecast(m,b',7);
        Pre(i,1)=ahat(end);
        for j=2:7
            Pre(i,j)=Pre(i,j-1)+k(j);
        end
    end
    if T(i)<=t_0
        m=ar(a,2,'ls')
        ahat=predict(m,a');
        Pre(i,1:7)=forecast(m,a',7);
    end
    figure,hold on
    plot(a);
    plot(ahat,'r');
    hold off
end
Pre
writematrix(Pre,'Pre.xlsx');
```

Problem2_1.m:

```
clc,clear
format compact
format short
%discount:2020年7月1日至2023年6月30日打折的情况
%Lcount:打折订单数量
%dSIC:附件4中的商品编码
%LossRate:对应的损耗率
```

```
%itemLossRate:第一列对应打折的商品序列(SIC 中),第二列对应其磨损率
discount=importdata('discount.txt');
search='是';
Ldiscount=sum(ismember(discount,search));
[numData1,textData1,rawData1] = xlsread('d1.xlsx');
dSIC=textData1;
LossRate=readmatrix('d2.xlsx');
[numData2,textData2,rawData2] = xlsread('a1.xlsx');
SIC=textData2;
[numData3,textData3,rawData3] = xlsread('a3.xlsx');
a2SIC=textData3;
flag=1;
for i=1:length(discount)
    if strcmp(discount(i),'是')
        temp1=find(ismember(SIC,a2SIC(i)));
        itemLossRate(flag,1)=temp1;
        temp2=find(ismember(dSIC,a2SIC(i)));
        itemLossRate(flag,2)=LossRate(temp2);
        flag=flag+1;
    end
end
[C,ia,ic]=unique(itemLossRate(:,1),'rows');
itemLossRate=itemLossRate(ia,:);
writematrix(itemLossRate,'itemlossrate.xlsx');
```

Problem2_2.m:

```
clc,clear
format compact
format short
%本文件是对附件 3 的数据进行处理
%CostTable:251*1095 的矩阵,用于记录每天各个单品的成本价格
%cSdate:附件 3 每条记录对应的天数
%cSIC:附件 3 每条记录对应的单品编号
[numData1,textData1,rawData1] = xlsread('a1.xlsx');
SIC=textData1;
cdate = importdata('c1.txt');
Lcdate=length(cdate);
baseDate = datetime('2020-06-30', 'Format', 'yyyy-MM-dd');
date= zeros(Lcdate,1);
for i = 1:Lcdate
    currentDate=datetime(cdate{i}, 'Format', 'yyyy-MM-dd');
    cSdate(i)=days(currentDate-baseDate);
end
[numData2,textData2,rawData2] = xlsread('c3.xlsx');
```

```
cSIC=textData2;
cost=readmatrix('c4.xlsx');
for i=1:Lcdate
    temp1=find(ismember(SIC,cSIC(i)));
    CostTable(temp1,cSdate(i))=cost(i);
end
writematrix(CostTable,'CostTable.xlsx');
```

Problem2_3.m:

```
clc,clear
format compact
format short
%该部分计算各品类的成本加成系数
%某品类在某天的成本=改品类下各单品销量乘上成本价格
%TypeCostTable:6*1095 的矩阵,记录每个品类每天的总进货成本
%TypePriceTable:6*1095 的矩阵,记录每个品类每天的总销售额
%TypeCPCTable:6*1095 矩阵,成本定价系数表
index=[226,227,580,855,857,883,884,885,886,935];
SISales=readmatrix('a4.xlsx');      %每一笔记录对应的销售量
Sdate=readmatrix('Sdate.xlsx');     %每一笔记录对应是第几天
SIPrice=readmatrix('a6.xlsx');      %每一笔记录对应的销售单价
[numData1,textData1,rawData1] = xlsread('a1.xlsx');
SIC=textData1;                      %SIC
corrType=readmatrix('corrType.xlsx');   %记录 SIC 对应的品类
TypeCostTable=zeros(6,1095);
TypePriceTable=zeros(6,1095);
CostTable=readmatrix('CostTable.xlsx'); %记录每个单品每天的进价
[numData2,textData2,rawData2] = xlsread('a3.xlsx');
a2SIC=textData2;                    %每一笔记录对应的单品编号
L2=length(a2SIC);
change=importdata('change.txt');
search='销售';
TypeCPCTable2=zeros(6,1095);
for i=1:L2
    if strcmp(change(i),search)
        temp1=find(ismember(SIC,a2SIC(i))); %找到这条记录对应的编号

