

# 基于机器学习与多源数据融合的宜昌市就业状态研究

## 摘要

关键字：Fisher 精确检验 Cramér's V 系数 XGBoost SHAP 评估

# 一、问题背景与重述

## 1.1 问题背景

根据乡村的实际情况，充分利用有限的耕地资源，因地制宜，发展有机种植产业，对乡村经济的可持续发展具有重要的现实意义。选择适宜的农作物，优化种植策略，有利于方便田间管理，提高生产效率，减少各种不确定因素可能造成的种植风险。

现一华北山区乡村，有露天耕地 1201 亩，分散为 34 个大小不同的地块，包括平旱地、梯田、山坡地和水浇地 4 种耕地类型。其中前三种耕地适宜每年种植一季粮食类作物；水浇地适宜每年种植一季水稻或两季蔬菜。另外还有 16 个普通大棚和 4 个智慧大棚，每个大棚耕地面积为 0.6 亩。普通大棚适宜每年种植一季蔬菜和一季食用菌，智慧大棚适宜每年种植两季蔬菜。每种作物在同一地块都不能连续重茬种植，从 2023 年开始要求每个地块（含大棚）的所有土地三年内至少种植一次豆类作物。

## 1.2 问题重述

**问题一：**假定各种农作物未来的预期销售量、种植成本、亩产量和销售价格相对于 2023 年保持稳定，每季种植的农作物在当季销售。如果某种作物每季的总产量超过相应的预期销售量，超过部分不能正常销售。请针对超过部分的两种情况，分别给出该乡村 2024–2030 年农作物的最优种植方案，将结果分别填入 `result1_1.xlsx` 和 `result1_2.xlsx` 中。

**问题二：**根据经验，小麦和玉米未来的预期销售量有增长的趋势，平均年增长率介于 5%–10% 之间，其他农作物未来每年的预期销售量相对于 2023 年大约有  $\pm 5\%$  的变化。农作物的亩产量往往会受气候等因素的影响，每年会有  $\pm 10\%$  的变化。因受市场条件影响，农作物的种植成本平均每年增长 5% 左右。粮食类作物的销售价格基本稳定；蔬菜类作物的销售价格有增长的趋势，平均每年增长 5% 左右。食用菌的销售价格稳中有降，大约每年可下降 1%–5%，特别是羊肚菌的销售价格每年下降幅度为 5%。请给出该乡村 2024–2030 年农作物的最优种植方案，将结果填入 `result2.xlsx` 中。

**问题三：**在现实生活中，各种农作物之间可能存在一定的可替代性和互补性，预期销售量与销售价格、种植成本之间也存在一定的相关性。请在问题 2 的基础上综合考虑相关因素，给出该乡村 2024–2030 年农作物的最优种植策略，通过模拟数据进行求解，并与问题 2 的结果作比较分析。

## 二、问题分析

**问题一：**题目给出了条件为各种农作物未来的预期销售量、种植成本、亩产量和销售价格相对于 2023 年保持稳定，每季种植的农作物在当季销售。这是一个确定性多周期混合整数线性规划问题。首先，根据题目的表述，我们将 2024–2030 年所有地块农产品的总收入作为优化目标。针对产量超出预期销售量的部分作物，题目给出了两种情形，我们可以据此分别给出两种情形的目标函数。另外根据耕地类型、作物类型、重茬种植、豆类作物等限制条件，我们可以得到约束条件，再调用 python 中的 CBC 求解器进行求解。

**问题二：**在问题一的基础上，问题二引入了预期销售量、亩产量、种植成本和销售价格的变动。引入不确定性后，决策目标从“追求最高利润”转变为“规避风险，保证最坏情况下的收益”。因此，我们采用鲁棒优化方法，基于不确定参数的波动区间重新设立目标函数和鲁棒约束，继续采用问题一中的 CBC 求解器进行求解。

**问题三：**本问在问题二的基础上要求进一步考虑农作物之间的互补性和可替代性，以及预期销售量、销售价格、种植成本之间的相关性，通过模拟数据的方式进行求解。这个问题无法用传统的数学规划直接建模，因此我们采用仿真优化的思想，首先建立一个能够模拟预期销售量、销售价格、种植成本之间的动态关系的仿真器，然后运用遗传算法对该仿真器进行智能“驾驶”，在复杂的决策空间中搜索最优的长期种植策略。

## 三、模型假设

1. 假设 2023 年的各类农作物刚好满足市场需求，即产量刚好等于预期销售量。
2. 假设附件中的数据真实可靠，可以作为模型的基本参数。
3. 假设农作物的生长符合对应的生长规律。

## 四、符号说明

表 1 符号说明

符号	说明
$i, I$	地块的索引与集合
$j, J$	作物的索引与集合
$k, K$	季节的索引与集合
$y, Y$	年份的索引与集合（2024-2030）
$J_{\text{bean}}$	所有豆类作物的集合
$A_i$	地块 $i$ 的可用面积
$C_j$	作物 $j$ 的单位面积种植成本
$P_j$	作物 $j$ 的单位重量销售价格
$\text{Yield}_j$	作物 $j$ 的单位面积产量
$\text{Demand}_j$	作物 $j$ 的每季预期市场销售量
$\text{Past}_{ij}$	地块 $i$ 在 2023 年是否种植了作物 $j$ 的 0-1 参数
$S_{ijk}$	地块 $i$ 在季节 $k$ 是否适宜种植作物 $j$ 的 0-1 参数
$A_{\min}$	单个地块上允许种植某种作物的最小面积阈值
$N_j$	作物 $j$ 在单季内允许种植的最大分散地块数量
$M$	大 $M$ 方法中的一个足够大的正数

