

蔬菜类商品的分类和销售策略问题的研究

袁文清, 朱晋铭, 孟铂凯, 鲁忠良, 张镇宇, 于洛刚

沈阳航空航天大学航空宇航学院, 辽宁 沈阳

收稿日期: 2024年3月19日; 录用日期: 2024年4月17日; 发布日期: 2024年4月24日

摘要

随着越来越多的人开始追求新鲜的蔬菜食品, 对于超市的运营方来说, 预测各个商品的销售情况对于超市的经营至关重要。本文从销售商品的历史记录出发, 对于未来各商品的销售情况进行分析与预测。根据已有的一批蔬菜商品销售数据中各商品的历史销售情况, 对之后的蔬菜类商品销售情况进行预测和分析, 使用spearman相关性及Pearson相关性分析法分析每年以及每季度中的不同单品之间的关系及不同品类之间的关系。使用线性回归模型预测并建立该品类的售价与其本身的批发价和销售量和与其相关性较强的品类的批发价和销售量有关的数学模型。由统计分析得出批发价在一年的周期上具有周期性, 使用偏最小二乘回归(PLSR)模型预测各品类未来一周的合适定价, 按此制定了补货策略和定价策略。其次我们深究了其他影响销售策略的因素, 由此更加准确地制定销售策略。

关键词

Spearman相关性分析, Pearson相关性分析, 偏最小二乘回归(PLSR)模型

Research on the Classification and Sales Strategy of Vegetable Product

Wenqing Yuan, Jinming Zhu, Bokai Meng, Zhongliang Lu, Zhenyu Zhang, Minggang Yu

School of Aeronautics and Astronautics, Shenyang Aerospace University, Shenyang Liaoning

Received: Mar. 19th, 2024; accepted: Apr. 17th, 2024; published: Apr. 24th, 2024

Abstract

As more and more people pursue fresh vegetable food, predicting the sales of various products is crucial for supermarket operations. This article analyzes and predicts the future sales of various products based on the historical records of sales. Based on the existing sales data of a batch of vegetable products, we predict and analyze the sales of vegetable products in the future. We use Spearman correlation and Pearson correlation analysis methods to analyze the relationships be-

tween different single products and categories in each year and each season. We use a linear regression model to predict and establish a mathematical model that relates the selling price of this category to its own wholesale price and sales volume, as well as the wholesale price and sales volume of categories with strong correlation. Statistical analysis shows that wholesale prices have periodicity over a one-year cycle. Partial least squares regression (PLSR) model is used to predict the appropriate pricing for each category in the next week, and replenishment and pricing strategies are formulated based on this. Secondly, we delved into other factors that affect sales strategies, in order to develop sales strategies more accurately.

Keywords

Spearman Correlation Analysis, Pearson Correlation Analysis, PLSR Model

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近年来,我国社会保持了和谐稳定的发展局面,为蔬菜行业的发展创造了良好的社会条件。居民健康意识和营养需求不断增强。随着我国社会不断进步,居民对健康和营养的关注度和需求度不断提高。由于商超销售的蔬菜品种众多、产地不尽相同,为此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。可靠的市场需求分析,对补货决策和定价决策尤为重要。现在有一家生鲜超市从2020年7月初到2023年6月末的各商品的相关数据,包含了6个蔬菜品类的商品信息、销售流水明细数据、蔬菜类商品的批发价格和蔬菜类商品的近期损耗率。蔬菜的定价一般采用“成本加成定价”方法,商超对运损和品相变差的商品通常进行打折销售。可靠的市场需求分析,对补货决策和定价决策尤为重要。从需求侧来看,蔬菜类商品的销售量与时间往往存在一定的关联关系;从供给侧来看,蔬菜的供应品种在4月至10月较为丰富,商超销售空间的限制使得合理的销售组合变得极为重要[1]。

2. 模型假设

- 1) 假设各品类售价与其本身及相关性较高的品类的批发价和时间相关。
- 2) 假设各品类批发价变化仅与时间有关。
- 3) 假设每个聚类中的每个单品的6~8月每周的销售量与此类中其他单品的批发价和时间(周数)有关。
- 4) 假设该超市为了保证新鲜程度,每天进货的蔬菜不会保留至第二天。

3. 蔬菜各品类及各单品销售量的分布规律

3.1. 数据预处理

销售数据包含销售菜品流水明细中的销售日期、扫码销售时间、单品编码、销量(千克)、销售单价(元/千克)、销售类型、是否打折销售的基本信息。由于数据数量过于庞大,故采用excel的数据透视表整理数据。先将数据分为2020-7至2021-6、2021-7至2022-6、2022-7至2023-6三组,分别生成数据透视表。其中将时间(月)作为列,单品编码为行,销量(kg)为值,生成表格,统计出每个单品在每个月的总销售量以及每大类的总销售量。

其中，因为考虑到打折销售会对产品销售量产生影响，故将正常销售的销售量按 1 倍计入，打折销售量按 0.5 倍计入。

经过统计，发现了有 5 种蔬菜 2 年内无出售记录，故删去(102900005116042 藕、102900005116776 本地菠菜、102900011016909 花茄子、102900011023648 芜湖青椒(2)、102900011032145 芜湖青椒(份))。

花叶类批发价格有两周存在异常值，使用插值替换后提高准确率。

茄类有四周无出售记录，我们将异常指删去以提高拟合准确率。

水生根茎类销售价格浮动过大，我们去除了销售价格远大于或远小于正常值范围的样本[2]。

3.2. 数据可视化

首先，我们将销售量数据经过 excel 表格中数据透视表处理后的统计数据建立图表，观察蔬菜各品类销售量的分布规律。

图 1 表示了蔬菜各品类销售量分布规律，由图得出规律：1) 花菜类和花叶类相关性较高；2) 花叶类和花菜类、水生根茎类相关性较高；3) 食用菌类和花叶类、水生根茎类、辣椒类相关性较高；4) 辣椒类和食用菌相关性较高；5) 茄类水和生根茎类相关性较高；6) 水生根茎类和食用菌类、花叶类、茄类相关性较高。

