

目錄

	頁次
一、研究背景	2
二、資料來源	2
三、自變數說明	2
四、皮爾森相關係數	7
五、試一階模型	7
六、Box-Cox transform	9
七、la_b(左心房大小)開根號後重試一階模型	9
八、模型選擇	11
九、模型比較	15
十、依心房顫動種類畫出散佈圖	17
十一、選定之一階模型	18
十二、追加變數圖	22
十三、迴歸分析診斷	24
十四、多重共線性診斷	29
十五、交互作用項	30
十六、結論	35
附註 ---- SAS 程式	36

一、研究背景

決定左心房(LA)大小的因子很多,例如 body size (indexed to a measure of body surface area), pressure and volume overload, underlying mitral valve diseases, left ventricular systolic and diastolic dysfunction, athlete, and AF (*JACC 2006;47:p 2357-2363*). 左心房擴大被認為與心房顫動的發生及治療預後有關 *Prospective data for the large population-based studies have established a relationship between LA diameter and the risk of developing AF (The Framingham Heart Study. Circulation 1994;89:724-730)*. In the Cardiovascular Heart Study, LA size >50 mm (正常 <38 mm) 會使心房顫動發生機率提高四倍(*Mayo Clin Proc 2001;76:467-75*).

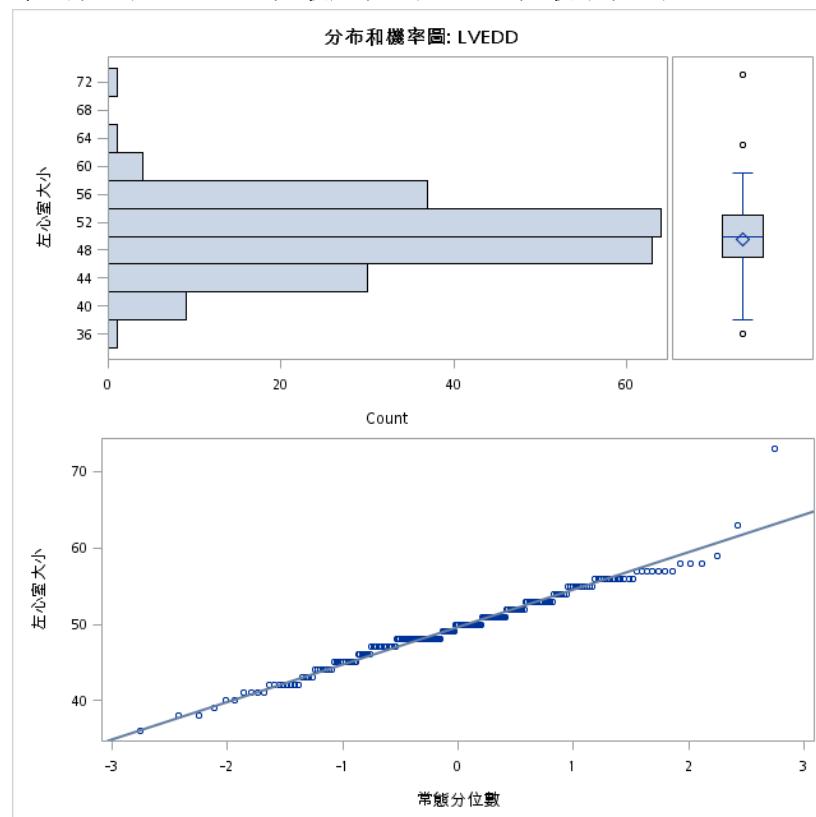
二、資料來源

林口長庚醫院心臟內科 2010/07/01~2014/10/30 因心房顫動接受電生理檢查及電燒手術治療病患基本資料。紀錄病患是否有高血壓、年齡、體重、左心房大小、左心室大小及功能、心室中膈厚度、心房顫動持續時間、二尖瓣逆流程度(輕或重)、性別、心房顫動種類(偶發性或慢性)、心房收縮量。共收集到 332 筆資料,隨機分成兩組,其中一組有 210 筆拿來做模型,另一組有 122 筆拿來做驗證。

三、自變數說明

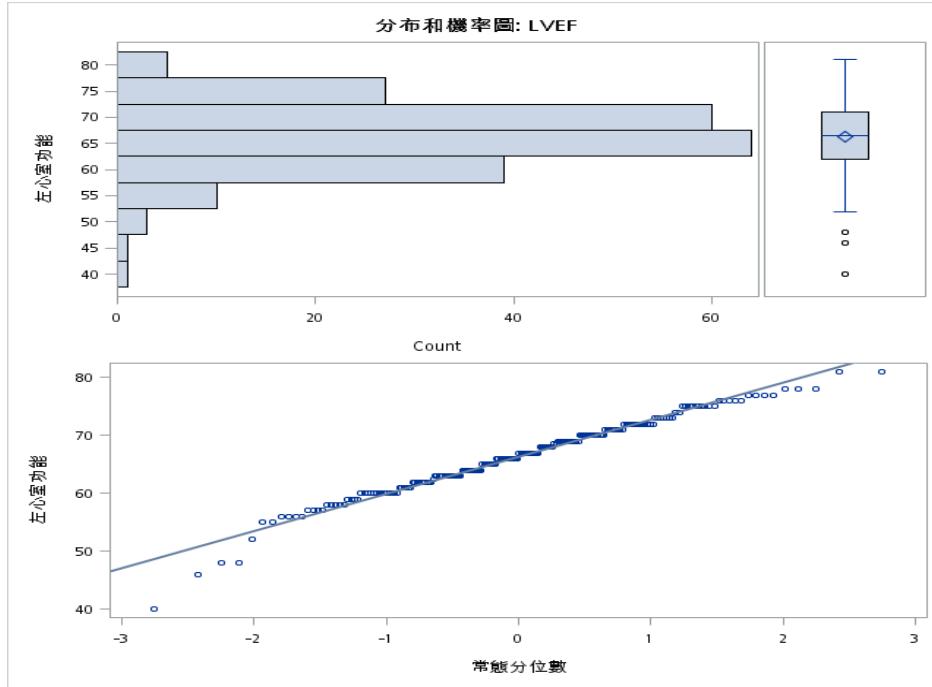
1. 左心室大小

平均值為 49.63mm, 最大值為 73mm, 最小值為 36mm.



