蔬菜类商品的自动定价与补货决策分析

摘要

题目强调了市场需求分析的重要性，对于补货和定价决策至关重要。商超需要综合考虑销售历史、需求趋势、供应情况以及蔬菜的保质期等因素，以确保蔬菜类商品的合理补货和定价策略，以最大程度地满足顾客需求，同时减少损失和浪费。这种复杂的决策过程需要商超拥有可靠的市场数据和分析能力，以提高经营效益。

针对问题 1，通过分析不同品类和不同单品的分布规律及相互关系，首先，利用 Python 软件对数据进行预处理并绘制不同品类的时间序列图和不同品类及不同单品的销售总量图等图形。通过这些图形发现销售量的周期性强，且部分蔬菜品类和单品销售量较大。最后，通过绘制不同品类的矩阵热力图和散点图矩阵，并利用 Spearman 相关性系数和层次聚类进行分析，得出花菜类和花叶类，食用菌类和水生根茎类具有强相关性；其他品类之间具有中等程度相关性或弱相关性；茄类与其他各类具有极弱相关性。单品分为三大类，第一类为大白菜，第二类为西兰花，净藕（1），芜湖青椒（1），第三类为其他剩下的单品品种。

针对问题 2，首先需要分析各品类销售总量与成本加成定价的函数关系，因为商超定价采用的是“成本加成定价”的方式，所以可以将问题转化为通过最小二乘法[6]来得出各品类销售总量与定价之间的函数关系，我们利用 Matlab 得出了各品类销售总量与定价之间的函数关系。接着我们用 Spss 软件建立了各品类进货量 ARIMA 模型以及各品类日进价的 ARIMA 模型，通过建立的 ARIMA 模型来预测接下来七天各品类的日进货量和日进价，最终确定使得商超利润最大的进货量和定价策略。

针对问题 3,利用粒子群算法，通过规划模型，每个粒子确定每个粒子个体的最优解并从这些个体最优解找到一个全局最优解。

针对问题 4，从消费者，销售环境，自身经营三个角度着手分析，包括销售高峰期，天气情况，消费者需求反馈等，这些数据可以更好地预测销售需求，制定更完善的补货计划，调整调价方案，提升市场竞争力等。

关键词：Python 软件 Matlab 软件 SPSS 软件 Spearman 相关性系数 层次聚类 ARIMA 模型

一、问题重述

在生鲜超市中，蔬菜类商品通常有短暂的保鲜期限，随着时间的推移，它们的品质会逐渐下降。很多蔬菜品种如果当天没有售出，第二天就无法再销售。因此，超市通常每天根据历史销售数据和需求情况进行蔬菜的补货。

由于蔬菜种类繁多，产地各异，而且进货通常发生在凌晨 3:00-4:00 之间，商家必须在不知道具体产品和进货价格的情况下，做出当日各种蔬菜的补货决策。一般来说，蔬菜的定价采用"成本加成定价"方法，商超通常会对受运输损失和品质下降影响的商品进行折扣销售。

准确的市场需求分析对于补货和定价决策至关重要。从需求方面来看，蔬菜类商品的销售量通常与时间有一定的关联。从供应方面来看，4 月至 10 月之间蔬菜供应丰富，但由于销售空间有限，合理的销售组合变得非常关键。

附件 1 给出了某商超经销的 6 个蔬菜品类的商品信息；附件 2 和附件 3 分别给出了该商超 2020 年 7 月 1 日至 2023 年 6 月 30 日各商品的销售流水明细与批发价格的相关数据；附件 4 给出了各商品近期的损耗率数据。

问题1 蔬菜类商品不同品类或不同单品之间可能存在一定的关联关系，分析蔬菜各品类及单品销售量的分布规律及相互关系。

问题2 考虑商超以品类为单位做补货计划，分析各蔬菜品类的销售总量与成本加成定价的关系，并给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3 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 4 为了更好地制定蔬菜商品的补货和定价决策，商超还需要采集哪些相关数据，这些数据对解决上述问题有何帮助，给出意见和理由。

二、问题分析

2.1 问题 1 的分析

问题 1 需要分析不同品类及不同单品的分布规律及相互关系，先使用 Python 进行数据处理[3][4]，再利用 Pearson 和 Spearman[9]相关系数以及层次聚类进行相关性分析。首先对附件 1 和附件 2 的数据进行预处理，合并成各品类、各单品的销售量数据，并绘制不同品类及不同单品的销售总量图和绘制矩阵热力图以及谱系图，从而得出不同品类及不同单品的分布规律和相互关系。

2.2 问题 2 的分析

需要分析多变量之间的函数关系以及根据提供的数据来预测未来发生的事

（1） 解决多变量之间的函数关系这类问题一般可以采用最小二乘法来得出变量之间的函数关系式，我们先对数据进行处理，将数据以品类为单位分为不同表格，表格内容包括各品类在一天中的销售总量和对应的平均定价，我们用 matlab[5]画出两者的关系图时发现他们之间存在一定的函数关系，接着我们通过 matlab 中的 polyfit 函数来进行多项式拟合，最终得出了销售总量与定价之间的函数关系。

（2） 对于解决预测未来发生的事这类问题一般可以采用时间序列预测模型来进行讨论，在用 spss 软件画出各品类销售总量的一阶差分时序图时我们发现变化较为平稳，所以我们采用时间序列预测传统模型中的 ARIMA 模型来进行预测，通过计算 p，d，q 值，最终得出的预测图像与实际图像吻合程度较高，说明预测模型成立。

（3） 通过研究分析问题 2，在不考虑极端因素的影响时，我们可以预测出长期以来最适合商超发展的进货与定价策略，有利于商超的长期经营，若只考虑短期的高回报率，会导致顾客的购买欲望下降，客户粘度降低，不利于商超的长期发展。

2.3 问题 3 的分析

问题 3 需要分析 6 月 24 日-30 日的可售品种，保证可以满足市场对不同品类的需求，且单品订购量满足最小陈列量 2.5 千克，预测日补货总量并制定合理的定价策略，使得商超收益最大化。

2.4 问题 4 的分析

商超需要寻找其他的数据，从消费者，销售环境，自身经营等角度分析考虑，进行更好地预测需求，调整补货策略和定价计划。

三、模型假设

1. 蔬菜的销售量数据在短期内不产生大波动，不受突发事件的影响。

2. 消费者的消费需求整体较为稳定，具有周期性。

四、定义与符号说明

表 4-1 符号说明

符号	说明
P_i	销售单价
saleAverage	花菜类平均销售量
avg_{values}	花叶类平均销售量
avg_values2	辣椒类平均销售量
avg_values4	茄类平均销售量
avg_values6	食用菌类平均销售量
avg_values8	水生根茎类平均销售量
R^2	模型拟合度统计
ρ	相关系数
x_i	独立变量
y_i	依赖变量

五、模型的建立与求解

5.1 数据的处理

1. 通过附件 1 和附件 2 的单品编码将附件 1 中的单品名称和分类名称合并到附件 2 中。

2. 整理出不同品类和不同单品总的销售量数据。

3. 把不同品类及不同单品的销售量整理成以天为单位的数据集。

4. 把不同品类销售量和不同单品销售量整理成以月为单位。

5.2 问题 1 模型的建立与求解

分析不同品类及不同单品之间销售量存在的关联关系，包括不同品类销售量的分布

规律及相互关系和不同单品销售量的分布规律及相互关系。通过分析附件数据，进行数据处理、可视化处理和相关性分析，并绘制不同品类和不同品单品销售总量直方图、不同品类的矩阵热力图和散点图矩阵，最后利用 Spearman 相关系数和层次聚类分析。

（一）分布规律

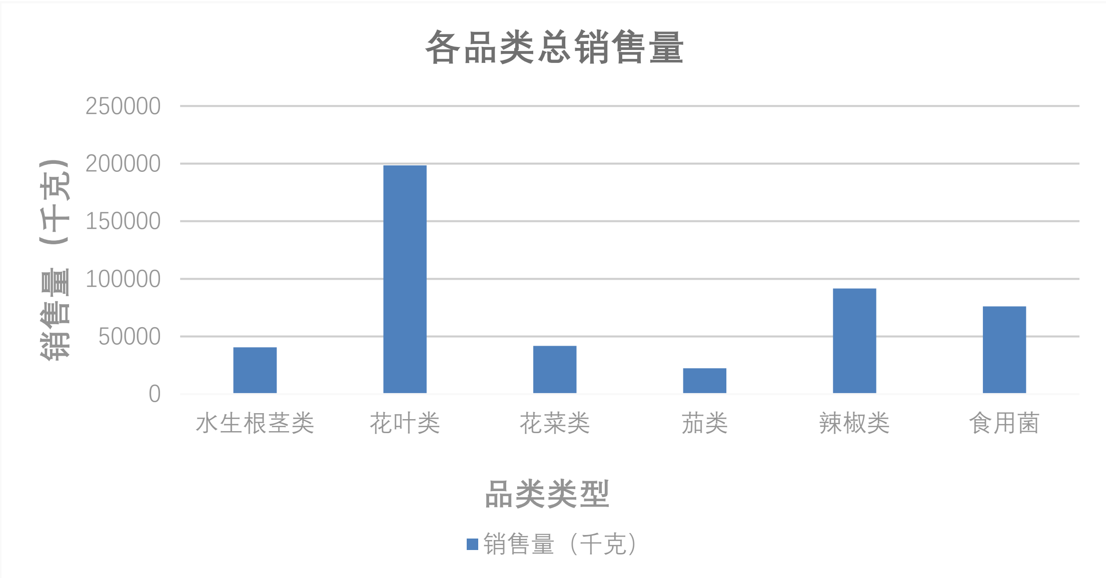


图 5 - 1 各品类总销售量

图 5 - 1 反映 2020 年 7 月 1 日至 2023 年 6 月 30 日三年间，花叶类蔬菜销售量最大，约等于其他类蔬菜销售量总和。其次是辣椒类和食用菌类销售量较大。

