

主效应、交互效应与模型

3.1 主效应与交互效应

我们举一个最简单的实验设计来说明主效应和交互效应的概念。

在半导体生产实验中，考虑两个因子，每个因子皆两个水平。
A：反应温度。低水平（-1）：反应温度低；高水平（+1）：反应温度高。
B：浓度。低水平（-1）：浓度低，高水平（+1）：浓度高。以产率 Y 为响应变量（按“%”计），列表如下（见表3-1）。

表3-1 实验数据表

		A	
		低	高
B	低	60	72
	高	52	64

我们把表3-1画成一个示意图，便于大家理解（见图3-1）。

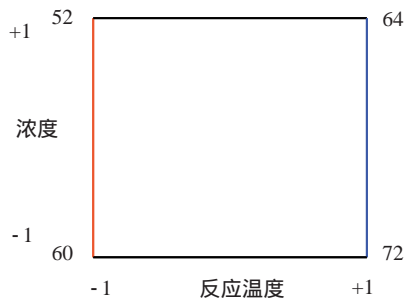


图3-1 因子效应示意图

如何分析因子A反应温度的效应呢？由于A处于低水平的情况（不考虑因子B）得到的产量的平均值是 $(60+52)/2 = 56$ ，A处于

高水平的情况（不考虑因子B）得到的产量的平均值是 $(72+64)/2 = 68$ 。产率由56提高到68完全是因子A的作用，这就是因子A的主效应。这时，我们得到因子A的主效应（main effect）为 $68 - 56 = 12$ ，表示随着从低温向高温的移动，收率平均增加12左右。

也可用下列算式计算：

$$\begin{aligned}\text{反应温度的主效应} &= [(+\text{符号的和}) - (-\text{符号的和})] / (+(-)\text{符号数}) \\ &= [(64 + 72) - (52 + 60)] / 2 \\ &= 24 / 2 = 12\end{aligned}$$

$$\text{B的主效应} = \text{B处于高水平时Y的平均值} - \text{B处于低水平时Y的平均值}$$

同样可以算出：

因子B的主效应 = $[(52 + 64)/2 - (60 + 72)/2] = 58 - 66 = -8$ ，表示随着低浓度向高浓度的转换，收率平均减少8左右。

也可用下列算式计算：

$$\begin{aligned}\text{浓度的主效应} &= [(+\text{符号的和}) - (-\text{符号的和})] / (+(-)\text{符号数}) \\ &= [(52 + 64) - (60 + 72)] / 2 \\ &= -16 / 2 = -8\end{aligned}$$

