C4 烯烃制备分析与试验设计

摘　要

C_4烯烃可用于化学工业生产与医药制造，是重要的基础化工原料，是石油化工产业的基础。近些年，随着国内化工产业的不断进步，C_4烯烃的综合利用越来越受到重视，乙醇催化偶合制备 C_4烯烃逐渐进入人们的视野。因此，选择合适的催化生产工艺，实现稳定、高效生产的目标，将有助于相关企业经济效益的提升，同时也能促进我国的化工产业的发展。

针对问题一，首先，经过**皮尔逊相关系数**分析，计算每种催化剂组合下乙醇转化率、C4 烯烃选择性与温度的相关性；根据计算结果所得的相关性系数表可知数据之间存在较强的相关性。以乙醇转化率、C4 烯烃选择性为因变量，温度作为自变量进行**非线性回归拟合**，所有组合的相关性系数R^2的均值在 95%以上，满足二项式非线性回归函数形式。再画出乙醇转化率以及烯烃选择性与温度的图像，对图像进行描述。其次对于第二小问，同时通过聚类分析对数据进行**降维处理**，将 5 个附加产物变量降维成 2 组，在给定温度和催化剂组合下对结果进行了定量分析，发现随着时间的增加，在一定温度和相同催化剂下，随着时间的增加，反应物乙醇转化率持续下降，逐渐趋于稳定；C4 烯烃选择性增加，并趋于稳定，与时间呈现出**弱负相关性**。

针对问题二，根据附件 1 中数据分析，得出“石英砂”对乙醇转化率以及 C4 烯烃选择性的影响可以忽略，因此剔除相关该数据，同时利用问题一中的聚类情况，对附件一中的数据进行**降维**。利用**多元线性回归**建模，并且通过偏最小二乘回归建模完善改进多元线性回归建模无法分析因变量之间的关系的缺点，得到两种模型对应的**残差平方和**都在 10 左右波动，因此两模型差异很小，得出乙醇转化率、C4 烯烃选择性这两个因变量之间呈现低相关性。并且得出温度在解释这两个因变量时，具有**极为重要**的作用，装料比的解释能力均为一般。

针对问题三，探究多个自变量与在相同的实验条件下，如何组合从而使得 C4 烯烃的收率最大。通过题中所给的 C4 烯烃收率计算式，写出基础**单目标最优化**模型，同时求出 C4 烯烃收率的多元回归函数对模型进行优化，引入**方差分析**探究多自变量之间的**交互作用**，从而更好地拟合数据最终求得温度在 450℃时 C4 烯烃的最大收率为 47.1%。当温度限制在 350℃时 C4 烯烃的最大收率为 28.2%。

针对问题四，为了能够在 5 次实验内更加充分的了解如何提高 C4 烯烃的收率，本文采取**均匀设计实验法**，以更少的实验次数获取更多的信息。本文根据问题二偏最小二乘模型所得相关系数判断各个催化剂组合各组分与 C4 烯烃收率的关系，以此来设置各因素水平及**排除**相关性极低可视为常量的因变量，满足均匀设计 5 此实验对自变量的要求，并根据参数水平表设置实验方案。为了化学实验操作安全，本文剔除了均匀设计使用表的最后一行，并以问题三所得最优结果代替。最终，根据均匀设计实验法排列出了再增加的 5 次实验。

关键词　多元回归模型 偏最小二乘回归分析 最优化 方差分析 均匀设计

1 问题提出

1.1 背景分析

C_4烯烃可用于生产聚丙烯等多项化工产品，是重要的基础化工原料，是优质的汽油调和组分，是石油化工产业的基础。近些年，随着国内煤化工业的不断发展，C_4烯烃的综合利用越来越受到重视，乙醇催化偶合制备 C_4烯烃逐渐进入人们的视野。[1]

因此，选择合适的催化生产工艺，达到稳定、高效生产的目的，提高 C_4烯烃的选择性和收率，不仅能为企业带来明显的经济效益，同时也能促进我国相关化工企业的发展。

1.2 问题重述

某化工实验室针对不同催化剂在不同温度下做了一组实验，结果如附件 1 和附件 2 所示。请通过数学建模完成下列问题：

问题一：针对附件 1 所给的催化剂组合，研究其乙醇转化率、C_4烯烃的选择性与温度的关系；并分析附件 2 所给定的催化剂组合在一次实验中 350℃下不同时间的实验结果进行分析。

问题二：研究不同催化剂组合和温度对乙醇转化率和 C_4烯烃选择性的影响。

问题三：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。

问题四：如果允许再增加五次实验，应如何设计，并给出详细理由。

2 问题分析

2.1 问题一的分析

问题一要求探讨对每种催化剂组合分别研究乙醇转化率、C4 烯烃的选择性与温度的关系并对附件 2 中确定的催化剂组合和温度下，分析反应物生成物随时间变化趋势。在第一小问中，观察附件 1 中的数据，发现温度与两项指标有着某种关系，利用皮尔逊相关系数判断数据是否存在相关性。通过对数据进行折线图的绘制以及非线性和线性拟合，最终确定温度对乙醇转化率以及 C4 烯烃选择性大小存在二次函数的关系。同时得出各个催化剂组合下的折线图，得出其中不同催化剂组合条件下确切的影响关系。其次，我们进行了第二部分，对附件二中的数据进行处理。因为给定了温度和催化剂组合，数据两项指标变为了与时间有关的变量，因为产生的无关产物也会对两项指标产生影响，对无关产物进行了聚类处理，就可以实现变量的降维。从而建立二次拟合方程对变量进行拟合，得到二

次函数形式的回归方程，从而得到两项指标与时间的具体函数关系。

2.2 问题二的分析

问题二要求探讨不同催化剂组合及温度对乙醇转化率以及 C4 烯烃选择性大小的影响。在问题一的基础上，我们发现数据之间都具有较强的相关性。不妨对附件一中不同组合的催化剂、温度、以及生成物的比率设置赋予变量值，对其进行多元线性回归。在进行相关性检验时，发现变量之间具有很强的相关性，并由此得到多元线性方程，得到的回归方程拟合程度较为理想。为了验证因变量之间也存在影响，同时又进行了偏最小二乘的回归分析进行对比，判断得到的回归方程拟合度与多元线性回归所得结果差别大小，若相差不大说明因变量之间对两项指标的影响不大，若相差较大，就需对比谁的显著性水平更高从而进行选择回归方程。得到的回归方程就可以作为不同催化剂组合和温度对两项指标的影响指标进行量化分析。

2.3 问题三的分析

问题三要探究一组催化剂组合与温度在相同实验条件下的最佳组合，以 C4 烯烃收率为目标函数，寻找使得目标函数最大值时的催化剂、温度配比。根据 C4 烯烃收率的求解式，因约束条件不充分，无法求解最优值，加入 C4 烯烃收率的多元线性回归分析进行模型优化，而 C4 烯烃收率不仅受到来自乙醇转化率与 C4 烯烃的回归系数的限制，还有可能受到其他变量的交互影响。因此，对多自变量进行方差分析，得到了温度与 Co 负载量以及乙醇浓度的组合交互时对因变量也产生一定的影响。将之作为注意力自变量加入自变量中，进而再利用 SPSS 进行拟合从而完成，得到了一个对温度信息更敏感的方差分析模型。

2.4 问题四的分析

问题四要求再增加 5 组实验来探索各个自变量（催化剂各组分）与因变量（C4 烯烃收率）之间的关系。为了在有限次的实验内更加充分的了解如何自变量与因变量之间的关系，本文采取均匀设计实验法。本文根据问题二偏最小二乘模型所得结果首先剔除了与因变量相关性极低的自变量，以满足均匀设计 5 次实验对自变量个数的要求，并根据化学实验安全原则剔除了均匀设计使用表中可能使化学反应过于剧烈而导致危险发生的组合，并以问题三所得最佳最优结果代替。最终，根据均匀设计实验法设计出了所求的 5 种实验。

3 模型假设

为了构建更为精确的数学模型，本文根据实际情况做出以下合理的假设或条件约束：

- ✧ 假设一：乙醇催化偶合制备 C_4 烯烃的过程中，除给定催化剂组合以及温度和反应时间外，其它条件对实验结果的影响可忽略不计。
- ✧ 假设二：在反应进行时，实验室内环境温度和气压不会对反应造成影响。

✧ 假设三：实验设备完好且在实验过程中不会出现漏气、破损等情况导致实验失败或数据出现差错。

4 符号申明

本文中涉及到的主要变量符号说明：

符号	符号含义及说明
y_i	多元线性回归方程的因变量，$i=1,2$
x_i	多元线性回归方程的自变量，$i=1,2$
y_j	偏最小二乘回归分析的原始因变量，$j=1,2,3,4$
x_j	偏最小二乘回归分析的原始自变量，$j=1,2,3,4,5$
$y\times j$	偏最小二乘回归分析的标准化因变量，$j=1,2,3,4$
$x\times j$	偏最小二乘回归分析的标准化自变量，$j=1,2,3,4,5$
$P_{x,y}$	皮尔逊相关系数
$\mathrm{U_i}$	二项式回归函数，$i=1,2$
βm	偏回归系数，$m=1,2,3,4,5$
μ_j	第j个自变量x_j样本均值
s_j	第j个自变量x_j样本标准差

注：未申明的变量以其在符号出现处的具体说明为准。

设y_1表示乙醇转化率，y_2表示 C4 烯烃选择性，x_1表示Co/SiO2质量，x_2表示 Co 负载量，x_3表示 HAP 质量，x_4表示乙醇浓度，x_5表示温度。

5 模型建立

5.1 问题一的求解与分析

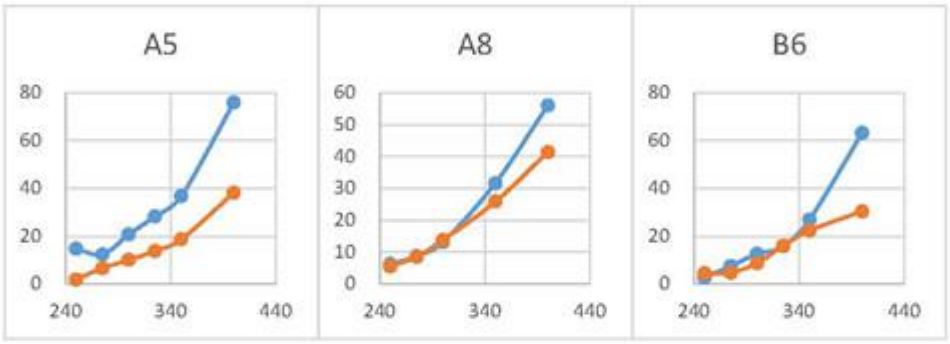
根据题意将问题分为两个小问。第一小问对实验类型分组，通过在数据的可视化以及非线性函数拟合来分析每种催化剂组合、温度对乙醇转化率、C4 烯烃的选择性影响分析。第二小问通过对数据聚合，可视化数据变化以及函数拟合关键变量从而对结果进行分析。

5.1.1 数据预处理

根据各个催化剂组合的组分不同，本文将所给的催化剂组合分成 4 类，如下表所示：

表 1：催化剂组合分类表

组别	催化剂组合	分类依据
1	A1—A6	Co 负载量
2	A7—A12	乙醇浓度
3	B1—B6	Co/SiO_2 和 HAP 装料比
4	A13—A14 A7—A8	特殊对照实验



注：如上所示为 Class 1 — Class 4 中抽取的具有代表性的图像，其它详见附录。

Class 1：

在 A1 至 A6 中，各组的装料比率相同，因此本文将其作为一类实验进行分析。

且 A4，温度 400℃下，乙醇转化率在实验组中达到最大，而在催化剂组合 A3，温度 400℃下，C4 烯烃选择性最高。

Class 2：

在 A7 至 A9 及 A12 中，催化剂组分中的 Co 负载量、装料比率都相同，因此本文同样将其作为一类实验进行分析。

在这些实验组中可以分析得出，随着乙醇加入浓度速率上升时，乙醇转化率逐渐降低。

Class 3：

在 B1 至 B6 中，B1、B2、B3、B4、B6 中以催化剂质量比为变量进行研究，B5 为其他变量的对照试验。因此本文同样将其作为一类实验进行分析。

实验组 B3、B4、B1、B6、B2 中质量比逐渐增加，由试验结果分析可知，随着温度的升高质量比越大 C4 烯烃选择性就越高。B1 与 B2 以乙醇浓度为变量，当升高乙醇浓度时，可以得知随着温度升高，乙醇转化率与 C4 烯烃选择性都略有下降。

Class 4：

在 A13、A14 的对比下，易知在该条件下，“乙醇转化率(%)”变化不大，但

“C4 烯烃选择性”波动。

A8 与 B7 之间仅存在着装料方式的不同，有结果分析，两种装袋方式对乙醇转化率、C4 选择性的变化趋势的影响很小，但是经过第（罗马数字二）装袋方式的效果总体优于第一种。