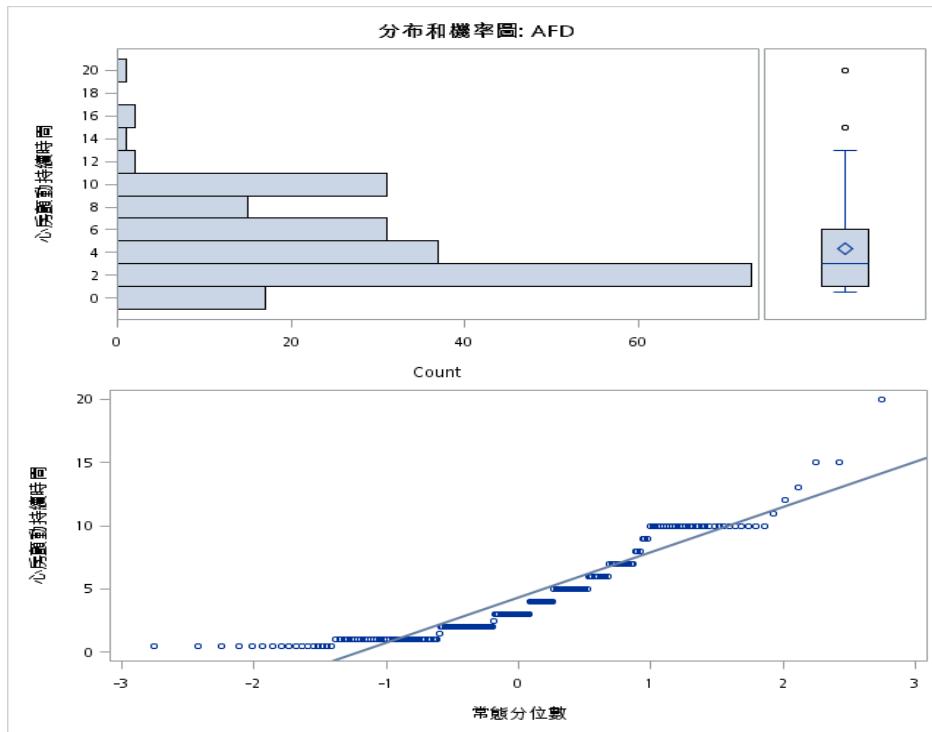
2. 左心室功能

平均值為 66%，最大值為 81%，最小值為 40%。



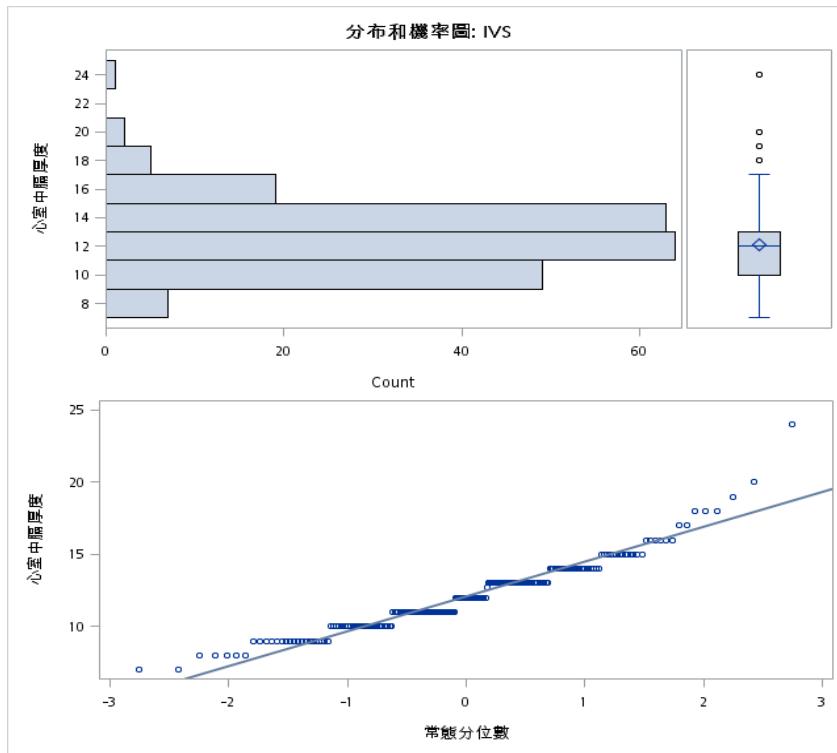
3. 心房顫動持續時間

平均值為 4.32 年，最大值為 20 年，最小值為 0.5 年



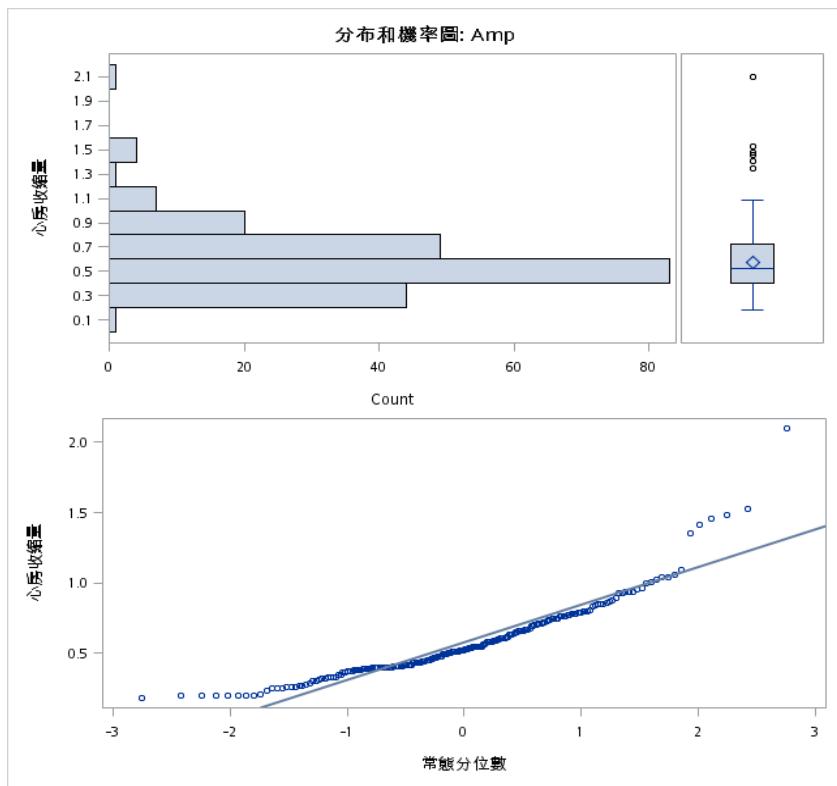
4. 心室中膈厚度

平均值為 12.08mm, 最大值為 24mm, 最小值為 7mm.



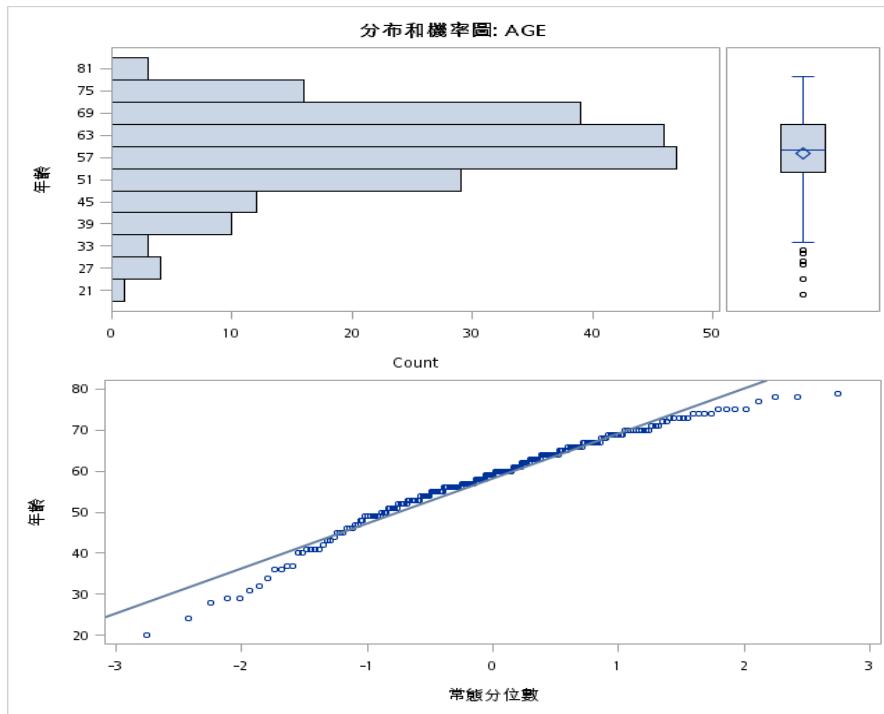
5. 心房收縮量

平均值為 0.58, 最大值為 2.1, 最小值為 0.18.



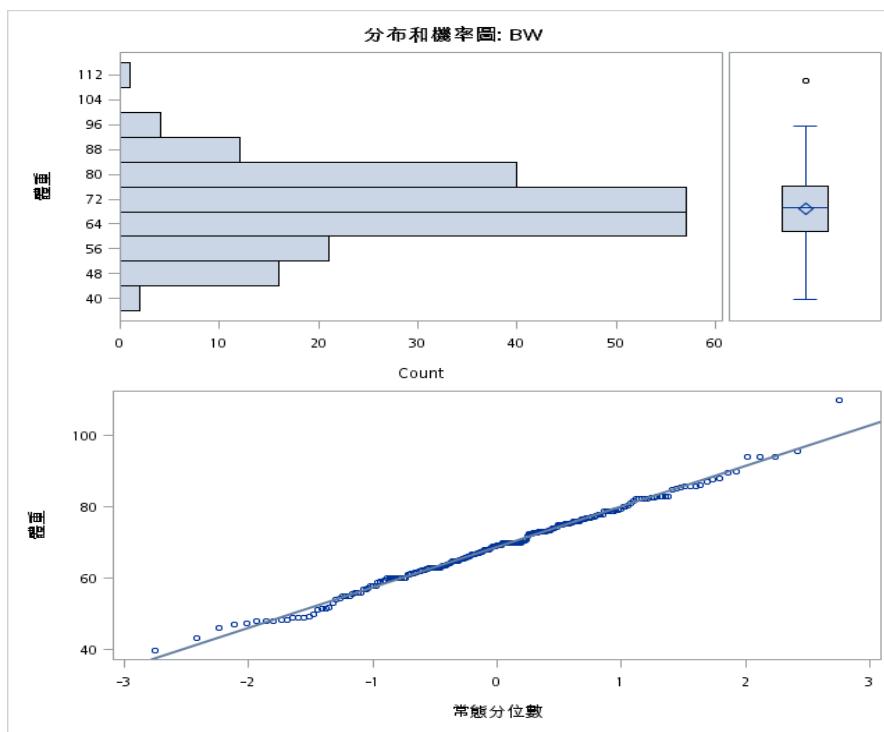
6. 年齡

平均值為 58 歲，最大值為 79 歲，最小值為 20 歲。



7. 體重

平均值為 68.8 公斤，最大值為 110 公斤，最小值為 39.7 公斤。



8. 二尖瓣逆流程度: 79.52%的人屬於輕度, 20.48%的人屬於重度
 高血壓: 52.86%的人沒有高血壓, 47.14%的人有高血壓
 心房顫動種類: 75.24%的人屬於偶發性, 24.76%的人屬於慢性
 性別: 68.57%的人屬於男性, 31.43%的人屬於女性.

