

基于多元线性回归的 C₄ 烯烃的制备研究

摘要

从乙醇出发制备 C₄ 烯烃是取代传统化石能源制备的有效路线，但由于转化过程中产物种类多、分布广导致乙醇转化率和 C₄ 烯烃选择性低，人们设计一种 Co/SiO₂-HAP 催化剂应用于催化乙醇偶合制备 C₄ 烯烃。本文采用催化剂组合、温度与乙醇转化率、C₄ 烯烃选择性、C₄ 烯烃收率的合理数值，建立统计回归模型，分析催化剂组合与温度对实验结果的影响，对探究 C₄ 烯烃制备工艺条件具有一定参考价值。

针对问题一：首先发现每组催化剂组合中乙醇转化率、C₄ 烯烃选择性与温度存在一定的关系，利用 MATLAB 的拟合工具箱 Curve Fitting，用线性函数、指数函数、多项式函数和 Fourier 函数对数据进行拟合，分析结果为：乙醇转化率会随着反应温度单调上升，C₄ 烯烃选择性在大部分情况下随温度上升，在小部分催化剂组合下会出现峰值再回落的现象。对附件 2 数据的折线图初步分析，利用 MATLAB 的拟合工具箱 Curve Fitting 得到各个产物选择性关于时间的拟合函数 $f_1(t) \sim f_5(t)$ ，分析结果为：乙醇转化率随时间推移降低，乙烯和选择性都随时间推移增加，其余产物变化较不规则。最后分析拟合效果：所有线性拟合函数的校正的 R^2 均值大于 0.9，非线性拟合函数均方根误差接近 0。

针对问题二：首先根据附件 1 数据，分析出乙醇转化率和 C₄ 烯烃选择性的 6 个影响因素，基于控制变量原理，作出折线图进行定性分析。建立关于催化剂组合与温度和乙醇转化率与 C₄ 烯烃选择性的多元线性回归模型，利用稳健标准误解决异方差问题，并得到方差膨胀因子 $VIF_{max}=1.39$ ，因变量间相关性小。最终得到，对乙醇转化率：催化剂质量和温度起正向作用，乙醇浓度起负向作用，装料方式、Co 负载量和装料比与其线性相关关系不显著；对 C₄ 烯烃选择性：A 类装料方式效果更好，乙醇浓度、催化剂质量和温度起正向作用，Co 负载量起负向作用，装料比与其线性相关水平不显著。

针对问题三：首先作出数据的柱状图进行定性分析，对自变量进行预处理并引入非线性自变量，建立影响 C₄ 烯烃收率的非线性回归模型，运用稳健标准误解决异方差问题。分析结果如下：欲使 C₄ 烯烃收率最高，应使用 A 类装料方式、Co/SiO₂ 质量 200mg、与 HAP 的混料比 1:1、Co 负载量 2wt%、乙醇浓度 0.9ml/min 和反应温度 400°C；当要求温度低于 350°C 时，应使用 A 类装料方式、Co/SiO₂ 质量 200mg、与 HAP 的混料比 1:1、Co 负载量 1wt% 和 350°C 的反应温度。

针对问题四：我们采用“局部搜索”的思路设计 5 次新实验的参数：以 C₄ 烯烃收率作为实验结果优劣的评判标准，将装料方式、Co/SiO₂ 质量、乙醇浓度和温度设定为问题三分析出的最优数值，仅对剩下 2 个在回归中不显著的变量进行微调。针对不同的实验情况，我们设计了两种方案：如果需要提前确定 5 次实验的全部参数，我们给出 5 组实验参数使 Co 负载量和装料比分别在 2wt% 和 1 附近浮动；如果可以在下一次实验参数确定前获得上一次实验结果，我们规定一套参数选取的规则来利用当前实验结果决定下一次实验的参数，同时保证收敛性和结果更优化。采用上述思路在已有的最优值附近进行搜索，有利于以较少的实验次数获得 C₄ 烯烃收率的进一步改进。

关键词：C₄ 烯烃制备 多元回归模型 函数拟合 MATLAB STATA

1 问题重述

C_4 烯烃作为重要的化工原料，广泛应用于化工产品和医药中间体的生产。在传统生产中， C_4 烯烃的制备来源于化石能源，而随着化石能源的短缺以及其带来的日益严重的环境问题，人们渴望找到一种可持续的绿色新型能源用于制备 C_4 烯烃。其中，乙醇因具有污染小、来源广泛等一系列优点而被广泛研究并开始被投入于 C_4 烯烃的制备。如今，乙醇催化转化为 C_4 烯烃的研究已经获得优异成果，但由链增长反应机理可知，在制备过程中会出现产物分布广、种类杂、目标产物选择性低、应用价值差等问题。因此，为了提高 C_4 烯烃的选择性，人们通过设计合适的化学催化剂组合和温度，研究在相同实验条件下产物选择性的提高情况，以期探索乙醇催化偶合制备 C_4 烯烃的最佳工艺条件。

基于上述背景，我们将建立数学模型来解决以下问题：

问题一：根据不同催化剂组合与温度下的结果，分析乙醇转化率、 C_4 烯烃的选择性与温度的关系；根据给定温度下某一次实验中各时间点的结果，分析各产物的占比和乙醇转化率与实验进行时间的关系。

问题二：根据收集的一系列实验结果，通过分析乙醇转化率和 C_4 烯烃选择性大小在不同催化剂组合与温度下的情况探究其中关系。

问题三：在相同的实验环境、装备等条件下，给出一个催化剂组合与温度设置方案，使 C_4 烯烃收率达到最高；当要求温度低于 350 度时，给出一个催化剂组合与温度设置方案以达到同样目的。

问题四：设计 5 个新的实验，以期获得更准确的 C_4 烯烃制备最优条件。

2 问题分析

预备知识：乙醇是生产制备 C_4 烯烃的原料，在制备的过程中，催化剂组合、温度、催化剂装料方式等将会影响 C_4 烯烃的选择性和 C_4 烯烃收率。其中，温度是指反应温度，选择性是指某个产物在所有产物中的占比，例如：在装料方式 I 下，250 度时使用 200mg 1wt%Co/SiO₂- 200mg HAP-乙醇浓度 1.68ml/min 作为催化剂，乙烯选择性（%）为 1.17，是指在此条件下乙烯占所有产物的 1.17%。反应时间是指催化剂在乙醇氛围下的反应时间。乙醇的转化率是指单位时间内乙醇的单程转化率， C_4 烯烃收率的计算公式为： C_4 烯烃收率 = 乙醇转化率 × C_4 烯烃的选择性。

关于拟合工具箱的说明：Curve Fitting 是 MATLAB 内置的函数拟合工具箱，可以用于将给定的数据拟合成各种函数（如多项式函数、指数函数、Fourier 函数、正态分布函数、样条函数等），并进行不同拟合结果之间的比较。同时，工具箱还集成了输出评价指标、作图、生成代码等功能。

2.1 问题一的分析

对于 C_4 烯烃的制备，催化剂组合、温度、催化剂装料方式等将会影响实验结果产生影响。本题针对控制催化剂不变，将温度作为自变量，研究乙醇转化率以及 C_4 烯烃的选择性与温度的关系。首先对附件 1 中的数据进行处理，仅保留催化剂、乙醇转化率、 C_4 烯烃的选择性、温度四个变量，控制催化剂不变，分别作出温度与乙醇转化率、温度与 C_4 烯烃的选择性的图像，观察图像后判断温度与乙醇转化率、温度与 C_4 烯烃的选择性的关系，并利用 MATLAB 进行拟合，得到两个函数关系。

再考虑附件 2 中的 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。由于时间序列的不均匀，无法采用时间序列分析，本题考虑以时间为自变量，乙醇转化率、乙烯选择性、 C_4 烯烃选择性、乙醛选择性、碳数为 4-12 脂肪醇选择性、甲基苯甲醛和甲基苯甲醇选择性以及其他乙醇转化率为因变量，进行函数的拟合，得到乙醇转化率关于时间的函数，以及各个产物选择性与时间的关系。在研究各个产物与时间的关系时，可能出现某一个产物选择性难以拟合的情况，此时考虑产物选择性之和为 100%，通过 100% 减去其余产物的选择性质和，亦可得到该产物的选择性与之间的关系函数。

2.2 问题二的分析

先对附件 1 中数据进行描述统计，了解数据大致的分布情况，得到下图：

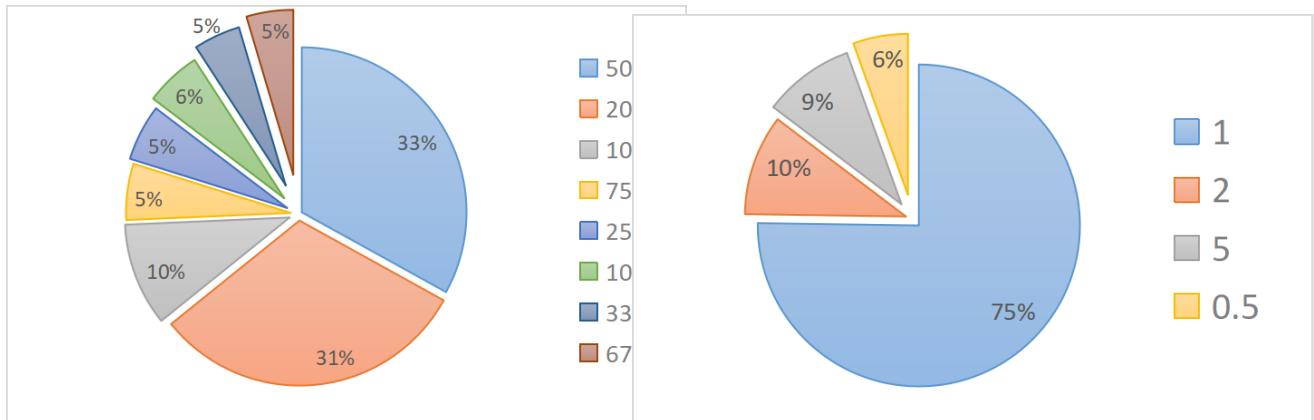


Figure 1: Co/SiO₂ 质量的饼状图

Figure 2: Co 负载量的饼状图

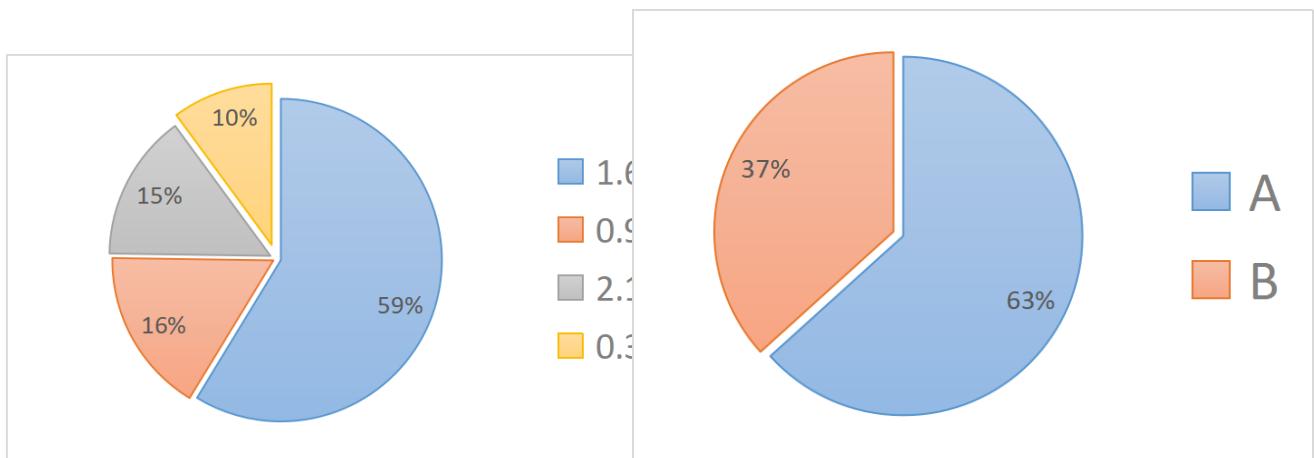


Figure 3: 乙醇浓度的饼状图

Figure 4: 装料方式的饼状图

问题二考虑乙醇转化率和 C₄ 烯烃选择性大小与不同催化剂以及温度的关系，将催化剂及其装料方式转化成五个变量，由这五个变量可以唯一确定一种催化剂与装料方式，并且五个变量之间互不相关，将温度也作为自变量，则有六个自变量共同决定乙醇转化率和 C₄ 烯烃的选择性。首先利用控制变量，画出折线图，观察乙醇转化率以及 C₄ 烯烃的选择性大小与催化剂的各个要素以及温度之间是否存在线性关系，**和具体呈正关系或负关系**；然后，可以利用多元线性回归对各自变量对乙醇转化率和 C₄ 烯烃的选择性大小的影响进行具体的分析；得到回归方程后，检验回归关系与参数的显著性，再利用怀特检验考察异方差，若扰动项存在异方差，则需要对异方差进行处理。

2.3 问题三的分析

问题三考虑 C₄ 烯烃收率与不同催化剂之间的关系。参照问题二的处理方式，用五个互不相关的变量唯一确定一种催化剂组合，加上温度，一共有六个可能的影响因素。运用多元线性回归模型，研究 C₄ 烯烃收率与这六个变量之间的线性关系。得到回归方程后，进行线性关系和回归系数的显著性以及异方差检验。同时，若回归的结果中出现了过多不显著的变量，可以尝试使用不同的指标刻画催化剂环境，比如用“总质量”和“装料比”替换“Co/SiO₂ 质量”和“HAP 质量”；或者构造交叉项、高次项进行回归。

回归及检验结束后，选择使得 C₄ 烯烃收率尽可能高的催化剂组合和温度的方法如下：对于回归结果

中显著的变量，我们选取实验中出现的最优值；对于回归结果中不显著的变量，我们将通过简单的比对后，选取实验中出现过的最优数值；或者查阅相关资料，参考最常用的数值。

如果给定了温度不高于 350°C 的条件，则选取温度不高于 350°C 的实验数据，以同样的方法和流程进行多元线性回归分析、显著性检验、异方差检验、催化剂组合及温度的选取。

2.4 问题四的分析

问题四研究增加的 5 次实验如何选择催化剂组合和温度的问题。由于我们的总目标是探索如何在化学反应中更高效地制备 C_4 烯烃，因此再增加 5 次实验的目标也应当是获得更高的 C_4 烯烃收率。在上一个问题中研究了 C_4 烯烃收率与催化剂组合、温度的关系，然而已有的实验数据总共是 109 次，远远大于增加的 5 次，所以如果把新的 5 次实验连同已有的实验数据，再次放到上一个问题的模型中进行求解，得到的新解与上一个问题的结果相比，产生的变化很可能是微小的。

因而，相比于补全已有的实验数据，新增加的 5 次实验的作用更可能是在已有的最优催化剂组合和温度附近进行局部搜索。换言之，在已有的最优数值的附近进行微调，收益可能远远高于填补原有实验数据的空缺。新的 5 次实验的参数也将按照“最优解附近局部搜索”的思路进行设计。

3 模型假设

1. 假设观测的数据的误差可忽略不计；
2. 假设每次实验中除了催化剂组合和温度，其余变量均相同；
3. 假设没有环境因素的影响：在相同大气压、空气湿度等情况下进行实验。

4 符号说明

符号	定义	单位
T	反应温度	$^{\circ}\text{C}$
t	反应时间	min
y_1	乙醇转化率	%
y_2	C_4 烯烃选择性	wt%
x_1	装料方式	\
x_2	Co 负载量	wt%
x_3	Co/SiO ₂ 质量	mg
x_4	装料比 (HAP 质量:Co/SiO ₂)	\
x_5	乙醇浓度	ml/min

