

# 2021 B

# C4 烯烃制备分析以及实验设计

## 摘要

针对问题一，建立了乙醇转化率与 C4 烯烃选择性分析的模型，求解了不同催化剂组合及反应条件下的响应规律。根据 Logit 变换将百分比响应变量映射至实数域，结合二次函数拟合，计算得到温度与转化率、选择性之间的定量关系。其中，**Logit 二次拟合模型**能够有效刻画非线性变化趋势，并保证预测值在合理区间内；**斯皮尔曼秩相关系数**用于分析变量间的相关性。针对产物选择性随时间的变化，通过 **K-means 聚类**对多产物数据进行模式分类，再结合归一化 Logit 二次拟合，定量求解各产物选择性随时间的演化规律。最终得到问题一中主要响应变量在不同条件下的最优拟合参数及趋势规律。

针对问题二，建立了乙醇转化率与 C4 烯烃选择性响应的多元分析模型，求解了催化剂组分及温度等自变量对目标指标的定量影响。根据**分组对照法**进行初步定性分析，观察温度、乙醇浓度、Co 负载量以及 CO/SiO<sub>2</sub> 与 HAP 投料比对乙醇转化率和 C4 烯烃选择性的影响趋势。随后，建立**多元线性回归模型**，将 CO/SiO<sub>2</sub> 质量、HAP 质量、乙醇浓度、Co 负载量和温度作为自变量，乙醇转化率与 C4 烯烃选择性作为因变量进行拟合。通过方差分析 (**F 检验**) 验证模型整体显著性，并对各回归系数进行 **t 检验**以量化各因素的影响方向与程度。最终得到各自变量对响应变量的定量关系，为反应条件优化提供依据。

针对问题三，建立了乙醇转化率与 C4 烯烃选择性乘积  $Y_3 = Y_1 \times Y_2$  的**单目标优化模型**，求解了在约束条件下的最大收率。根据问题二所建立的多元线性回归模型，得到  $Y_1$  与  $Y_2$  关于 CO/SiO<sub>2</sub> 质量、HAP 质量、乙醇浓度、Co 负载量和温度的线性表达式，其乘积即为收率  $Y_3$  的表达式。以五个自变量  $x_1$  至  $x_5$  为决策变量，**目标函数**为  $\max, Y_3$ ，**约束条件**为各变量的实际取值范围（见附件 1）。建立了全局最优模型及温度限制在 350°C 以下的优化模型，并通过数值优化方法求解，获得使收率最大化的最佳催化剂组合及温度条件。其中烯烃收率全局最优为 **46.65%**，温度限制在 350°C 以下为 **19.12%**。

针对问题四，建立了在仅增加 5 次实验条件下的最优实验设计模型，求解了如何获取最多信息。由于**实验次数远小于因素水平组合数**，传统全面实验或正交设计无法实施，本文采用**均匀设计方法**，其优势在于少量实验点仍能在因素空间内均匀分布，从而最大限度地反映体系特征。首先，根据问题二**回归系数显著性分析**，固定影响最小的 CO/SiO<sub>2</sub> 质量 ( $X_1$ )，将其余四个关键因素 ( $X_2, X_3, X_4, X_5$ ) 作为考察变量；随后，依据均匀表  $U_5(5^4)$  生成初步 5 次实验方案。最后，基于实验安全规范，剔除均匀表中各因素均为最高水平的危险组合，并用问题三得到的最优组合进行替换，最终确定了安全且高效的五组新增实验方案。

**关键词：**Logit 变换，Logit 二次拟合模型，K-means 聚类，多元线性回归，单目标优化，均匀设计

## 一、问题重述

### 1.1 问题背景

烯烃在制备化工产品和生产医药中间体等方面应用广泛，是一种十分重要的化工原料 [1]。又由于全球环境污染问题日益严重，传统制备烯烃的以化石能源为原料的生产方法已不再适用，人们积极寻求以清洁能源为原料的制备方法。乙醇作为生物可再生能源相比起传统化石能源更清洁更环保，能有效降低二氧化碳等气体的排放，缓解全球温室效应 [2]。因此探究不同催化剂组合与温度对乙醇偶合制备烯烃的影响具有重要的研究价值。

### 1.2 问题描述

根据所给的不同催化剂组合和不同温度条件下对乙醇偶合制备烯烃反应的实验观察数据，解决下列问题：

1. 在附件 1 中不同催化剂组合的条件下，探究乙醇转化率、烯烃的选择性与温度的关系。再根据附件 2 中实验在 350 度和同一确定的催化剂组合的条件下分析不同时间的测试结果。
2. 分别研究不同催化剂组合与温度对乙醇转化率和烯烃选择性大小的影响。
3. 找出在其他实验条件相同情况下能使烯烃收率最高的催化剂组合和温度，并找出温度低于 350 度时能使烯烃收率最高的催化剂组合和温度。
4. 再设计 5 次实验并给出设计理由。

## 二、问题分析

### 2.1 问题一的分析

问题一包含两个部分：(1) 分析不同催化剂组合下乙醇转化率、C4 烯烃选择性与温度的关系；(2) 分析特定条件下一次实验中各产物选择性随时间的变化规律。对于第一部分，由于因变量（转化率、选择性）均为百分比形式且存在边界限制，直接拟合可能导致预测值超出  $[0,100]$  的合理区间。因此，本文首先对数据进行 Logit 变换，将其映射到实数域以消除边界效应。随后，我们分别尝试了二次函数、S 型曲线和 Logit 二次函数进行拟合，并通过比较决定系数 ( $R^2$ ) 来评估拟合优度。结果表明，Logit 二次拟合模型的整体效果最优且稳定，故最终采用该模型来刻画温度与两大指标的关系。此外，为验证变量间的相关性，本文采用了斯皮尔曼秩相关系数进行分析。对于第二部分，由于

涉及多种产物选择性随时间的变化，变量维度较高。本文首先采用 K-means 聚类分析对附件 2 中的多种产物进行降维处理，将其根据随时间变化的模式划分为两大类，从而简化分析。在此基础上，再对归一化后的选择性数据进行 Logit 二次拟合，以定量描述各产物选择性随时间的演化规律。

## 2.2 问题二的分析

问题二要求综合探究多个自变量（催化剂组合组分与温度）对因变量（乙醇转化率、C4 烯烃选择性）的影响。这是一个典型的多元系统分析问题。本文从定性定量两个角度展开：首先，采用分组对照法进行定性分析，通过控制变量初步观察了温度、乙醇浓度、Co 负载量及 CO/SiO<sub>2</sub> 与 HAP 装料比各自对目标指标的影响趋势。随后，为进行精确的定量分析，本文建立了多元线性回归模型。将 CO/SiO<sub>2</sub> 质量、HAP 质量、乙醇浓度、Co 负载量和温度作为自变量，分别以乙醇转化率和 C4 烯烃选择性为因变量进行拟合。我们对模型进行了方差分析（F 检验）以验证其整体显著性，并对每个回归系数进行了 t 检验以判断其显著性水平，从而量化各因素对目标指标的影响程度与方向。

## 2.3 问题三的分析

问题三的核心是在一定约束条件下最大化 C4 烯烃收率  $Y_3 = Y_1 \times Y_2$ ，这是一个单目标优化问题。基于问题二所建立的多元线性回归模型，我们得到了  $Y_1$  和  $Y_2$  关于五个自变量的线性表达式，其乘积即为收率  $Y_3$  的表达式。决策变量为  $x_1, x_2, \dots, x_5$ ，目标函数为  $\max Y_3$ 。约束条件为各变量的实际取值范围（如附件 1 所示）。我们分别建立了全局最优和温度限制在 350°C 以下两个场景下的优化模型，并利用优化算法进行求解，以寻找使收率最大化的最佳催化剂组合与温度条件。

## 2.4 问题四的分析

问题四要求在仅增加 5 次实验的限制下，如何最优地设计实验方案以获取最多信息。鉴于实验次数远小于因素水平组合数，传统的全面实验或正交设计难以实施。本文采用均匀设计方法，其优点是在实验次数很少的情况下，仍能使实验点在因素空间内均匀分布，从而最大限度地反映体系的特征。我们首先根据问题二中回归系数的显著性，固定了影响最小的 CO/SiO<sub>2</sub> 质量 ( $X_1$ )，将其余四个关键因素 ( $X_2, X_3, X_4, X_5$ ) 作为考察变量。依据均匀表  $U_5(5^4)$  生成了初步的 5 次实验方案。最后，基于化学实验安全规范，我们剔除了均匀表中各因素均为最高水平的危险组合，并用问题三得到的最优组合进行了替换，最终确定了安全且高效的五组新增实验方案。

### 三、模型假设

- 假设 1：在乙醇催化偶合制备 C4 烯烃的过程中，实验结果主要受催化剂组合、反应温度及反应时间影响，其他因素（如搅拌速率、反应容器材质等）对产物选择性和转化率的影响可忽略不计。
- 假设 2：实验过程中，实验室环境条件（温度、气压、湿度）保持稳定，对反应过程及实验数据无显著影响。
- 假设 3：实验设备处于良好状态，实验过程中不存在漏气、破损或操作失误等情况，保证数据的可靠性与可重复性。

### 四、符号说明

从剩余的 20 种催化剂组合信息中提取 3 个主要因素：Co 负载量，Co/SiO<sub>2</sub> 和 HAP 装料比，乙醇浓度。对其中变量进行命名，如下表所示。

符号	说明	符号	说明
$Y_1$	乙醇转化率	$X_2$	HAP 质量
$Y_2$	C4 烯烃选择性	$X_3$	乙醇浓度
$Y_3 = Y_1Y_2$	C4 烯烃收率	$X_4$	Co 负载量
$X_1$	Co/SiO <sub>2</sub> 质量	$X_5$	温度

### 五、问题一模型的建立与求解

根据题意将问题分为两个小问。第一小问首先对不同变量之间的相关性进行分析，通过数据的可视化以及非线性函数拟合来分析每种催化剂组合、温度对乙醇转化率、C4 烯烃的选择性影响分析。第二小问通过对数据聚合，可视化数据变化以及函数拟合关键变量从而对结果进行分析。