二尖瓣逆流程度				
nMR	次數	百分比	累計次數	累計百分比
0	167	79.52	167	79.52
1	43	20.48	210	100.00

高血壓				
HTN	次數	百分比	累計次數	累計百分比
無	111	52.86	111	52.86
有	99	47.14	210	100.00

心房顫動種類				
aft	次數	百分比	累計次數	累計百分比
偶發性	158	75.24	158	75.24
慢性	52	24.76	210	100.00

性別				
nsex	次數	百分比	累計次數	累計百分比
男性	144	68.57	144	68.57
女性	66	31.43	210	100.00

四、皮爾森相關係數

Pearson 相關係數, N = 210 Prob > r (位於 H0 底下): Rho=0												
	LA_B	LVEDD	LVEF	AFD	IVS	Amp	AGE	BW	HTN	aft	nsex	nMR
LA_B	1.00000	0.31696	-0.07567	0.06104	0.28104	-0.05274	0.09844	0.29778	0.13465	0.43314	0.02761	0.17667
左心房大小		<.0001	0.2750	0.3788	<.0001	0.4471	0.1552	<.0001	0.0514	<.0001	0.6908	0.0103
LVEDD	0.31696	1.00000	-0.28084	-0.07331	0.13732	-0.17538	0.02469	0.34646	0.09776	0.10804	-0.22112	-0.04143
左心室大小	<.0001		<.0001	0.2903	0.0469	0.0109	0.7220	<.0001	0.1581	0.1185	0.0013	0.5504
LVEF	-0.07567	-0.28084	1.00000	0.06906	0.13878	0.16391	0.06948	0.02027	0.05989	-0.17276	0.01096	-0.07173
左心室功能	0.2750	<.0001		0.3193	0.0445	0.0174	0.3163	0.7703	0.3879	0.0122	0.8745	0.3008
AFD	0.06104	-0.07331	0.06906	1.00000	-0.07260	-0.01376	0.05221	-0.08500	-0.01949	-0.03299	-0.02065	-0.10316
心房顫動持續時間	0.3788	0.2903	0.3193		0.2950	0.8429	0.4517	0.2200	0.7789	0.6345	0.7661	0.1362
IVS	0.28104	0.13732	0.13878	-0.07260	1.00000	0.01833	0.10075	0.42291	0.27523	0.07137	-0.19784	-0.01772
心室中隔厚度	<.0001	0.0469	0.0445	0.2950		0.7917	0.1457	<.0001	<.0001	0.3033	0.0040	0.7986
Amp	-0.05274	-0.17538	0.16391	-0.01376	0.01833	1.00000	0.11100	-0.17758	0.03117	-0.30629	0.25512	0.04041
心房收縮量	0.4471	0.0109	0.0174	0.8429	0.7917		0.1087	0.0099	0.6533	<.0001	0.0002	0.5603
AGE	0.09844	0.02469	0.06948	0.05221	0.10075	0.11100	1.00000	-0.27197	0.14586	-0.02105	0.12455	0.26431
年齡	0.1552	0.7220	0.3163	0.4517	0.1457	0.1087		<.0001	0.0347	0.7617	0.0717	0.0001
BW	0.29778	0.34646	0.02027	-0.08500	0.42291	-0.17758	-0.27197	1.00000	0.19194	0.10532	-0.41307	-0.17424
體重	<.0001	<.0001	0.7703	0.2200	<.0001	0.0099	<.0001		0.0053	0.1282	<.0001	0.0114
HTN	0.13465	0.09776	0.05989	-0.01949	0.27523	0.03117	0.14586	0.19194	1.00000	0.01073	-0.04345	0.04086
高血壓	0.0514	0.1581	0.3879	0.7789	<.0001	0.6533	0.0347	0.0053		0.8771	0.5312	0.5559
aft	0.43314	0.10804	-0.17276	-0.03299	0.07137	-0.30629	-0.02105	0.10532	0.01073	1.00000	0.06315	0.14633
心房顫動種類	<.0001	0.1185	0.0122	0.6345	0.3033	<.0001	0.7617	0.1282	0.8771		0.3625	0.0341
nsex	0.02761	-0.22112	0.01096	-0.02065	-0.19784	0.25512	0.12455	-0.41307	-0.04345	0.06315	1.00000	0.19029
性別	0.6908	0.0013	0.8745	0.7661	0.0040	0.0002	0.0717	<.0001	0.5312	0.3625		0.0057
nMR	0.17667	-0.04143	-0.07173	-0.10316	-0.01772	0.04041	0.26431	-0.17424	0.04086	0.14633	0.19029	1.00000
二尖瓣逆流程度	0.0103	0.5504	0.3008	0.1362	0.7986	0.5603	0.0001	0.0114	0.5559	0.0341	0.0057	

從皮爾森相關係數可看到 P<0.05 的自變數有 LVEDD(左心室大小),IVS(心室中隔厚度),BW(體重),AFT(心房顫動種類),NMR(二尖瓣逆流程度).

五、試一階模型

變異數分析					
來源	自由度	平方和	平均值 平方	F 值	Pr > F
模型	11	3681.99884	334.72717	11.55	<.0001
誤差	198	5738.78212	28.98375		
已校正的總計	209	9420.78095			

根 MSE	5.38366	R 平方	0.3908
應變平均值	42.92381	調整 R 平方	0.3570
變異係數	12.54235		

參數估計值												
變數	標籤	自由度	參數 估計值	標準 誤差	t 值	Pr > t	第二型 SS	標準化 估計值	平方 偏 相關第二型	95% 信賴界限		
Intercept	Intercept	1	4.09681	6.91633	0.59	0.5543	10.16937	0	.	-9.54232	17.73594	
LVEDD	左心室大小	1	0.31729	0.08670	3.66	0.0003	388.21958	0.23251	0.06336	0.14633	0.48826	
LVEF	左心室功能	1	0.00822	0.06345	0.13	0.8970	0.48696	0.00784	0.00008485	-0.11690	0.13335	
AFD	心房顫動持續時間	1	0.25328	0.10564	2.40	0.0174	166.61387	0.13550	0.02821	0.04496	0.46161	
IVS	心室中膈厚度	1	0.39794	0.18106	2.20	0.0291	139.99827	0.14344	0.02381	0.04088	0.75500	
Amp	心房收縮量	1	2.48302	1.54908	1.60	0.1106	74.46764	0.09933	0.01281	-0.57179	5.53783	
AGE	年齡	1	0.04845	0.03847	1.26	0.2094	45.96969	0.07918	0.00795	-0.02742	0.12431	
HTN	高血壓	1	0.12474	0.78977	0.16	0.8747	0.72307	0.00930	0.00012598	-1.43271	1.68219	
BW	體重	1	0.14249	0.04412	3.23	0.0015	302.31142	0.24077	0.05004	0.05548	0.22949	
aft	心房顫動種類	1	5.92532	0.94486	6.27	<.0001	1139.84309	0.38185	0.16571	4.06204	7.78859	
nsex	性別	1	1.78771	0.92562	1.93	0.0549	108.11353	0.12391	0.01849	-0.03764	3.61305	
nMR	二尖瓣逆流程度	1	2.33309	0.99188	2.35	0.0196	160.36237	0.14056	0.02718	0.37709	4.28909	

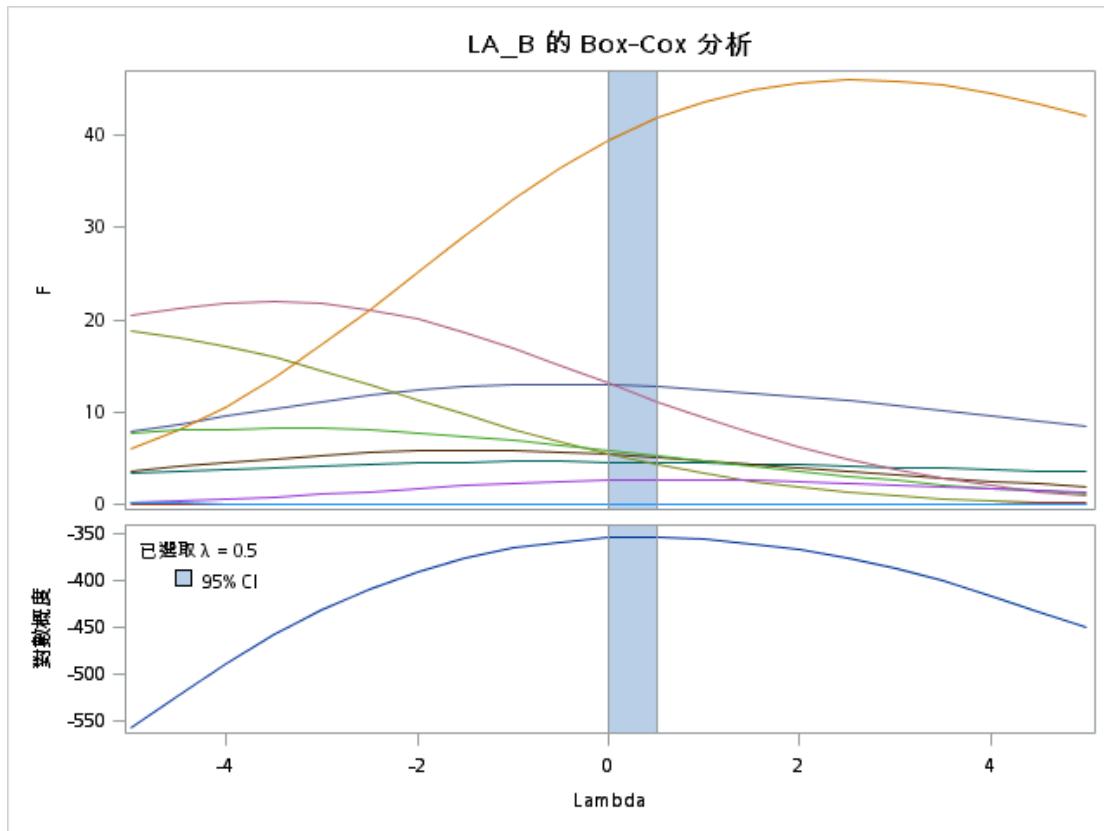
從一階模型可看到 $P<0.05$ 的自變數有 LVEDD(左心室大小), AFD(心房顫動持續時間), IVS(心室中膈厚度), BW(體重), AFT(心房顫動種類), NMR(二尖瓣逆流程度).