TypeCostTable(corrType(temp1),Sdate(i))=TypeCostTable(corrType(temp1
),Sdate(i))+SISales(i)*CostTable(temp1,Sdate(i));

TypePriceTable(corrType(temp1),Sdate(i))=TypePriceTable(corrType(tem
p1),Sdate(i))+SISales(i)*SIPrice(i);
    end
end
```

```
TypeSales=readmatrix('TypeSales0.xlsx');
for i=1:L2
    if strcmp(change(i),search)
        temp1=find(ismember(SIC,a2SIC(i))); %找到这条记录对应的编号
        p=(SIPrice(i)-
CostTable(temp1,Sdate(i)))/CostTable(temp1,Sdate(i));

TypeCPCTable2(corrType(temp1),Sdate(i))=TypeCPCTable2(corrType(temp1
),Sdate(i))+p*SISales(i)/TypeSales(corrType(temp1),Sdate(i));
    end
end
writematrix(TypeCostTable,'TypeCostTable0.xlsx');
writematrix(TypePriceTable,'TypePriceTable0.xlsx');
TypeCostTable(:,index)=[];
TypePriceTable(:,index)=[];
writematrix(TypeCostTable,'TypeCostTable.xlsx');
writematrix(TypePriceTable,'TypePriceTable.xlsx');
TypeCPCTable=(TypePriceTable-TypeCostTable)./TypeCostTable;
writematrix(TypeCPCTable,'TypeCPCTable.xlsx');
TypeCPCTable2(:,index)=[];
writematrix(TypeCPCTable2,'TypeCPCTable2.xlsx')
```

Problem2_4.m:

```
clc,clear
format compact
format short
%本文件用于建立售价和销售量的回归模型以及建立成本和销售量的关系
%beta:2*6 的矩阵,用于存储 6 个回归模型的回归系数
%Fa:模型检验系数
%F:1*6 的矩阵用于存储 F 统计量,用于检验模型
TypePriceTable=readmatrix('TypePriceTable0.xlsx');
TypeCostTable=readmatrix('TypeCostTable0.xlsx');
TypeSales=readmatrix('TypeSales0.xlsx');
m=1;n=1085;
Fa=finv(0.99,m,n-m-1);
for i=1:6
    c=regstats(TypePriceTable(i,:),TypeSales(i,:));
    beta(i,:)=c.beta;
    F(i)=c.fstat.f;
end
writematrix(beta,'beta.xlsx');
for i=1:6
    c=regstats(TypeCostTable(i,:),TypeSales(i,:));
    bbeta(i,:)=c.beta;
```

```
    FF(i)=c.fstat.f;
end
Fa
F,FF
beta,bbeta
writematrix(bbeta,'bbeta.xlsx');
```

Problem2_5.m:

```
clc,clear
format compact
format short
%本文件用于指定定价策略
%变量 x:42 个,即接下来一周每天 6 种品类蔬菜的进货量
%变量 y:42 个,定价方案,接下来一周每天 6 种品类蔬菜的定价
%约束条件 1:成本定价系数比如介于以往定价系数的最低和最大值之间
%约束条件 2:进货量大于等于销售量(此处的销售量是预测的结果)
format long
Pre=readmatrix('Pre.xlsx');
minAndmax=[0.4674,0.8768;0.2742,0.7982;0.2895,0.6433;0.3275,0.8622;0
.3620,0.9213;0.4150,0.7540];
bbeta=readmatrix('bbeta.xlsx');
prob=optimproblem('ObjectiveSense','max');
x=optimvar('x',6,7,'LowerBound',0);
y=optimvar('y',6,7,'LowerBound',0);
prob.Objective=sum(sum(y))-7*sum(bbeta(:,1))-bbeta(:,2)'*x(:,1)-
bbeta(:,2)'*x(:,2)-...
    bbeta(:,2)'*x(:,3)-bbeta(:,2)'*x(:,4)-bbeta(:,2)'*x(:,5)-
bbeta(:,2)'*x(:,6)-...
    bbeta(:,2)'*x(:,7);
prob.Constraints.con1=x<=Pre;
con2=[];
for i=1:6
    for j=1:7

con2=[con2;y(i,j)<=(bbeta(i,1)+bbeta(i,2)*x(i,j))*(1+minAndmax(i,2))
];
    end
end
prob.Constraints.con2=con2;
con3=[];
for i=1:6
    for j=1:7
```

