

基于乙醇偶合制备 C4 烯烃问题的研究

摘要

针对问题 1 本文建立了一元线性回归分析模型的变量相关关系判别体系，首先通过散点分布图对线性关系进行初步判别，通过 **pearson 相关性检验** 客观分析了温度与乙醇转化率以及 C4 烯烃选择性的线性相关程度。使用 MATLAB 软件编程求解出相应组合的**线性回归方程**，后通过残差分析、**拟合优度检验** 等多个评价体系深入分析回归方程的合理性与拟合程度。分析结果：所有催化剂组合中，温度与乙醇转化率以及 C4 烯烃选择性均呈线性正相关关系。针对附件 2 实验结果分析，本文对乙醇转化率、C4 烯烃选择性以及 C4 烯烃收率分别进行了**曲线拟合**，并分别从制备产物角度以及制备过程角度确定了最佳反应时间节点，分别为**20min**(C4 烯烃收率最高)、**240min**(反应物反应最充分)，并对数据进行了深入解读。

针对问题 2 探究多自变量与多因变量之间的关系，本文首先对自变量进行提取分类(核心变量，**交互变量**，逻辑变量)，然后分别建立**多元线性回归模型**与**PLS 回归模型**。通过对比 MPAE 以及 R 方值分析模型的优劣，最终选定 PLS 回归模型对乙醇转化率以及 C4 烯烃选择性进行探究分析。并得出以下**结论**：Co/SiO₂ 和 HAP 的装料比以及乙醇浓度对乙醇转化率影响非常显著，且均与乙醇转化率呈负相关关系。Co 负载量对乙醇转化率影响显著，为负相关关系。其余自变量对于乙醇转化率影响较小。Co 负载量与催化剂组合质量对 C4 烯烃选择性影响非常显著，呈负相关关系。Co/-SiO₂和 HAP 的装料比对 C4 烯烃选择性影响非常显著，呈正相关关系。温度对 C4 烯烃选择性影响显著，呈正相关关系。

针对问题 3 在第二问的基础上建立**单目标非线性规划优化模型**，选取决策变量、制定目标函数、挖掘约束条件，最终通过**LINGO**进行求解，结果如下：温度上限 450° 时，最佳催化剂组合及温度为：Co 负载量:0.5wt% Co/SiO₂和 HAP 装料比:0.5 乙醇浓度:0.43ml/min Co/SiO₂质量:200mg 温度：450°，此时 C4 烯烃收率为 **4371.051**；温度上限 350° 时，最佳催化剂组合及温度为：Co 负载量:0.5wt% Co/SiO₂和 HAP 装料比:0.5 乙醇浓度:0.38ml/min Co/SiO₂质量:200mg 温度：350°，此时 C4 烯烃收率为 **1792.488**；最后本文对模型进行了**灵敏度分析**，发现在微小变动下，最优解基本未发生变化，模型稳定性强。

针对问题 4 基于本实验实验次数少，多因素多水平的特点，我们选择均匀设计实验。在本题中，选取五个变量中除 Co 负载量以外的其余四个自变量进行实验，共进行四次实验，Co/SiO₂和 HAP 装料比的水平为：0.45、0.65、0.85；乙醇浓度的水平为：0.4、0.6、0.8、1.0；Co/SiO₂质量的水平为：150、200、250、300；温度的水平为：380、410、440、470；第五次实验各自变量取值（按照上方列举顺序）分别为：0.5、0.43、200、450。

关键词：线性回归分析 PLS 回归分析 单目标非线性规划 灵敏度分析

一、问题重述

1.1 问题背景

C4烯烃^[1]对于化工和医药的生产有着十分重要的作用， C_2H_5OH 是制备C4烯烃的原材料。在生产活动中，催化剂组合(组合内容见后文)与温度对C4烯烃的选择性和C4烯烃收率将产生影响。那么通过合理设计催化剂组合方案，进而探索 C_2H_5OH 催化偶合制备C4烯烃工艺条件的实验是十分具有现实意义与实际价值的。

1.2 问题提出

已知某实验室针对不同催化剂在不同温度下做了一系列实验，实验结果存放在附件1、2中。请解决下列问题：

(1) 对附件1中每种催化剂组合，分别研究 C_2H_5OH 转化率、C4烯烃的选择性与温度的关系，并对附件2中350度时给定的催化剂组合在一次实验不同时间的测试结果进行分析。

(2) 探讨不同催化剂组合及温度对 C_2H_5OH 转化率以及C4烯烃选择性大小的影响。

(3) 如何选择催化剂组合与温度，使得在相同实验条件下C4烯烃收率尽可能高。若使温度低于350度，又如何选择催化剂组合与温度，使得C4烯烃收率尽可能高。

(4) 如果允许再增加5次实验，应如何设计，并给出详细理由。

二、问题分析

2.1 针对问题一的分析

通过对问题1解读以及附件内容的分析可知，在固定催化剂组合的情况下（即：将催化剂组合看做常量处理），温度分别与乙醇转化率以及C4烯烃选择性符合某种类型的相关关系。基于本题有催化剂A1、A2……等若干组合，因此本文随机采取某一催化剂组合（以A4组为例），并将温度定义为自变量，乙醇转化率、C4烯烃选择性定义为因变量，研究这两个因变量与自变量温度之间的关系。

首先采用散点图方式对附件1中的相关数据进行可视化处理，以此来简要判别因变量与自变量之间的线性相关程度。然后通过检验是否符合线性回归的基本条件的方式，判别回归方程类型（线性与非线性）并建立回归方程。最后，对已建立的回归方程进行拟合优度检验，符合条件（ $R^2 > 0.6$ 即为符合要求）则证明该方程可以较为理想的体现乙醇转化率、C4烯烃选择性与温度的关系。其余各催化剂组合沿用上述模型即可求解相应回归方程。整体思路如图1

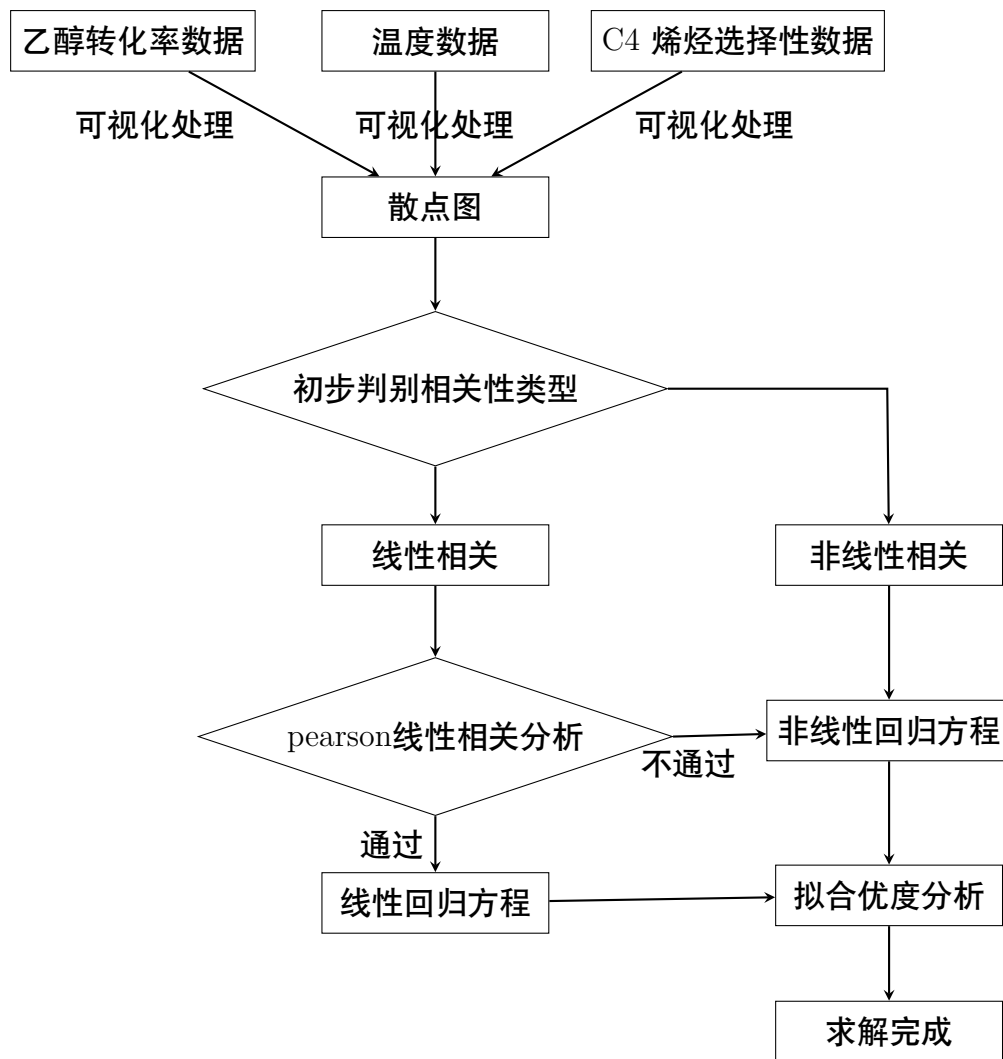


图 1: 问题 1 流程图

2.2 针对问题二的分析

问题 2 实质上是探究多自变量与多因变量之间的关系，本文首先对自变量进行提取分类（核心变量，交互变量，逻辑变量），建立多元线性回归模型，通过 T 检验显著性分析剔除非显著性变量，进行多次回归分析直到显著性理想为止。然后建立对比模型——偏最小二乘回归模型，针对两个模型通过分析 MP AE 以及 R 方值判别模型的优劣，最终选定偏最小二乘回归模型对乙醇转化率以及 C4 烯烃选择性进行分析，解读出各因变量与各自变量的关系。整体思路如图 2