殘差

可看到一階模型的殘差不是常態分配(Shapiro-Wilk 檢定 $P=0.0162 <0.05$)

常態性檢定				
檢定		統計值		p 值
Shapiro-Wilk		W	0.983752	Pr < W 0.0162
Kolmogorov-Smirnov		D	0.080835	Pr > D <0.0100
Cramer-von Mises		W-Sq	0.178498	Pr > W-Sq 0.0098
Anderson-Darling		A-Sq	1.079292	Pr > A-Sq 0.0081

六、Box-Cox transform



從 Box-Cox transform 可知應選取 $\lambda = 0.5$, 也就是將 y 值開根號.

七、la_b(左心房大小)開根號後重試一階模型

變異數分析					
來源	自由度	平方和	平均值 平方	F 值	Pr > F
模型	11	22.07018	2.00638	11.95	<.0001
誤差	198	33.23143	0.16784		
已校正的總計	209	55.30161			

根 MSE	0.40968	R 平方	0.3991
應變平均值	6.53150	調整 R 平方	0.3657
變異係數	6.27233		

參數估計值											
變數	標籤	自由度	參數 估計值	標準 誤差	t 值	Pr > t	第二型 SS	標準化 估計值	平方 偏 相關第二型	95% 信賴界限	
Intercept	Intercept	1	3.45476	0.52631	6.56	<.0001	7.23165	0	.	2.41687	4.49265
LVEDD	左心室大小	1	0.02440	0.00660	3.70	0.0003	2.29510	0.23333	0.06460	0.01139	0.03741
LVEF	左心室功能	1	0.00041905	0.00483	0.09	0.9309	0.00126	0.00521	0.00003804	-0.00910	0.00994
AFD	心房顫動持續時間	1	0.01938	0.00804	2.41	0.0168	0.97524	0.13530	0.02851	0.00353	0.03523
IVS	心室中膈厚度	1	0.03139	0.01378	2.28	0.0238	0.87093	0.14766	0.02554	0.00422	0.05856
Amp	心房收縮量	1	0.18915	0.11788	1.60	0.1102	0.43215	0.09876	0.01284	-0.04331	0.42161
AGE	年齡	1	0.00443	0.00293	1.51	0.1314	0.38510	0.09459	0.01146	-0.00134	0.01021
HTN	高血壓	1	0.01027	0.06010	0.17	0.8645	0.00490	0.00999	0.00014751	-0.10824	0.12879
BW	體重	1	0.01180	0.00336	3.51	0.0005	2.07290	0.26022	0.05872	0.00518	0.01842
aft	心房顫動種類	1	0.44097	0.07190	6.13	<.0001	6.31304	0.37090	0.15964	0.29918	0.58276
nsex	性別	1	0.14563	0.07044	2.07	0.0400	0.71742	0.13174	0.02113	0.00672	0.28453
nMR	二尖瓣逆流程度	1	0.17364	0.07548	2.30	0.0225	0.88829	0.13654	0.02603	0.02480	0.32249

可看到 $P<0.05$ 的自變數有 LVEDD(左心室大小), AFD(心房顫動持續時間), IVS(心室中膈厚度), BW(體重), AFT(心房顫動種類), NMR(二尖瓣逆流程度). 以及 nsex(性別)(原先的一階模型性別的 $P>0.05$).

la_b(左心房大小)開根號後配適之殘差

可看到殘差是常態分配(Shapiro-Wilk 檢定 $P=0.1419 >0.05$)

常態性檢定				
檢定	統計值		p 值	
Shapiro-Wilk	W	0.989777	Pr < W	0.1419
Kolmogorov-Smirnov	D	0.067403	Pr > D	0.0203
Cramer-von Mises	W-Sq	0.112936	Pr > W-Sq	0.0791
Anderson-Darling	A-Sq	0.669476	Pr > A-Sq	0.0831

八、模型選擇

有許多種方法,我們先試電腦自動選擇法(Stepwise, Forward, Backward)

Stepwise selection

逐步選擇 的摘要									
步驟	輸入的變動	移除的變動	標籤	動目 Vars In	偏 R 平方	模型 R 平方	C(p)	F 值	Pr > F
1	aft		心房顫動種類	1	0.1805	0.1805	64.0225	45.82	<.0001
2	LVEDD		左心室大小	2	0.0781	0.2586	40.3024	21.79	<.0001
3	IVS		心室中膈厚度	3	0.0522	0.3108	25.0950	15.61	0.0001
4	nMR		二尖瓣逆流程度	4	0.0179	0.3287	21.2037	5.46	0.0204
5	BW		體重	5	0.0168	0.3454	17.6821	5.22	0.0233
6	nsex		性別	6	0.0180	0.3635	13.7389	5.75	0.0174
7	AFD		心房顫動持續時間	7	0.0192	0.3827	9.3996	6.30	0.0129
8	Amp		心房收縮量	8	0.0088	0.3915	8.5094	2.90	0.0903

Forward selection

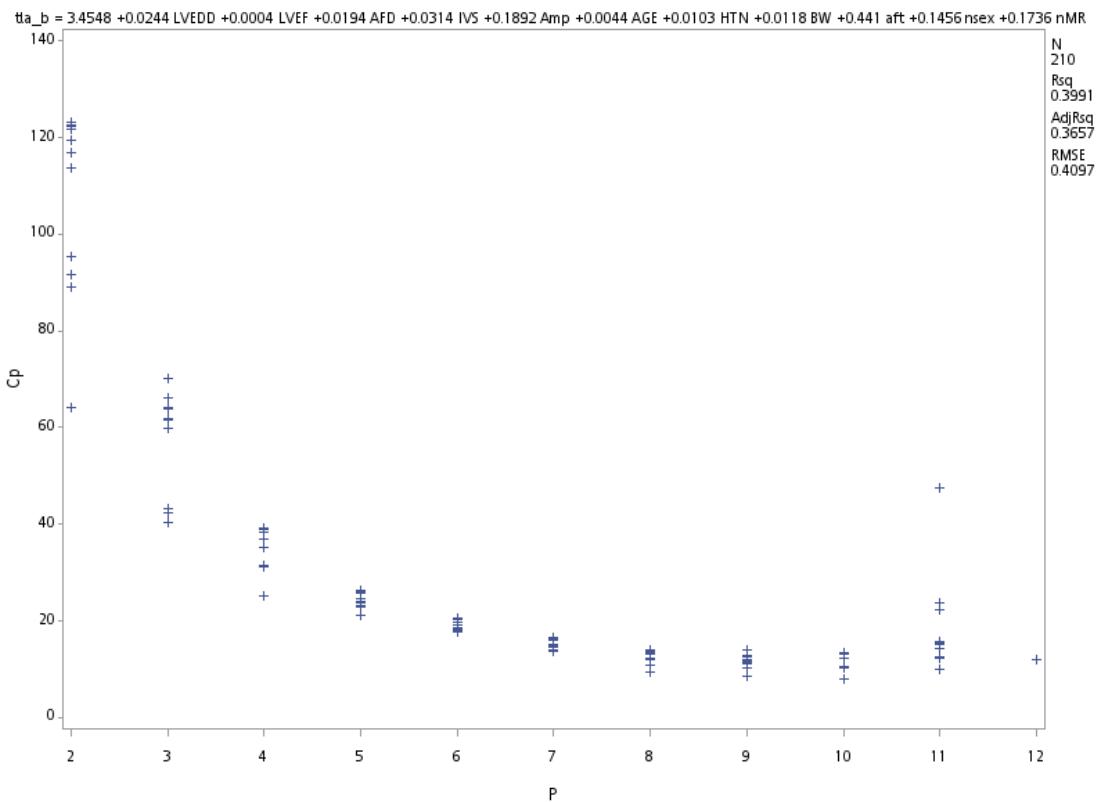
前進選擇 的摘要									
步驟	輸入的變動	標籤	動目 Vars In	偏 R 平方	模型 R 平方	C(p)	F 值	Pr > F	
1	aft	心房顫動種類	1	0.1805	0.1805	64.0225	45.82	<.0001	
2	LVEDD	左心室大小	2	0.0781	0.2586	40.3024	21.79	<.0001	
3	IVS	心室中膈厚度	3	0.0522	0.3108	25.0950	15.61	0.0001	
4	nMR	二尖瓣逆流程度	4	0.0179	0.3287	21.2037	5.46	0.0204	
5	BW	體重	5	0.0168	0.3454	17.6821	5.22	0.0233	
6	nsex	性別	6	0.0180	0.3635	13.7389	5.75	0.0174	
7	AFD	心房顫動持續時間	7	0.0192	0.3827	9.3996	6.30	0.0129	
8	Amp	心房收縮量	8	0.0088	0.3915	8.5094	2.90	0.0903	