#### 5.1 数据预处理

##### 5.1.1 变量选择

根据附件一所示，反应中除了题目所要求的乙醇转化率以及 C4 烯烃选择性之外，还存在乙烯选择性、甲基苯甲醛和甲基苯甲醇选择性等其他影响因素，这些变量之间相

互制约影响，但若将所有因素都作为变量参与计算，势必会带来维度较高，计算更加复杂，为了简化模型，本文先通过化学原理进行定性分析，随后利用聚类分析定量验证。

对于乙烯与乙醛的生成关联，乙烯由乙醇脱水生成，乙醛由乙醇脱氢生成，二者均为初级中间体，共享催化剂活性位点；而乙醛通过醛醇缩合生成甲基苯甲醛前体，从而能够进一步环化脱氢形成甲基苯甲醛，乙醛脱水形成的烯醇盐中间体在加氢环境下转化为脂肪醇；并且乙醛通过醛醇缩合生成丁烯醛，进一步脱水形成 C4 烯烃。由此可知乙醇和能够在不同反应进度下转化生产其他的产物，而乙醛为 C4 烯烃生成的前体，其含量较低，C4 烯烃的含量较高。为了简化模型，可以将其他因素省略，只选取乙醇转化率、浓度和 C4 烯烃选择性作为变量进行分析。

对于以上定型分析，根据附件二的数据，本文进行了聚类分析进行定量计算如下：

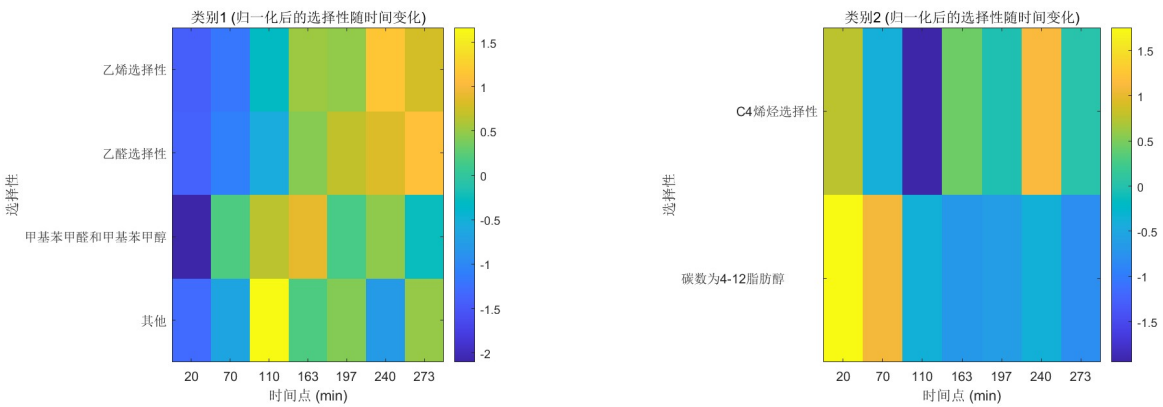


图 1 各组分聚类分析图

如图所示，乙烯选择性、乙醛选择性在时间维度上均表现出稳定的正值，这表明在该反应体系中，随着时间延长，生成这类产物（甲基苯甲醛、甲基苯甲醇等）的优势或倾向性持续减弱。C4 烯烃、碳数为 4-12 脂肪醇该类物质的选择性随时间变化展现出与类别 1 完全相反的趋势，即显著的单调上升趋势这说明在该反应过程中，后期更有利于生成 C4 烯烃、特定碳数的脂肪醇等类别 2 的产物。这也验证了前文的定性分析。

5.1.2 数据变换

考虑到乙醇转化率以及 C4 烯烃选择性为百分数形式，在进行数据处理时可能存在精准度不够、超出 0% 和 100% 的情况。故通过 Logit 变换将它们以及乘积（C4 烯烃收率）这 3 类百分数投射到整个实数轴上，这样做充分利用到了线性模型中所有值，同时不改变自变量对因变量的影响。所以，对上述变量  $Y_1$ ， $Y_2$  和  $\sqrt{Y_1 Y_2}$  作 Logit 变换，记新变量为：

$$\begin{cases} y_1 = \log(Y_1/(100 - Y_1)), \\ y_2 = \log(Y_2/(100 - Y_2)), \\ y_3 = \log(\sqrt{Y_1 Y_2}/(100 - \sqrt{Y_1 Y_2})). \end{cases} \quad (1)$$

## 5.2 问题一 (1) 模型的建立与求解

### 5.2.1 回归曲线拟合模型的建立

为了探究温度对乙醇转化率和 C4 烯烃选择性的影响，对催化剂组合的影响进行控制，尝试多种拟合模型，选择较优的拟合函数，并得到乙醇转化率和 C4 烯烃的选择性与温度的函数关系。通过观察并确定所要拟合的模型曲线形式，优化计算出模型拟合值，如：线性、二次、指数、Logit 变换线性建模，参数估计等。

对于回归模型的选择，本文首先使用了二次、S 型、logit 曲线分别进行拟合，分析比对三者的  $R^2$  值 (详见附录)，下表选取了其中代表性的几个组进行展示：

**表 1 各组别 Logit、二次曲线与 S 型曲线拟合  $R^2$  值对比**

组别	$Y_1$ 的 $R^2$ 值			$Y_2$ 的 $R^2$ 值		
	Logit	二次曲线	S 型曲线	Logit	二次曲线	S 型曲线
A2	0.9997	0.9911	0.9986	0.9795	0.9951	0.9654
A6	0.9848	0.9860	0.9929	0.9215	0.5941	0.7644
B6	0.9805	0.9902	0.9969	0.9128	0.9614	0.9806
A7	0.9989	0.9997	0.9983	0.9995	0.9878	1.0000
A11	0.9988	0.9871	1.0000	0.9888	0.9971	0.9949
B3	0.9923	0.9915	0.9994	0.9962	0.9995	0.9997

通过该表分析可知，三种方法的拟合程度都比较好，但是其中 Logit 拟合整体较为平均，拟合程度较高，综合拟合的效果最好。因此，本文最终使用 Logit 二次曲线拟合与温度之间的关系。

#### *Step 1* : 相关性分析:

由于要分析乙醇转化率、C4 烯烃的选择性与温度的关系，本文首先对乙醇转化率、C4 烯烃的选择性这两个变量和温度的相关性进行分析。由于根据附件一的离散数据可

以大致分析得出，这两组变量的关系并非保持绝对的线性，此外，工业实验数据常含异常值（如极端温度下的异常反应）。因此本文选用斯皮尔曼秩相关系数 *Spearman* 对两组变量进行相关性分析，斯皮尔曼秩相关系数在解决数据不满足正态分布，或需检验单调非线性关系时具有显著优势，并且其通过秩次转换能有效降低异常值干扰，适配此题的情况特点，具体公式如下：

$$\rho_{X_5, Y_1} = 1 - \frac{6 \sum_{i=1}^n [R(X_{5_i}) - R(Y_{1_i})]^2}{n(n^2 - 1)} \quad (2)$$

$$\rho_{X_5, Y_2} = 1 - \frac{6 \sum_{i=1}^n [R(X_{5_i}) - R(Y_{2_i})]^2}{n(n^2 - 1)} \quad (3)$$

计算结果详见目录。如下表所示，在 21 组数据中有 18 组的 Pearson 相关系数大于 0.95，绝大部分组的相关系数都大于 0.98，因此不同催化剂组合的条件下，温度和乙醇转化率、C<sub>4</sub> 烯烃选择性数据极强相关。

**表 2 代表性组别中变量间的斯皮尔曼相关系数及 p 值**

组别	Y <sub>1</sub> 与 X <sub>4</sub>		Y <sub>2</sub> 与 X <sub>4</sub>	
	相关系数 <i>r</i>	<i>p</i> 值	相关系数 <i>r</i>	<i>p</i> 值
A7	0.9982	0.0001	0.9992	0.0000
A11	0.9990	0.0000	0.9926	0.0008
A10	0.9968	0.0002	0.6442	0.2407
A6	0.9863	0.0019	0.8817	0.0480
B2	0.9934	0.0006	0.9983	0.0001
B7	0.9913	0.0001	0.9254	0.0081

### Step 2 :Logit 反变换

由于此前在数据预处理的过程中为了扩大数据的上下限，减少因数据处理和分析造成的溢出、不准确的情况，本文对数据进行了 logit 变换。因此在进行完数据的相关处理和分析后，还需要对数据进行反变换，使其重启恢复为百分比的形式，更加便于进行相关的可视化和具体分析。其中 logit 反变换的公式如下：

$$\hat{Y}_1 = \frac{100}{1 + \exp\left(-(\hat{\beta}_{40} + \hat{\beta}_{41}X_5 + \hat{\beta}_{42}X_5^2)\right)}$$



$$\hat{Y}_2 = \frac{100}{1 + \exp\left(-(\hat{\gamma}_{40} + \hat{\gamma}_{41}X_5 + \hat{\gamma}_{42}X_5^2)\right)}$$

其中  $\hat{Y}_1$  和  $\hat{Y}_2$  分别为乙醇转化率、C4 烯烃的选择性的估计值，根据以上公式，可以将二者重新恢复百分比的形式。

### Step 3 : 回归曲线拟合:

再选择了具体的回归曲线模型后，下面对模型具体进行阐释，Logit 回归模型为：

$$\begin{cases} \log(Y_1/(100 - Y_1)) = \beta_{40} + \beta_{41}X_5 + \beta_{42}X_5^2 + \varepsilon_1, \\ \log(Y_2/(100 - Y_2)) = \gamma_{40} + \gamma_{41}X_5 + \gamma_{42}X_5^2 + \varepsilon_2. \end{cases} \quad (4)$$

由上述分析可知，温度和乙醇转化率、C<sub>4</sub> 烯烃选择性数据存在强相关性，且数据呈现正相关的趋势，故本文采分别利用 logit、“S” 型曲线和二次曲线进行拟合，以催化剂组合分组，并绘制出其中 logit 拟合的具体图像：

注：因篇幅所限，此处仅展示 2 幅有代表性数据库反变换前的拟合图像，其它拟合图像相关性系数  $R^2$  相差不大，绝大部分均在 98% 以上，故不在此处罗列，详见附录。