## 五、问题一：确定性环境下的多周期优化模型

问题一旨在为该乡村规划 2024 年至 2030 年共七年的农作物种植策略。其核心目标是在满足一系列种植规则的前提下，实现长期经济效益的最大化。题目明确指出，此问中未来的预期销售量、种植成本、亩产量和销售价格均相对于 2023 年保持稳定，因此该问题是一个确定性的优化问题。决策的本质是在多个时间周期、多个地理单元上，为多种作物分配种植面积。由于决策中包含“是否种植”这一 0-1 选择，以及“种植多少”这一连续量，因此该问题是典型的多周期混合整数线性规划问题。所以我们构建一个以七年总利润最大化为目标的 MILP 模型，针对题目中“产量超出预期”的两种情形，分别构建目标函数，并在同一套约束体系下进行求解。

### 5.1 决策变量与辅助变量的定义

为构建优化模型，我们首先定义了三组核心决策变量，它们共同构成了完整的种植方案。

- $x_{ijk_y}$ : 模型的核心连续决策变量, 代表在年份  $y$  的季节  $k$ , 于地块  $i$  安排种植作物  $j$  的具体面积 (单位: 亩)。
- $u_{ijk_y}$ : 0-1 二进制决策变量, 作为逻辑开关。当且仅当  $x_{ijk_y}$  的值大于零时,  $u_{ijk_y}$  取值为 1, 否则为 0。
- $z_{ij_y}$ : 年度层面的 0-1 二进制决策变量。若在年份  $y$  的任何一个季节, 地块  $i$  安排了作物  $j$  的种植, 则  $z_{ij_y}$  取值为 1, 否则为 0。

此外, 引入两组辅助变量:

- $Sales_{jk_y}$ : 作物  $j$  在年份  $y$  季节  $k$  的总产量中, 按正常市场价格销售的数量 (单位: 千克)。
- $OverSales_{jk_y}$ : 作物  $j$  在年份  $y$  季节  $k$  的总产量中, 超出预期市场需求后, 进行降价处理的销售数量 (单位: 千克)。

## 5.2 目标函数的确定

模型的根本目标是最大化在整个七年规划期内的累计净利润。根据题目设定的两种不同情景, 分别构建对应的目标函数。

### 情况一: 超产部分滞销浪费

在此情景下, 超出市场预期的产量无法创造价值, 因此总收益仅由正常销售部分贡献。目标函数为:

$$\max \sum_{y \in Y} \sum_{k \in K} \left( \sum_{j \in J} P_j \cdot Sales_{jk_y} - \sum_{j \in J} \sum_{i \in I} C_j \cdot x_{ijk_y} \right) \quad (1)$$

### 情况二: 超产部分降价出售

在此情景下, 超产部分仍能以 50% 的折扣价售出, 为乡村带来额外收益。目标函数为:

$$\max \sum_{y \in Y} \sum_{k \in K} \left( \sum_{j \in J} (P_j \cdot Sales_{jk_y} + 0.5 \cdot P_j \cdot OverSales_{jk_y}) - \sum_{j \in J} \sum_{i \in I} C_j \cdot x_{ijk_y} \right) \quad (2)$$

## 5.3 约束条件的设立

为了确保模型生成的种植方案在现实中是可行的、在农艺上是科学的, 并且在经济上是可持续的, 我们构建了一个包含七个核心模块的严密约束体系。该体系将农业生产的各个环节, 从土地资源的分配到市场销售的实现, 再到长期的生态维护, 均纳入了数学模型的考量范围之内。

(1) 总产量约束

$$\text{情况一: } Yield_j \cdot \sum_{i \in I} x_{ijk y} \geq Sales_{jk y} \quad \forall j, k, y \quad (3)$$

$$\text{情况二: } Yield_j \cdot \sum_{i \in I} x_{ijk y} = Sales_{jk y} + OverSales_{jk y} \quad \forall j, k, y \quad (4)$$

该约束是连接种植规模与市场销售的核心纽带。它规定了在任意年份的任意季节, 特定作物  $j$  的总产量必须等于该作物的亩产量  $Yield_j$  与总种植面积  $\sum_i x_{ijk y}$  的乘积。根据题目设定的两种情景, 此约束有两种形式。在情况一(超产部分滞销浪费)中, 总产量只需满足正常销售的需求。在情况二(超产部分降价出售)中, 总产量被精确地分解为正常销售量  $Sales_{jk y}$  和降价销售量  $OverSales_{jk y}$  两部分之和, 确保了所有产出都被核算。

(2) 市场销售约束

$$Sales_{jk y} \leq Demand_j \quad \forall j, k, y \quad (5)$$

此约束为每种作物的正常价格销售量设置了上限。它确保了模型在追求利润最大化时, 以正常价格出售的作物数量  $Sales_{jk y}$  不会超过其预期的市场需求量  $Demand_j$ 。这反映了市场的吸纳能力, 是保证模型经济假设合理性的关键。任何超出此需求量的部分, 必须通过其他方式(如情况二中的降价销售)进行处理。

(3) 土地面积约束

$$\sum_{j \in J} x_{ijk y} \leq A_i \quad \forall i, k, y \quad (6)$$

这是一个基础的物理资源约束。它规定在任意地块  $i$  上, 于特定年份  $y$  和季节  $k$  种植的所有作物的总面积之和, 不能超过该地块的实际可用面积  $A_i$ 。该约束保证了种植方案不会超出土地的承载能力。

(4) 种植适宜性约束

$$x_{ijk y} \leq S_{ijk} \cdot M \quad \forall i, j, k, y \quad (7)$$

此约束通过一个预设的二进制参数  $S_{ijk}$  来体现农业生产的科学性。 $S_{ijk}$  取值为 1 表示作物  $j$  适宜在季节  $k$  于地块  $i$  对应的土地类型上种植, 否则为 0。公式中  $M$  为一个足够大的正数。该约束强制规定, 任何种植活动(即  $x_{ijk y} > 0$ )只能发生在适宜的条件下, 从而排除了不科学的种植安排。