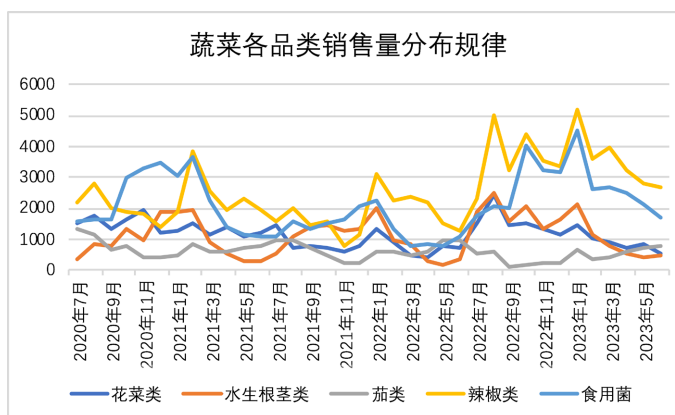


Figure 1. Sales line chart for each category

图 1. 各品类销售折线图

3.3. 相关性分析

鉴于以上分析，我们首先进行斯皮尔曼相关性分析每品类每年销量的数据。利用 SPSSPRO 对花叶类，花菜类，水生根茎类，茄类，辣椒类，食用菌进行相关性检验，数据分别取于一整年与四个季度，得到结果见表 1、表 2、图 2 和图 3。

表 1 为使用 Spearman 相关性分析得到的一整年内各品类蔬菜销售量的相关系数表。

图 2 为 Spearman 相关性分析优化后的相关系数热力图。

表 2 为使用 Pearson 相关性分析得到的一整年内各品类蔬菜销售量的相关系数表。

图 3 为 Pearson 相关性分析优化后的相关系数热力图。

在一整年的跨度中进行分析，初步判定其中辣椒类和食用菌、食用菌和花叶类、食用菌和水生根茎类、花菜类和花叶类的销售量两两正相关。

由于蔬菜具有季节性，不同季节内蔬菜的销售情况往往不尽相同。我们将一年划分四个季度：3 至 5 月为春季，6 至 8 月为夏季，9 至 11 月为秋季，12 至次年 2 月为冬季。将一年内的蔬菜销售量按季节划

Table 1. Spearman correlation coefficient table for sales volume of various categories of vegetables
表 1. 各品类蔬菜销售量 Spearman 相关系数表

	花叶类	花菜类	水生根茎类	茄类	食用菌	辣椒类
花叶类	1	0.695	0.448	-0.043	0.578	0.488
花菜类	0.695	1	0.427	0.076	0.462	0.31
水生根茎类	0.448	0.427	1	-0.467	0.669	0.316
茄类	-0.043	0.076	-0.467	1	-0.447	-0.17
食用菌	0.578	0.462	0.669	-0.447	1	0.49
辣椒类	0.488	0.31	0.316	-0.17	0.49	1

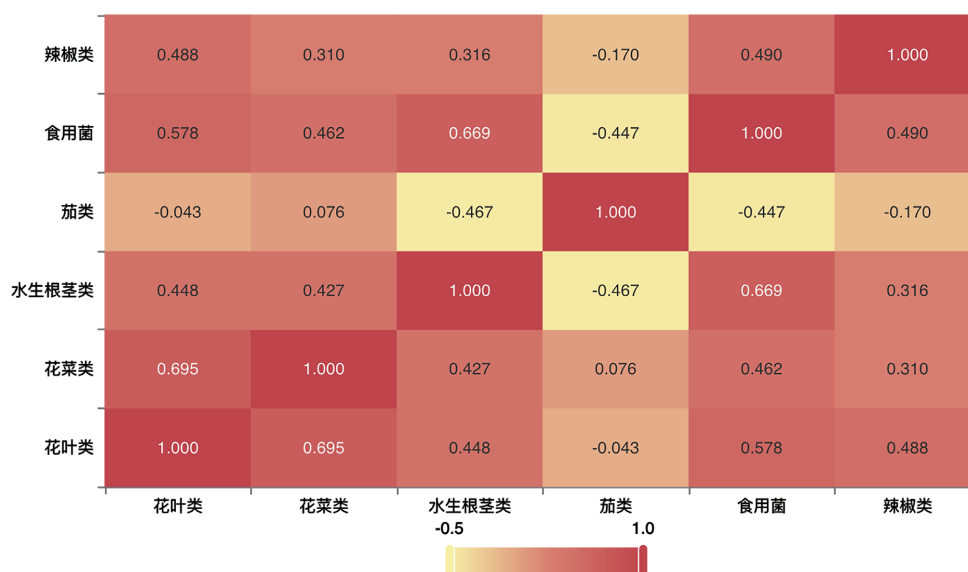


Figure 2. Heat map of Spearman correlation coefficient for sales volume of various categories of vegetables

图 2. 各品类蔬菜销售量 Spearman 相关系数热力图

Table 2. Pearson correlation coefficient table for sales volume of various categories of vegetables
表 2. 各品类蔬菜销售量 Pearson 相关系数表

	花叶类	花菜类	水生根茎类	茄类	辣椒类	食用菌
花叶类	1	0.746	0.484	-0.035	0.624	0.546
花菜类	0.746	1	0.472	0.058	0.425	0.429
水生根茎类	0.484	0.472	1	-0.466	0.459	0.653
茄类	-0.035	0.058	-0.466	1	-0.191	-0.409
辣椒类	0.624	0.425	0.459	-0.191	1	0.576
食用菌	0.546	0.429	0.653	-0.409	0.576	1

分。由图 2 和图 3 对比得知 Spearman 分析与 Pearson 分析结果近似。使用同样的方法得到各季节内各蔬菜品类之间的相关性，对比发现各季节内两种分析法呈现的相关性也近似，故下文只展示 Spearman 相关