通过MINITAB可以很直观地画出主效应图（见图3-2）。

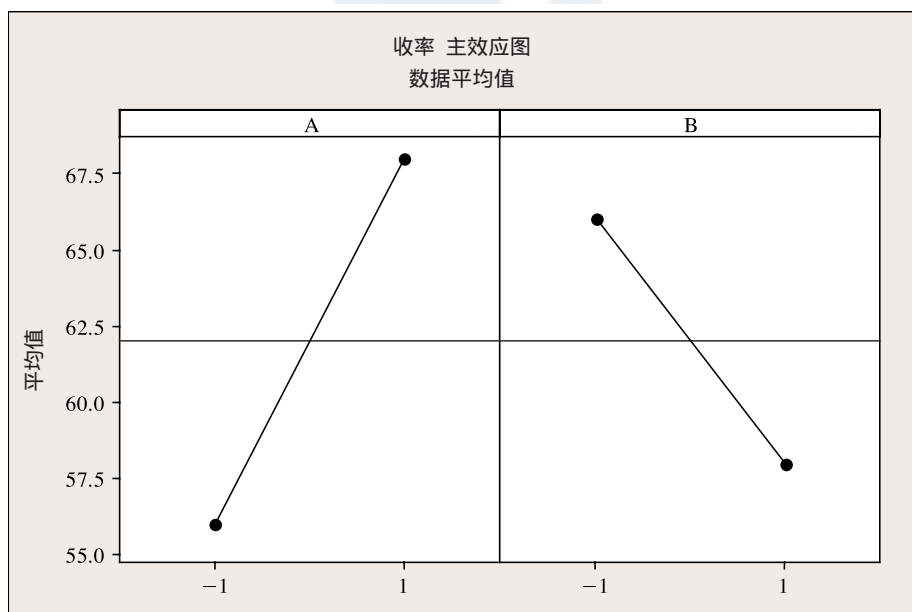


图3-2 主效应图

我们发现：当B（浓度）处于高水平时，因子A的效应为 $64 - 52 = 12$ ，当B（浓度）处于低水平时，因子A的效应为 $72 - 60 = 12$ ，二者相同。以因子B为横轴，以响应变量（产率）为纵轴作图（见图3-3），可以看出，两条线是平行的。

36 实验设计(DOE)应用指南

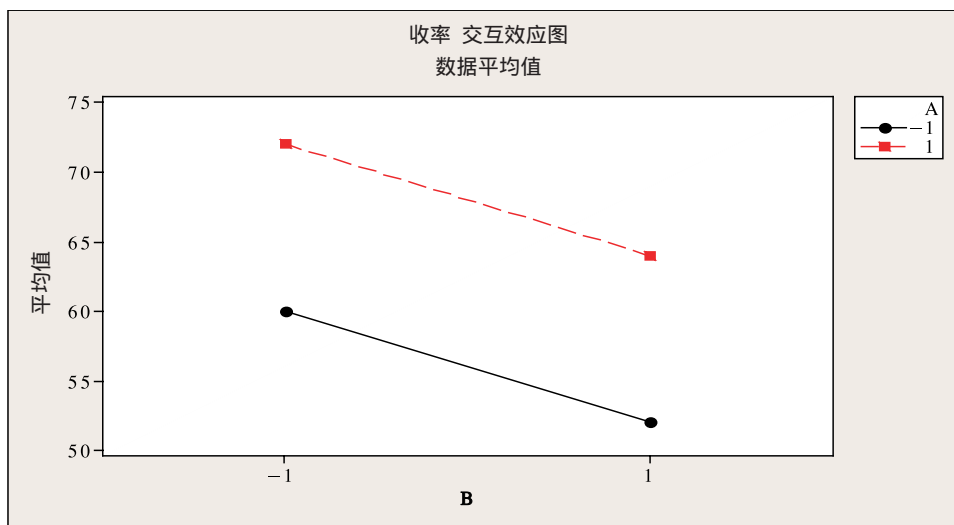


图3-3 无交互效应时的效应图

如果数据换成另一组 (见表3-2):

表3-2 有交互效应的数据表

		A	
		反应温度低 (-1)	反应温度高 (+1)
B	浓度低 (-1)	60	72
	浓度高 (+1)	52	83

我们把表3-2画成一个示意图 (见图3-4), 便于大家理解。

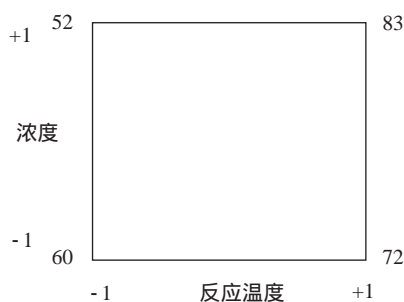


图3-4 因子效应示意图

如何分析因子A反应温度的效应呢? 由于A处于低水平的情况 (不考虑因子B) 得到的产量的平均值是 $(60+52) / 2 = 56$, A处于高水平的情况 (不考虑因子B) 得到的产量的平均值是 $(72+83) / 2 = 77.5$ 。产率由56提高到77.5完全是因子A的作用。这时, 我们得到因子A的主效应为 $77.5 - 56 = 21.5$, 表示随着从低温向高温的移动, 收率平均增加21.5左右。