5.1.2 问题一第(1)问模型的建立与求解

以下为本文在进行数据分析的流程图：

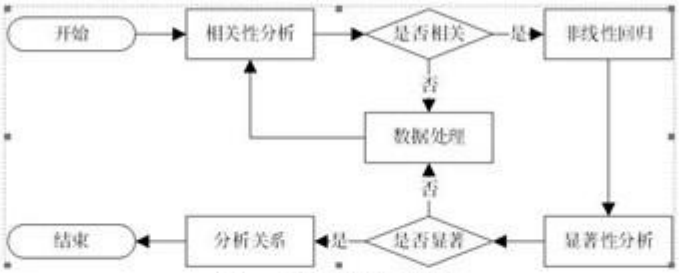


图 1：数据分析流程图

Step 1：相关性分析

本文首先通过计算皮尔逊相关系数，每种催化剂组合下乙醇转化率、C4 烯烃选择性与温度的相关性。

在公式(5-1)中，x 代表温度，y 分别代表乙醇转化率和 C4 烯烃选择性。

$$P_{x,y}=\frac{cov(x,y)}{\sigma_x\sigma_y}=\frac{E[(x-x_i)(y-y_i)]}{\sigma_x\sigma_y} \tag{5-1}$$

计算结果详见目录。在 21 组数据中有 19 组的 *Pearson* 相关系数大于 0.9，因此，不同催化剂组合的条件下，温度和乙醇转化率、C_4烯烃选择性数据**极强相关**。

表 2：乙醇转化率、C4 烯烃选择性与温度的相关系数表

催化剂组合	乙醇转化率(%)	C_4烯烃选择性(%)	催化剂组合	乙醇转化率(%)	C_4烯烃选择性(%)
A1	0.965	0.887	A12	0.963	0.983
A2	0.995	0.914	A13	0.937	0.988
A3	0.982	0.955	A14	0.964	0.959
A4	0.998	0.958	B1	0.962	0.986
A5	0.934	0.97	B2	0.962	0.985
A6	0.984	0.885	B3	0.92	0.971
A7	0.999	0.968	B4	0.91	0.895
A8	0.977	0.992	B5	0.913	0.978
A9	0.921	0.997	B6	0.94	0.982

A10	0.923	0.861	B7	0.936	0.994
A11	0.903	0.989			

Step 2：**非线性回归曲线拟合**

由上述分析可知，温度和乙醇转化率、C_4烯烃选择性数据存在**强相关性**，且数据呈现正相关的趋势，故本文采用非线性回归的方式进行曲线拟合，分别利用线性方程、"S"型曲线和二次曲线进行拟合，以催化剂组合分组，可得出 21 张拟合图像：

注：因篇幅所限，此处仅展示 2 幅有代表性的拟合图像，其它拟合图像相关性系数R^2相差不大，均在 93%以上，故不在此处罗列，详见附录。

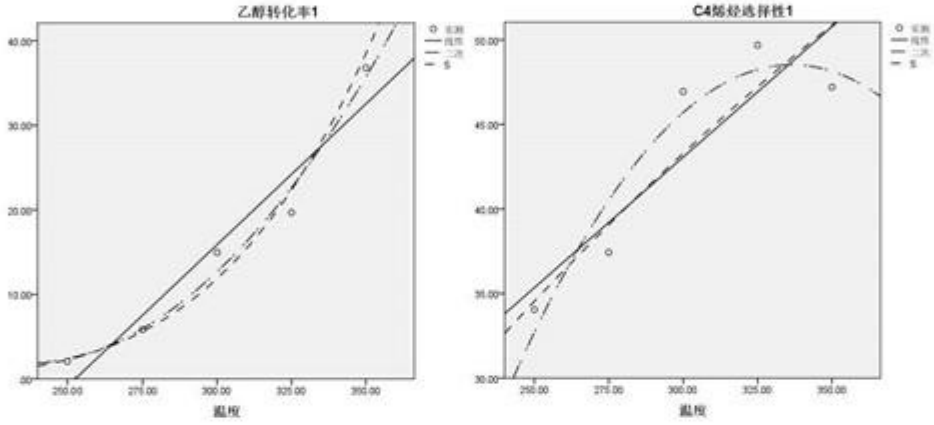


图 2：拟合图像

观察所有拟合图像可知，三种拟合方式均有**较强的相关性**。

Step 3：**相关性检验**

表 3：不同拟合方式的相关性系数表

方程	R^2	方程	R^2
线性	0.932	线性	0.943
二次	0.980	二次	0.976
S 型	0.976	S 型	0.967

注：因篇幅所限，此处仅展示上述 2 幅有代表性的拟合图像的相关性系数R^2计算结果，其它拟合图像的相关性系数详见附录。

由完整的相关性系数表可知，所有图像的相关性系数R^2的均值在 95%以上，且始终大于线性和"S"型曲线的相关性系数。因此本文认为，乙醇转化率、C4 烯烃的选择性与温度的具有**二次相关性**，且为**正相关**。

5.1.3 问题一第(2)问模型的建立与求解

Step 1：数据预处理

首先，本文为获得时间与乙醇转化率和反应产物的量之间的关系，考虑到该反应副产物选择性之间的数值关系，但是变量过于冗余，可以将相关性高的分为

一类，对反应产物进行了降维处理，也就是对各个反应产物进行了分类合并处理，利用 SPSS 软件进行系统 Q 型聚类，做出了所有副反应生成物数据谱系图（如下图所示），并针对反应产生的副产物进行分类。

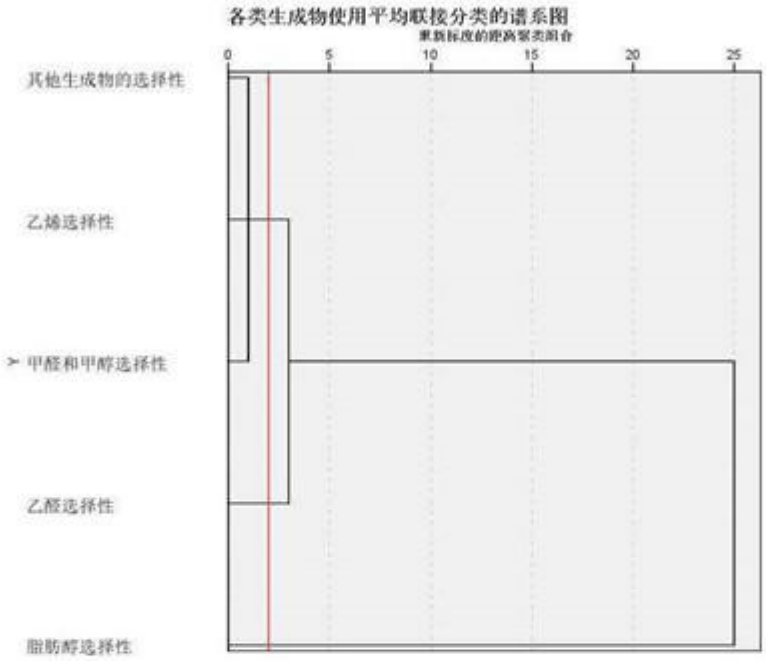


图 3：各反应物聚类谱系图

分析系谱图可以发现，我们可以将乙烯选择性、苯甲基甲醛和苯甲基甲醇的选择性、乙醛选择性以及其他生成物的选择性分为一类称作次要附加产物；脂肪醇的选择性分为一类称作主要附加产物。从而实现数据的降维，因此在六种副生成物中，我们只需要考虑两类生成物对乙醇转化率和 C4 烯烃选择性的影响即可。

分析附件二中数据得出下图乙醇转化率与各产物选择性图随时间变化图（如图 4 所示）。

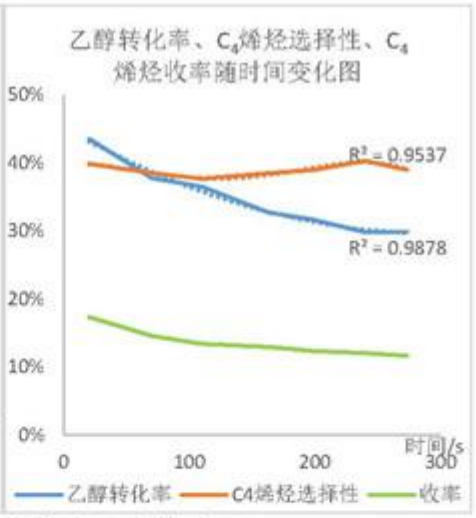


图 4：乙醇转化率与各产物选择性图随时间变化图

图 5：乙醇转化率、C4 烯烃选择性、C4 烯烃收率随时间变化图

由图 4 可知，目标生成物中，C_4烯烃的选择性在 40%附近波动，且趋向于稳定，其方差为1.1768×10^{-4}；主要附加产物的含量在时间区间$[20s, 110s)$之间呈减少的趋势，次要附加产物在该时间区间内呈现增加的趋势，且增加和减少的变化量相比 C_4烯烃选择性变化较大，而在 110s 之后，两附加产物的选择性趋于稳定；此外，乙醇转化率在整个实验中均呈减小的趋势。

假设乙醇转化率以及 C4 烯烃收率随时间变化的函数关系式分别为：

$$U_1 = \alpha_{11}t^2 + \alpha_{12}t + \alpha_{13}$$
$$U_2 = \alpha_{21}t^2 + \alpha_{22}t + \alpha_{23}$$

经过 SPSS 软件的分析，得到

表 4：相关系数表

表a：乙醇转化率

R	R^2	调整后R^2	估算标准误差	系数α_{1m}
.994	.988	.982	.679	-0.105、0.00017、45.258

表b：烯烃选择性

R	R^2	调整后R^2	估算标准误差	系数α_{2m}
.968	.937	.906	.063	0.004、$-0.6*10^{-5}$、4.112

其中 R^2达到了 98.8%和 93.7%非常接近 1，假设成立，可以说乙醇转化率以及 C4 烯烃收率与时间具有二次函数的关系，满足下列函数关系式

$$U_1 = -0.105t^2 + 0.00017t + 45.258$$
$$U_2 = 0.004 - 0.6 \times 10^{-5}t + 4.112$$

在一定温度和相同催化剂下，随着时间的增加，反应物乙醇转化率持续下降，逐渐趋于稳定；C4 烯烃选择性增加，并趋于稳定，与时间呈现出弱负相关性。

5.2 问题二的求解与分析

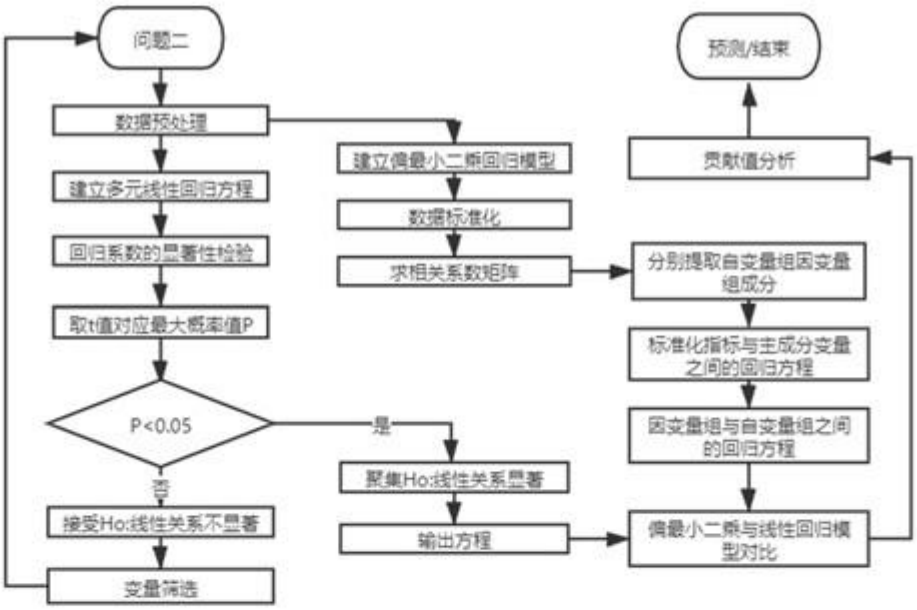


图 6：问题二解题流程图

5.2.1 数据预处理

(1) 考虑到乙醇在催化偶合制备 C4 烯烃的过程中，所给定的催化剂组合每一种都有相应的对照实验，因此，本文选取所有类型的实验，通过 Python 编程软件中的 pandas 库对附件 1 中的催化剂组合按组分进行拆分，其中装料比是比值的形式在数据中存在，不利于进行数据处理，因而将装料比拆分成 Co/SiO2 质量和 HAP 质量，得到单独的两个自变量。最终我们得到 5 个自变量。

(2) 石英砂在实验中充当防暴沸的效果[3]，在实验中起对照作用，并且在加入石英砂的实验组中乙醇转化率和 C4 烯烃选择性并效率没有明显增大，我们认为石英砂不具有催化作用。因此我们舍弃石英砂这一变量组，详细变量数据见附录。

(3) 同时，根据问题一中对反应产物的聚类结果，我们仍将杂质分为主要副产物和次要副产物，并得到如下表所示的变量表。

表 5：因变量自变量的选择表

自变量		因变量					
		目标产物		主要副产物		次要副产物	
编码	变量	编码	变量	编码	变量	编码	变量
$x1$	Co/SiO2 质量	$y2$	C4 烯烃选择性	$y3$	碳数为 4-12 脂肪醇	$y4$	甲基苯甲醛和甲基苯甲醛选择性
$x2$	Co 负载量	$y1$	乙醇转化率				乙醛选择性
$x3$	HAP 质量						其他生成物选择性
$x4$	乙醇浓度						乙烯选择性