注：未列出符号及重复符号以出现位置为准

Z

5 模型的建立与求解

5.1 问题一模型的建立与求解

5.1.1 第一问模型的建立与求解

首先对附件 1 中的每种催化剂组合，研究乙醇转化率与温度、 C_4 烯烃的选择性与温度的关系，分别拟合函数，并分析拟合结果。

将每组数据分别导入到 MATLAB 中后，使用 MATLAB 的函数拟合工具箱进行数据拟合。对于拟合函数的选择如下：线性函数、指数函数、不超过三次的多项式函数以及 Fourier 函数（Fourier 级数的有限和）。前两个函数可以对单调增长（下降）的数据进行不同增长（下降）速率的刻画，后两个函数可以拟合一些更复杂的数据，表现数据的波动趋势。对于每组数据，我们依据均方误差，从上述四种拟合函数中选择最优的拟合函数进行拟合；若出现指标近似的情况，则优先选择简单函数。

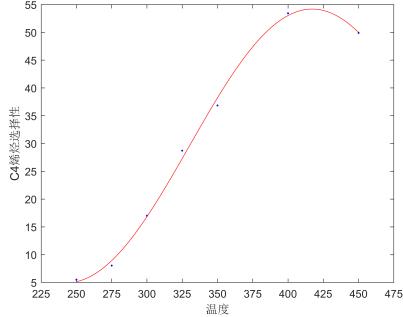


Figure 5: A₃ 催化剂中 C₄ 烯烃选择性关于 T 的拟合函数 $g_1(T)$

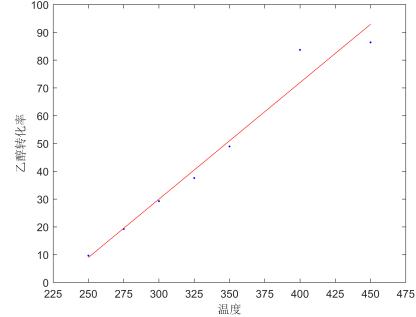


Figure 6: A₃ 催化剂中乙醇转化率关于 T 的拟合函数 $g_2(T)$

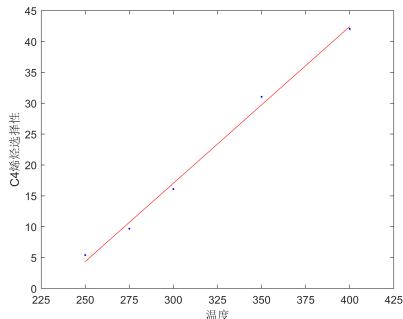


Figure 7: A₉ 催化剂中 C₄ 烯烃选择性关于 T 的拟合函数 $g_3(T)$

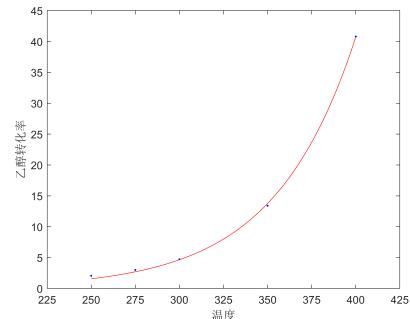


Figure 8: A₉ 催化剂中乙醇转化率关于 T 的拟合函数 $g_4(T)$

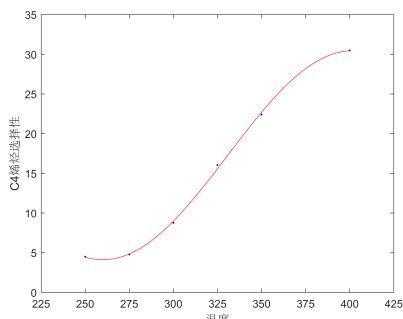


Figure 9: B₆ 催化剂中 C₄ 烯烃选择性关于 T 的拟合函数 $g_5(T)$

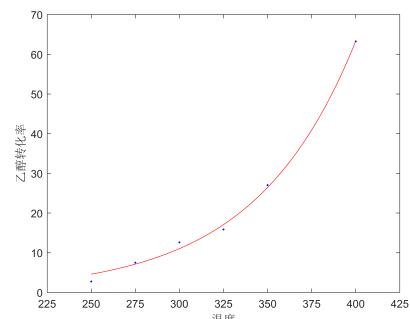


Figure 10: B₆ 催化剂中乙醇转化率关于 T 的拟合函数 $g_6(T)$

$$g_1(T) = 29.57 + 7.83 \cos(0.02T) + 23.34 \sin(0.02T)$$

$$g_2(T) = 0.42T - 95.88$$

$$g_3(T) = 0.25T - 59.1$$

$$g_4(T) = 0.007e^{0.02T}$$

$$g_5(T) = 17.29 - 11.31 \cos(0.02T) + 6.735 \sin(0.02T)$$

$$g_6(T) = 0.06e^{0.02T}$$

以上为部分的拟合结果，完整的结果见附录。在拟合的过程中，我们发现：在所有的催化剂组合下，乙醇转化率都会随着反应温度单调上升，有的环境下增长速率为线性，有的为指数；而在大部分催化剂组合下，C₄ 烯烃的选择性会随温度近似线性地上升，在小部分催化剂组合下，C₄ 烯烃的选择性会随温度的升高而出现波动，即会在低于某一温度时随温度升高而上升，在此温度达到峰值，随后下降。

同时，拟合结果的分析也对第三个问题有着一定程度的启发：若想获得最大的 C₄ 烯烃的收率，必须同时获得较大的乙醇转化率和 C₄ 烯烃的选择性，而不同的催化剂环境将会对结果有着很大的影响。在第三问中，我们最理想的催化剂组合应该是使得乙醇转化率随温度以指数增长，并使得 C₄ 烯烃的选择性随温度近似线性增长。

5.1.2 第二问模型的建立与求解

考虑 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，并对结果进行分析。

利用 MATLAB 做出乙醇转化率关于时间的变化曲线以及各种生成物关于时间的变化折线，得到下图：

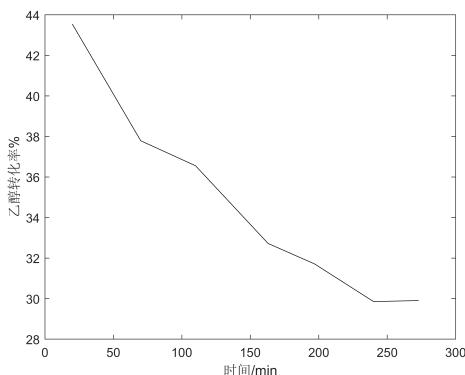


Figure 11: 乙醇转化率关于 t 的折线

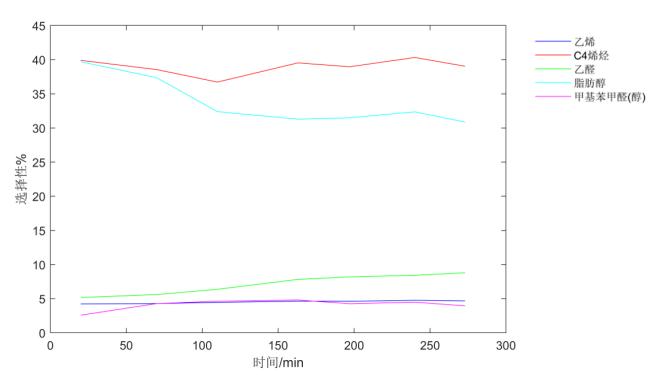


Figure 12: 各产物选择性关于 t 的折线

观察图 11 可以得到，乙醇转化率与时间有明显的关系，乙醇转化率随时间的增加而逐渐减小，乙醇转换速率放缓；观察图 12 可以得到，C₄ 烯烃的选择性与脂肪醇的选择性明显高于其余三种产物，可以猜测 C₄ 烯烃与脂肪醇是该实验的主要产物，观察随时间增加各产物的变化，发现乙醛的选择性随时间增加而增加，脂肪醇的选择性随时间增加而减少，而 C₄ 烯烃与乙烯、甲基苯甲醛（醇）的选择性与时间并没有明显的线性关系。下面考虑利用 MATLAB 分别拟合各个产物选择性关于时间的函数，从而进一步分析时间对于实验结果的影响。

对于不同产物选择性与时间的变化关系，使用 MATLAB 拟合工具箱进行拟合，拟合函数的选取方式与 5.1.1 中相同。考虑到所有产物的选择性之和为 100%，故而 C₄ 烯烃选择性关于时间的关系受到其他产物的影响，除乙烯选择性、C₄ 烯烃选择性、乙醛选择性、碳数为 4-12 脂肪醇选择性、甲基苯甲醛和甲基苯甲醇选择性以外，还有其它产物，受到此类产物的影响，C₄ 烯烃选择性难以获得较好的函数拟合。在多次实验之后，得到均方误差最小的拟合函数图像如下：

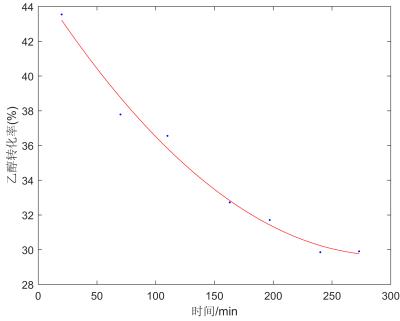


Figure 13: 乙醇转化率关于 t 的拟合函数 $f_1(t)$

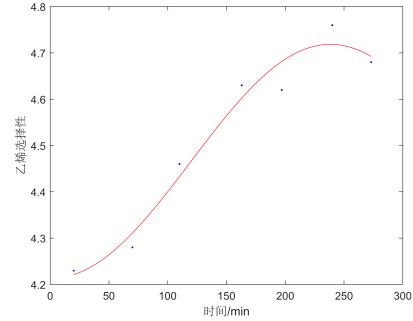


Figure 14: 乙烯选择性关于 t 的拟合函数 $f_2(t)$

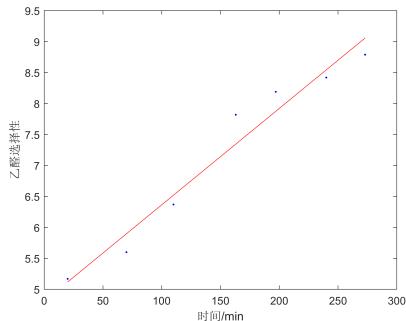


Figure 15: 乙醛选择性关于 t 的拟合函数 $f_3(t)$

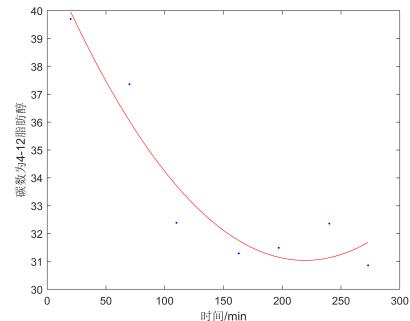


Figure 16: 脂肪醇选择性关于 t 的拟合函数 $f_4(t)$

函数关系式分别为：

$$f_1(t) = 0.0001772t^2 - 0.1051t + 45.25$$

$$f_2(t) = 4.466 - 0.2524 \cos(0.01325t) - 0.004488 \sin(0.01325t)$$

$$f_3(t) = 0.01557t + 4.808$$

$$f_4(t) = 0.0002245t^2 - 0.09848t + 41.84$$

$$f_5(t) = 4.753 \times 10^{-7}t^3 - 0.0002985t^2 + 0.05497t + 1.631$$

考察上述拟合函数的拟合优度，发现对于所有的函数关系，拟合优度 $R^2 > 0.91$ （具体回归评价指标见模型的分析与检验部分），故而可以利用上述函数作为各产物选择性关于时间的近似关系。观察图象以及函数关系式可以得到：乙醇转化率随时间的增加而降低，而乙烯选择性以及乙醛选择性都随时间增加有明显的增加；脂肪醇的选择性随时间增加而先降低，而到 230min 左右，脂肪醇的选择性又随时间的增加而产生一定的上升趋势；甲基苯甲醛（醇）的选择性在 130min 之前随时间的增加有显著的增长，而 130min 后增长放缓甚至开始下降；C₄ 烯烃的选择性在 100min 之前随时间增加有明显的增加，而 100min 之后增

长放缓，并于 150min 后下降。

5.2 问题二模型的建立与求解

5.2.1 模型的建立

step 1 对乙醇转化率与 C₄ 烯烃选择性的影响因素进行分析：

从附件 1 给出的性能数据中可以看出，在不同的催化剂组合和温度下，乙醇的转化率和各产物的选择性不同。催化剂组合中的变量包含：Co 负载量、Co/SiO₂ 和 HAP 的装料比、乙醇浓度和装料方式。同时，根据对附件 1 中数据的观察，在装料比相同的情况下，催化剂质量不同对结果的影响较大，为了防止变量的共线性，我们选取 Co/SiO₂ 和 HAP 中一中的质量作为解释变量。因装料方式为定型变量，所以我们将它设置为虚拟变量。基于上述分析，我们选取温度、Co 负载量、Co/SiO₂ 和 HAP 的装料比、Co/SiO₂、乙醇浓度和装料方式（虚拟变量）作为乙醇转化率与 C₄ 烯烃选择性的影响因素。

step 2 建立影响乙醇转化率与 C₄ 烯烃选择性的多元线性回归模型：

(1) 建立多元线性回归模型：

为了研究催化剂组合与温度和乙醇转化率与 C₄ 烯烃选择性之间的相关性，我们建立多元线性回归模型来解释预测其中关系，k 元线性回归模型的表达形式如下：

$$y = \epsilon + \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$$

其中 y 为被解释变量， ϵ 为随机误差项， x_i 为自变量（解释变量）， β_i 为回归系数。

由于我们无法了解上述表达式的具体特征，故通过样本考察总体，得到 n 组观察值 $y_i, x_{1i}, x_{2i} \dots$ ，他们分别满足表达式：

$$y_i = \epsilon_i + \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki}$$

为了方便，模型可以改写为向量形式，引入矩阵记号：

$$y = X\beta + \xi,$$

其中 $y = [y_1, \dots, y_n]^T$ ，待估计的参数向量 $\beta = [\beta_0, \dots, \beta_n]^T$ ，不可观测的随机误差向量 $\xi = [\epsilon_1, \dots, \epsilon_n]^T$ ，都是未知待定的常数向量， X 为模型设计矩阵，是常数矩阵。

(2) 回归参数的最小二乘估计：

当随机误差服从正态分布和 X 列满秩的假定下，经过最小二乘估计，希望选择 β 的估计值 b 使得残差平方和 $L = \sum(y_i - b_1 x_{1i} - \dots - b_k x_{ki})^2$ 最小，因此须满足求极值的必要条件：

$$\frac{\partial L}{\partial b_i}, i = 1, \dots, k$$

根据此关系式，可以得到关于带估计参数的 k 个方程，其矩阵表示为： $(X^T X)b = X^T Y$ 。再根据列满秩假设，可以得到 β 的估计值： $\hat{\beta} = (X^T X)^{-1} X^T Y$ 。进一步，为了满足充分性，目标函数的 Hessian 矩阵需要满足正定条件。当矩阵满秩假定成立时，可知 $\frac{\partial^2}{\partial b \partial b^T} L(\hat{\beta}) = 2X^T X$ 正定。因此，OLS 的估计值为

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

于是 y 的估计值为：

$$\hat{y} = X\hat{\beta}$$

(3) 回归模型的求解：基于上述确定的影响因素与因变量，利用 STATA 中的回归指令 regress $y x_1 \dots x_k$ 生成回归参数估计值等数据。可以先对回归系数进行解释：

对于 $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \dots + \hat{\beta}_k x_{ki}$ ，斜率系数 $\hat{\beta}_p$ 表示在其他条件保持不变的情况下，自变量 x_p 每增加 1 个单位，因变量 y 平均地提高 $\hat{\beta}_p$ 个单位（若 $\hat{\beta}_p$ 为负值，则意味着平均地减少 $|\hat{\beta}_p|$ 个单位）。

截距 $\hat{\beta}_0$ 表示, 当各自变量均为 0 时, 因变量 y 平均值大约为 $\hat{\beta}_0$, 但当样本取值不包含 0 值时这样的解释并没有太大的实际意义。截距最好的解释是表示回归模型中被省略的变量对被解释变量 y 的影响。

(4) 模型的回归诊断: 得到回归模型参数的估计值后, 需要检验扰动项是否存在异方差, 变量是否存在完全多重共线性, 被解释变量与解释变量之间线性回归关系的显著性, 解释变量对被解释变量的影响的显著性和预报值的置信区间, 进而检验模型的优劣。