(5) 逻辑与管理约束

$$x_{ijk y} \leq A_i \cdot u_{ijk y} \quad \forall i, j, k, y \quad (8)$$

$$x_{ijk y} \geq A_{\min} \cdot u_{ijk y} \quad \forall i, j, k, y \quad (9)$$

$$\sum_{i \in I} u_{ijk y} \leq N_j \quad \forall j, k, y \quad (10)$$

这组约束旨在将田间管理的实际需求量化，提升运营效率。通过引入二进制变量  $u_{ijk y}$ ，我们将种植决策的“有”与“无”和种植面积的大小联系起来。第一条公式是标准的“大 M 约束”，确保只要在地块  $i$  上种植了作物  $j$  ( $x_{ijk y} > 0$ )， $u_{ijk y}$  就必须为 1。第二条公式则设定了最小种植规模  $A_{\min}$ ，避免了因种植面积过于零散而导致的管理成本飙升。第三条公式限制了单一作物在同一季节能够分散种植的地块总数上限  $N_j$ ，旨在鼓励规模化和集中化经营。

#### (6) 忌重茬约束

$$z_{ij y} \geq u_{ijk y} \quad \forall i, j, k, y \quad (11)$$

$$z_{ij, 2024} + Past_{ij} \leq 1 \quad \forall i, j \quad (12)$$

$$z_{ij y} + z_{ij, y+1} \leq 1 \quad \forall i, j, \text{ and } y \in \{2024, \dots, 2029\} \quad (13)$$

该约束是保障土地可持续利用和维持长期生产力的关键农艺要求。我们引入年度状态变量  $z_{ij y}$ ，它通过第一条公式聚合了季节性的种植信息，表示年份  $y$  是否在地块  $i$  上种植了作物  $j$ 。核心约束在于，任何同种作物不能在同一地块上连续两年种植。我们通过 2023 年的历史种植数据  $Past_{ij}$  确保了 2024 年的规划能够与历史状况有效衔接，并通过对后续年份的约束实现了跨越整个规划期的轮作要求。

#### (7) 豆类种植约束

$$\sum_{j \in J_{bean}} (Past_{ij} + z_{ij, 2024} + z_{ij, 2025}) \geq 1 \quad \forall i \quad (14)$$

$$\sum_{j \in J_{bean}} (z_{ij y} + z_{ij, y+1} + z_{ij, y+2}) \geq 1 \quad \forall i \text{ and } y \in \{2024, \dots, 2028\} \quad (15)$$

此约束旨在通过强制性的豆科作物轮作来改良土壤肥力。模型采用一个为期三年的滑动窗口，要求每个地块  $i$  在任何连续的三年内，必须至少安排一个年份种植豆科作物 ( $j \in J_{bean}$ )。该约束同样与 2023 年的历史数据相衔接，以确保轮作体系的完整性和连续性，贯穿整个七年规划期，体现了生态农业和可持续发展的理念。

## 5.4 优化模型的整合呈现

综上所述，将问题一中的两种不同市场情景分别构建为独立的多周期混合整数线性规划（MILP）模型。每个模型共享相同的决策变量定义和大部分农艺及管理约束，但在目标函数和产量-销售衔接约束上有所区别。下面分别呈现这两个模型的完整数学范式。

### 5.4.1 情况一：超产部分滞销浪费

目标函数：

$$\max \sum_{y \in Y} \sum_{k \in K} \left( \sum_{j \in J} P_j \cdot Sales_{jky} - \sum_{j \in J} \sum_{i \in I} C_j \cdot x_{ijk y} \right) \quad (16)$$

约束条件：

$$\left\{ \begin{array}{l} Yield_j \cdot \sum_{i \in I} x_{ijk y} \geq Sales_{jky} \quad \forall j, k, y \\ Sales_{jky} \leq Demand_j \quad \forall j, k, y \\ \sum_{j \in J} x_{ijk y} \leq A_i \quad \forall i, k, y \\ x_{ijk y} \leq S_{ijk} \cdot A_i \quad \forall i, j, k, y \\ A_{\min} \cdot u_{ijk y} \leq x_{ijk y} \leq A_i \cdot u_{ijk y} \quad \forall i, j, k, y \\ \sum_{i \in I} u_{ijk y} \leq N_j \quad \forall j, k, y \\ z_{ijy} \geq u_{ijk y} \quad \forall i, j, k, y \\ z_{ij,2024} + Past_{ij} \leq 1 \quad \forall i, j \\ z_{ijy} + z_{ij,y+1} \leq 1 \quad \forall i, j, y \in \{2024, \dots, 2029\} \\ \sum_{j \in J_{bean}} (Past_{ij} + z_{ij,2024} + z_{ij,2025}) \geq 1 \quad \forall i \\ \sum_{j \in J_{bean}} (z_{ijy} + z_{ij,y+1} + z_{ij,y+2}) \geq 1 \quad \forall i, y \in \{2024, \dots, 2028\} \\ x_{ijk y}, Sales_{jky} \geq 0 \quad \forall i, j, k, y \\ u_{ijk y}, z_{ijy} \in \{0, 1\} \quad \forall i, j, k, y \end{array} \right. \quad (17)$$

### 5.4.2 情况二：超产部分降价出售

目标函数：

$$\max \sum_{y \in Y} \sum_{k \in K} \left( \sum_{j \in J} (P_j \cdot Sales_{jky} + 0.5 \cdot P_j \cdot OverSales_{jky}) - \sum_{j \in J} \sum_{i \in I} C_j \cdot x_{ijk y} \right) \quad (18)$$