$x5$　　温度

5.2.2 模型建立及误差分析

根据附件 1 中的拆解数据作为自变量，进行多元回归分析，从而得到标准回归系数，进而分析对乙醇转化率、C4 烯烃选择性大小的主要影响成分。针对本题中两个因变量，我们通过偏最小二乘法来改进原有的分析方法，探究多因变量的分析。最终通过回归系数来分析乙醇转化率以及 C4 烯烃转化率的主要影响因素。

模型 1：多元线性回归模型

Part 1：方差分析

根据问题一，可知各个催化剂组合对最终反应生成物的影响存在某种线性关系。此时则对不同的催化物和温度进行分析，判断两者对乙醇转化率以及 C4 烯烃选择性大小的影响。以 x_1, x_2, x_3, x_4, x_5 作为自变量，以 y_1, y_2 作为因变量进行多元线性回归分析。

首先对变量进行方差分析。利用 SPSS 软件进行方差分析，得到如下表所示方差齐次检验的显著性指标：

表 6：方差齐次检验表

乙醇转化率（y1）					
自变量	Co/SiO2 质量（x1）	Co 负载量（x2）	HAP 质量（x3）	乙醇浓度（y4）	温度（y5）
显著性	0.125	0.043	0.041	0.05	0.08
C4 烯烃选择性（y2）					
自变量	Co/SiO2 质量（x1）	Co 负载量（x2）	HAP 质量（x3）	乙醇浓度（y4）	温度（y5）
显著性	0.125	0.043	0.041	0.05	0.08

分析图表可知，方差的显著性均大于 0.05，可以认为本次方差分析是齐的。因此进一步进行单因素方差分析得到如表所示单因素方差分析表：

表 7：主体间效应检验

源	III 类平方和	自由度	均方	F	显著性
修正模型	98469877.840[a]	103	956018.231	51.598	.000
截距	20101179.490	1	20101179.490	1084.889	.000
x_1	.000	0	.	.	.
x_2	1881119.655	3	627039.885	33.842	.000
x_3	531008.332	1	531008.332	28.659	.000
x_4	861063.569	3	287021.190	15.491	.000
x_5	38683517.781	6	6447252.964	347.967	.000

$x_2 * x_5$	2014792.493	13	154984.038	8.365	.001
$x_3 * x_5$	993214.612	4	248303.653	13.401	.001
$x_4 * x_5$	1360694.082	14	97192.434	5.246	.006
误差	185283.350	10	18528.335		
总计	137924965.473	114			
修正后总计	98655161.190	113			

关于多个控制变量的独立作用部分，可以得出对因变量影响排序为x_5，x_4，x_3，x_1，x_2.说明温度的贡献量最大，即对 C_4烯烃收率的大小影响最大，Co 的负载量贡献量最小，即对其大小影响最小。因此我们可以进行回归分析模型的建立。

Part 2：建立多元回归模型

多元线性回归分析的模型为：

$$\begin{cases} y = \beta_0 + \beta_1 x_1 + \beta_2 x_1 + \cdots \beta_m x_m + \varepsilon \\ \varepsilon \sim N(0, \sigma^2), \end{cases} i = 1,2,\ldots,n \qquad (5-2)$$

其中，$\beta_0, \beta_1, \beta_2, \cdots, \beta_m, \varepsilon^2$是偏回归系数，与$x_1, x_2, x_3, \ldots, x_m$无相关性。$\varepsilon$为随机误差项。

假设，因变量与各自变量之间存在线性关系，那他们之间的线性总体回归模型可以表示为：

$$\begin{cases} y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \beta_3 x_{13} + \beta_4 x_{14} + \beta_5 x_{15} + \varepsilon_1 \\ y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \beta_3 x_{23} + \beta_4 x_{24} + \beta_5 x_{25} + \varepsilon_2 \end{cases} \qquad (5-3)$$

其中，ε为随机误差项$\varepsilon \sim N(0, \sigma^2)$。

Part 3：求解回归系数及分析

下表是回归系数分析表。

表 8：回归系数分析表

模型		未标准化系数		标准化系数	t	显著性
		B	标准误差	Beta		
y_1	(常量)	-80.509	7.231		-11.134	.000
	x_1	-.034	.071	-.104	-.476	.635
	x_2	.134	.898	.007	.149	.882
	x_3	.141	.069	.448	2.045	.043
	x_4	-8.765	2.065	-.199	-4.244	.000
	x_5	.333	.019	.763	17.530	.000
y_2	(常量)	-48.732	5.112		-9.532	.000
	x_1	.003	.050	.015	.059	.953
	x_2	-3.164	.635	-.271	-4.983	.000
	x_3	.086	.049	.462	1.769	.080
	x_4	2.673	1.460	.102	1.831	.070
	x_5	.181	.013	.701	13.489	.000

并且得到的多元线性回归模型可以表示为：

$$y_1 = -80.509 - 0.034x_1 + 0.134x_2 + 0.141x_3 - 8.765x_4 + 0.333x_5$$

$$y_2 = -48.732 + 0.003x_1 - 3.164x_2 + 0.086x_3 + 2.673x_4 + 0.181x_5$$

Part 4：误差分析

利用 SPSS 算变量之间的相关系数，得表 4，y_1，y_2与x_1, x_2, x_3, x_4和x_5之间显著相关。

表 9：回归方程的显著性指标

模型	R	R 方	调整后 R 方	标准估算的误差	德宾-沃森
y_1	.892	.796	.786	10.55848	.937
y_2	.842	.709	.696	7.46480	.580

从表中结果可以得出，经过逐步线性回归分析后拟合出的模型y_1和y_2的R^2值分别为 0.796 和 0.709，其值接近 1，可以认为模型拟合程度较好。

表 10：残差统计表

乙醇转化率残差统计

	最小值	最大值	平均值	标准偏差	个案数
预测值	-13.5101	83.1577	21.9846	20.36060	114
残差	-18.60947	28.80237	.00000	10.32224	114
标准预测值	-1.743	3.004	.000	1.000	114
标准残差	-1.763	2.728	.000	.978	114

烯烃选择性残差统计

	最小值	最大值	平均值	标准偏差	个案数
预测值	-9.1596	49.9478	16.4508	11.39512	114
残差	-17.30414	22.09034	.00000	7.29778	114
标准预测值	-2.247	2.940	.000	1.000	114
标准残差	-2.318	2.959	.000	.978	114

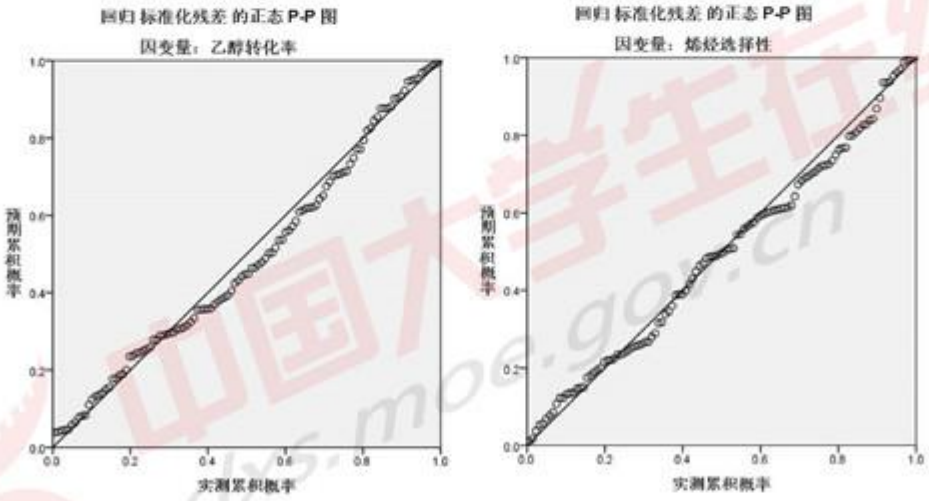


图 7：回归标准化残差的正态 P-P 图

由表 9 和图 6 的残差分布表图，残差最大值均在 25 左右，且标准化残差都集中分布在直线附近，可以认为标准化残差满足正态分布，误差通过，建立的回归方程合理且误差较小。

模型 2：偏最小二乘法回归模型

Part 1：模型优化机理

上文在利用多元线性回归建立模型时，分别建立了$y1, y2$ 所对应的多元线性回归模型，而对于在化学背景下的多因变量问题中，因变量之间很有可能产生相互影响，针对本题中两个因变量，我们通过偏最小二乘法来改进原有的分析方法，探究多因变量的分析，在原有的基础上探究因变量之间是否存在影响。

Part 2：构造偏最小二乘法模型

依据节 5.2.1 中因变量和自变量的分类，存储在下面两个矩阵中：

自变量的观测数据矩阵$A=(a_{ij})_{\{108\times5\}}$，因变量的观测数据矩阵为$B=(b_{ij})_{108\times4}$.

Step 1：数据标准化

将各指标a_{ij}转化成指标值a_{ij}^{*}，有

$$a_{ij}^{*}=\frac{a_{ij}-\mu_j^{(1)}}{s_j^{(1)}}, i=1,2,\cdots,108, j=1,2,3,4. \tag{5-4}$$

其中，$\mu_j^{(1)}=\frac{1}{108}\sum_{i=1}^{108}a_{ij}, s_j^{(1)}=\sqrt{\frac{1}{108-1}\sum_{i=1}^{108}\left(a_{ij}-\mu_j^{(1)}\right)^2}, j=1,2,3,4$，即$\mu_j^{(1)}, s_j^{(1)}$为第$j$个自变量$x_j$的样本均值和样本标准差。对应地，称

$$x_j^{*}=\frac{x_j-\mu_j^{(1)}}{s_j^{(1)}}, j=1,2,3,4, \tag{5-5}$$

为标准化变量。

同理，有标准化指数值

$$\mathrm{b}_{\mathrm{ij}}^{*}=\frac{b_{ij}-\mu_j^{(2)}}{s_j^{(2)}}, i=1,2,\cdots,108, j=1,2,3,4. \tag{5-6}$$

其中，$\mu_j^{(2)}=\frac{1}{108}\sum_{i=1}^{108}b_{ij}, s_j^{(2)}=\sqrt{\frac{1}{108-1}\sum_{i=1}^{108}\left(a_{ij}-\mu_j^{(2)}\right)^2}, \mathrm{j}=1,2,3,4$，即$\mu_j^{(1)}, s_j^{(1)}$为第$j$个自变量$x_j$的样本均值和样本标准差。对应地，称

$$y_j^{*}=\frac{x_j-\mu_j^{(2)}}{s_j^{(2)}}, j=1,2,3,4, \tag{5-7}$$

为标准化变量。

Step 2：求相关系数矩阵

下表给出了这 9 个变量的简单相关系数矩阵。

表 11：简单相关系数矩阵

	$x1$	$x2$	$x3$	$x4$	$x5$	$y1$	$y2$	$y3$	$y4$
$x1$	1	0.2089	0.9800	-0.2992	-0.0018	0.3946	0.3799	-0.1748	-0.0351
$x2$	0.2089	1	0.2136	0.1413	-0.0142	0.0418	-0.1650	0.0979	-0.0137
$x3$	0.9800	0.2136	1	-0.3017	-0.0014	0.4065	0.3876	-0.0799	-0.1653
$x4$	-0.2992	0.1413	-0.3017	1	-0.0385	-0.3312	-0.1071	0.0237	0.0435
$x5$	-0.0018	-0.0142	-0.0014	-0.0385	1	0.7706	0.6998	-0.7032	0.4370
$y1$	0.3946	0.0418	0.4065	-0.3312	0.7706	1	0.7353	-0.5935	0.2679
$y2$	0.3799	-0.1650	0.3876	-0.1071	0.6998	0.7353	1	-0.6634	0.1752
$y3$	-0.1748	0.0979	-0.0799	0.0237	-0.7032	-0.5935	-0.6634	1	-0.8529
$y4$	-0.0351	-0.0137	-0.1653	0.0435	0.4370	0.2679	0.1752	-0.8529	1

从相关系数矩阵可以看出 Co/SiO2 质量与 Co 负载量正相关，Co 负载量与 HAP 质量、乙醇浓度正相关，HAP 与乙醇浓度负相关，乙醇转化率与 Co/SiO2 质量、Co 负载量、HAP 质量和温度呈正相关，与乙醇浓度呈负相关；C4 烯烃选择性与 Co/SiO2 质量、HAP 质量、温度呈正相关，与 Co 负载量、乙醇浓度呈负相关。