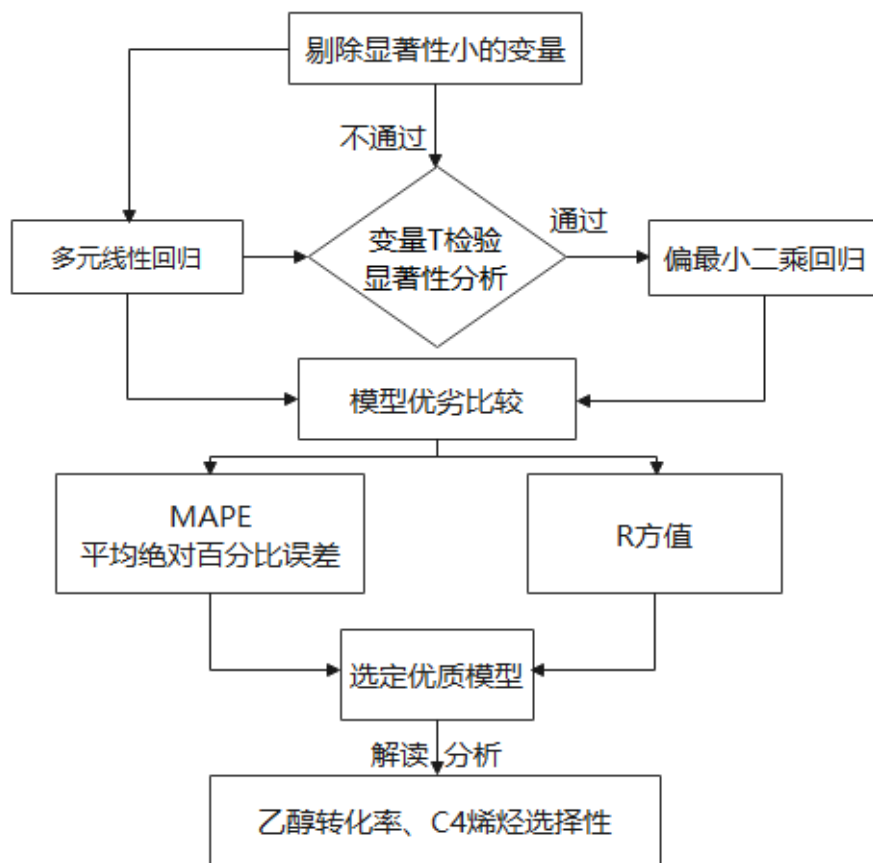


图 2: 问题 2 流程图

2.3 针对问题三的分析

问题三是优化问题，可以直接在问题二结果的基础上建立单目标非线性规划优化模型，选取决策变量、制定目标函数、挖掘约束条件，最终通过 LINGO 进行求解，分别求出在温度上限不同的情况下，所对应的 C4 烯烃收率以及最佳催化剂组合及温度。问题最后，对模型进行灵敏度分析，在最优解附近给予一个微小变动，研究最优的变化幅度与趋势，判断模型的稳定性以及模型最优解的可信度。整体思路如下图 3

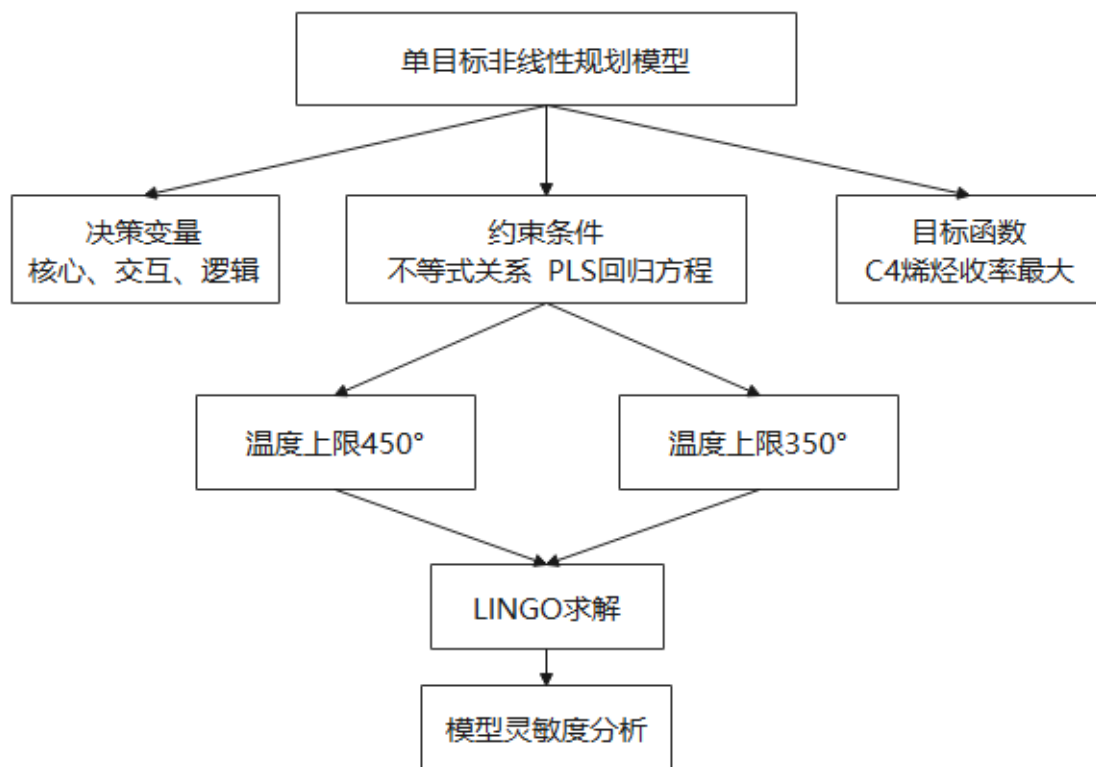


图 3: 问题 3 流程图

2.4 针对问题四的分析

三、基础性模型假设与合理性说明

假设 1: 化学反应均不受除题中出现因素之外的其他外界因素影响;

假设 2: 各反应生成物之间有较强稳定性, 不会进行下一步反应

假设 3:

四、变量说明

符号	符号说明
x_0	温度
y_1	乙醇转化率回归方程
y_2	C ₄ 烯烃选择性回归方程
y_3	乙烯选择性回归方程
y_4	乙醛选择性回归方程
y_5	碳数为 4-12 脂肪醇选择性回归方程
y_6	甲基苯甲醛和甲基苯甲醇选择性回归方程

注: 未说明的变量以其在符号出现处的具体说明为准。

五、基于线性回归方程的相关性分析模型——问题一

5.1 模型建立

5.1.1 一元线性回归模型原理

中心思想：

一元线性回归分析是在理想条件下（忽略其他因素），分析某一个因素（自变量）对另外的一个因素（因变量）所产生的影响的过程。因为因变量、自变量的个数均为一个，所以可以认为 y 随着 x 变化，每一个已知的 x 都会有一个已知的 y 值，而一元线性回归就是要求解出一条直线，它不仅能够使得每一个已知的 x 都对应一个直线上的预测值 y' ，还使得每个实际的 y 值与预测值 y' 之差的平方和最小，即达到一元线性回归的最终结果。

数学表达：

一元线性回归方程如下：

$$y = \beta_0 + \beta_1 x_0 + \epsilon \quad (1)$$

β_0 : 常数, β_1 : 回归系数, x : 回归变量, ϵ : 随机变量

因为 $E(\epsilon) = 0$, 所以方程可以简化为：

$$y = \beta_0 + \beta_1 x_0 \quad (2)$$

β_0, β_1 可以利用最小二乘法得到，估计公式为：

$$\begin{cases} \beta_0 = \bar{y} - \bar{x}_0 \beta_1 \\ \beta_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})} \end{cases} \quad (3)$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n X_i \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