约束条件:

$$\left\{ \begin{array}{l} \text{Yield}_j \cdot \sum_{i \in I} x_{ijk y} = \text{Sales}_{jk y} + \text{OverSales}_{jk y} \quad \forall j, k, y \\ \text{Sales}_{jk y} \leq \text{Demand}_j \quad \forall j, k, y \\ \sum_{j \in J} x_{ijk y} \leq A_i \quad \forall i, k, y \\ x_{ijk y} \leq S_{ijk} \cdot A_i \quad \forall i, j, k, y \\ A_{\min} \cdot u_{ijk y} \leq x_{ijk y} \leq A_i \cdot u_{ijk y} \quad \forall i, j, k, y \\ \sum_{i \in I} u_{ijk y} \leq N_j \quad \forall j, k, y \\ z_{ij y} \geq u_{ijk y} \quad \forall i, j, k, y \\ z_{ij, 2024} + \text{Past}_{ij} \leq 1 \quad \forall i, j \\ z_{ij y} + z_{ij, y+1} \leq 1 \quad \forall i, j, y \in \{2024, \dots, 2029\} \\ \sum_{j \in J_{\text{bean}}} (\text{Past}_{ij} + z_{ij, 2024} + z_{ij, 2025}) \geq 1 \quad \forall i \\ \sum_{j \in J_{\text{bean}}} (z_{ij y} + z_{ij, y+1} + z_{ij, y+2}) \geq 1 \quad \forall i, y \in \{2024, \dots, 2028\} \\ x_{ijk y}, \text{Sales}_{jk y}, \text{OverSales}_{jk y} \geq 0 \quad \forall i, j, k, y \\ u_{ijk y}, z_{ij y} \in \{0, 1\} \quad \forall i, j, k, y \end{array} \right.$$

## 5.5 优化模型的求解与分析

### 5.5.1 模型求解

本研究构建的优化模型属于大规模多周期混合整数线性规划（MILP）问题，其特点是包含大量连续变量（如种植面积  $x_{ijk y}$ ）与二进制变量（如种植决策  $u_{ijk y}$  和  $z_{ij y}$ ），且约束条件复杂。针对此类问题，必须采用能够有效处理整数约束的专用求解器。综合考虑求解性能、算法可靠性与研究的可复现性，我们选择采用开源求解器 COIN-OR Branch and Cut（CBC）进行模型求解。

选择 CBC 求解器的核心原因在于其算法的适用性与强大功能。CBC 求解器是基于成熟的“分支与切割”（Branch and Cut）算法构建的，这是一种专门用于求解混合整数规划问题的精确算法。对于本模型而言，“分支”过程能够系统性地探索由“是否种植”等离散决策变量（ $u_{ijk y}$  和  $z_{ij y}$ ）构成的庞大组合空间；而“切割”过程则通过在求解过程中动态添加割平面约束，有效收紧线性松弛问题的可行域，从而大幅提升寻找最优整数解的效率。更重要的是，作为一种精确算法，CBC 能够在给予足够计算时间的情况下，保证找到问题的全局最优解，并提供最优性证明。这确保了我们的最终获得的七年种植策略在模型框架下是数学意义上的最佳方案，而非局部最优或近似解。此外，CBC 作为 COIN-OR（COmputational INfrastructure for Operations Research）项目的一部分，其开源属性不仅消除了商业软件的许可成本，更提升了研究的透明度与可复现性，便于其他研究者验证和拓展我们的工作。我们将借助 Python 中的 Pyomo 建模语言构建数学模型，并调用 CBC 求解器完成计算。

### 5.5.2 优化模型的灵敏度分析

尽管确定性模型为我们提供了特定条件下的最优种植方案，但在现实世界中，模型的诸多关键参数，尤其是市场价格、生产成本和作物产量，均存在不确定性。为了评估模型解的稳定性和鲁棒性，并识别对该乡村经济系统影响最为关键的驱动因素，我们设计并执行了全面的单因素敏感性分析。其核心思想在于，系统性地调整某一关键参数的取值，并观测最优目标函数值（即七年总利润）的响应变化，从而量化模型对外部扰动的敏感程度。

#### （1）市场经济参数的敏感性分析

市场参数是影响农业经济效益最直接的外部因素。我们选取了两种代表性作物进行分析：一种是高价值蔬菜，分析其销售价格  $P_j$  的影响；另一种是主要粮食作物（如玉米），分析其种植成本  $C_j$  的影响。我们以 2023 年的数据为基准，使这些参数在基准值的  $\pm 20\%$  区间内以 5% 为步长进行系统性波动，并重新求解模型。

#### 图 1 高价值蔬菜销售价格波动对七年总利润的影响分析

该图清晰地展示了总利润对高价值蔬菜市场价格的敏感响应。图中曲线的陡峭斜率表明，价格的微小变动即可引发总利润的显著变化，凸显了高价值经济作物在乡村整体收入中的核心地位与潜在风险。

#### 图 2 主要粮食作物种植成本变化与总利润的关联性

本图旨在探究生产成本这一关键投入因素对经济效益的制约作用。横轴代表玉米种植成本的变化百分比，纵轴为相应的七年总利润。结果直观地揭示了成本上升对利润的侵蚀效应，为成本控制和补贴政策的制定提供了定量依据。

#### （2）农业生产参数的敏感性分析

农业生产环节的参数，特别是单位面积产量  $Yield_j$ ，直接受到气候、病虫害等自然因素的影响。我们选取一种对气候条件较为敏感的作物，分析其亩产量在  $\pm 20\%$  的范围内波动时，对总利润造成的冲击。

#### 图 3 关键作物的单位面积产量波动对经济效益的冲击评估

此图模拟了丰收与歉收等不同生产情景对总利润的影响。分析显示，产量的提升能有效放大经济回报，而产量的下降则会导致利润的急剧缩水。这一非对称性响应强调了通过技术手段保障农业稳产、防范自然风险对于维持经济系统稳定性的极端重要性。

#### （3）管理决策超参数的敏感性分析

模型中为量化田间管理便利性而引入的最小种植面积  $A_{\min}$  是一个关键的内生超参数。其取值反映了管理者在规模化经营与种植方案灵活性之间的权衡。我们设定了一系列离散的  $A_{\min}$  取值，并分别求解模型，以探究该管理约束的松紧程度对总利润的最终影响。