Step 3：分别提出自变量组和因变量组的成分

由 Matlab 程序（详见附录），可以得到三对成分[u1,v1],[u2,v2],[u3,v3]，[u4,v4]，[u5,v5]。

$$\begin{cases} u_1 = -0.0296x_1 + 0.0063x_2 - 0.0265x_3 + 0.015x_4 - 0.0708x_5 \\ v_1 = -9.2291y_1 - 11.8003y_2 - 7.799y_4 \end{cases}$$

$$\begin{cases} u_2 = -0.0218x_1 + 0.0158x_2 - 0.037x_3 + 0.0229x_4 + 0.0602x_5 \\ v_2 = -2.5653y_1 - 2.2759y_2 + 6.7408y_4 \end{cases}$$

$$\begin{cases} u_3 = -0.025x_1 + 0.0821x_2 - 0.0325x_3 - 0.068x_4 + 0.0106x_5 \\ v_3 = 1.8275y_1 - 2.3454y_2 + 0.4012y_4 \end{cases}$$

$$\begin{cases} u_4 = -0.1090x_1 + 0.0515x_2 + 0.0829x_3 - 0.0681x_4 - 0.0106x_5 \\ v_4 = 0.9458y_1 - 0.2164y_2 - 2.5390y_4 \end{cases}$$

$$\begin{cases} u_5 = -0.4577x_1 + 0.0083x_2 + 0.463x_3 + 0.0159x_4 + 0.0034x_5 \\ v_5 = 2.3562y_1 - 0.1258y_2 - 8.218y_4 \end{cases}$$

而前四个成分解释自变量的的比率为 98.69%，因此只要选取前四对成分即可满足分析需求。

Step 4：求两个成分对时标准化指标变量与成分变量之间的回归方程

求得自变量和因变量组与u_1，u_2，u_3，u_4，u_5，之间的回归方程分别为

$$\begin{cases} x_1 = -6.6226u_1 - 7.7166u_2 - 2.1872u_3 - 2.0490u_4 \\ x_2 = -0.2756u_1 - 2.9268u_2 + 6.7982u_3 - 7.4630u_4 \\ x_3 = -6.6197u_1 - 7.7633u_2 - 2.1411u_3 - 1.6233u_4 \\ x_4 = 4.006u_1 + 4.0661u_2 - 4.4631u_3 - 7.6109u_4 \\ x_5 = -8.0602u_1 + 6.7369u_2 + 1.3742u_3 - 0.8111u_4 \\ y_1 = -9.2291u_1 + 1.6398u_2 + 1.2505u_3 + 0.3285u_4 \\ y_2 = -8.3245u_1 + 2.2226u_2 - 2.3624u_3 - 0.1029u_4 \\ y_3 = 6.5551u_1 - 4.1319u_2 + 0.6684u_3 + 1.495u_4 \\ y_4 = -2.8172u_1 + 3.886u_2 + 0.7688u_3 - 1.8953u_4 \end{cases} \quad (5-8)$$

Step 5：求因变量组与自变量组之间的回归方程。

把步骤 2.5 中成分u_i代入 Step 3 中y_j^*的回归方程，得到标准化指标变量之间的回归方程

将标准化变量$y_i^*, x_j^*(j = 1,2,3,4)$分别还原成原始变量$y_j, x_j$，得到回归方程

$$\begin{cases} y_1 = -79.8445 + 0.0554x_1 + 0.0351x_2 - 0.0538x_3 - 9.1856x_4 + 0.3323x_5 \\ y_2 = -48.3595 + 0.0517x_1 - 3.2176x_2 + 0.0387x_3 + 2.4439x_4 + 0.1807x_5 \\ y_3 = 184.2684 - 0.1032x_1 - 1.8552x_2 + 0.0287x_3 - 7.066x_4 - 0.3518x_5 \\ y_4 = -35.909 + 0.0515x_1 + 1.3624 - 0.0674x_3 + 4.6221x_4 + 0.1711x_5 \end{cases} \quad (5-9)$$

并且根据$y1, y2$ 分别计算出R^2为 0.812 和 0.765，其值接近于 1，可以认为拟合情况较好。

模型的对比与检验

Part 1：模型对比

当比较两个模型在同一组数据上的拟合效果时，更倾向于比较他们的残差平方和来确定，当残差平方和越小时，说明该模型的拟合效果越好。因此利用 Python 程序分别计算两模型所对应回归函数的残差平方和，进行对比。如下表所示：

表 12：均方误差模型参数对比表

		多元线性回归
因变量 y	R^2	均方误差（MSE）
乙醇转化率（y1）	0.796	10.348
C4 烯烃选择性（y2）	0.709	9.474
		偏最小二乘回归分析
乙醇转化率（y1）	0.812	10.404
C4 烯烃选择性（y2）	0.765	7.057

本文认为，在本题中可以直接用多元线性回归分别对因变量进行建模，可以减少考虑本文中多个因变量之间的联系。因为由图中数据可知，两模型所对应的$y1, y2$ 的残差平方和相差很小，R^2的差值也很小，说明无论是通过多元线性回归建模，还是通过偏最小二乘法建模得到的回归系数都可以很好的解释多个自变量

与因变量之间的相关性。

Part 2：模型的解释

利用偏最小二乘法建模时，根据预测值和实际值做出关于中心轴线的离散图，得到乙醇转化率预测图以及烯烃转化率预测图，根据下图数据分布情况得知拟合效果良好。

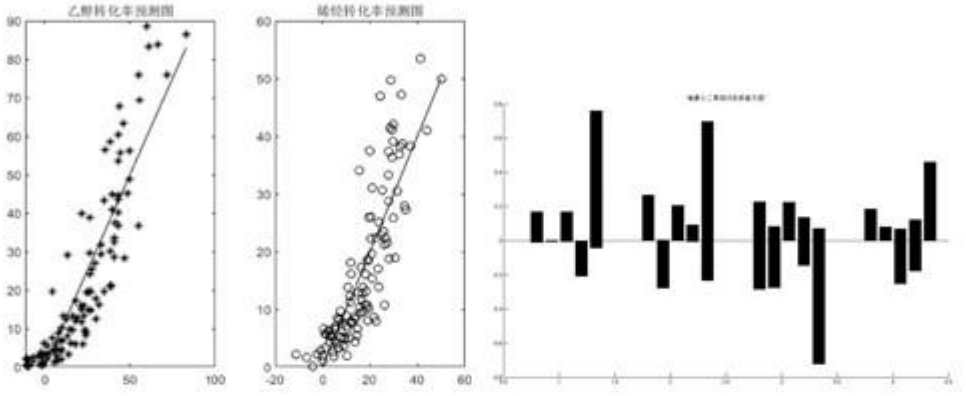


图 8：乙醇转化率、烯烃转化率预测图与偏最小二乘回归系数直方图

从回归系数图中可以观察到，温度变量在解释两个乙醇转化率、C4 烯烃回归方程中起到了**极为重要**的作用，当温度和其它自变量相组合时，对因变量的结果也会产生重要的影响。且 Co/SiO2 与 HAP 的装料比的解释能力在两个回归方程中差别不大，均为正相关。而 Co 负载量在乙醇转化率中基本没有解释性，但是在 C4 烯烃选择性中有较强的**负相关**解释性。而乙醇浓度则对乙醇转化率起负相关解释性，而对 C4 烯烃选择性很少有解释性。

5.3 问题三的求解与分析

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，我们首先根据化学反应关系得出 C4 烯烃收率的直接影响因素，又基于问题二中的结论，推导出 C4 烯烃收率与多个自变量之间的影响关系，建立了单目标最优化模型，并且对多个自变量进行方差分析，对原有模型进行改进，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。

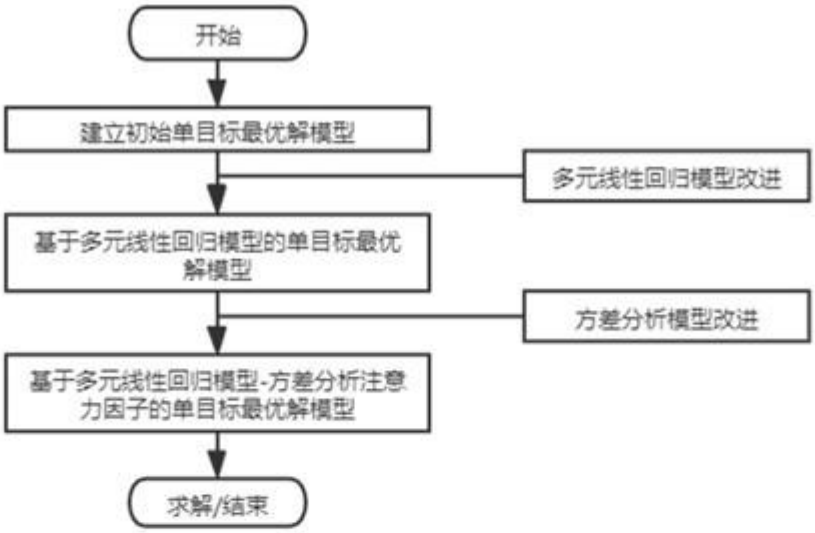


图 9：问题三解题流程图

5.3.1 C4 烯烃收率的模型建立

5.3.1.1 烯烃收率的基础模型

Step 1：数据处理

继续参照问题一得到的聚类结果，将生成物分为 C4 烯烃、主要附加产物（脂肪醇）、次要附加产物，进行数据分析。

由收率公式：C4 烯烃收率=乙醇转化率×C4 烯烃选择性，可求得任意一项催化剂组合所对应的 C4 烯烃收率，因此后两项为 C4 烯烃收率的直接影响因素。考虑到本题中需要拟合优良数据，从而得到 C4 烯烃最大收率，而实验中存在部分拟合效果欠佳的数据，因此通过对 109 组 C4 烯烃收率进行排序，确定 C4 烯烃收率值阈值为 1%，以阈值作为筛选条件对数据进行初步筛选。同时对于一些效果不佳的对照实验，也进行数据剔除，最终得到 61 组积极数据。

Step 2：基础模型建立

设 ρ 表示 C4 烯烃收率，可以得到下列关系式：

$$\rho = y_1 \times y_2 \tag{5-10}$$

5.3.1.2 最大 C4 烯烃收率的单目标优化模型

Part 1：目标函数确定

$$max\rho = y_1 \times y_2 \tag{5-11}$$

Part 2：约束条件确定

①约束条件 1：所有生成物选择性之和等于 100%，则有

$$\sum_{i=1}^{5} y_i = 100 \tag{5-12}$$

②约束条件 2：所有自变量满足已有实验设计的取值范围。根据问题二中对附件一所进行的数据拆分，分析得到每一种自变量的实际取值范围，因此一个约束条件是：

$$Xl_i \le x_i \le Xl_r i = 1,2,\dots,5 \tag{5-13}$$

其中，Xl_i, Xl_r分别表示自变量的左右区间。

Part 3：模型整合

综上所述，我能得到了最大 C4 烯烃收率的单目标优化模型：

$$max\rho = y_1 \times y_2 \tag{5-14}$$

$$s.t \begin{cases} \sum_{i=1}^{5} y_i = 100 \\ Xl_i \le x_i \le Xl_i \end{cases} \quad i = 1,2,\dots 5 \tag{5-15}$$

5.3.2 基于多元线性回归的模型优化

Part 1：模型改进机理分析

根据题目中的描述以及附件 1 中数据的特点，可知本数据为化学反应中的记录值，无法提过足够有效的约束条件，经过测试无法直接利用上述的最优化模型进行最优化求解。因此我们提出基于多元线性回归的 C4 烯烃最大收率模型。由于 C4 烯烃选择率来自乙醇转化率与 C4 烯烃选择性的乘积，其两者的函数表达式是通过多元线性回归来拟合，因此可以直接用拆分得到的多个自变量对 C4 烯烃收率来做多元线性回归分析，从而对原有的模型进行有效的改进。

Part 2：基于多元线性回归拟合 C4 烯烃收率

参照问题二中的多个自变量参数，使用 SPSS 软件对 C4 烯烃收率进行同等方法的数据拟合，可得到以下的分析结果：

$$R^2 = 0.751$$

$$\rho' = -3619.647 + 1.236x_1 - 69.431x_2 + 4.121x_3 - 1276.736x_4 + 11.995x_5$$

由于$R\text{-}squeare > 0.7$，则我们认为该多元线性回归模型具有较高的可信性。

Part 3：获得多元线性回归改进的 C4 烯烃最大收率的最优化模型

因此基于多元线性回归拟合的改进下，由上述分析可得 C4 烯烃最大收率的单目标最优化模型：

$$\max\ \rho'$$

$$s.t. \begin{cases} \sum_{i=1}^{5} y_i = 100 \\ Xl_i \le x_i \le Xl_i \end{cases} \quad i = 1,2,\dots,5$$

5.3.3 基于方差分析的 C4 烯烃收率模型优化

Part 1：优化机理分析

由于 C4 烯烃收率的是由乙醇转化率与 C4 烯烃选择性相乘所得，因此对于 C4 烯烃收率来说，受到了乙醇转化率与 C4 烯烃转化率的系数的直接限制，因此，我们这此引入方差分析，来看这些变量之间是否含有相互作用关系，从而来对 C4 烯烃的单目标优化模型做优化。