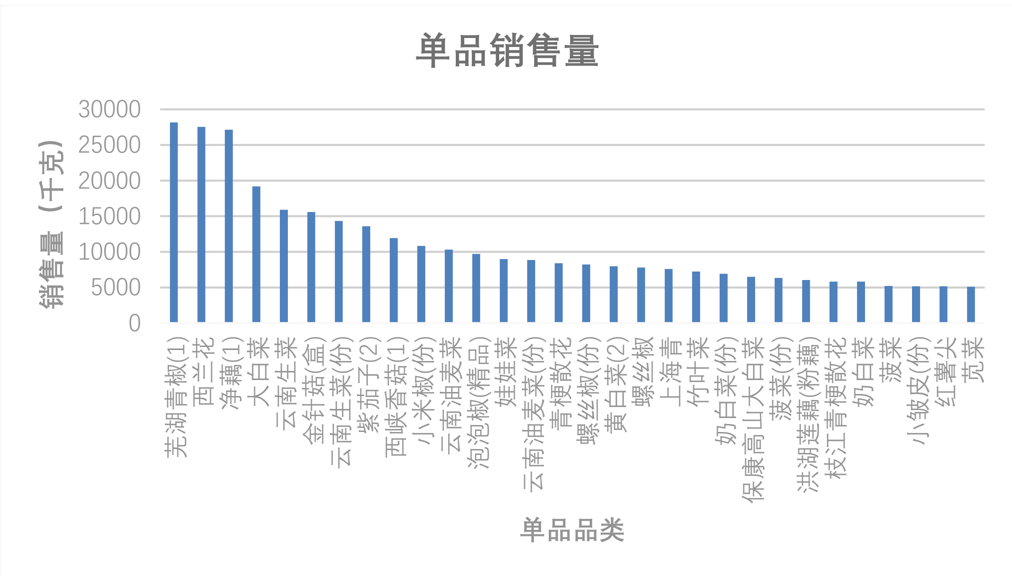


图 5 - 2 单品销售量

图 5 - 2 为 2020 年 7 月 1 日至 2023 年 6 月 30 日三年间，单品销售量前 30 的总销售量。由图表得出芜湖青椒（1），西兰花，净藕为三个销售量较大的单品。

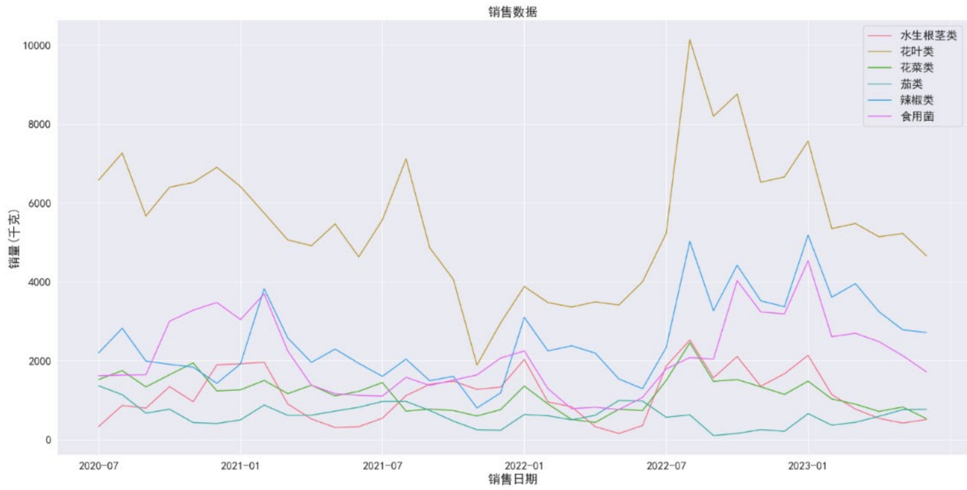


图 5 - 3 不同品类总销售量

图 5 - 3 为 2020 年 7 月 1 日至 2023 年 6 月 30 日期间 36 个月，每个月不同品类的销售总量图。

由于数据量很大，我们选取月为单位，把每月的销售总量求出来，绘制出不同品类的销售总量随时间变化的关系，方便观察并对销售量的周期性进行分析。

由图表得出各品类的销售量都呈现出较明显的季节性，其中水生根类和食用菌类尤为明显，在 4 月到 10 月销售量明显增大，与题目中蔬菜的供应品种在 4 月至 10 月较为丰富相符。

（二）相关性分析

首先以月为单位，整理了每个月与各品类销售量的数据，并导入 SPSS 中进行相关性检验。

然后用 Spearman 相关性系数进行不同品类间的相关性分析。

相关性系数：

$$\rho = \frac{\sum_i (x_i - \bar{y})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \tag{5 - 1}$$

表 5 - 1 相关强度等级表

数值范围	**相关程度**
0.8-1.0	极强相关
0.6-0.8	强相关
0.4-0.6	中等程度相关
0.2-0.4	弱相关
0.0-0.2	极弱相关或无相关

利用 Spearman 相关性系数分析结果如下：

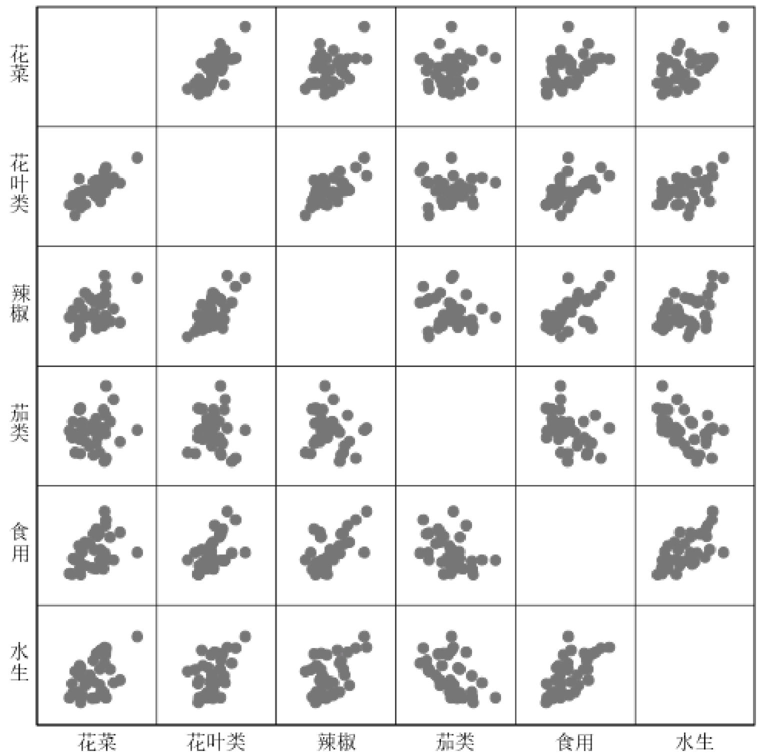


图 5 - 4 不同品类散点图矩阵

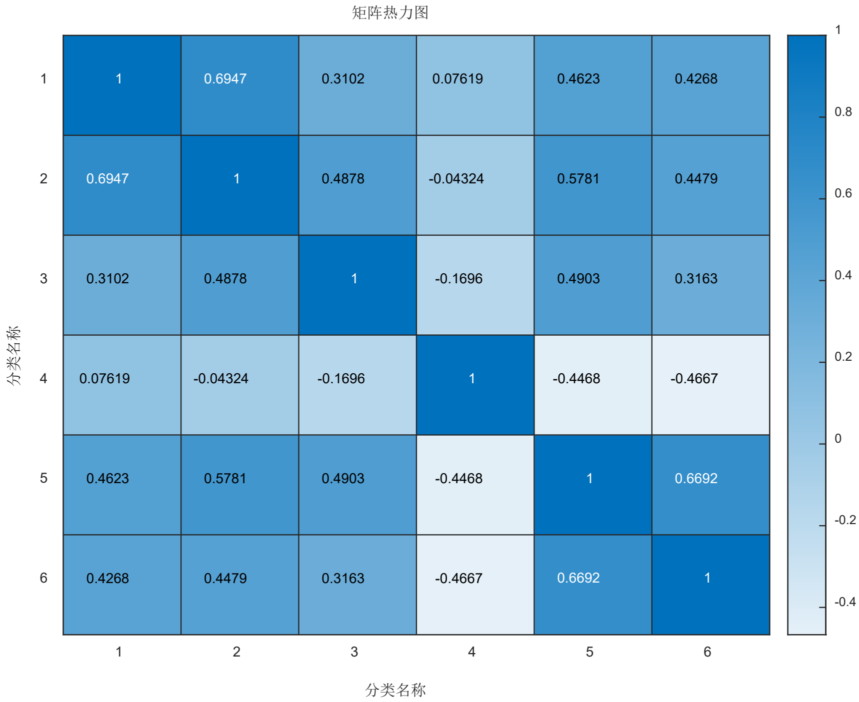


图 5 - 5 不同品类矩阵热力图

1-花菜类，2-花叶类，3-辣椒类，4-茄类，5-食用菌类，6-水生根茎类

利用 Spearman 相关性系数进行相关性分析，绘制出不同品类销售量之间散点图矩阵和接矩阵热力图，分析得出花菜类和花叶类，食用菌类和水生根茎类具有强相关性；花叶类和辣椒类，花菜类和食用菌类，花叶类和食用菌类，辣椒类和食用菌类，花菜类和水生根茎类，花叶类和水生根茎类具有中等程度相关性；花菜类和辣椒类，辣椒类和水生根茎类具有弱相关性；而茄类与其他各类具有极弱相关性。