5.1.2 数据预处理

数据可视化：

以 A4 组为例，分别将乙醇转化率、C4 烯烃选择性数据与温度数据进行可视化处理得到相应的散点分布图，通过散点图的数据分布对乙醇转化率、C4 烯烃选择性与温度之间的模型关系有一个初步的认识，如下图 4

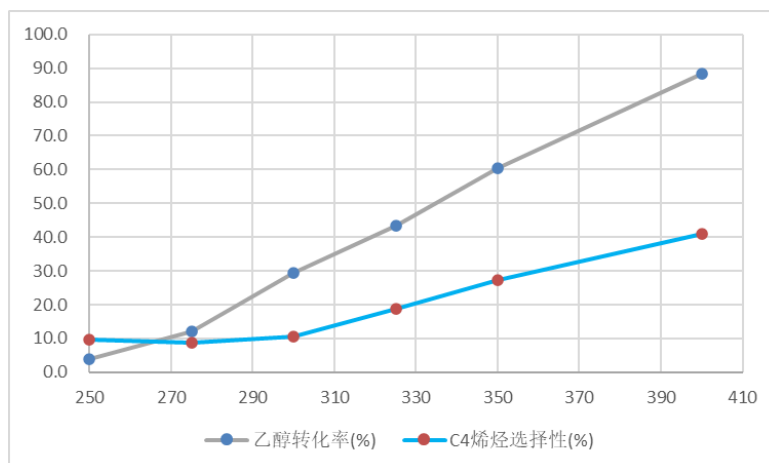


图 4: 温度与乙醇转化率以及 C4 烯烃选择性的散点图

由上图可以看出，乙醇转化率与 C4 烯烃选择性均与温度呈线性关系，且为正相关。那么就可以进行线性回归分析，为确保结果的严谨性与合理性，本文在进行线性回归分析前，对自变量与因变量进行相关性分析检验。

线性相关分析：

通过 SPSS 软件处理，得到自变量温度与乙醇转化率以及 C4 烯烃选择性的线性相关程度。如下表 2。由表 2 可以看到，温度对乙醇转化率以及 C4 烯烃选择性有着极强的线性相关程度（最小相关性为 0.943）。那么 A4 组的乙醇转化率以及 C4 烯烃选择性与温度的关系通过了线性相关检验，可以进行线性回归分析。

表 2: 线性相关分析结果

分组	相关性系数	
	pearson	spearman
乙醇转化率与温度	0.998**	1.000**
C4 烯烃选择性与温度	0.958**	0.943**

** 在 0.01 级别（双尾），相关性显著。

5.2 模型求解

5.2.1 线性回归方程求解

综上所述，本文通过 Matlab 进行合理分析（残差分析、误差分析见附录）后得到了 y_1 、 y_2 的具体数学形式：

$$y_1 = -144.6 + 0.5817x_0 \quad (4)$$

$$y_2 = -52.41 + 0.2266x_0 \quad (5)$$

y_1 : 乙醇转化率 y_2 : C4 烯烃选择性 x : 温度

5.2.2 回归方程拟合优度检验

为了验证上述线性关系 (y_1 、 y_2) 的正确性, 本问采用 SPSS 软件进行线性回归分析, 如表 3。由表得 y_1 、 y_2 的 R 方值分别为 0.995、0.9173, 远大于 0.6, 充分说明在上述的回归方程中, 温度变量可以较为理想的解释乙醇转化率以及 C4 烯烃选择性的变化趋势。

表 3: R 方值

分组	R 方
乙醇转化率和温度 (y_1)	0.995
C4 烯烃选择性和温度 (y_2)	0.9173

5.3 结果呈现

上文以 A4 为例进行建模求解得出了相应的回归方程, 现在对所有催化剂组合进行线性回归分析处理得到其他回归方程如下表 4(部分结果的呈现, 详见附录)。

表 4: 回归方程

组合	乙醇转化率和温度	R 方	C4 烯烃选择性和温度	R 方
A1	$y_1=0.3332x_0-84.08$	0.9322	$y_2=0.1544x_0-3.242$	0.7869
A2	$y_1=0.663x_0-161.9$	0.99	$y_2=0.2216x_0-41.55$	0.8359
A3	$y_1=0.4196x_0-95.88$	0.9643	$y_2=0.2612x_0-59.19$	0.9128
A4	$y_1=0.5817x_0-144.6$	0.995	$y_2=0.2266x_0-52.41$	0.9173
A5	$y_1=0.4079x_0-97.62$	0.8734	$y_2=0.2297x_0-57.81$	0.9401
A6	$y_1=0.5015x_0-119.8$	0.9674	$y_2=0.2027x_0-50.75$	0.784
A7	$y_1=0.3775x_0-74.26$	0.9988	$y_2=0.1869x_0-44.26$	0.9374
...

5.4 回归方程结果分析

上表中所有的回归方程均是严格按照模型步骤进行求解所得, 已经进行过 pearson 相关性分析、残差分析、误差分析、拟合优度检验。因此在结果分析处不在进行重复操作。

本文通过分析所有的回归方程发现, 所有的 y_1 、 y_2 均为线性回归方程, 且乙醇转化率与 C4 烯烃选择性均与温度呈正比关系。

其中针对乙醇转化率, 受温度影响最高的组合是 A2, 回归系数为 0.663。拟合优度最高的是 A7 组合, R 方值为 0.9988; 针对 C4 烯烃选择性受温度影响最高的组合是

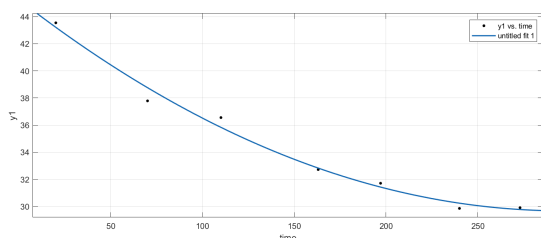
A3, 回归系数为 0.2612。拟合优度最高的是 A9 组合, R 方值为 0.9948; 综合对比发现, 回归方程的拟合优度最低为 0.7421>0.7, 整体回归方程均保持高度拟合, 温度对乙醇转化率以及 C4 烯烃选择性拥有较高可信度的解释性, 结果比较理想。

5.5 测试结果分析

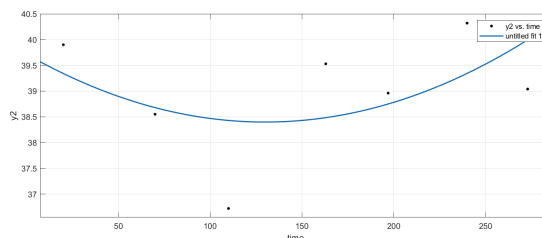
针对附件 2 所给的数据, 本文分别对乙醇转化率、C4 烯烃选择性进行拟合, 结果如下图 5(a)、5(b)。其中对应拟合方程分别为:

$$y_1 = 0.0001772x_0^2 - 0.1051x_0 + 45.25 \quad (6)$$

$$y_2 = 7.798 \times 10^{-5}x_0^2 - 0.02027x_0 + 39.72 \quad (7)$$



(a) 乙醇转化率的拟合曲线



(b) C4 烯烃的拟合曲线

图 5: 拟合曲线

结合方程以及拟合曲线可得, 随着时间推移乙醇转化率呈下降趋势, C4 烯烃选择性呈先降低后升高的趋势。考虑到实验最终目的是制备 C4 烯烃, 所以本文引入 C4 烯烃收率的定义, 对 C4 烯烃收率进行上述操作得下图 6。拟合方程为:

$$y = 0.00998x_0^2 - 4.935x_0 + 1798 \quad (8)$$

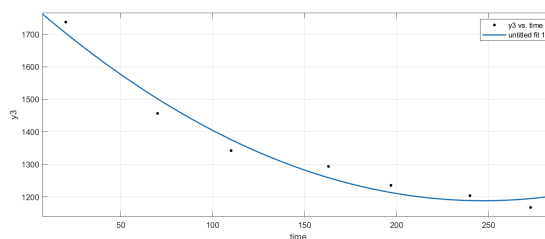


图 6: C4 烯烃收率的拟合曲线

综上, 本文对测试结果解读如下: 针对制备产物角度, 在所给时间节点里, 最佳的反应时间节点为 20min 此时 C4 烯烃收率最大 (1737.541)。针对制备过程角度, 为避免实验原材料的浪费, 应可能的选择后面的时间节点, 因此 240min 时停止实验比较理想, 此时可以最大化的利用乙醇进行 C4 烯烃的制备, 并且该时段 C4 烯烃选择性最大, 占比也高。