也可用下列算式计算:

$$\begin{aligned}\text{反应温度的主效应} &= [(+\text{符号的和}) - (-\text{符号的和})] / (+(-)\text{符号数}) \\ &= [(83 + 72) - (52 + 60)] / 2 \\ &= 43 / 2 = 21.5\end{aligned}$$

B的主效应 = B处于高水平时Y的平均值 - B处于低水平时Y的平均值

同样可以算出：

因子B的主效应 = $[(52+83)/2 - (60+72)/2] = 67.5 - 66 = 1.5$ ，表示随着低浓度向高浓度的转换，收率平均增加1.5左右。

也可用下列算式计算：

$$\begin{aligned}\text{浓度的主效应} &= [(+\text{符号的和}) - (-\text{符号的和})] / (+(-)\text{符号的和}) \\ &= [(52 + 83) - (60 + 72)] / 2 \\ &= 3 / 2 = 1.5\end{aligned}$$

通过MINITAB可以很直观地画出主效应图（见图3-5）。

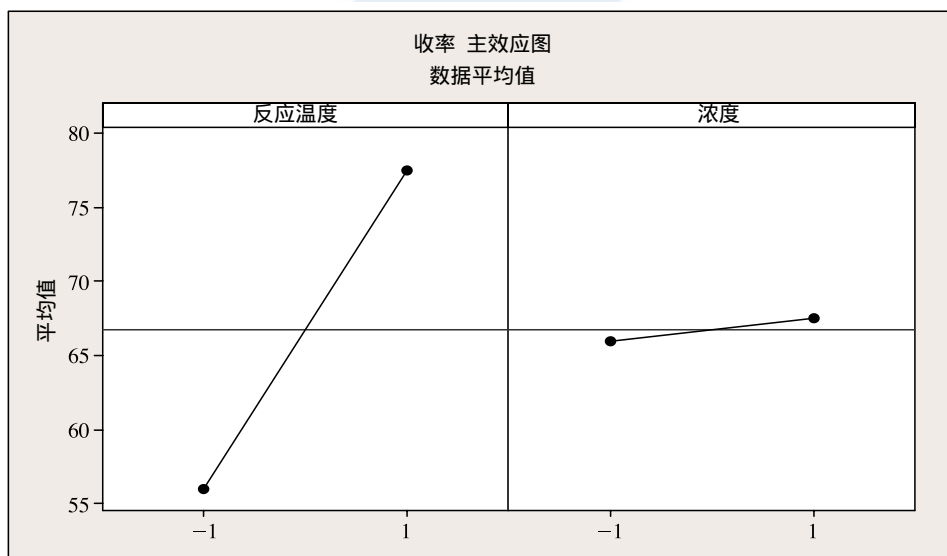


图3-5 主效应图

反应温度对收率的影响很大，但浓度对其几乎不构成影响。但由于因子间的交互效应有可能歪曲结果，所以只有当没有交互效应的前提成立时才是正确的。

不但如此，我们还发现；当B（浓度）处于高水平时，因子A的效应为 $83 - 52 = 31$ ；当B（浓度）处于低水平时，因子A的效应为 $72 - 60 = 12$ ，两者大不相同。如果以因子B为横轴，以响应变量（产率）为纵轴作图，可得图3-6。

