

實驗設計與統計績效改善
之
期末報告---乳酪實驗

指導教授：郭 瑞 祥

組 員：B88701202 蔡韋鋒

B88701211 陳永豐

B88701229 高宜群

〔一〕 **實驗動機**：

在許多餐廳都有非常好吃的小點心，其中相當有名的一種小點心就是「提拉米蘇」，它滑嫩的口感及濃郁的乳香味都讓人吃了還想再吃，令人相當好奇的是這種看似簡單的點心為何有如此奇妙的口感及香味，於是我們便想瞭解它的製作成分，進而發現其中一項成分便是他香純味道的來源：乳酪。我們又發現其實製作乳酪的方法相當簡單，利用牛奶加入食用醋攪拌加熱便可析出牛奶的蛋白質，再將牛奶的蛋白質經過處理便成為了乳酪，而在我們選擇了許多實驗對象中，我們認為由「食品」下手，是最切實際、最實用，可以造就今後長遠的「口福」；所以本組選擇大家都熟悉的乳酪作為實驗對象，希望能找出造就「最佳自製乳酪」的所有環境變異。

〔二〕 **前置作業**：

前置作業有三個最主要的任務：

1. 決定 response 的衡量方式

即牛奶的蛋白質析出量，單位以（g）計算。

2. 找變因(factor)

我們從實際的從牛奶析出蛋白質的過程中，觀察出一些可能的因子，可能會對蛋白質數量的結果產生影響：

可能影響之因子	觀察結果
1. 牛奶粉及水的比例	當固定數量的牛奶粉加入不同數量的冷水中，會造成其中的比例不同，可能會影響析出的蛋白質量。
2. 放入沸騰水中的時間（隔水加熱）	在實驗的過程中，受熱時間過長或過短，會造成牛奶中蛋白質的凝結程度的不同。

3. 醋的加入量

一次加入的量不同，可能會影響析出牛奶蛋白質析出的數量，因為太少可能造成蛋白質無法凝聚，太多又可能使牛奶變質，導致失敗。

4. 攪拌的時間

我們攪拌的時間在於當我們裝牛奶的容器剛放進煮沸的水中便開始攪拌，而這樣的過程可能使牛奶的受熱更加均勻，與醋的反應更有效率，進而影響實驗結果。

可能影響之其他因素

觀察結果

隔水加熱時，鍋子內的水量 因為在隔水加熱的過程中，可能會因為水量的多寡造成熱量傳導的差異，因此我們認為不同的水量所傳導的熱量也許不同，造成實驗的誤差。

3. 設定實驗中各個變因的水準(level)

在了解析出牛奶中的蛋白質過程的大概情況之後，我們已有一些結果影響因子的概念，因此下一步我們將要實際訂出實驗的各個因子之水準值。在此，我們重新審視了我們的實驗目的：是想要在一個不是太離譜的悠關範圍內，試著找出對析出蛋白質結果有顯著影響的因子，以及進一步建議最適的組合，來幫助大家乳酪自製的更成功，也就是使得製造出來的乳酪的量最多！

區分因子的種類

從以上的表中，可以看到我們所列出的可能影響的因子，而隨著不同因子的來源不同，代表的意義也不同，所以我們把它分為以下兩類：

Factors (因子)

Factor1 : 牛奶及水的比例 (A)

Factor2 : 隔水加熱的時間 (B)

Factor3 : 醋的加入量 (C)

Factor4：攪拌時間 (D)

以上 A、B、C、D 四個影響因子，我們把它列為本次實驗的主要研究因素，原因是這幾項變因是在析出牛奶中蛋白質的過程中，可以控制且可能具重要性的因子，也是我們想探討的地方。

Blocks (集區)

Block1：隔水加熱時鍋中水量的因素

上面這個因素，我們認為是可能對試驗造成影響的隨機因子，但是卻是無法控制的，因此我們並不多花時間來討論它，只是藉由集區化的設計，達到把這類誤差分離出來的效果，才能有效的隔絕這種隨機的干擾，使其不影響實驗結果的分析。

設計各因子之水準(level)

在確定我們想研究的因子之後，接下來我們必須來設計實驗應有的因子水準。因此在這樣的考慮下，我們認為實驗的因子水準，應該設定在具攸關的範圍之內才有意義，也就是在一般可能被採用的範圍之內，選擇具代表性的水準 (level) 來做實驗。