5.2.2 模型的求解

根据控制变量的原理, 利用 EXCEL 作出在各指标不同取值时实验随温度变化的乙醇转化率与 C₄ 烯烃选择性百分比的折线, 得到下图:

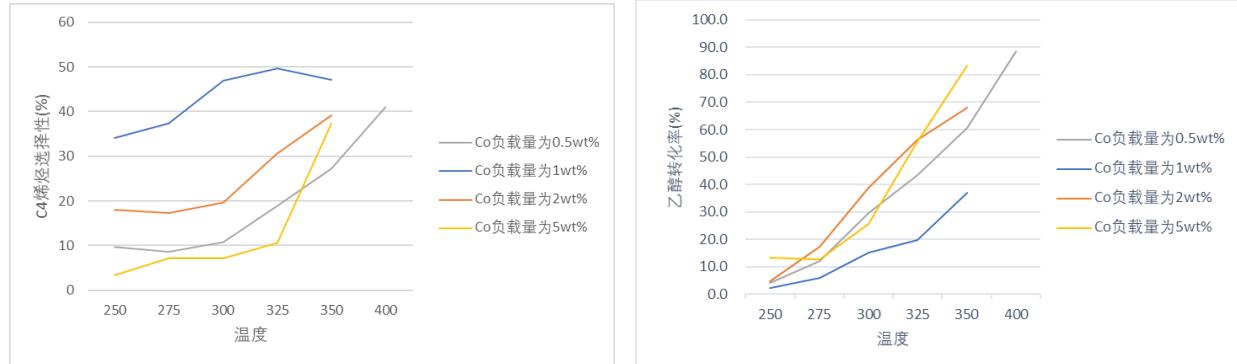


Figure 17: 仅改变 Co 负载量,C₄ 烯烃选择性和乙醇转化率关于 T 的折线

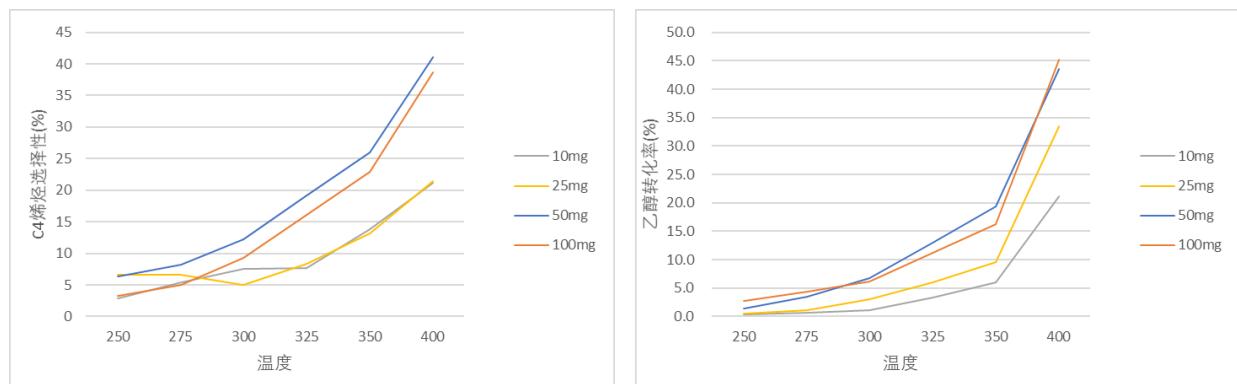


Figure 18: 仅改变催化剂质量,C₄ 烯烃选择性和乙醇转化率关于 T 的折线

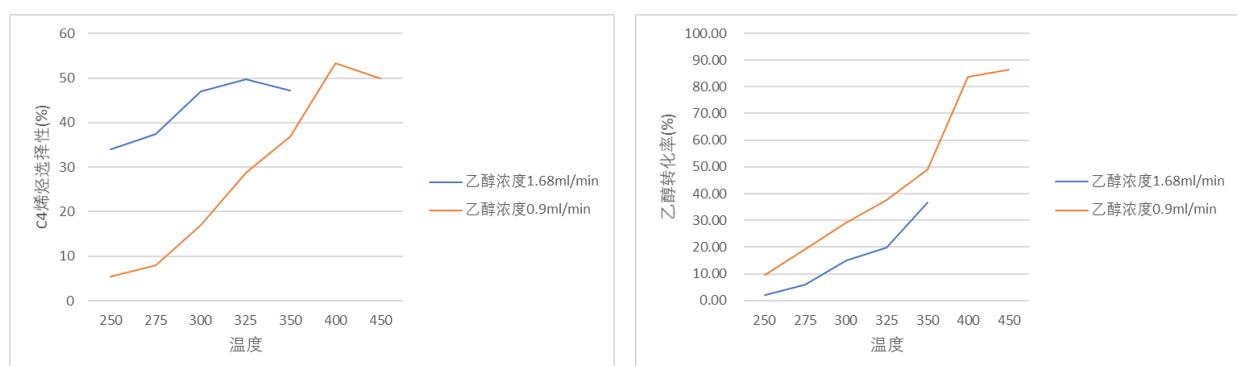


Figure 19: 仅改变乙醇浓度,C₄ 烯烃选择性和乙醇转化率关于 T 的折线

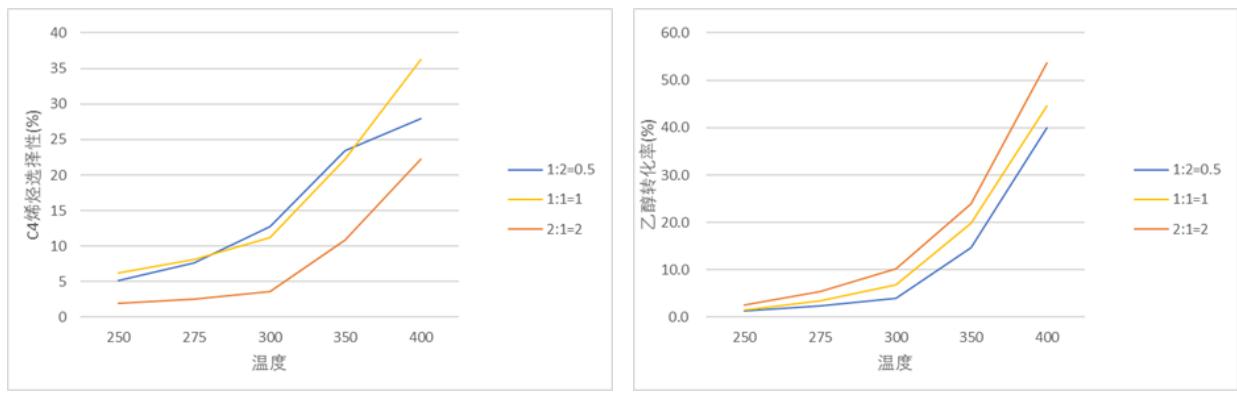


Figure 20: 仅改变装料比,C₄ 烯烃选择性和乙醇转化率关于 T 的折线

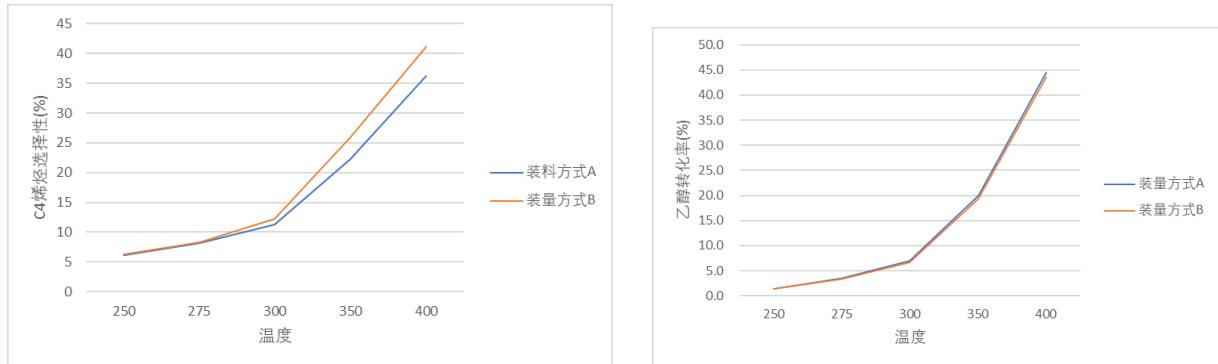


Figure 21: 仅改变装料方式,C₄ 烯烃选择性和乙醇转化率关于 T 的折线

观察五组图片可以得到，当温度小于 400°C 时，乙醇转化率与 C₄ 烯烃选择性均随着温度升高而显著增加，当温度大于 400°C 时（仅有一组实验温度），由于乙烯选择性的增加，C₄ 烯烃选择性出现了下降。

观察图 15 可以得到，通过纵向对比可知，温度 400°C, Co/SiO₂ 与 HAP 质量比为 200mg:200mg, 乙醇浓度 1.68ml/min 时，当 Co 负载量从 0.5wt% 增加到 5wt%，C₄ 烯烃选择性百分比呈现先增加后减少的趋势，而乙醇转化率的变化趋势与它不同，当 Co 负载量为 1wt% 时，乙醇转化率总体最小，而对于其他负载量，没有明显的关系。

观察图 16 可以得到，当催化剂装料比 1:1 时，经各温度时的纵向对比，Co/SiO₂ 装料量越多，乙醇转化率和 C₄ 烯烃选择性明显越高，可以猜测它们与 Co/SiO₂ 质量呈正向关系。

观察图 17 可以得到，在相同反应条件下，乙醇浓度较高时的乙醇转化率显著地低于较低时的结果，可以猜测其与乙醇浓度呈负向关系，但是 C₄ 烯烃选择性的影响不明显，甚至可以根据折线图走势猜测，若增加乙醇浓度 1.68ml/min 的实验样本，温度大于 350°C 时，乙醇浓度 1.68ml/min 的 C₄ 烯烃选择性转而低于乙醇浓度 0.9ml/min 的 C₄ 烯烃选择性。

观察图 18 可以得到，在确定 Co/SiO₂ 与 HAP 总质量为 100mg 时，在反应条件相同时，HAP 比 Co/SiO₂ 的值越高，乙醇转化率越高，而 C₄ 烯烃选择性相反，此时 1:1 装料比的百分比总体最高，2:1 装料比的选择性百分比远低于另外两种情况。

观察图 19 可以得到，在相同的反应条件下，两种装量方式具有相似的乙醇转化率和 C₄ 烯烃选择性，说明装料方式对催化剂的性能并没有显著影响。

有了上述定性分析的基础，下面考虑利用 STATA 软件进行多元线性回归，进一步分析各指标对目标对影响情况：

根据上述分析得到的 6 个指标对附件 1 中的数据进行整理，将催化剂组合中的变量提取出来，其中，装料方式为一时，设值为 1，装料方式为 2 时，设值为 0。将整合好的 EXCEL（见附录表：性能数据表）

导入 STATA 软件，利用 STATA 软件中的 *regress* 语句回归，回归目标分别为乙醇转化率 (y_1) 与 C_4 烯烃选择性 (y_2)，自变量为：装料方式 (x_1)、Co 负载量 (x_2)、 Co/SiO_2 质量 (x_3)、装料比 (x_4)、乙醇浓度 (x_5)、温度 (T)。在实验后，得到标准化的回归方程如下：

$$y_1 = 0.091236x_1 - 0.012322x_2 + 0.301581x_3 + 0.055253x_4 - 0.188642x_5 + 0.771521T$$

$$y_2 = 0.136174x_1 - 0.315466x_2 + 0.390606x_3 - 0.086818x_4 + 0.130576x_5 + 0.726611T$$

考察上述多元线性回归方程，发现两个函数的拟合优度 $R^2 \approx 0.8$ （具体回归评价指标见模型的分析与检验部分），在第一个回归方程中 Co/SiO_2 质量 (x_3)、乙醇浓度 (x_5)、温度 (T) 的系数在 95% 的置信水平下显著；在第二个回归方程中装料方式 (x_1)、Co 负载量 (x_2)、 Co/SiO_2 质量 (x_3)、乙醇浓度 (x_5)、温度 (T) 在 95% 的置信水平下显著（具体见模型的分析与检验部分），故而可以利用上述函数作为乙醇转化率与 C_4 烯烃选择性关于六项自变量的近似线性函数。根据折线图和函数关系式可以得出，当其他条件不变时，在一定范围内，斜率系数可以表明：

温度对变量影响明显：每升高 1 个单位，乙醇转化率平均增加 0.771521 个单位， C_4 烯烃选择性平均增加 0.726611 个单位；

同时，两者均随着催化剂质量的增加而增加： Co/SiO_2 质量每增加 1 个单位，乙醇转化率约增加 0.305181 个单位， C_4 烯烃选择性约增加 0.390606 个单位；

乙醇转化率随着乙醇浓度的增加而有下降趋势，将乙醇浓度提高 1 个单位，乙醇转化率平均减少 0.188642 个单位， C_4 烯烃选择性平均增加 0.130576 个单位；

两个方程的 x_1 的系数表示装料方式 I 和装料方式 II 时乙醇转化率与 C_4 烯烃选择性的平均差距，如果我们进行装料比、催化剂质量等变量均相同的实验，将催化剂以 I 和 II 的方式装料，则 I 装料实验的乙醇转化率平均比 II 装料实验高出 0.091236 个单位，I 装料实验的 C_4 烯烃选择性平均比 II 装料实验高出 0.136174 个单位；

Co 负载量的系数均为负，说明每增加 1 个单位，乙醇转化率平均减少 0.012322 个单位， C_4 烯烃选择性平均增加 0.315466 个单位，但结合折线图，发现仅负向关系较难解释样本中的现象，尚可确定其为起到正作用的较为复杂的变量；

在催化剂总量相同时，乙醇转化率随着装料比的增加而增加， C_4 烯烃选择性大体上与装料比的变化趋势相反：如果装料比提高一个单位，则乙醇转化率平均增加 0.055253 个单位， C_4 烯烃选择性平均减少 0.086818 个单位，但根据折线图综合考量，超过 350°C 时，1:1 的装料比导致的 C_4 烯烃选择性最大，催化剂具有较好的催化效果。

5.3 问题三的建模与求解

5.3.1 问题三的建模

与问题二的建模类似，仍然将催化剂组合及其装料方式转化为 Co 负载量、 Co/SiO_2 质量、 Co/SiO_2 和 HAP 的装料比、乙醇浓度和装料方式（虚拟变量）五个变量，加上温度共六个解释变量，作为 C_4 烯烃收率大小的影响因素，建立影响 C_4 烯烃收率大小的多元线性回归模型。在刻画 Co/SiO_2 和 HAP 的转料比与质量时需要两个变量，问题二中使用了 Co/SiO_2 质量与 Co/SiO_2 和 HAP 的装料比，满足关系式：HAP 质量 = Co/SiO_2 质量 × 装料比，在本题中可以考虑总质量 (HAP 质量 + Co/SiO_2 质量) 与装料比，替代上述两个量用于描述 Co/SiO_2 和 HAP 的装料比与质量。**同时， C_4 烯烃收率 = 乙醇转化率 × C_4 烯烃选择性。对于 350 度以下时，选择催化剂与温度使得 C_4 烯烃收率最大的问题，可以考虑去除 350 度以上数据后再进行回归分析，得到在 350 度以下较为显著的影响因子。**

进行回归后观察各变量的显著性与系数，结合实际考察各变量对 C_4 烯烃收率大小的影响。若显著性较差，可以考虑对各变量进行处理后再进行回归，最终得到较好的回归方程。

