

實驗設計作業一

(課本例子做 CRD)

指導教授：沈葆聖 老師

學生：王奎賢

```

data one; /* 建立一個新的資料集one */
input y plant; /* 放入變數 */
cards;
12.2 10
11.4 10
12.4 10
16 20
15.5 20
16.5 20
18.6 30
20.2 30
18.2 30
17.6 40
19.3 40
17.1 40
18 50
16.4 50
16.6 50
; /* 放入值 */

proc means data = one noprint; var y; by plant; /* 依照資料集one，在不同plant下針對y變數做計算 */
output out = two mean = my; /* 算y的平均數，並輸出結果檔命名為two */
proc sort; by plant; /* 根據plant做排序 */
proc print; run;

data three; merge one two; by plant; /* 建立一個新的資料檔three，將one 和 two 資料檔合併 (依據plant) */
drop _type_ _freq_; /* 把沒必要的變數拿掉 (因為資料集one沒有這兩個變數) */

data four; set three; /* 建立新的資料檔four，把資料檔three放入 */
/* 建立新變數 */
x1 = plant;
x2 = plant**2;
x3 = plant**3;
x4 = plant**4;
proc print; run;

/* 將變數全放入會有過度配適的問題 */
/* regression (with quadratic cubic quartic terms implemented) */

```

```

proc reg data = four; /* 用資料four跑回歸分析 */
model y = x1 x2 x3 x4; /* 模型 */
output out = out1 p = py1; /* 將輸出結果命名為out1，並產生出預測值 */
/* 針對y my py1設定圖的樣式 */
symbol1 interpol=none value=square color=blue;
symbol2 interpol=none value=dot color=red;
symbol3 interpol=sm45 value=triangle color=green;
proc gplot; plot (y my py1)*plant/overlay; /* 將三條線劃在一起 */
title 'fitted plot';

/* 變數只放一些進模型 */
/* regression (with quadratic terms implemented) */
proc reg noprint data = four; /* 針對資料four做迴歸分析 */
model y = x1 x2; /* 模型 */
output out = out2 p = py2; /* 輸出結果命名為out2，並產生出預測值 */
/* 針對y my py2設定圖的樣式 */
symbol1 interpol=none value=square color=blue;
symbol2 interpol=none value=dot color=red;
symbol3 interpol=sm45 value=triangle color=green;
proc gplot; plot (y my py2)*plant/overlay;
proc print;run;

/* 另外一種畫圖的方法 */
proc sgplot data = out2; /* 針對資料out2畫圖 */
scatter x = plant y = y; /* 將值點在圖上 */
scatter x = plant y = my;
reg x=plant y=py2 / degree=2; /* 畫一條回歸線，多項式的次數為2 */
proc print;run;

/* experimental design */
/* 用於檢測資料是否有多項式之趨勢 */
proc glm data = out1 plot = none ;
class x1; /* 宣告x1是名目變數 */
model y = x1 ; /* 放入模型 */
contrast 'Linear' x1 -2 -1 0 1 2;
contrast 'Quadratic' x1 2 -1 -2 -1 2;
contrast 'Cubic' x1 -1 2 0 -2 1;
contrast 'Quartic' x1 1 -4 6 -4 1;

```

```
proc print;run;  
proc rsreg data = out1 ;  
model y = x1 / lackfit ;  
proc print;run;
```