實驗設計的因子水準如下表：

實驗因子	High Level (+)	Low Level (-)
Factor1：牛奶及水的比例 (A)	水 200 毫升	水 100 毫升
Factor2：隔水加熱的時間 (B)	2 分鐘	1 分鐘
Factor3：醋的加入量 (C)	2 匙	1 匙
Factor4：攪拌的時間 (D)	20 秒	10 秒

牛奶粉固定為一湯匙約 25 公克，醋一匙約 2 毫升

ABCD 為正值時，鍋中水量為 600 毫升；ABCD 為負值時，鍋中水量為 900 毫升。

總實驗次數為 16 次，一種水量進行 8 次的實驗。Replicate=1。

〔四〕 實驗結果：

<i>No</i>	<i>Run</i>	<i>Response (g)</i>
1	11	5.85
2	5	3.675
3	1	7.95
4	9	3.75
5	7	8.325
6	15	7.275
7	14	8.925
8	4	8.3625
9	2	5.85
10	10	3.5625
11	13	10.8375
12	6	5.625
13	16	6
14	3	6.825
15	8	6.7125
16	12	8.025

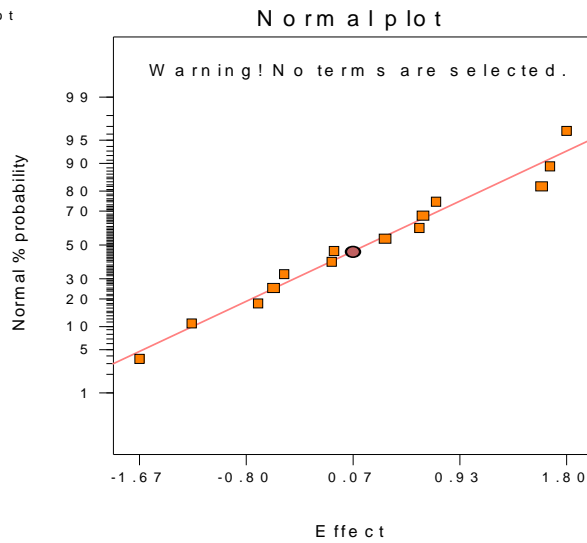
〔五〕 實驗分析

*找出顯著的因子

由於這是個沒有重複實驗的 2^4 設計，所以有兩種方法可以找出顯著的因子，第一個就是看 Normal Probability Plot，第二個就是把高次項的當作 error 來作 ANOVA Table 而下圖就是我們所做出來的 Normal Probability Plot，

DESIGN-EXPERT Plot
Response 1

A : A
B : B
C : C
D : D



由於看不太出來哪一個因子較顯著，所以我們採用第二種方法，把三次項以上的當 error，也就是把 ABC,ABD,ACD,BCD 都當作是 error，



然後經由 Design Expert 軟體的運算後我們可以得到 ANOVA Table，然後把不顯著的因素去除掉之後我們可以得到一個經過消除不顯著因子的 ANOVA Table。

把 3 次項的因子當作是 error

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F
Block	0.050625	1	0.050625		
Model	45.51785	4	11.37946	6.959866	0.0060
A	11.13891	1	11.13891	6.812738	0.0260
B	10.28004	1	10.28004	6.287441	0.0310
C	11.13891	1	11.13891	6.812738	0.0260
AC	12.96	1	12.96	7.926549	0.0183
Residual	16.35012	10	1.635012		
Cor Total	61.91859	15			

由上圖可知我們的 Model 的 p-value<0.05，顯著，代表我們所用的這個 Model 蠻成功的，而且所得到因子 A,B,C,AC 也都顯著，但是 MS(block)不顯著

* * 由 R-Square 來看 Model 的適合度

Std. Dev.	1.278676	R-Squared	0.735726
Mean	6.721875	Adj R-Squared	0.630016
C.V.	19.02261	Pred R-Squared	0.323458
PRESS	41.8563	Adeq Precision	6.62095

我們的 adjusted R-square 有 0.630016，代表我們這個 Model 的解釋能力有 63.0016%算是還可以的一個數值，而為什麼不是 100%呢？這就是由於有許多 noise 所造成的，以下是我們本組的推論：