Part 2：通过方差分析研究 C4 烯烃收率

根据问题二中的方差分析结果，$x_2 * x_5$、$x_4 * x_5$与$x_3 * x_5$之间存在交互作用

同时，根据两者的 F 值与均方差值可知其对 C4 烯烃收率的大小产生显著影响。因此我们将该三个交互作用部分作为相互作用因子，令 $x_6 = x_2x_5$, $x_7 = x_3x_5$, $x_8 = x_4x_5$. 通过附件 1 中数据的合并处理，我们得到了 8 个自变量，进而通过 SPSS 软件分析计算多元线性回归模型如下所示：

$$\rho = \frac{1}{100}(-3362.489 + 5.238x_1 + 67.621x_2 - 27.568x_3 + 206.598x_4 + 11.56x_5 - 0.529x_6 + 0.086x_7 - 1.15x_8)$$

Part 3：获得方差分析改进的 C4 烯烃最大收率的最优化模型

因此当考虑温度因素时，则通过方差分析对模型进行改进，由上述分析可得 C4 烯烃最大收率的单目标最优化模型：

$$\max \ \rho$$

$$\text{s.t.} \begin{cases} x_6 = x_2x_5 \\ x_7 = x_5x_3 \\ x_8 = x_5x_4 \\ Xl_i \le x_i \le Xl_i \end{cases} \quad i = 1,2,\ldots,5$$

5.3.3.1 C4 烯烃最大收率时的催化剂最佳配比模型的求解

Case 1：当在相同的实验条件时，使用多元线性回归优化的模型进行求解

在相同实验条件下时，所有的自变量的取值与附件 1 中的数据取值限制相同，说明对于温度等主要影响 C4 烯烃收率的变量没有受到特殊限制，因此我们直接选取基于多元线性回归改进的 C4 烯烃最大收率的最优化模型进行求解。

经过 Lingo 程序求得最优值为：47.1%

Case 2：当在限制温度的条件时，使用经方差分析二次优化模型进行求解

由上分析同理可得，此时题目中对温度进行了限制，因此我们采用对温度等主要影响系数有增强作用的方差分析改进的 C4 烯烃最大收率的最优化模型进行求解。

经过 Lingo 程序（见附录）求得最优值为：28.2%

表 13：主体间效应检验

	x_1	x_2	x_3	x_4	x_5
Case 1	200	0.5	200	0.3	450
Case 2	200	0.5	200	0.3	350

5.4 问题四的求解与分析

为了能够在有限的实验次数内更加充分地了解如何提高乙醇选择性和 C4 烯烃转化率，本文在实验设计时选择均匀设计的方法。

相比于正交设计，均匀设计只考虑试验点在实验范围内的均匀散布。因此，相比于正交设计，均匀设计减少了进行实验的次数，且更加适合在较少的实验中获取更多信息，能够在相同的实验次数下更迅速的找到最优解[8, 9]。

在乙醇催化偶合制备 c4 烯烃的实验中，对其产物的造成主要影响的因素有 $x1$、$x2$、$x3$、$x4$、$x5$ 等五种因素。

由于均匀设计的设计原理约束了 U_5 最多只能包含 4 个自变量，故本文参考了前面的问题中做出的偏最小二乘回归系数直方图，将相关性足够低且在直方图中回归系数最接近 0 的自变量 $x4$ 乙醇浓度定为常量，在设计时着重研究其它 4 种自变量与 C4 烯烃收率的关系。

根据附件 1 所给实验结果，本文在 4 种因素内选取 5 种水平，表示为U_5。

均匀设计$U_5(5^4)$表如下所示：

表 74：U_5 与 U_4^* 均匀设计使用表

列号 / 实验号	1	2	3	4
1	1	2	3	4
2	2	4	1	3
3	3	1	4	2
4	4	3	2	1
5	5	5	5	5

列号 / 实验号	1	2	3	4
1	1	2	3	4
2	2	4	1	3
3	3	1	4	2
4	4	3	2	1

各因素水平表如下所示，波浪线表示该数据来自问题三所求最终结果。在设计各因素水平的过程中，观察到偏最小二乘回归系数直方图中，反应温度的相关系数远大于其他几项，因此在设定“温度”因素的水平时，超过附件 1 所给最大值后增加的步进值较小，但通过研究“温度”因素更能够更好地确定使收率更高的因素，以提高反应速率；相关性最小的因素x_2在超过问题三所得最佳结果后的步进值较大；而x_1、x_3的相关性在x_2与x_5之间，但总体更接近x_2，其因素水平取值也比较靠近实验 1 中已存在的数据。

表 85：均匀设计中各因素水平

水平	$x1$	$x2$	$x3$	$x5$
1	10	0.5	50	350
2	50	1	100	380

3	75	2	200	450
4	100	3.5	300	475
5	200	5	400	500

因化学试验存在特殊性，若将最高水平组合在一起可能会导致化学反应过于剧烈，从而使反应容器龟裂，甚至发生爆炸等危险，故在均匀设计使用表中，移除最后一行，得到更加均匀、安全的使用表，以防止事故的发生。此时，设计方案变为U_4^*。

为了补全第 5 次实验，现有 2 种方法可以选择：①将问题三使用多元线性回归优化的模型所得结果作为第 5 次实验；②将 U_6 的均匀设计使用表的最后一行删除，化作 U_5^* 使用。但②因均匀设计原理决定了进行 6 次实验最多只能存在 2 个自变量，故舍弃该方法。

表 16：使用均匀设计增加的 5 组新实验表

实验号	$x1$	$x2$	$x3$	$x5$
1	10	2	300	500
2	50	5	50	475
3	75	1	400	450
4	100	3.5	100	350
5	200	0.5	200	380

小结 设计的实验方案如上表所示。

第一次实验条件：500℃下，10mg 2wt% Co/SiO_2，300mg HAP。

第二次实验条件：470℃下，50mg 5wt% Co/SiO_2，50mg HAP。

第三次实验条件：450℃下，75mg 1wt% Co/SiO_2，400mg HAP。

第四次实验条件：350℃下，100mg 3.5wt% Co/SiO_2，100mg HAP。

第五次实验条件：380℃下，200mg 0.5wt% Co/SiO_2，200mg HAP。

6 灵敏度分析

根据问题二中的分析可知，温度等自变量对因变量的变化有着强烈的影响，因此我们通过对 C4 烯烃收率的回归方程中温度的系数进行灵敏度分析，使其值经过上下 5%的数据波动，绘制出 C4 烯烃收率的变化情况，如下图所示：

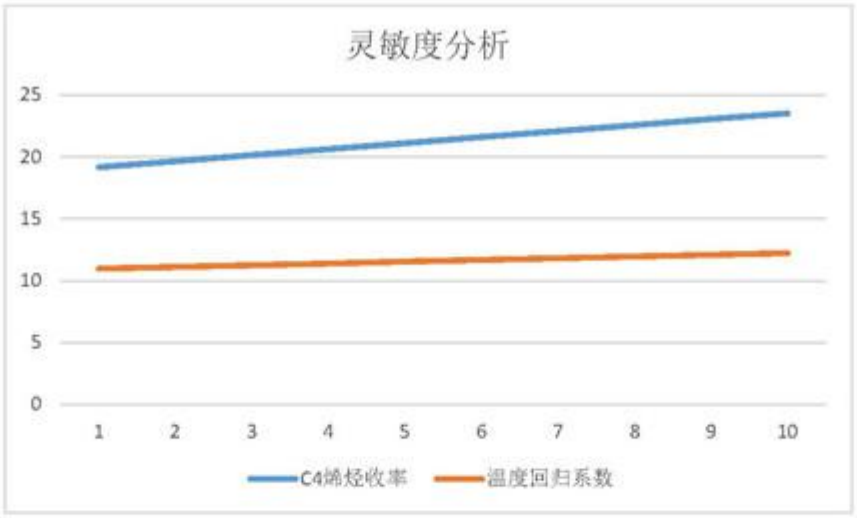


图 10：灵敏度分析图

温度的回归系数在 10%的变化区间内引起了 C4 烯烃收率 18%的变化，比较稳定，因此表明模型具有较好的灵敏性。

表 17：灵敏度数据变化表

C4 烯烃收率	温度回归系数	C4 烯烃收率	温度回归系数
19.176759	10.982	21.604359	11.6756
19.662279	11.12072	22.089879	11.81432
20.147799	11.25944	22.575399	11.95304
20.633319	11.39816	23.060919	12.09176
21.118839	11.53688	23.546439	12.23048

7 模型的评价与推广

7.1 模型的优点

1. 方差分析将有助于探究自变量之间的相互作用，从而探究更深层次自变量之间的联系，以此将存在相互作用关系的每一组变量存为新的一个变量，反过来作为多元函数中的一员，提供更好的拟合效果。
2. 偏最小二乘法和多元线性回归两者相结合进行对比分析，可以探究多个因变量之间是否存在着更深层次的联系，更有利于化学实验型数据的拟合，使其得到更为准确的拟合方程，探究实验物之间的联系，对实验结果进行预测。

7.2 模型的缺点

1. 基于多元线性回归与方差分析的 C4 烯烃最大收率模型中，各个组成部分的权重的确定仍然是按照变量贡献率特点进行分析，并无一个完整的体系，含有一定的主观性。
2. 通过偏最小二乘法进行研究时，只是对应线性多元回归中的乙醇转化率和 C4 烯烃选择性这两个因变量进行分析，得出了因变量之间相关性不大，但模型没有分析其他因变量与只是对应线性多元回归中的乙醇转化率和 C4 烯烃选择性是否存在着隐含关系。存在一定的偶然性。
3. 在聚类分析时，只是单纯的量化考虑了数值上的聚类，而忽略了其内部化学结构之间联系，使结果偏离于真实情况。

7.3 模型的推广

本文问题二中主要应用了多元线性回归以及偏最小二乘回归分析模型。从横向来看，在加入了偏最小二乘法回归分析对于多个因变量，有利于研究多因变量之间的关系，更深层的关注多个因变量之间的潜在相关性，有助于我们分析在化学反应时，更全面的分析各个生成物之间的关系，此时可以再对该模型进行延伸，在医学诊断、化学实验预测中发挥更大的作用，做到“未卜先知”。

8 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 C_4 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 C_4 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.

[6] 石彦国,单彤彤,曾剑华,刘琳琳,朱秀清. 基于主成分分析和偏最小二乘法的蒸煮大豆食味品质评价[J]. 中国食品学报, 2019,19(10):265-277.

[7] 李会芳,刘静婷,郎霞,路晓娟. 基于偏最小二乘法关联分析栀子不同炮制品化学成分与肝肾毒性[J]. 药物评价研究,2021,44(09):1890-1896.

[8] 浦仕彪,王钺涵,杨竹雅,刘录,周志宏. 均匀设计法对临床方剂胃肠炎宁颗粒组方优化研究[J]. 云南民族大学学报(自然科学版),2021,30(04):343-346.

[9] 均匀设计及其应用[J]. 方开泰. 数理统计与管理. 1994(01)

附　录

A．问题一第(1)问结论

催化剂组合分类表

组别	催化剂组合	分类依据
1	A1—A6	Co 负载量
2	A7—A12	乙醇浓度
3	B1—B6	Co/SiO2 和 HAP 装料比
4	A13—A14 A7—A8	特殊对照实验

Class 1：

在 A1 到 A6 的实验组中，催化剂组合各分组的质量比为 1，因此本文将其作为一类实验进行分析。

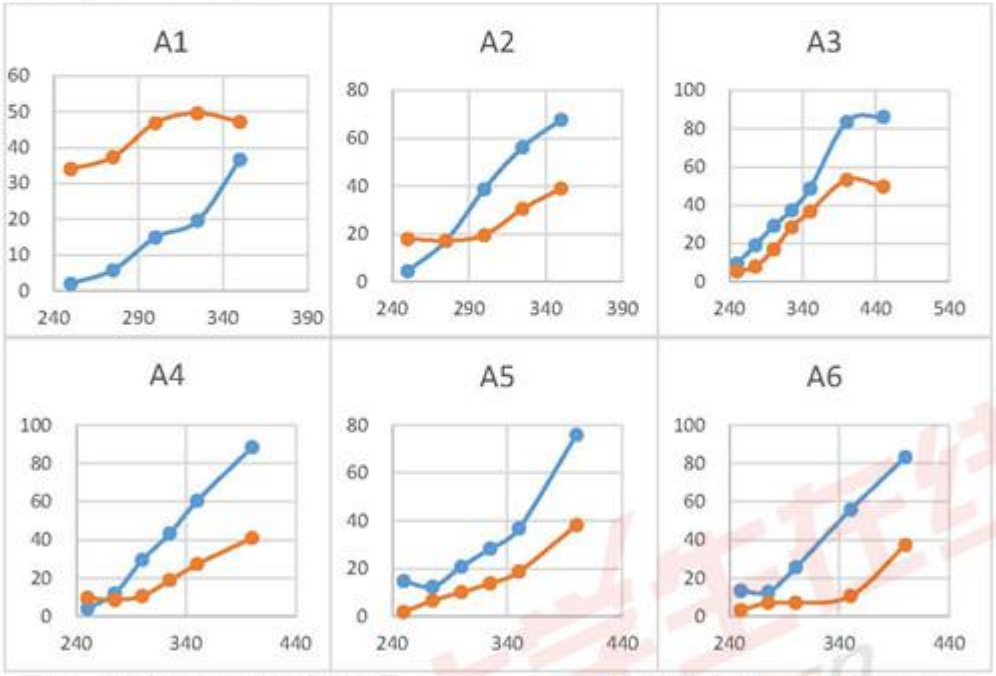


图 11：催化剂组合 A1-A7 下相关性图

根据催化剂组合特点，可以将 A1、A2、A4、A6 作为一组的对照实验，对照变量为 Co/SiO2 的含量，根据图像变化特点可以分析得出，当 Co/SiO2 的负载量为 1wt%且在温度区间为(250℃, 300℃)时，C4 烯烃的选择性远高于其他含量的情况，但随着温度不断升高，其他组呈上升趋势，A1 组虽趋于平缓，但仍多余其他组。