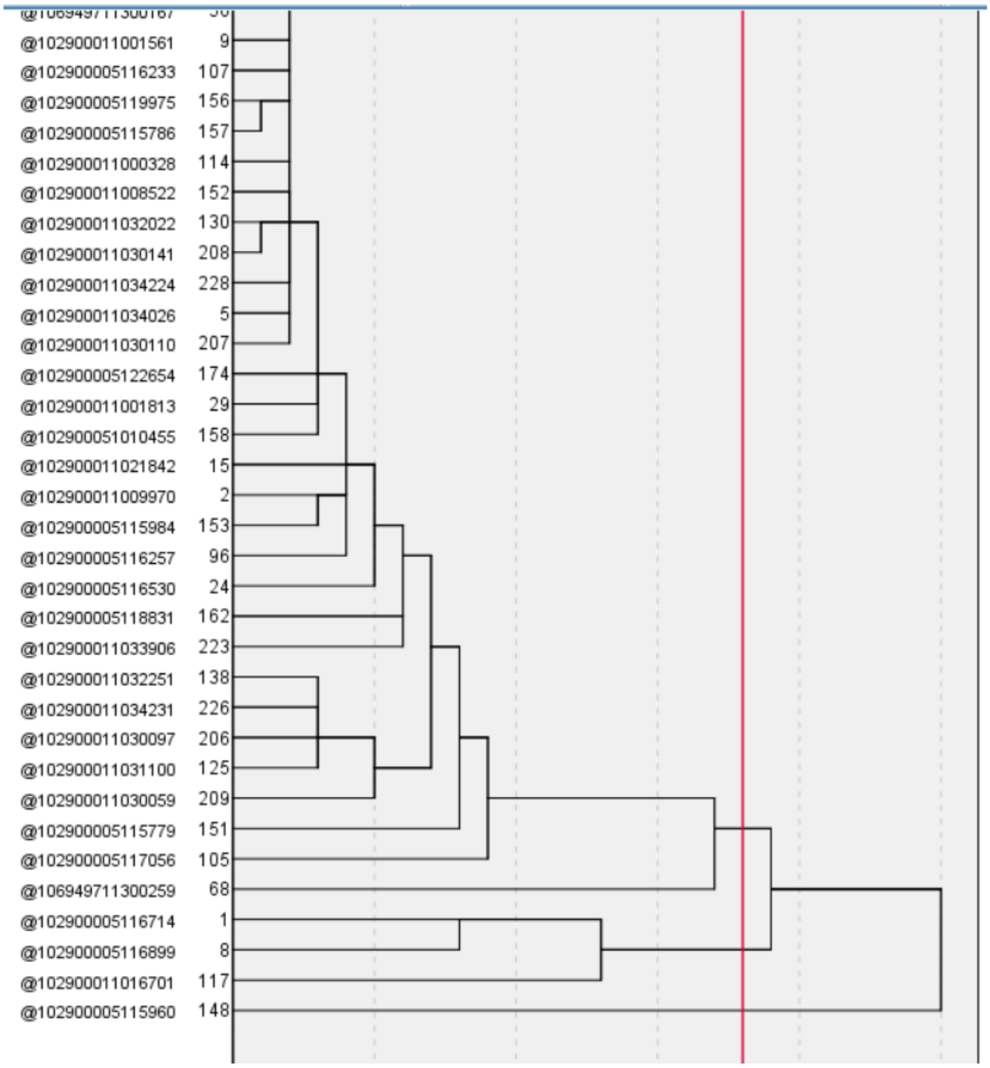


图 5 – 6 不同单品销售量谱系图

图 5 – 6 是以天为单位的不同单品销售量的聚类分析，并得出的不同单品销售量谱系图，反映不同单品销售量之间的相关关系。上图可以得出，蔬菜单品可分为三类，第一类为大白菜，第二类为西兰花，净藕（1），芜湖青椒（1），第三类为其他剩下的单品品种。对于第一类，大白菜作为家里的常用菜，家庭做饭经常都有大白菜，因此需求量很大，销售量也非常大，属于家庭常用菜；对于第二类，是做饭常见的配菜，与其他主菜搭配烹饪，因此这几类单品销售量关联系很大，属于家庭常用配菜；对于第三类，则各单品之间的关联系不高，所以属于其他菜品。

s

5.3 问题 2 模型的建立与求解

将问题 2 分为求解函数关系以及建立 ARIMA[8]模型两部分来解决：

1. 求解销售总量与定价之间的函数关系：

首先将附件二中的数据进行处理，得出各品类在的定价以及对应的销售总量平均值表格，接着将数据导入 matlab 中画出对应的二维线图，图 2-1 到图 2-6 为各品类的定价以及对应的销售总量平均值之间的二维线图。