1. 時間的精確度

比如說在加熱的時候，雖然我們是以沸騰後才開始計算時間不過由於人為的疏失很容易就會有一些時間拿捏不好的問題

2. 溫度的一致性

在實驗時雖然我們是雖然是以沸騰之後才將我們的杯子放入鍋子中，由於沸騰後的溫度我們沒辦法控制，所以傳導到杯子中的熱量無法一致，所以這也會造成 noise

3. 蛋白質的流失

我們實驗攪拌的時候筷子上附著了一些未凝固的蛋白質，還有在用濾網過濾時也有許多細微的蛋白質隨之而流失

4. 水汽的蒸發

我們在做實驗時由於為了減少實驗的時間我們並沒有每一次都將水重新煮沸，所以水汽一直隨著不停的沸騰而蒸發造成水量的誤差

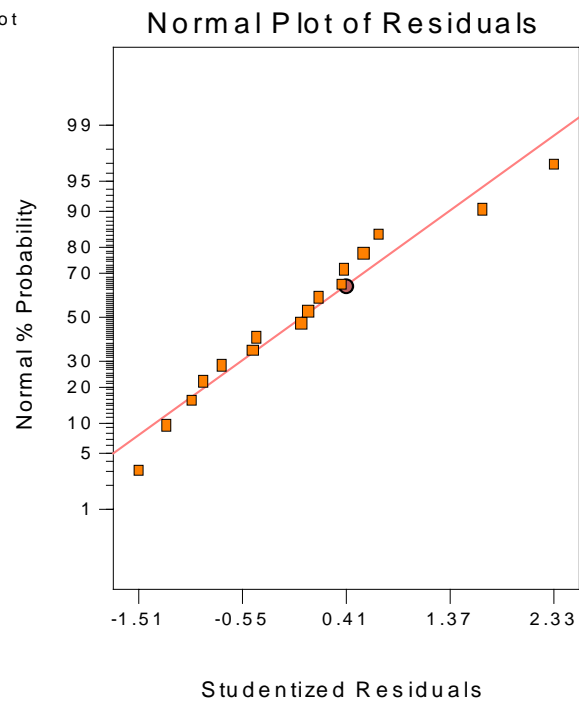
*** regression model

$$\text{response} = 6.72 - 0.83 * A + 0.80 * B + 0.83 * C + 0.90 * A * C$$