六、基于多元线性回归与 PLS 回归的多变量关系模型——问题二

6.1 建模准备

6.1.1 多元线性回归^[2] 模型原理

中心思想: 由于问题一已经对线性回归（一元线性回归）的中心思想进行了详细的阐述，此处不再赘述，直接进入数学公式的表达环节。

数学表达: 多元线性回归方程结构如下：

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad (9)$$

β_0 : 常数 β_1, \dots, β_k : 回归系数 x_1, \dots, x_k : 回归变量 ε : 随机变量

又因为 $E(\varepsilon) = 0$ ，所以方程可化简为：

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (10)$$

其中回归系数采用最小二乘法进行参数估计，处理方式与一元线性回归同理，此处不再赘述。

6.1.2 PLS回归^[3] 模型原理

中心思想: PLS 回归的中心思想是首先在自变量集合中提出第一成分 g_1 (g_1 是 x_1, \dots, x_m 的线性组合)；同理在因变量集合中也提取第一成分 u_1 ，且满足 g_1 与 u_1 相关程度达到最大的要求。然后建立多因变量与 g_1 的回归，若回归方程已达到理想的精度，则中止算法。否则继续上述步骤（即提取第二对成分），知道方程满足所需精度为止。设最终对自变量集提取 z 个中间变量 t_1, t_2, \dots, t_z ，则建立多因变量与中间变量的回归式，然后进行数学表达变换得到多因变量与原自变量的回归方程式，即偏最小二乘回归方程式。

数学表达: 由 PLS 的中心思想可知，问题的关键在于求解 g_1 和 u_1 的协方差最大，即求解下面的最优。

$$\max\{\text{Cov}(g_1, u_1)\} = \max \langle E_0 w_1, F_0 c_1 \rangle \quad (11)$$

$$\text{s.t.} \begin{cases} w_1^T w_1 = 1 \\ c_1^T c_1 = 1 \end{cases} \quad (12)$$

利用拉格朗日乘数法求出 w_1 和 c_1

$$\begin{cases} E_0^T F_0 F_0^T E_0 w_1 = \theta_1^2 w_1 \\ F_0^T E_0 E_0^T F_0 c_1 = \theta_1^2 c_1 \end{cases} \quad (13)$$

E_0, F_0 : X 与 Y 的标准化数据 w_1 : $E_0^T F_0 F_0^T E_0$ 的单位特征向量 θ_1^2 : $F_0^T E_0 E_0^T F_0 c_1$ 的最大特征值 c_1 : θ_1^2 的单位特征向量。

求出 w_1 和 c_1 即可得到成分 g_1 和 u_1 , 然后分别求 E_0 和 F_0 对 g_1 的的回归方程

$$\begin{aligned} E_0 &= g_1 p_1^T + E_1 \\ F_0 &= g_1 r_1^T + F_1 \end{aligned} \quad (14)$$

$p_1 = \frac{E_0^T t_1}{\|t_1\|^2}$ $r_1 = \frac{F_0^T t_1}{\|t_1\|^2}$ E_1, F_1 : 回归方程的残差矩阵。

则由 E_1, F_1 , 可求解 w_2 和 c_2 以及第二个主成分 g_2, u_2 有 $g_2 = E_1 w_2, u_2 = F_1 c_2$

建立回归方程:

$$E_1 = g_2 p_2^T + E_2 F_1 = g_2 r_2^T + F_2 \quad (15)$$

以此类推即可得

$$F_0 = g_1 r_1^T + \dots + g_m r_m^T + F_m = E_0 \left[\sum_{j=1}^m w_j^* r_j^T \right] + F_m \quad (16)$$

$w_j^* = \prod_{i=1}^{j-1} (I - w_i p_i^T) w_j$, $\sum_{j=1}^A w_j^* r_j^T$: PLS 回归系数向量 A : X 的秩
变换后即可得 n 个因变量的 PLS 回归方程式:

$$y_j = c_{j1} x_1 + \dots + c_{jm} x_m, j = 1, 2, \dots, n \quad (17)$$

6.1.3 自变量选取

通过对实验数据的全局解读, 本文将自变量分为三种, 如下表 5。其中自变量为催化剂以及温度, 属于此项化学实验项目的探究核心内容, 对制备 C4 烯烃影响最大; 考虑到化学实验经常涉及各因素之间的交互影响, 本文定义了三种交互变量以探究交互变量对 C4 制备的影响; 0-1 自变量即逻辑变量, 探讨装料方式的不同对 C4 烯烃制备的影响。

表 5: 自变量定义

自变量类型	数学表达	自变量说明
核心自变量	x_1	C_o 负载量
	x_2	C_o/SiO_2 和 HAP 的装料比
	x_3	乙醇浓度
	x_4	催化剂组合质量
	x_5	温度
交互自变量	$x_3 x_5$	乙醇流速与温度的交互影响
	x_4 / x_3	催化剂组合质量与乙醇流速的交互影响
	x_3^2	乙醇浓度的平方
0-1 自变量	x_6	装料方式, 1 对应方式 A, 0 对应方式 B

6.1.4 因变量选取

由于实验主要探究的对象是乙醇转化率以 C4 烯烃选择性，故本文将这两种因变量作为 y_1 、 y_2 ，作为重点去研究分析。其余因变量作为 $y_3, y_4 \dots$ 等，此处对于次要因变量不再详细展开，具体内容见附录。

6.2 模型建立与求解

6.2.1 多元线性回归模型

将前文已经提炼出来的自变量与因变量数据导入到 SPSS 中进行初次多元线性回归，得到了相应的参数估计值以及 T 检验显著性数据（表格见附录），分析发现其中的交互自变量 x_3^2 以及逻辑变量 x_6 在 y_1 、 y_2 回归分析中没有通过 T 检验，因此将上述逻辑自变量剔除后进行第二次多元线性回归分析，结果如下表 6(以 y_1 为例，其余因变量的相关数据见附录)。可以看到其余自变量均通过了 T 检验显著性分析（考虑到 $x_1 - x_5$ 为核心自变量，此处不对其进行 T 检验显著性分析）

表 6: 参数估计

参数	未标准化系数	t	显著性
常数	-103.701	-5.806	0.000
x_1	0.886	1.025	0.308
x_2	-3.261	-0.878	0.382
x_3	9.219	0.820	0.414
x_4	0.143	8.117	0.000
x_5	0.451	8.510	0.000
$x_3 * x_5$	-0.081	-2.367	0.020
x_4 / x_3	-0.043	-3.645	0.000

则可求得对应的回归方程:

$$y_1 = -103.701 + 0.886x_1 - 3.261x_2 + 9.291x_3 + 0.143x_4 + 0.451x_5 - 0.081x_3x_5 - 0.043x_4/x_3 \quad (18)$$

同理可以求得 y_2 的回归方程:

$$y_2 = -72.315 - 2.720x_1 + 4.684x_2 + 15.619x_3 + 0.107x_4 + 0.260x_5 - 0.054x_3x_5 - 0.023x_4/x_3 \quad (19)$$

6.2.2 多个因变量的多元线性回归模型

将已提炼出的自变量与因变量做成 EXCEL 数据表，使用 MATLAB 软件编程得到相关系数矩阵如下表 7(篇幅受限，这里只展示部分数据。)从相关系数矩阵中，可以看出因变量与自变量之间的相关程度以及相关关系，可以对回归方程进行简单的预估。

表 7: 相关系数矩阵

	x1	x2	x3	x4	x5	...	y6
x1	1.0000	-0.0035	0.1413	0.2089	-0.0142	...	-0.2252
x2	-0.0035	1.0000	0.0037	0.0570	-0.0001	...	0.1840
x3	0.1413	0.0037	1.0000	-0.2992	-0.0385	...	-0.0814
x4	0.2089	0.0570	-0.2992	1.0000	-0.0018	...	0.1290
x5	-0.0142	-0.0001	-0.0385	-0.0018	1.0000	...	0.2695
x3*x5	0.1226	0.0034	0.8940	-0.2806	0.3869	...	0.0437
x4/x3	0.1143	0.0134	-0.7210	0.6120	0.0204	...	0.1273
y1	0.0418	-0.0164	-0.3312	0.3946	0.7706	...	0.2858
...
y6	-0.2252	0.1840	-0.0814	0.1290	0.2695	...	1.0000

由于 y_7 满足一个恒等式:

$$y_2 + y_3 + y_4 + y_5 + y_6 + y_7 = 100 \quad (20)$$

所以不在对 y_7 进行单独求解。

下列各式即为 MATLAB 编程所得回归方程 (重点分析 y_1, y_2 , 其余方程见附录):

$$y_1 = -65.8743 - 1.7114x_1 - 9.9651x_2 - 9.1408x_3 + 0.1144x_4 + 0.3139x_5 + 0.0068x_3x_5 + 0.0044x_4/x_3 \quad (21)$$

$$y_2 = -42.7073 - 3.3831x_1 + 3.7492x_2 - 3.2979x_3 + 0.0678x_4 + 0.1747x_5 + 0.0082x_3x_5 - 0.0048x_4/x_3 \quad (22)$$

