

# 2023 年高教社杯全国大学生数学建模竞赛题目

(请先阅读“全国大学生数学建模竞赛论文格式规范”)

## C 题蔬菜类商品的自动定价与补货决策

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，商超通常会根据各商品的历史销售和需求情况每天进行补货。

由于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，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。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。

背景分析：前两段中，第一段给出了背景，没有什么可用的。第二段中，提出：“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。蔬菜的定价一般采用“成本加成定价”方法”，这就意味着，我们后续可能需要对单品和进货价格做预测，预测常用的是机器学习算法进行预测，或者如果这道题目和时间序列有关，那就需要用到时间序列算法，比如 ARIMA 模型。

在预测模型结束后，根据背景分析，我们下一步可能要做补货决策模型和定价决策模型（这里大家可以看看之前的一道国赛题，机场出租车问题，也是需要很多个决策模型，大家可以从那个的优秀论文里找到一些思路），这里的模型需要兼顾过去的历史数据，以及一些动态变化的数据，比如预测未来的销量，品相对补货情况的影响等等。

题目提到了，可靠的市场需求分析，对补货决策和定价决策尤为重要。所以要有一个小节对市场需求进行详细分析。这里题目也给明了，是要对销量与时间做关联分析，从而寻找到时间对于销量的影响，这里其实也可以用拟合来做，多项式拟合等各种拟合的方式，具体后面看到题目再分析。接着呢，又写了商超销售空间的限制，所以还要对销售进行组合。如果是空间组合问题，这里我想到的是装箱问题，比如二维、三维装箱问题，不过还是需要看题目具体怎么给，可能并不一定这么复杂。

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。请根据附件和实际情况建立数学模型解决以下问题：

分析：下面我们先对这几个附件进行分析，题目在分析完附件后再说。

附件 1 给出了某商超经销的 6 个蔬菜品类的商品信息，具体的：

1	102900005115823	上海青	1011010101	花叶类		
---	-----------------	-----	------------	-----	--	--

可以看出，有两个信息和他们对应的索引，这里可以用类别来做聚类分析，就看后续的要求了。聚类分析可以用 kmeans 或者 DBSCAN 算法。或者简单的做统计分析也可以，看题目要求。

2020-07-01	09:15:07.924	102900005117056	0.396	7.60	销售	否
2020-07-01	09:17:27.295	102900005115960	0.849	3.20	销售	否

这里，要做如下几件事：

1 将销售日期转换为时间格式 (python 的话)

2 对于扫码销售时间, 应该改变其格式, 推荐用正则表达式提取需要的数字, 将其转换为数值型数据。我看了下销售类型和是否打折销售, 这两个指标需要量化, 推荐用 0 或 1 来量化即可。

附件 3 是关于批发价格的。这个表可以新增三列, 根据单品编号来后面追加 3 列, 也就是和附件 1 对应起来, 以便后续数据处理。这里可能需要对同一单品做关于时间的分析, 也就是随着时间, 这个单品的批发价格情况如何变化, 可以针对情况做统计分析以及可视化

附件 4 可以和附件 1 合并, 在附件 1 里追加一列, 这样, 附件 3 又可以多追加一列了。

对于题目的数据处理工作, 很重要的一点就是要对多个表进行合并, 以便后续分析了。

下面我们再介绍一些你们可以做的其他 EDA (探索性数据分析) 工作。

建议大家使用一些可视化方法, 可以使用常见的 EDA 可视化方法:

- 直方图和密度图: 展示数值变量的分布情况。
- 散点图: 展示两个连续变量之间的关系。
- 箱线图: 展示数值变量的分布情况和异常值。
- 条形图和饼图: 展示分类变量的分布情况。
- 折线图: 展示随时间或顺序变化的趋势。
- 热力图: 展示不同变量之间的相关性。
- 散点矩阵图: 展示多个变量之间的散点图矩阵。
- 地理图: 展示地理位置数据和空间分布信息。

分析完附件, 就可以来做题了。

**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, 请分析蔬菜各品类及单品销售量的分布规律及相互关系。

问题一分析: 前半句告诉我们要用怎么样的方法, 这道题目我的建议是针对蔬菜的品类做聚类分析 (比如 kmeans 等聚类算法), 针对单品销量做时间序列分析, 来观察世界对销量的影响。

相互的关系, 可以用拟合来做, 比如:

1 线性回归: 使用线性回归模型来拟合销售量与其他特征之间的关系, 例如时间、价格等。可以分别对不同品类或单品进行回归分析, 得到各自的拟合曲线。

2 非线性回归: 如果线性模型无法很好地拟合数据, 则可以尝试使用非线性回归模型, 如多项式回归、指数函数拟合等。这些模型能够更灵活地适应可能存在的非线性关系。

3 时间序列模型: 对于销售数据中具有时间信息的情况, 可以使用时间序列模型, 如 ARIMA 模型或季节性模型, 来拟合销售量的变化趋势和周期性。这些模型可以帮助揭示销售量随时间的演变规律。

4 面板数据模型: 如果想同时考虑品类和单品之间的关联关系, 可以使用面板数据模型, 如固定效应模型或随机效应模型。这些模型考虑了跨个体和跨时间的相关性, 可以更准确地估计各品类或单品之间的关系。

这样做, 可以更好的获得各品类或单品之间销售量的关联性及其规律。这些模型可以帮助商超更好地理解不同品类或单品之间的相互影响, 并为补货计划和定价决策提供指导。

**问题 2** 考虑商超以品类为单位做补货计划, 请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。

2-4 问 8 号更新

**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，请给出你们的意见和理由。

附件 1 6 个蔬菜品类的商品信息

附件 2 销售流水明细数据

附件 3 蔬菜类商品的批发价格

附件 4 蔬菜类商品的近期损耗率

**注**（1）附件 1 中，部分单品名称包含的数字编号表示不同的供应来源。

（2）附件 4 中的损耗率反映了近期商品的损耗情况，通过近期盘点周期的数据计算得到。