**** 殘差分析圖

- normal probability plot

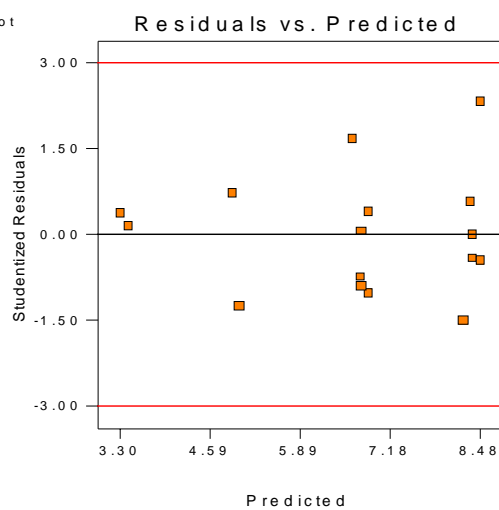
DESIGN-EXPERT Plot
Response 1



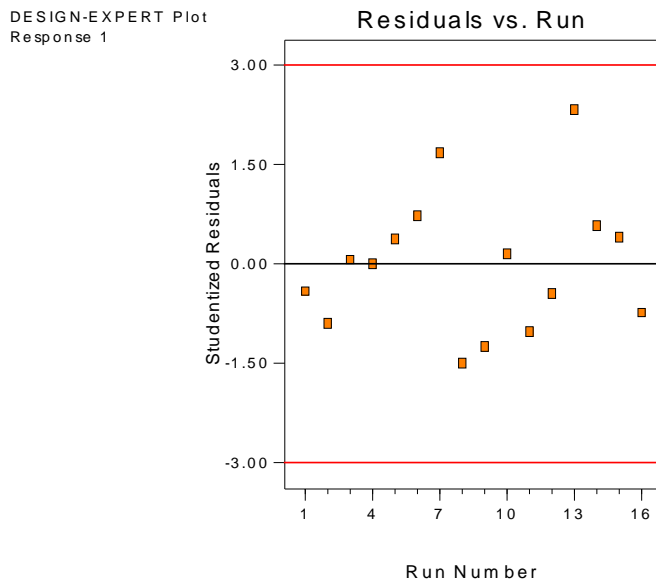
由圖看來並沒有違反 normal distribution 的假設

- constant variance

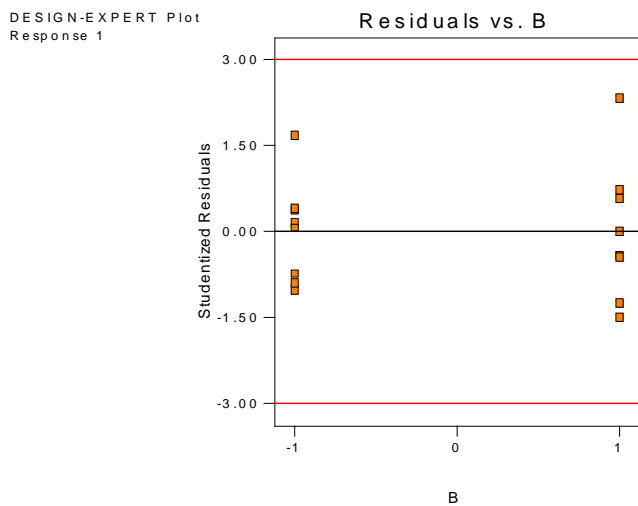
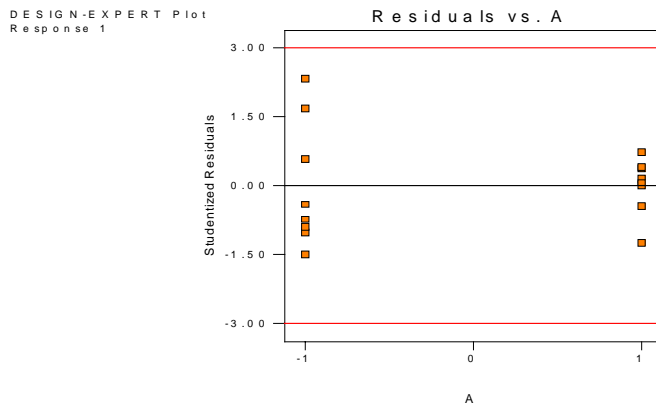
DESIGN-EXPERT Plot
Response 1

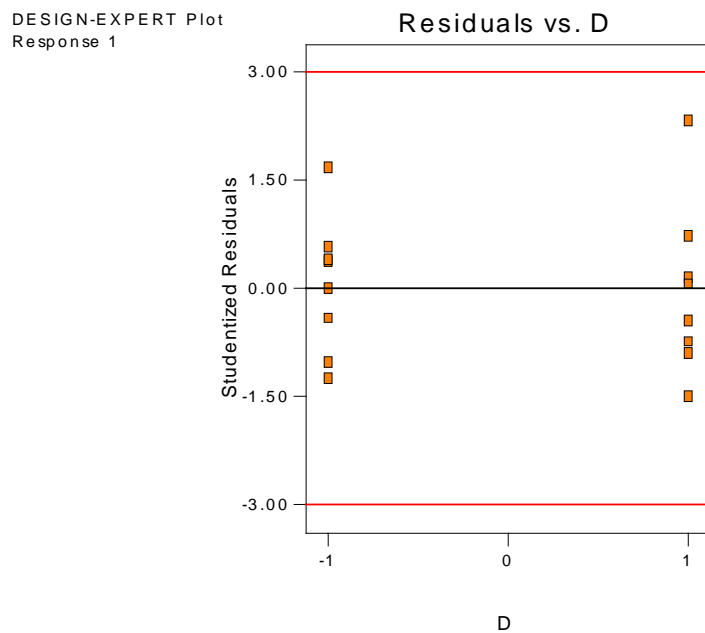
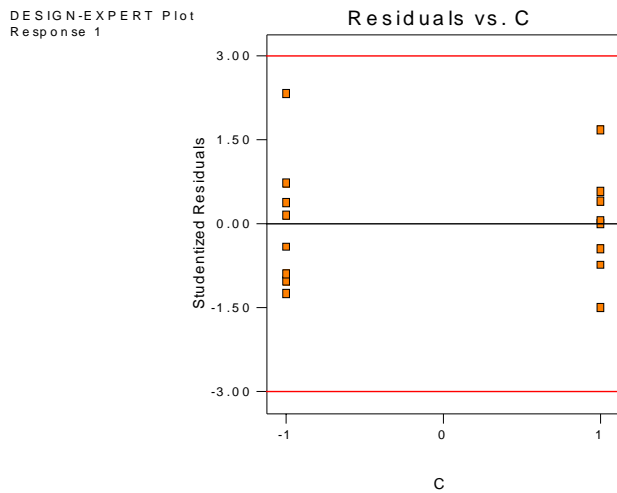