而 A5 中的 Co/SiO2 含含量比 A3 中的多一倍，但乙醇相对其他组过少，导致最终 C4 选择性少于大部分对照实验，但与 Co/SiO2 过多的 A6 组的选择性类

似。

Class 2:

在催化剂组合 A7 至 A9 以及 A12 中，催化剂组分中的 Co 负载量均为 1wt%，且 1wt%Co/SiO2 与 HAP 含量均为 50mg，因此本文同样将其作为一类实验进行分析。

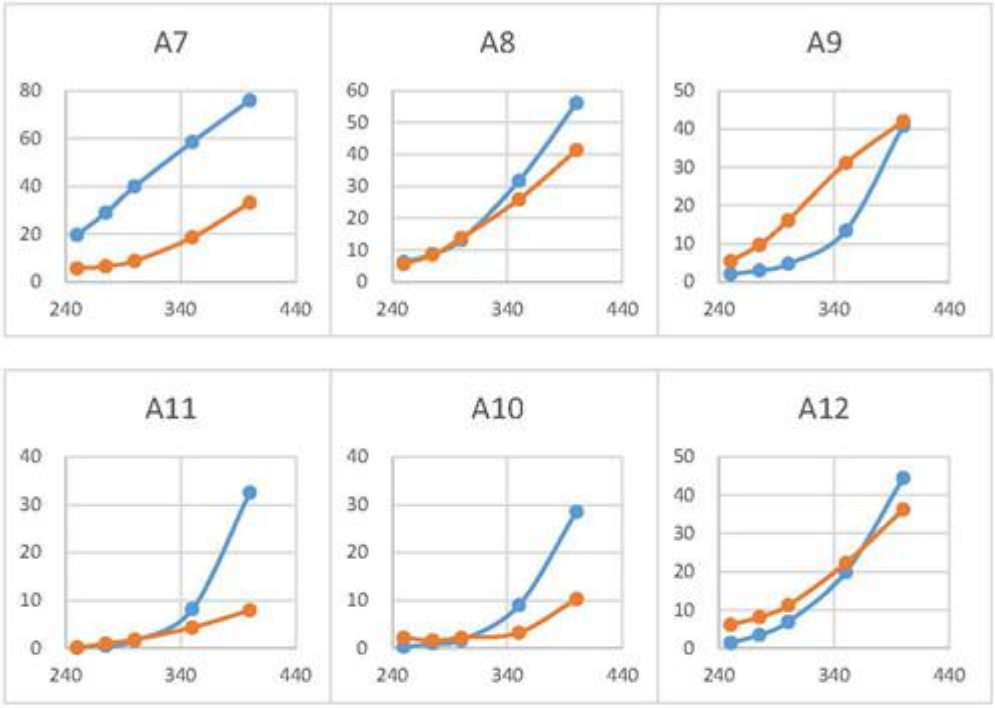


图 12：催化剂组合 A7-A12 下相关性图

根据附件 1 易知，从 A7、A8、A12 与 A9 这四种催化剂组合中，其乙醇每分钟加入速度从 0.3 ml/min 递增至 2.1ml/min。由上图可以得出，在乙醇加入速度越快的情况下，乙醇转化率逐渐降低，在温度区间(250℃, 350℃]上，C_4 选择性基本不变，在温度区间(350℃, 400℃]呈现先升高后降低的趋势。

Class 3:

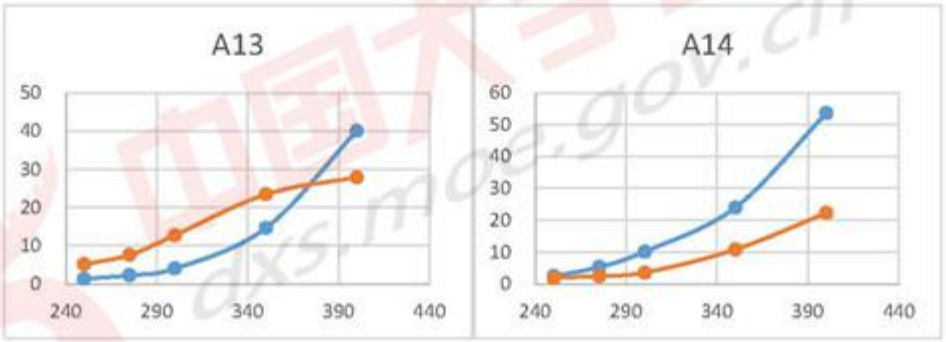


图 13：催化剂组合 A13、A14 下相关性图

在第Ⅱ种装袋方式，实验组 B1 至 B6 中，B1、B2、B3、B4、B6 中以催化剂质量比为变量进行研究，B5 为其他变量的对照试验。因此此本文同样将其作为一类实验进行分析。

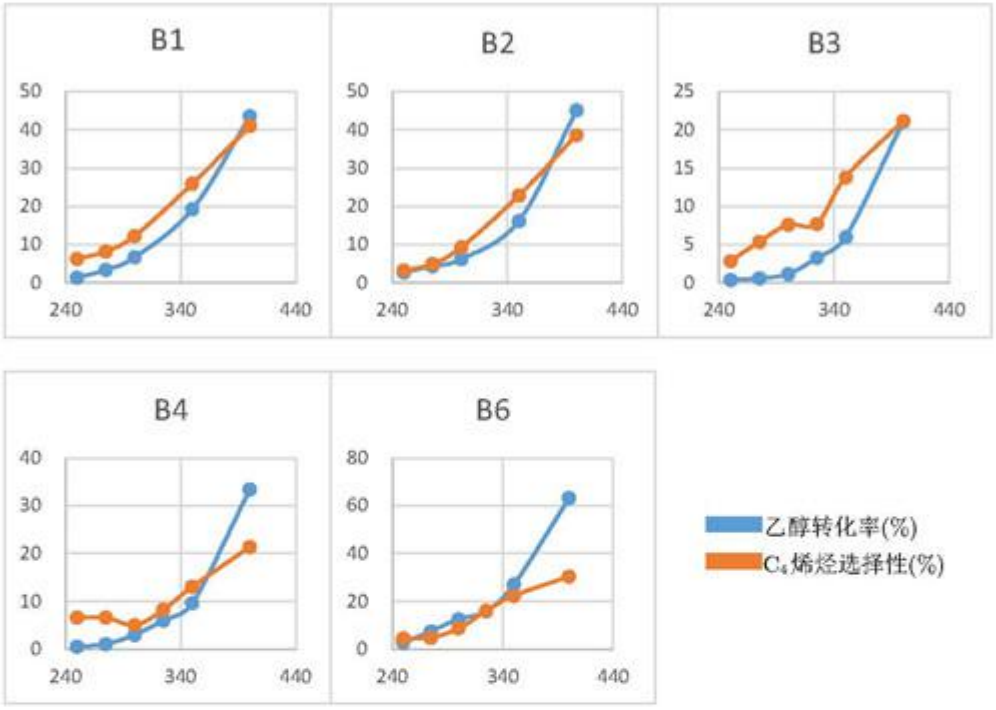


图 14：催化剂组合 B1-B4、B6 下相关性图

实验组 B3、B4、B1、B6、B2 中质量比逐渐增加，由试验结果分析可知，随着温度的升高质量比越大 C4 烯烃选择性就越高。B1 与 B2 以乙醇浓度为变量，当升高乙醇浓度时，可以得知随着温度升高，乙醇转化率与 C4 烯烃选择性都略有下降。

Class 4：

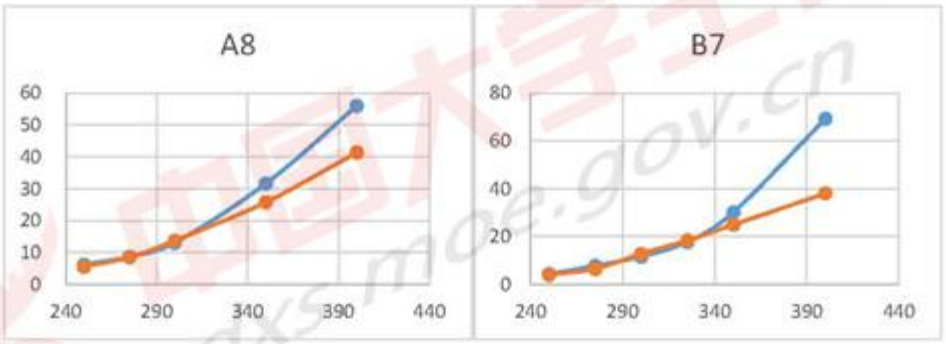


图 15：催化剂组合 A8、B7 下相关性图

在催化剂组合 A13、A14 的条件下，易知其组合的变量 1wt%Co/SiO2 与 HAP-乙醇浓度的含量不同，其含量“互换”。在该条件下，“乙醇转化率(%)”变化不大，但“C4 烯烃选择性”波动。

A8 与 B7 之间仅存在着装料方式的不同，有结果分析，两种装袋方式对乙醇转化率、C4 选择性的变化趋势的影响很小，但是经过第（罗马数字二）装袋方式的效果总体优于第一种。

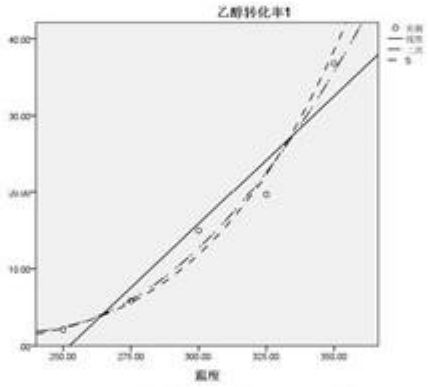
B．问题一拟合图像与相关性系数

模型摘要和参数估算值

因变量: 乙醇转化率 1

	模型摘要					参数估算值		
方程	R^2	F	自由度 1	自由度 2	显著性	常量	b1	b2
线性	.932	41.204	1	3	.008	-84.074	.333	
二次	.980	48.340	2	2	.020	141.807	-1.194	.003
S	.983	178.598	1	3	.001	10.708	-2470.152	

自变量为 温度。

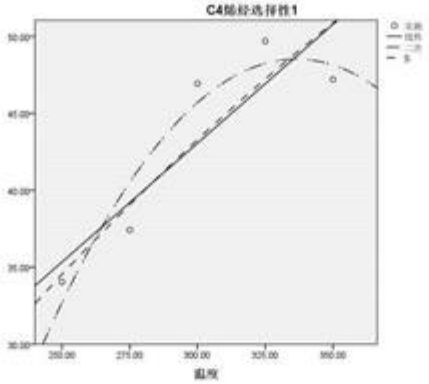


模型摘要和参数估算值

因变量: C4 烯烃选择性 1

	模型摘要					参数估算值		
方程	R^2	F	自由度 1	自由度 2	显著性	常量	b1	b2
线性	.787	11.079	1	3	.045	-3.242	.154	
二次	.916	10.899	2	2	.084	-190.783	1.422	-.002
S	.848	16.784	1	3	.026	4.898	-338.868	

自变量为 温度。

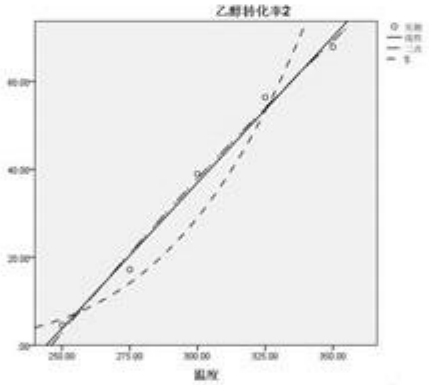


模型摘要和参数估算值

因变量: 乙醇转化率 2

	模型摘要					参数估算值		
方程	R^2	F	自由度 1	自由度 2	显著性	常量	b1	b2
线性	.990	297.825	1	3	.000	-161.892	.663	
二次	.991	111.381	2	2	.009	-227.415	1.106	-.001
S	.944	50.190	1	3	.006	11.258	-2366.823	

自变量为 温度。

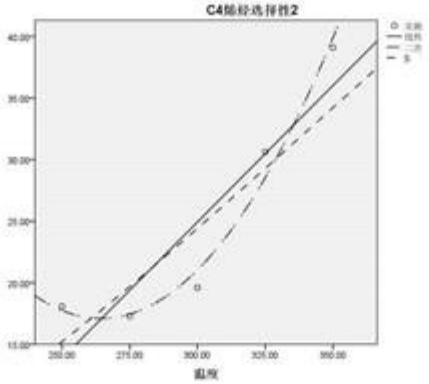


模型摘要和参数估算值

因变量: C4 烯烃选择性 2

	模型摘要					参数估算值		
方程	R^2	F	自由度 1	自由度 2	显著性	常量	b1	b2
线性	.836	15.285	1	3	.030	-41.546	.222	
二次	.980	49.722	2	2	.020	234.745	-1.646	.003
S	.782	10.758	1	3	.046	5.559	-709.059	