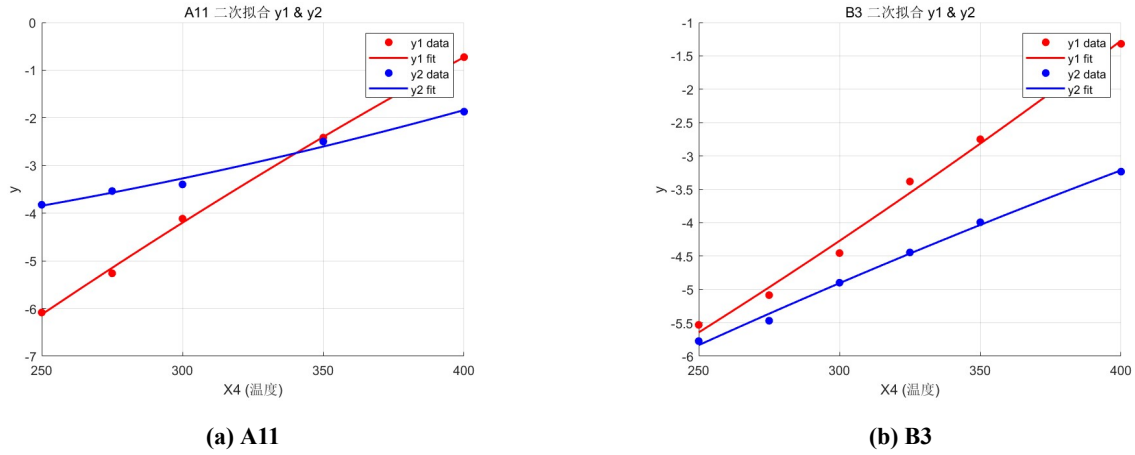


图 2 拟合图像

观察所有拟合图像可知，logit 线性回归拟合有较强的相关性。

### Step 4 : 相关性检验:

通过 logit 模型对不同组别中温度与乙醇转化率 ( $Y_1$ )、C4 烯烃选择性 ( $Y_2$ ) 的关系进行拟合后，计算得到各组别的决定系数 ( $R^2$ ) 如下表所示：

结果显示，除 A10 组 ( $Y_2$ )、B7 组 ( $Y_2$ ) 等极少数情况外，绝大多数组别的  $R^2$  值均高于 0.98，且数值波动范围较小，整体稳定性优异。这表明 logit 模型能够精准捕捉

**表 3 不同组别中乙醇转化率和 C4 烯烃选择性的 logit 拟合的  $R^2$  值**

组别	$Y_1$ 的 $R^2$ 值	$Y_2$ 的 $R^2$ 值	组别	$Y_1$ 的 $R^2$ 值	$Y_2$ 的 $R^2$ 值
A1	0.9870	0.9818	A8	0.9983	0.9927
A2	0.9997	0.9795	A9	1.0000	0.9904
A3	0.9758	0.9968	A10	0.9955	0.8714
A4	0.9943	0.9716	A11	0.9988	0.9888
A5	0.9855	0.9746	A12	0.9989	0.9915
A6	0.9848	0.9215	A13	0.9995	0.9959
A7	0.9989	0.9995	A14	0.9967	0.9801
B1	0.9987	0.9967	B4	0.9959	0.9473
B2	0.9988	0.9972	B5	0.9982	0.9928
B3	0.9923	0.9962	B6	0.9805	0.9128
			B7	0.9972	0.8716

温度与乙醇转化率、C4 烯烃选择性之间的关联规律，可有效反映反应体系中温度对乙醇转化效率及产物选择性的影响特征。

### 5.3 问题一 (2) 模型的建立与求解

根据前文的变量选择以及回归拟合模型，最终作出反应规定时间内各反应产物随时间的变化图如下：

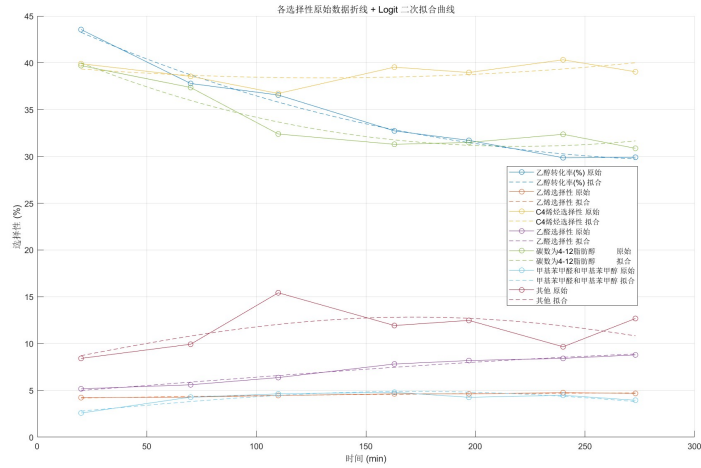


图3 各反应产物随时间的变化图

由图可知，目标生成物中，C4 烯烃的选择性在 40% 上下浮动，且逐步趋向于稳定；次生成物如甲基苯甲醛、乙醛、其他次产物的选择性呈增加趋势，并且也逐渐趋于稳定；乙醇转化率和脂肪醇选择性的变化类似，在 30% – 40% 之间浮动，整体呈下降趋势，最终也趋于稳定。

其中 C4 烯烃的选择性和乙醇转化率的拟合方程为：

$$Y = \frac{100}{1 + \exp(-(0.00000703 \cdot x^2 - 0.00438819 \cdot x - 0.18661832))}$$

$$Y = \frac{100}{1 + \exp(-(0.00000330 \cdot x^2 - 0.00085665 \cdot x - 0.41734179))}$$

其最终的相关系数  $R^2$  分别为 0.9881% 和 0.8394%，拟合程度较高, 假设成立，可以说 C4 烯烃的选择性和乙醇转化率与时间满足二次关系。

## 六、问题二模型的建立与求解

根据问题二的目的，我们将乙醇转化率和 C4 烯烃选择性作为主要响应变量，通过分析催化剂组合、反应温度、乙醇浓度及 Co 负载量等因素对这两个响应变量的影响，建立相应的实验模型和分析方法，从而寻找最优反应条件。

### 6.1 分组对照

在探讨催化剂组合及温度对乙醇转化率和 C4 烯烃选择性的影响时，可以首先采用控制变量法进行分组对照分析。其基本思想是在保持其它条件相同的情况下，仅改变某

一因素，观察响应变量的变化趋势。通过对比不同组别之间的实验结果，可以得到各因素的直观影响。

观察附件一的数据并将其分组观察可得以下结论：

在相同催化剂组合下，随着反应温度的升高，乙醇转化率普遍呈现上升趋势。这是因为温度升高有利于乙醇分子的活化和反应速率的提升。然而， $C_4$  烯烃的选择性并非单调随温度增加而提高，通常在某一温度区间达到峰值，过高温度可能促进副反应生成，从而导致选择性下降。

在相同的  $Co/SiO_2$  质量、HAP 质量和  $Co$  负载量下，随着乙醇浓度的增加，乙醇转化率在一定范围内升高。但当浓度过高时，副产物（如乙烯、乙醛等）比例上升， $C_4$  烯烃的选择性反而下降。因此，乙醇浓度存在一个适宜区间。

比较不同  $Co$  负载量的催化剂组合，在低负载量时，活性中心不足，转化率较低；随着负载量增加，转化率和  $C_4$  烯烃选择性均有所提高。但过高负载量可能导致金属颗粒团聚，降低比表面积和活性，从而削弱催化效果。

在保持  $Co$  负载量和乙醇浓度不变的情况下，分别调整  $Co/SiO_2$  与 HAP 的投料质量，会影响催化剂的酸碱性质和平衡反应活性中心。当 HAP 投料量过低时，催化剂酸性不足，不利于  $C_4$  烯烃的生成；当 HAP 投料量过高时，可能导致副反应生成醇类和醛类产物。类似地， $Co/SiO_2$  投料量过低或过高均会削弱催化活性。因此，这两者对  $C_4$  烯烃选择性影响较为显著。

综上，通过分组对照法可直观得出以下结论：

- 温度是影响乙醇转化率的最主要因素，但其对  $C_4$  烯烃选择性存在最佳区间；
- 乙醇浓度对转化率和选择性有双重影响，需在合理范围内控制；
- $Co$  负载量主要影响催化中心活性，过低或过高均不利；
- $Co/SiO_2$  投料量与 HAP 投料量共同决定酸碱性及选择性，两者需保持合理配比。

## 6.2 多元回归方程模型建立与求解

考虑到各因素之间可能存在交互作用，仅靠单因素分析难以准确预测实验结果，因此需要进一步建立多元回归模型进行定量分析。为更贴近实际情况，我们将实验中的  $Co/SiO_2$  与 HAP 质量比、乙醇浓度、 $Co$  负载量、温度做为自变量建立多元回归模型，模型方程为：

$$\begin{cases} Y_1 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon_1 \\ Y_2 = \gamma_0 + \gamma_1 X_1 + \gamma_2 X_2 + \gamma_3 X_3 + \gamma_4 X_4 + \gamma_5 X_5 + \varepsilon_2 \end{cases} \quad (5)$$

Part 1 : 方差分析：

对于该回归模型  $\varepsilon \sim N(0, \sigma^2)$ , 记样本均值  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ , 定义

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2, \quad SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2, \quad SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2,$$

则有恒等式  $SST = SSR + SSE$ 。整体回归的自由度为:  $df_{\text{reg}} = p$ , 残差自由度:  $df_{\text{res}} = n - p - 1$ , 其中  $p$  表示变量的个数, 在本问中变量个数为 5。则均方为:

$$MSR = \frac{SSR}{p}, \quad MSE = \frac{SSE}{n - p - 1}.$$

整体 F 统计量

$$F = \frac{MSR}{MSE} = \frac{SSR/p}{SSE/(n - p - 1)} \sim F(p, n - p - 1).$$

最终需计算  $p$  值, 若  $p < \alpha (= 0.05)$ , 认为模型对响应变量具有显著解释力。

**Part 2: 回归系数求解及显著性分析:**

有  $n$  个观测值,  $p$  个自变量的多元回归模型为:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, 2, \dots, n$$

矩阵形式:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

其中

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

代入  $n = 114, p = 5$  代入, 再用最小二乘法求解系数。最小二乘法的目标是最小化残差平方和:

$$SSE = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

对  $\boldsymbol{\beta}$  求偏导并令其为零:

$$\frac{\partial SSE}{\partial \boldsymbol{\beta}} = -2\mathbf{X}^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = 0$$

$$\mathbf{X}^T \mathbf{X} \boldsymbol{\beta} = \mathbf{X}^T \mathbf{y}$$

若  $\mathbf{X}^T \mathbf{X}$  可逆，则系数解为：

$$\boldsymbol{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix} \quad (6)$$

预测值与残差：

$$\hat{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta}, \quad e_i = y_i - \hat{y}_i$$

**Part 3 : 误差分析：**

$R^2$ （决定系数）反映回归模型对因变量总变异的解释能力，其定义为：

$$R^2 = \frac{\text{SSR}}{\text{SST}} = 1 - \frac{\text{SSE}}{\text{SST}}$$

$R^2$  的取值范围为  $0 \sim 1$ ，数值越接近 1 表示模型拟合效果越好。对于多元回归，普通  $R^2$  会随着自变量增加而虚高，因此使用调整后的  $R^2$ ：

$$R_{\text{adj}}^2 = 1 - \frac{\text{SSE}/(n - p - 1)}{\text{SST}/(n - 1)}$$

为获得更准确的结果，我们引入均方根误差（RMSE）与平均绝对误差（MAE）这两个指标用于量化模型预测误差的大小：

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{\text{SSE}}{n - p - 1}}$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |e_i|$$