性分析的结果。

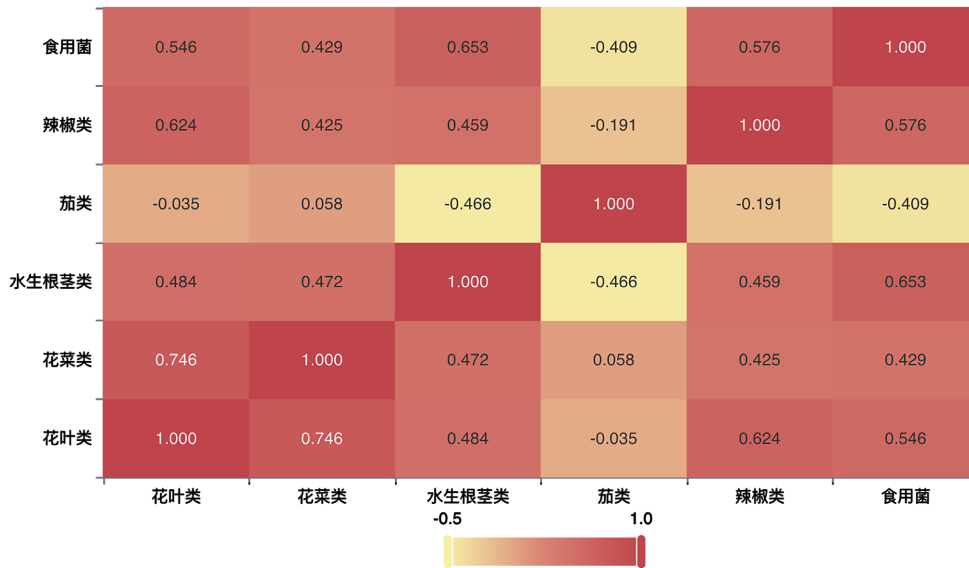


Figure 3. Heat map of Pearson correlation coefficient for sales volume of various categories of vegetables

图 3. 各品类蔬菜销售量 Pearson 相关系数热力图

- 图 4 为春季各品类蔬菜 Spearman 相关性分析优化后的相关系数热力图。
- 图 5 为夏季各品类蔬菜 Spearman 相关性分析优化后的相关系数热力图。
- 图 6 为秋季各品类蔬菜 Spearman 相关性分析优化后的相关系数热力图。
- 图 7 为冬季各品类蔬菜 Spearman 相关性分析优化后的相关系数热力图。

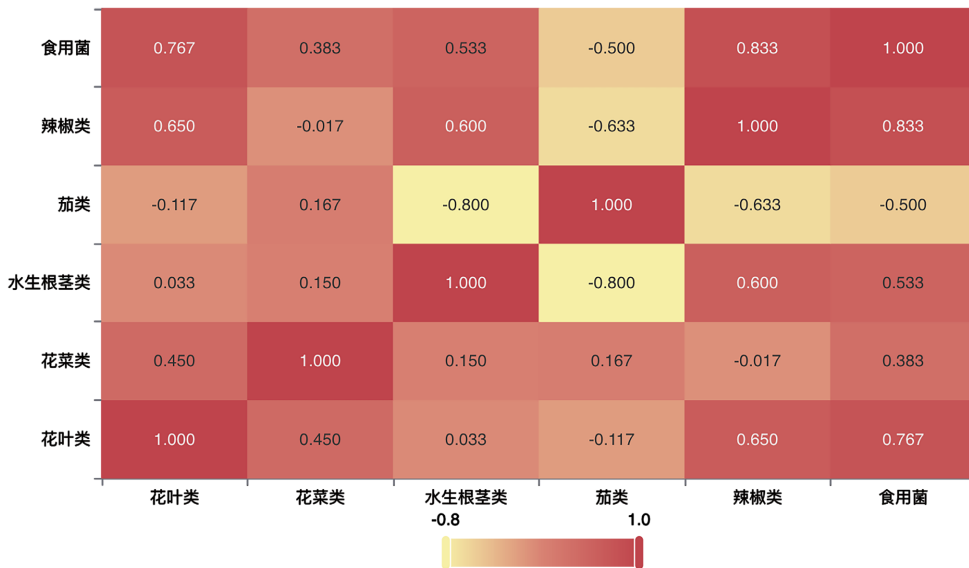


Figure 4. Heat map of Spearman correlation coefficient for sales volume of various categories of vegetables in spring

图 4. 春季各品类蔬菜销售量 Spearman 相关系数热力图

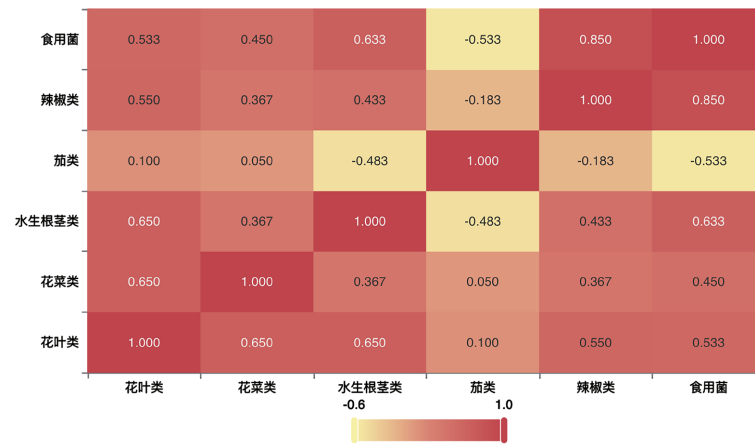


Figure 5. Heat map of Spearman correlation coefficient for sales volume of various categories of vegetables in summer

图 5. 夏季各品类蔬菜销售量 Spearman 相关系数热力图

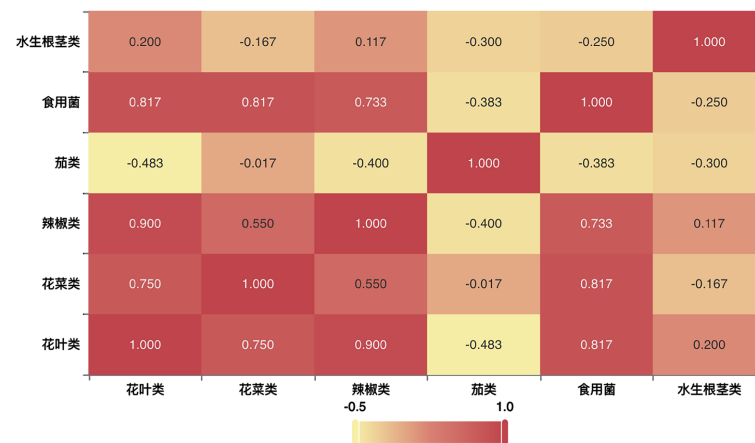


Figure 6. Heat map of Pearson correlation coefficient for sales volume of various categories of vegetables in autumn