6.3 模型对比

本文分别通过多元线性回归以及偏最小二乘回归对实验产物进行了分析, 并求出了相应的回归方程。为保证对实验结果 (因变量) 最好的回归分析, 需要对两种模型所求得的方程进行对比。

本文选取了两种指标进行判别模型的好坏, 如下表 8 所示, 由表可以看出, 偏最小二乘回归的 $MAPE$ 值以及 R^2 值均优于多元线性回归。说明相较于多元线性回归而言, 偏最小二乘回归对样本数据的解释性更强, 预测可信度更高, 因此本文采用偏最小二乘回归模型进行解释分析。

表 8: 指标数据

模型	因变量	$MAPE$	R^2
偏最小二乘回归	y_1	1.8375	0.821
	y_2	0.6415	0.721
多元线性回归	y_1	2.1085	0.7681

y_2	0.7405	0.6888
<hr/>		
MAPE:平均绝对百分比误差	R^2 : 拟合优度	

6.4 结果分析

本文最终选取偏最小二乘回归模型，下表 9 为从回归方程中提取出的系数阵，由表可得：

Co/SiO₂ 和 HAP 的装料比以及乙醇浓度对乙醇转化率影响非常显著，且均与乙醇转化率呈负相关关系。C₀ 负载量对乙醇转化率影响显著，为负相关关系。其余自变量对于乙醇转化率影响较小。

Co 负载量与催化剂组合质量对 C₄ 烯烃选择性影响非常显著，呈负相关关系。Co/SiO₂ 和 HAP 的装料比对 C₄ 烯烃选择性影响非常显著，呈正相关关系。温度对 C₄ 烯烃选择性影响显著，呈正相关关系。

表 9: 系数矩阵

x_i	y_1	y_2
Co 负载量	-1.7114*	-3.3831**
Co/SiO ₂ 和 HAP 的装料比	-9.9651**	3.7492**
乙醇浓度	-9.1408**	0.0678
催化剂组合质量	0.1144	-3.2979**
温度	0.3139	0.1747*
乙醇浓度与温度的交互影响	0.0068	0.0082
催化剂组合质量与乙醇浓度的交互影响	0.0044	-0.0048

*: 显著性强 **: 显著性非常强

综上所述，要想使得乙醇转化率尽可能的大，则需要尽量减小 Co/SiO₂ 和 HAP 的装料比、乙醇浓度以及 C₀ 负载量的值，增大其余变量值；要想使得 C₄ 烯烃选择性尽可能的大，则需要尽量减小 Co 负载量与催化剂组合质量的值，增大 Co/SiO₂ 和 HAP 的装料比的值。

综合分析，针对 y_1 、 y_2 ，自变量中既有同性相关变量（自变量对 y_1 、 y_2 产生的影响相同），也有异性相关变量。则说明存在一组自变量矩阵使得 y_1 、 y_2 达到整体理想值。

七、基于单目标非线性规划的优化类模型——问题三

7.1 建模准备

7.1.1 非线性规划^[4]模型

非线性规划中心思想：基于解决目标函数或者约束条件中至少有一个是非线性函数

时的最优化问题即为非线性规划问题。非线性规划问题的求解思路同线性相同，即通过建立目标函数、决策变量、约束条件的关系，去求解最优目标函数。

数学表达:

非线性规划优化问题，首先需要定义决策变量：

$$x_1, x_2, x_3 \dots x_i \quad (23)$$

然后结合决策变量去解读实际问题，并从中提取建立目标函数：

$$\max = f(x) \quad (24)$$

深入分析决策变量与目标函数的关系，找出其中的约束条件（等式或不等式关系）：

$$\text{s.t.} \begin{cases} A \cdot x \leq b \\ A_{\text{eq}} \cdot x = b_{\text{eq}} \\ c(x) \leq 0 \\ \text{ceq}(x) = 0 \\ \text{lb} \leq x \leq \text{ub} \end{cases} \quad (25)$$

最后通过 MATLAB 编程或者 LINGO 编程求解即可。

7.1.2 定义决策变量

问题三实在问题二的基础上进行的一部延伸，因此基于此题的决策变量只需要选取前文回归方程中的自变量即可。如下表 10

表 10: 决策变量表

决策变量	决策变量说明
x_1	C_o 负载量
x_2	C_o/SiO_2 和HAP的装料比
x_3	乙醇浓度
x_4	催化剂组合质量
x_5	温度
x_3x_5	乙醇浓度与温度的交互影响
x_4/x_3	催化剂组合质量与乙醇浓度的交互影响

7.1.3 建立目标函数

题目要求探究 C4 烯烃收率与决策变量之间的关系，C4 烯烃收率数学表达式如下：

$$Z = y_1 * y_2 \quad (26)$$

那么关于 C4 烯烃收率最优问题即可转换为求解 Z_{max} 问题，即目标函数为：

$$Z_{max} = y_1 * y_2 \quad (27)$$

7.1.4 挖掘约束条件

由已给数据（附件 1）能够等到一个不等式约束，即

$$y_2 + y_3 + u_4 + y_5 + y_6 \leq 100 \quad (28)$$

显然只仅此一个约束条件，无法求解最优值，因此需要从 21 中催化剂组合共 114 组实验的实验数据中挖掘潜在约束条件。通过对 114 组数据进行分析发现，决策变量可分为催化剂类变量以及温度变量，其中催化剂类变量也可进行细分，每种变量所对应的变量值均是分类变量，即离散变量。那么就可以针对决策变量进行值域的确定，即可得到每个变量上下限取值。即

$$\begin{cases} 0.5 \leq x_1 \leq 5 \\ 0.5 \leq x_2 \leq 2 \\ 0.3 \leq x_3 \leq 2.1 \\ 10 \leq x_4 \leq 200 \\ 250 \leq x_5 \leq 450 \end{cases} \quad (29)$$

同时通过对化学反应的产物指标分析易得，对于实际化学实验中的百分比数值一定遵循实验反应准则，即所有因变量的值均应该小于等于 100。即

$$0 \leq y_i \leq 100 \quad i = 1, 2, \dots, 6 \quad (30)$$

7.2 模型建立与求解

由上述分析可直接建立单目标非线性规划模型。

7.2.1 温度上限 450° 时模型求解

目标函数：

$$Z_{max} = y_1 * y_2 \quad (31)$$

约束条件:

$$\text{st.} \left\{ \begin{array}{l} y1 = -65.8743 - 1.7114x1 - 9.9651x2 - 9.1408x3 \\ \quad + 0.1144x4 + 0.3139x5 + 0.0068x3x5 + 0.0044x4/x3 \\ y2 = -42.7073 - 3.3831x1 + 3.7492x2 - 3.2979x3 \\ \quad + 0.0678x4 + 0.1747x5 + 0.0082x3x5 - 0.0048x4/x3 \\ y3 = -8.2906 - 0.0772x1 - 1.4555x2 - 0.4917x3 \\ \quad + 0.0088x4 + 0.0340x5 + 0.0023x3x5 + 0.00021773x4/x3 \\ y4 = 15.6469 + 1.6917x1 - 15.5341x2 + 0.5485x3 \\ \quad - 0.0107x4 + 0.0297x5 + 0.0032x3x5 - 0.0095x4/x3 \\ y5 = 153.9014 + 3.2638 * x1 + 9.4221x2 + 4.3213x3 \\ \quad - 0.0941x4 - 0.3154x5 - 0.0192x3x5 + 0.0130x4/x3 \\ y6 = -2.8899 - 0.4839x1 + 2.0813x2 - 0.1618x3 \\ \quad + 0.0060x4 + 0.0120x5 + 0.00093335x3x5 + 0.00046965x4/x3 \end{array} \right. \quad (32)$$

$$\text{st.} \left\{ \begin{array}{l} y2 + y3 + y4 + y5 + y6 \leq 100 \\ 0 \leq y_i \leq 100 \quad i = 1, 2, \dots, 6 \\ 0.5 \leq x_1 \leq 5 \\ 0.5 \leq x_2 \leq 2 \\ 0.3 \leq x_3 \leq 2.1 \\ 10 \leq x_4 \leq 200 \\ 250 \leq x_5 \leq 450 \end{array} \right. \quad (33)$$

将上述数学表达转化为 LINGO 语言，即可得到最优解。如下表 11。

表 11: 最优解数据

全局最优解	变量	变量值	变量	变量值
4371.051	y1	91.87873	x5	450.00000
	y2	47.57414	y3	8.337247
	x1	0.50000	y4	16.35936
	x2	0.50000	y5	3.725449
	x3	0.4277443	y6	4.838841
	x4	200.00000		