5.3.2 问题三的求解

根据控制变量的原理，利用 EXCEL 作出各指标不同取值时实验随温度变化的 C₄ 烯烃收率。

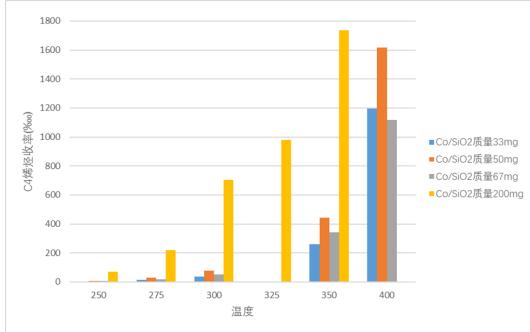


Figure 22: 仅改变 CoSiO₂ 质量 C₄ 烯烃收率与温度的关系

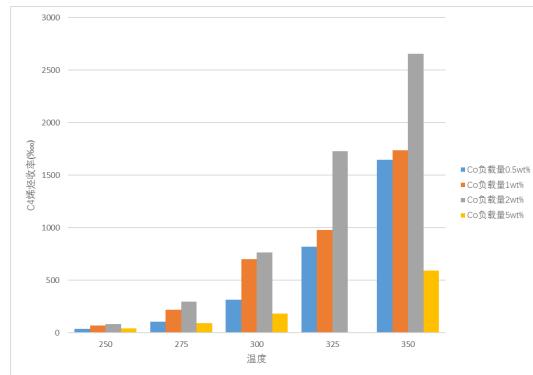


Figure 23: 仅改变 Co 负载量 C₄ 烯烃收率与温度的关系

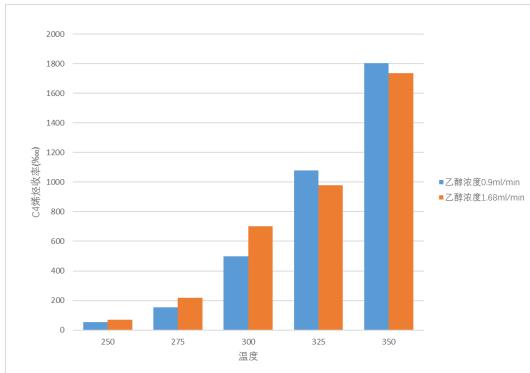


Figure 24: 仅改变乙醇浓度 C₄ 烯烃收率与温度的关系

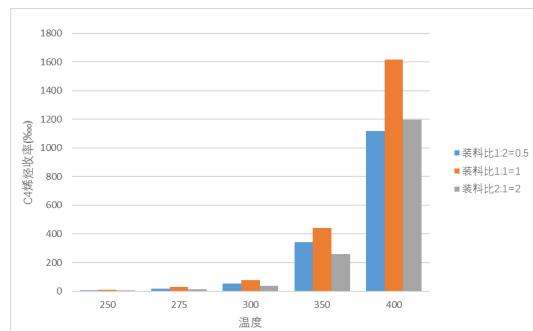


Figure 25: 仅改变装料比 C₄ 烯烃收率与温度的关系

观察六张图片，考察 C₄ 烯烃收率与各变量的关系，发现控制催化剂不变时，几乎所有实验中 C₄ 烯烃收率均随温度升高而有明显上升，这与问题二中我们研究所得的结论相符（乙醇转化率与烯烃选择性在同种催化剂下均随温度升高而增大）。

观察图 10，发现 CoSiO₂ 质量为 200mg 时，C₄ 烯烃收率最大。控制装料比相同 (1:1)，CoSiO₂ 质量分别为 200mg 与 50mg 时，C₄ 烯烃收率有显著差别，CoSiO₂ 质量越大（即总质量越大），C₄ 烯烃收率越高。

观察图 10，发现控制其他变量相同时，Co 负载量为 1wt% 时，C₄ 烯烃收率最大，并且 C₄ 烯烃收率随 Co 负载量与 1wt% 的差异增大而减小。

观察图 10，发现乙醇浓度在不同温度下大小关系并不一致，并且差异不大，猜测乙醇浓度对 C₄ 烯烃收率的影响不大。

观察图 10，发现控制总质量相同，仅改变装料比时，装料比为 1:1 且温度为 400 度时，C₄ 烯烃收率明显大于其它两类，猜测装料比为 1:1 时 C₄ 烯烃收率有最大值。

观察图 10，发现装料方式对于 C₄ 烯烃收率影响较小，装料方式 A 对应的 C₄ 烯烃收率略高于装料方式 B。

观察图 10，发现总质量对 C₄ 烯烃收率有较明显影响，即总质量越大 C₄ 烯烃收率越大，这与上述对 CoSiO₂ 质量的分析有关并且一致。

有了上述分析的基础，再考虑进行多元线性回归，获得各解释变量与 C₄ 烯烃收率较为准确的关系。考虑以 Co 负载量、CoSiO₂ 和 HAP 的装料比、乙醇浓度、装料方式（虚拟变量）和温度作为解释变量对 C₄ 烯烃收率进行回归。在 EXCEL 中对数据进行预处理后，在 STATA 中导入变量并计算 C₄ 烯烃收率，利用 STATA 中 regress 语句回归，回归目标为 C₄ 烯烃收率 (z)，自变量为：装料方式 (x_1)、Co 负载量 (x_2)、装料比 (x_3)、Co/SiO₂ 质量 (x_4)、乙醇浓度 (x_5)、温度 (T)，得到线性多元回归方程系数如下表：

Table 1: C₄ 烯烃收率关于装料方式等的回归结果

C ₄ 烯烃收率	系数	标准误差	t	P > t	[95% 置信区间]
装料方式	158.30180	126.51300	1.25	0.214	-92.6361 409.2398
Co 负载量	-87.77591	47.46761	-1.85	0.067	-181.9277 6.375873
CoSiO ₂ 质量	4.19203	0.86203	4.86	0.000	2.482194 5.901855
装料比	15.31978	215.91570	0.07	0.944	-412.948 443.5876
乙醇浓度	-113.51390	106.91380	-1.06	0.291	-325.577 98.54916
温度	13.46720	0.99419	13.55	0.000	11.49523 15.43918
常数项	-3878.95000	441.06640	-8.79	0.000	-4753.803 -3004.097

观察该表可知，将各变量作为解释变量直接进行回归的效果并不好：装料方式、Co 负载量、装料比、乙醇浓度在 95% 的置信水平下均不显著，进行怀特检验后发现存在异方差，考虑使用最小二乘法与稳健的标准误：

Table 2: C₄ 烯烃收率关于装料方式等使用稳健的标准误的回归结果

C ₄	系数	稳健的标准误	t	P > t	[95% 置信区间]
装料方式	158.30180	98.61435	1.61	0.112	-37.29927 353.9029
Co 负载量	-87.77591	56.25768	-1.56	0.122	-199.3627 23.81093
CoSiO ₂ 质量	4.19203	0.90661	4.62	0.000	2.393769 5.99028
装料比	15.31978	169.36770	0.09	0.928	-320.6202 351.2598
乙醇浓度	-113.51390	105.29180	-1.08	0.284	-322.3598 95.33192
温度	13.46720	1.35176	9.96	0.000	10.78599 16.14842
常数项	-3878.95000	497.91370	-7.79	0.000	-4866.56 -2891.341

发现显著性出现了明显降低，下面考虑对各变量进行预处理后再进行回归。

根据上述六张图片的分析，并且装料方式显然已不显著，故考虑将 Co 负载量、装料比、乙醇浓度分别转换为三者与 2wt%、1、0.9ml/min 的距离，可以通过差的绝对值、差的平方、差值开根等进行刻画，再结合温度与总质量进行回归。

在进行多次试验后得到如下两次较好的回归结果（系数为标准化后的系数）：

$$z = -0.039|x_2 - 2| + 0.304x_3(x_4 + 1) - 0.023|x_4 - 1| - 0.126|x_5 - 0.9| + 0.738T$$

$$z = -0.039|x_2 - 2| + 0.305x_3(x_4 + 1) - 0.021|x_4 - 1|^2 - 0.126|x_5 - 0.9| + 0.738T$$

回归方程中温度 (T)、总质量 ($x_3(x_4 + 1)$) 以及乙醇浓度与 0.9 的距离 ($|x_5 - 0.9|$) 的系数在 95% 的置信水平下显著异于 0，故而下面探讨这几种变量对于 C₄ 烯烃收率的影响。根据柱状图和函数关系式可以得出，当其他条件不变时，在一定范围内，斜率系数可以表明：

温度对变量影响明显：每升高 1 个单位，C₄ 烯烃收率平均增加 0.738 个单位；

同时，C₄ 烯烃收率随催化剂质量的增加而增加：总质量每增加 1 个单位，C₄ 烯烃收率约增加 0.304 个单位；

C₄ 烯烃收率随着乙醇浓度与 0.9 距离的增加而有下降趋势，将乙醇浓度与 0.9 的距离提高 1 个单位，C₄ 烯烃收率平均减少 0.126 个单位；

Co 负载量与 2 的距离的系数为负，说明 Co 负载量与 2 的距离每增加 1 个单位，乙醇转化率平均减少 0.039 个单位；

查阅参考文献可得，装料比为 1:1 时，C₄ 烯烃收率有最优值，观察柱状图亦可得到，装料方式 B 的 C₄ 烯烃收率略高于装料方式 A。

故而考虑选取装料方式 A、Co 负载量为 2wt% 时、装料比为 1:1、总质量为 400mg (Co/SiO₂ 质量为 200mg)、乙醇浓度为 0.9ml/min、温度为 400 度时，C₄ 烯烃收率有最大值。

接着筛选出温度不高于 350 °C 的数据，共 90 条，进行多元线性回归分析。在之前结果的基础上，我们猜测温度在 350 °C 时大部分影响因子的最优值变化不大。因此在经过尝试后，我们选择装料方式 (x_1)、Co 负载量与 1 的距离 (x_{21})、装料比与 1 的距离的 $\frac{1}{4}$ 次方 (x_{31})、Co/SiO₂ 的质量 (x_4)、乙醇浓度 (x_5)、温度 (T) 作为自变量，C₄ 烯烃收率作为因变量，进行标准化多元线性回归。

得到回归结果后，进行怀特检验。从怀特检验的结果中发现存在异方差的问题（详见模型三的分析与检验）。因此我们选择使用 OLS + 稳健的标准误来进行回归，避免异方差的问题。结果如下：

Table 3: 350 度以下 C₄ 烯烃收率关于装料方式等使用稳健的标准误的回归结果

C ₄	系数	稳健的标准误	t	P > t	标准化后的系数
装料方式	216.196	63.92737	3.38	0.001	0.2170724
CoSiO ₂	2.675124	0.6541741	4.09	0.000	0.3969398
温度	7.555571	1.152191	6.56	0.000	0.5701567
Co 负载量与 1 距离	-91.30024	27.26143	-3.35	0.001	-0.2171236
装料比与 1 的四分之一距离	-208.6582	94.77646	-2.20	0.030	-0.1150267
乙醇浓度	57.26344	72.15989	0.79	0.430	0.0621213
常数项	-2376.192	422.2409	-5.63	0.000	

从回归结果中发现，除了乙醇浓度之外，其余自变量在 95% 的置信区间下均显著。观察回归系数发现，在固定其余条件不变的情况下，装料方式从 0 改变到 1 将会使 C₄ 烯烃收率增加 216.2 个单位；Co/SiO₂ 的质量每增加一个单位，C₄ 烯烃收率将增加 2.7 个单位；温度每增加一个单位，C₄ 烯烃收率将增加 7.6 个单位。同时，Co 负载量与 1 的距离、装料比与 1 的距离的 $\frac{1}{4}$ 次方的回归系数都为负值，说明 Co 负载量和装料比越接近 1，C₄ 烯烃收率就会越大。

因此，当温度低于 350 °C 时，欲使得 C₄ 烯烃收率最大，我们应当使用 A 类装料方式以及实验中最大的 Co/SiO₂ 质量 (200 mg) 和温度 (350 °C)，同时选择装料比为 1，Co 负载量为 1 wt%。最终得到

两种条件下 C_4 烯烃收率最大时催化剂组合与温度的选择:

Table 4: C_4 烯烃收率最大时催化剂组合与温度的选择

	装料方式	Co 负载量	Co/SiO ₂	装料比	乙醇浓度	温度
温度不限	B	2wt%	200mg	1	0.9ml/min	400
350 度以下	A	1wt%	200mg	1	0.9ml/min	350

5.4 问题四的求解

在问题四中，我们将按照“最优解附近局部搜索”的思路进行实验参数设计。首先，由于实验次数的限制，我们能进行的参数调整非常有限。参考上一个问题中没有温度限制时的回归结果，装料方式 (x_1)、Co 负载量 (x_2)、装料比 (x_3) 这三个自变量对 C_4 烯烃收率的影响不显著，而装料方式只有 A 与 B 两种，通过简单的控制变量比对发现，A 类装料方式在大多数情况下能获得更好的 C_4 烯烃收率，因此新增的 5 次实验将集中在对于 Co 负载量和装料比的局部搜索上。

为了在新增的实验中获得更好的 C_4 烯烃收率，我们固定装料方式、Co/SiO₂ 的质量、乙醇浓度和温度分别为上一个问题中得出的最优值，然后改变 Co 负载量和装料比进行局部搜索。《乙醇偶合制备丁醇及 C_4 烯烃》中指出，Co 负载量的最优值为 2 wt%，装料比的最优值为 1，我们将这一组值作为局部搜索的起始值。至此，已经将实验参数的设计转化为在二维平面上（以 Co 负载量为 x 轴，以装料比为 y 轴），以 (2, 1) 为初值的局部搜索决策。

自然地，第一次实验应当以起始值参数，这组参数在已有的实验中并没有出现：

Table 5: 第一次实验催化剂组合及温度选择

实验次数	装料方式	Co/SiO ₂ 的质量	乙醇浓度	温度	Co 负载量	装料比
1	A	200 mg	0.9 mg/min	400	2 wt%	1

剩下的 4 次实验，我们拟给出在两种不同实验条件下的方案。第一种实验条件是：在实验前设计好全部的 5 组参数，再进行实验。在这种条件下，我们给出的 5 组实验参数如下：

Table 6: 第二至四次实验催化剂组合及温度选择

实验次数	装料方式	Co/SiO ₂ 的质量	乙醇浓度	温度	Co 负载量	装料比
2	A	200 mg	0.9 ml/min	400	2 wt%	0.75
3	A	200 mg	0.9 ml/min	400	2 wt%	1.25
4	A	200 mg	0.9 ml/min	400	1.5 wt%	1
5	A	200 mg	0.9 ml/min	400	0.5 wt%	1

以上参数的给出，同时参考了回归结果、控制变量的对比结果以及《乙醇偶合制备丁醇及 C_4 烯烃》，使得局部搜索的范围在起始值的较小邻域内。

第二种实验条件是：在每一次实验结束后，可以得知实验结果，并以此决定下一次实验的参数。在这种条件下，我们设计了一套规则，可以依据并利用上一组实验的结果来选择下一组实验的参数，并且可以保证这样的局部搜索不会获得更坏的结果。这套参数选择规则的叙述如下：

默认所有的装料方式、Co/SiO₂ 的质量、乙醇浓度和温度均取上一个问题中的最优值。记 g_1 表示前一次实验的 C_4 烯烃收率， g_2 表示当前实验的 C_4 烯烃收率， $g(x, y)$ 表示当 Co 负载量为 x wt%，装料比

为 y 时, 实验得到的 C_4 烯烃收率, $\alpha = 0.5$ 为初始步长。开始时, $g_1 = g(1, 1)$ (此为已有实验数据中 C_4 烯烃收率的最大值), $g_2 = g(2, 1)$ 。每一次实验后, 更新 $\alpha \leftarrow \frac{1}{2}\alpha$, 并比对结果: 若此次实验相较于上一次实验, C_4 烯烃收率增大或保持不变, 那么以 α 的步长改变参数中在上一轮中保持不变的变量, 分别以 0.5 的概率决定正负; 而若此次实验相较于上一次实验, C_4 烯烃收率减小, 那么将参数回退至上一次实验的参数, 并重复上面的更新操作。参数选择规则的流程图展示如下:

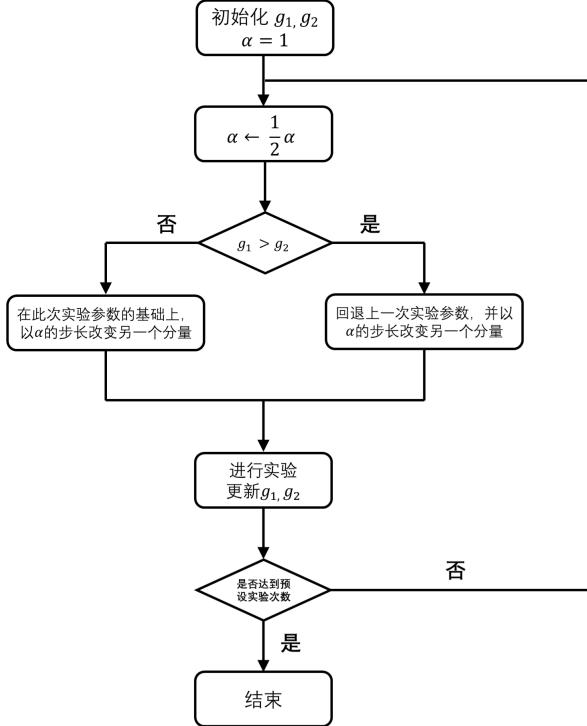


Figure 26: 实验变量的选择流程

例如, 开始时从 $(1, 1)$ 改变到了 $(2, 1)$, 改变了参数的 x 分量。如果 $g(2, 1) \geq g(1, 1)$, 则下一次的实验参数为:

$$\begin{cases} (2, 1 - \frac{1}{2}\alpha), & p = 0.5 \\ (2, 1 + \frac{1}{2}\alpha), & p = 0.5 \end{cases}$$

反之, 则下一次实验的参数为:

$$\begin{cases} (1, 1 - \frac{1}{2}\alpha), & p = 0.5 \\ (1, 1 + \frac{1}{2}\alpha), & p = 0.5 \end{cases}$$

由规则的设置不难看出, 每当获得了比前一次差的实验结果, 实验参数就会回退至上一次, 然后向正交的方向继续探索, 保证了只会获得更优的实验参数; 同时每次更新后, 步长都会相应地减小, 保证了局部搜索的收敛性。

6 模型的分析与检验

拟合优度 R^2 , 也称可决系数, 通常用于评价回归结果或者拟合函数对参数成线性时的拟合结果。 R^2 的计算公式如下:

$$\begin{aligned} SST &= \sum_{i=1}^n (y_i - \bar{y})^2 \\ SSE &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ SSR &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \\ R^2 &= \frac{SSR}{SST} = 1 - \frac{SSE}{SST}. \end{aligned}$$

其中, SST 称为总体平方和, SSE 称为误差平方和, SSR 称为回归平方和。 R^2 越接近 1, 说明误差平方和越接近 0, 说明拟合效果越好。

均方根误差 (RMSE) 的计算公式如下:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} SSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

6.1 问题一模型的分析与检验

在第一个问题的求解中, 先固定催化剂组合, 分别拟合乙醇转化率和 C₄ 烯烃的选择性与温度的函数, 再固定催化剂组合及温度, 拟合了各项产物的比例与反应时间的函数。在函数拟合中, 通常选择 R^2 、校正的 R^2 (Adjusted R^2) 和均方根误差 (RMSE) 作为拟合优劣的评价指标。

乙醇转化率和 C₄ 烯烃的选择性与温度的函数的拟合指标如下:

(表格篇幅过长, 此处仅列出指标的平均值, 由于 Fourier 函数和指数函数关于参数不是线性的, 所以 R^2 和校正的 R^2 仅计算了多项式函数的平均值。每个拟合函数及其拟合优度见附录)

Table 7: 两类拟合函数的评价指标平均值

	R^2	校正的 R^2	均方根误差
温度-乙醇转化率	0.991300882	0.988534154	1.726823919
温度-C ₄ 烯烃选择性	0.992736609	0.986497081	0.811586509

各项产物的比例与反应时间的函数的拟合指标如下:

Table 8: 产物选择性关于时间 t 的拟合函数与评价指标

	乙醇	乙烯	乙醛	脂肪醇	甲基苯甲醛 (醇)
拟合函数	二次多项式	Fourier 函数	线性函数	二次多项式	三次多项式
SSE	1.835	0.008172	0.5255	6.081	0.1753
R^2	0.9878	0.9674	0.9588	0.9151	0.9468
Adjusted R^2	0.9817	0.9348	0.9505	0.8727	0.8936
RMSE	0.6773	0.05219	0.3242	1.233	0.2417

由结果不难发现, 所有线性拟合函数的校正的 R^2 的平均值都在 0.9 以上, 说明拟合的效果较好。同时, 均方根误差的平均值也在可控范围内, 说明很好地抑制了过拟合的现象。

非线性
函数的
均方根
误差接
近0

6.2 问题二模型的分析与检验

在第二个问题的求解中，利用 STATA 对乙醇转化率与 C₄ 烯烃选择性和分析出的六个变量进行多元线性回归。下面基于回归得到的各项指标，对回归结果进行分析与检验。

由结果可知，在对乙醇转化率的回归模型中，拟合优度 $R^2=0.8021$ ，校正的拟合优度 Adjusted $R^2=0.7905$ ；在对 C₄ 烯烃选择性的回归模型中，拟合优度 $R^2=0.7486$ ，校正的拟合优度 Adjusted $R^2=0.7338$ 。调整后的 R^2 越大，拟合效果越好。本题中的回归属于解释型的回归，更多关注的是模型整体和回归参数的显著性如何。

接下来考虑回归方程的显著性，研究弃真概论 F 值。乙醇转化率的回归模型中的 $F=68.91$ ，C₄ 烯烃选择性的回归模型中的 $F=50.63$ ，均大于 F 检验临界值，因此回归方程显著，解释变量和被解释变量间存在明显函数关系。

下表中给出了两个回归模型中对 6 个自变量的偏回归系数的 T 检验结果。从表中可知，在对乙醇转化率的解释中，概率 p 都小于显著性水平 0.05 的变量是 Co/SiO₂ 质量、乙醇浓度和温度；在对 C₄ 烯烃选择性的解释中，显著性水平小于 0.05 的变量是 Co 负载量、Co/SiO₂ 质量和温度。因此认为这些相关系数 β 显著地不为 0，与因变量线性相关水平显著。

Table 9: 乙醇转化率关于装料方式等的回归结果

乙醇转化率	系数	标准误差	t	$P > t $	[95% 的置信区间]	标准化后的系数
装料方式	4.336637	2.471729	1.75	0.082	-0.56603 9.23930	0.091236
Co 负载量	-0.239550	0.927391	-0.26	0.797	-2.07903 1.59993	-0.012322
Co/SiO ₂ 质量	0.097801	0.016842	5.81	0.000	0.06440 0.13121	0.301581
装料比	5.170628	4.218421	1.23	0.223	-3.19659 13.53785	0.055253
乙醇浓度	-8.227098	2.088813	-3.94	0.000	-12.37025 -4.08395	-0.188642
温度	0.339908	0.019424	17.50	0.000	0.30138 0.37844	0.771521
常数项	-89.987520	8.617269	-10.440	0.000	-107.07980 -72.89520	

Table 10: C₄ 烯烃选择性关于装料方式等的回归结果

C ₄ 烯烃选择性	系数	标准误差	t	$P > t $	[95% 的置信区间]	标准化后的系数
装料方式	3.799291	1.635252	2.32	0.022	0.5557775 7.0428050	0.136174
Co 负载量	-3.599927	0.613546	-5.87	0.000	-4.8168920 -2.3829630	-0.315466
Co/SiO ₂ 质量	0.074353	0.011142	6.67	0.000	0.0522526 0.0964536	0.390606
装料比	-4.768830	2.790832	-1.71	0.091	-10.3044300 0.7667717	-0.086818
乙醇浓度	3.342660	1.381921	2.42	0.017	0.6016259 6.0836940	0.130576
温度	0.187904	0.012851	14.62	0.000	0.1624146 0.2133925	0.726611
常数项	-46.906050	5.701031	-8.230	0.000	-93.8121000 -35.5980800	

在问题二的模型建立中，我们假定了扰动项 ξ 同方差和无自相关两条性质，而本文中的横截面数据很容易出现异方差的现象，因此我们利用怀特检验 (White Test) 检验是否存在异方差，防止方差较大的扰动项对我们模型稳定性的破坏。原假设 $H_0=$ 不存在异方差，在回归结束后利用 STATA 命令：`estat imest, white`。乙醇转化率的回归模型得到怀特检验的统计量 $chi2(21)=49.09$ ，且在 C₄ 烯烃选择性的回归模型得到怀特检验的统计量 $chi2(21)=53.60$ ，说明存在异方差。下面考虑使用最小二乘法与稳健的标准误对模型进行回归，得到下表，显著性变量不变：

Table 11: 使用稳健的标准误后乙醇转化率关于装料方式等的回归结果

乙醇转化率	系数	稳健的标准误	t	$P > t $	[95% 置信区间]	标准化后的系数	
装料方式	4.336637	2.079950	2.08	0.040	0.21107	8.46221	0.091236
Co 负载量	-0.239550	1.039114	-0.23	0.818	-2.30063	1.82153	-0.012322
CoSiO ₂ 质量	0.097801	0.019215	5.09	0.000	0.05969	0.13591	0.301581
装料比	5.170628	2.660127	1.94	0.055	-0.10572	10.44698	0.055253
乙醇浓度	-8.227098	2.645042	-3.11	0.002	-13.47353	-2.98067	-0.188642
温度	0.339908	0.020754	16.38	0.000	0.29874	0.38107	0.771521
常数项	-89.987520	9.260551	-9.720	0.000	-108.35580	-71.61926	

Table 12: 使用稳健的标准误后 C₄ 烯烃选择性关于装料方式等的回归结果

C ₄	系数	稳健的标准误	t	$P > t $	[95% 置信区间]	标准化后的系数	
装料方式	3.799291	1.212365	3.13	0.002	1.3945700	6.2040120	0.136174
Co 负载量	-3.599927	0.733714	-4.91	0.000	-5.0552460	-2.1446090	-0.315466
CoSiO ₂ 质量	0.074353	0.013918	5.34	0.000	0.0467467	0.1019595	0.390606
装料比	-4.768830	1.982468	-2.41	0.018	-8.7010460	-0.8366141	-0.086818
乙醇浓度	3.342660	1.422809	2.35	0.021	0.5205246	6.1647960	0.130576
温度	0.187904	0.012479	15.06	0.000	0.1631509	0.2126562	0.726611
常数项	-46.906050	4.714759	-9.950	0.000	-56.2577500	-37.5543500	

为了防止自变量之间存在多重共线性而导致的 $(X^T X)^{-1}$ 的不存在和参数的难以识别，虽然本次实验中选取的解释变量看上去独立，不存在任何关系，但为了保证检验的完整性，下面我们通过方差膨胀因子 VIF 来检验：在回归结束以后，利用 STATA 软件各自变脸的 VIF 命令：`estat vif`，由于两次模型中自变量仙童，所以得到同一组 VIF 数据，见下表。从表中可见，各变量的 VIF 值均远小于 10，说明构建模型的变量之间相关性很小。

Table 13: 装料方式等变量间的回归的多重共线性检验

Variable	VIF	1/VIF
装料方式	1.39	0.71742
Co/SiO ₂ 质量	1.39	0.719289
乙醇浓度	1.18	0.845701
Co 负载量	1.17	0.852538
装料比	1.05	0.954699
温度	1.00	0.998048
Mean VIF	1.20	

6.3 问题三模型的分析与检验

在第三个问题的求解中，利用 STATA 对 C₄ 烯烃收率和分析出的六个变量进行多元线性回归。下面基于回归得到的各项指标，对回归结果进行分析与检验。

首先对温度不设限制的模型进行分析与检验：

在变量选择同问题二的回归初尝试中，拟合优度 $R^2=0.6947$ ，校正的拟合优度 Adjusted $R^2=0.6768$ ，拟合优度不高。接下来看模型整体和回归参数的显著性：弃真概论 $F=38.69$ ，大于 F 检验的临界值，因此此回归方程显著，自变量与因变量间函数关系明显；下表中给出了回归模型中对 6 个自变量的偏回归系数的 T 检验结果，其中，装料方式、Co 负载量、装料比和乙醇浓度在 95% 的置信区间下均不显著，因此本次回归效果并不理想。

Table 14: C₄ 烯烃收率关于装料方式等的回归结果

C ₄ 烯烃收率	系数	标准误差	t	P > t	[95% 置信区间]
装料方式	158.30180	126.51300	1.25	0.214	-92.6361 409.2398
Co 负载量	-87.77591	47.46761	-1.85	0.067	-181.9277 6.375873
CoSiO ₂ 质量	4.19203	0.86203	4.86	0.000	2.482194 5.901855
装料比	15.31978	215.91570	0.07	0.944	-412.948 443.5876
乙醇浓度	-113.51390	106.91380	-1.06	0.291	-325.577 98.54916
温度	13.46720	0.99419	13.55	0.000	11.49523 15.43918
常数项	-3878.95000	441.06640	-8.79	0.000	-4753.803 -3004.097

根据结果和文献资料，我们剔除装料方式这一不显著的变量，其余变量中，我们作出适当改进并且引入了二次项：Co 负载量转化为 Co 负载量与 2 的距离、Co/SiO₂ 质量转化为总质量、装料比分别转化为装料比与 1 的距离和距离的平方、乙醇浓度转化为乙醇浓度为 0.9 的距离。改进模型后的拟合优度 R^2 分别为 0.6934 和 0.6933，F 分别为 46.58 和 46.56，均大于 F 检验的临界值，说明该回归拟合优度不高但回归方程显著；下面分析回归参数的显著性，结果如下表，发现：不显著的解释变量减少，剩下 Co 负载量与 2 的距离，分别地装料比与 1 距离和距离的平方不显著。

Table 15: C₄ 烯烃收率关于装料比与 1 的距离等变量的回归结果

C ₄	系数	标准误差	t	P > t	标准化后的系数
Co 负载量	-53.86023	77.68063	-0.69	0.490	-0.03909
总质量	2.039664	0.38868	5.25	0.000	0.30444
装料比与 1 的距离	-93.50157	224.14630	-0.42	0.677	-0.02334
乙醇浓度与 0.9 的距离	-340.3238	158.26870	-2.15	0.034	-0.12609
温度	13.40681	0.99288	13.50	0.000	0.73845
常数项	-3723.666	361.88430	-10.29	0.000	

Table 16: C₄ 烯烃收率关于装料比与 1 的距离的平方等变量的回归结果

C ₄	系数	标准误差	t	P > t	标准化后的系数
Co 负载量	-53.42792	77.65723	-0.69	0.493	-0.038779
总质量	2.046101	0.38715	5.29	0.000	0.305399
装料比与 1 的距离的平方	-87.20724	232.57570	-0.37	0.708	-0.020875
乙醇浓度与 0.9 的距离	-340.72	158.27950	-2.15	0.034	-0.126240
温度	13.4069	0.99304	13.50	0.000	0.738453
常数项	-3726.499	361.61100	-10.31	0.000	

下面类似问题二，我们利用检验 (White Test) 检验是否存在异方差，在回归结束后利用 STATA 命令：

`estat imest, white`, 得到怀特检验统计量 $chi2(17)$ 均为 53.23, 否定了原假设, 说明存在异方差。所以我们使用 OLS+ 稳健的标准误来解决异方差问题, 利用 STATA 命令: `regress y x1,...,xk, robust`, 得到下表, 显著性参量不变。

Table 17: 使用稳健的标准误后 C₄ 烯烃收率关于装料比与 1 的距离等变量的回归结果

C ₄	系数	稳健的标准误	t	P > t	[95% 置信区间]
Co 负载量	-53.42792	92.40445	-0.58	0.561	-237.1227 129.4022
总质量	2.046101	.4240847	4.81	0.000	1.198592 2.880736
装料比与 1 的距离	-87.20724	165.5944	-0.56	0.574	-421.919 234.9158
乙醇浓度与 0.9 的距离	-340.72	168.9822	-2.01	0.047	-675.4602 -5.187478
温度	13.4069	1.308772	10.24	0.000	10.81117 16.00245
常数项	-3726.499	438.3132	-8.50	0.000	-4592.957 -2854.375

Table 18: 使用稳健的标准误后 C₄ 烯烃收率关于装料比与 1 的距离的平方等变量的回归结果

C ₄	系数	稳健的标准误	t	P > t	[95% 置信区间]
Co 负载量	-53.42792	92.37342	-0.58	0.564	-236.6288 129.773
总质量	2.046101	.4225174	4.84	0.000	1.208137 2.884064
装料比与 1 的距离的平方	-87.20724	171.3969	-0.51	0.612	-427.1326 252.7181
乙醇浓度与 0.9 的距离	-340.72	168.9499	-2.02	0.046	-675.7923 -5.647712
温度	13.4069	1.309184	10.24	0.000	10.81044 16.00336
常数项	-3726.499	438.2984	-8.50	0.000	-4595.76 -2857.237