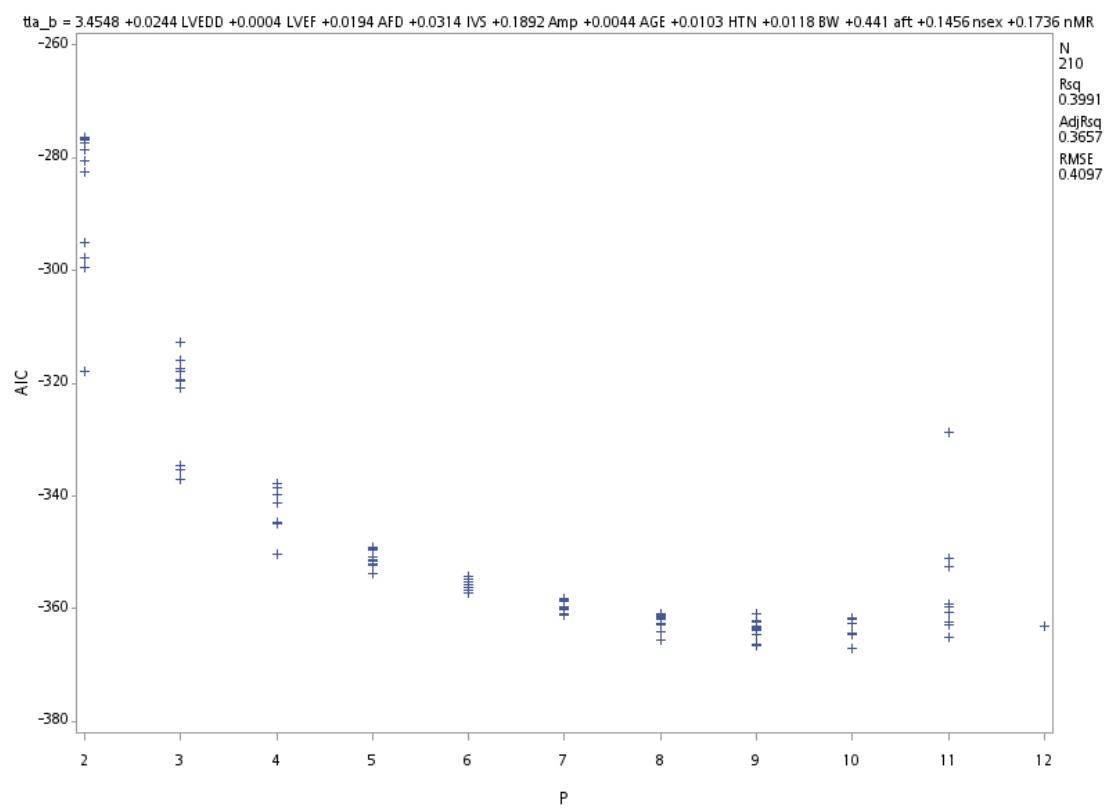
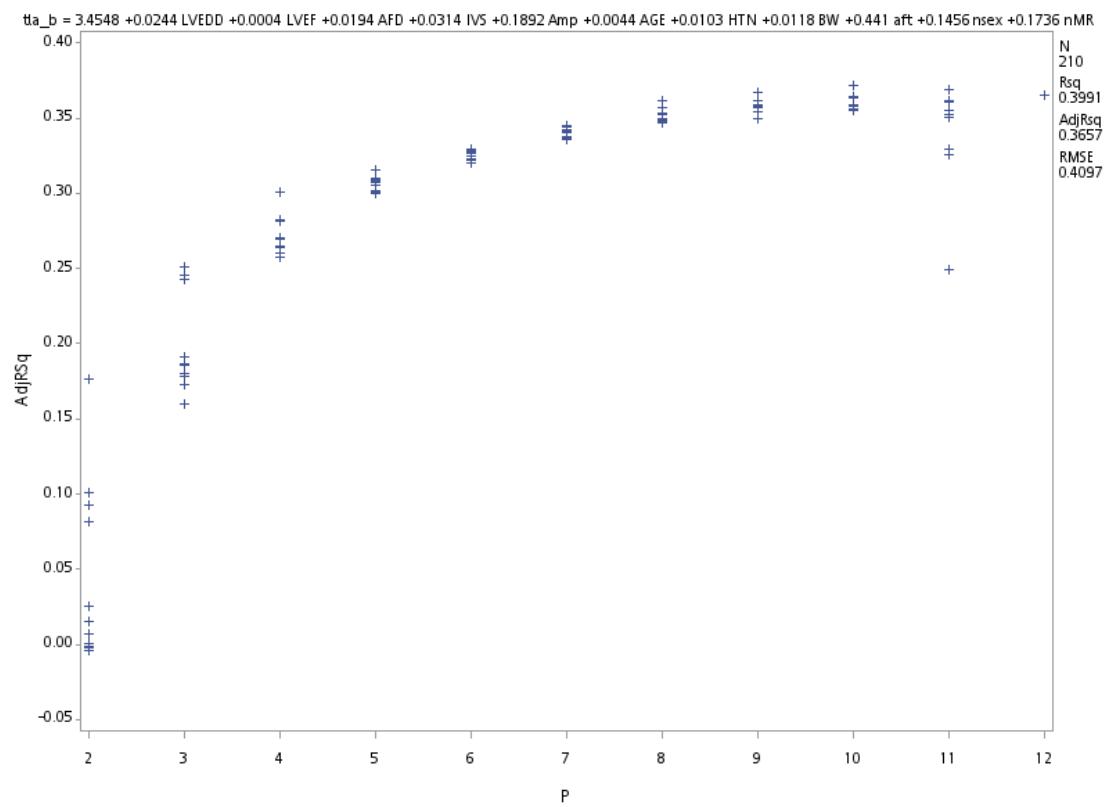
Backward selection

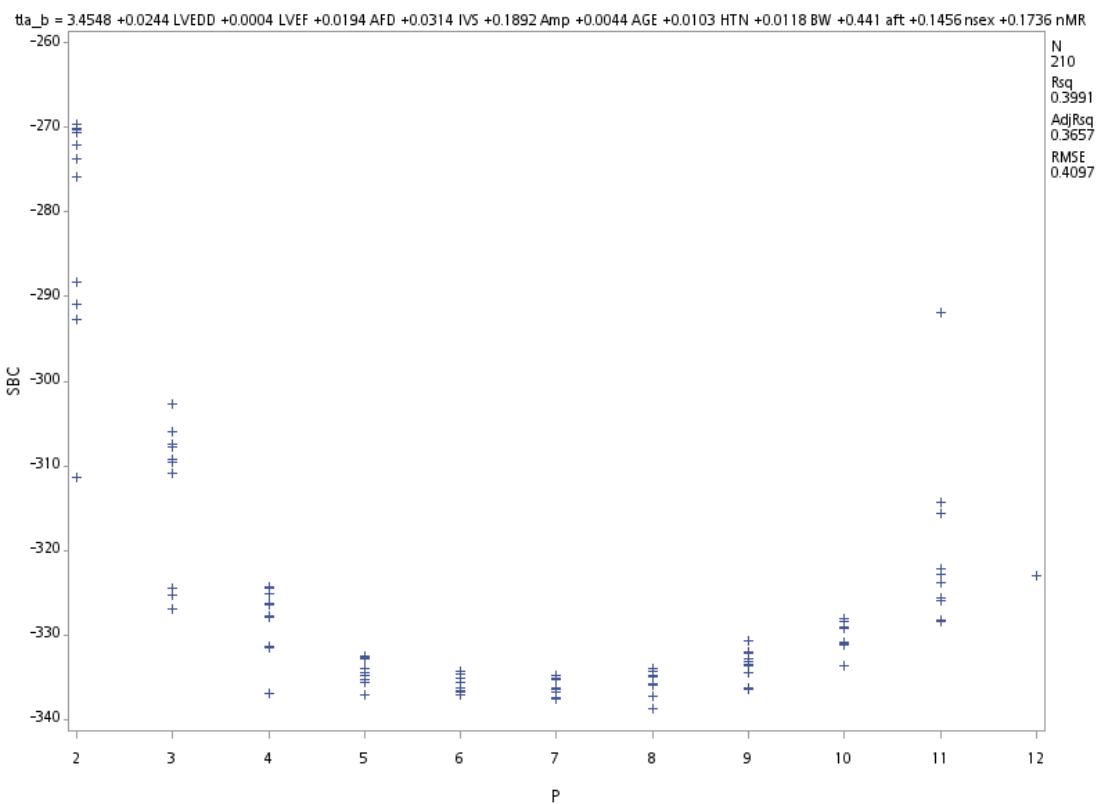
向後消去的摘要								
步驟	移除的變數	標籤	數目 Vars In	偏 R 平方	模型 R 平方	C(p)	F 值	Pr > F
1	LVEF	左心室功能	10	0.0000	0.3991	10.0075	0.01	0.9309
2	HTN	高血壓	9	0.0001	0.3990	8.0374	0.03	0.8626
3	AGE	年齡	8	0.0075	0.3915	8.5094	2.50	0.1157

從以上 3 個表格可知道選出來的都是同樣的 8 個自變數.