```
con3=[con3;y(i,j)>=(bbeta(i,1)+bbeta(i,2)*x(i,j))*(1+minAndmax(i,1))
];
    end
end
prob.Constraints.con3=con3;
[sol,fval,flag,out]=solve(prob)
xx=sol.x,yy=sol.y
writematrix(xx,'xx.xlsx');
writematrix(yy,'yy.xlsx');
```

Problem2_6.m:

```
clc,clear
format compact
format short
%本文件用于分析利率
index=[226,227,580,855,857,883,884,885,886,935];
TypeSales=readmatrix('TypeSales.xlsx');
SingleSales=readmatrix('SingleSales.xlsx');
TypeCostTable=readmatrix('TypeCostTable.xlsx');
TypePriceTable=readmatrix('TypePriceTable.xlsx');
m1=min(sum(TypePriceTable-TypeCostTable));
m2=max(sum(TypePriceTable-TypeCostTable));
plot(sum(TypePriceTable-TypeCostTable))
```

Problem3_1.m:

```
clc,clear
format compact
format short
%本文件用于提取 2023 年 6 月 24 日-6 月 30 日的可售品种信息和损耗率的相关信息
%时间 t 范围:1089-1095
%index:用于记录售出商品的编号
%itemnum:售出的商品数量
%SICostTable:itemnum*8 的矩阵,用于记录 7 日单品的成本,第一列记录商品编号
%SIPriceTable:itemnum*8 的矩阵,用与记录 7 日单品的收入,第一列记录商品编号
%SISalesTable:itemnum*8 的矩阵,用于记录 7 日单品的销售量,第一列记录商品编号
%p4SingleCost:251*7 的矩阵,最后一周所有单品的批发价格
%p4SingleSales:251*7 的矩阵,最后一周所有单品的销售量
%p4SinglePrice:251*7 的矩阵,最后一周所有单品的售价
[numData1,textData1,rawData1] = xlsread('a1.xlsx');
SIC=textData1;
temp=zeros(251,1);      %如果商品有售出就置为 1
index=[];
N=55629;                %附件 3 中从 55629 条数据开始
```

```
[numData2,textData2,rawData2] = xlsread('c3.xlsx');
cSIC=textData2;
SIcost=readmatrix('c4.xlsx');
cdate = importdata('c1.txt');
Lcdate=length(cdate);
baseDate = datetime('2020-06-30', 'Format', 'yyyy-MM-dd');
date= zeros(Lcdate,1);
for i = 1:Lcdate
    currentDate=datetime(cdate{i}, 'Format', 'yyyy-MM-dd');
    cSdate(i)=days(currentDate-baseDate);
end
p4SingleCost=zeros(251,7);
for i=N:length(cSIC)
    temp1=find(ismember(SIC,cSIC(i)));
    index=[index;temp1];
    temp(temp1)=1;
    p4SingleCost(temp1,cSdate(i)-1088)=SIcost(i);
end
index=unique(sort(index));
itemnum=length(index);
SICostTable=zeros(itemnum,8);
SICostTable(:,1)=index;
SIPriceTable=zeros(itemnum,8);
SIPriceTable(:,1)=index;
CostTable=readmatrix('CostTable.xlsx');
SingleSales=readmatrix('SingleSales0.xlsx');
p4SingleSales=SingleSales(:,[1089:1095]);
writematrix(p4SingleSales,'p4SingleSales.xlsx');
M=875002;               %附件 2 中从 875002 条数据开始
[numData3,textData3,rawData3] = xlsread('a3.xlsx');
a2SIC=textData3;
SISales=readmatrix('a4.xlsx');
Sdate=readmatrix('Sdate.xlsx');
change=importdata('change.txt');
search='销售';
p4SinglePrice=zeros(251,7);
SIprice=readmatrix('a6.xlsx');
for i=M:length(a2SIC)
    if strcmp(change(i),search)
        temp1=find(ismember(SIC,a2SIC(i)));
        p4SinglePrice(temp1,Sdate(i)-1088)=SIprice(i);
    end
end
writematrix(p4SinglePrice,'p4SinglePrice.xlsx');
```