很明顯的我們可以 residuals 隨著跟 predicted value 變大而增大，所以我們可以採用 transformation 來改進



這是由 residual 跟每一次的實驗所得到的圖，可以看得出來並沒有特定的形狀





由以上四個圖可以看出 residual 對各個 factor 的關係的確違反了 constant variance 的假設所不適合來作 Anova 檢定，所以我們決定用 transformation 來改進了經過 try & error 之後，發現到用 inverse sqrt 之後的結果較好，也就是所得到的 residual plot 變得較不違反假設，而 ANOVA table 如下：

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F
Block	0.0002075	1	0.0002075		
Model	0.059	4	0.059	12.65	0.0006
A	0.014	1	0.014	12.37	0.0056
B	0.008591	1	0.008591	7.35	0.0219
C	0.019	1	0.019	16.00	0.0025

AC	0.017	1	0.017	14.88	0.0032
Residual	0.012	10	0.001168		
Cor Total	0.071	15			

由上圖可知我們的 Model 的 p-value<0.05，顯著，代表我們所用的這個 Model 蠻成功的，而且所得到因子 A,B,C,AC 也都顯著，但是 MS(block)不顯著

* 由 R-Square 來看 Model 的適合度

Std. Dev.	0.034	R-Squared	0.8350
Mean	0.40	Adj R-Squared	0.7690
C.V.	8.54	Pred R-Squared	0.5776
PRESS	0.030	Adeq Precision	8.973

Adj R-squared 值變大成為 0.7690，也就是經過 transformation 後的 model 較適合！

* * * regression model

$$\text{response} = 0.4 + 0.03 * A - 0.023 * B - 0.034 * C - 0.033 * A * C$$

接著我們就來解釋顯著的因素。

- A. 代表牛奶及水的比例，這是蠻符合我們剛開始的假設，因為在這個化學實驗中反應物的濃度越大越有利反應的發生，所以一開始我們才會設計這個因子作為有可能影響 output 的實驗因子，而且我們看所得到的 regression model，A 的係數是正 0.03 的，代表水量越多越不利於反應的發生，完全符合一般的化學常識
- B. 代表沸騰後的加熱時間，這個也是我們做出來顯著的因素，一般來說加熱時間越多越能使奶粉融入水中，使溶液的溶度增大，加快反應的速度，但是溶解度不可能無限變大，所以我們所得到的 regression model 所得到的 B 是負的 0.023，即表示過多的加熱時間並不會使乳酪的量增加，相反的，會使得量減少。
- C. 代表加入醋的量，這個因素作出來顯著也是蠻符合常理的，因為

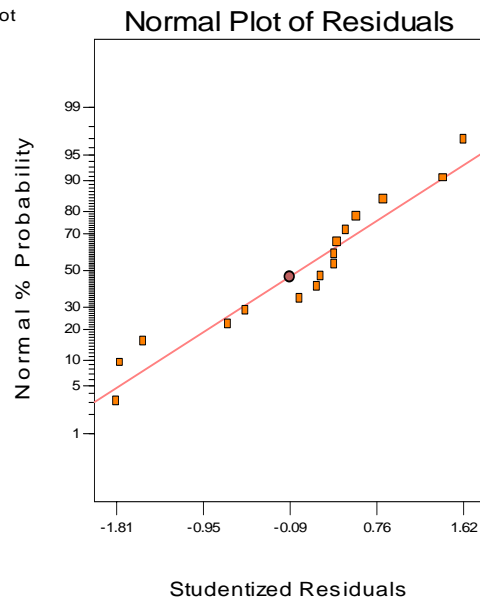
反應物的量越多，接觸面積越大，越有利於反應的發生，但是我們得到的 regression model 的 C 是負 0.034，這讓我們的覺得很不同與事先的想法！