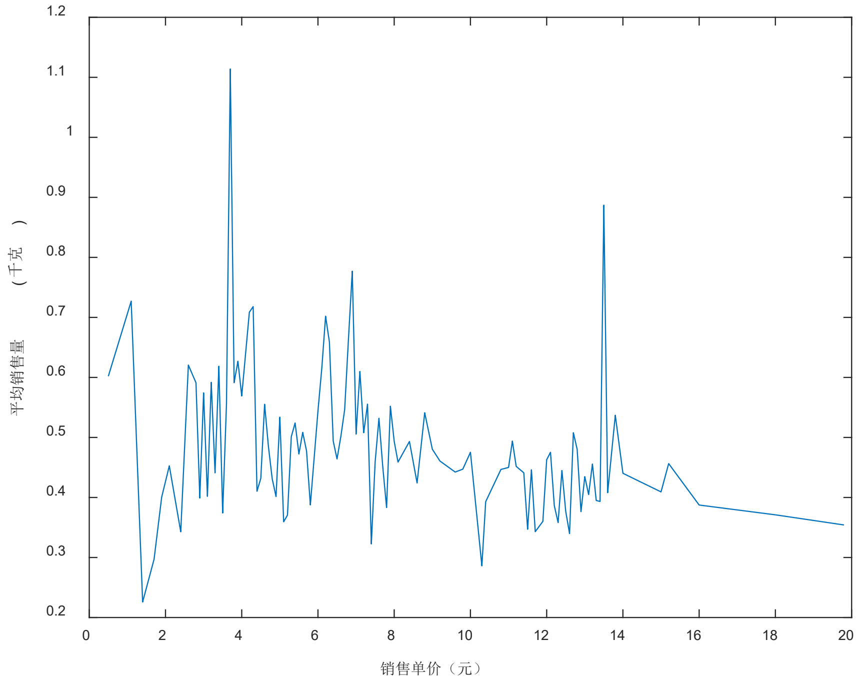


图 5-7 花菜类销售单价与平均销售量二维线图

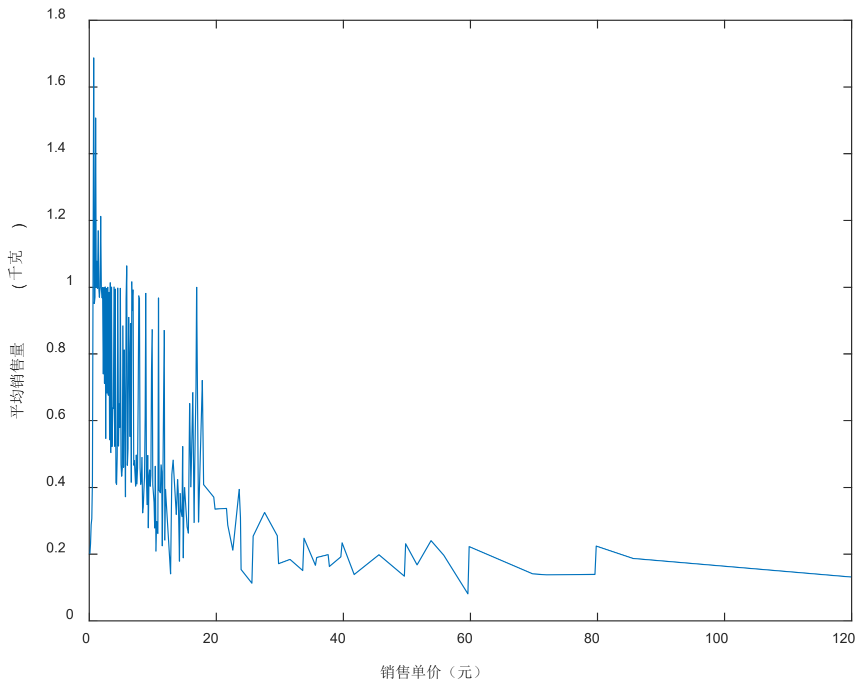


图 5-8 花叶类销售单价与平均销售量二维线图

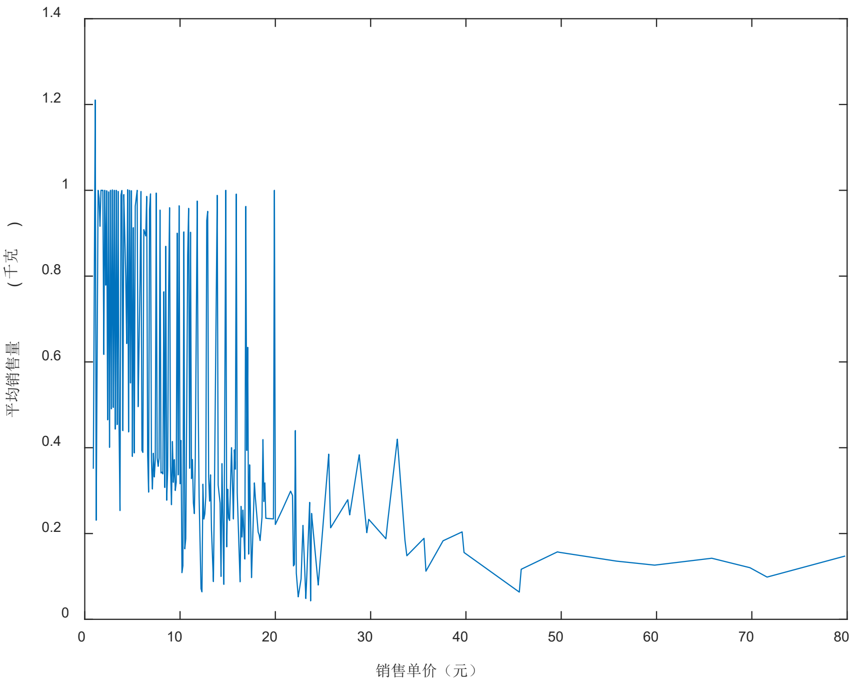


图 5-9 辣椒类销售单价与平均销售量二维线图

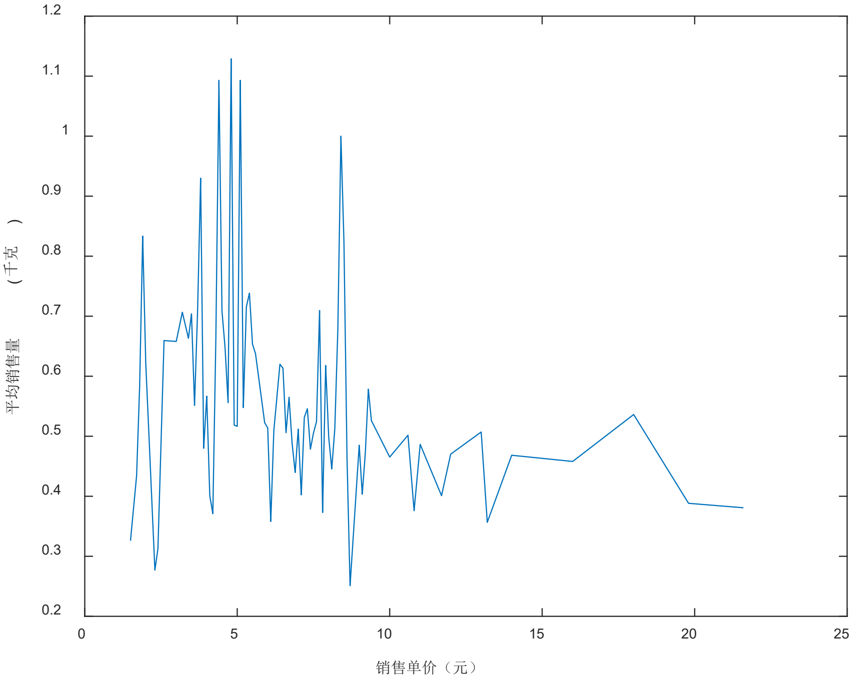


图 5-10 茄类销售单价与平均销售量二维线图

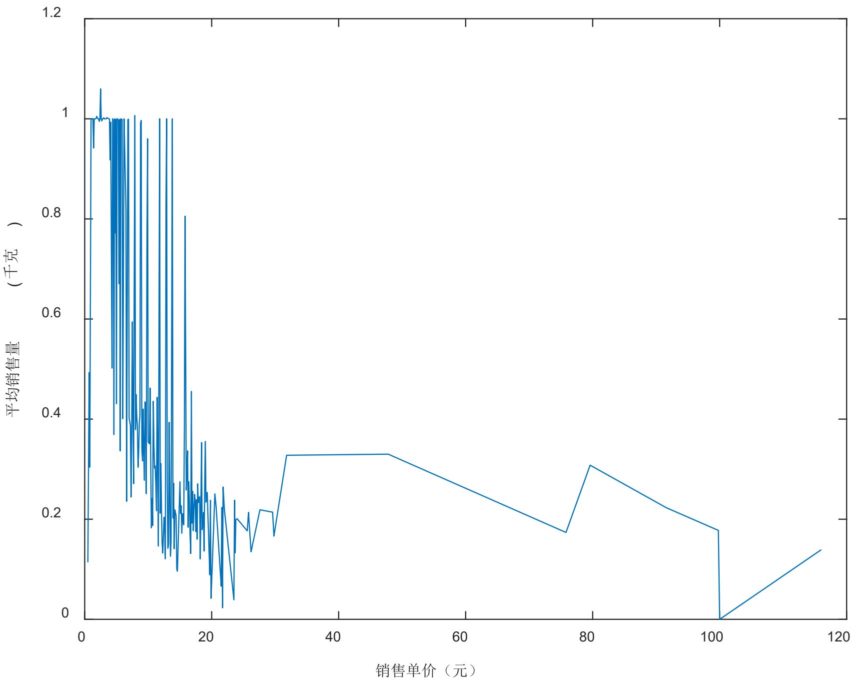


图 5-11 食用菌类销售单价与平均销售量二维线图

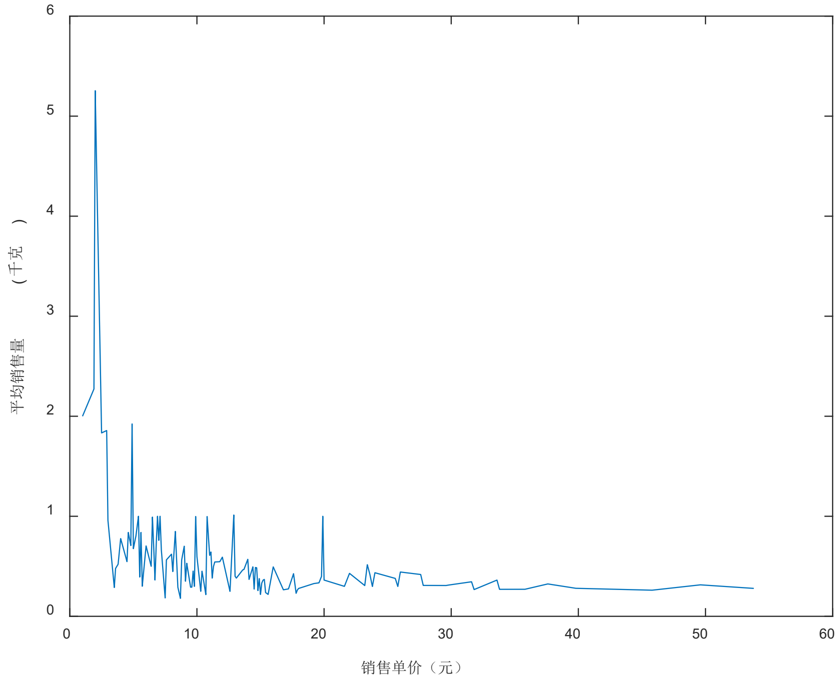


图 5-12 水生根茎类销售单价与平均销售量二维线图

通过二维线图我们发现两者之间存在线性关系，所以我们尝试运用多项式拟合来分析函数关系式，调节拟合次数来得出函数关系式，并运用均方根误差判断[7]和皮尔逊相关系数来判断拟合程度是否合适，最终的发现各品类在拟合 19~20 次后的拟合程度最好，如果小于 19~20 次会出现皮尔逊相关系数较小，如果大于 19~20 次会出现过拟合的情况，图 2-8 到图 2-13 为各品类拟合后曲线与实际值的混合图。

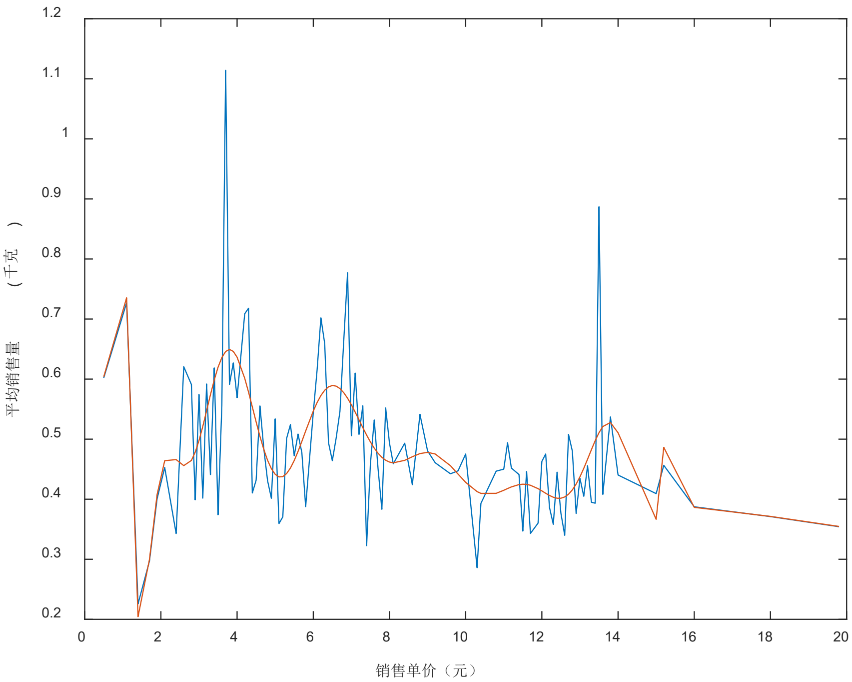


图 5-13 花菜类拟合曲线与实际值混合图

函数关系式如下：

$$
\begin{aligned}
saleAverage = {} & 3.5453e-15P_i^{19} - 8.6659e-13\,P_i^{18} + 9.0805e-11\,P_i^{17} - 5.5769e-09\,P_i^{16} \\
& + 2.2812e-07\,P_i^{15} - 6.6481e-06\,P_i^{14} + 0.00014364\,P_i^{13} - 0.0023573\,P_i^{12} + P_i^{11} \\
& - 0.29277\,P_i^{10} + 2.2363\,P_i^{9} - 13.2443\,P_i^{8} + 60.3209\,P_i^{7} - 208.3701\,P_i^{6} \\
& + 534.6559\,P_i^{5} - 988.2825\,P_i^{4} + 1256.8428\,P_i^{3} - 1021.5181\,P_i^{2} + 464.0181\,P_i \\
& - 85.2391
\end{aligned}
$$

