响应面分析之 python 方法（一）最速上升法

响应曲面法（Response Surface Methodology,RSM）是数学方法和统计方法结合的产物，用于对感兴趣的响应受多个变量影响的问题进行建模和分析，以优化这个响应。例如，设一位化学工程师想求出温度（x_1）和压强（x_2）的水平以使得过程的产率（y）达到最大值。产率是温度水平和压强水平的函数，比方说

$$y = f(x_1 + x_2) + \varepsilon$$

其中ε表示响应 y 的观测误差或噪音。如果记期望响应为$E(y) = f(x_1, x_2) = \eta$,则由

$$\eta = f(x_1, x_2)$$

表示的曲面称为响应面。

例 11.1 一位化学工程师要确定使过程产率最高的操作条件。影响产率的两个可控变量是反应时间和反应温度。工程师当前使用的操作条件是反应时间为 35 分钟，温度为 155℉，产率约为 40%，因为此区域不大可能包含最优值，于是她拟合一阶模型并应用最速上升法。

这位工程师认为，所拟合的一阶模型的探测区域应该是反应时间为（30，40）分钟，反应温度为（150，160）℉，为简化计算，将自变量规范在（-1，1）区间内。于是，如果记$\boldsymbol{\xi_1}$为自然变量时间，$\boldsymbol{\xi_2}$为自然变量温度，则规范变量是

$$\boldsymbol{x_1 = \frac{\xi_1 - 35}{5}} \quad \text{和} \quad \boldsymbol{x_2 = \frac{\xi_2 - 155}{5}}$$

实验设计列在表 11.1 中，用来收集这些数据的设计是附加 5 个中心点的$\boldsymbol{2^2}$析因设计，在中心点处的重复试验用于估计实验误差，并可以用于检测一阶模型的合适性，而且，过程的当前运行条件也就在设计的中心点处。

表 11.1 拟合一阶模型的过程数据

自然变量		规范变量		响应
$\boldsymbol{\xi_1}$	$\boldsymbol{\xi_2}$	$\boldsymbol{x_1}$	$\boldsymbol{x_2}$	
30	150	−1	−1	39.3
30	160	−1	1	40.0
40	150	1	−1	40.9
40	160	1	1	41.5
35	155	0	0	40.3
35	155	0	0	40.5
35	155	0	0	40.7
35	155	0	0	40.2
35	155	0	0	40.6

使用最小二乘法，以一阶模型来拟合这些数据，用二水平设计的方法，可以求得规范变量表示的下列模型：

$$\boldsymbol{\hat{y} = 40.44 + 0.775x_1 + 0.325x_2}$$

在沿着最速上升路径探测之前，应研究一阶模型的适合性，有中心点的设计使实验者能够

（1） 求出误差的一个估计量。

（2） 检测模型中的效互作用（效叉乘积项）。

（3） 检测二次效应（弯曲性）。

中心点处的重复试验观测值可用于计算误差的估计量：

$$\hat{\sigma}^2=\frac{(40.3)^2+(40.5)^2+(40.7)^2+(40.2)^2+(40.6)^2-(202.3)^2/5}{4}$$

$$=0.0430$$

一阶模型假定变量 x_1 和 x_2 对响应有可加效应，变量间的交互作用可用已添加到模型中的交叉乘积项 x_1x_2 的系数β_{12}来表示。此系数的最小二乘估计恰好是按普通 2^2 析因设计算得的交互效应的 1/2，或

$$\hat{\beta}_{12}=\frac{1}{4}[(1\times 39.3)+(1\times 41.5)+(-1\times 40.0)+(-1\times 40.9)]=\frac{1}{4}\times(-0.1)=-0.025$$