```
writematrix(p4SingleCost,'p4SingleCost.xlsx');
SISalesTable=zeros(itemnum,8);
SISalesTable(:,1)=index;
for i=1:itemnum
    for j=2:8
        temp1=SISalesTable(i,1);
        SISalesTable(i,j)=p4SingleSales(temp1,j-1);
        SICostTable(i,j)=p4SingleSales(temp1,j-
1)*p4SingleCost(temp1,j-1);
        SIPriceTable(i,j)=p4SingleSales(temp1,j-
1)*p4SinglePrice(temp1,j-1);
    end
end
writematrix(SISalesTable,'SISalesTable.xlsx');
writematrix(SICostTable,'SICostTable.xlsx');
writematrix(SIPriceTable,'SIPriceTable.xlsx');
```

Problem3_2.m:

```
clc,clear
format compact
format short g
SISalesTable=readmatrix('SISalesTable.xlsx');
SIPriceTable=readmatrix('SIPriceTable.xlsx');
SICostTable=readmatrix('SICostTable.xlsx');
%本文件的作用在于选出利润贡献率最高的前 33 种,并预测它们在 7 月 1 号的销量
%排名
SIProfitTable=zeros(61,2);
SIProfitTable(:,1)=SISalesTable(:,1);
SIProfitTable(:,2)=sum(SIPriceTable(:,2:end)-
SICostTable(:,2:end),2);
[temp,I]=sort(SIProfitTable(:,2),'descend');
SIProfitTable33=SIProfitTable(I,:);
%delta:承接相对误差
%uu:33*3 的矩阵,用于承接系数
%yuece:33*7 的矩阵,用于承接预测值
%pre:承接预测值
%Eps:33*7 的矩阵,用于承接残差
%Del:33*7 的矩阵,用于承接误差比
for i=1:61
    temp1=SIProfitTable33(i,1);
    SISalesTable33(i,1)=SIProfitTable33(i,1);
    temp2=find(ismember(SISalesTable(:,1),temp1));
    x0=SISalesTable(temp2,2:end);
    SISalesTable33(i,2:8)=x0;
```

```
%GM(2,1)
    n=length(x0);
    x1=cumsum(x0);
    a_x0=diff(x0)';        %计算一次累减序列
    z=0.5*(x1(1:end-1)+x1(2:end))';
    B=[-x0(2:end)',-z,ones(n-1,1)];
    u=B\a_x0;
    uu(i,:)=u';
    if i==3
        x1,a_x0,z,B,u
    end
    syms x(t)

x=dsolve(diff(x,2)+u(1)*diff(x)+u(2)*x==u(3),x(0)==x1(1),x(6)==x1(7)
);
    %利用边界条件求符号解(dsolve 求符号解的时候既可以是初值条件也可以是边界条件)
    xt=vpa(x,6);
    yuce=subs(x,t,0:n-1);
    yuce=double(yuce);
    x0_hat=[yuce(1),diff(yuce)];
    pre(i,:)=x0_hat;
    y=subs(x,t,0:n);
    y=double(y);
    xx=[y(1),diff(y)];
    p(i,:)=xx;
    epsilon=x0-x0_hat;
    Eps(i,:)=epsilon;
    delta=abs(epsilon./x0);  %求相对误差
    Del(i,:)=delta;
    if i==3
        xt,y,xx,epsilon,delta
    end
end
p(:,end+1)=SISalesTable33(:,1);
writematrix(SISalesTable33,'SISalesTable33.xlsx');
writematrix(p,'ppt.xlsx');
```

Problem3_3.m:

```
clc,clear
format compact
format short
%本文件用于制定单品策略
pre=readmatrix('ppt.xlsx');
index33=pre(:,end);
```