再來看以下的數個圖形,可發現電腦自動選擇法選出的組合有最佳的 Cp 值(P=9),而 adjR² 的最高點出現在 P=10 時,AIC 的最低點出現在 P=10 時, SBC 的最低點出現在 P=8 時, 這些圖型選出來的最佳組合似乎並不一致.







先試電腦自動選擇法跑出的模型 (P=9)

變異數分析					
來源	自由度	平方和	平均值 平方	F 值	Pr > F
模型	8	21.64901	2.70613	16.16	<.0001
誤差	201	33.65260	0.16743		
配適不足	201	33.65260	0.16743	.	.
純誤差	0	0	.		
已校正的總計	209	55.30161			

根 MSE	0.40918	R 平方	0.3915
應變平均值	6.53150	調整 R 平方	0.3673
變異係數	6.26467		

參數估計值																
變數	標籤	自由度	參數 估計值		標準 誤差	t 值	Pr > t	第一型 SS	第二型 SS	標準化 估計值	平方 偏 相關第一型		平方 偏 相關第二型		變異數 膨脹	95% 信賴界限
			參數 估計值	標準 誤差							平方 偏 相關第一型	平方 偏 相關第二型	免差			
Intercept	Intercept	1	3.70065	0.35346	10.47	<.0001	8958.69839	18.35224	0	.	.	.	0	3.00367	4.39762	
aft	心房顫動種類	1	0.43647	0.07123	6.13	<.0001	9.98225	6.28607	0.36712	0.18051	0.15739	0.84339	1.18569	0.29601	0.57693	
LVEDD	左心室大小	1	0.02572	0.00621	4.14	<.0001	4.31675	2.86826	0.24601	0.09525	0.07854	0.85698	1.16689	0.01347	0.03798	
IVS	心室中膈厚度	1	0.03704	0.01306	2.84	0.0050	2.88801	1.34653	0.17426	0.07043	0.03847	0.80184	1.24713	0.01129	0.06279	
nMR	二尖瓣逆流程度	1	0.20057	0.07318	2.74	0.0067	0.98877	1.25752	0.15772	0.02594	0.03602	0.91418	1.09388	0.05626	0.34487	
BW	體重	1	0.01028	0.00314	3.27	0.0012	0.92672	1.79557	0.22681	0.02496	0.05065	0.63118	1.58434	0.00409	0.01648	
nsex	性別	1	0.14635	0.07030	2.08	0.0386	0.99747	0.72563	0.13239	0.02756	0.02111	0.74858	1.33587	0.00773	0.28497	
AFD	心房顫動持續時間	1	0.02042	0.00799	2.55	0.0114	1.06396	1.09292	0.14257	0.03022	0.03146	0.97224	1.02855	0.00466	0.03618	
Amp	心房收縮量	1	0.19947	0.11719	1.70	0.0903	0.48508	0.48508	0.10414	0.01421	0.01421	0.80872	1.23652	-0.03161	0.43055	

可看到 Amp(心房收縮量) P=0.0903>0.05, 暗示我們這個參數的 95% 信賴界線可能包含 0, 也就是不具統計顯著性, 由於自變數太多, 不符合模型簡約的原則, 我們先排除 Amp(心房收縮量) 這個變數, 以 P=8 開始進行模型比較, 再根據試跑數次不同模型時產生的參數 P 值是否顯著, 逐一比較數個模型, 希望能找到較適當的模型,

九、模型比較

statistic	N=210	N=122	N=210	N=122	N=210	N=122	N=210	N=122
p	8	8	6	6	5	5	4	4
intercept	3.85	4.06	4.34	4.09	4.45	4.14	4.52	4.06
S{interce pt}	0.34	0.45	0.32	0.39	0.32	0.40	0.32	0.41
心房顫動 種類	0.40	0.40	0.42	0.40	0.43	0.43	0.45	0.47
S{心房顫 動種類}	0.07	0.09	0.07	0.09	0.07	0.09	0.07	0.09
左心室大 小	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.03
S{左心室 大小}	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
心室中膈 厚度	0.04	0.05	0.04	0.05	0.05	0.06	0.05	0.06
S{心室中 膈厚度}	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
二尖瓣逆 流程度	0.20	0.37	0.20	0.37	0.17	0.32		
S{二尖瓣	0.07	0.10	0.07	0.10	0.07	0.10		

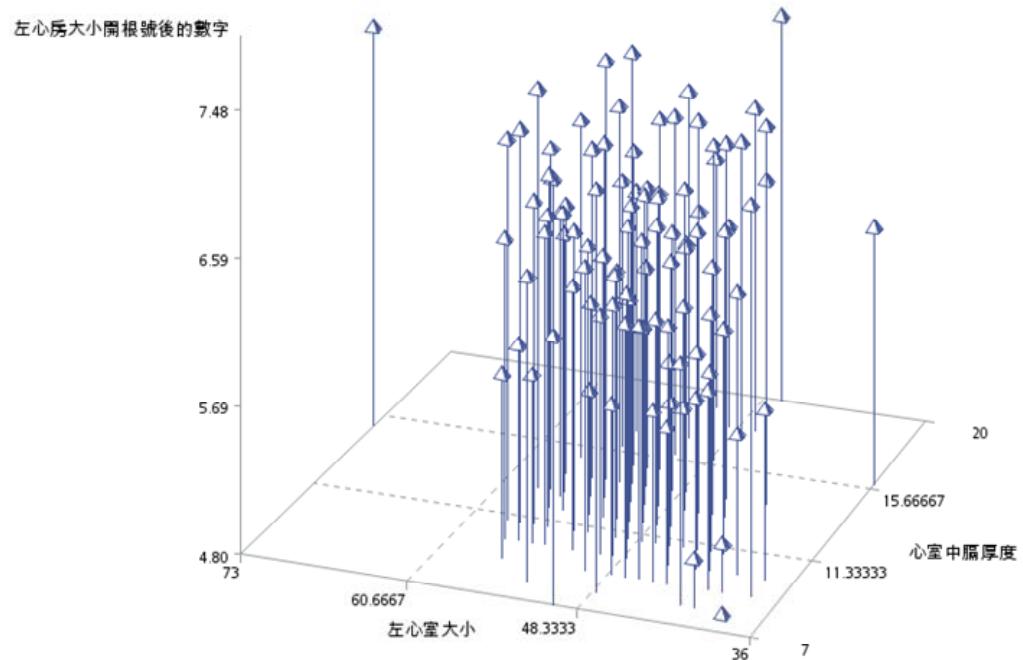
逆流程度)							
體重	0.01	0.01	0.01	0.01			
S{體重}	0.00	0.00	0.00	0.00			
性別	0.18	0.02					
S{性別}	0.07	0.09					
心房顫動持續時間	0.02	0.00					
S{心房顫動持續時間})	0.01	0.01					
SSEp	34.1	16.4	36.2	16.4	37.1	17.3	38.1
Cp	9.4	23.9 (8)	17.7	20.0 (4.1)	21.2	25.0 (8.2)	25.1 (17.0)
MSEp	0.17	0.14	0.18	0.14	0.18	0.15	0.19
MSPR	0.14		0.14		0.15		0.16
adjR2	0.36	0.40	0.33	0.41	0.32	0.39	0.30
PRESSp	37.4	19.0	39.0	18.3	40.0	19.0	40.1
配適不足	.	.	.	0.53	0.03	0.42	0.07
AIC	-366	-229	-357	-233	-354	-228	-350
SBC	-339	-206	-337	-216	-337	-214	-337
							-220

從以上的表格可知，雖然 P 的數目不同，參數估計值及標準誤差並沒有正負號互相矛盾的情況，彼此差異不大，PRESSp的值很接近 SSEp的值，MSPR 的值很接近 MSEp的值，表示這四個候選模型的預測能力可以被肯定，雖然 AIC, SBC 可能不是最低的，但彼此差異不大，adjR² 彼此也差異不大，配適不足在 P 較多時跑不出來，在 P=4 時 P 值>0.05 表示 P=4 時沒有配適不足的情況，只有 Cp 值產生較不一致的狀態，除了第一個模型，其他的模型 Cp 值都偏高，但奇怪的是用不同程式跑出來的 Cp 值有的一樣有的不一樣，黑色字體是用 supportsas 網頁中可跑出 press 值的程式跑的，紅字是用無法跑出 PRESS 值的 SAS 程式跑的。