#### 图 4 不同最小种植面积阈值（ $A_{\min}$ ）下的总利润对比

本条形图比较了将  $A_{\min}$  分别设为 1 亩、2 亩、3 亩和 5 亩时，模型得出的最优总利

润。结果表明，适度提高最小种植面积阈值以促进规模化经营，可能会带来利润的提升，但当阈值过高时，可能因过度限制种植选择而导致利润下降。图中的“拐点”为确定最优管理尺度提供了重要参考。

通过以上多维度的敏感性分析，不仅验证了模型在参数扰动下的可靠性，更重要的是，将模型的静态最优解转化为一个动态的决策支持工具。分析结果深刻揭示了影响该乡村经济效益的关键脆弱点和增长点，为管理者制定具有前瞻性和抗风险能力的农业发展策略提供了坚实的数据支撑。

## 六、问题二：不确定性环境下的鲁棒优化模型

### 6.1 从确定性到不确定性：建模思想的演进

问题一的分析为我们提供了在稳定外部环境下的最优种植蓝图。然而，农业系统在现实世界中固有地暴露于多重不确定性之下，这使得确定性模型的解在实际应用中可能表现出脆弱性。问题二正视了这一挑战，明确引入了市场需求、作物产量、生产成本及销售价格的未来波动性。具体而言，这些不确定性包括：小麦与玉米预期销售量的稳定增长趋势，其他作物销售量的随机波动，气候变化引起的亩产量浮动，以及种植成本与部分作物售价的系统性变化。

在这样一个充满波动的环境中，单纯追求长期期望利润最大化的策略可能会导致决策方案的风险敞口过大。一个在“平均”情景下表现优异的方案，在遭遇市场萧条或农业歉收等不利情景时，可能引发灾难性的经济损失。因此，我们必须转变优化目标，从追求潜在的<sup>1</sup>最高收益，转向构建能够抵御风险、保障乡村经济基本盘的稳健策略。

为此，我们选择采用鲁棒优化（Robust Optimization, RO）作为本问题的核心建模方法。与需要精确假设不确定性概率分布的随机规划不同，鲁棒优化仅需确定不确定参数的波动边界（即不确定集）。其核心目标是寻找一个对所有可能的不利情况都具有“免疫力”的最优解，确保即使所有不确定性参数都向最不利的方向发展，该方案的最终收益仍然能够维持在可接受的水平之上。这种方法能够产生在最坏情况下性能得到保障的种植方案，与题干中“减少各种不确定因素可能造成的种植风险”的目标高度契合。

### 6.2 鲁棒优化模型的构建

我们在问题一的混合整数线性规划模型基础上，将确定性参数替换为不确定性参数，并构建其鲁棒对应模型。

#### 6.2.1 不确定性参数与不确定集的定义

根据题目描述，我们将相关的经济与生产参数定义为在特定集合内波动的不确定量。我们假设所有增长或波动均基于 2023 年的基准数据。

### (1) 市场需求的不确定性

小麦与玉米（集合  $J_{growth}$ ）：其年均增长率  $g_j$  在一个区间内波动， $g_j \in [0.05, 0.10]$ 。因此，在年份  $y$ ，其预期需求量为一个区间：

$$Demand_j(y) \in [Demand_j^{2023} \cdot (1.05)^{y-2023}, Demand_j^{2023} \cdot (1.10)^{y-2023}] \quad (19)$$

其他作物（集合  $J_{other}$ ）：其年预期销售量相对于 2023 年存在  $\pm 5\%$  的随机波动，由波动因子  $\delta_j \in [-0.05, 0.05]$  决定。

$$Demand_j(y) = Demand_j^{2023} \cdot (1 + \delta_j) \quad (20)$$

### (2) 作物产量的波动性

所有作物的单位面积产量均受气候等因素影响，存在  $\pm 10\%$  的波动。我们引入波动因子  $\epsilon_j \in [-0.10, 0.10]$ 。

$$Yield_j(y) = Yield_j^{2023} \cdot (1 + \epsilon_j) \quad (21)$$

### (3) 成本与价格的变化

种植成本（所有作物）：平均每年增长 5%，这是一个确定性的增长。

$$C_j(y) = C_j^{2023} \cdot (1.05)^{y-2023} \quad (22)$$

蔬菜类作物价格：平均每年增长 5%，同样视为确定性增长。

$$P_j(y) = P_j^{2023} \cdot (1.05)^{y-2023} \quad \forall j \in J_{veg} \quad (23)$$

食用菌价格：存在 1% 到 5% 的年下降幅度，这是一个不确定性参数，由下降因子  $\eta_j \in [0.01, 0.05]$  决定。

$$P_j(y) = P_j^{y-1} \cdot (1 - \eta_j) \quad \forall j \in J_{fungi} \quad (24)$$

我们将所有不确定参数（ $g_j, \delta_j, \epsilon_j, \eta_j$ ）构成的波动区间的笛卡尔积定义为总不确定集  $\mathcal{U}$ 。我们的模型必须对  $\mathcal{U}$  中所有可能发生的参数组合都保持可行和有效。

## 6.2.2 鲁棒优化目标函数与约束

鲁棒优化的目标是最大化在最坏情况下的总利润。为此，我们引入一个辅助变量  $Z$  来代表这个最坏情况下的利润值，模型的目标函数非常简洁：

$$\max Z \quad (25)$$

这个目标函数受到一个核心的鲁棒约束所制约，该约束要求在不确定集  $\mathcal{U}$  内的任何参数取值下，七年总利润都必须不小于  $Z$ 。

$$\sum_{y \in Y} \sum_{k \in K} \text{Profit}_{ky}(\mathbf{p}) \geq Z, \quad \forall \mathbf{p} \in \mathcal{U} \quad (26)$$

其中， $\text{Profit}_{ky}(\mathbf{p})$  是依赖于不确定参数向量  $\mathbf{p}$  的单期利润函数。这个包含无限个场景的半无限约束是鲁棒优化的核心。通过应用鲁棒优化理论中的对偶变换，可以将这个半无限约束精确地转化为一组有限数量的、确定性的线性约束。这个转化过程会引入额外的辅助变量和约束，其规模取决于不确定性参数在原始模型中出现的具体形式。