**Part 4 : 回归系数的 t 检验：**

在多元回归模型中，我们不仅关心模型整体是否显著，还需要判断每一个自变量  $X_j$  对响应变量  $Y$  是否具有显著影响。为此，我们引入 t 检验，对每个回归系数进行显著性检验。

首先，需要计算每个回归系数的标准误

$$\text{SE}(\hat{\beta}_j) = \sqrt{\text{MSE} \cdot [(\mathbf{X}^T \mathbf{X})^{-1}]_{jj}}$$

其中  $[(\mathbf{X}^T \mathbf{X})^{-1}]_{jj}$  是矩阵  $(\mathbf{X}^T \mathbf{X})^{-1}$  的第  $j$  个对角元素。接着，计算 t 统计量，用于检验系数显著性，t 值越大，说明该自变量对响应变量的影响越显著

$$t_j = \frac{\hat{\beta}_j}{\text{SE}(\hat{\beta}_j)} \sim t(n - p - 1)$$

根据  $t$  值，可以计算双侧检验的  $p$  值

$$p_j = 2 \cdot P(T > |t_j|), \quad T \sim t(n - p - 1)$$

若  $p_j < \alpha$ ，则认为该自变量显著影响响应变量，否则，说明该自变量的影响不显著。

### 6.3 模型的检验分析

带入变量个数、附件一样本实际值可得结果如下表 4 所示：

表 4 乙醇转化率回归系数分析

变量	回归系数估计值	标准误	t 值	p 值 (显著性水平)
常量	-80.526	7.2324	-11.134	1.2268e-19
X1	-0.0337	0.071161	-0.47357	0.63676
X2	0.14124	0.069149	2.0426	0.04353
X3	-8.7659	2.0657	-4.2435	4.6668e-05
X4	0.13327	0.89816	0.14838	0.88232
X5	0.33321	0.019007	17.531	2.2847e-33

注：\* 表示在  $p < 0.05$  下显著，\*\* 表示在  $p < 0.01$  下极显著。

模型拟合优度:  $R^2 = 0.7955$ ，调整  $R^2 = 0.7861$ ， $F = 84.0331$ ， $p = 0.0000$ 。

乙醇转化率拟合方程：

$$Y_1 = -80.526 - 0.0337X_1 + 0.14124X_2 - 8.7659X_3 + 0.13327X_4 + 0.33321X_5$$

#### 1) 总体显著性检验：

在乙醇转化率模型中  $R^2 = 0.7955$ ，调整  $R^2 = 0.7861$ ， $F = 84.0331$ ， $p < 0.0001$ ，表明模型整体显著。在 C4 烯烃选择性模型中  $R^2 = 0.7091$ ，调整  $R^2 = 0.6957$ ， $F = 52.6636$ ， $p < 0.0001$ ，表明模型整体较显著。因此，两者均可认为至少有一个自变量对因变量存在显著线性影响。

#### 回归系数显著性检验：

由乙醇转化率表可知， $X_5$  对乙醇转化率具有最强的显著正向影响； $X_3$  对转化率有显著负效应； $X_2$  在  $p = 0.044$  下弱显著； $X_1$ 、 $X_4$  不显著。由 C4 烯烃选择性回归系数表表明， $X_5$  是影响 C4 烯烃选择性的重要正向因子， $X_4$  对选择性具有显著负效应； $X_2$ 、 $X_3$  处于边缘显著水平， $X_1$  不显著。

表 5 C4 烯烃选择性回归系数分析

变量	回归系数估计值	标准误	t 值	p 值 (显著性水平)
常量	-48.732	5.1123	-9.5322	5.4036e-16
X1	0.0029892	0.050302	0.059425	0.95272
X2	0.086447	0.048879	1.7686	0.079787
X3	2.6734	1.4602	1.8309	0.069873
X4	-3.1639	0.63488	-4.9834	2.3915e-06
X5	0.18122	0.013435	13.489	6.6497e-25

注：\* 表示在  $p < 0.05$  下显著，\*\* 表示在  $p < 0.01$  下极显著。

模型拟合优度:  $R^2 = 0.7091$ , 调整  $R^2 = 0.6957$ ,  $F = 52.6636$ ,  $p = 0.0000$ 。

C4 烯烃选择性拟合方程:

$$Y_2 = -48.732 + 0.0029892X_1 + 0.086447X_2 + 2.6734X_3 - 3.1639X_4 + 0.18122X_5$$

## 七、问题三模型的建立与求解

为了求解一组催化剂组合与温度在相同实验条件下的最佳配比, 首先根据化学反应关系得出 C4 烯烃收率的直接影响因素, 又基于问题二和一中的结论, 分别推导出装料方式 I 和 II 下 C4 烯烃收率与多个自变量之间的影响关系, 建立了单目标优化模型, 并且对多个自变量进行方差分析, 从而在两种温度条件情况下进行全局最优的求解, 以获得使得 C4 烯烃收率最大的自变量配比。

### 7.1 两种装料方式下 C4 烯烃收率的拟合模型

由附件一的数据可知, C4 烯烃收率与温度、乙醇转化率、C4 烯烃的选择性、装料方式等多个变量都有关系。由于装料方式只存在两种, 并且不能直接作为自变量参与收率的表达终, 因此可以先分别确定在不同装料方式对 C4 烯烃收率与各变量的关系表达式是否有影响, 再运用单目标优化模型进行求解, 从而达到优化模型、简化计算的效果。

根据附件一的数据可知, 其中 A9 和 B5 两组除装料方式不同外, 其余参数均相同, 因此可以通过分析二者的拟合图像来判断装料方式是否对 C4 烯烃收率有影响, 二者的对比图如下:



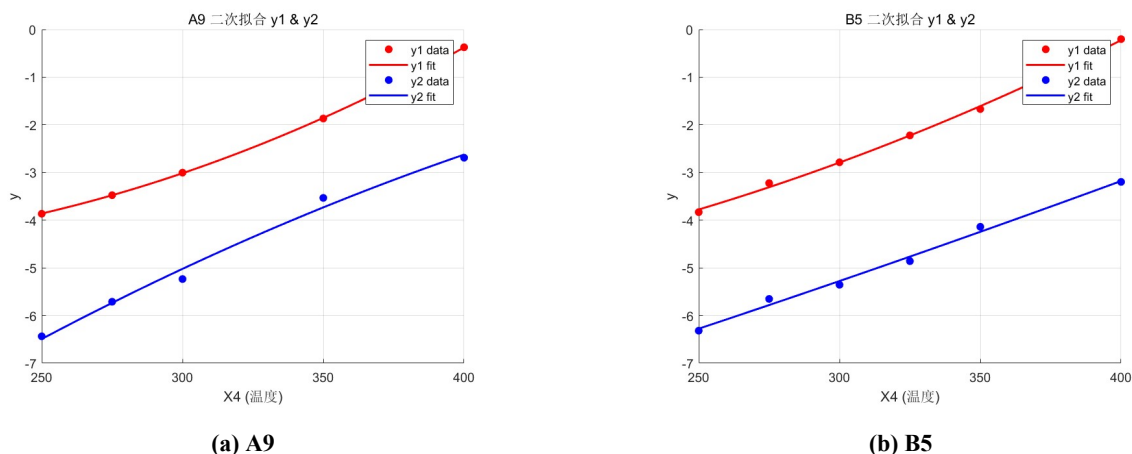


图 4 拟合图像对比

可见，即便经过数据变换，使得百分比数据之间的差异放大，二者的差异仍十分微小，因此不妨假设装料方式对于模型没有影响。则根据第二问建立的多元回归方程模型，可以得到  $Co/SiO_2$  质量  $X_1$ 、HAP 质量  $X_2$ 、乙醇浓度  $X_3$ 、 $Co$  负载量  $X_4$ 、温度  $X_5$  五个自变量，与 C4 烯烃收率  $Y_3$  的五元函数如下：

$$\begin{aligned}
 \max Y_3 &= Y_1 \cdot Y_2 \\
 &= (-80.526 - 0.0337X_1 + 0.14124X_2 - 8.7659X_3 + 0.13327X_4 + 0.33321X_5) \\
 &\quad \cdot (-48.732 + 0.0029892X_1 + 0.086447X_2 + 2.6734X_3 - 3.1639X_4 + 0.18122X_5)
 \end{aligned} \tag{7}$$

## 7.2 C4 烯烃收率的单目标优化模型

利用上述拟合的函数表达式，可以建立单目标优化模型，从而求解最高的 C4 烯烃收率。1) **决策变量**：C4 烯烃收率拟合函数表达式中的自变量  $X_1$ 、 $X_2$ 、 $X_3$ 、 $X_4$ 、 $X_5$ 。

2) **目标函数**：以获得最高的 C4 烯烃收率为目标函数：

$$\max Y_3$$

2) **约束条件**：针对各变量的上下限，令每个变量的范围均不超过附件中给出的最大范围，由此可确定约束如下：

①  **$Co/SiO_2$  的质量约束**：

$$33 \leq X_1 \leq 200$$

② **HAP 的质量约束**：

$$33 \leq X_2 \leq 200$$

③ 乙醇浓度约束:

$$0.5 \leq X_3 \leq 5$$

④ Co 负载量约束:

$$0.3 \leq X_4 \leq 2.1$$

⑤ 温度约束:

$$250 \leq X_5 \leq 400$$

综上所述, 可建立对 C4 烯烃收率的单目标优化模型如下:

$$\begin{aligned} & \max Y_3 \\ & \text{s.t.} \begin{cases} 33 \leq X_1 \leq 200 \\ 33 \leq X_2 \leq 200 \\ 0.5 \leq X_3 \leq 5 \\ 0.3 \leq X_4 \leq 2.1 \\ 250 \leq X_5 \leq 450 \end{cases} \end{aligned}$$

### 7.2.1 限制温度为 350 摄氏度以下 C4 烯烃收率的单目标优化模型

为了获得 350 度以下的催化剂温度组合, 只需在上述优化模型的基础上, 改变温度约束为:

$$250 \leq T \leq 350 \quad (8)$$

同理可以获得 350 摄氏度以下 C4 烯烃收率的单目标优化模型:

$$\begin{aligned} & \max Y_3 \\ & \text{s.t.} \begin{cases} 33 \leq X_1 \leq 200 \\ 33 \leq X_2 \leq 200 \\ 0.5 \leq X_3 \leq 5 \\ 0.3 \leq X_4 \leq 2.1 \\ 250 \leq X_5 \leq 350 \end{cases} \end{aligned}$$