单自由度的交互作用的平方和是

$$SS_{交互作用}=\frac{(-0.1)^2}{4}=0.0025$$

比较$SS_{交互作用}$和$\hat{\sigma}^2$，得到下列拟合不足统计量：

$$\mathrm{F}=\frac{SS_{交互作用}}{\hat{\sigma}^2}=\frac{0.0025}{0.0430}=0.058$$

它很小，表示交互作用可以忽略。

对直线模型适合性的另一检测方法，是应用 6.6 节中所介绍的对纯二次弯曲效应的检测。回忆一下它的构成，也就是，比较设计的析因部分 4 个点处的平均响应，即$\bar{y}_F=40.425$，以及在设计的中心点处的平均响应$\bar{y}_C=40.46$。如果在真实响应函数中存在二次弯曲性，则$\bar{y}_F-\bar{y}_C$是这种弯曲性的度量。如果β_{11}与β_{22}是纯二次项${x_1}^2$ 与${x_2}^2$的系数，则$\bar{y}_F-\bar{y}_C$是β_{11}+β_{22}的一个估计。在我们的例子中，纯二次项的一个估计是

$$\hat{\beta}_{11}+\hat{\beta}_{22}=\bar{y}_F-\bar{y}_C=40.425-40.46=-0.035$$

与零假设$H_0{:}\,\beta_{11}+\beta_{22}=0$有关的单自由度的平方和是

$$SS_{纯二次}=\frac{n_P n_C(\bar{y}_F-\bar{y}_C)^2}{n_F+n_C}=\frac{(4)(5)(-0.035)^2}{4+5}=0.0027$$

其中n_F与n_C分别是析因部分的点数和中心点数。因为

$$\mathrm{F}=\frac{SS_{纯二次}}{\hat{\sigma}^2}=\frac{0.0027}{0.0430}=0.063$$

很小，没有显示出纯二次项的影响。

此模型的方差分析概括在表 11.2 中，交互作用和弯曲性的检验都不显著，而总回归的 F 检验是显著的。此外，$\hat{\beta}_1$和$\hat{\beta}_2$的标准差是

$$\mathrm{se}(\hat{\beta}_i)=\sqrt{\frac{MSE}{4}}=\sqrt{\frac{\hat{\sigma}^2}{4}}=\sqrt{\frac{0.0430}{4}}=0.10 \quad i=1,2$$

回归系数$\hat{\beta}_1$和$\hat{\beta}_2$相对于它们的标准差都较大，此时我们没有理由怀疑一阶模型的合适性。

表 11.2 一阶模型的方差分析

方差来源	平方和	自由度	均方	F_0	P 值
回归($\hat{\beta}_1,\hat{\beta}_2$)	2.8250	2	1.4125	47.83	0.0002
残差	0.1772	6			
（交互作用）	（0.0025）	1	0.0025	0.058	0.8215
（纯二次）	（0.0027）	1	0.0027	0.063	0.8142
（纯误差）	（0.1720）	4	0.0430		
总和	3.0022	8			

以上内容参见《实验设计与分析》（第 6 版）。

下面提供 python 代码（文件链接为 http://www.aluoyun.cn/images/code11-1.txt）：

```
#code11-1.py
#本代码作者李绍安联系电话 13712566524，转载请注明出处：www.aluoyun.cn
#from pyDOE2 import *
import statsmodels.formula.api as smf
import pandas as pd
import  statsmodels.api  as  sm
from sklearn import linear_model
import seaborn as sns
import matplotlib.pyplot as plt
Viscosity =[39.3,40.0,40.9,41.5,40.3,40.5,40.7,40.2,40.6]
Time =[30,30,40,40,35,35,35,35,35]
Temperature =[150,160,150,160,155,155,155,155,155]
data= {"Viscosity":Viscosity,"Temperature":Temperature,"Time": Time}
df =pd.DataFrame(data)
model = smf.ols('df.Viscosity ~df.Time+ df.Temperature', data=df).fit()
print(model.summary2())
print(model.params)
anovatable=sm.stats.anova_lm(model)

ax = sns.residplot(x=model.predict(df.Temperature), y=df.Viscosity, lowess=False,
color='black')
ax.set_xlabel('Fitted value')
ax.set_ylabel('Residuals')
plt.show()
ax = sns.residplot(x=model.predict(df.Time), y=df.Viscosity, lowess=False, color='black')
ax.set_xlabel('Fitted value')
ax.set_ylabel('Residuals')
plt.show()
fig=plt.figure(figsize=(8,12))
fig=sm.graphics.plot_ccpr_grid(model,fig=fig)

Viscosity =[39.3,40.0,40.9,41.5,40.3,40.5,40.7,40.2,40.6]
```

```
Time=[-1,-1,1,1,0,0,0,0,0]
Temperature =[-1,1,-1,1,0,0,0,0,0]
data= {"Viscosity":Viscosity,"Temperature":Temperature,"Time": Time}
df =pd.DataFrame(data)
model = smf.ols('df.Viscosity ~ df.Time+ df.Temperature', data=df).fit()
print(model.summary2())
print(model.params)

model = smf.ols('df.Viscosity ~df.Temperature +df.Time +df.Temperature :df.Time ',
data=df).fit()
print(model.summary2())
print(model.params)
anovatable=sm.stats.anova_lm(model)
sqsigmahat=((40.3)^2+(40.5)^2+(40.7)^2+(40.2)^2+(40.6)^2-(202.3)^2/5)/4
> sqsigmahat
[1] 0.043
F=4*5*(40.425-40.46)^2/(4+5)/sqsigmahat=0.06330749
```