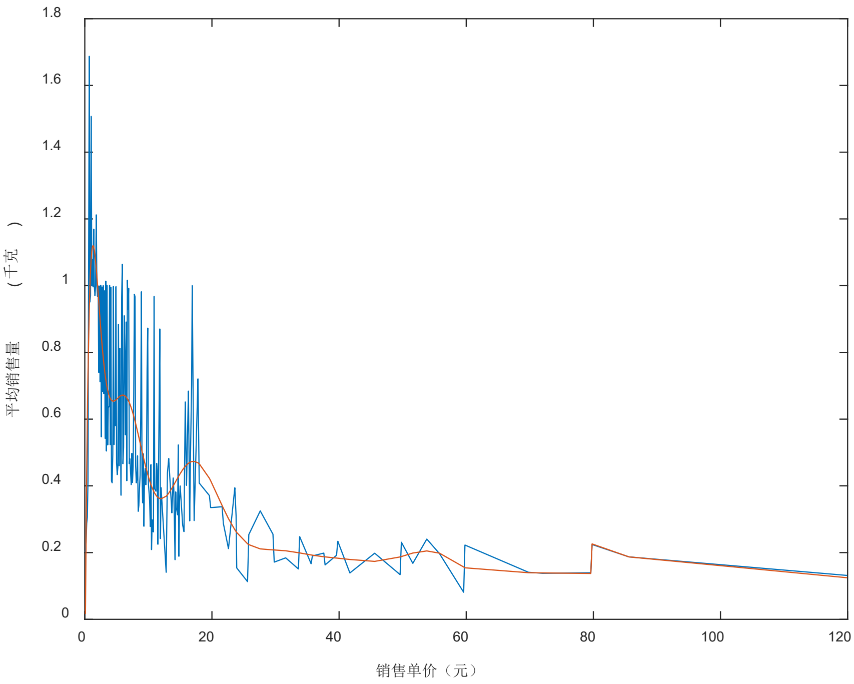


图 5-14 花叶类拟合曲线与实际值混合图

函数关系如下：

$$
\begin{aligned}
avg\_values = {} & 4.2195e-28\,P_i^{19} - 3.6736e-25\,P_i^{18} + 1.4744e-22\,P_i^{17} - 3.6208e-20\,P_i^{16} \\
& + 6.0887e-18\,P_i^{15} - 7.4328e-16\,P_i^{14} + 6.8128e-14\,P_i^{13} - 4.7827e-12\,P_i^{12} \\
& + 2.5994e-10\,P_i^{11} - 1.0983e-08\,P_i^{10} + 3.6006e-07\,P_i^{9} - 9.0927e-06\,P_i^{8} \\
& + 0.0001745\,P_i^{7} - 0.0024925\,P_i^{6} + 0.025703\,P_i^{5} - 0.18312\,P_i^{4} + 0.84469\,P_i^{3} \\
& - 2.2771\,P_i^{2} + 2.9275\,P_i - 0.25473
\end{aligned}
$$

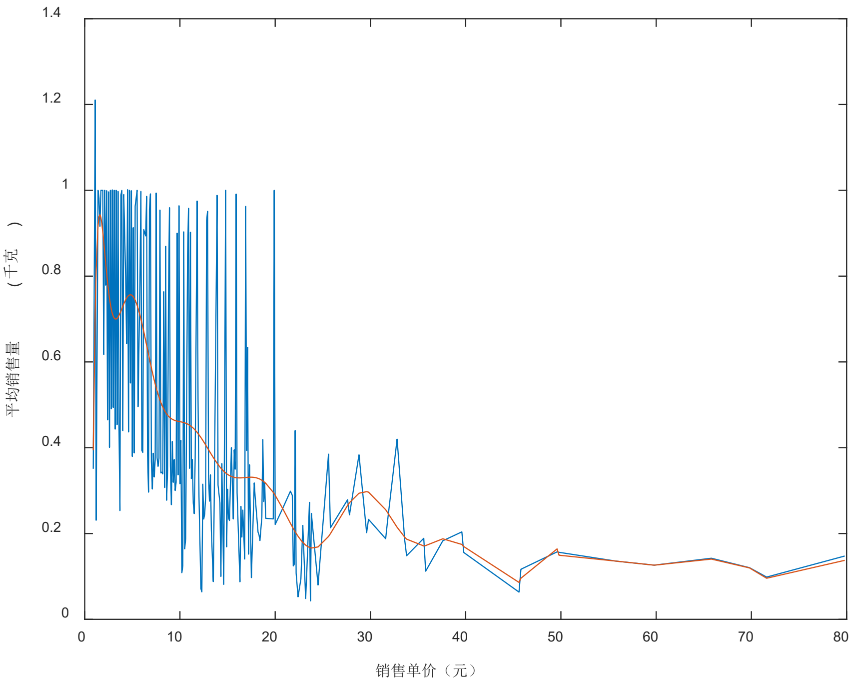


图 5-15 辣椒类拟合曲线与实际值混合图

函数关系如下：

$$
\begin{aligned}
avg\_values2 = & -6.7472e-27\ P_i^{20} + 4.7959e-24 P_i^{19} - 1.5768e-21\ P_i^{18} + 3.1823e-19\ P_i^{17} \\
& - 4.413e-17\ P_i^{16} + 4.4595e-15\ P_i^{15} - 3.3987e-13\ P_i^{14} + 1.9946e-11\ P_i^{13} \\
& - 9.1249e-10\ P_i^{12} + 3.2736e-08\ P_i^{11} - 9.2204e-07\ P_i^{10} + 2.0321e-05\ P_i^{9} \\
& - 0.00034776\ P_i^{8} + 0.0045639\ P_i^{7} - 0.045101\ P_i^{6} + 0.32708\ P_i^{5} - 1.6787\ P_i^{4} \\
& + 5.7868\ P_i^{3} - 12.3749\ P_i^{2} + 14.\ P_i - 5.7416
\end{aligned}
$$

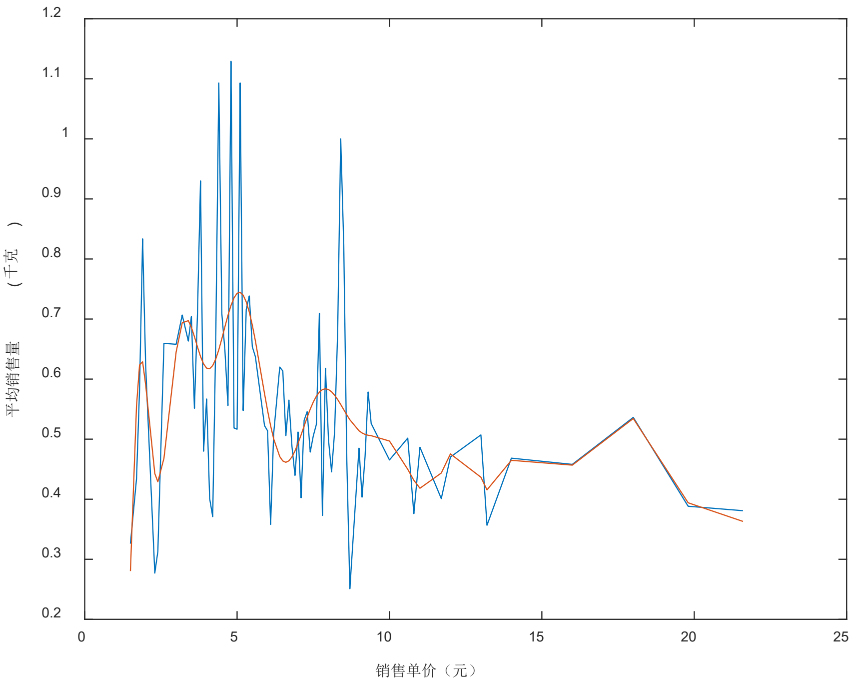


图 5-16 茄类拟合曲线与实际值混合图

函数关系如下：

$$\begin{aligned}avg\_values4 = &-7.255e-14\,P_i^{19} + 1.3488e-11\,P_i^{18} - 1.1614e-09\,pP_i^{17} + 6.1502e-08\,P_i^{16}\\ &- 2.2429e-06\,P_i^{15} + 5.9769e-05\,P_i^{14} - 0.0012052\,P_i^{13} + 0.01879P_i^{12}\\ &- 0.22942\,P_i^{11} + 2.2089\,P_i^{10} - 16.8048\,P_i^{9} + 100.7908\,P_i^{8} - 473.4642\,P_i^{7}\\ &+ 1722.2745\,P_i^{6} - 4768\,P_i^{5} + 9790.7689\,P_i^{4} - 14342.6456\,P_i^{3} + 14074.1392\,P_i^{2}\\ &- 8235.5032\,P_i + 2159.3026\end{aligned}$$