程式結果

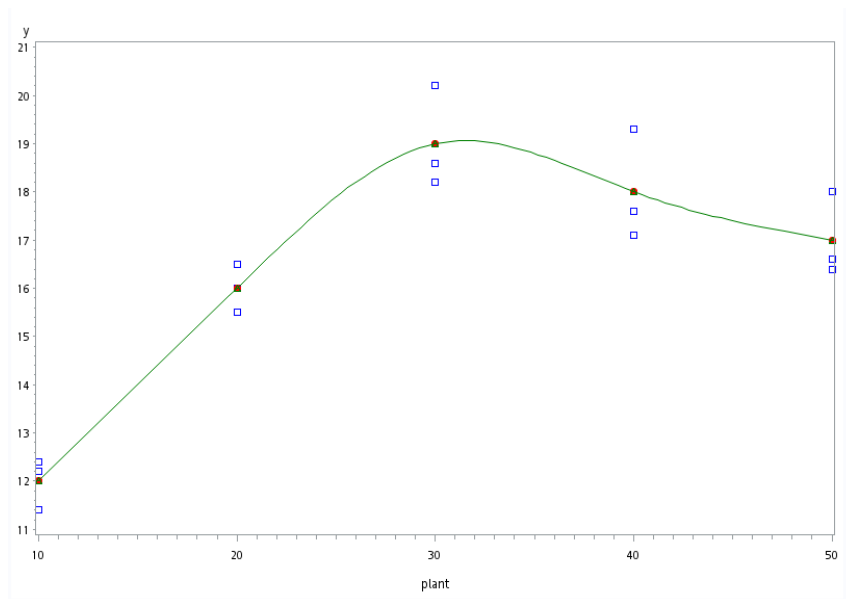


圖 1：此為模型配適圖（過度配適）

由於將所有變數放入模型裡，因此發生了過度配適的現象。

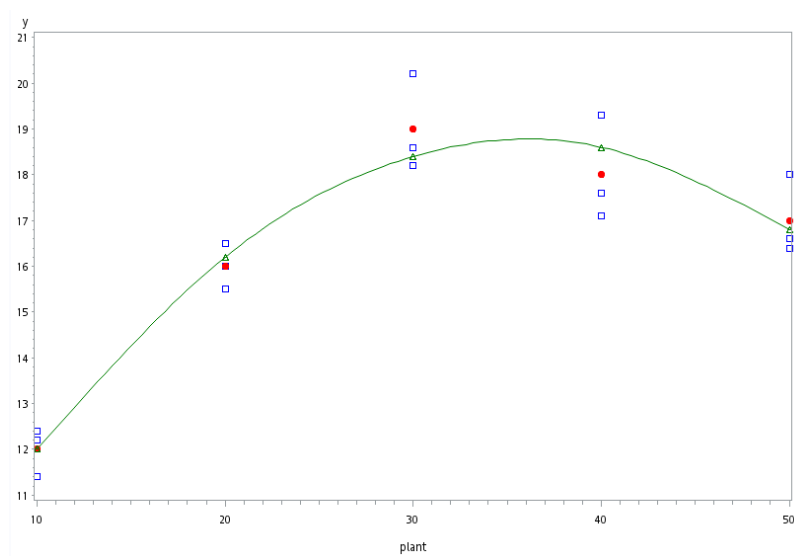


圖 2：此為模型配適圖（僅放兩個變數）

只有兩個變數放入模型裡，並無發生過度配適的現象。

對比	DF	對比 SS	均方	F 值	Pr > F
Linear	1	43.20000000	43.20000000	57.75	<.0001
Quadratic	1	42.00000000	42.00000000	56.15	<.0001
Cubic	1	0.30000000	0.30000000	0.40	0.5407
Quartic	1	2.10000000	2.10000000	2.81	0.1248

圖 3: 此為對比的 anova 表

對比 anova 表用於查看資料是否呈現線性、二項式等趨勢。從此結果可以看到在線性及二項式的 pvalue 皆小於 0.05，因此此資料可能有三次到四次多項式的趨勢。

迴歸	DF	類型 I 平方和	R 平方	F 值	Pr > F
線性	1	43.2000000	0.4544	52.47	<.0001
二次	1	42.0000000	0.4417	51.01	<.0001

殘差	DF	平方和	均方	F 值	Pr > F
缺少配適	2	2.400000	1.200000	1.60	0.2487
純誤差	10	7.480000	0.748000		
總計誤差	12	9.880000	0.823333		

圖 4: 缺適性(lack-of-fit)檢定的結果

缺適性用於檢查模型是否需要增加新的變數已達到更好的解釋力，從上面那張表可以看到在線性及二項式的 pvalue 皆小於 0.05，因此在這筆資料中，沒有線性跟二次的效應；而在下面的表顯示，此模型並不需要加入更多的變數(因為 pvalue 大於 0.05)