```
p4SingleCost=readmatrix('p4SingleCost.xlsx');
p4SinglePrice=readmatrix('p4SinglePrice.xlsx');
p4SICPC=(p4SinglePrice-p4SingleCost)./p4SingleCost;
p4SICPC=p4SICPC(index33,:);
mm=zeros(61,2);
mm(:,1)=min(p4SICPC,[],2);
mm(:,2)=max(p4SICPC,[],2);
weightlimit=readmatrix('p4SingleSales.xlsx');
weightlimit=weightlimit(index33,:);
weightlimit=max(weightlimit,[],2);
cc=zeros(61,1);
for i=1:61
    flag=0;
    temp=index33(i);
    for j=1:7
        if p4SingleCost(temp,j)~=0
            flag=flag+1;
            cc(i)=cc(i)+p4SingleCost(temp,j);
        end
    end
    cc(i)=cc(i)/flag;
end
pre=pre(:,end-1);
%x 是单品进货量
%y 是单品定价(1-33 种商品的定价)
[numData1,textData1,rawData1] = xlsread('a1.xlsx');
SIC=textData1;
[numData2,textData2,rawData2] = xlsread('d1.xlsx');
dSIC=textData2;
LossRate=readmatrix('d2.xlsx');
LRate=zeros(61,1);
for i=1:61
    temp=index33(i);
    temp2=find(ismember(dSIC,SIC(temp)));
    LRate(i)=LossRate(temp2);
end
LRate=LRate/100;
for j=27:33
prob=optimproblem('ObjectiveSense','max');
x=optimvar('x',j,'LowerBound',0);
y=optimvar('y',j,'LowerBound',0);
prob.Objective=(1-LRate(1:j)')*y-cc(1:j)'*x;
con1=[];
for i=1:j
```

```
    con1=[con1;x(i)*(1-LRate(i))>=2.5];
end
prob.Constraints.con1=con1;
con2=[];
for i=1:j
    if pre(i)>2.5
        con2=[con2;x(i)*(1-LRate(i))<=pre(i)];
    else
        con2=[con2;x(i)*(1-LRate(i))<=3];
    end
end
con3=[];
for i=1:j
    con3=[con3;y(i)<=(1+mm(i,2))*cc(i)*x(i)];
end
con4=[];
for i=1:j
    con4=[con4;y(i)>=(1+mm(i,1))*cc(i)*x(i)];
end
prob.Constraints.con2=con2;
prob.Constraints.con3=con3;
prob.Constraints.con4=con4;
[sol,fval,flag,out]=solve(prob)
xx=sol.x;yy=sol.y;
end
xx,yy
writematrix(xx,'xxx.xlsx');
writematrix(yy,'yyy.xlsx');
Name=importdata('SIngleName.txt');
Name=Name(index33);
Name=Name(1:33);
writecell(Name,'Name.txt');
```

全部代码执行完毕后:文件夹中会多出下列文件:
“TypeSales0.xlsx”, “TypeSales.xlsx”, “SingleSales0.xlsx”, “SingleSales.xlsx” “Sdate.xlsx”, “corrType.xlsx”, “SingleCoref.xlsx”, “weak.xlsx”, “middle.xlsx”, “strong.xlsx”, “itemlossrate.xlsx”,”CostTable.xlsx”,”TypeCostTable0.xlsx”, “TypePriceTable0.xlsx”, “TypeCostTable.xlsx”, “TypePriceTable.xlsx”, “TypeCPCTable.xlsx”, “TypeCPCTable2.xlsx”, “beta.xlsx”, “bbeta.xlsx”, “Pre.xlsx” “xx.xlsx”, “yy,xlsx”, “p4SingleSales.xlsx”, “p4SinglePrice.xlsx”, “p4SingleCost.xlsx”, “SISalesTable.xlsx”, “SICostTable.xlsx”, “SIPriceTable.xlsx” “SISalesTable33.xlsx”, “ppt.xlsx”, “xxx.xlsx”, “yyy.xlsx”, “Name.txt”

除此之外,我们还利用 SPSS 获得了一些图片,具体的操作如下:

从 SPSS 导入 Excel 表格后点击菜单中的分析选项，找到里面的描述分析，选择 Q-Q 图选项来作出某一列数据的 Q-Q 图以拟合正态分布曲线，选择频率选项后对一列数据进行描述性分析得到均值、最大值、最小值等信息;除此之外利用图形选项中的旧对话框子选项，随后选择直方图和散点图进行绘制，接着要选择某一列数据或同时选中某两列数据进行绘制，在绘制直方图时，勾选”显示正态分布曲线“选项即可拟合出正态分布曲线，并且显示出正态分布曲线的均值和标准差等参数。