要离开设计中心--点(x_1=0, x_2=0)--沿最速上升路径移动,对应于沿 x2 方向每移动 0.325 个单位，则应沿 x_1 方向移动 0.775 个单位，于是，最速上升路径经过点（x_1=0, x_2=0）且斜率为 0.325/0.775。工程师决定用 5 分钟反应时间作为基本步长。由ξ_1与 x_1 之间的关系式，知道 5 分钟反应时间等价于规范变量 x_1 的步长为Δx_1=1。因此，沿最速上升路径的步长是Δx_1=1.0000 和Δx_2=（0.325/0.775）Δx_1=0.42。

工程师计算了沿此路径的点，并观测了在这些点处的产率直到响应有下降为止。其结果见表 11.3，表中既列出了规范变量，也列出了自然变量，虽然规范变量在数学上容易计算，但是过程运行中必须用自然变量。图 11.5 画出了沿最速上升路径的每一步处的产率图，一直到第 10 步所观测到的响应都是增加的；但是，这以后的每一步收率都是减少的，因此，另一个一阶模型应该在点（ξ_1=85，ξ_2=175）的附近区域进行拟合。

表 11.3 例 11.1 的最速上升实验

步长	规范变量		自然变量		响应 y
	x_1	x_2	ξ_1	ξ_2	
原点	0	0	35	155	
Δ	1.00	0.42	5	2	
原点+**Δ**	1.00	0.42	40	157	41.0
原点+2**Δ**	2.00	0.84	45	159	42.9
原点+3**Δ**	3.00	1.26	50	161	47.1
原点+4**Δ**	4.00	1.68	55	163	49.7
原点+5**Δ**	5.00	2.10	60	165	53.8
原点+6**Δ**	6.00	2.52	65	167	59.9
原点+7**Δ**	7.00	2.94	70	169	65.0
原点+8**Δ**	8.00	3.36	75	171	70.4
原点+9**Δ**	9.00	3.78	80	173	77.6
原点+10**Δ**	10.00	4.20	85	175	80.3
原点+11**Δ**	11.00	4.62	90	177	76.2