则求得最佳催化剂组合及温度为：

$$\left\{ \begin{array}{l} \text{Co负载量: } 0.5wt\% \\ \text{Co/SiO}_2\text{和HAP装料比:}0.5 \\ \text{乙醇浓度: } 0.43ml/min \\ \text{Co/SiO}_2\text{质量: } 200mg \\ \text{温度: } 450^\circ \end{array} \right. \quad (34)$$

7.2.2 温度上限 350° 时模型求解

针对问题三后半部分，可以等价于是改变温度约束条件的非线性规划。即将前文的温度约束上限改为 350° 即可，由此可得最优解如下表 12。由图表可得

表 12: 350° 最优解数据

全局最优解	变量	变量值	变量	变量值
1792.488	y1	60.78920	x5	350.00000
	y2	29.48695	y3	4.836690
	x1	0.50000	y4	12.60311
	x2	0.50000	y5	36.98095
	x3	0.3792175	y6	3.619017
	x4	200.00000		

则求得温度上限位 350° 的最佳催化剂组合及温度为：

$$\left\{ \begin{array}{l} \text{Co负载量: } 0.5wt\% \\ \text{Co/SiO}_2\text{和HAP装料比:}0.5 \\ \text{乙醇浓度: } 0.38ml/min \\ \text{Co/SiO}_2\text{质量: } 200mg \\ \text{温度: } 350^\circ \end{array} \right. \quad (35)$$

7.3 模型灵敏度检验

为探究模型的稳定性，本文对 PLS 回归模型进行灵敏度分析，在以最优解为中心的小范围内给予微小变动，观察对应的最优解的变化幅度。如图 8。可以看出，随着参数值的逐渐减小，最优解呈上升趋势，说明参数与最优解保持局部的线性关系。又因为最优解变化量小，连续 7 次的参数值变动（减小），只引起最优解 2.854 的数值变动，由此可说明，此模型保持高稳定性，因此最优解的可信度较高。

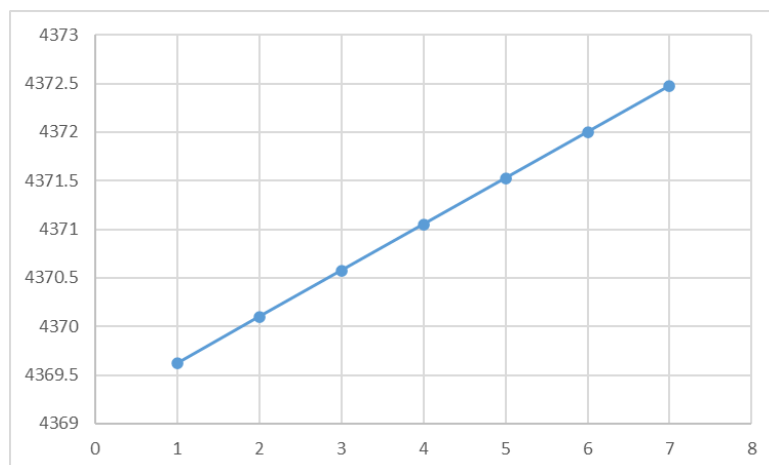


图 7: 最优解趋势

八、均匀设计实验——问题四

8.1 均匀设计实验简介

实验原理: 均匀设计实验是一种只考虑试验点在试验范围内均匀散布的一种试验设计方法。此试验法由方开泰教授提出，是一种将数论方法与多元统计试验相结合的创新试验设计法。最重要原则之一为均匀性原则。均匀设计的实验原理为一致分布原理，属于伪蒙特卡罗数学方法的范畴。此实验使每个因素的每个水平仅通过一次实验来完成，并通过均匀设计表选择适当的属性来安排实验。

均匀设计表介绍及使用方法: 均匀设计表是一种利用同余法建立的规格化列表，其表示形式为: $U_n(m^r)$ 其中, U 为均匀表的代号, n 表示实验次数, 在表中即为行数; m 表示因素的水平数, r 表示因素数。 U^* 则表示此类设计表均匀性更佳, 应优先选择。在这里, 行与列的构造原则不再详细说明。下图 9(a)为 $U_5(5^4)$ 的均匀设计表^[6]及使用表:

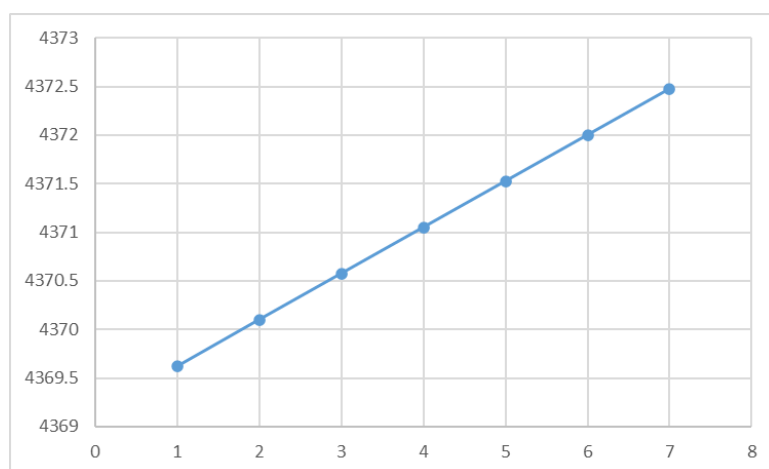


图 8: 最优解趋势

8.2 均匀设计实验的适用性

试验特点分析 本问设计的五组实验需要考虑到四个变量因素的影响，属于多因素多水平实验。在设计实验时，在保证少量实验次数的前提下，确保所设计实验具有普适性与灵活性^[5]，能够更加科学的反应多个水平因素的影响。而均匀设计实验能够在选取试验点数时具有灵活多样性的特点并能较为均匀客观的把设计变量点散布在试验区域内，因此我们选择均匀设计试验。

正交设计实验与均匀设计实验对比

1. 均匀设计实验实验水平数即为实验次数，正交设计实验至少为水平数的平方次。

均匀设计实验能预测实验最佳效果时各因素水平数值，正交设计主要计算因素主效应和交互效应。可以归纳出：正交设计实验的主要特点为其实验范围为全体试验，实验量大，实验次数多，虽具有一定客观性，但不利于本题化工实验的设计，而均匀设计实验特点是影响因素及设计的水平数量多，试验点的代表性强，在已有的试验范围内最大幅度的减少实验次数来获得预期的实验结果，最大的优点是通过以上分析表明，均匀实验设计最为符合本次实验的设计原则与条件。

8.3 实验设计过程

Step1: 自变量的选取 在设计实验的过程中，共可以设计五次实验，因此本文采用均匀设计实验法，所设计的表记为 $U_5(5^4)$ ，在此过程中，由于设计方法原理限制，只能选取 4 个自变量来用于设计均匀实验，而我们共有 5 个自变量，因此要从 5 个自变量中挑选出 4 个自变量，由于在实验中 Co 负载量为 1wt% 的实验占总实验次数的 87/115，所以我们决定将 Co 负载量定值处理为 1wt%，之后对剩下的 4 个自变量进行均匀设计实验。

Step2: 不同自变量水平的选取 由第三问中得出基于 5 个决策变量的全局最优解分别为 Co 负载量: 0.5wt%、Co/SiO₂ 和 HAP 装料比:0.5、乙醇浓度: 0.43ml/min、Co/SiO₂ 质量: 200mg、温度: 450°C，而从所做过的 114 次实验中得出基于 5 个决策变量的局部最优解为: Co 负载量: 1wt%、Co/SiO₂ 和 HAP 装料比:1、乙醇浓度: 0.9ml/min、Co/SiO₂ 质量: 200mg、温度: 400°C，由于我们不再考虑 Co 负载量这个自变量，因此选取 Co/SiO₂ 和 HAP 装料比的 4 个水平为:0.45、0.65、0.85、1.05；乙醇浓度的 4 个水平为:0.4 ml/min、0.6 ml/min、0.8 ml/min、1.0 ml/min；Co/SiO₂ 质量的 4 个水平为:150 mg、200 mg、250 mg、300 mg；温度的 4 个水平为:380°C、410°C、440°C、470°C；

8.4 实验设计方案

由于 Co/SiO₂ 和 HAP 装料比的全局最优解和局部最优解分别是 0.5 和 1.0，因此设计中间两个水平位于 0.5 到 1 之间，然后另外两个水平分别大于 1 和小于 0.5 以确保实验的严谨。乙醇浓度和温度采取和 Co/SiO₂ 和 HAP 装料比的一样的水平选取方法。对于 Co/SiO₂ 质量的水平选择，我们已知全局最优解和局部最优解均为 200mg，因此

选取一个水平为 200mg，另外选取一个水平小于 200mg，由于之前没有做过 Co/SiO₂ 质量大于 200mg 的实验，因此剩余两个水平用来探索 Co/SiO₂ 质量大于 200mg。这样设计四组实验，再加上全局最优解：Co 负载量：0.5wt%、Co/SiO₂ 和 HAP 装料比:0.5、乙醇浓度：0.43ml/min、Co/SiO₂ 质量：200mg、温度：450°C，共五组实验。