图 6. 秋季各品类蔬菜销售量 Spearman 相关系数热力图

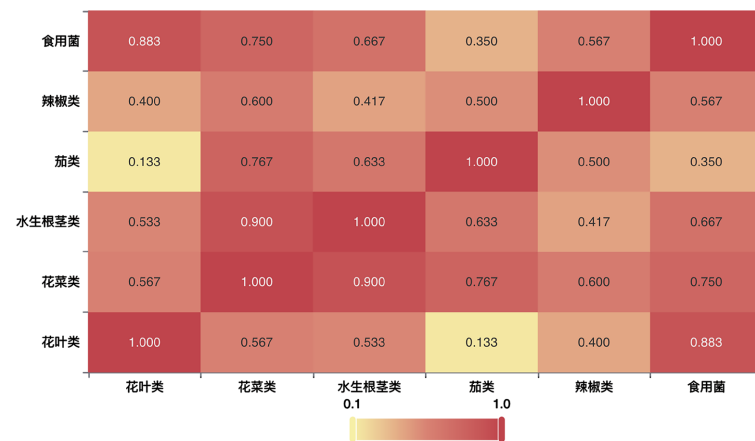


Figure 7. Heat map of Pearson correlation coefficient for sales volume of various categories of vegetables in winter

图 7. 冬季各品类蔬菜销售量 Spearman 相关系数热力图

综上所述：蔬菜各品类销售量之间的相关性在不同季节有较大差异，其中春季食用菌类和花叶、辣椒类正相关；辣椒类和食用菌、花叶类正相关；茄类和水生根茎类负相关；花菜类和花叶类正相关；花叶类与花菜类、食用菌类正相关。夏季食用菌类和辣椒、花叶、水生根茎类正相关；辣椒类和食用菌类正相关；茄类和水生根茎类负相关；水生根茎类和花叶、食用菌类正相关，与茄类负相关；花菜类和花叶类正相关；花叶类和花菜、水生根茎、食用菌类正相关。秋季食用菌类和花叶、花菜、辣椒类正相关；辣椒类和花叶、食用菌类正相关，与茄类负相关；茄类与辣椒类负相关；花菜类和花叶类、食用菌类正相关；花叶类与花菜类、食用菌类、辣椒类正相关。冬季食用菌类和花叶类、花菜类、水生根茎类正相关；辣椒类和花菜类正相关；茄类和花菜类正相关；水生根茎类和食用菌、花菜、花叶类正相关；花菜类和花叶、水生根茎、茄类、辣椒类、食用菌类正相关；花叶类和花菜类、水生根茎类、食用菌类正相关[3]。

3.4. 模型建立

我们发现各品类的售价不仅与其本身的批发价和销售量有关，还和与该品类相关性较强的其他品类的批发价和销售量有关。因此，我们准备建立各品类的售价与该品类及其相关品类的批发价和销售量的关系函数。

首先，统计出3年内每品类每周的平均售价和3年内每品类每周的总销售量。然后计算出各品类的售价与该品类及其相关品类的批发价和销售量的关系函数。对同一品种蔬菜的销售量，由于不仅可能受到本类产品的批发价格和销售价格的影响，还可能受到其他品类产品的批发价格和销售价格的影响。需要建立各品类蔬菜成本加成定价关于每一品类蔬菜的批发价格和销售量的预测模型。采用偏最小二乘回归(PLSR)模型对各类品种成本加成定价进行预测[4] [5] [6]。结果如下：

1) 花菜类

算法：线性回归(最小二乘法) $R^2 = 0.801$

变量：自变量 X: {花菜类平均批发价格, 花叶类平均批发价格, 花菜类周销售量, 花叶类周销售量};

因变量 Y: {花菜类周平均销售价}

模型的公式如下：

$$y = 1.565 + 1.324 * \text{花菜类平均批发价格} + 0.005 * \text{花叶类平均批发价格} - 0.001 * \text{花菜类周销售量} + 0.0 * \text{花叶类周销售量}。$$

图8展示了该模型的原始数据图、模型预测值。可以发现拟合结果较好。

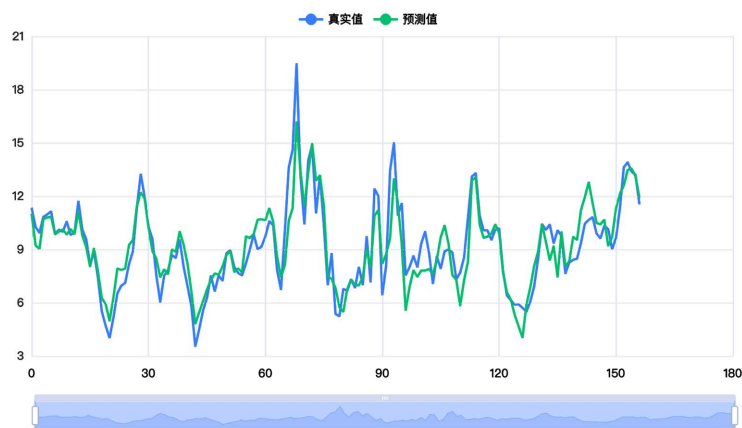


Figure 8. Fit effect diagram

图8. 拟合效果图

2) 花叶类:

算法: 线性回归(最小二乘法) $R^2 = 0.892$

变量: 自变量 X: {花叶类平均批发价格, 花菜类平均批发价格, 水生根茎类平均批发价格, 食用菌类平均批发价格, 花叶类周销售量, 花菜类周销售量, 水生根茎类周销售量, 食用菌类周销售量};