自变量为 温度。

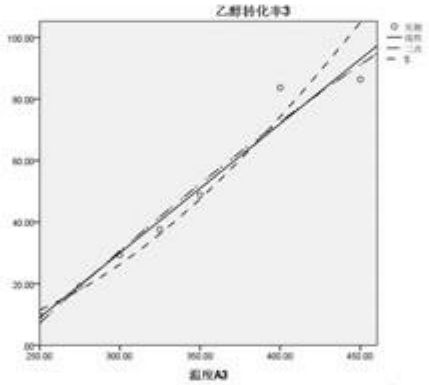


模型摘要和参数估算值

因变量: 乙醇转化率 3

	模型摘要					参数估算值		
方程	R^2	F	自由度 1	自由度 2	显著性	常量	b1	b2
线性	.964	134.968	1	5	.000	-95.828	.419	
二次	.966	57.371	2	4	.001	-134.189	.647	.000
S	.974	185.577	1	5	.000	7.421	-1245.588	

自变量为 温度 A3。

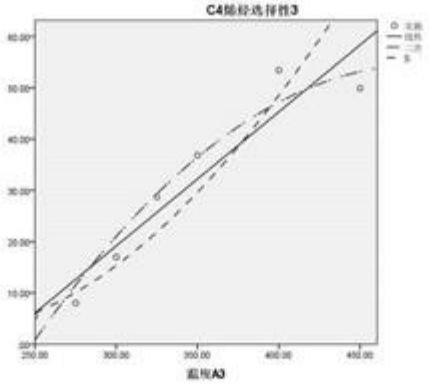


模型摘要和参数估算值

因变量: C4 烯烃选择性 3

	模型摘要					参数估算值		
方程	R^2	F	自由度 1	自由度 2	显著性	常量	b1	b2
线性	.913	52.349	1	5	.001	-59.195	.261	
二次	.955	42.519	2	4	.002	-171.076	.924	-.001
S	.931	67.340	1	5	.000	7.333	-1380.400	

自变量为 温度 A3。

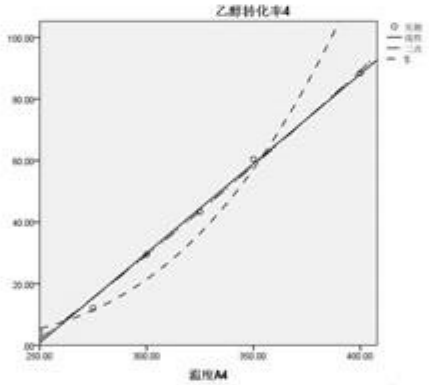


模型摘要和参数估算值

因变量: 乙醇转化率 4

	模型摘要					参数估算值		
方程	R^2	F	自由度 1	自由度 2	显著性	常量	b1	b2
线性	.995	801.582	1	4	.000	-144.540	.582	
二次	.996	362.876	2	3	.000	-107.611	.349	.000
S	.951	77.767	1	4	.001	9.953	-2066.750	

自变量为 温度 A4。

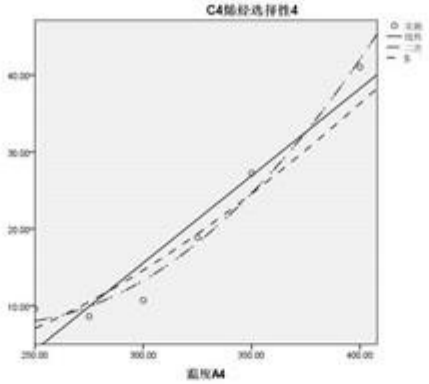


模型摘要和参数估算值

因变量: C4 烯烃选择性 4

	模型摘要					参数估算值		
方程	R^2	F	自由度 1	自由度 2	显著性	常量	b1	b2
线性	.917	44.368	1	4	.003	-52.411	.227	
二次	.977	62.381	2	3	.004	72.773	-.562	.001
S	.871	27.095	1	4	.006	6.317	-1090.270	

自变量为 温度 A4。

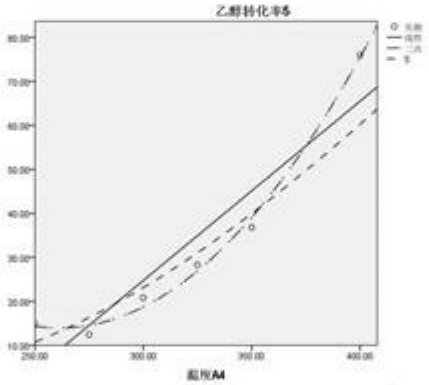


模型摘要和参数估算值

因变量: 乙醇转化率 5

	模型摘要					参数估算值		
方程	R^2	F	自由度 1	自由度 2	显著性	常量	b1	b2
线性	.873	27.503	1	4	.006	-97.593	.408	
二次	.994	249.910	2	3	.000	232.407	-1.672	.003
S	.882	29.952	1	4	.005	6.973	-1149.781	

自变量为 温度 A4。

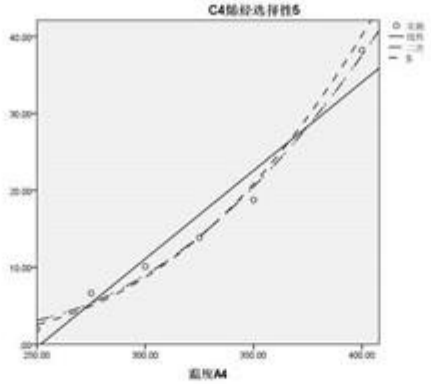


模型摘要和参数估算值

因变量: C4 烯烃选择性 5

	模型摘要					参数估算值		
方程	R^2	F	自由度 1	自由度 2	显著性	常量	b1	b2
线性	.940	62.768	1	4	.001	-57.813	.230	
二次	.991	157.173	2	3	.001	57.884	-.500	.001
S	.963	103.385	1	4	.001	8.276	-1833.200	

自变量为 温度 A4。

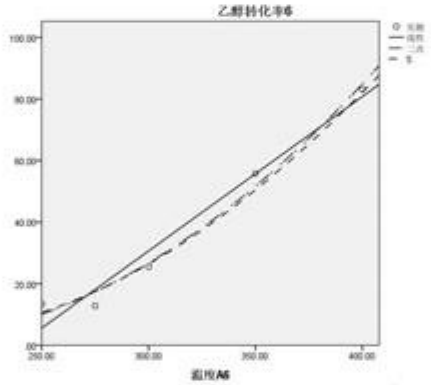


模型摘要和参数估算值

因变量: 乙醇转化率 6

	模型摘要					参数估算值		
方程	R^2	F	自由度 1	自由度 2	显著性	常量	b1	b2
线性	.968	89.414	1	3	.003	-119.731	.501	
二次	.986	70.012	2	2	.014	51.004	-.577	.002
S	.943	49.586	1	3	.006	7.814	-1362.320	

自变量为 温度 A6。

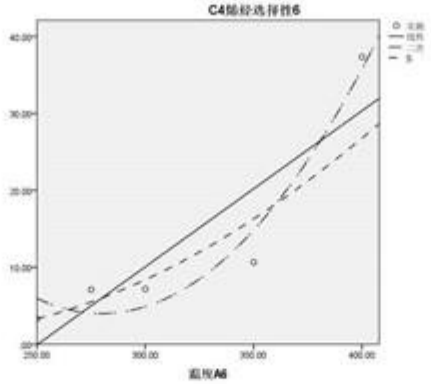


模型摘要和参数估算值

因变量: C4 烯烃选择性 6

	模型摘要					参数估算值		
方程	R^2	F	自由度 1	自由度 2	显著性	常量	b1	b2
线性	.784	10.886	1	3	.046	-50.748	.203	
二次	.945	17.316	2	2	.055	176.616	-1.234	.002
S	.880	21.968	1	3	.018	6.770	-1392.868	

自变量为 温度 A6。

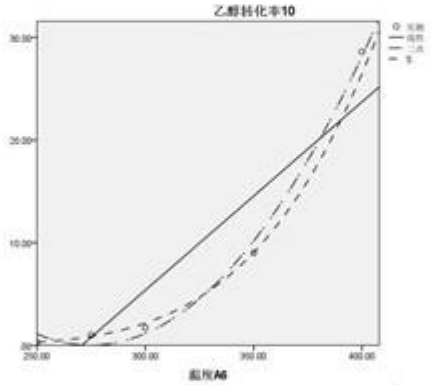


模型摘要和参数估算值

因变量: 乙醇转化率 10

	模型摘要					参数估算值		
方程	R^2	F	自由度 1	自由度 2	显著性	常量	b1	b2
线性	.852	17.250	1	3	.025	-49.688	.184	
二次	.994	162.912	2	2	.006	135.513	-.986	.002
S	.994	461.766	1	3	.000	10.792	-3006.158	

自变量为 温度 A6。

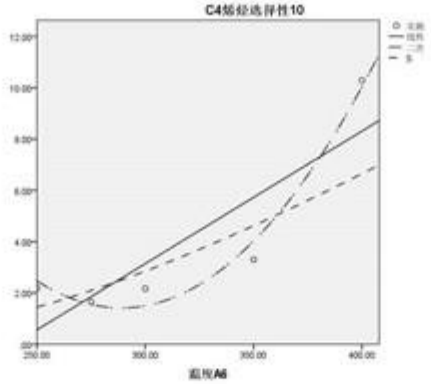


模型摘要和参数估算值

因变量: C4 烯烃选择性 10

	模型摘要					参数估算值		
方程	R^2	F	自由度 1	自由度 2	显著性	常量	b1	b2
线性	.742	8.633	1	3	.061	-12.357	.052	
二次	.978	45.444	2	2	.022	59.711	-.404	.001
S	.709	7.312	1	3	.074	4.434	-1015.963	

自变量为 温度 A6。

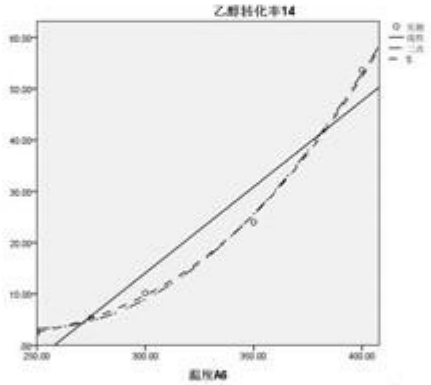


模型摘要和参数估算值

因变量: 乙醇转化率 14

	模型摘要					参数估算值		
方程	R^2	F	自由度 1	自由度 2	显著性	常量	b1	b2
线性	.929	39.286	1	3	.008	-86.644	.336	
二次	.997	340.958	2	2	.003	137.880	-1.083	.002
S	.999	2735.459	1	3	.000	9.004	-2018.312	

自变量为 温度 A6。

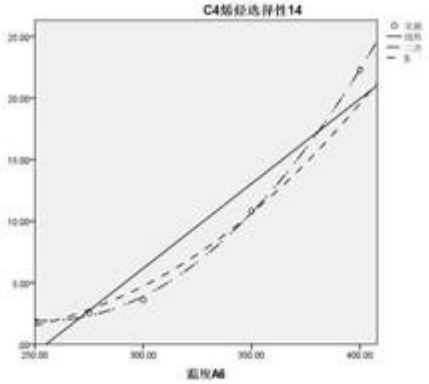


模型摘要和参数估算值

因变量: C4 烯烃选择性 14

	模型摘要					参数估算值		
方程	R^2	F	自由度 1	自由度 2	显著性	常量	b1	b2
线性	.920	34.486	1	3	.010	-35.116	.138	
二次	.999	1862.218	2	2	.001	64.863	-.494	.001
S	.966	84.314	1	3	.003	7.238	-1705.611	

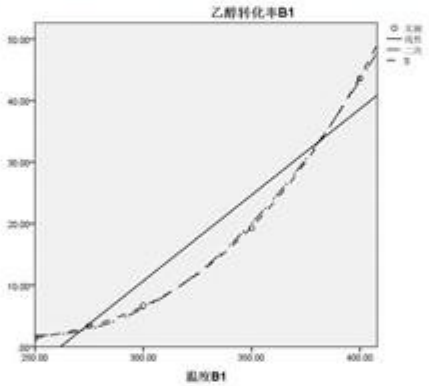
自变量为 温度 A6。



模型摘要和参数估算值

因变量: 乙醇转化率 B1

	模型摘要					参数估算值	
方程	R^2	F	自由度 1	自由度 2	显著性	常量	b1
线性	.925	37.225	1	3	.009	-73.190	.280
二次	.999	899.904	2	2	.001	121.497	-.950
S	1.000	10189.576	1	3	.000	9.477	-2277.798