可以看出，两条线是不平行的。

我们将两因子间有交互效应定义为：如果因子A的效应依赖于因子B所处的水平时，则我们称A与B之间有交互效应。

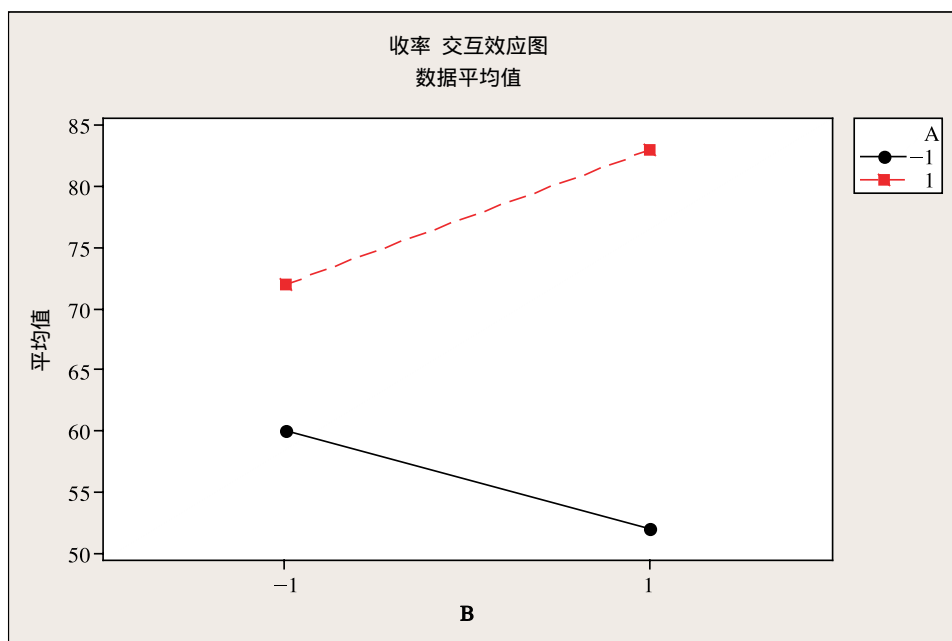


图3-6 有交互效应时的效应图

显然在前组数据中，因子A与因子B无交互效应，在后组数据中，因子A与因子B是有交互效应的。那么，如何来度量交互效应的大小呢？如果没有交互效应，当因子B处于不同水平时，因子A的效应是固定不变的，因此我们要考虑的定义交互效应的出发点就是，当因子B处于不同水平时，因子A的效应到底差了多少？

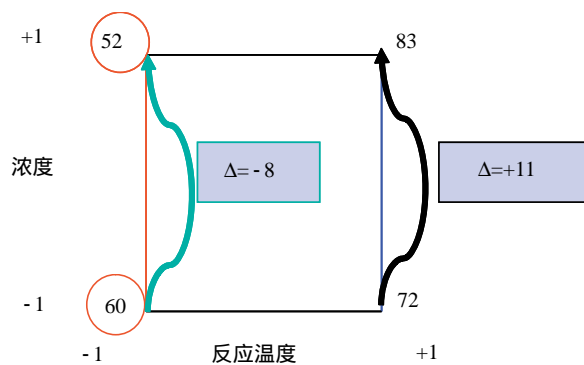


图3-7 因子效应示意图

当反应温度处于高水平时：随着浓度从低水平向高水平移动收率增加 11左右；

当反应温度处于低水平时：随着浓度从低水平向高水平移动时收率减小8左右。

所以，浓度的效应随着温度水平的不同而不同，温度和浓度存在交互效应。

因此，我们将A、B二因子的交互效应（interaction effect）表示为A × B，简称为AB，MINITAB常用A*B表示，其定义为：

$$AB交互效应 = (B处于高水平时A的效应 - B处于低水平时A的效应)/2$$

在后组数据中，交互效应 $A*B = [(83 - 52) - (72 - 60)]/2 = 9.5$ 。同样，交换AB顺序，可以得到：

$$BA交互效应 = (A处于高水平时B的效应 - A处于低水平时B的效应)/2$$

交互效应 $B*A = [(83 - 72) - (52 - 60)]/2 = 9.5$ 。显然二者是相同的， $A*B = B*A$ 。以后，我们不再区分 $A*B$ 或 $B*A$ 。另外，我们还可以得到变形的公式：

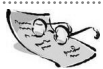
$$AB交互效应 = [(A高B高 + A低B低) - (A高B低 + A低B高)]/2$$