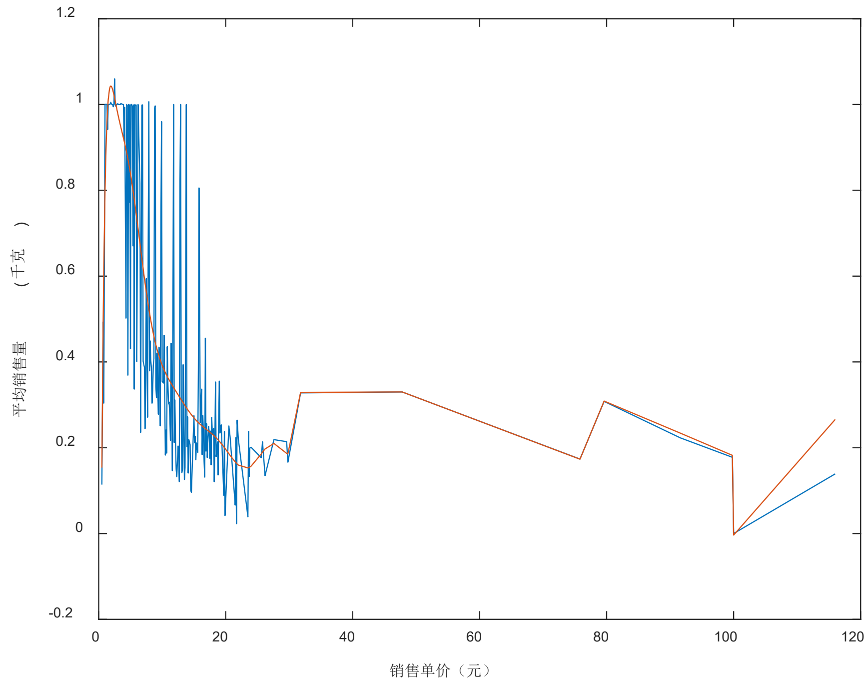


图 5-17 食用菌类拟合曲线与实际值混合图

函数关系如下：

$$
\begin{aligned}
avg\_values6 = {} & 1.2499e-26\ P_i^{19} - 1.0235e-23\ P_i^{18} + 3.8041e-21\ P_i^{17} - 8.5056e-19 P_i^{16} \\
& + 1.2788e-16\ P_i^{15} - 1.3692e-14\ P_i^{14} + 1.0793e-12\ P_i^{13} - 6.3884e-11\ pP_i^{12} \\
& + 2.8714e-09\ P_i^{11} - 9.8519e-08\ P_i^{10} + 2.5806e-06\ P_i^{9} - 5.1373e-05\ P_i^{8} \\
& + 0.00076989\ P_i^{7} - 0.0085528 P_i^{6} + 0.068846\ P_i^{5} - 0.3887\ P_i^{4} + 1.4704\ P_i^{3} \\
& - 3.4892\ P_i^{2} + 4.5652\ P_i - 1.4182
\end{aligned}
$$

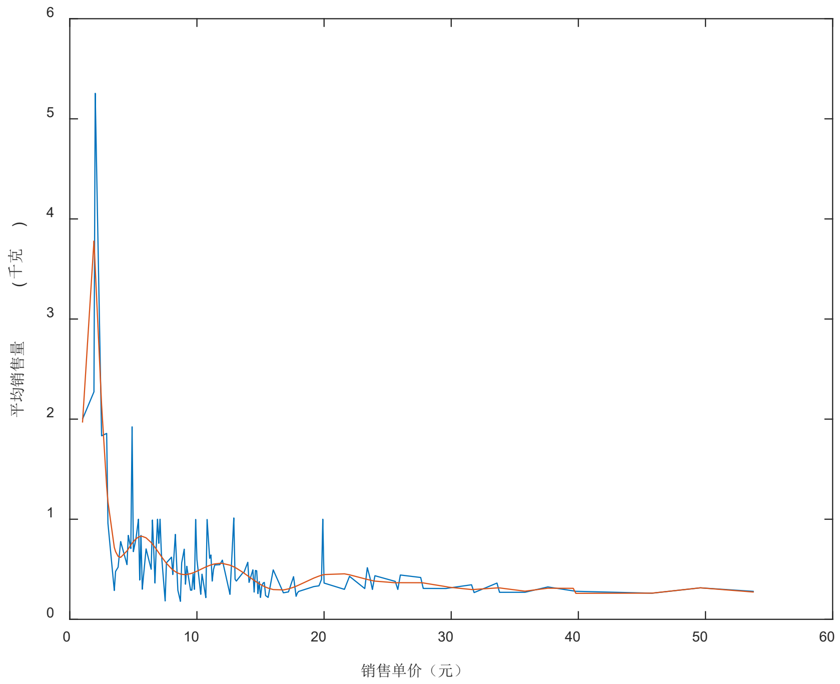


图 5-18 水生根茎类拟合曲线与实际值混合图

函数关系如下：

$$
\begin{aligned}
avg\_values8 = & -3.825e-24\ P_i^{20} + 1.9294e-21\ P_i^{19} - 4.5124e-19\ P_i^{18} + 6.4994e-17\ P_i^{17} \\
& - 6.459e-15\ P_i^{16} + 4.7028e-13\ P_i^{15} - 2.6003e-11\ P_i^{14} + 1.117e-09\ P_i^{13} \\
& - 3.7829e-08\ P_i^{12} + 1.019e-06\ P_i^{11} - 2.1933e-05\ P_i^{10} + 0.00037741 P_i^{9} \\
& - 0.0051732\ P_i^{8} + 0.056001\ P_i^{7} - 0.4717\ P_i^{6} + 3.0199\ P_i^{5} - 14.1822\ P_i^{4} + 46.2955\ P_i^{3} \\
& - 96.3917\ P_i^{2} + 109.5232\ P_i - 45.8784
\end{aligned}
$$

2. 建立 ARIMA 模型：

首先将附件二与附件三的数据进行处理，得到各品类每天的销售平均量与平均进价的表格，将表格数据导入 spss 中后分析日期与销售总量平均值的时序图，并分析一阶差分时的自相关性与偏相关性，建立传统 ARIMA 模型，发现平稳 R 方的数值过低，说明拟合程度不好（因分析图片过多，故本小题中只以花菜类为例，其他品类的图片放在附录之中）

表 5-2 花菜类一阶差分时 ARIMA 模型统计

模型	预测变量数	模型拟合度统计	杨-博克斯 Q(18)			离群值数
		平稳 R 方	统计	DF	显著性	
销量(千克)-模型_1	0	.255	57.618	9	.000	0

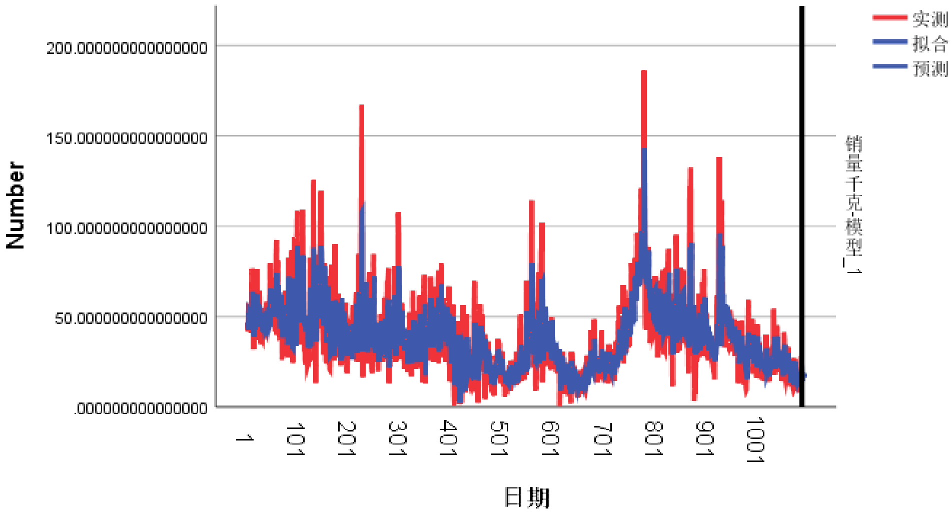


图 5-19 花菜类一阶差分时 ARIMA 模型

所以我们对数据进行三次差分后发现，平稳R方数值接近1，说明拟合程度较好

表 5-3 花菜类三阶差分时 ARIMA 模型统计

模型	预测变量数	模型拟合度统计	杨-博克斯 Q(18)			离群值数
		平稳 R 方	统计	DF	显著性	
销量(千克)-模型_1	0	.875	125.771	5	.000	0