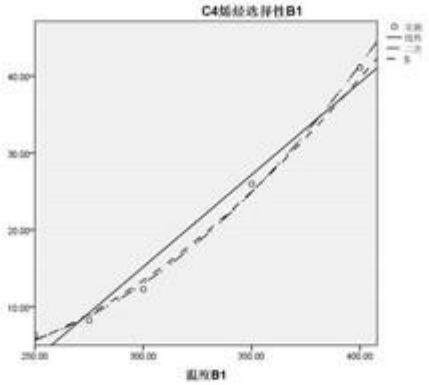


模型摘要和参数估算值

因变量: C4 烯烃选择性 B1

	模型摘要					参数估算值		
方程	R^2	F	自由度 1	自由度 2	显著性	常量	b1	b2
线性	.972	103.253	1	3	.002	-56.728	.240	
二次	.997	353.204	2	2	.003	39.068	-.365	.001
S	.987	225.337	1	3	.001	6.934	-1300.576	

自变量为 温度 B1。

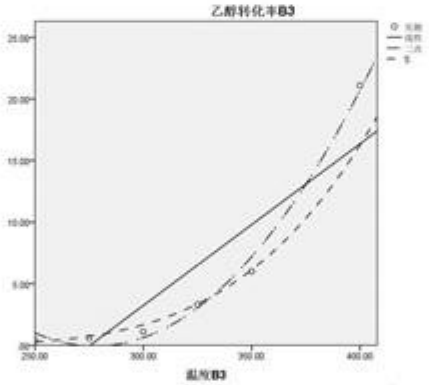


模型摘要和参数估算值

因变量: 乙醇转化率 B3

	模型摘要					参数估算值		
方程	R^2	F	自由度 1	自由度 2	显著性	常量	b1	b2
线性	.792	15.230	1	4	.018	-36.184	.131	
二次	.992	177.563	2	3	.001	107.209	-.772	.001
S	.962	100.473	1	4	.001	9.617	-2731.411	

自变量为 温度 B3。

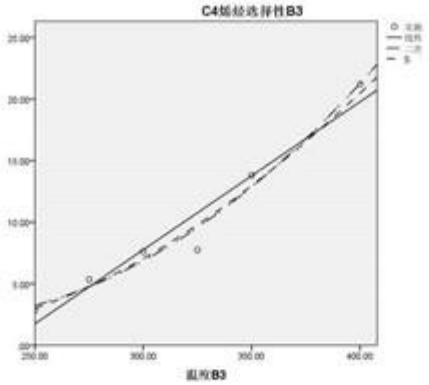


模型摘要和参数估算值

因变量: C4 烯烃选择性 B3

	模型摘要					参数估算值		
方程	R^2	F	自由度 1	自由度 2	显著性	常量	b1	b2
线性	.943	65.903	1	4	.001	-28.294	.120	
二次	.976	61.782	2	3	.004	20.970	-.190	.000
S	.967	117.810	1	4	.000	6.219	-1279.868	

自变量为 温度 B3。

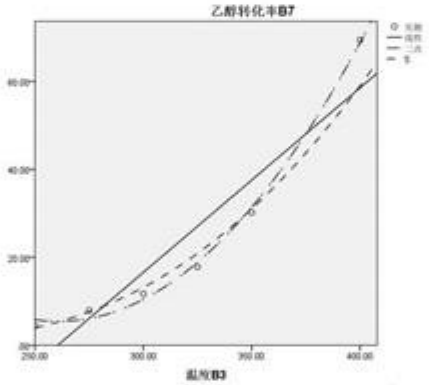


模型摘要和参数估算值

因变量: 乙醇转化率 B7

	模型摘要					参数估算值		
方程	R^2	F	自由度 1	自由度 2	显著性	常量	b1	b2
线性	.876	28.342	1	4	.006	-109.343	.420	
二次	.997	439.470	2	3	.000	228.711	-1.711	.003
S	.984	241.789	1	4	.000	8.581	-1801.328	

自变量为 温度 B3。

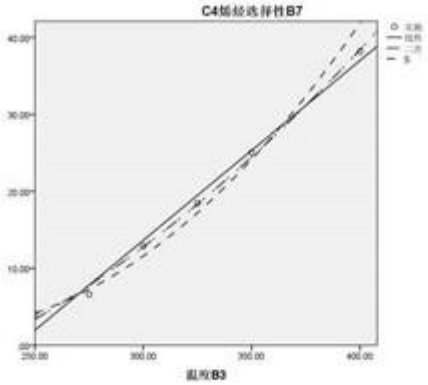


模型摘要和参数估算值

因变量: C4 烯烃选择性 B7

	模型摘要					参数估算值		
方程	R^2	F	自由度 1	自由度 2	显著性	常量	b1	b2
线性	.989	353.933	1	4	.000	-56.451	.234	
二次	.997	523.678	2	3	.000	-9.858	-.060	.000
S	.990	402.233	1	4	.000	7.582	-1539.007	

自变量为 温度 B3。



C. 模型建立过程中所使用的源代码

1. lingo 程序求得最优解：

```
sets:
parm/all/1..6/:p1,p2,p3,p4;
endsets

data:
p1=-79.844,0.05542,0.03508,0.05383,-9.1856,0.33232;
p2=-48.3594,0.0517,-3.217,0.0386,2.4439,0.1807;
p3=184.268,-0.1032,1.85520,0.0287,-7.0659,-0.3518;
p4=-35.908,0.05148,1.3623,-0.0673,4.6220,0.17108;
enddata

!max = (-3619.647+1.236*x1-69.431*x2+4.121*x3-
1276.736*x4+11.995*x5)*a+b*y1*y2+(x5*x2+x5*x4)*c;
!max = (-3619.647+1.236*x1-69.431*x2+4.121*x3-
1276.736*x4+11.995*x5)*0.7+0.6*y1*y2+(x5*x2+x5*x4)*0.5;
max = ((-6314.470+3.910*x1-154.727*x2+3.933*x3-
312.063*x4+21.375*x5)*0.5+0.4*y1*y2+0.55*(x5*x2+x5*x4))/100;
!max = (-6314.470+3.910*x1-154.727*x2+3.933*x3-
312.063*x4+21.375*x5)*0.5+0.5*y1*y2;

y1=p1(1)+p1(2)*x1+p1(3)*x2+p1(4)*x3+p1(5)*x4+p1(6)*x5;
y2=p2(1)+p2(2)*x1+p2(3)*x2+p2(4)*x3+p2(5)*x4+p2(6)*x5;
```

```
y3=p3(1)+p3(2)*x1+p3(3)*x2+p3(4)*x3+p3(5)*x4+p3(6)*x5;
y4=p4(1)+p4(2)*x1+p4(3)*x2+p4(4)*x3+p4(5)*x4+p4(6)*x5;


y2+y3+y4¡=100;
@bnd(10,x1,200);
@bnd(0.5,x2,5);
@bnd(10,x3,200);
@bnd(0.3,x4,2.1);
@bnd(250,x5,350);
```

2. 第二问偏最小二乘代码.m

```
clc,clear
ab0= xlsread('pz.xls'); % 原始数据存放在纯文本文件 pz.txt 中
mu = mean( ab0);sig=std(ab0); % 求均值和标准差
rr= corrcoef(ab0)%求相关系数矩阵
ab= zscore(ab0); %数据标准化
a=ab(:,[1:5]);b=ab(:,[6:end]); % 提出标准化后的自变量和因变量数据
[XL,YL,XS,YS,BETA,PCTVAR,MSE,stats] =plsregress(a,b)
xw=a"XS %求自变量提出成分系教每列对应一个成分,这里 xw 等于 stats.w
yw =b"YS %求因变量提出成分的系数
ncomp =input('ncomp =');
[XL2,YL2,XS2,YS2,BETA2,PCTVAR2,MSE2,stats2] =plsregress(a,b,ncomp)
n=size(a,2);m=size(b,2);%n 是自变量的个数,m 是因变量的个数
beta3(1,:) = mu(n +1:end)-mu(1:n)./sig(1:n)*BETA2([2:end],:).*sig(n+1:end);%原始数据回归方程的常数项
beta3([2:n+1],:) =(1./sig(1:n))'*sig(n +1:end).*BETA2([2:end],:)%计算原始变量 x1,…; xn 的系数,每一列是一个回归方程
hold on;
%x3str=-'乙醇转化率','C4 烯烃转化率','乙醛选择性','脂肪醇','甲醛甲醛','其他生成物','乙烯选择性'"; %新坐标的值
%c11 = categorical();
bar(BETA2','k')%画直方图
```

```
    title('''偏最小二乘回归系数直方图''')
    yhat = repmat(beta3(1,:),[size(a,1),1]) +ab0(:,[1:n])* beta3([2:end],:)%求
y1, , ym 的预测值
    ymax =max([yhat,ab0(:,[n+1;end])]);
    %求预测值和观测值的最大值
    %下面画 y1,y2,y3 的预测图，并画直线 y=X
    figure, subplot(1,2,1)
    a1=[0:ymax(1)];
    a2=[0:ymax(1)];
    b1=[0:ymax(2)];
    b2=[0:ymax(2)];
    plot(yhat(:,1),ab0(:,n+1),'*',a1,a2,'color','k')
    title('乙醇转化率预测图')
    subplot(1,2,2)
    plot(yhat(:,2),ab0(:,n+2),'O',b1,b2,'color','k')
    title('烯烃转化率预测图')
  %legend('烯烃',2)
```

3. mse.ipynb

```
{
 "cells": [
  {
   "cell_type": "code",
   "execution_count": 66,
   "metadata": {
    "collapsed": true
   },
   "outputs": [],
   "source": [
    "import pandas as pd\n",
    "import numpy as np\n",
    "import matplotlib"
   ]
  },
```

```
-
 "cell type": "code",
 "execution count": 68,
 "outputs": [],
 "source": [
  "# 自己把所有的 X Y（只有两列，乙醇转化率 烯烃选择性）“n",
  "X self = pd.read excel(“"./X.xlsx“")“n",
  "X = X self.values“n",
  "Y self = pd.read excel(“"./Y.xlsx“")“n",
  "Y = Y self.values“n",
  "# print(3^2)"
 ],
 "metadata": -
  "collapsed": false,
  "pycharm": -
   "name": "#%% 读取数据“n"
  "
 "
",
-
 "cell type": "code",
 "execution count": 90,
 "outputs": [],
 "source": [
  "sum 2 = 0.0“n",
  "sum 1 = 0.0“n",
  "sum 3 = 0.0“n",
  "sum 4 = 0.0“n",
  "for i in range(109):“n",
  "      y 1 = -79.8445 + X[i][0]*0.0554 + X[i][1]*(-0.0351) +
X[i][2]*(0.0538) + X[i][3]*(-9.1856) + X[i][4]*0.3323“n",
  "      y 2 = -48.3595 + X[i][0]*0.0517 + X[i][1]*(-3.2176) +
X[i][2]*(0.0387) + X[i][3]*(-2.4439) + X[i][4]*0.1807“n",
```

```
        "      sy'1 = -80.218 + X[i][0]*(-0.033) + X[i][1]*(0.134) +
X[i][2]*(0.141) + X[i][3]*(-8.765) + X[i][4]*0.333\n",
        "      sy'2 = -54.118 + X[i][0]*0.005 + X[i][1]*(-3.173) + X[i][2]*(0.091)
+ X[i][3]*(2.670) + X[i][4]*0.181\n",
        "      delta'1 = (y'1 - Y[i][0])**2\n",
        "      delta'2 = (y'2 - Y[i][1])**2\n",
        "      delta'3 = (sy'1 - Y[i][0])**2\n",
        "      delta'4 = (sy'2 - Y[i][1])**2\n",
        "      sum'1 += delta'1\n",
        "      sum'2 += delta'2\n",
        "      sum'3 += delta'3\n",
        "      sum'4 += delta'4\n",
        "print(\"MSE'y1, MSE'y2 In PLS ::\", (sum'1/109)**(1/2),
(sum'2/109)**(1/2))\n",
        "print(\"MSE'y1, MSE'y2 In SPSS::\", (sum'3/109)**(1/2),
(sum'4/109)**(1/2))"
       ],
       "metadata": {
        "collapsed": false,
        "pycharm": {
         "name": "#%% 求解 MSE\n"
        }
       }
      },
      {
       "cell'type": "code",
       "execution'count": 72,
       "outputs": [
        {
         "name": "stdout",
         "output'type": "stream",
         "text": [
          "0\n",
          "1\n",
```

```
      "3\"n"
     ]
    "
   ],
   "source": [],
   "metadata": {
    "collapsed": false,
    "pycharm": {
     "name": "#%%\"n"
    }
   }
  },
  {
   "cell_type": "code",
   "execution_count": null,
   "outputs": [],
   "source": [],
   "metadata": {
    "collapsed": false,
    "pycharm": {
     "name": "#%%\"n"
    }
   }
  }
 ],
 "metadata": {
  "kernelspec": {
   "display_name": "Python 3",
   "language": "python",
   "name": "python3"
  },
  "language_info": {
   "codemirror_mode": {
```

```
      "name": "ipython",
      "version": 2
    },
    "file_extension": ".py",
    "mimetype": "text/x-python",
    "name": "python",
    "nbconvert_exporter": "python",
    "pygments_lexer": "ipython2",
    "version": "2.7.6"
  }
 },
 "nbformat": 4,
 "nbformat_minor": 0
}
```