用本例数据代入，即交互效应 $A*B = [(60+83) - (52+72)]/2 = 9.5$ 。此公式可以帮助我们理解后面要讲到的从正交表中直接分析交互效应的计算公式。

我们注意的是，如果两个因子间存在显著的交互效应时，我们就不能只用主效应大小来作为该因子是否重要的判断依据。有时一个因子主效应很小，只要某个含它的交互效应显著，则这个因子就是重要的，就得予以保留。

实验设计示例简介

为了说明实验设计有关概念，我们先研究一个非常简单的例子。



例3-1 某半导体制造工程中得出以下数据，请计算各因子的主效应和交互效应。

在对工艺的研究过程中，发现有3个因子很重要，它们是因子A——反应温度；因子B——浓度；因子C——压力。对每个因子都设定了高低两个水平。我们希望考察这3个因子中，哪些因子的主效应及交互效应是显著的。其具体取值如下：

A因子——反应温度，低水平：160°C；高水平：180°C

B因子——浓度，低水平：20%；高水平：40%

C因子——压力，低水平：5psi；高水平：10psi

如何安排实验呢？3个因子按二水平全面搭配应该有8种组合。我们先看下面这张正交表（见表3-3）。

表3-3 三因子正交实验代码表

反应温度	浓度	压力	收率
-1	-1	-1	60
1	-1	-1	72
-1	1	-1	54
1	1	-1	68
-1	-1	1	52
1	-1	1	83
-1	1	1	45
1	1	1	80

我们将按照此表来安排实验，具体方法是这样的：凡表中列为-1者，我们实验时取低水平；凡表中列为1者，我们实验时取高水平。总共可以进行8次实验。如表3-3所示，数据

(续)

区分	A × B	A × C	B × C	A × B × C
总和 -	254	237	257	256
总和 +	260	277	257	258
差	6	40	0	2
平均效应	1.5	10	0	0.5

3.2 交互效应的进一步探讨

交互效应在自然界中广泛存在，正是人们对它的认识越来越深入，实验设计也得到较大的发展。在演讲时，为了说明交互效应，博克斯往往会举下面的例子。

响应 = 小兔子数

因子A，雌兔；低水平：不在笼子里；高水平：在笼子里。

因子B，雄兔；低水平：不在笼子里；高水平：在笼子里。

常识告诉我们，因子A对响应的影响，不仅仅由因子A的水平决定，还受到因子B的水平的影响。同理，因子B对响应的影响，不仅仅由因子B的水平决定，还受到因子A的水平的影响。因为响应只有在A因子与B因子都取高水平时，才可能不等于0，否则都为0。即，笼中的小兔子数，只有在雌兔与雄兔同在一只笼子里时才有可能产生小兔子，此例很好地说明了生活中的交互效应现象（见图3-8）。

响应 = 小兔子数

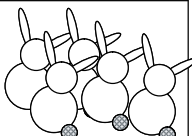
		因子A：雌兔	
		不在笼子里	在笼子里
因子B：雄兔	不在笼子里	0	0
	在笼子里	0	