綜合以上所述，在主客觀的考慮之下，我們決定選擇 P=4 的模型，自變數較少較簡約，且各個層面也和較多 P 的模型類似。

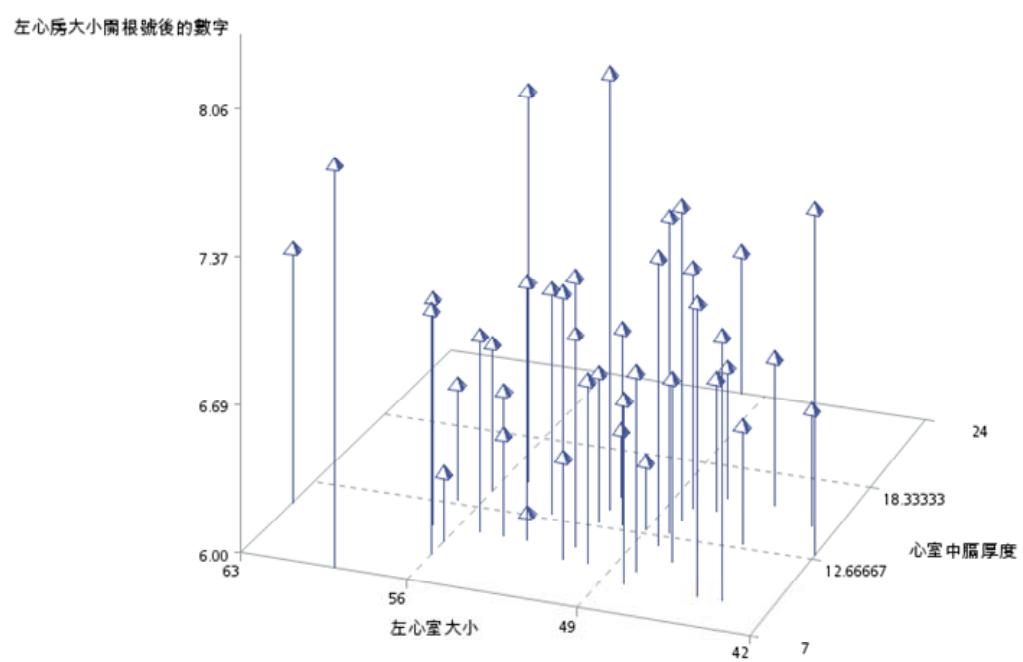
十、依心房顫動種類畫出散佈圖

心房顫動種類:偶發性



偶發性的散佈圖看起來像是一個平面，慢性的散佈圖看起來像彎曲的平面。

心房顫動種類:慢性



十一、選定之一階模型(P=4)

我們得到的迴歸方程式是

左心房大小開根號後之數值= **4.52452+0.45330*心房顫動種類(偶發性 0 或慢性 1)**
+0.02622*左心室大小+ 0.04912*心室中膈厚度

意即，如果心房顫動種類為**偶發性**，則方程式為

左心房大小開根號後之數值= **4.52452+0.02622*左心室大小+ 0.04912*心室中膈厚度**

如果心房顫動種類為**慢性**，則方程式為

左心房大小開根號後之數值= **4.97782+0.02622*左心室大小+ 0.04912*心室中膈厚度**

是截距不同的兩個方程式。從方程式可看出左心室越大，心室中膈越厚，左心房越大。慢性心房顫動的病人左心房較偶發性心房顫動的病人大。

變異數分析					
來源	自由度	平方和	平均值 平方	F 值	Pr > F
模型	3	17.18701	5.72900	30.96	<.0001
誤差	206	38.11459	0.18502		
配適不足	137	27.85142	0.20330	1.37	0.0746
純誤差	69	10.26318	0.14874		
已校正的總計	209	55.30161			

根 MSE	0.43014	R 平方	0.3108
應變平均值	6.53150	調整 R 平方	0.3007
變異係數	6.58566		

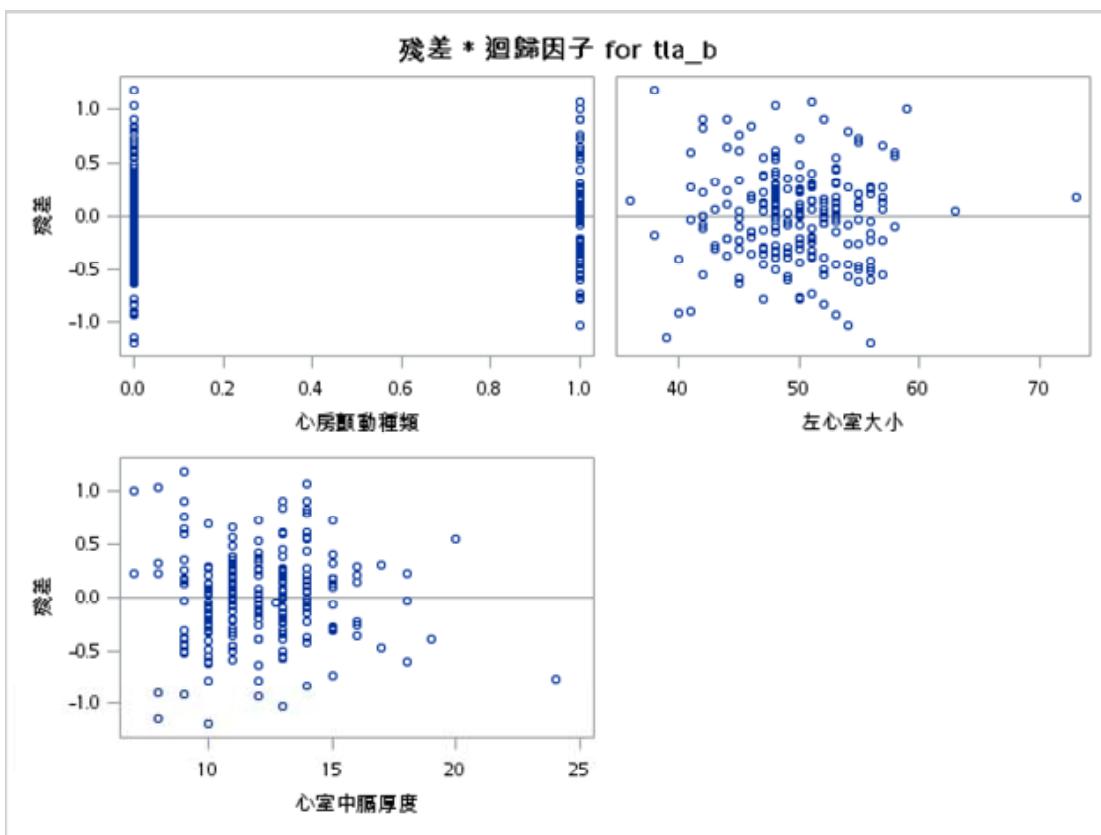
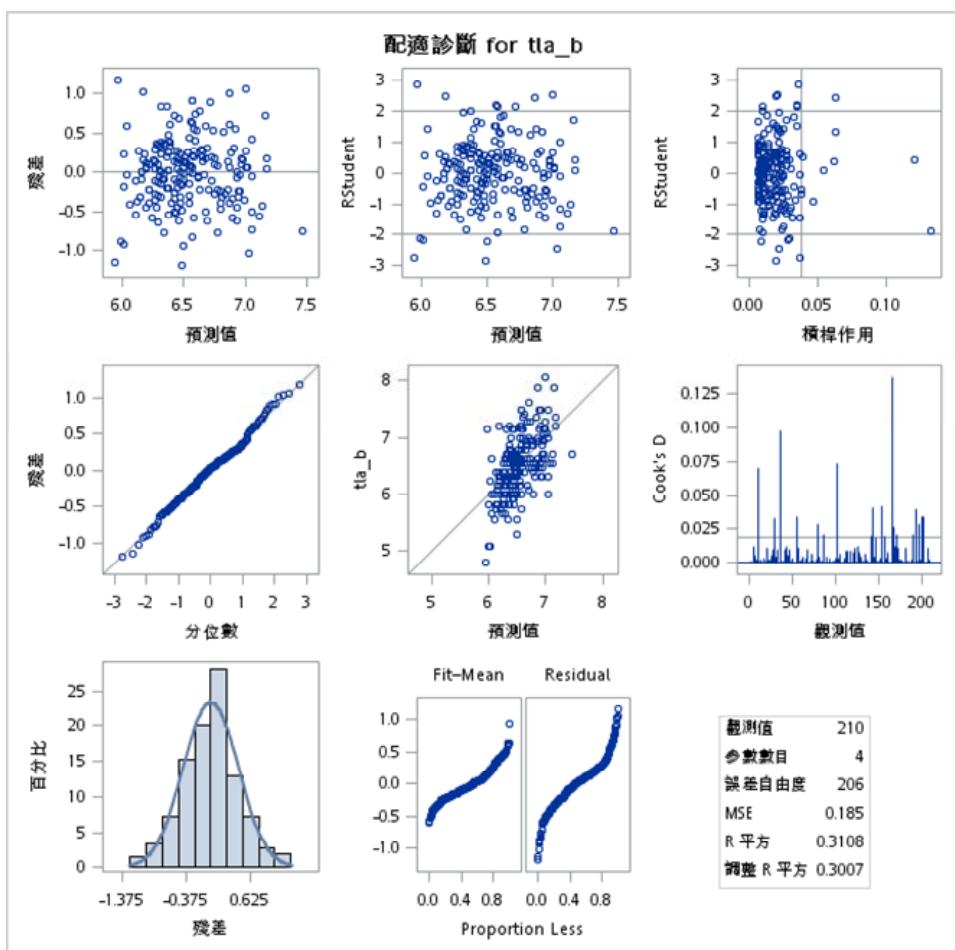
參數估計值																
變數	標籤	自由度	參數 估計值	標準 誤差	t 值	Pr > t	第一型 SS	第二型 SS	標準化 估計值	相關第一型	平方 偏	相關第二型	平方 偏	變異數 方差	變異數 膨脹	95% 信賴界限
Intercept	Intercept	1	4.52452	0.32125	14.08	<.0001	8958.69839	36.70151	0	0	3.89116	5.15788
aft	心房顫動種類	1	0.45330	0.06929	6.54	<.0001	9.98225	7.91931	0.38128	0.18051	0.17203	0.98507	1.01516	0.31670	0.58991	
LVEDD	左心室大小	1	0.02622	0.00614	4.27	<.0001	4.31675	3.37729	0.25073	0.09525	0.08140	0.97144	1.02940	0.01412	0.03831	
IVS	心室中隔厚度	1	0.04912	0.01243	3.95	0.0001	2.88801	2.88801	0.23109	0.07043	0.07043	0.97791	1.02259	0.02461	0.07363	