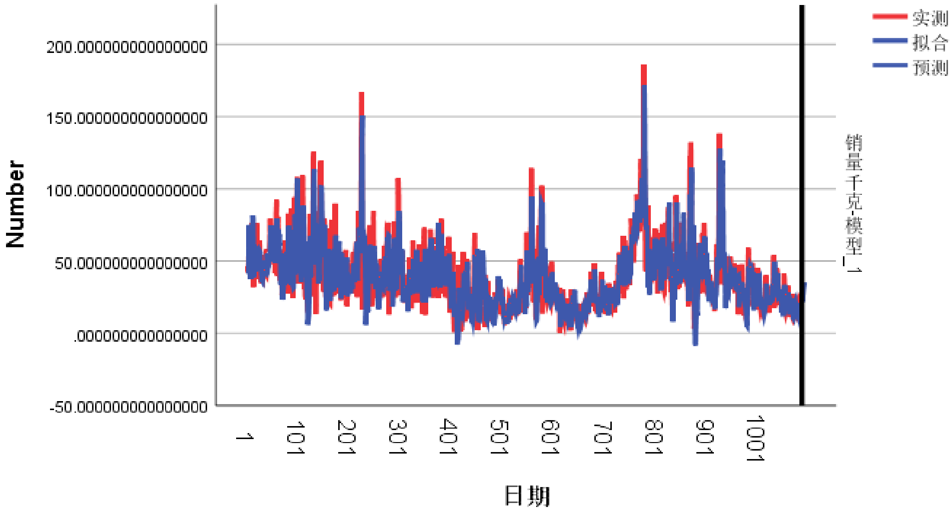


图 5-20 花菜类三阶差分时 ARIMA 模型

所以，根据三阶差分时的ARIMA模型可以预测出之后七天的销售量数据。

接着分析各品类平均进价的ARIMA模型，运用同上方法发现，进行一次差分就可以得到模拟程度较高的ARIMA模型。

表 5-4 花菜类一阶差分时 ARIMA 模型统计

		模型拟合度统计	杨-博克斯 Q(18)			
模型	预测变量数	平稳 R 方	统计	DF	显著性	离群值数
平均进价-模型_1	0	8.882E-16	66.376	18	.000	0

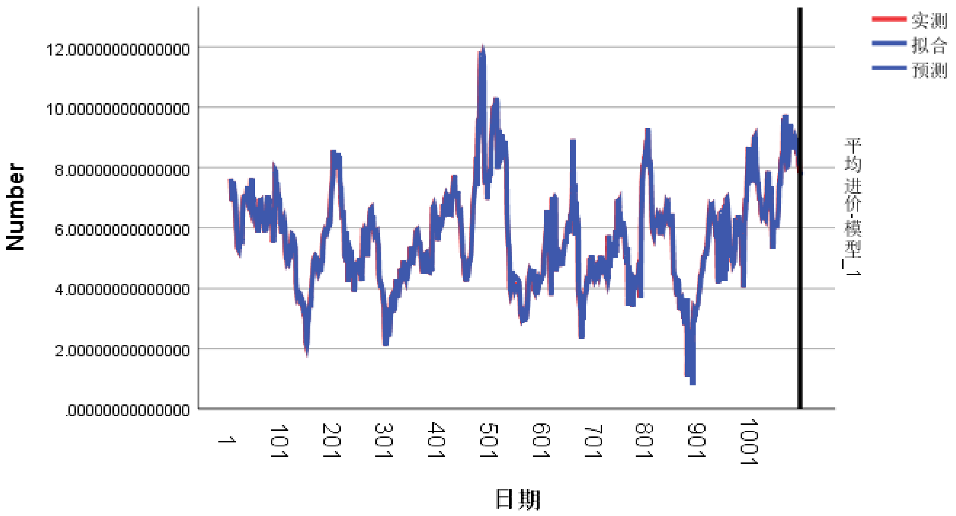


图5-21 花菜类一阶差分时ARIMA模型

所以，根据一阶差分时的ARIMA模型可以预测之后七天的进价数据。

通过预测的销售量数据得出对应天数定价策略，但因为拟合次数过多导致求解时的解有不同值，如果采用较高的定价策略会导致商超的客户粘度下降，虽然短期内会有较高的收益，但是不利于商超的长期经营，过低的定价策略会导致商超的花费入不敷出，可能最终倒闭，所以我们认为取解中的中位数作为定价策略有利于商超的长期发展并能获得较大的利润。

表 5-4 2023.07.01-202307.07 各品类进货量

日期	花菜类进货量（千克）	花叶类进货量（千克）	辣椒类进货量（千克）	茄类进货量（千克）	食用菌类进货量（千克）	水生根茎类进货量（千克）
2023/7/1	23.81	132.16	72.07	20.84	34.35	18.12
2023/7/2	22.01	117.76	69.08	15.22	30.63	15.54
2023/7/3	22.46	94.77	61.08	10.90	27.59	14.18
2023/7/4	25.58	98.09	60.58	8.36	29.25	13.64
2023/7/5	29.66	115.68	61.65	8.81	31.29	18.33
2023/7/6	33.25	121.69	67.45	9.63	28.86	22.60
2023/7/7	35.36	117.28	63.38	11.32	24.83	22.61

表 5-5 2023.07.01-2023.07.01 各品类定价策略

日期	花菜类	花叶类	辣椒类	茄类	食用菌类	水生根茎类
2023/7/1	0.5647	0.5877	0.624	0.6032	0.5882	0.5333
2023/7/2	0.5637	0.5877	0.6241	0.5983	0.5882	0.5247
2023/7/3	0.564	0.5854	0.6241	0.5928	0.5883	0.52
2023/7/4	0.5656	0.5855	0.6241	0.5883	0.5711	0.5012
2023/7/5	0.5673	0.5858	0.6241	0.5892	0.5724	0.534
2023/7/6	0.5686	0.5859	0.6241	0.5907	0.5709	0.5371
2023/7/7	0.5693	0.5859	0.6241	0.5934	0.5677	0.5371

5.4 问题 3 模型的建立与求解

问题 3 需要分析 6 月 24 日-30 日的可售品种，保证可以满足市场对不同品类的需求，且单品订购量满足最小陈列量 2.5 千克，预测日补货总量并制定合理的定价策略，使得商超收益最大化。首先利用预处理后的数据，计算各类单品补货量与定价之间的关系，得出不同单品的线性回归模型。然后引入开关变量 qi=0 或 1，i=1,2,3,...251，qi 为 1 则表示该单品当天有销售，为 0 则表示没有进行销售。

$$q_i=\begin{cases}0, & 第i个单品当天没有出售\\ 1, & 第i个单品当前进行出售\end{cases}$$

因此，单品总数为：

$$27 \le \sum_{i=1}^{251} q_i \le 33$$

各单品订购量满足最小陈列量约束总量为 e_i，i=1,2,3，...，251 需要大于 2.5 千克，即

$$e_i \ge 2.5, i=1,2,3,\dots,251$$

利润计算，涉及到题目二的成本加成定价，假设 Wi 为利润率，则单品定价为

$$y_i = (1 + w_i)c_i$$

其中，yi 为第 i 中单品定价，ci 为第 i 种单品的成本价，则最终盈利为

5.5 问题 4 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，商超除了参考以往的商品销售流水明细，批发价格，损坏情况等还可以参考以下相关信息：

（一）从消费者的角度考虑

（1）消费者销售习惯数据

通常情况下，消费者都会购买自己习惯搭配的菜式，收集消费者的购买数据，购物车组合，购买偏好，这样可以更好地了解消费者的购买需求，优化蔬菜的组合和促销策略。

（2）消费者需求反馈数据

通过消费者评价反馈平台，实时关注消费者的需求，可以更好地稳定客源，提升消费者好感，更好地调整补货量和调整定价策略。

（二）从销售环境的角度考虑

（1）供应链数据

供应商的交货时间，可靠性会影响销售时间和销售量，所以了解供应链的信息，供应商的可靠性，具体的交货时间，这样可以优化补货计划。

（2）市场销售趋势

市场销售趋势反映了批发货物的趋势和消费者的消费趋势，所以实时了解关注市场的变化趋势，包括批发价格趋势，市场销售价格趋势，季节性变化趋势，新品种上市趋势等，有助于调整补货和销售策略，更好地适应市场环境。

（3）竞争对手销售数据

竞争对手会影响自身在周边的销售环境，影响销售量等，所以了解附近超市的销售价格，促销活动等，可以更好地了解市场竞争情况，改变经营策略。

（4）销售高峰时间数据

节假日，过年期间，消费者的购买量会大大增加，这样会提供商超的销售量，所以收集销售数据，整理出销售量高的时间，包括每天的销售高峰情况，周六日的销售高峰情况，节假日，过年的销售高峰情况等，这样更好地对补货和定价进行调整。

（5）天气数据

时刻关注天气信息，天气会影响消费者的出门购买欲望，所以天气数据可以帮助商场更好地调整补货计划。

（6）线上网络平台销售的影响

如今线上购买蔬菜越来越便利，线上销售平台的比重越来越高，但是线上不便于消费者观察蔬菜新鲜度，因此关注线上的销售情况也更好地了解市场的销售方式的趋势。

（三）从自身经营策略考虑

（1）促销活动数据

促销活动可以吸引消费者的购买欲望，包括打折销售，促销时段，销售渠道等，收集促销活动的数据可以调整更好地的促销方式，增加销售量

六、参考文献

[1] 童宇，物化探测量元素基于 R 型聚类分析的分类[J]，《西部探矿工程》，2021 年第 7 期，138-140；

[2] 笨牛慢耕，斯皮尔曼相关(Spearman correlation)系数概述及其计算例，https://blog.csdn.net/chenxy_bwave/article/details/121427036，2023.9.10;

[3] Matplotlib API Reference *API Reference — Matplotlib 3.7.2 documentation*;

[4] pandas documentation *pandas documentation — pandas 2.1.0 documentation (pydata.org)*;

[5] matlab MathWorks 帮助和文档 https://ww2.mathworks.cn/help/matlab/learn_matlab/help.html；

[6] 司守奎，孙玺菁.数学建模算法与应用[M]北京:国防工业出版社，2016-2.；

[7] 张玲.单因素及双因素方差分析及检验的原理及统计应用 IJ.数学学习与研究 2010，07:94-96.；

[8] Shumway, R. H. & Stoffer, D. S. ARIMA models. In Time Series Analysis and Its Applications 75 – 163 (Springer, 2017).；

[9] de Winter, J. C. F., Gosling, S. D., & Potter, J. (2016). Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. Psychological Methods, 21(3), 273 – 290. https://doi.org/10.1037/met0000079；