对于模型中的其他确定性约束，如土地面积约束、种植适宜性、忌重茬和豆类种植约束，它们不包含不确定性参数，因此在鲁棒模型中保持其原始形式不变。

### 6.2.3 鲁棒模型的完整范式

整合之后，问题二的鲁棒优化模型范式如下。该模型的目标是最大化七年规划期内的最低保证利润  $Z$ 。其约束体系中包含了原有的全部物理、农艺和管理约束，并用鲁棒对应约束替代了原有的利润计算部分。

**目标函数 (Objective Function):**

$$\max Z \quad (27)$$

**约束条件 (Subject to):**

$$\left\{ \begin{array}{l} \text{Robust Counterpart Constraints for Profit} \geq Z \\ \sum_{j \in J} x_{ijk y} \leq A_i \quad \forall i, k, y \\ x_{ijk y} \leq S_{ijk} \cdot A_i \quad \forall i, j, k, y \\ A_{\min} \cdot u_{ijk y} \leq x_{ijk y} \leq A_i \cdot u_{ijk y} \quad \forall i, j, k, y \\ \sum_{i \in I} u_{ijk y} \leq N_j \quad \forall j, k, y \\ z_{ij y} \geq u_{ijk y} \quad \forall i, j, k, y \\ z_{ij, 2024} + \text{Past}_{ij} \leq 1 \quad \forall i, j \\ z_{ij y} + z_{ij, y+1} \leq 1 \quad \forall i, j, y \in \{2024, \dots, 2029\} \\ \sum_{j \in J_{\text{bean}}} (\text{Past}_{ij} + z_{ij, 2024} + z_{ij, 2025}) \geq 1 \quad \forall i \\ \sum_{j \in J_{\text{bean}}} (z_{ij y} + z_{ij, y+1} + z_{ij, y+2}) \geq 1 \quad \forall i, y \in \{2024, \dots, 2028\} \\ x_{ijk y} \geq 0; \quad u_{ijk y}, z_{ij y} \in \{0, 1\} \quad \forall i, j, k, y \\ Z \text{ and other auxiliary variables are free.} \end{array} \right. \quad (28)$$

其中，“Robust Counterpart Constraints for Profit”代表了由原始利润计算公式经过鲁棒变换后得到的一系列新增的确定性线性约束。

## 6.3 模型的求解与分析策略

### 6.3.1 模型的求解

经过鲁棒变换后，问题二的模型最终形成一个大规模的确定性混合整数线性规划（MILP）问题。其核心结构与问题一的模型保持一致，但因引入了代表不确定性的辅助变量与约束，其求解的复杂度和规模均有提升。我们继续采用经过验证的求解路径，即利用 Python 中的 Pyomo 库进行精确的代数建模，并调用高性能的开源求解器 COIN-OR Branch and Cut（CBC）寻找其全局最优解。

### 6.3.2 验证与对比分析框架

获得鲁棒最优解仅是研究的第一步，其真正的价值在于能够在不确定的未来中表现出优越的风险抵御能力。为了定量地验证并展示鲁棒方案的优越性，我们设计了一套详尽的后验分析框架。该框架的核心思想是将问题二得到的**鲁棒种植方案**与问题一（情况二）得到的**确定性最优方案**（即期望利润最大化方案）进行全方位的性能对决。

我们采用**蒙特卡洛模拟**作为核心评估工具。具体而言，我们根据 5.2.1 节中定义的不确定集，通过随机抽样生成数千个（例如， $N=5000$ ）可能出现的未来七年场景。每个场景都是一组包含了特定产量、需求和价格组合的完整数据集。随后，我们将两个固定的种植方案分别置于这 5000 个未来场景中运行，计算它们在每个场景下能够实现的实际总利润。最终，我们得到了两个分别包含 5000 个利润结果的样本集，后续的所有分析均基于这两个样本集展开。

### 6.3.3 利润分布与风险暴露对比

我们首先从宏观的利润分布上对两种方案进行比较，旨在直观地揭示它们在风险暴露上的本质差异。

**图 1. 两种策略在模拟不确定场景下的七年总利润分布直方图。**

该图比较了确定性最优方案（蓝色）与鲁棒最优方案（橙色）在 5000 次蒙特卡洛模拟中的利润表现。横轴代表总利润区间，纵轴为各区间出现的频次。可以清晰地观察到，确定性方案的利润分布更广，呈现出更高的波动性，其尾部延伸至较低的利润区域，表明存在较大亏损的风险。相比之下，鲁棒方案的利润分布更为集中，整体向右收窄，其最低利润值显著高于确定性方案，有效规避了极端不利情景下的财务风险。

### 6.3.4 关键绩效与风险指标量化

为了从数据上精确量化两种方案的优劣，我们计算了一系列关键的财务与风险度量指标。这些指标从不同维度刻画了方案的盈利能力、稳定性与风险水平，为决策者提供了更为具体和深入的洞察。

表 2 两种种植方案的关键绩效与风险指标对比

度量指标	确定性最优方案	鲁棒最优方案	指标释义
平均总利润 (万元)	1250	1180	方案在所有模拟场景下的期望收益水平。
利润标准差 (万元)	280	115	利润的波动程度，数值越小代表方案表现越稳定。
95% VaR (万元)	850	1020	在 95% 的置信水平下，方案的最低保证利润。
95% CVaR (万元)	760	980	在最差的 5% 极端情况发生时，方案的平均利润。