因变量 Y: {花叶类周平均销售价}

模型的公式如下:

$$y = 1.654 + 1.347 * \text{花叶类平均批发价格} - 0.006 * \text{花菜类平均批发价格} - 0.038 * \text{水生根茎类平均批发价格} + 0.046 * \text{食用菌类平均批发价格} + 0.0 * \text{花叶类周销售量} + 0.0 * \text{花菜类周销售量} - 0.0 * \text{水生根茎类周销售量} - 0.001 * \text{食用菌类周销售量}.$$

图 9 展示了该模型的原始数据图、模型预测值。可以发现拟合结果较好。

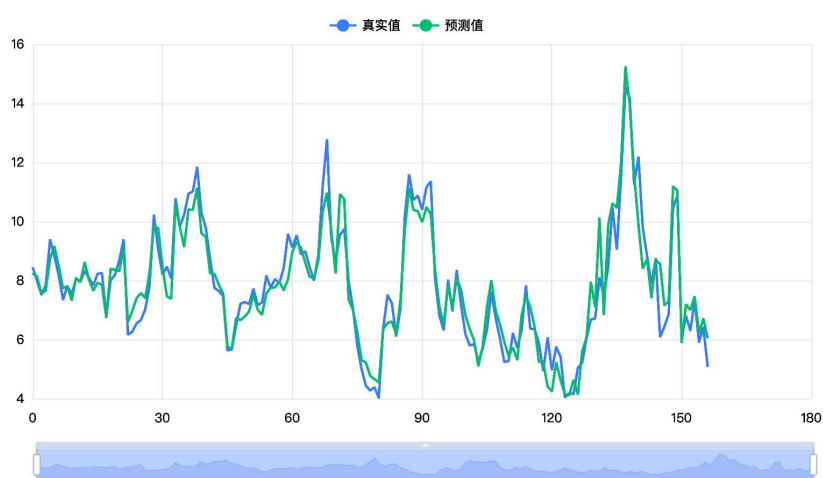


Figure 9. Fit effect diagram

图 9. 拟合效果图

3) 食用菌类:

算法: 线性回归(最小二乘法) $R^2 = 0.745$

变量: 自变量 X: {食用菌平均批发价格, 辣椒类平均批发价格, 花叶类周销售量, 花叶类平均批发价格, 水生根茎类平均批发价格, 水生根茎类周销售量, 食用菌类周销售量, 辣椒类周销售量};

因变量 Y: {食用菌周平均销售价}

模型的公式如下:

$$y = -0.672 + 1.396 * \text{食用菌平均批发价格} + 0.032 * \text{辣椒类平均批发价格} + 0.002 * \text{花叶类周销售量} - 0.012 * \text{花叶类平均批发价格} - 0.037 * \text{水生根茎类平均批发价格} - 0.005 * \text{水生根茎类周销售量} + 0.001 * \text{食用菌类周销售量} - 0.001 * \text{辣椒类周销售量}.$$

图 10 展示了该模型的原始数据图、模型预测值。可以发现拟合结果较好。

4) 辣椒类:

变量: 自变量 X: {辣椒类平均批发价格, 辣椒类周销售量, 食用菌平均批发价格, 食用菌类周销售量};

因变量 Y: {辣椒周平均销售价}

算法: 线性回归(最小二乘法) $R^2 = 0.882$

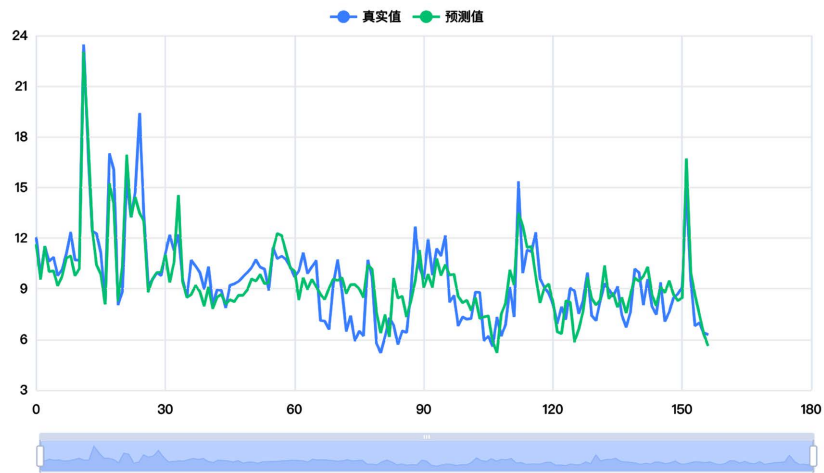


Figure 10. Fit effect diagram
图 10. 拟合效果图

模型的公式如下:

$$y = 0.629 + 1.364 * \text{辣椒类平均批发价格} - 0.001 * \text{辣椒类周销售量} + 0.158 * \text{食用菌平均批发价格} + 0.002 * \text{食用菌类周销售量}。$$

图 11 展示了该模型的原始数据图、模型预测值。可以发现拟合结果较好。

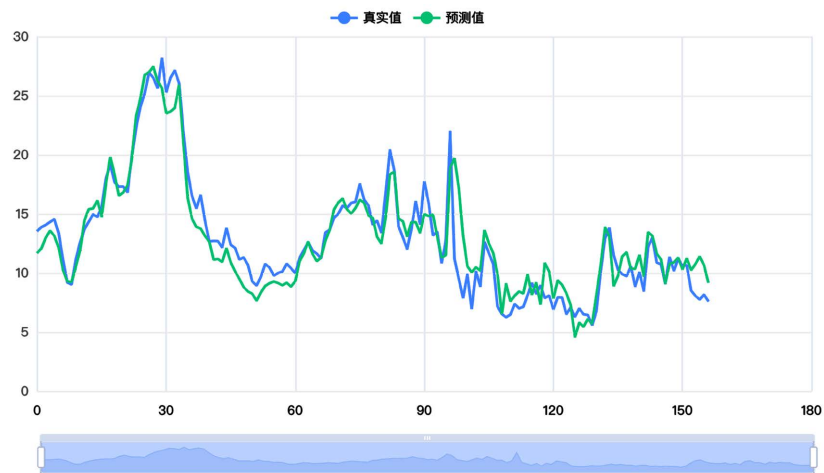


Figure 11. Fit effect diagram
图 11. 拟合效果图

5) 茄类:

变量: 自变量 X: {茄类平均批发价格, 水生根茎类平均批发价格, 茄类周销售量, 水生根茎类周销售量};

因变量 Y: {茄类周平均销售价}

算法: 线性回归(最小二乘法) $R^2 = 0.853$

模型的公式如下:

$$y = 1.219 + 1.306 * \text{茄类平均批发价格} + 0.049 * \text{水生根茎类平均批发价格} - 0.001 * \text{茄类周销售量} +$$

0.001*水生根茎类周销售量。

图 12 展示了该模型的原始数据图、模型预测值。可以发现拟合结果较好。



Figure 12. Fit effect diagram

图 12. 拟合效果图

6) 水生根茎类:

变量: 自变量 X: {食用菌类平均批发价格, 花叶类平均批发价格, 水生根茎类平均批发价格, 茄类平均批发价格, 水生根茎类周销售量, 花叶类周销售量, 食用菌类周销售量, 茄类周销售量};

因变量 Y: {水生根茎类周平均销售价}

算法: 线性回归(最小二乘法) $R^2 = 0.744$

模型的公式如下:

$$y = 6.827 - 0.151 * \text{食用菌类平均批发价格} - 0.002 * \text{花叶类平均批发价格} + 0.838 * \text{水生根茎类平均批发价格} - 0.241 * \text{茄类平均批发价格} - 0.003 * \text{水生根茎类周销售量} + 0.0 * \text{花叶类周销售量} + 0.0 * \text{食用菌类周销售量} + 0.006 * \text{茄类周销售量}.$$

图 13 展示了该模型的原始数据图、模型预测值。可以发现拟合结果较好。



Figure 13. Fit effect diagram

图 13. 拟合效果图

4. 蔬菜补货与定价决策模型构建

4.1. 偏最小二乘法构建销售量预测模型

我们发现，各品类的销售量对于周数来说有周期关系，与其本身和相关性强的品类的批发价有线性关系。此外，我们还发现了各品类的批发价近与日期有关。因此，统计每年各品类每周的批发价，使用一周的平均批发价代替本周的批发价。

图 14 显示过去 3 年内以周为单位，每个品类的批发价存在周期关系，周期为一年。

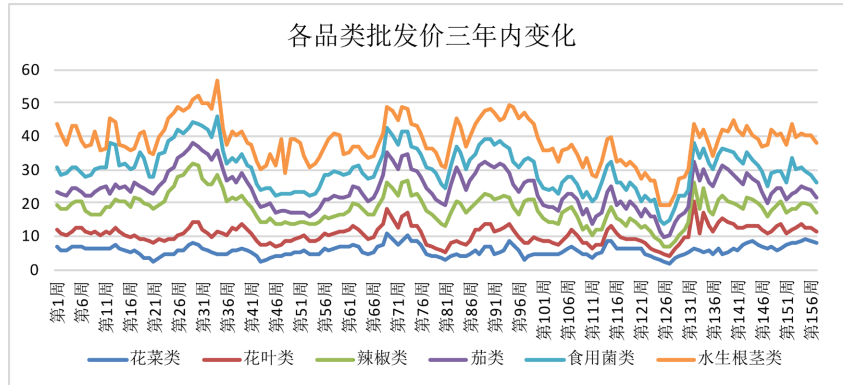


Figure 14. Wholesale price line chart

图 14. 批发价格折线图

采用过去 3 年的平均值预测今年 7 月 1 日到 7 日的各品类批发价。并通过过去 3 年的统计数据建立各品类的周销量与包括自己在内的 6 大品类的批发价、周数的预测函数，结果如下：

算法：偏最小二乘法(PLSR)

变量：变量 Y:{食用菌类周销售量，水生根茎类周销售量，辣椒类周销售量，茄类周销售量，花菜类周销售量，花叶类周销售量}；变量 X:{周数，食用菌平均批发价格，辣椒类平均批发价格，花菜类平均批发价格，花叶类平均批发价格，茄类平均批发价格，水生根茎类平均批发价格}

表 3 是经过偏最小二乘法模型计算后得到的模型系数。

Table 3. Model coefficient result table

表 3. 模型系数结果表

	食用菌类周销售量	水生根茎类周销售量	辣椒类周销售量	茄类周销售量	花菜类周销售量	花叶类周销售量	食用菌类周销售量	水生根茎类周销售量	辣椒类周销售量	茄类周销售量	花菜类周销售量	花叶类周销售量
常数	221.443	185.27	-470.701	-131.655	89.953	997.274	0	0	0	0	0	0
周数	-1.169	-2.069	27.247	-0.046	4.059	6.155	-0.016	-0.054	0.277	-0.002	0.158	0.054
食用菌平均批发价格	47.073	-8.772	-11.863	-5.228	8.619	49.574	0.243	-0.088	-0.046	-0.101	0.128	0.166
辣椒类平均批发价格	-0.057	5.377	-33.835	2.149	5.188	5.447	-0.001	0.181	-0.441	0.138	0.258	0.061
花菜类平均批发价格	-2.094	0.87	-16.787	-10.773	-14.343	-44.202	-0.012	0.01	-0.074	-0.233	-0.24	-0.166
花叶类平均批发价格	57.305	3.428	51.125	3.347	6.573	144.666	0.368	0.043	0.247	0.08	0.121	0.601

续表

茄类平均 批发价格	8.705	17.507	57.289	25.601	15.001	-75.359	0.047	0.186	0.235	0.52	0.236	-0.266
水生根茎类 平均批发 价格	-24.002	19.148	34.427	19.746	-15.137	-38.456	-0.113	0.175	0.122	0.346	-0.205	-0.117

模型的标准化公式为:

食用菌类周销售量 = 221.443 - 1.169*周数 + 47.073*食用菌平均批发价格 - 0.057*辣椒类平均批发价格 - 2.094*花菜类平均批发价格 + 57.305*花叶类平均批发价格 + 8.705*茄类平均批发价格 - 24.002*水生根茎类平均批发价格