图3-8 交互效应示意图

因子间的交互效应有三种方式：一种为无交互效应，一种为正向交互效应，另一种为反向交互效应（见图3-9）。

因子间的交互效应项随着因子个数的增加而增加。如四个因子A、B、C、D间的交互效应有以下几类：

二阶交互效应有6个：AB、AC、AD、BC、BD、CD

三阶交互效应有4个：ABC、ABD、ACD、BCD

四阶交互效应有1个：ABCD

42 实验设计(DOE)应用指南

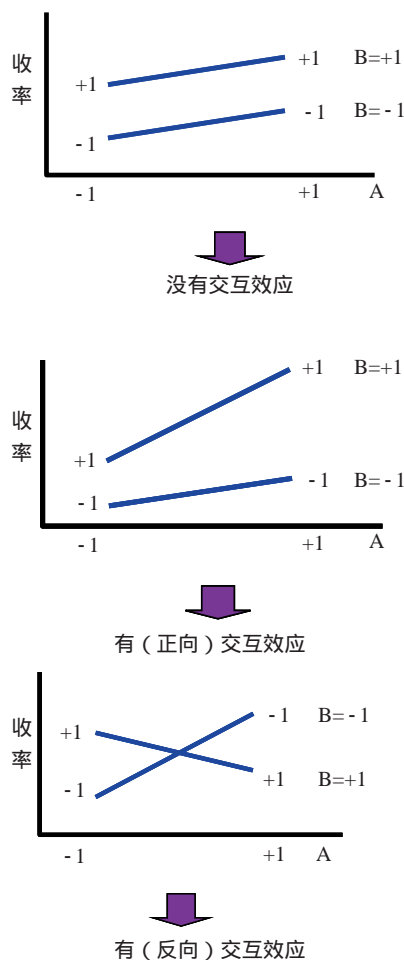


图3-9 因子A与B的交互效应示意图

共有11个，比因子个数还多。实际经验表明，多数交互效应是不存在的或很小以至可以忽略不计，实际中主要考虑部分二阶交互效应，具体考察哪些二阶交互效应还要依赖专业知识来决定。

3.3 模型

实验设计是一种有效的建立模型的方法，一旦你通过实验了解了因子变化对产品或工艺的响应的变化，你就可以建立一个模型来对未来的取值进行预估。完全析因实验与部分析因实验都是一阶模型，也称线性模型，如：

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_{12}x_1x_2 + \dots + \varepsilon$$

其中， β_1x_1 ， β_2x_2 同主效应有关， $\beta_{12}x_1x_2$ 同二阶交互效应有关， ε 为误差。

误差常用 ε 表示，除了包含有非可控因子（或噪声）所造成的“实验误差”（experimental error）外，它还可能包含“失拟误差”（lack of fit）。这里，失拟误差是指我们所采用的模型函数 f 与真实函数间的差异。实验误差与失拟误差这两种误差性质是不同的，分析时也要分别处理。有时为了简化，常假定函数关系 f 是准确的，从而可以忽略失拟误差。从上述概念中还可以看到，实验误差本身包含了测量误差，为了不使测量误差影响我们的分析结果，通常要在实验进行前，先进行测量系统的分析，只有测量误差满足了对测量系统的最低要求后，实验才能开始进行。有关测量系统的分析我们就不在这里讨论了。

如果曲率P-数值的统计显著性表明一阶模型已经不够用了，我们为了尽快找出最佳条件的时候，往往需要二阶模型，如：

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_{12}x_1x_2 + \beta_{11}x_1^2 + \beta_{22}x_2^2 \dots + \varepsilon$$

上式中增加了 $\beta_{11}x_1^2$ ， $\beta_{22}x_2^2$ 二次项。

所以，通过建立模型，我们能够确定：

- 哪些 x 对 y 有显著影响
- 知道 x_1, x_2, \dots 的值， y 会是什么值
- ▶ 最优化 y ，相对应的 x 是什么
- ▶ 如果知道 x 的变异，通过水晶球工具蒙特卡罗模拟，能求得 y 的变异
- ▶ 知道了客户的规格，能求得工程能力指数 C_p, C_{pk}
-

练习题

1. 请计算各因子的主效应以及二因子的交互效应（见图3-10）。
2. 利用MINITAB画出主效应图，交互效应图。

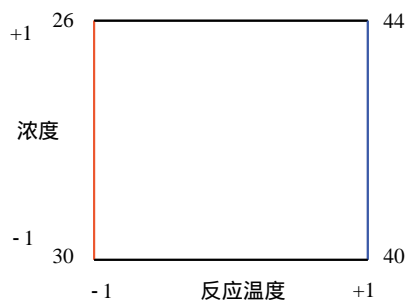


图3-10 因子效应示意图