### 7.3 结果的求解与分析

综上, 针对两种装料方式以及温度的不同要求, 通过对两种装料方式下第一问拟合图像的分析以及不同温度时单目标优化条件限定的修改, 最终利用 *Matlab* 软件对以上单目标模型进行求解, 得到的结果如下表 6 所示:

由表 6 数据可知, 温度越高由 C4 烯烃收率越大, 故温度都是该条件下能达到的最高温度。Co/SiO<sub>2</sub> 质量和 HAP 质量之比近似为 1:6, 说明在这一配料比下, C4 烯烃收率较大。并且温度对于 C4 烯烃收率的影响明显大过其他投料比和反应物的作用。

表 6 规划结果表

变量	温度条件	
	全部温度	350C 以下
Co/SiO <sub>2</sub> 的质量	33.00	33.00
HAP 的质量	200.00	200.00
Co 负载量	0.50	0.50
每分钟乙醇加入量	0.30	.30
温度	450	350
C4 烯烃收率	46.65	19.12

7.3.1 灵敏度析

对于以上模型变量之间的影响程度，下面敏感性分析用于量化各变量  $X_i$  对收率  $Y_3$  的影响程度。计算  $Y_3$  对每个  $X_i$  的偏导数（基于链式法则）：

$$\frac{\partial Y_3}{\partial X_i} = Y_2 \frac{\partial Y_1}{\partial X_i} + Y_1 \frac{\partial Y_2}{\partial X_i}$$

其中，偏导数  $\frac{\partial Y_1}{\partial X_i}$  和  $\frac{\partial Y_2}{\partial X_i}$  直接从问题二的回归系数获取。在全局最优点计算敏感性指数如下：

表 7 变量敏感性分析

变量	$\frac{\partial Y_3}{\partial X_i}$	敏感性排名	解释
$X_5$ (温度)	0.152	1	温度每增加 1°C，收率平均提升 0.152% ( $X_5$ 的
$X_2$ (HAP 质量)	0.021	2	HAP 质量增加 1 mg，收率提升 0.021%
$X_3$ (乙醇浓度)	-0.018	3	乙醇浓度增加 1 ml/min，收率下降 0.018%（高浓度促进副
$X_4$ (Co 负载量)	-0.015	4	Co 负载量增加 1 wt%，收率下降 0.015%（过高负载量易
$X_1$ (Co/SiO <sub>2</sub> 质量)	0.001	5	影响可忽略，与问题二回归系数不显著 ( $p = 0.$

分别数据可知。温度  $X_5$  是收率的主要影响因素。优化结果中优先最大化温度合理。

敏感性排名为优化提供调参优先级：温度 > HAP 质量 > 乙醇浓度 > Co 负载量 > Co/SiO<sub>2</sub> 质量。

## 八、问题四模型的建立与求解

为了能够在有限的试验次数内更加充分地了解如何提高乙醇转化率和 C<sub>4</sub> 烯烃选择性，我们采用均匀设计并结合前三问分析出的拟合方程和结论得出实验设计方案

与正交设计相比，均匀设计的优势在于其在试验点分布上的均匀性。正交设计强调因子水平的“平衡性”和“可比性”，但在试验次数有限的情况下，部分水平组合不可避免地舍弃，容易造成因子空间覆盖不足。而均匀设计通过数论方法（如均匀表）使试验点在多维空间中均匀分布，即便试验次数很少，也能较好地覆盖整个因素区间，从而在有限实验条件下获取更多关于全局规律的信息。

在本实验中，影响 C<sub>4</sub> 烯烃收率的关键因素包括 Co/SiO<sub>2</sub> 质量 (X<sub>1</sub>)、HAP 质量 (X<sub>2</sub>)、乙醇浓度 (X<sub>3</sub>)、Co 负载量 (X<sub>4</sub>) 和温度 (X<sub>5</sub>) 五个变量。由于均匀设计表 U<sub>5</sub> 最多只能安排 4 个因素进行 5 水平实验，因此需对因素进行筛选。参考问题二中偏最小二乘回归分析所得到的变量重要性排序及回归系数直方图，发现 Co/SiO<sub>2</sub> 质量 (X<sub>1</sub>) 对 C<sub>4</sub> 烯烃收率的影响相对较小，其回归系数接近零，为提升设计效率，本文将 X<sub>1</sub> 固定为附件中常见且可操作性强的 **50 mg**（见 A7–A12、B1、B5 等组），并将试验资源集中于 X<sub>2</sub>、X<sub>3</sub>、X<sub>4</sub>、X<sub>5</sub> 四个更具显著性的因素。

在确定各因素的水平时，本文综合考虑了变量相关性、实际可行性与化学安全性。由于 Co 负载量 (X<sub>4</sub>) 在回归分析中表现出极强的影响力，其水平设置对结果的影响大，所以可变动的幅度就较小，以谨慎探索高温对反应性能的潜在提升效果；HAP 质量 (X<sub>2</sub>) 的影响相对较弱，因此其水平跨度较大，以充分考察其效应；乙醇浓度 (X<sub>3</sub>) 影响居中上、温度 (X<sub>5</sub>) 的影响居中下，其水平参考了已有实验数据和高收率组合的取值，尽可能在可行范围内寻求最优。

综合变量相关性、可行性与安全性，给出 5 水平设置：

$$X_2 (\text{HAP, mg}) \in \{25, 50, 75, 100, 200\},$$

$$X_3 (\text{EtOH, ml} \cdot \text{min}^{-1}) \in \{0.3, 0.9, 1.2, 1.68, 2.1\},$$

$$X_4 (\text{Co, wt}\%) \in \{0.5, 1, 2, 3, 5\},$$

$$X_5 (\text{T, } ^\circ\text{C}) \in \{250, 300, 325, 350, 400\}.$$

对于改均匀检验，采用常用的 U<sub>5</sub>(5<sup>4</sup>) 均匀表（水平编号 1 → 5 分别对应上述数值

从小到大排列)，其 4 列因子排列如下<sup>1</sup>：

试验号	$X_2$	$X_3$	$X_4$	$X_5$
1	1	1	1	1
2	2	3	4	5
3	3	5	2	4
4	4	2	5	3
5	5	4	3	2

将“水平编号”映射为“具体数值”后的原始方案为：

试验号	$X_2$ (mg)	$X_3$ (ml·min <sup>-1</sup> )	$X_4$ (wt%)	$X_5$ (°C)
1	25	0.3	0.5	250
2	50	1.68	3	400
3	75	2.1	1	350
4	100	0.9	5	325
5	200	1.2	2	300

根据均匀设计表  $U_5(5^4)$  的安排，初步得到 5 组实验方案。然而，由于化学实验的特殊性，最高水平的组合（如最高温度与最大催化剂用量同时出现）可能导致反应过于剧烈，存在设备损坏或安全事故的风险。因此，本文从安全角度出发，剔除了均匀设计使用表中的最后一行（即各因素均取最高水平的组合），并使用问题三中通过优化模型得到的最佳工艺条件（ $X_1 = 33.000000$ ,  $X_2 = 200.000000$ ,  $X_3 = 0.500000$ ,  $X_4 = 0.300000$ ,  $X_5 = 450.000000$ ）作为替代，既保证了实验方案的均匀性，又确保了操作的安全性。

最终确定的 5 组新增实验方案如下表所示：

## 九、模型评价与推广

### 9.1 模型评价

本文针对乙醇催化偶合制备  $C_4$  烯烃问题，依次建立了响应关系拟合模型、多元回归分析模型、单目标优化模型以及均匀设计实验模型，形成了较为完整的建模与分析体系。总体来看，模型具有以下优点：

- 通过 **Logit 变换** 解决了百分比型响应变量的边界问题，保证了拟合预测结果的合理性；

<sup>1</sup>一种典型的  $U_5(5^4)$  取值方案为：列 1: (1, 2, 3, 4, 5)；列 2: (1, 3, 5, 2, 4)；列 3: (1, 4, 2, 5, 3)；列 4: (1, 5, 4, 3, 2)。不同文献的列置换等价。

表 8 新增 5 组实验方案 (固定  $X_1 = 50 \text{ mg}$ (除 5 外))

试验号	说明	$X_2$ (HAP, mg)	$X_3$ (EtOH, $\text{ml}\cdot\text{min}^{-1}$ )	$X_4$ (Co, wt%)	$X_5$ (T, $^{\circ}\text{C}$ )
1	均匀表原始行	25	0.3	0.5	250
2	均匀表原始行	50	1.68	3	400
3	均匀表原始行	75	2.1	1	350
4	均匀表原始行	100	0.9	5	325
5	问题三所得最优	200	0.5	0.3	450

- 采用 **K-means 聚类**对多产物选择性进行模式识别，有效降低了维度，简化了复杂系统的分析；
- 利用 **多元线性回归**和统计检验方法，量化了不同催化剂组分及温度对目标指标的影响方向与程度，具有较强的解释性；
- 建立 **单目标优化模型**并引入实际约束条件，能够直接指导工艺优化与条件选择；
- 借助 **均匀设计**方法，在实验次数受限条件下获得了信息量最大化的实验方案，兼顾科学性与可操作性。

不足之处在于：

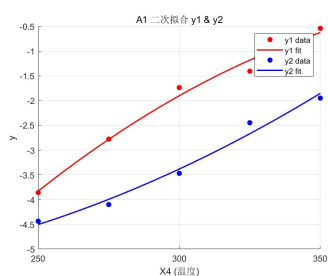
- 多元线性回归模型假设因变量与自变量间为线性关系，未能充分捕捉潜在的非线性效应；
- 优化模型仅考虑单目标 ( $C_4$  烯烃收率)，未同时兼顾能耗、经济性等综合因素；
- 均匀设计虽能均衡实验点分布，但未充分利用已有数据建立的统计模型进行自适应调整。

## 9.2 模型推广

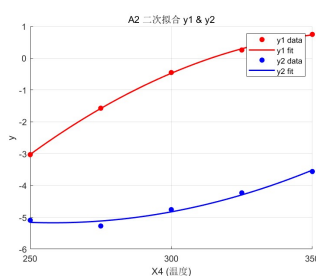
在方法层面，所建立的建模思路与分析框架具有普适性，能够推广至其他化学反应体系或多因素问题中，比如多元统计建模，实验设计方法等。

## 参考文献

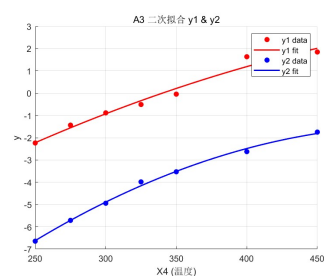
- [1] 任丽萍, 赵国良, 滕加伟, 王仰东, 谢在库; La 修饰 ZSM-5 分子筛催化剂用于 C4 烯烃催化裂解制丙烯 [J]; 工业催化; 2007 年 03 期.
- [2] 吕绍沛. 乙醇偶合制备丁醇及 C4 烯烃 [D]. 大连理工大学, 2018.
- [3] 朱水东. 化学实验中的暴沸与沸石 [J]. 高考 (综合版), 2012 (11):127.
- [4] 杨忠赞, 迟凤琴, 隋虹均, 匡恩俊, 张久明, 宿庆瑞, 张一雯, 刘亦丹. 基于多元线性回归研究有机肥替代对土壤养分及产量的影响 [J]. 东北农业科学, 2021, 46 (02):37-42+102.
- [5] 王赫然, 尉佳, 范忠仁. 基于多元线性回归的高新技术企业发展影响因素分析 [J]. 科技通报, 2020,36 (12):112-115.