水生根茎类周销售量 = 185.27 - 2.069*周数 - 8.772*食用菌平均批发价格 + 5.377*辣椒类平均批发价格 + 0.87*花菜类平均批发价格 + 3.428*花叶类平均批发价格 + 17.507*茄类平均批发价格 + 19.148*水生根茎类平均批发价格

辣椒类周销售量 = - 470.701 + 27.247*周数 - 11.863*食用菌平均批发价格 - 33.835*辣椒类平均批发价格 - 16.787*花菜类平均批发价格 + 51.125*花叶类平均批发价格 + 57.289*茄类平均批发价格 + 34.427*水生根茎类平均批发价格

茄类周销售量 = - 131.655 - 0.046*周数 - 5.228*食用菌平均批发价格 + 2.149*辣椒类平均批发价格 - 10.773*花菜类平均批发价格 + 3.347*花叶类平均批发价格 + 25.601*茄类平均批发价格 + 19.746*水生根茎类平均批发价格

花菜类周销售量 = 89.953 + 4.059*周数 + 8.619*食用菌平均批发价格 + 5.188*辣椒类平均批发价格 - 14.343*花菜类平均批发价格 + 6.573*花叶类平均批发价格 + 15.001*茄类平均批发价格 - 15.137*水生根茎类平均批发价格

花叶类周销售量 = 997.274 + 6.155*周数 + 49.574*食用菌平均批发价格 + 5.447*辣椒类平均批发价格 - 44.202*花菜类平均批发价格 + 144.666*花叶类平均批发价格 - 75.359*茄类平均批发价格 - 38.456*水生根茎类平均批发价格

在以上模型中输入周数为 26.5, 并输入采用过去 3 年的平均值预测出的今年 7 月 1 日到 7 日的各品类预测批发价, 便可预测出各品类的 7 月 1~7 日的预测周销售量。将各品类的 7 月 1~7 日的预测周销售量再次输入问题 2.1 得出的各品类的定价与销售量的关系模型中, 得出各品类的预测定价。由此来制定未来一周(2023 年 7 月 1~7 日)的日补货总量和定价策略(补货量 = 销售量/(1 - 损耗率))。

图 15 以图表形式清晰的展示了未来一周(2023 年 7 月 1~7 日)的日补货总量和定价策略。

品类	平均批发价(元)	预计周销售量(kg)	预计定价(元)	补货量(kg)
食用菌类	6.759237	629.3987972	9.411433325	686.36
水生根茎类	6.281654	342.2171749	9.104651108	377.89
辣椒类	6.096706	629.9881134	10.64167591	684.76
茄类	5.689148	73.35984661	9.225447043	78.67
花菜类	5.092381	231.4165454	8.185278509	254.78
花叶类	4.076522	1222.868867	7.816133121	1376.48

Figure 15. Sales price forecast table

图 15. 销售价格预测表

4.2. 构建销售价格模型

首先，选出 2023 年 6 月 24 日~30 日的出售单品，为了缩小可售商品种类，我们删去其中在每年 7 月 1 日统计单品批发价时从没有出现的单品，同时删去其中平均日销售量小于 1.5 kg 的单品，最终筛选出以下单品，作为 7 月 1 日预计出售商品，共 29 种。其次，使用聚类分析(K-Means)以上单品历年 6~8 月每天的销售量，由此得出 6 大类。

图 16 显示聚类汇总图，可直观的看出各组份占比。

图 17 显示各聚类下的组成成分。

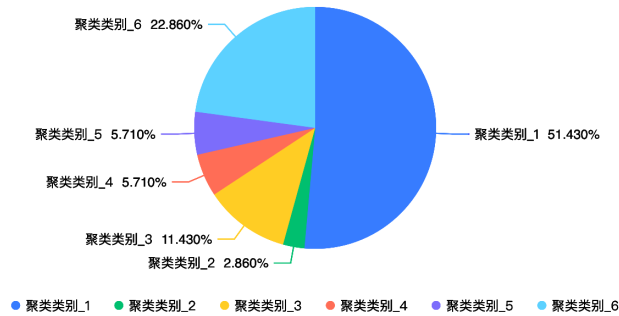


Figure 16. Cluster summary chart

图 16. 聚类汇总图

聚类类别	单品名称	单品编号	品类
类别1	西兰花	102900005116714	花菜类
类别1	菠菜	102900005118817	花叶类
类别1	菜心	102900005115908	花叶类
类别1	红薯尖	102900005119975	花叶类
类别1	云南生菜	102900005115779	花叶类
类别1	云南生菜(份)	102900011030059	花叶类
类别1	云南油麦菜	102900005115984	花叶类
类别1	竹叶菜	102900005115786	花叶类
类别1	金针菇(盒)	106949711300259	食用菌
类别1	双孢菇(盒)	102900011034330	食用菌
类别1	高瓜(1)	102900005118824	水生根茎类
类别1	菱角	102900011001691	水生根茎类
类别2	木耳菜	102900005115946	花叶类
类别3	苋菜	102900005115762	花叶类
类别3	娃娃菜	102900005118831	花叶类
类别3	紫茄子(2)	102900005116257	茄类
类别3	小米椒(份)	102900011031100	辣椒类
类别4	青茄子(1)	102900005116509	茄类
类别4	云南油麦菜	102900005115984	花叶类
类别5	净藕(1)	102900005116899	水生根茎类
类别5	奶白菜	102900011008164	花叶类
类别6	洪湖藕带	102900051000944	水生根茎类
类别6	上海青	102900005115823	花叶类
类别6	长线茄	102900011022764	茄类
类别6	小青菜(1)	102900011023464	花叶类
类别6	枝江青梗散花	102900011034026	花菜类
类别6	螺丝椒	102900011000328	辣椒类
类别6	菠菜(份)	102900011030110	花叶类
类别6	菜心	102900005115908	花叶类

Figure 17. Cluster result graph

图 17. 聚类结果图

由于上述各聚类中的每个单品都属于各大品类，所以我们认为这些单品的周销售量与定价规律近似的符合 4.1 所建立的预测模型得出的预测函数。为了使用 4.1 结果中得出的单品的周销售量与定价预测函数，我们做出以下规定：