接下来筛选温度低于 350°C 的数据后, 对模型进行分析检验:

经过尝试, 我们在自变量中引入了四次根项等非线性项: 装料方式、Co 负载量与 2 的距离、装料比与 1 的距离的 $\frac{1}{4}$ 次方、Co/SiO₂ 质量、乙醇浓度和温度。对这些变量进行标准化多元线性回归后, 首先利用怀特检验 (White Test) 检验是否存在异方差, 得到怀特检验统计量 $chi2(21) = 38$, 否定原假设, 存在异方差现象, 所以使用 OLS+ 稳健的标准误解决异方差问题, 得到下表的回归参数。此时, 在 95% 的置信区间下, 仅乙醇浓度这一自变量不显著。

综上, 为了优化参数不显著的问题, 通过多次回归实验, 我们运用到了非线性回归, 在回归方程中分别增添二次项和四次根项, 即使拟合优度没有太大改进, 但更好地保证了回归参数的显著性和回归方程的显著性。

7 模型的评价、改进与推广

模型的优点: 问题一中的模型对于拟合函数的选择范围较广, 使得拟合函数能反映数据不同的增长速率和波动趋势。同时, 线性拟合函数的修正的 R^2 、非线性拟合函数的均方根误差都较低, 说明拟合的效果较好。

问题二中的模型运用多元线性回归分别寻找乙醇转化率、C₄ 烯烃选择性与催化剂组合和温度之间的线性关系。通过线性回归可以发现乙醇转化率及 C₄ 烯烃选择性与催化剂的选择与温度存在显著关系, 通过此方法可以科学精准地判断催化剂组合与温度对乙醇转化率及 C₄ 烯烃选择性的影响。同时, 对回归结果也进行了异方差检验, 并利用较为合适的方法对异方差进行了处理。

问题三中的模型同样运用多元线性回归探究 C₄ 烯烃收率与催化剂组合与温度之间的关系, 先通过定

性分析再进行定量回归，再进行回归之后针对回归结果对数据进行预处理，以获得最佳的回归结果，最终得到较好的线性关系。对回归结果同样进行了异方差检验，并对异方差进行了处理。

模型四中的模型考虑实验结果中 C₄ 烯烃收率最高为最佳，不仅基于已有实验数据以及第三问的结果给出了五个较优的结果，还设计了针对可获得实验结果的情况下的优化算法。

模型的缺点：问题一中部分拟合函数较为复杂，虽然能较好地反应函数关系，但是由于大量非线性函数的存在，无法均使用拟合优度 R^2 对拟合结果进行评价。

问题二中进行多元线性回归时未考虑对数据进行预处理，虽然回归结果较好，但如若进行预处理，或许可以将现不显著的变量转为显著变量。

问题三中的多元线性回归未能探究出 Co 负载量以及装料比对 C₄ 烯烃转化率的影响。

模型的改进：问题二中可以尝试通过对数据进行预处理、设置计算列、交叉项等方法，获得更好的回归结果。

模型的推广：第四问中建立了可以广泛使用的实验方案设计算法，通过对实验数据的评估，获得下一次实验的方案，以获得更高的实验效率，在降低实验成本、节约实验实验的同时，也更可能获得较好的实验结果。

参考文献

- [1] Damodar N.Gujarati,Dawn C.Porter. 经济计量学精要(第4版) [M]. 北京:机械工业出版社.2010:29-31.
- [2] 吕绍沛. 乙醇偶合制备丁醇及 C₄ 烯烃 [D]. 大连理工大学,2018:1,49.
- [3] Stock, J.S. and Watson, M.W. (2011) Introduction to Econometrics. 3rd Edition, Addison-Wesley, Boston.

附录

附录 1

Table 19: 问题一中各催化剂组合中 C₄ 烯烃选择性与乙醇转化率关于温度 T 的拟合函数的评价指标

	R^2	修正的 R^2	均方根误差
A1	0.932173462	0.909564616	4.102248141
	0.990938461	0.963753845	1.309517342
A2	0.990006333	0.98667511	3.040247226
	0.991428419	0.965713675	1.774008678
A3	0.964275046	0.957130055	6.236827456
	0.997757166	0.995514332	1.29086263
A4	0.995030385	0.993787982	2.482272776
	0.998945197	0.997362992	0.656216297
A5	0.994026643	0.990044405	2.351737722
	0.990546623	0.984244372	1.606029512
A6	0.967449309	0.956599079	6.395717173
	0.999930017	0.999720067	0.230651252
A7	0.998786271	0.998381695	0.914970499
	0.999970517	0.999882067	0.126211811
A8	0.999038755	0.998077511	0.917566718
	0.999716392	0.998865569	0.495451017
A9	0.999557711	0.999410281	0.395468684
	0.994805866	0.993074488	1.274977124
A10	0.99988664	0.999546559	0.25480516
	0.995727062	0.982908247	0.472141378
A11	0.999441175	0.9992549	0.376299401
	0.978216292	0.970955056	0.538406546
A12	0.994406448	0.99254193	1.543758255
	0.994610044	0.992813392	1.063658858
A13	0.998612545	0.99815006	0.701110248
	0.999828075	0.999312301	0.260038114
A14	0.997023722	0.996031629	1.321759214
	0.999503116	0.998012464	0.385140733
B1	0.994907607	0.993210143	1.441179352
	0.999976737	0.99990695	0.14122526
B2	0.999288492	0.999051322	0.544177591
	0.999986876	0.999947503	0.108141139
B3	0.998898878	0.998623597	0.295569112
	0.942778012	0.928472515	1.787709828
B4	0.998900838	0.998626047	0.462475033
	0.985796548	0.964491369	1.163513373
B5	0.999828838	0.999786048	0.235145727
	0.998844014	0.997110036	0.434109519

	R^2	修正的 R^2	均方根误差
B6	0.99675588	0.99594485	1.404008857
	0.999338613	0.998346533	0.425496999
B7	0.999023535	0.998779418	0.84595795
	0.988824738	0.986030923	1.499809274