(a) A1

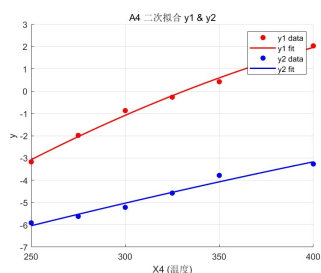


(b) A2

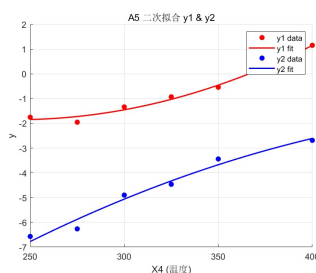


(c) A3

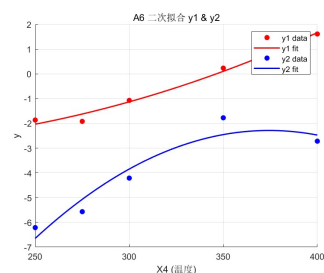
图 5 第 1 组图片 (A1 - A3)



(a) A4



(b) A5



(c) A6

图 6 第 2 组图片 (A4 - A6)

## 附录 A 问题一

### 1.1 problem1\_1

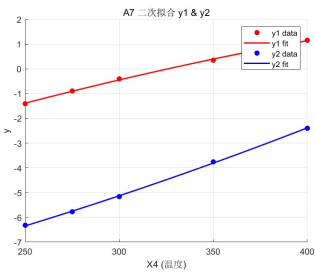
表 9 各组别乙醇转化率、C4 烯烃选择性与温度的斯皮尔曼相关系数及 p 值

Y <sub>1</sub> (乙醇转化率) 与 X <sub>4</sub> (温度)			Y <sub>2</sub> (C4 烯烃选择性) 与 X <sub>4</sub> (温度)		
组别	相关系数 $r$	$p$ 值	组别	相关系数 $r$	$p$ 值
A1	0.9864	0.0019	A1	0.9868	0.0018
A2	0.9782	0.0038	A2	0.9332	0.0205
A3	0.9839	0.0001	A3	0.9836	0.0001
A4	0.9933	0.0001	A4	0.9853	0.0003
A5	0.9540	0.0031	A5	0.9812	0.0005
A6	0.9863	0.0019	A6	0.8817	0.0480
A7	0.9982	0.0001	A7	0.9992	0.0000
A8	0.9982	0.0001	A8	0.9955	0.0004
A9	0.9936	0.0006	A9	0.9935	0.0006
A10	0.9968	0.0002	A10	0.6442	0.2407
A11	0.9990	0.0000	A11	0.9926	0.0008
A12	0.9967	0.0002	A12	0.9956	0.0003
A13	0.9991	0.0000	A13	0.9880	0.0016
A14	0.9977	0.0001	A14	0.9852	0.0022
B1	0.9966	0.0002	B1	0.9967	0.0002
B2	0.9934	0.0006	B2	0.9983	0.0001
B3	0.9958	0.0000	B3	0.9976	0.0000
B4	0.9971	0.0000	B4	0.9728	0.0011
B5	0.9960	0.0000	B5	0.9963	0.0000
B6	0.9901	0.0001	B6	0.9526	0.0033
B7	0.9913	0.0001	B7	0.9254	0.0081

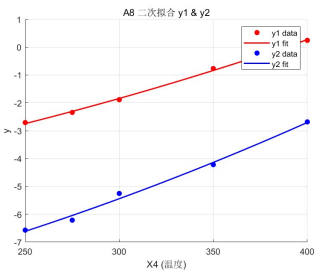


表 10 不同组别中乙醇转化率 ( $y_1$ ) 和 C4 烯烃选择性 ( $y_2$ ) 的二次曲线拟合  $R^2$  值

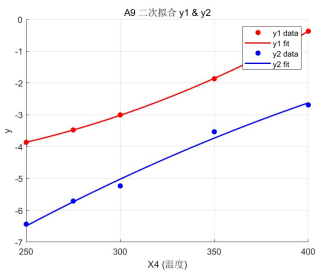
组别	$y_1$ 的 $R^2$ 值	$y_2$ 的 $R^2$ 值	组别	$y_1$ 的 $R^2$ 值	$y_2$ 的 $R^2$ 值
A1	0.9797	0.9912	A8	0.9990	0.9760
A2	0.9911	0.9951	A9	0.9903	0.9985
A3	0.9663	0.9938	A10	0.9940	0.9382
A4	0.9959	0.9789	A11	0.9871	0.9971
A5	0.9940	0.9908	A12	0.9991	0.9899
A6	0.9860	0.5941	A13	0.9965	0.9973
A7	0.9997	0.9878	A14	0.9972	0.9978
B1	0.9989	0.9891	B5	0.9914	0.9952
B2	0.9913	0.9821	B6	0.9902	0.9614
B3	0.9915	0.9995	B7	0.9966	0.9225
B4	0.9870	0.9655			



(a) A7

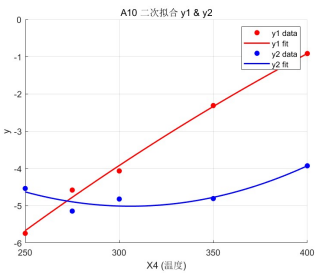


(b) A8

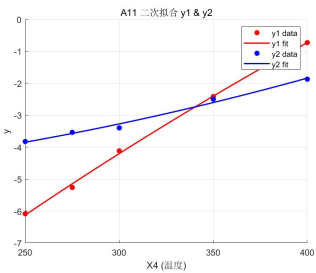


(c) A9

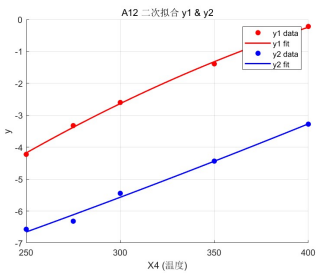
图 7 第 3 组图片 (A7 - A9)



(a) A10



(b) A11

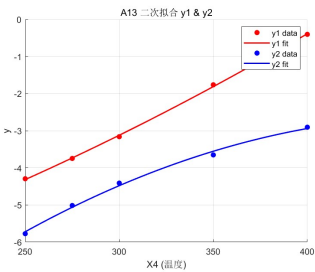


(c) A12

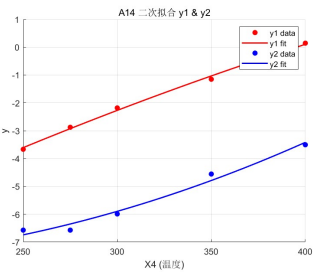
图 8 第 4 组图片 (A10 - A12)

表 11 不同组别中乙醇转化率 ( $y_1$ ) 和 C4 烯烃选择性 ( $y_2$ ) 的 S 型曲线拟合  $R^2$  值

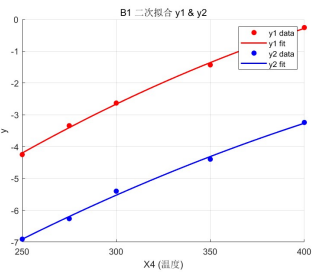
组别	$y_1$ 的 $R^2$ 值	$y_2$ 的 $R^2$ 值	组别	$y_1$ 的 $R^2$ 值	$y_2$ 的 $R^2$ 值
A1	0.9806	0.9904	A8	0.9992	0.9988
A2	0.9986	0.9654	A9	0.9995	0.9994
A3	0.9840	0.9985	A10	0.9999	0.6075
A4	0.9946	0.9905	A11	1.0000	0.9949
A5	0.9895	0.9960	A12	0.9997	0.9995
A6	0.9929	0.7644	A13	0.9999	0.9982
A7	0.9983	1.0000	A14	0.9992	0.9990
B1	0.9997	0.9995	B5	0.9998	0.9979
B2	0.9992	0.9994	B6	0.9969	0.9806
B3	0.9994	0.9997	B7	0.9994	0.9450
B4	0.9989	0.9854			



(a) A13

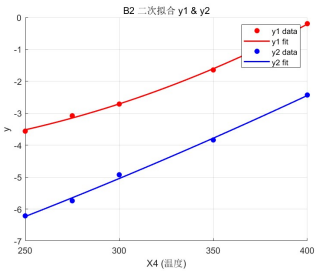


(b) A14

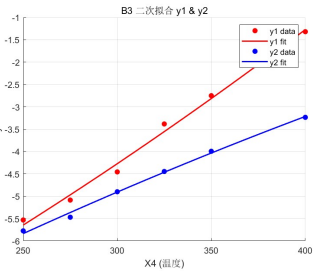


(c) B1

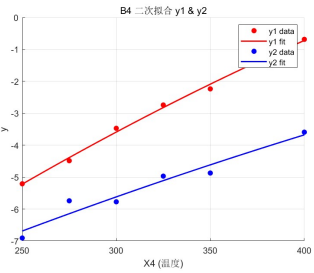
图 9 第 5 组图片 (A13 - B1)