七、附件

8.1 文件清单

数据处理 Python 代码

问题一相关文件

各单品天销售量.xlsx

各单品总销售量.xlsx

各品类月销售量.xlsx

各品类总销售量.xlsx

聚类分析.sav

谱系图.spv

问题二相关文件

matlab 代码（文件夹）

spss 各品类平均进价按天数预测（文件夹）

spss 各品类销售总量按天数预测（文件夹）

各品类价格与销售量关系列表（文件夹）

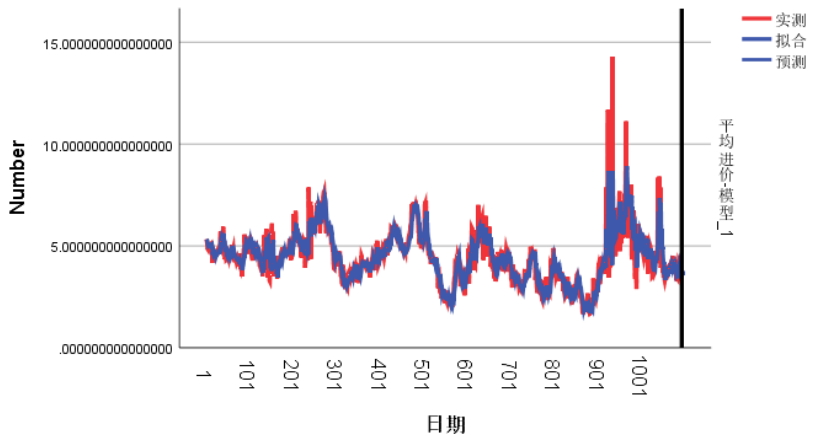
各品类平均进价与时间列表（文件夹）

各品类销售量按天数求和（文件夹）

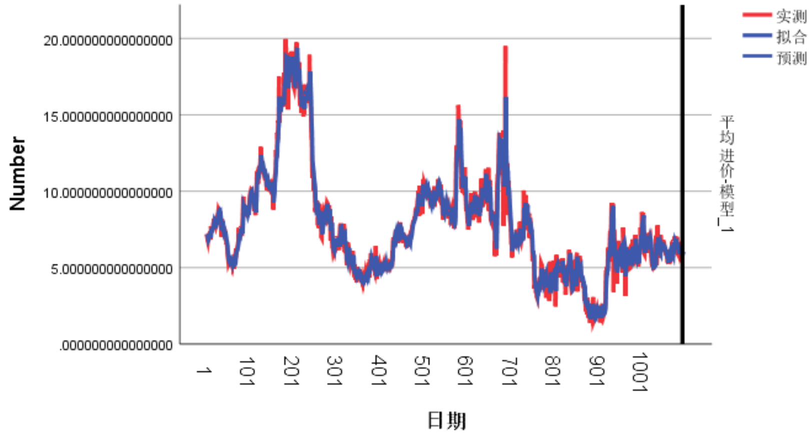
定价策略.xlsx

进货量预测数据.xlsx

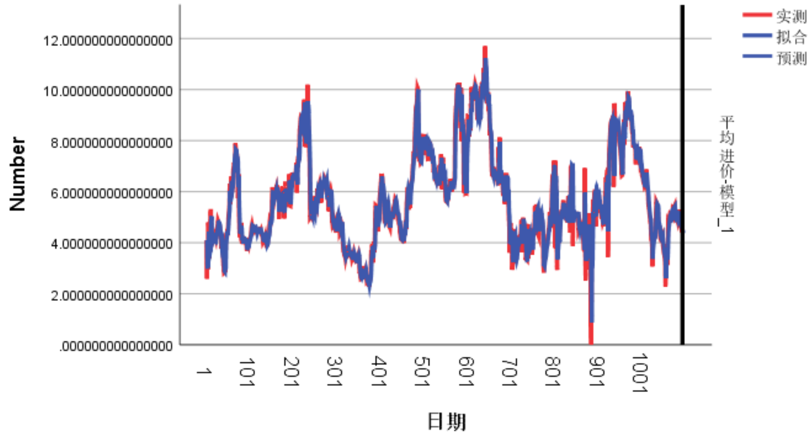
花叶类一阶差分的 ARIMA 模型



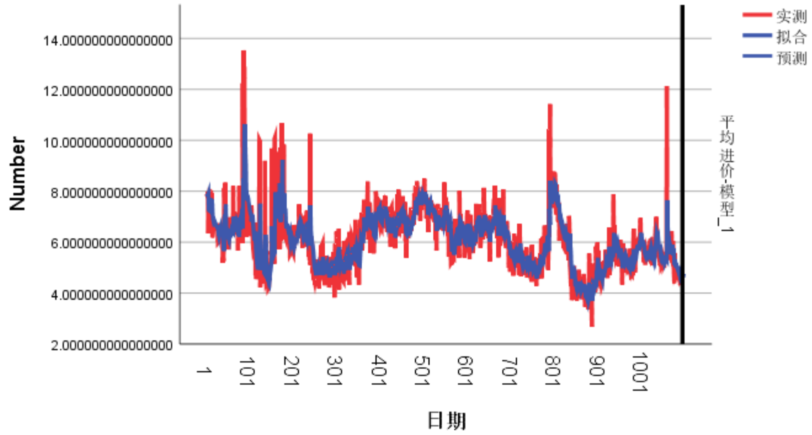
辣椒类一阶差分的 ARIMA 模型



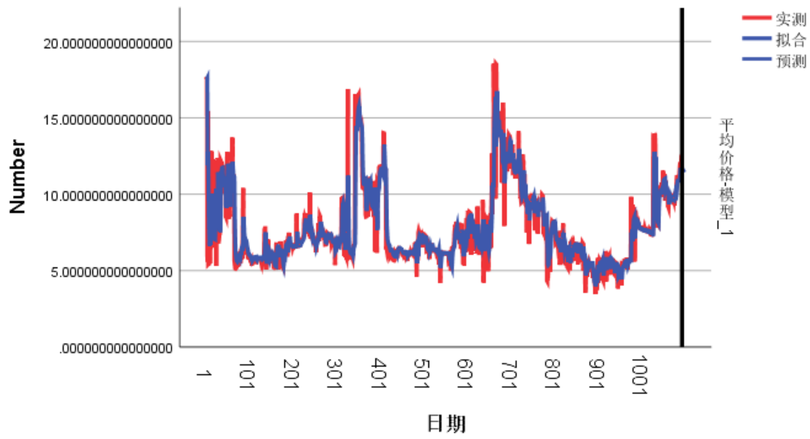
茄类一阶差分的 ARIMA 模型



食用菌类一阶差分的 ARIMA 模型



水生根茎类一阶差分的 ARIMA 模型



8.3 核心代码

程序编译器版本：

代码 1 至代码 2：MATLAB2020a

代码 1

```
plot(price,saleaverage)
a0=polyfit(price,saleaverage,19)
saleaverage1=polyval(a0,price)
hold on
plot(price,saleaverage1)
xlabel('销售单价（元）');
ylabel('平均销售量(千克)');
corrcoef(saleaverage,saleaverage1)
sqrt(mean((saleaverage -saleaverage1).^2))
% 给定因变量值
given_saleaverage =
[23.806739232337247,22.013595991848984,22.456808327912510,25.575863746142200,29.657139890678817,33.254591051630680,35.361497594847960];
%第一天
given_saleaverage1 = given_saleaverage(1,1)
% 使用 polyval 函数求解对应的自变量值
```

```
price_values1 = roots(flip(a0) - given_saleaverage1);
% 打印自变量值
disp(price_values1);
%第二天
given_saleaverage2 = given_saleaverage(1,2)
% 使用 polyval 函数求解对应的自变量值
price_values2 = roots(flip(a0) - given_saleaverage2);
% 打印自变量值
disp(price_values2);
%第三天
given_saleaverage3 = given_saleaverage(1,3)
% 使用 polyval 函数求解对应的自变量值
price_values3 = roots(flip(a0) - given_saleaverage3);
% 打印自变量值
disp(price_values3);
%第四天
given_saleaverage4 = given_saleaverage(1,4)
% 使用 polyval 函数求解对应的自变量值
price_values4 = roots(flip(a0) - given_saleaverage4);
% 打印自变量值
disp(price_values4);
%第五天
given_saleaverage5 = given_saleaverage(1,5)
% 使用 polyval 函数求解对应的自变量值
price_values5 = roots(flip(a0) - given_saleaverage5);
% 打印自变量值
disp(price_values5);
%第六天
given_saleaverage6 = given_saleaverage(1,6)
```

```
% 使用 polyval 函数求解对应的自变量值
price_values6 = roots(flip(a0) - given_saleaverage6);
% 打印自变量值
disp(price_values6);
%第七天
given_saleaverage7 = given_saleaverage(1,7)
% 使用 polyval 函数求解对应的自变量值
price_values7 = roots(flip(a0) - given_saleaverage7);
% 打印自变量值
disp(price_values7);
% 构造拟合函数关系式
func = poly2str(a0,'price'); % 'x'表示变量名称
disp(['拟合函数表达式：saleaverage= ' func]);
```

代码 2

```
%使用 unique 函数获取 data 表格中的所有唯一日期。
unique_dates = unique(costdata1.Date);
%创建成本平均值表格
average_data = table();
average_data.Date = unique_dates;
%使用循环遍历每个唯一日期，在每次循环中计算该日期对应的值的平均值，将结果存储到 average_data 表格中。
for i = 1:length(unique_dates)
   date = unique_dates(i);
   values = costdata1.Value(costdata1.Date == date);
   avg_value = mean(values);
   average_data.AverageValue(i) = avg_value;
end
```