AC.代表交互項的影響，由於水量越多跟加入的醋量越多雖然是有相抵銷效果可是總合起來極有可能會讓整個反應溶液濃度減小使得反應變慢，反之，加入的水量小跟加入的醋越少雖然不曉得會是有利還是不利於反應的發生，但是有可能因共同的溶液溶度減小使得反應加快，所以這個交互項才會顯著。

*** 殘差分析圖

- normal probability plot

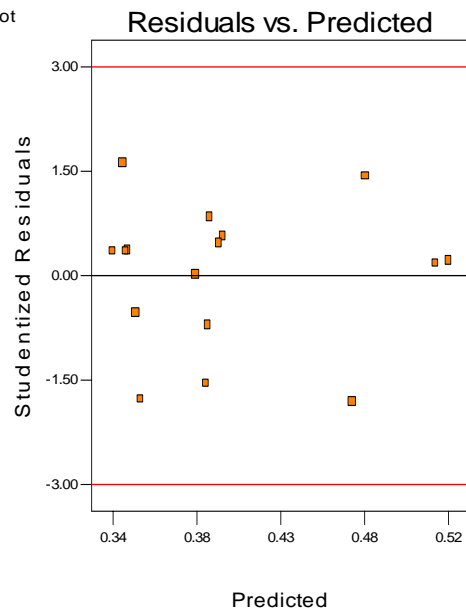
DESIGN-EXPERT Plot
1.0/Sqrt(Response 1)



由圖看來並沒有違反 normal distribution 的假設

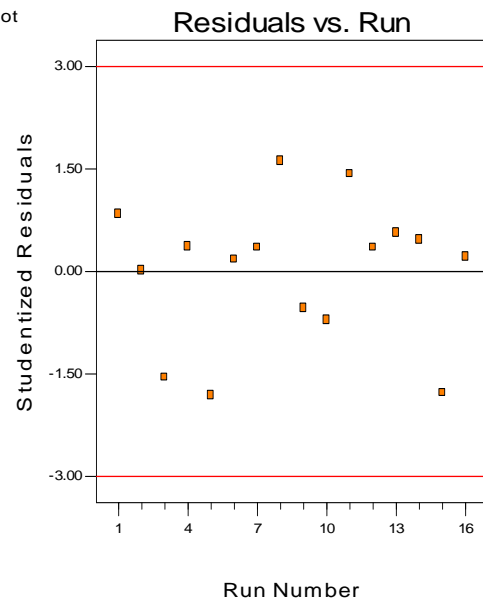
- constant variance

DESIGN-EXPERT Plot
1.0/Sqrt(Response 1)



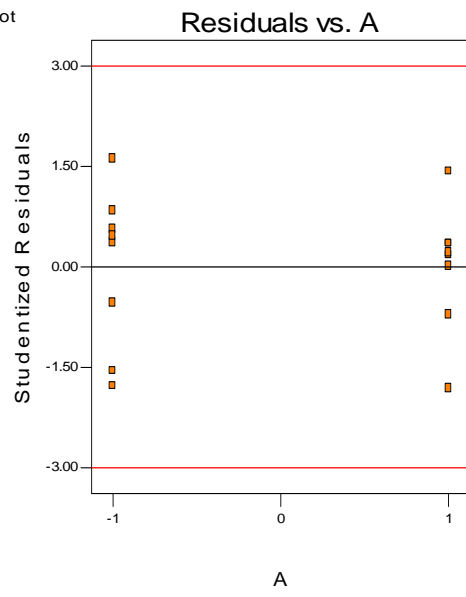
似乎沒有特別的規則！

DESIGN-EXPERT Plot
1.0/Sqrt(Response 1)