一、预测 a 单品的周销售量与定价时，a 在过去 3 年的 7 月 1 日数据的平均值代替 a 所属公式中 A 品类的数据。若 a、b 属于 A 品类，依然只使用 a 的数据代替 A 品类数据。

二、预测 a 单品的周销售量与定价时，需要 b、c 各单品的数据。a 属于 A 品类；b、c 属于 B 品类。若 b、c 和 a 属于同一聚类中的单品，用 b、c 数据的平均值代替公式中 B 品类的数据。

三、预测 a 单品的周销售量与定价时，若需要 C 品类数据，而该聚合大类中并没有出现属于 C 品类的单品，则使用问题 2 得出的各大类的数据。

图 18 显示经过预测计算得到的结果，为最佳定价采购策略。

单品名称	预计批发价	7月1日预计定价	7月1日销售量	7月1日损耗率	7月1日补货量	7月1日预测利润
西兰花	7.2	10.73	55.05	9.26	60.66	153.87
菠菜	7.98	11.79	39.24	18.51	48.16	78.39
菜心	3.7	6.02	111.02	13.7	128.65	192.36
红薯尖	2.53	4.45	86.84	8.42	94.83	146.54
云南生菜	5.35	8.25	145.12	15.25	171.23	281.15
云南生菜(份)	2.06	3.82	77.13	9.43	85.16	119.2
云南油麦菜	3.64	5.94	109.78	12.81	125.91	193.79
竹叶菜	2.4	4.27	84.15	13.62	97.42	125.52
金针菇(盒)	1.9	0.23	49.23	0.45	49.45	-82.64
双孢菇(盒)	3.63	2.73	60.86	0.2	60.99	-55.22
高瓜(1)	6	9.37	53.39	29.25	75.47	47.49
菱角	10.36	12.77	65.32	9.61	72.27	85.48
木耳菜	3.19	5.36	154.27	7.61	166.97	294.23
苋菜	2.49	4.42	173.29	18.52	212.68	236.43
娃娃菜	3.74	6.1	199.13	2.48	204.19	451.81
紫茄子(2)	5.17	8.56	8.45	6.07	8.99	25.83
小米椒(份)	2.36	5.51	94.89	9.43	104.77	275.63
青茄子(1)	4.39	7.54	8.25	5.01	8.69	24.12
云南油麦菜	3.64	5.97	194.38	12.81	222.94	348.83
净藕(1)	10.11	25.49	64.65	5.54	68.45	956.37
奶白菜	1.8	3.35	135.32	15.68	160.49	163.83
洪湖藕带	22.81	24.64	105.7	24.05	139.17	-569.87
上海青	3.94	5.99	150.53	14.43	175.91	207.91
长线茄	10.34	8.53	73.62	6.9	79.07	-189.84
小青菜(1)	3.16	4.93	134.41	10.33	149.89	189.64
枝江青梗散花	5.13	7.33	6.5	9.43	7.18	10.82
螺丝椒	6.15	10.49	122.41	10.18	136.29	445.63
菠菜(份)	3.21	5	135.44	9.43	149.55	197.48
菜心	3.7	5.66	144.41	13.7	167.34	198.56

Figure 18. Pricing and procurement strategy

图 18. 定价与采购策略

5. 补充影响因素

我们认为无论是销售量和批发价格都会受到天气气温和降水量的影响。天气因素，例如天气状况、风力等级、空气质量指数等隐变量也会影响运输成本和消费者消费意愿从而对销量产生间接的影响。天

气气温和降水量会影响农产品的生长质量和保质期, 气温还会改变顾客的食欲和选择倾向[7]。

同时, 电商平台的好评率也是重要参考依据之一。消费者为减少购物过程中的不确定性, 往往会参考电商平台提供的在线评论。由于蔬菜类农产品本身保质期短, 易在运输或储存过程中受到损坏, 因此与其他商品相比, 在线评论在降低消费者感知风险、提高消费者购买意愿方面发挥了更重要的作用[8]。

6. 结论

本文基于销量和批发价数据分析对蔬菜类商品对预期销量和成本加成定价进行研究, 并针对不同时节不同品类的蔬菜进行销售策略的制定。结合蔬菜类产品具有时令性的销售特性, 按季节划分数据, 研究不同季节下的产品的相关性和销售量预测。进行线性相关性检验后, 采用 Spearman 与 Pearson 相关性检测通过分析各蔬菜品类的月平均销量, 分析每个季节下的各蔬菜品类间的相关性。同时采用线性回归模型对各蔬菜品类的周销售量进行预测。建立各品类蔬菜成本加成定价关于每一品类蔬菜的批发价格和销售量的预测模型。采用偏最小二乘回归(PLSR)模型对各类品种成本加成定价进行预测。

参考文献

- [1] 林晓东. 成本加成定价法的弊端分析及其改革[J]. 引进与咨询, 2005(11): 13-14.
- [2] 李伟伟, 易平涛, 李玲玉. 综合评价中异常值的识别及无量纲化处理方法[J]. 运筹与管理, 2018, 27(4): 173-178.
- [3] Saroj, K. (2016) Review: Study on Simple K Mean and Modified K Mean Clustering Technique. *International Journal of Computer Science Engineering and Technology*, 6, 279-281.
- [4] Scientific Platform Serving for Statistics Professional 2021. SPSSPRO. Version 1.0.11. <https://www.spsspro.com>
- [5] 徐维超. 相关系数研究综述[J]. 广东工业大学学报, 2012, 29(3): 12-17.
- [6] 梁北辰, 戴景民. 偏最小二乘法在系统故障诊断中的应用[J]. 哈尔滨工业大学学报, 2020, 52(3): 156-164. <https://doi.org/10.11918/201805149>
- [7] 林东平. 考虑天气因素的生鲜农产品销量预测模型研究[D]: [硕士学位论文]. 大连: 大连理工大学, 2020. <https://doi.org/10.26991/d.cnki.gdlu.2020.002773>
- [8] 戴文旭. 京东在线评论对生鲜农产品销量影响的实证研究[D]: [硕士学位论文]. 烟台: 烟台大学, 2023. <https://doi.org/10.27437/d.cnki.gytd.2023.000711>