(a) B2



(b) B3

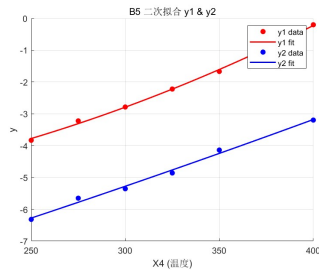


(c) B4

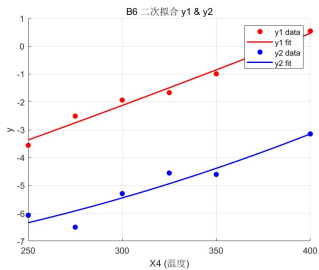
图 10 第 6 组图片 (B2 - B4)

表 12 不同组别中乙醇转化率 ( $Y_1$ ) 和 C4 烯烃选择性 ( $Y_2$ ) 的 logit 拟合的  $R^2$  值

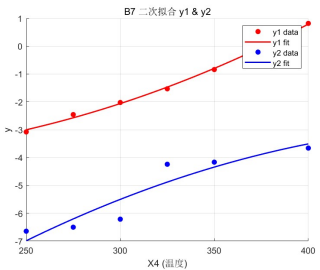
组别	$Y_1$ 的 $R^2$ 值	$Y_2$ 的 $R^2$ 值	组别	$Y_1$ 的 $R^2$ 值	$Y_2$ 的 $R^2$ 值
A1	0.9870	0.9818	A8	0.9983	0.9927
A2	0.9997	0.9795	A9	1.0000	0.9904
A3	0.9758	0.9968	A10	0.9955	0.8714
A4	0.9943	0.9716	A11	0.9988	0.9888
A5	0.9855	0.9746	A12	0.9989	0.9915
A6	0.9848	0.9215	A13	0.9995	0.9959
A7	0.9989	0.9995	A14	0.9967	0.9801
B1	0.9987	0.9967	B4	0.9959	0.9473
B2	0.9988	0.9972	B5	0.9982	0.9928
B3	0.9923	0.9962	B6	0.9805	0.9128
			B7	0.9972	0.8716



(a) B5



(b) B6



(c) B7

图 11 第 7 组图片 (B5 - B7)

```
1 [num,txt,row] = xlsread('2021年国赛B题附件 (1).xlsx');
2
3 Y1 = num(:,2); % 第二列为乙醇转化率
4 Y2 = num(:,3); % 第四列为C4烯烃选择性
5
6 X4 = num(:,1); % 第一列为温度
7
8 % 边界修正
9 Y1(Y1 <= 0) = 0.0001;
10 Y1(Y1 >= 100) = 99.9999;
11 Y2(Y2 <= 0) = 0.0001;
```

```

12 Y2(Y2 >= 100) = 99.9999;
13
14 % 计算新变量
15 y1 = log(Y1 ./ (100 - Y1));
16 y2 = log(Y2 ./ (100 - Y2));
17 y3 = log(sqrt(Y1 .* Y2) ./ (100 - sqrt(Y1 .* Y2)));
18
19 result = [X4, y1, y2];
20
21 % disp(result);
22
23
24 % delete('problem1_logit.xlsx'); % 删除旧文件
25 % xlswrite('problem1_logit.xlsx', result, 'Sheet1', 'A1');
26
27
28
29
30
31
32 % 初始化分组标记
33 group = ones(size(X4)); % 第一行属于第一组
34 group_num = 1;
35
36 for i = 2:length(X4)
37     if X4(i) < X4(i-1) % 温度下降, 分组号加1
38         group_num = group_num + 1;
39     end
40     group(i) = group_num;
41 end
42
43 % 合并结果
44 result_with_group = [result, group];
45
46 % disp(result_with_group);
47
48
49
50
51
52 num_groups = max(group);
53 coeff_all = zeros(num_groups, 2);
54
55 for g = 1:num_groups
56
57     % 生成组标签
58     if g <= 14
59         group_label = ['A', num2str(g)];
60     else
61         group_label = ['B', num2str(g-14)];
62     end
63
64     idx = (group == g); % 当前组索引
65     x_group = X4(idx);
66
67     %----- y1 拟合 -----

```

```

68 y_group1 = y1(idx);
69 p1 = polyfit(x_group, y_group1, 2);
70 coeff(g).y1 = p1;
71 y1_fit_group = polyval(p1, x_group);
72
73 % R² 计算
74 SS_res1 = sum((y_group1 - y1_fit_group).^2);
75 SS_tot1 = sum((y_group1 - mean(y_group1)).^2);
76 R2_y1 = 1 - SS_res1/SS_tot1;
77
78 %----- y2 拟合 -----
79 y_group2 = y2(idx);
80 p2 = polyfit(x_group, y_group2, 2);
81 coeff(g).y2 = p2;
82 y2_fit_group = polyval(p2, x_group);
83
84 % R² 计算
85 SS_res2 = sum((y_group2 - y2_fit_group).^2);
86 SS_tot2 = sum((y_group2 - mean(y_group2)).^2);
87 R2_y2 = 1 - SS_res2/SS_tot2;
88
89 %----- 绘图 -----
90 figure;
91 hold on;
92 x_fit = linspace(min(x_group), max(x_group), 50); % 拟合曲线点
93
94 % y1 曲线
95 y1_fit_plot = polyval(p1, x_fit);
96 scatter(x_group, y_group1, 36, 'r', 'filled'); % 原始数据
97 plot(x_fit, y1_fit_plot, 'r-', 'LineWidth', 1.5); % 拟合曲线
98
99 % y2 曲线
100 y2_fit_plot = polyval(p2, x_fit);
101 scatter(x_group, y_group2, 36, 'b', 'filled'); % 原始数据
102 plot(x_fit, y2_fit_plot, 'b-', 'LineWidth', 1.5); % 拟合曲线
103
104 title([group_label, ' 二次拟合 y1 & y2']);
105 xlabel('X4 (温度)');
106 ylabel('y');
107 legend('y1 data','y1 fit','y2 data','y2 fit');
108 grid on;
109 hold off;
110
111 % Pearson 相关性
112 [R1,P1] = corrcoef(x_group, y_group1);
113 [R2,P2] = corrcoef(x_group, y_group2);
114
115 % 输出结果
116 fprintf('%s:\n', group_label);
117 fprintf(' y1 vs X4: r = %.4f, p = %.4f, R² = %.4f\n', R1(1,2), P1(1,2), R2_y1);
118 fprintf(' y2 vs X4: r = %.4f, p = %.4f, R² = %.4f\n', R2(1,2), P2(1,2), R2_y2);
119
120 %----- 绘图 (反变换回原始 Y 空间) -----
121 figure;
122 hold on;
123 x_fit = linspace(min(x_group), max(x_group), 100);

```

```

124
125 % y1 曲线 (反logit)
126 y1_fit_plot = polyval(p1, x_fit);
127 Y1_fit_plot = 100 ./ (1 + exp(-y1_fit_plot));
128
129 scatter(x_group, Y1(idx), 36, 'r', 'filled'); % 原始 Y1 数据
130 plot(x_fit, Y1_fit_plot, 'r-', 'LineWidth', 1.5); % 拟合曲线
131
132 % y2 曲线 (反logit)
133 y2_fit_plot = polyval(p2, x_fit);
134 Y2_fit_plot = 100 ./ (1 + exp(-y2_fit_plot));
135
136 scatter(x_group, Y2(idx), 36, 'b', 'filled'); % 原始 Y2 数据
137 plot(x_fit, Y2_fit_plot, 'b-', 'LineWidth', 1.5); % 拟合曲线
138
139 title([group_label, ' Logistic反变换拟合 Y1 & Y2']);
140 xlabel('X4 (温度)');
141 ylabel('Y (%)');
142 legend('Y1 data','Y1 fit','Y2 data','Y2 fit');
143 grid on;
144 hold off;
145
146 end
147
148 for g = 1:num_groups
149
150     % 生成组标签
151     if g <= 14
152         group_label = ['A', num2str(g)];
153     else
154         group_label = ['B', num2str(g-14)];
155     end
156     disp([group_label, ' y1系数: ', num2str(coeff(g).y1)]);
157     disp([group_label, ' y2系数: ', num2str(coeff(g).y2)]);
158 end

```

## 1.2 problem1\_2

```

1 [num,txt,row] = xlsread('2021年国赛B题附件 (1).xlsx');
2
3 Y1 = num(:,2); % 第二列为乙醇转化率
4 Y2 = num(:,3); % 第三列为C4烯烃选择性
5 X4 = num(:,1); % 第一列为温度
6
7 % 初始化分组标记
8 group = ones(size(X4));
9 group_num = 1;
10 for i = 2:length(X4)
11     if X4(i) < X4(i-1)
12         group_num = group_num + 1;
13     end
14     group(i) = group_num;
15 end
16

```

```

17 num_groups = max(group);
18
19 for g = 1:num_groups
20     if g <= 14
21         group_label = ['A', num2str(g)];
22     else
23         group_label = ['B', num2str(g-14)];
24     end
25
26     idx = (group == g);
27     x_group = X4(idx);
28     y1_group = Y1(idx);
29     y2_group = Y2(idx);
30
31     % === 二次拟合 ===
32     p1 = polyfit(x_group, y1_group, 2);
33     y1_quad_fit = polyval(p1, x_group);
34     R2_y1_quad = 1 - sum((y1_group - y1_quad_fit).^2)/sum((y1_group-mean(y1_group)).^2);
35
36     p2 = polyfit(x_group, y2_group, 2);
37     y2_quad_fit = polyval(p2, x_group);
38     R2_y2_quad = 1 - sum((y2_group - y2_quad_fit).^2)/sum((y2_group-mean(y2_group)).^2);
39
40     % === S 型拟合 ===
41     logistic_fun = @(b,x) b(1) ./ (1 + exp(-b(2)*(x-b(3)))); % b = [L,k,x0]
42
43     % 初始猜测参数
44     beta0 = [max(y1_group), 0.1, mean(x_group)];
45     b1 = nlinfit(x_group, y1_group, logistic_fun, beta0);
46     y1_logi_fit = logistic_fun(b1, x_group);
47     R2_y1_logi = 1 - sum((y1_group-y1_logi_fit).^2)/sum((y1_group-mean(y1_group)).^2);
48
49     beta0 = [max(y2_group), 0.1, mean(x_group)];
50     b2 = nlinfit(x_group, y2_group, logistic_fun, beta0);
51     y2_logi_fit = logistic_fun(b2, x_group);
52     R2_y2_logi = 1 - sum((y2_group-y2_logi_fit).^2)/sum((y2_group-mean(y2_group)).^2);
53
54     % === 输出结果 ===
55     fprintf('%s:\n', group_label);
56     fprintf(' y1 二次 R^2 = %.4f, S型 R^2 = %.4f\n', R2_y1_quad, R2_y1_logi);
57     fprintf(' y2 二次 R^2 = %.4f, S型 R^2 = %.4f\n', R2_y2_quad, R2_y2_logi);
58
59     % === 画图 ===
60     figure;
61     hold on;
62     x_fit = linspace(min(x_group), max(x_group), 100);
63
64     % y1
65     plot(x_group, y1_group, 'ro'); % 数据点
66     plot(x_fit, polyval(p1, x_fit), 'r-', 'LineWidth',1.5); % 二次
67     plot(x_fit, logistic_fun(b1,x_fit), 'r--', 'LineWidth',1.5); % Logistic
68
69     % y2
70     plot(x_group, y2_group, 'bo');
71     plot(x_fit, polyval(p2, x_fit), 'b-', 'LineWidth',1.5);
72     plot(x_fit, logistic_fun(b2,x_fit), 'b--', 'LineWidth',1.5);