附录 2

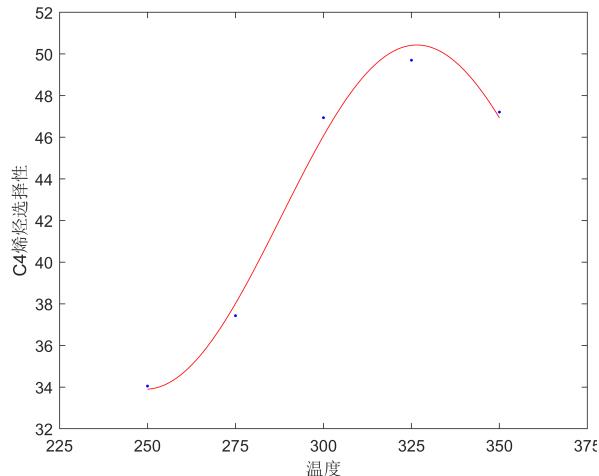


Figure 27: A_1 催化剂中 C₄ 烯烃选择性关于温度的拟合函数

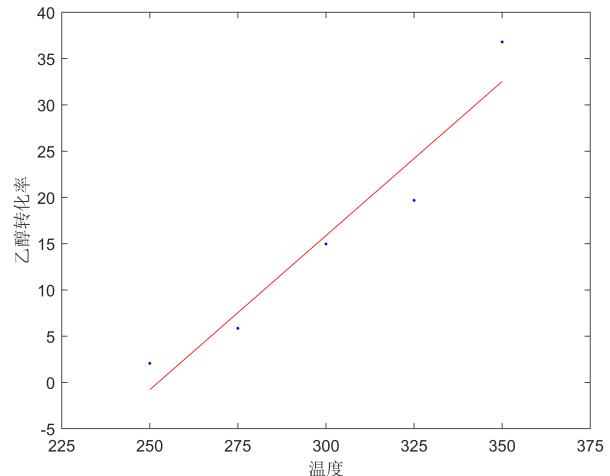


Figure 28: A_1 催化剂中乙醇转化率关于温度的拟合函数

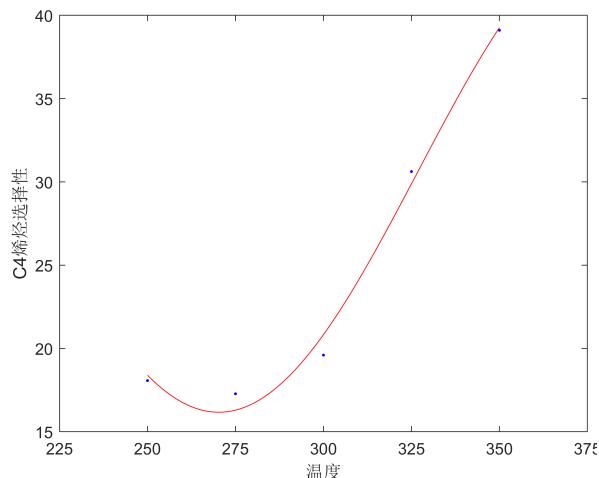


Figure 29: A_2 催化剂中 C₄ 烯烃选择性关于温度的拟合函数

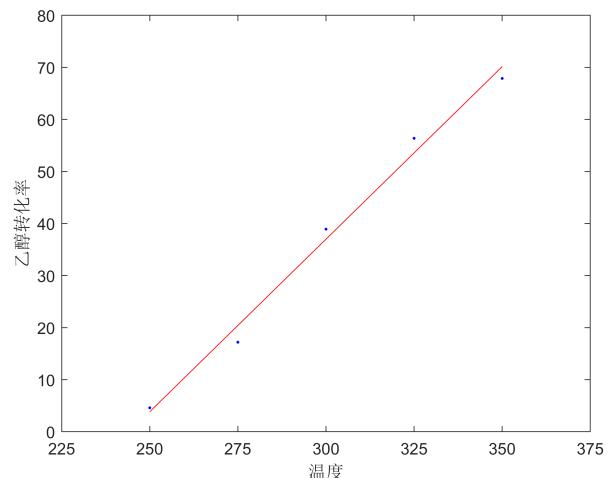


Figure 30: A_2 催化剂中乙醇转化率关于温度的拟合函数

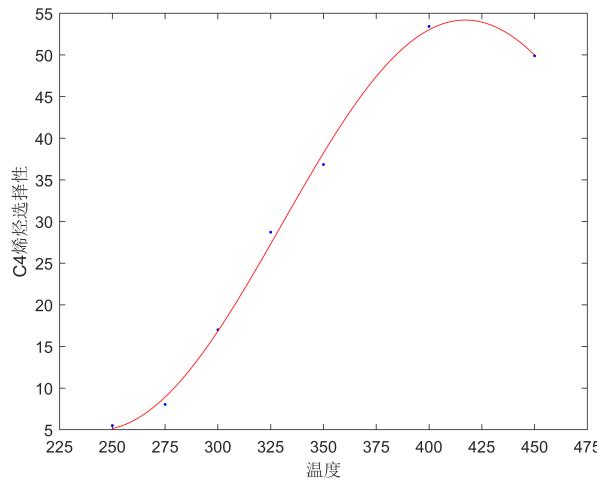


Figure 31: A_3 催化剂中 C_4 烯烃选择性关于温度的拟合函数

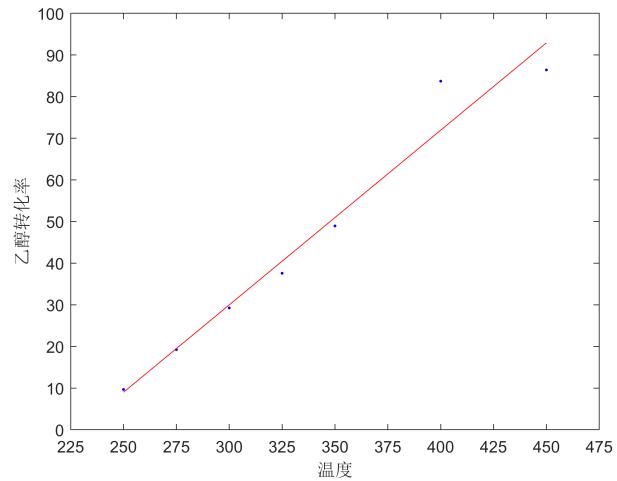


Figure 32: A_3 催化剂中乙醇转化率关于温度的拟合函数

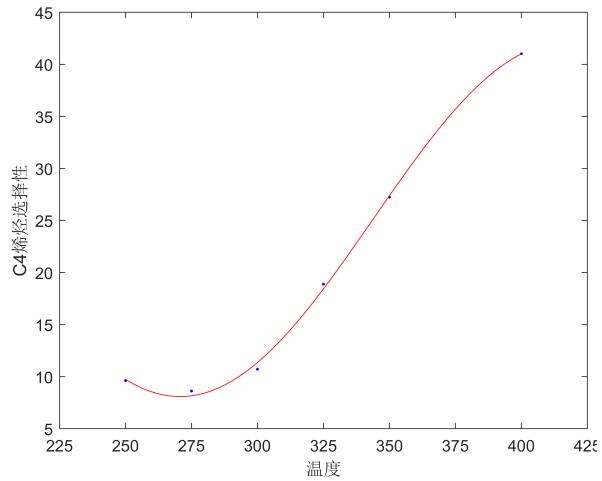


Figure 33: A_4 催化剂中 C_4 烯烃选择性关于温度的拟合函数

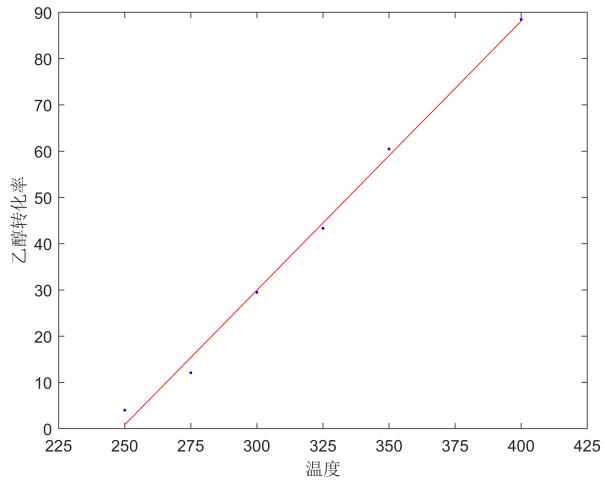


Figure 34: A_4 催化剂中乙醇转化率关于温度的拟合函数

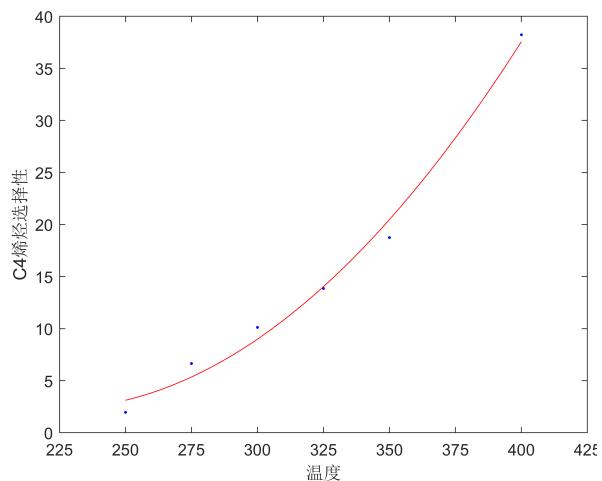


Figure 35: A_5 催化剂中 C_4 烯烃选择性关于温度的拟合函数

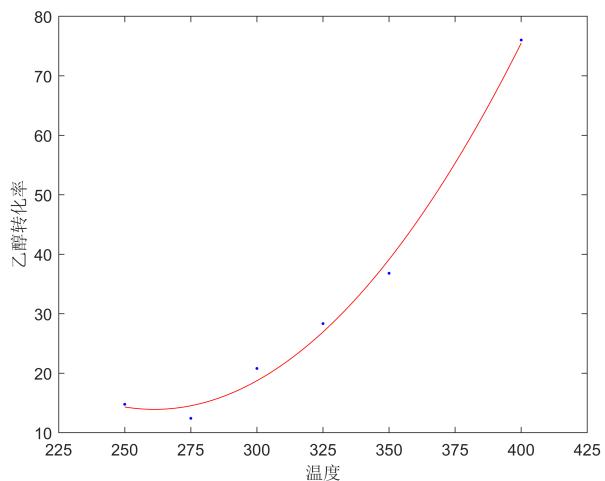


Figure 36: A_5 催化剂中乙醇转化率关于温度的拟合函数

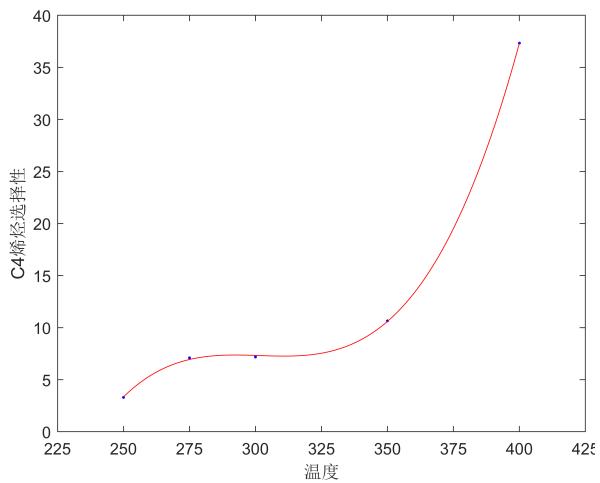


Figure 37: A_6 催化剂中 C₄ 烯烃选择性关于温度的拟合函数

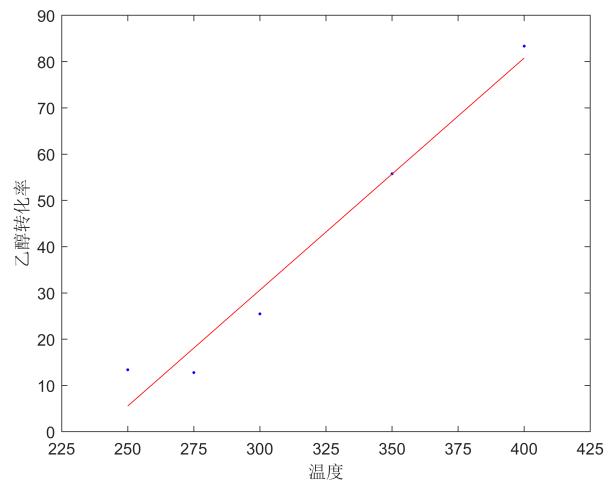


Figure 38: A_6 催化剂中乙醇转化率关于温度的拟合函数

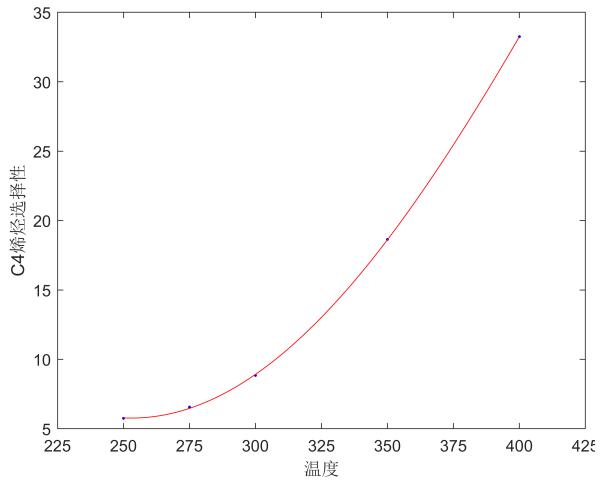


Figure 39: A_7 催化剂中 C₄ 烯烃选择性关于温度的拟合函数

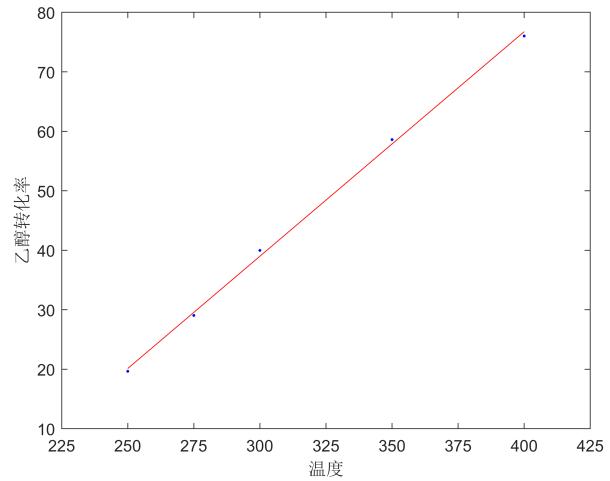


Figure 40: A_7 催化剂中乙醇转化率关于温度的拟合函数

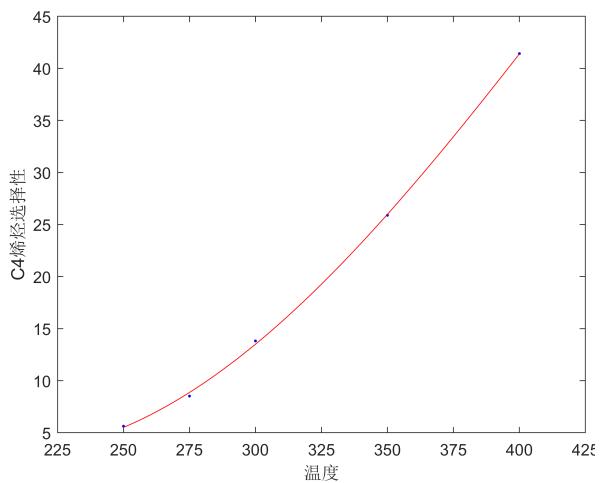


Figure 41: A_8 催化剂中 C₄ 烯烃选择性关于温度的拟合函数

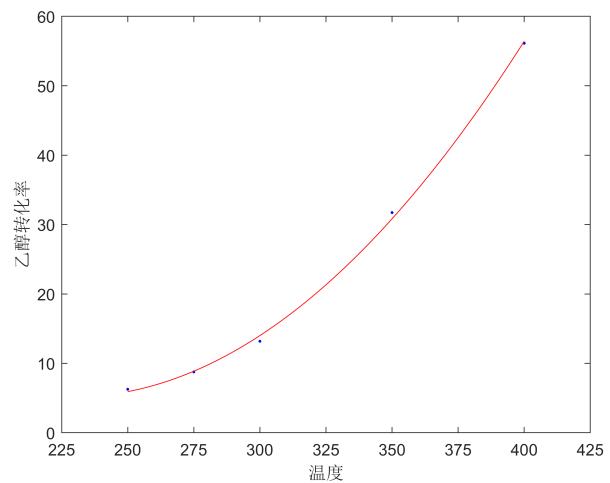


Figure 42: A_8 催化剂中乙醇转化率关于温度的拟合函数

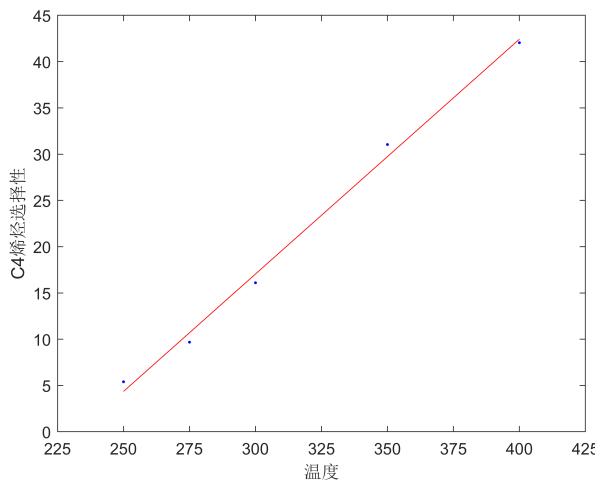


Figure 43: A_9 催化剂中 C_4 烯烃选择性关于温度的拟合函数

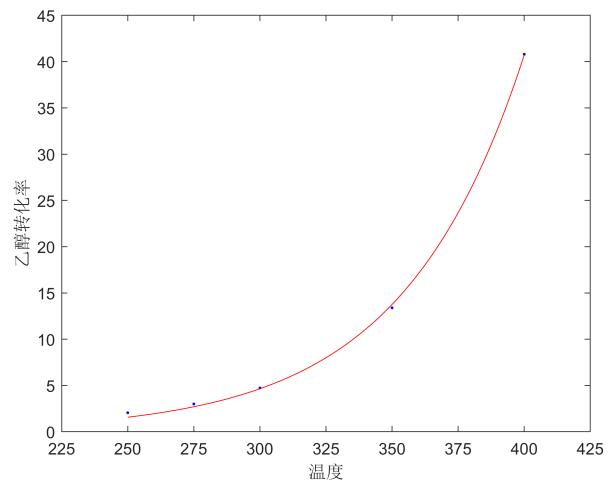


Figure 44: A_9 催化剂中乙醇转化率关于温度的拟合函数

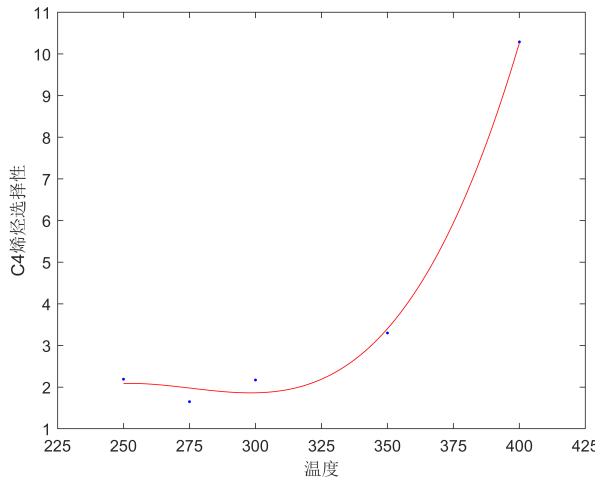


Figure 45: A_{10} 催化剂中 C_4 烯烃选择性关于温度的拟合函数

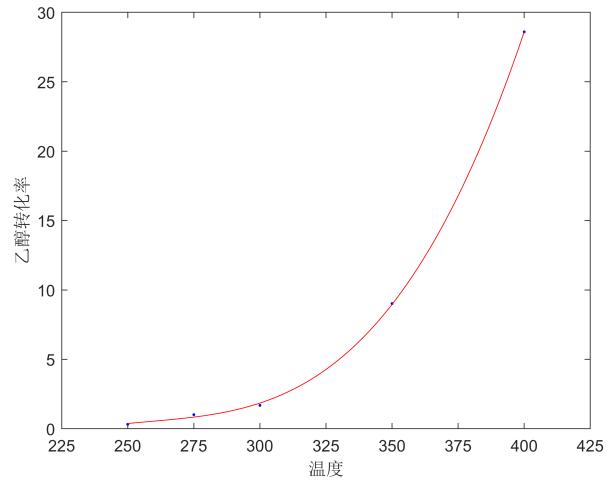


Figure 46: A_{10} 催化剂中乙醇转化率关于温度的拟合函数