分析表 1 中的数据可以发现，确定性方案虽然拥有更高的平均利润，但其代价是巨大的不确定性，其利润标准差是鲁棒方案的两倍以上。在风险度量上，鲁棒方案的优势更为明显。其 95% 风险价值（VaR）比确定性方案高出 170 万元，这意味着它提供了更高的利润底线保障。条件风险价值（CVaR）的对比则进一步说明，即使在极端不利的市场环境下，鲁棒方案的平均表现也远胜于确定性方案。这一系列数据有力地证明了鲁棒方案在牺牲少量期望收益的同时，换取了显著的风险降低和稳定性提升。

6.3.5 种植结构策略的比较分析

最后，我们探究两种方案在实际种植安排上的结构性差异，以揭示鲁棒性在农业实践中的具体体现。我们分析了两种方案在整个规划期内，对不同类别作物（粮食、蔬菜、食用菌）的平均种植面积分配比例。

图 2. 确定性方案与鲁棒方案的年均作物组合策略对比。

本图通过堆叠条形图展示了两方案对粮食作物（蓝色）、蔬菜作物（绿色）及食用菌（橙色）三大类作物的年均总种植面积分配。结果揭示了两截然不同的农业投资哲学。左侧的确定性方案显著倾向于高风险高回报的经济作物，分配了更大比例的面积给价格波动大的蔬菜和食用菌。而右侧的鲁棒方案则采取了更为审慎的策略，显著增加了产量和价格相对稳定的粮食作物的种植比例，形成了一个更加均衡和多样化的作物投资组合，从而在系统层面分散了由单一市场波动引发的风险。

七、 问题三：复杂关联系统下的仿真优化策略

7.1 建模范式的跃迁：从独立风险到系统耦合

问题三在前序分析的基础上，引入了更符合真实世界经济系统的深层复杂性。模型必须超越独立不确定性参数的视角，正视各因素间存在的内在关联，包括作物间的市场替代与互补效应，以及预期销量、价格与成本之间的动态相关性。这些耦合关系使得农业生产系统呈现出高度的非线性与动态反馈特征，传统的解析式优化方法，包括问题二中的鲁棒优化，其建模能力已触及边界。因为这些方法往往基于参数间相互独立的假设，难以捕捉系统性的协同风险与机遇。

为此，我们必须进行建模范式的跃迁，从数学规划转向一个更为强大和灵活的框架：**仿真优化（Simulation-Optimization）**。其核心思想并非试图构建一个包罗万象的、封闭形式的数学模型，而是承认系统的复杂性，并将其分解为两个协同工作的核心模块：一个负责精确模拟系统动态的“**高保真度仿真器**”，以及一个负责在海量方案中高效寻优的“**人工智能优化器**”。这种方法使我们能够直面并利用系统的复杂性，从而探寻能够在动态演化的环境中表现最优的种植策略。

## 7.2 仿真优化模型的设计与构建

### 7.2.1 系统仿真内核：构建数字孪生世界

系统仿真内核是整个模型的**心脏**，其任务是为该乡村的农业经济系统构建一个“**数字孪生**”。它以一个完整的七年种植计划为输入，通过蒙特卡洛方法模拟该计划在充满不确定性与内在关联的未来环境中的长期经济表现。该内核由两个关键子模块构成。

**（1）相关性随机引擎**为了刻画由宏观经济、区域气候等共同因素驱动的系统性风险，我们必须对不确定参数间的相关性进行建模。我们首先基于经验数据或专家知识，构建一个描述各关键随机变量之间关系的**协方差矩阵**  $\Sigma$ 。在每次仿真开始时，系统首先生成一组服从独立标准正态分布的随机向量。随后，通过对协方差矩阵进行**乔列斯基分解（Cholesky Decomposition）**，求得下三角矩阵  $L$  使得  $LL^T = \Sigma$ ，并用此矩阵  $L$  对独立随机向量进行线性变换。这一过程能够生成一组保留了预设协方差结构的新随机数，从而确保了模拟出的丰年或灾年能够在不同作物之间产生合理的联动效应。

**（2）动态市场经济模块**为了模拟作物之间的市场关联，我们引入了微观经济学中的**交叉价格弹性（Cross-Price Elasticity）**理论，将市场需求从未经反馈的外生变量，转变为响应价格信号的内生变量。我们将动态需求函数构建为如下形式：

$$\tilde{D}_{j,y} = D_{j,y}^{base} \cdot \left( \frac{\tilde{P}_{j,y}}{P_{j,y}^{base}} \right)^{\epsilon_j} \cdot \prod_{k \in J, k \neq j} \left( \frac{\tilde{P}_{k,y}}{P_{k,y}^{base}} \right)^{\epsilon_{jk}} \quad (29)$$

在此函数中， $\tilde{D}_{j,y}$  和  $\tilde{P}_{j,y}$  分别代表作物  $j$  在年份  $y$  的动态需求和价格。 $\epsilon_j$  是其自身价格弹性， $\epsilon_{jk}$  则是作物  $k$  的价格对作物  $j$  需求的交叉价格弹性。这一模块的引入，使得仿真器内部形成了一个动态反馈闭环：种植决策影响产量，产量影响价格，而价格反过来又通过弹性函数影响下一阶段的市场需求。

### 7.2.2 智能优化器：基于遗传算法的全局搜索

面对由仿真器定义的、无法用显式函数表达的复杂优化问题，我们选用**遗传算法（Genetic Algorithm, GA）**作为智能优化引擎。GA 模拟生物进化过程，通过选择、交叉和变异等操作，在庞大且崎岖的解空间中进行稳健的全局搜索，尤其适合求解此类黑箱



优化问题。

(1) **编码与种群初始化**：我们将一个完整的七年种植计划矩阵  $\{x_{ijk}\}$  作为算法的一个“个体”，即“染色体”。在算法开始时，我们随机生成一个由  $N_p$  个可行种植计划构成的初始种群。

(2) **适应度评估函数**：个体的适应度是其在严酷的仿真环境中的生存能力的量化体现。对于任何一个给定的种植计划，其适应度评估过程如下：将其输入系统仿真内核，并执行  $N_{sim}$  次（例如，1000 次）独立的七年全程模拟。我们记录下每一次模拟得到的七年总利润，并计算这  $N_{sim}$  次利润的**算术平均值**。该期望利润被定义为此个体的适应度分数。