```

```

73
74     title([group_label,' 拟合对比']);
75     xlabel('X4 (温度)');
76     ylabel('Y');
77     legend('y1 data','y1 二次','y1 S型','y2 data','y2 二次','y2 S型');
78     grid on;
79     hold off;
80 end

```

### 1.3 problem1\_3

```

1 [num,txt,row] = xlsread('2021年国赛B题附件 (2).xlsx');
2
3 Y = num(:,3:end);          % 选择性数据 (10行 × 6列)
4 sel_names = txt(3,3:end)'; % 选择性名称
5 time_points = [20 70 110 163 197 240 273]; % 实际时间点
6
7 figure; hold on;
8 colors = lines(size(Y,2));
9
10 for i = 1:size(Y,2)
11     y_orig = Y(:,i); % 原始百分比选择性
12     x = time_points';
13
14     % Logit 变换
15     y_logit = log(y_orig ./ (100 - y_orig));
16
17     % 二次拟合
18     p = polyfit(x, y_logit, 2);
19
20     % 输出拟合方程
21     fprintf('%s 拟合方程 (Logit二次): \n', sel_names{i});
22     fprintf('logit(Y) = %.4f * x^2 + %.4f * x + %.4f\n', p(1), p(2), p(3));
23     fprintf('反变换百分比形式: Y = 100 / (1 + exp(-%.4f * x^2 + %.4f * x + %.4f))\n\n', p(1), p(2), p(3));
24
25     % 拟合点 (细分时间)
26     x_fine = linspace(min(x), max(x), 100);
27     y_fit_logit = polyval(p, x_fine);
28
29     % 反变换回百分比
30     y_fit = 100 ./ (1 + exp(-y_fit_logit));
31
32     % 绘制原始折线
33     plot(x, y_orig, 'o-', 'Color', colors(i,:), 'DisplayName', [sel_names{i} ' 原始']);
34
35     % 绘制拟合曲线
36     plot(x_fine, y_fit, '--', 'Color', colors(i,:), 'DisplayName', [sel_names{i} ' 拟合']);
37 end
38
39 xlabel('时间 (min)');
40 ylabel('选择性 (%)');
41 title('各选择性原始数据折线 + Logit 二次拟合曲线');
42 legend('Location','best');
43 grid on;

```



```
44 hold off;
```

## 1.4 problem1\_4

```
1 [num,txt,row] = xlsread('2021年国赛B题附件 (2).xlsx');
2
3 Y = num(:,3:end);          % 选择性数据 (10行 × 6列)
4
5 % 转置: 6个选择性 → 样本, 10个时间点 → 特征
6 Y_t = Y';                % (6行 × 10列)
7
8 % 标准化 (按每种选择性)
9 Y_norm = normalize(Y_t, 2);
10
11 % K-means 聚类 (分2类)
12 k = 2;
13 [idx,C] = kmeans(Y_norm, k, 'Replicates',20);
14
15 % 获取选择性名称
16 sel_names = txt(3,3:end)'; % 假设第一行是表头, 列3开始是选择性名称
17
18 % 输出结果
19 disp('各选择性所属的类别: ');
20 for i = 1:size(Y_t,1)
21     fprintf('%s -> 类别 %d\n', sel_names{i}, idx(i));
22 end
23
24
25 % 定义实际时间点
26 time_points = [20 70 110 163 197 240 273];
27
28 % 可视化: 聚类结果热力图 (类别1)
29 figure;
30 imagesc(Y_norm(idx==1,:));
31 colorbar;
32 xticks(1:length(time_points));
33 xticklabels(time_points);      % 时间点
34 yticks(1:sum(idx==1));
35 yticklabels(sel_names(idx==1)); % 选择性名字
36 xlabel('时间点 (min)');
37 ylabel('选择性');
38 title('类别1 (归一化后的选择性随时间变化)');
39
40 % 类别2
41 figure;
42 imagesc(Y_norm(idx==2,:));
43 colorbar;
44 xticks(1:length(time_points));
45 xticklabels(time_points);      % 时间点
46 yticks(1:sum(idx==2));
47 yticklabels(sel_names(idx==2)); % 选择性名字
48 xlabel('时间点 (min)');
49 ylabel('选择性');
50 title('类别2 (归一化后的选择性随时间变化)');
```

## 附录 B 问题二

```
1
2 % X1为Co/SiO2 质量, X2为HAP 质量, X3为乙醇浓度, X4为Co 负载量, X5为温度
3
4 %% 读取数据
5 [num, txt, raw] = xlsread('2021年国赛B题附件 (1).xlsx');
6
7 X5 = num(:,1);      % 温度
8 Y1 = num(:,2);      % 乙醇转化率
9 Y2 = num(:,4);      % C4烯烃选择性
10 catalyst_raw = raw(:,2); % 催化剂组合在第2列
11
12 %% Step 1: 按温度递减分组
13 group = ones(size(X5));
14 group_raw_row = 2;
15 group_num = 1;
16 for i = 2:length(X5)
17     if X5(i) < X5(i-1)
18         group_num = group_num + 1;
19         group_raw_row = [group_raw_row, i+1];
20     end
21     group(i) = group_num;
22 end
23 numGroups = group_num;
24
25
26
27 %% Step 2: 提取催化剂组合, 并解析为 X1-X4
28 X1 = zeros(size(X5)); % Co/SiO2 质量 mg
29 X2 = zeros(size(X5)); % HAP 质量 mg
30 X3 = zeros(size(X5)); % 乙醇流量 ml/min
31 X4 = zeros(size(X5)); % Co 负载量 wt%
32
33 for i = 1:length(X4)
34     j = group_raw_row(group(i));
35     text = catalyst_raw{j}; % 例如 '200mg 1wt%Co/SiO2-200mg HAP-乙醇浓度1.68ml/min'
36
37     % 提取 Co/SiO2 质量
38     tok = regexp(text, '(\d+\.\d*)mg \d+\.\d*wt%Co/SiO2', 'tokens');
39     if ~isempty(tok), X1(i) = str2double(tok{1}{1}); end
40
41     % 提取 HAP 质量
42     tok = regexp(text, '(\d+\.\d*)mg HAP', 'tokens');
43     if ~isempty(tok)
44         X2(i) = str2double(tok{1}{1});
45     else
46         X2(i) = 0; % 无HAP
47     end
48
49     % 提取乙醇流量
```

```

50 tok = regexp(text, '乙醇浓度(\d+\.?\d*)ml/min', 'tokens');
51 if ~isempty(tok), X3(i) = str2double(tok{1}{1}); end
52
53 % 提取 Co 负载量
54 tok = regexp(text, '(\d+\.?\d*)wt%Co/SiO2', 'tokens');
55 if ~isempty(tok), X4(i) = str2double(tok{1}{1}); end
56 end
57
58 %% Step 3: 构建设计矩阵
59 X = [X1, X2, X3, X4, X5];
60
61 %% Step 4: 多元回归
62 mdl1 = fitlm(X, Y1, 'linear');
63 mdl2 = fitlm(X, Y2, 'linear');
64
65 %% Step 5: 可视化
66 figure;
67 subplot(1,2,1); plot(mdl1); title('乙醇转化率 多元回归');
68 subplot(1,2,2); plot(mdl2); title('C4烯烃选择性 多元回归');
69
70 %% Step 6: 导出回归系数表格
71 tbl1 = mdl1.Coefficients; % 包含 Estimate, SE, tStat, pValue
72 tbl2 = mdl2.Coefficients;
73
74 writetable(tbl1, '回归结果_乙醇转化率.xlsx');
75 writetable(tbl2, '回归结果_C4烯烃选择性.xlsx');
76
77 %% Step 7: 打印回归系数分析表
78 disp('==== 乙醇转化率 回归系数分析 =====');
79 disp(tbl1);
80
81 disp('==== C4烯烃选择性 回归系数分析 =====');
82 disp(tbl2);
83
84 %% Step 8: 输出模型指标
85 fprintf('乙醇转化率: R²=%.4f, 调整R²=%.4f, F=%.4f, p=%.4f\n', ...
86         mdl1.Rsquared.Ordinary, mdl1.Rsquared.Adjusted, ...
87         mdl1.ModelFitVsNullModel.Fstat, mdl1.ModelFitVsNullModel.Pvalue);
88
89 fprintf('C4烯烃选择性: R²=%.4f, 调整R²=%.4f, F=%.4f, p=%.4f\n', ...
90         mdl2.Rsquared.Ordinary, mdl2.Rsquared.Adjusted, ...
91         mdl2.ModelFitVsNullModel.Fstat, mdl2.ModelFitVsNullModel.Pvalue);

```

## 附录 C 问题三

```

1
2 % 定义目标函数 (负 Y3, 因为 fmincon 默认最小化)
3 objfun = @(X) -( (-80.526 -0.0337*X(1) +0.14124*X(2) -8.7659*X(3) +0.13327*X(4) +0.33321*X(5)) ...
4               * (-48.732 +0.0029892*X(1) +0.086447*X(2) +2.6734*X(3) -3.1639*X(4) +0.18122*X(5)) );
5
6
7 % 边界
8 lb = [33, 33, 0.5, 0.3, 250];

```

```

9  ub = [200, 200, 5, 2.1, 450];
10
11  % 初始猜测
12  x0 = mean([lb; ub]);
13
14  % 调用 fmincon
15  options = optimoptions('fmincon','Display','iter','Algorithm','sqp');
16  [x_opt, fval] = fmincon(objfun, x0, [], [], [], [], lb, ub, [], options);
17
18  % 输出最大 Y3
19  Y3_max = -fval;
20  Y3_pct = Y3_max / 100;
21  fprintf('最优自变量 X = [%f, %f, %f, %f, %f]\n', x_opt);
22  fprintf('最大 Y3 = %f\n', Y3_pct);

```