原点+12**Δ**	12.00	5.04	95	179	75.1

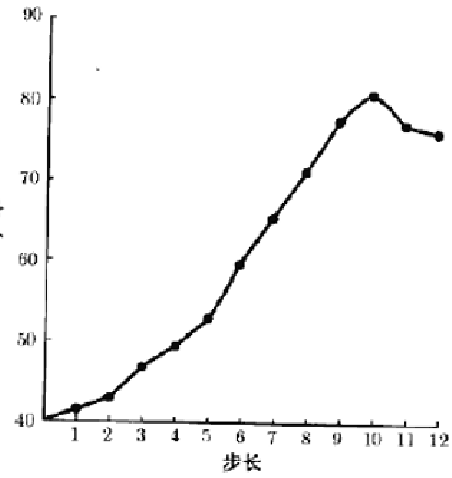

图 11.5 例 11.1 中沿最速上升路径的产率与步长的关系图

一个新的一阶模型在点（ξ_1=85，ξ_2=175）附近拟合，探测的区域对ξ_1是[80,90],对于ξ_2是[170，180]，于是，规范变量是

$$x_1=\frac{\xi_1-85}{5} \quad 和 \quad x_2=\frac{\xi_2-175}{5}$$

再次用 5 个中心点的 2^2 设计，实验设计列在表 11.4 中。

表 11.4 第 2 个一阶模型的数据

自然变量		规范变量		响应 y
ξ_1	ξ_2	x_1	x_2	
80	170	−1	−1	76.5
80	180	−1	1	77.0
90	170	1	−1	78.0
90	180	1	1	79.5
85	175	0	0	79.9
85	175	0	0	80.3
85	175	0	0	80.0
85	175	0	0	79.7
85	175	0	0	79.8

拟合表 11.4 的规范数据的一阶模型是

$$\hat{y}=78.97+1.00x_1+0.50x_2$$

此模型的方差分析，包括交互作用和纯二次项的检测，如表 11.5 所示，交互作用和纯二次项的检测表明，一阶模型不是合适的近似，真实曲面的弯曲性指明了我们已接近最优点，为更精细地确定最优点，在该点必须做进一步的分析。

表 11.5 第 2 个一阶模型的方差分析

方差来源	平方和	自由度	均方	F_0	P 值
回归($\hat{\beta}_1,\hat{\beta}_2$)	5.00	2			
残差	11.120	6			
（交互作用）	（0.2500）	1	0.2500	4.72	0.0955
（纯二次）	（10.6580）	1	10.6580	201.09	0.0001
（纯误差）	（0.2120）	4	0.0530		
总和	16.120	8			

以上内容参见《实验设计与分析》（第 6 版）。

下面提供 python 代码（文件链接为 http://www.aluoyun.cn/images/code11-2.txt）：

```
#code11-2.py
#本代码作者李绍安联系电话 13712566524，转载请注明出处：www.aluoyun.cn
#from pyDOE2 import *
import statsmodels.formula.api as smf
import pandas as pd
import statsmodels.api as sm
from sklearn import linear_model
import seaborn as sns
import matplotlib.pyplot as plt

Viscosity =[76.5,77.0,78.0,79.5,79.9,80.3,80.0,79.7,79.8]
Time=[-1,-1,1,1,0,0,0,0,0]
Temperature=[-1,1,-1,1,0,0,0,0,0]
data= {"Viscosity":Viscosity,"Temperature":Temperature,"Time": Time}
df =pd.DataFrame(data)
model = smf.ols('df.Viscosity ~df.Time+df.Temperature', data=df).fit()
print(model.summary2())
print(model.params)
anovatable=sm.stats.anova_lm(model)

ax = sns.residplot(x=model.predict(df.Temperature), y=df.Viscosity, lowess=False,
color='black')
ax.set_xlabel('Fitted value')
ax.set_ylabel('Residuals')
plt.show()

ax = sns.residplot(x=model.predict(df. Time), y=df.Viscosity, lowess=False, color='black')
ax.set_xlabel('Fitted value')
ax.set_ylabel('Residuals')
plt.show()

fig=plt.figure(figsize=(8,12))
fig=sm.graphics.plot_ccpr_grid(model,fig=fig)
```

```
model = smf.ols('df.Viscosity ~df.Temperature +df.Time +df.Temperature :df.Time ',
data=df).fit()
print(model.summary2())
print(model.params)
anovatable=sm.stats.anova_lm(model)
```

```
sqsigmahat=((79.9)**2+(80.3)**2+(80.0)**2+(79.7)**2+(79.8)**2-(79.9+80.3+80.0+79.7+79.8)**2/5)/4
>sqsigmahat
[1] 0.053
```

```
F=4*5*((76.5+77.0+78.0+79.5)/4-(79.9+80.3+80.0+79.7+79.8)/5)**2/(4+5)/sqsigmahat
> F
[1] 201.09
```