九、模型优缺点及推广

9.1 模型优点

(1) 一元线性回归分析模型：此模型在问题一中能够明了直观地表示乙烯转化率和 C₄ 烯烃与温度之间的关系，将数据进行了可视化处理，帮助更好的分析两者与温度的关系。

(2) 多个因变量的多元线性回归模型：此模型将问题二所给出的多个因变量与多个自变量很好的结合，充分考虑了各变量之间的关系，能够较为客观地表现出多变量之间的线性关系，呈现出其密切程度，得出变量间唯一值，使题目的分析更为清晰。

(3) 非线性规划优化模型：非线性规划在第三问中最大化的应用题目所给条件，其约束条件在最大程度上考虑多方面因素，灵活性强，使所得最优解客观性与科学性较强。本问中 C₄ 烯烃收率结果较为理想。此模型是决策系统的静态最优化数学规划模型规划方法之一

9.2 模型缺点

(1) 一元线性回归模型：模型具有一定局限性，只有当两变量呈直线关系时才能较为准确的表达变量之间的关系，对数据进行回归的精度较差。

(2) 多个因变量的多元线性回归模型：回归分析中多种变量的关系表达存在一定不确定性与不可测性，在某些条件下可能受限。

9.3 模型推广

(1) 多因素的多元线性回归模型可应用在多领域的行情评估及预测未来发展趋势方面；

(2) 非线性规划模型：非线性规划与优化控制算法的结合应用；非线性规划模型与遗传算法的结合。

十、参考文献

- [1] 吕绍沛. 乙醇偶合制备丁醇及 C₄ 烯烃 [D]. 大连理工大学,2018.
- [2] 李小明, 王申令. 带线性约束的多元线性回归模型参数估计 [J]. 统计研究,2016,33(11):85-92.

- [3] 陆洪涛. 偏最小二乘回归数学模型及其算法研究 [D]. 华北电力大学,2014.
- [4] 汤龙. 多参数非线性优化方法关键技术研究及应用 [D]. 湖南大学,2013.
- [5] 寇林元. 均匀试验设计效果评价及优化分析方法研究 [D]. 山西医科大学,2009.
- [6] 方开泰. 均匀设计及其应用 [J]. 数理统计与管理,1994(01):57-63+56.

十一、附录

问题一：

图表支撑材料：

表 13: A1-B7 全部回归方程

组合	乙醇转化率和温度	R 方	C4 烯烃选择性和温度	R 方
A1	$y_1=0.3332x -84.08$	0.9322	$y_2= 0.1544x -3.242$	0.7869
A2	$y_1= 0.663x -161.9$	0.99	$y_2= 0.2216x -41.55$	0.8359
A3	$y_1= 0.4196x -95.88$	0.9643	$y_2= 0.2612x -59.19$	0.9128
A4	$y_1= 0.5817x -144.6$	0.995	$y_2= 0.2266x -52.41$	0.9173
A5	$y_1= 0.4079x -97.62$	0.8734	$y_2= 0.2297x -57.81$	0.9401
A6	$y_1= 0.5015x -119.8$	0.9674	$y_2= 0.2027x -50.75$	0.784
A7	$y_1= 0.3775x -74.26$	0.9988	$y_2= 0.1869x -44.26$	0.9374
A8	$y_1= 0.3396x -83.78$	0.9548	$y_2= 0.2423x -57.26$	0.9832
A9	$y_1= 0.2491x -65.67$	0.8482	$y_2= 0.2538x -59.1$	0.9948
A10	$y_1= 0.1834x -49.65$	0.8517	$y_2= 0.05167x -12.36$	0.7421
A11	$y_1= 0.2068x -56.5$	0.8155	$y_2= 0.0519x -13.31$	0.9782
A12	$y_1= 0.2859x -74.81$	0.9275	$y_2= 0.2049x -47.73$	0.9666
A13	$y_1= 0.2535x -67.37$	0.8766	$y_2= 0.1628x -35.89$	0.9768
A14	$y_1= 0.3359x -86.69$	0.9291	$y_2= 0.1367x -35.12$	0.92
B1	$y_1=0.2795x -73.14$	0.9256	$y_2= 0.2397x -56.73$	0.9718
B2	$y_1= 0.2727x -70.95$	0.8635	$y_2= 0.2441x -61.07$	0.9698
B3	$y_1= 0.1313x -36.16$	0.7924	$y_2= 0.1202x -28.29$	0.9428
B4	$y_1= 0.2078x -56.81$	0.8087	$y_2= 0.1023x -22.21$	0.801
B5	$y_1=0.2716x -72.27$	0.8325	$y_2= 0.1462x -34.6$	0.9557
B6	$y_1= 0.3837x -100$	0.8834	$y_2= 0.1903x -45.76$	0.9646
B7	$y_1= 0.4197x -109.3$	0.8763	$y_2= 0.2337x -56.45$	0.9888

代码支撑材料:

问题 1 的求解代码

```
1 问题 1 的求解代码
2
3 clear,clc,close all
4 %数据预处理，提取到matlab中
5 xlsname="附件1.xlsx";
6 sheet="性能数据表";
7 xlsRange="C2:J115";
8 ori_data=xlsread(xlsname,sheet,xlsRange);%xls数据读入
9 A1=[ori_data([1:5],1),ori_data([1:5],2),ori_data([1:5],4)];
10 %催化剂组合A1的温度_乙醇转化率(%)_C4烯烃选择性(%)
11 A2=[ori_data([6:10],1),ori_data([6:10],2),ori_data([6:10],4)];
12 %催化剂组合A2的温度_乙醇转化率(%)_C4烯烃选择性(%)
13 A3=[ori_data([11:17],1),ori_data([11:17],2),ori_data([11:17],4)];
14 %催化剂组合A3的温度_乙醇转化率(%)_C4烯烃选择性(%)
15 A4=[ori_data([18:23],1),ori_data([18:23],2),ori_data([18:23],4)];
16 %催化剂组合A4的温度_乙醇转化率(%)_C4烯烃选择性(%)
17 A5=[ori_data([24:29],1),ori_data([24:29],2),ori_data([24:29],4)];
18 %催化剂组合A5的温度_乙醇转化率(%)_C4烯烃选择性(%)
19 A6=[ori_data([30:34],1),ori_data([30:34],2),ori_data([30:34],4)];
20 %催化剂组合A6的温度_乙醇转化率(%)_C4烯烃选择性(%)
21 A7=[ori_data([35:39],1),ori_data([35:39],2),ori_data([35:39],4)];
22 %催化剂组合A7的温度_乙醇转化率(%)_C4烯烃选择性(%)
23 A8=[ori_data([40:44],1),ori_data([40:44],2),ori_data([40:44],4)];
24 %催化剂组合A8的温度_乙醇转化率(%)_C4烯烃选择性(%)
25 A9=[ori_data([45:49],1),ori_data([45:49],2),ori_data([45:49],4)];
26 %催化剂组合A9的温度_乙醇转化率(%)_C4烯烃选择性(%)
27 A10=[ori_data([50:54],1),ori_data([50:54],2),ori_data([50:54],4)];
28 %催化剂组合A10的温度_乙醇转化率(%)_C4烯烃选择性(%)
29 A11=[ori_data([55:59],1),ori_data([55:59],2),ori_data([55:59],4)];
30 %催化剂组合A11的温度_乙醇转化率(%)_C4烯烃选择性(%)
31 A12=[ori_data([60:64],1),ori_data([60:64],2),ori_data([60:64],4)];
32 %催化剂组合A12的温度_乙醇转化率(%)_C4烯烃选择性(%)
33 A13=[ori_data([65:69],1),ori_data([65:69],2),ori_data([65:69],4)];
34 %催化剂组合A13的温度_乙醇转化率(%)_C4烯烃选择性(%)
35 A14=[ori_data([70:74],1),ori_data([70:74],2),ori_data([70:74],4)];
36 %催化剂组合A14的温度_乙醇转化率(%)_C4烯烃选择性(%)
37 B1=[ori_data([75:79],1),ori_data([75:79],2),ori_data([75:79],4)];
38 %催化剂组合B1的温度_乙醇转化率(%)_C4烯烃选择性(%)
```



```

39 B2=[ori_data([80:84],1),ori_data([80:84],2),ori_data([80:84],4)];
40 %催化剂组合B2的温度_乙醇转化率(%)_C4烯烃选择性(%)
41 B3=[ori_data([85:90],1),ori_data([85:90],2),ori_data([85:90],4)];
42 %催化剂组合B3的温度_乙醇转化率(%)_C4烯烃选择性(%)
43 B4=[ori_data([91:96],1),ori_data([91:96],2),ori_data([91:96],4)];
44 %催化剂组合B4的温度_乙醇转化率(%)_C4烯烃选择性(%)
45 B5=[ori_data([97:102],1),ori_data([97:102],2),
46 ori_data([97:102],4)];
47 %催化剂组合B5的温度_乙醇转化率(%)_C4烯烃选择性(%)
48 B6=[ori_data([103:108],1),ori_data([103:108],2),
49 ori_data([103:108],4)];
50 %催化剂组合B6的温度_乙醇转化率(%)_C4烯烃选择性(%)
51 B7=[ori_data([109:114],1),ori_data([109:114],2),
52 ori_data([109:114],4)];
53 %催化剂组合B7的温度_乙醇转化率(%)_C4烯烃选择性(%)
54 %%对数据进行回归分析
55 temp=A4(:,1);%将温度提取到temp中
56 [r,c]=size(temp);
57 p=[ones(r,c),temp];%回归预处理数据
58 count=1:r;
59 trans_yichun=A4(:,2);%乙醇转化率
60 ratio_C4=A4(:,3);%C4烯烃选择性
61 %b1, b2: 多元线性回归的系数估计值
62 %bint1, bint2: 系数的置信区间
63 %r1, r2: 残差
64 %rint1, rint2: 残差的置信区间
65 [b1,bint1,r1,rint1,stats1]=regress(trans_yichun,p);%乙醇转化率与温度
66 [b2,bint2,r2,rint2,stats2]=regress(ratio_C4,p);%C4与温度
67 figure(1);
68 yichun_fit=p(count,:)*b1;
69 C4_fit=p(count,:)*b2;
70 plot(count,yichun_fit,'r-',count,trans_yichun,'g-',
71 count,abs(yichun_fit-trans_yichun),'k-','LineWidth',1);
72 legend('拟合值','实际值','误差值');
73 figure(6);
74 C4_fit=p(count,:)*b2;
75 plot(count,C4_fit,'r-',count,ratio_C4,'g-',
76 count,abs(C4_fit-ratio_C4),'k-','LineWidth',1);
77 legend('拟合值','实际值','误差值');
78 %下面画出C4烯烃选择性和温度的图像