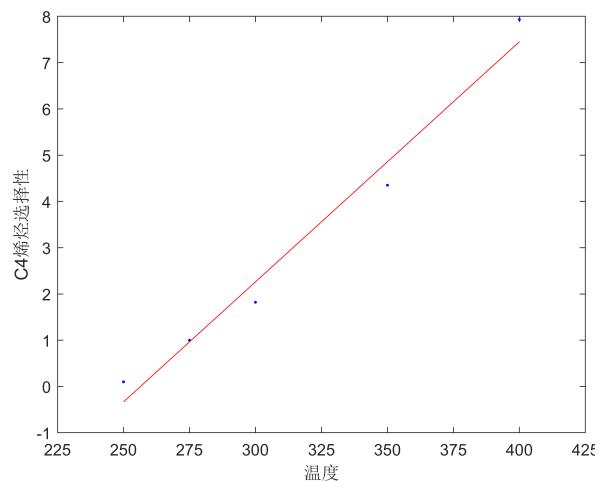


Figure 47: A_{11} 催化剂中 C_4 烯烃选择性关于温度的拟合函数

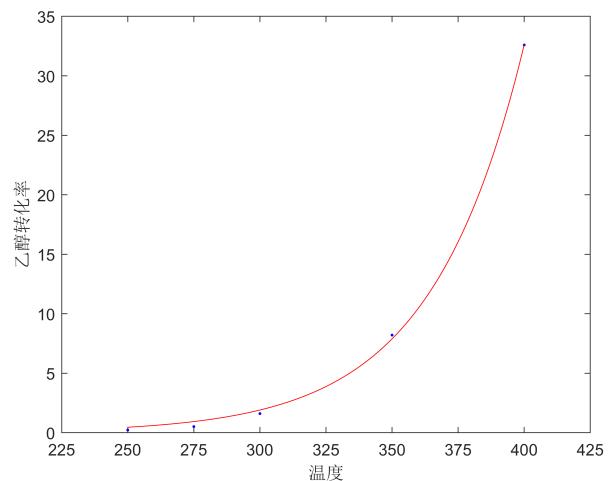


Figure 48: A_{11} 催化剂中乙醇转化率关于温度的拟合函数

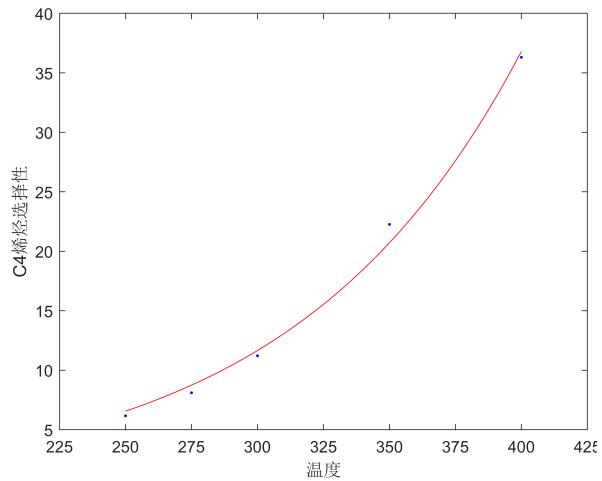


Figure 49: A_{12} 催化剂中 C_4 烯烃选择性关于温度的拟合函数

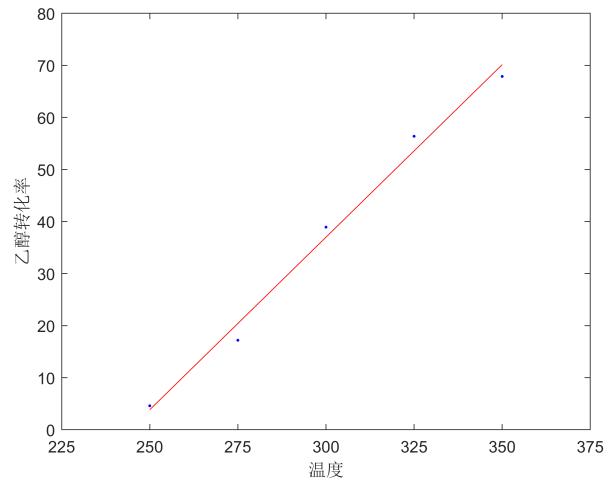


Figure 50: A_{12} 催化剂中乙醇转化率关于温度的拟合函数

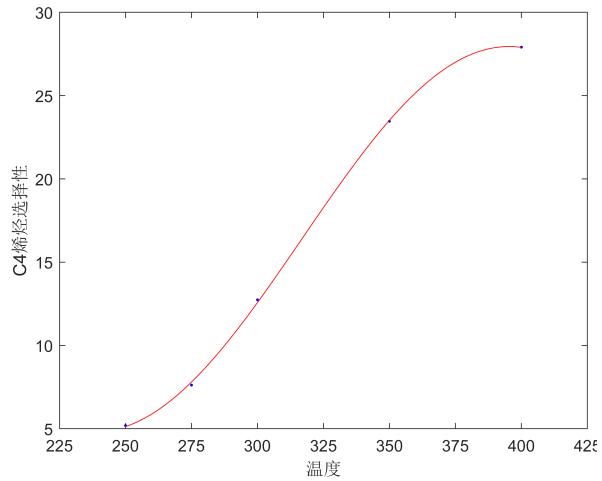


Figure 51: A_{13} 催化剂中 C_4 烯烃选择性关于温度的拟合函数

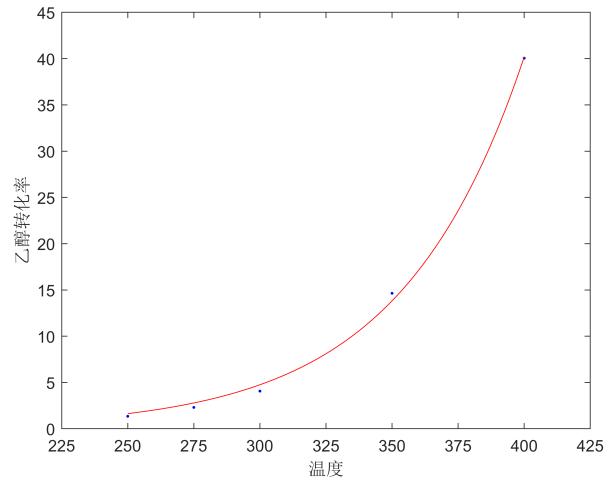


Figure 52: A_{13} 催化剂中乙醇转化率关于温度的拟合函数

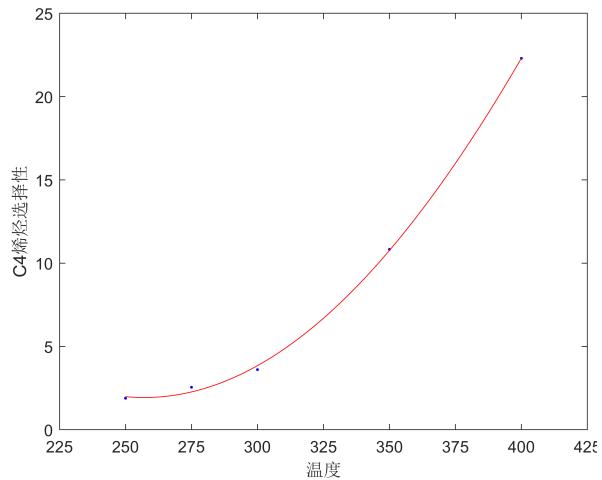


Figure 53: A_{14} 催化剂中 C_4 烯烃选择性关于温度的拟合函数

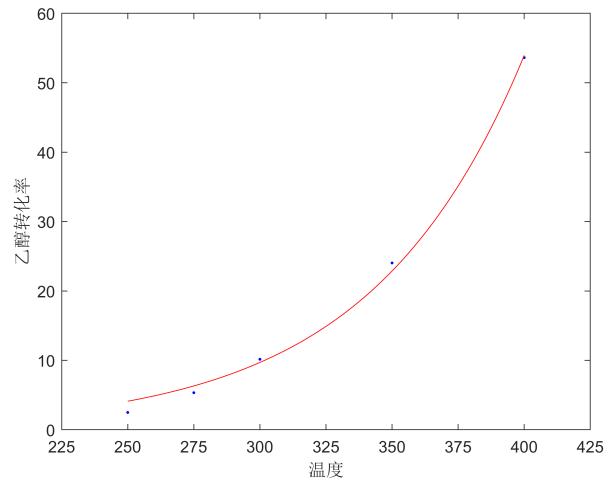


Figure 54: A_{14} 催化剂中乙醇转化率关于温度的拟合函数

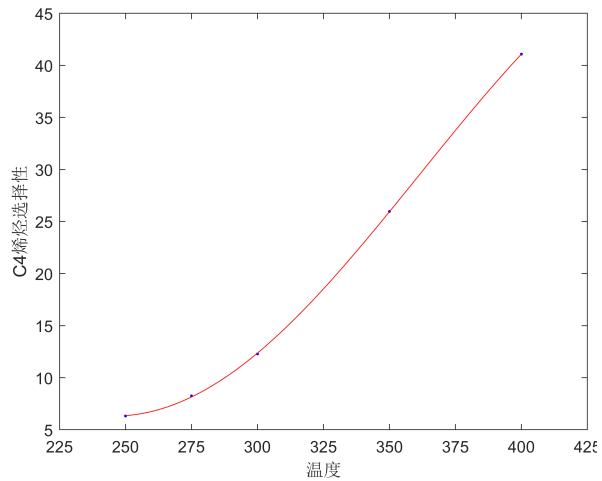


Figure 55: B_1 催化剂中 C_4 烯烃选择性关于温度的拟合函数

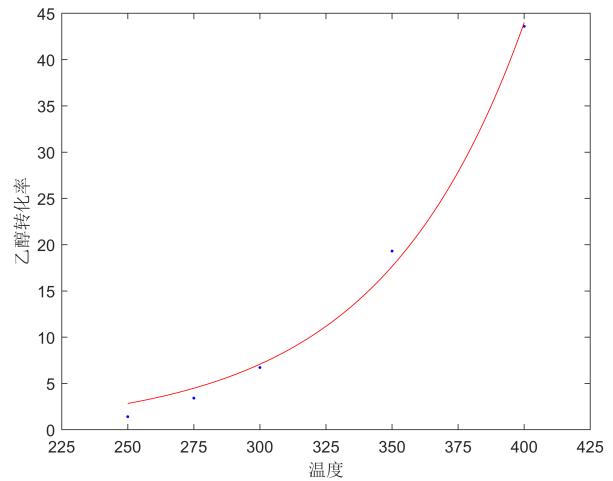


Figure 56: B_1 催化剂中乙醇转化率关于温度的拟合函数

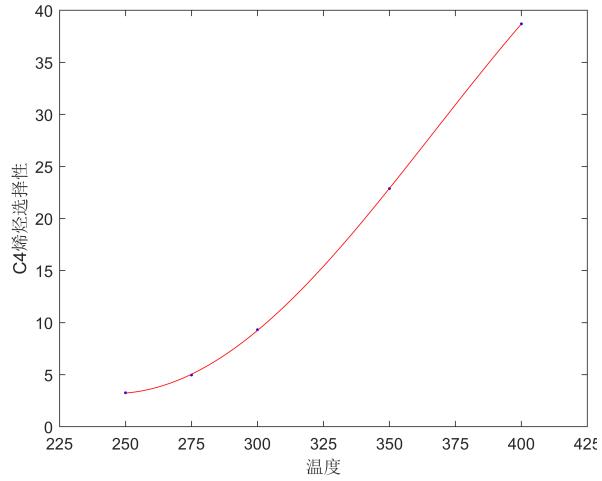


Figure 57: B_2 催化剂中 C_4 烯烃选择性关于温度的拟合函数

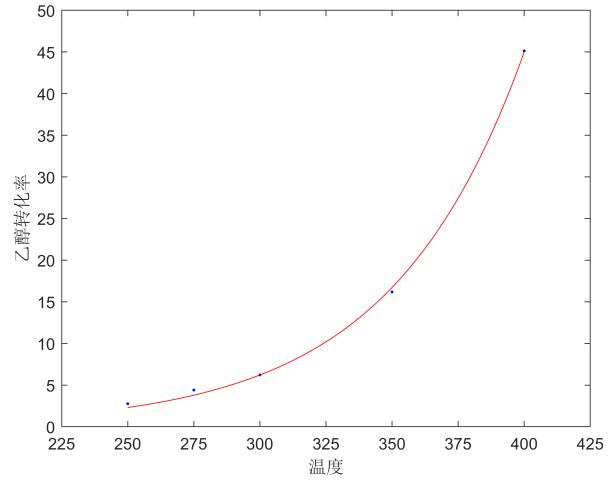


Figure 58: B_2 催化剂中乙醇转化率关于温度的拟合函数

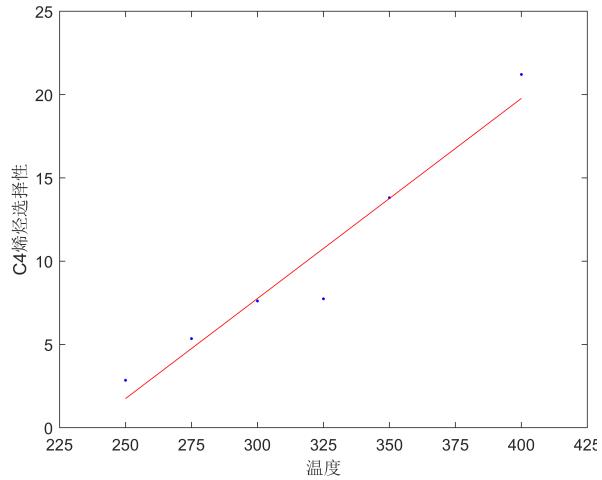


Figure 59: B_3 催化剂中 C_4 烯烃选择性关于温度的拟合函数

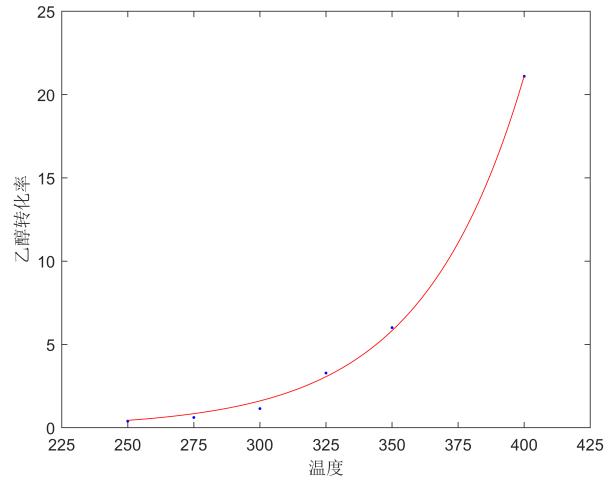


Figure 60: B_3 催化剂中乙醇转化率关于温度的拟合函数

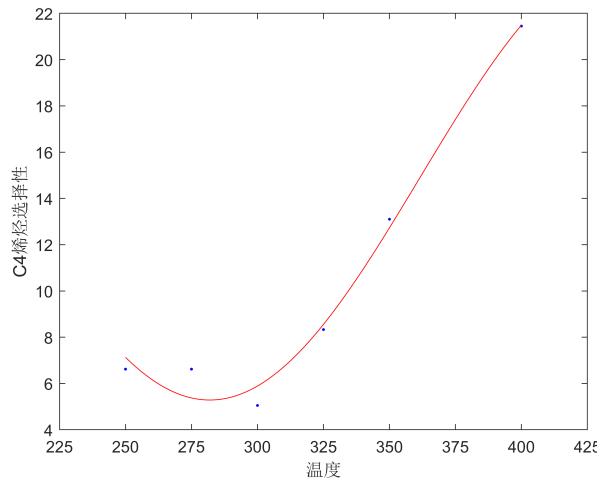


Figure 61: B_4 催化剂中 C_4 烯烃选择性关于温度的拟合函数

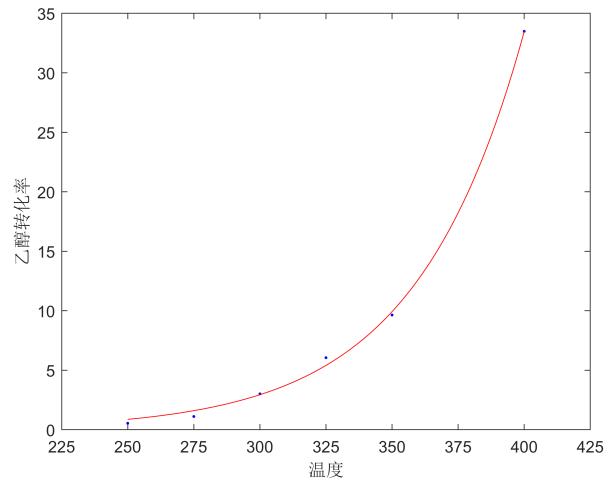


Figure 62: B_4 催化剂中乙醇转化率关于温度的拟合函数

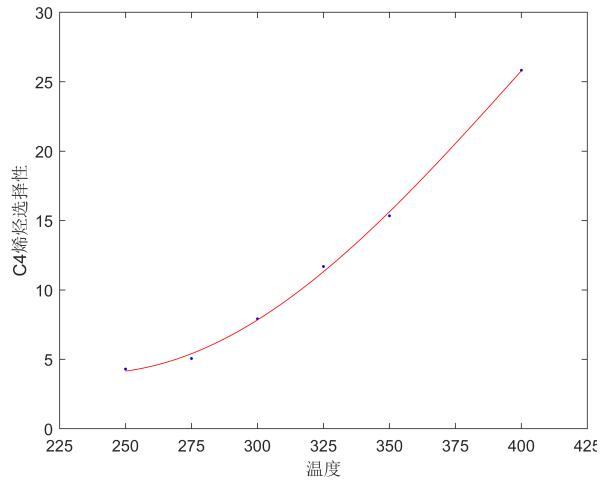


Figure 63: B_5 催化剂中 C_4 烯烃选择性关于温度的拟合函数

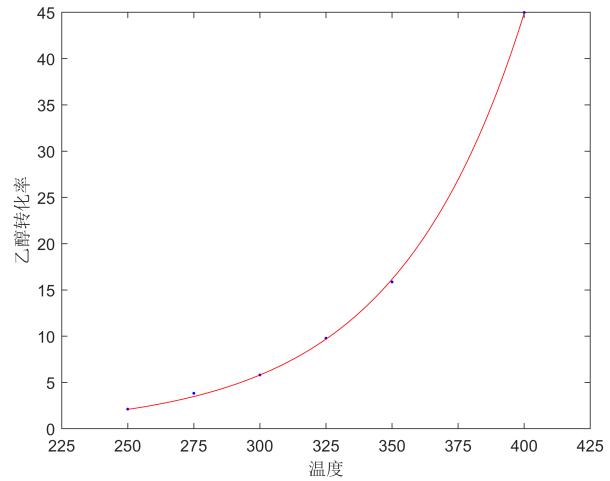


Figure 64: B_5 催化剂中乙醇转化率关于温度的拟合函数

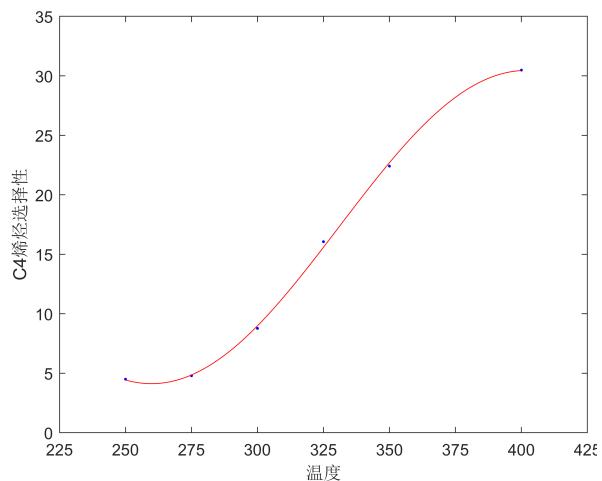


Figure 65: B_6 催化剂中 C_4 烯烃选择性关于温度的拟合函数

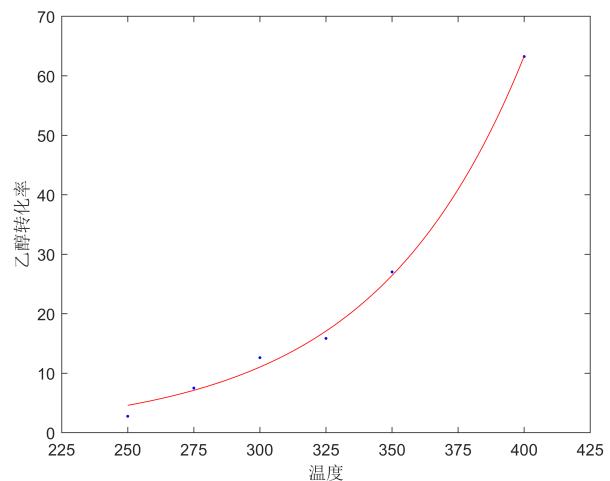


Figure 66: B_6 催化剂中乙醇转化率关于温度的拟合函数