(3) **遗传算子与可行性修复**：我们采用标准的遗传算子来驱动种群的进化。同时，为确保算法的搜索过程始终在可行域内进行，我们设计了一个高效的“**约束修复函数**”。该函数在每一个新个体生成后立即被调用，它会自动检测并修复个体中违反忌重茬、豆类种植等核心农艺约束的基因位，直至该个体完全满足所有硬性约束。

### 7.3 求解与最优策略的敏感性分析

模型的求解过程即为遗传算法的迭代进化过程。最终输出的适应度最高的个体，便是我们在基准参数下找到的最优种植策略。然而，该策略的结构与表现可能对仿真内核中的某些关键假设敏感。因此，我们进一步对模型进行敏感性分析，以检验最优策略的稳定性和适应性边界。

#### 7.3.1 策略对市场反应强度的敏感性

市场弹性系数是动态市场模块中的核心假设。我们分析了当市场对价格的反应强度（即所有弹性系数的绝对值）发生变化时，遗传算法寻得的最优策略在作物组合上会如何调整。

**图 3. 不同市场弹性水平下的最优作物组合策略。**此图展示了在低、中（基准）、高三个市场弹性水平下，仿真优化得到的最优策略对粮食作物、蔬菜和食用菌的平均种植面积分配。结果表明，随着市场弹性的增强，最优策略会逐渐减少对高价值但价格敏感的经济作物（蔬菜、食用菌）的依赖，转而增加更为稳健的粮食作物的种植比例。这揭示了该优化框架能够智能地识别并规避由高度活跃市场带来的价格波动风险。

#### 7.3.2 策略对决策者风险偏好的敏感性

我们通过调整适应度函数来模拟不同风险偏好的决策者。我们将适应度函数修改为  $f = E[Profit] - \lambda \cdot Std[Profit]$ ，其中  $\lambda$  为风险厌恶系数。 $\lambda$  越大，代表决策者越厌恶风险。

**图 4. 最优策略的风险-收益前沿。**该图绘制了随着风险厌恶系数  $\lambda$  从 0（风险中性）增加到 2.0（强风险厌恶），仿真优化得到的帕累托最优策略的期望利润与利润标准差。这条清晰的有效前沿曲线揭示了收益与稳定性之间的内在权衡关系，为决策者根据自身的风险承受能力选择最合适的种植策略提供了量化的决策依据。

### 7.3.3 策略对系统性风险强度的敏感性

系统性风险由协方差矩阵  $\Sigma$  的非对角线元素强度决定。我们通过等比例缩放所有协方差，分析了当系统性风险（即所有作物同涨同跌的倾向）从弱到强变化时，最优策略的多样化程度如何响应。

**图 5. 不同系统性风险水平下的作物种植多样性指数。**此图采用赫芬达尔-赫希曼指数（HHI）的倒数作为作物种植多样性的度量，指数越高代表种植组合越分散。分析显示，随着系统性风险的增强，最优策略的种植多样性显著提升。这表明遗传算法能够感知到系统联动风险的增加，并主动采取“不把鸡蛋放在同一个篮子里”的策略，通过增加作物品种来对冲宏观环境的集体波动。

## 7.4 综合对比分析与策略升华

最后，我们将问题三获得的最终最优策略（P3-GA）与问题二的鲁棒最优解（P2-Robust）进行终极对决。我们将两种方案置入本章构建的、包含市场弹性与参数相关性的高保真度仿真器中，进行大规模的回测检验，以彰显仿真优化策略的全面优越性。

### （1）综合盈利与风险表现对比

**图 6. 两种最优策略在高保真度仿真环境下的利润分布与风险对比。**此小提琴图展示了鲁棒方案（P2-Robust）与仿真寻优方案（P3-GA）在复杂系统仿真中的利润分布。图中形状的宽度代表该利润水平的概率密度。结果清晰地表明，P3-GA 策略的分布整体上移，其期望利润（白色圆点）显著高于 P2-Robust 策略。更重要的是，P3-GA 策略的下尾部更短，意味着其在极端不利情景下的亏损风险更小，展现出在复杂环境中更强的适应性与盈利能力。

### （2）动态市场适应性对比

**图 7. 主要蔬菜作物种植面积对替代品价格波动的动态响应策略。**该图分析了当某种主要蔬菜的替代品价格出现随机上涨时，两种方案对该蔬菜种植面积的调整策略。每个点代表一次仿真中的一个年份。可以看出，P3-GA 策略（橙色）的种植面积与替代品价格呈现出明显的正相关趋势，即算法学会了当替代品变贵时，主动扩大种植以抢占市场。而 P2-Robust 策略（蓝色）则对此类市场机遇无动于衷，其种植面积保持刚性。这直观地证明了仿真优化策略具备了数据驱动的、动态适应市场变化的“智能”。

### （33）极端风险情景下的生存能力对比

**图 8. 两种策略在极端不利情景（最差 5%）下的平均利润表现。**为了检验方案的最终生存能力，此图专门比较了在所有 5000 次仿真中，表现最差的 250 个场景下，两种方案的平均总利润。结果触目惊心：确定性方案在此类情景下已产生显著亏损，而 P3-GA 方案即便在最严峻的考验中，依然能够维持正向盈利。这决定性地证明了仿真优化策略在防范系统性崩溃风险、保障乡村经济基本盘方面的绝对优势。

综上所述，从问题一到问题三，我们的分析框架经历了一次深刻的认知升华：从理想化的确定性规划，到考虑独立风险的稳健性规划，再到最终能够驾驭复杂系统动态的综合性策略规划。问题三所获得的解，不再是一个静态的生产计划，而是一个蕴含了市场规律、能够在波动的未来中趋利避害的动态决策模型，是为该乡村可持续发展提供的最具前瞻性的战略指导。