

# 迴歸分析作業一

(模擬資料)

指導教授：沈葆聖 老師

學生：王奎賢

```

dm'log; clear; output; clear;';
data one; /* 建立一個資料集one*/
rep = 1; n = 100; sigama = 1; /* 產生rep * n個觀察值 */
beta0 = 1; beta1 = 2; beta2 = 2; beta3 = 5; /* 設定參數 */
mu = 1; sigma = 1; mu2 = 2;
do r = 1 to rep;
do j = 1 to n;
seed1 = 1234; seed2 = 3456; seed3 = 899;
x1 = mu + sigma * rannor(seed1); /* x1隨機從N(1, 1)抽出 */
u = ranuni(seed2);
x2 = (mu2) / 2 * u; /* x2由u做轉換所得 */
epsilon = sigama * rannor(seed3); /* epsilon隨機從N(0, 1)抽出 */
ey = beta0 + beta1*x1 + beta2*x2 + beta3*x1**2;
y = ey + epsilon; /* 建立模型 */
output;
end;
end;
data one; set one; /* 建立資料集one，將剛才的資料集one放入 */
x3 = x1**2; /* 產生一個新變數x3 = x1^2 */
proc sort; by r; /* 根據r排序 */
proc print; run;
proc reg outest = tcorr; by r; /* 建立一個統計摘要表叫做tcorr */
model y = x1 x2 x3; /* 模型 */
output out = out1 p=py r=ry student = ry1 rstudent = ry2; /* 輸出資料檔並命名為
out1，放入了預測值(p)、殘差(r)、學生化殘差、去除化學生化殘差*/
title 'residual plot when model is correct';
data out1; set out1; if r = 1;
proc gplot; plot ry1*py = ' * ' / vref = 0; /* y軸為學生化殘差，x軸為預測值 */
title 'plot for studentized residual';
proc gplot; plot ry2*py = ' * ' / vref = 0; /* y軸為去除學生化殘差，x軸為預測值
*/
title 'plot for deleted residual';
proc means noprint; var y py ry; /* 針對y py ry 計算*/
output out = two mean = my mpy mry; /* 算平均，並輸出資料檔命名為two */
proc print; var my mpy mry;
run;
/* since model is correct, the bias of the estimated coefficient is very small */
proc means data = tcorr noprint; var x1 x2 x3;

```

```

output out = test mean = mx1 mx2 mx3;
proc print; var mx1 mx2 mx3;
proc reg data = one outest = tincorr; by r; /* 用資料集one做回歸分析，並輸出統計摘要表命名為tincorr */
model y = x1 x2; /* 模型 */
output out = out2 p = py r = ry student = ry1 rstudent = ry2; /* 輸出資料檔命名為out2，產生預測值、殘差、學生化殘差、去除學生化殘差 */
title 'residual plot when model is incorrect';
data out2; set out2; if r = 1; /* 建立資料檔命名為out2，將上一步的out2放入 */
proc gplot; plot ry1*py = ' * '/vref = 0;
title 'plot for studentized residual';
proc gplot; plot ry2*py = ' * '/vref = 0;
title 'plot for deleted residual';
*proc print; *var r x1 y py ry;
proc means noprint; var y py ry;
output out = two mean = my mpy mry;
proc print; var my mpy mry;
/* since model is incorrect, the bias of the estimated coefficient is large */
proc means data = tincorr noprint; var x1 x2;
output out = test mean = inmx1 inmx2;
proc print; var inmx1 inmx2;
run;

```

# 程式結果

## 一、當模型正確時：

變異數的分析					
來源	DF	平方和	均方	F 值	Pr > F
模型	3	29384	9794.66146	12128.4	<.0001
誤差	96	77.52464	0.80755		
已校正的總計	99	29462			

根 MSE	0.89864	R 平方	0.9974
應變平均值	15.88101	調整 R 平方	0.9973
變異係數	5.65856		

參數估計值					
變數	DF	參數估計值	標準誤差	t 值	Pr >  t
Intercept	1	0.88399	0.21448	4.12	<.0001
x1	1	2.15095	0.15606	13.78	<.0001
x2	1	1.92587	0.32601	5.91	<.0001
x3	1	5.01731	0.05221	96.09	<.0001

圖 1：迴歸分析結果(模型正確下)