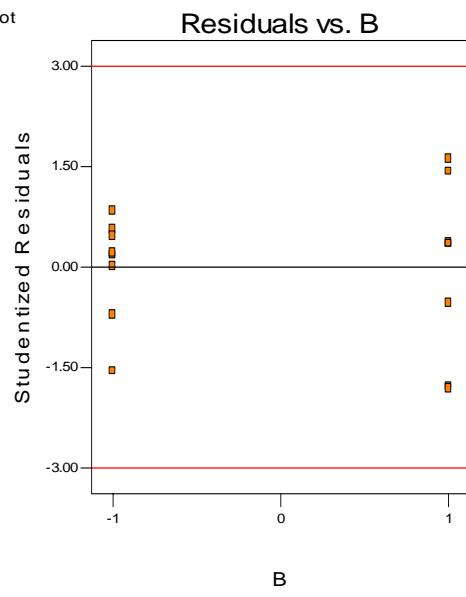


這是由 residual 跟每一次的實驗所得到的圖，可以看得出來並沒有特定的形狀

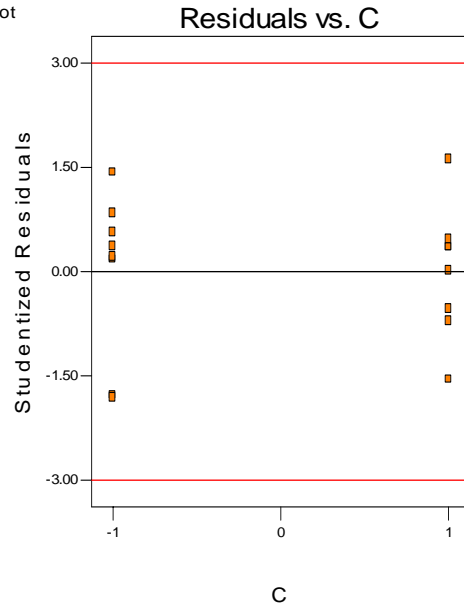
DESIGN-EXPERT Plot
1.0/Sqrt(Response 1)



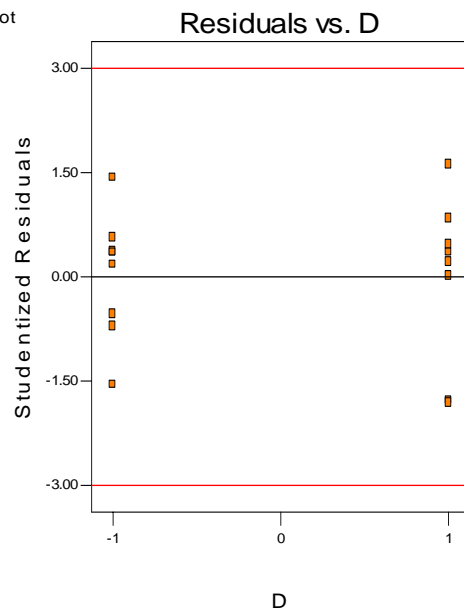
DESIGN-EXPERT Plot
1.0/Sqrt(Response 1)



DESIGN-EXPERT Plot
1.0/Sqrt(Response 1)



DESIGN-EXPERT Plot
1.0/Sqrt(Response 1)



根據上述的 model，我們本組做出了一個對於乳酪製造量最大值的數學預測值：

A	B	C	D*	1.0/Sqrt(Response 1)
<u>1.00</u>	<u>-1.00</u>	<u>-1.00</u>	<u>-0.19</u>	<u>0.521</u>

***** 改進

1. 於我們所作出來的數據 MS(block) 很小，不顯著，所以如果我們還要在做類似的實驗時，就可以不用這項 block 了
2. D 因子的不顯著代表說，用筷子攪拌的有無對我們的實驗沒有影響的所以如

果要做類似的實驗時可以不要考慮這個因素了，而考慮其他的因素了

- 3 . 雖然我們得到的 model 顯著可是由於違反了 constant variance 的假設所以我們可以用 transformation 的方式來改進
- 4 . 在實驗的最後我們發現當我們在把沸騰的蛋白質放入濾網要析出時由於很燙所以我們在放到濾網的過濾過程，蛋白質沉澱的時間變得不固定，所以在下一次的實驗我們就可以把這個因子納入考慮