```

```

79 figure(2);
80 shangjie=rint1(:,1);%残差95%置信区间下限
81 xiajie=rint1(:,2);%残差95%置信区间上限
82 plot(count,xiajie,'b-',count,r1,'*',count,shangjie,
83 'r-','LineWidth',1)%温度与乙醇转化率之间的残差图
84 legend('上限','残差值','下限');
85 xlabel('样本值');
86 ylabel('残差值');
87 figure(3);
88 rcoplot(r1,rint1);%温度与乙醇转化率的残差个案分析图
89 %下面画出乙醇转化率和温度的图像
90 figure(4);
91 shangjie=rint2(:,1);%残差95%置信区间下限
92 xiajie=rint2(:,2);%残差95%置信区间上限
93 plot(count,xiajie,'b-',count,r2,'*',count,shangjie,'r-','LineWidth',1)
94 legend('上限','残差值','下限');
95 xlabel('样本值');
96 ylabel('残差值');
97 figure(5);
98 rcoplot(r2,rint2);%温度与C4烯烃选择性的残差个案分析图

```

问题二：

图表支撑材料：

表 14: 因变量定义表

数学表达	因变量说明
y_1	乙醇转化率
y_2	C4 烯烃选择性
y_3	乙烯选择性
y_4	乙醛选择性
y_5	碳数为 4-12 脂肪醇选择性
y_6	甲基苯甲醛和甲基苯甲醇选择性
y_7	其他生成物的选择性

代码支撑材料：

1 问题 2 的求解代码

```

2
3     clc,clear,close all
4     %%xls 数据导入
5     xlsname1="问题2数据表_通过显著性剔除两个变量之后.xlsx";
6     sheet1="不带变量命名含义";
7     xlsRange1="A2:M115";
8     ori_data=xlsread(xlsname1,sheet1,xlsRange1);
9     xlsname2="一对多 mape.xlsx";
10    sheet2="Sheet1";
11    xlsRange2="A1:D114";
12    yiduiduo=xlsread(xlsname2,sheet2,xlsRange2);
13    mu=mean(ori_data);sig=std(ori_data);%求均值和标准差
14    R=corrcoef(ori_data);%求相关系数矩阵
15    Std_data=zscore(ori_data);%数据标准化
16    X_squire=Std_data(:,[1:7]);%提取出X矩阵
17    Y_squire=Std_data(:,[8:end]);%提取出Y矩阵
18    Y_ori=ori_data(:,[8:end]);
19    [n,m]=size(Y_ori);
20    [XL,YL,XS,YS,BETA,PCTVAR,MSE,stats] = plsregress(X_squire,Y_squire);
21    %pls 回归
22    xw=X_squire\XS;
23    yw=Y_squire\YS;
24    ncomp=input("输入nump值");
25    [XL2,YL2,XS2,YS2,BETA2,PCTVAR2,MSE2,stats2]
26    = plsregress(X_squire,Y_squire,ncomp);%pls 回归
27    u=size(X_squire,2);%u是自变量个数
28    v=size(Y_squire,2);%v是因变量个数
29    beta3(1,:)=mu(u+1:end)-mu(1:u)./sig(1:u)*BETA2([2:end],:)
30    .*sig(u+1:end);
31    beta3([2:u+1],:)=(1./sig(1:u))*sig(u+1:end).*BETA2([2:end],:);
32    bar(BETA2','k');    %画直方图
33    Ynihe= repmat(beta3(1,:),[size(X_squire,1),1])
34    +ori_data(:,[1:u])*beta3([2:end],:);    %求因变量的预测值
35    Y_max=max([Ynihe;ori_data(:,[u+1:end])]);
36    %求预测值和预测值的最大值
37    mape_duo=mean(abs((Y_ori-Ynihe)./Y_ori));
38
39
40    y1nihe=Ynihe(:,1);%Y1的拟合值
41    y1_means=mean(Y_ori(:,1));%Y1的拟合值

```

```

42     y1_ori=Y_ori(:,1);%Y1 的拟合值
43     r2=1-(sum((y1nihe-y1_ori).^2)./sum((y1_ori-y1_means).^2));%求R^2
44
45     y2nihe=Ynihe(:,2);%Y1 的拟合值
46     y2_means=mean(Y_ori(:,2));%Y1 的拟合值
47     y2_ori=Y_ori(:,2);%Y1 的拟合值
48     r22=1-(sum((y2nihe-y2_ori).^2)./sum((y2_ori-y1_means).^2));%求R^2
49
50     a1=yiduiduo(:,1);
51     b1=yiduiduo(:,2);
52     a22=yiduiduo(:,3);
53     b2=yiduiduo(:,4);
54     mape_yiduiduo_y1=mean(abs((a1-b1)./a1));%求y1的mape值
55     mape_yiduiduo_y2=mean(abs((a22-b2)./a22));%求y2的mape值

```

问题三：

图表支撑材料：

代码支撑材料：

```

1  问题 3 的求解代码
2  LINGO 代码
3  model:
4  max=y1*y2;
5  y1=-65.8743-1.7114*x1-9.9651*x2-9.1408*x3+0.1144*x4+0.3139*x5
6  +0.0068*x3*x5+0.0044*x4/x3;
7  y2=-42.7073-3.3831*x1+3.7492*x2-3.2979*x3+0.0678*x4+0.1747*x5
8  +0.0082*x3*x5-0.0048*x4/x3;
9  y3=-8.2906-0.0772*x1-1.4555*x2-0.4917*x3+0.0088*x4+0.0340*x5
10 +0.0023*x3*x5+0.00021773*x4/x3;
11 y4=15.6469+1.6917*x1-15.5341*x2+0.5485*x3-0.0107*x4+0.0297*x5
12 +0.0032*x3*x5-0.0095*x4/x3;
13 y5=153.9014+3.2638*x1+9.4221*x2+4.3213*x3-0.0941*x4-0.3154*x5
14 -0.0192*x3*x5+0.0130*x4/x3;
15 y6=-2.8899-0.4839*x1+2.0813*x2-0.1618*x3+0.0060*x4+0.0120*x5
16 +0.00093335*x3*x5+0.00046965*x4/x3;
17 y2+y3+y4+y5+y6<100;
18 y1<100;
19 y1>0;

```

```
20  y2<100;
21  y2>0;
22  y3<100;
23  y3>0;
24  y4<100;
25  y4>0;
26  y5<100;
27  y5>0;
28  y6<100;
29  y6>0;
30  x5>250;
31  x5<450;
32  x4>10;
33  x4<200;
34  x3>0.3;
35  x3<2.1;
36  x2>0.5;
37  x2<2;
38  x1>0.5;
39  x1<5;
40  end
```