從 ANOVA 結果顯示所有的參數估計值皆不等於 0，這結果相當合理，因為參數是給定的。另外，在模型正確下，參數的估計值跟實際的參數差異很小。(實際

$$\beta_0 = 1, \beta_1 = 2, \beta_2 = 2, \beta_3 = 5)$$

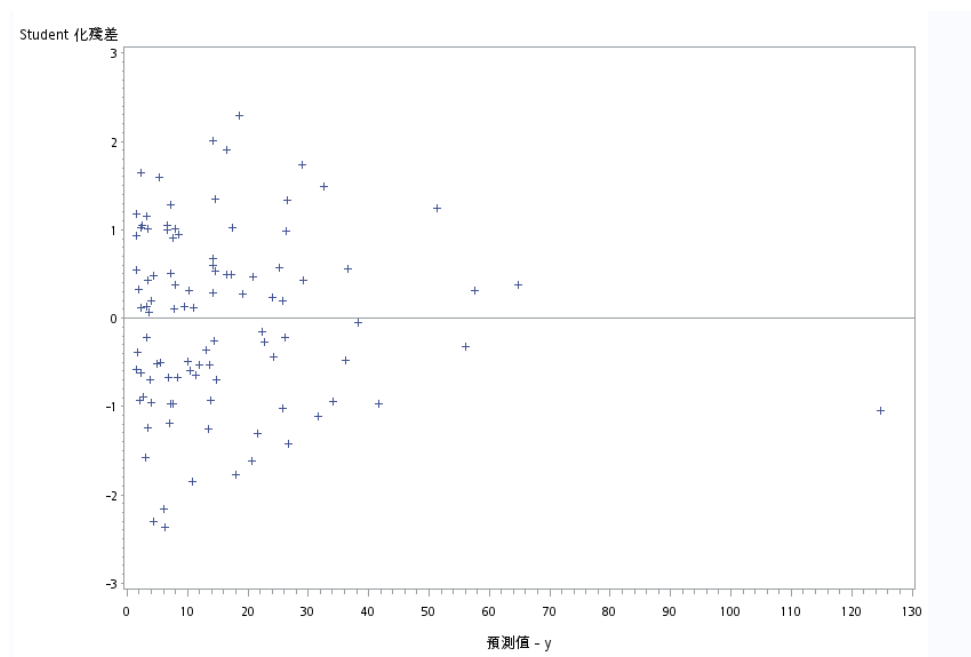


圖 2：殘差圖(模型正確下)

大部分的殘差都落在正負 2 之間，因此滿足模型變異數同質性。

## 二、當模型不正確時：

變異數的分析					
來源	DF	平方和	均方	F 值	Pr > F
模型	2	21928	10964	141.16	<.0001
誤差	97	7533.89572	77.66903		
已校正的總計	99	29462			

根 MSE	8.81300	R 平方	0.7443
應變平均值	15.88101	調整 R 平方	0.7390
變異係數	55.49396		

參數估計值					
變數	DF	參數估計值	標準誤差	t 值	Pr >  t
Intercept	1	-4.00020	2.04346	-1.96	0.0532
x1	1	14.52393	0.86478	16.79	<.0001
x2	1	7.51011	3.14597	2.39	0.0189

圖 3：迴歸分析結果(模型不正確下)

從 ANOVA 結果顯示所有的參數估計值皆不等於 0，因為參數是給定的。另外，在模型不正確下，參數的估計值跟實際的參數差異很大。(實際 $\beta_0 = 1$ ， $\beta_1 = 2$ ， $\beta_2 = 2$ ， $\beta_3 = 5$ )

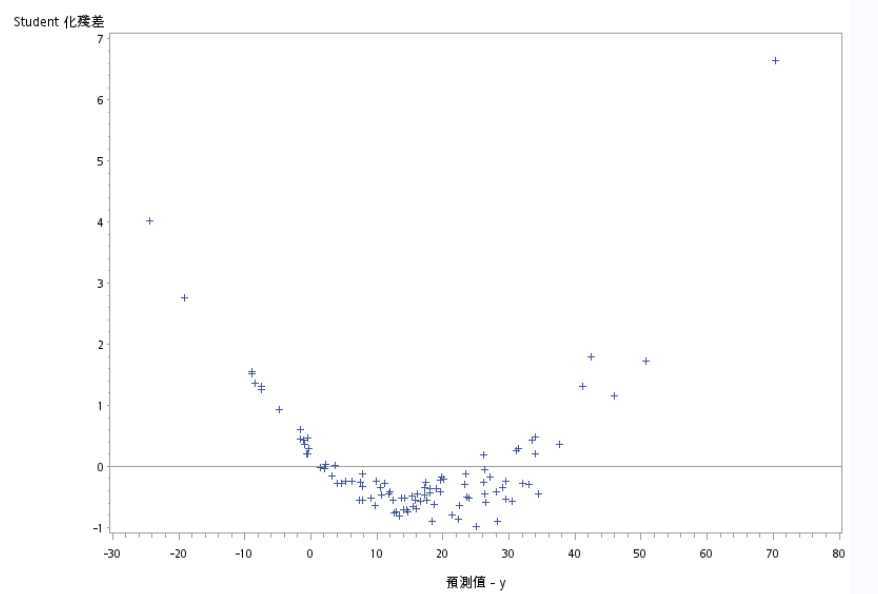


圖 4：殘差圖(模型不正確下)

殘差為二項式的形式，因此不滿足模型變異數同質性。