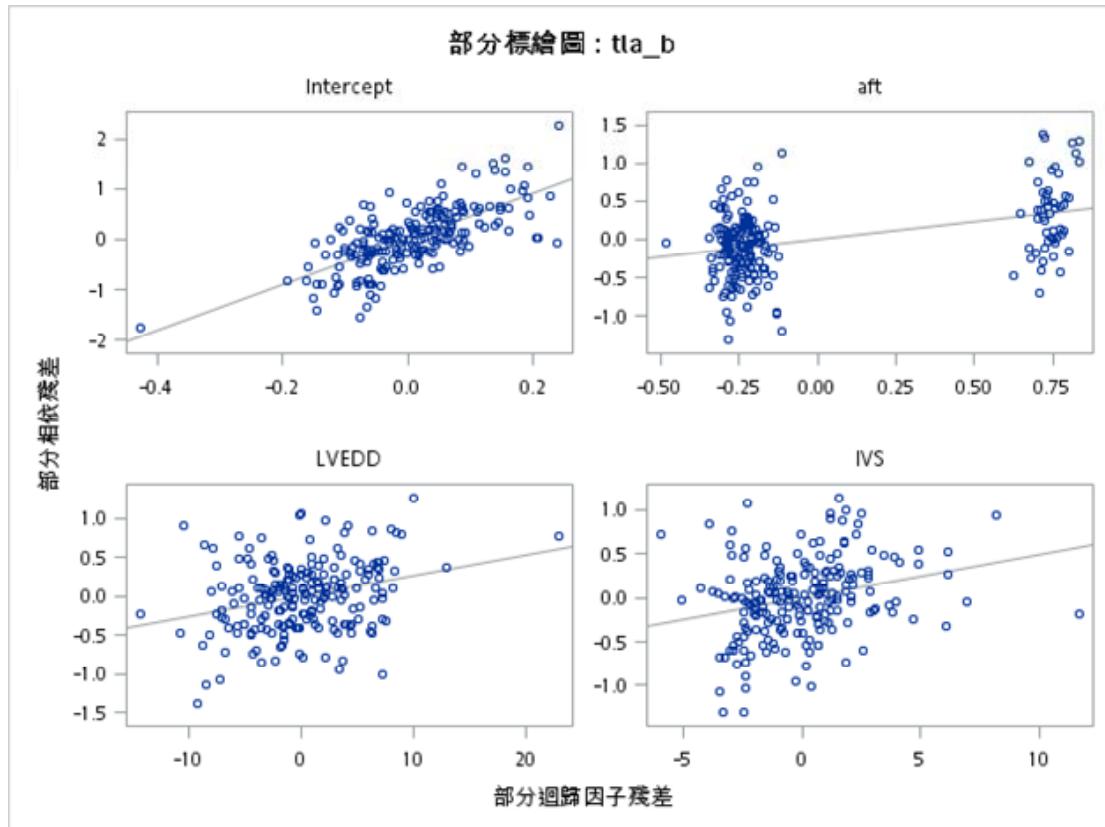
共變性診斷						
變目	特徵值	條件索引	變異的比例			
			Intercept	aft	LVEDD	IVS
1	3.29698	1.00000	0.00075740	0.02805	0.00083578	0.00330
2	0.67259	2.21402	0.00056084	0.96438	0.00055963	0.00227
3	0.02568	11.33081	0.04409	0.00006910	0.07337	0.97503
4	0.00475	26.35627	0.95459	0.00750	0.92524	0.01940

估計值的共變異數

- $S^2\{aft\}=0.0048008274$, $S\{aft\}=0.069$
- $S^2\{lvedd\}=0.0000376497$, $S\{lvedd\}=0.0061$
- $S^2\{ivs\}=0.0001545767$, $S\{ivs\}=0.012$

估計值的共變異數						
變數	標籤	Intercept	aft	LVEDD	IVS	
Intercept	Intercept	0.1032013699	0.0015071024	-0.001737723	-0.001360816	
aft	心房顫動種類	0.0015071024	0.0048008274	-0.000042276	-0.000049453	
LVEDD	左心室大小	-0.001737723	-0.000042276	0.0000376497	-9.970864E-6	
IVS	心室中隔厚度	-0.001360816	-0.000049453	-9.970864E-6	0.0001545767	





皮爾森相關係數

Pearson 相關係數, N = 210				
Prob > r (位於 H0 底下): Rho=0				
	tla_b	aft	LVEDD	IVS
tla_b	1.00000	0.42486 <.0001	0.32366 <.0001	0.29273 <.0001
aft	0.42486	1.00000	0.10804	0.07137
心房顫動種類	<.0001		0.1185	0.3033
LVEDD	0.32366	0.10804	1.00000	0.13732
左心室大小	<.0001	0.1185		0.0469
IVS	0.29273	0.07137	0.13732	1.00000
心室中隔厚度	<.0001	0.3033	0.0469	

可看到這三個自變數的 P 值都 < 0.0001 , 表示它們與 y 值有相關.

殘差

殘差屬於常態分配

常態性檢定				
檢定	統計值		p 值	
Shapiro-Wilk	W	0.992882	Pr < W	0.4065
Kolmogorov-Smirnov	D	0.058236	Pr > D	0.0817
Cramer-von Mises	W-Sq	0.088489	Pr > W-Sq	0.1635
Anderson-Darling	A-Sq	0.533812	Pr > A-Sq	0.1778

殘差符合恆常性

Brown-Forsythe test

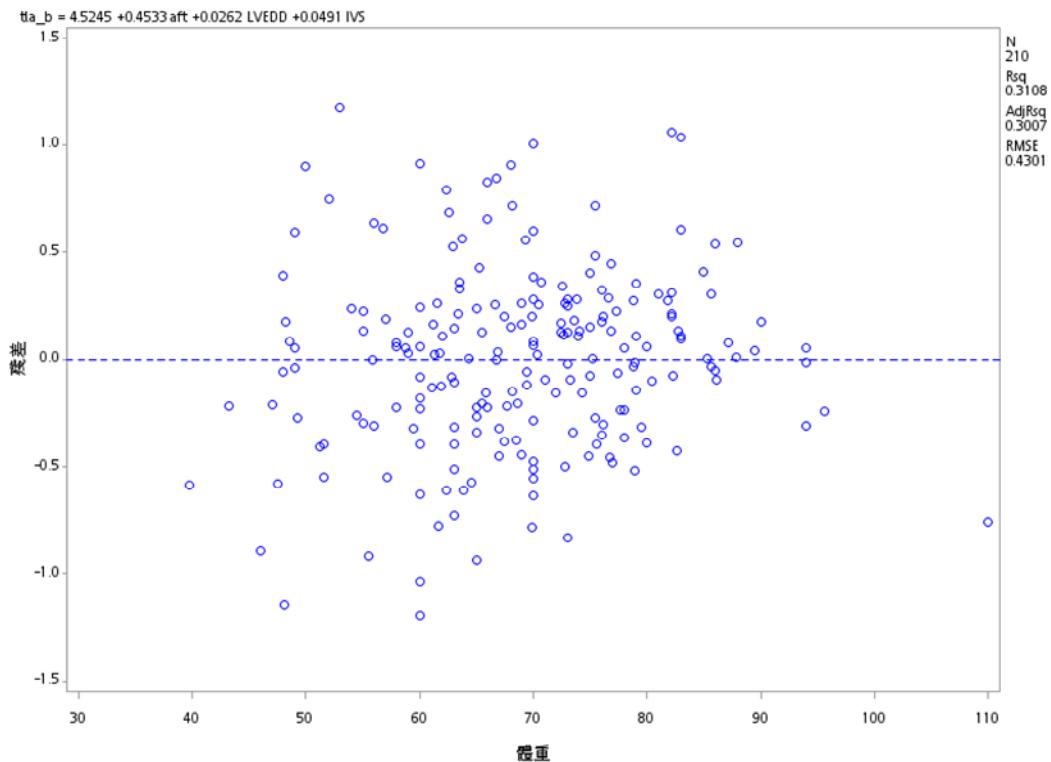
方法	變異數	自由度	t 值	Pr > t
集區	均等	208	-1.47	0.1434
Satterthwaite	不均等	82.253	-1.42	0.1598

變異數相等性				
方法	分子自由度	分母自由度	F 值	Pr > F
Folded F	51	157	1.15	0.5173

十二、追加變數圖

原本我們以為體重應該是個重要的變數，現在探討體重要不要加入模型中。

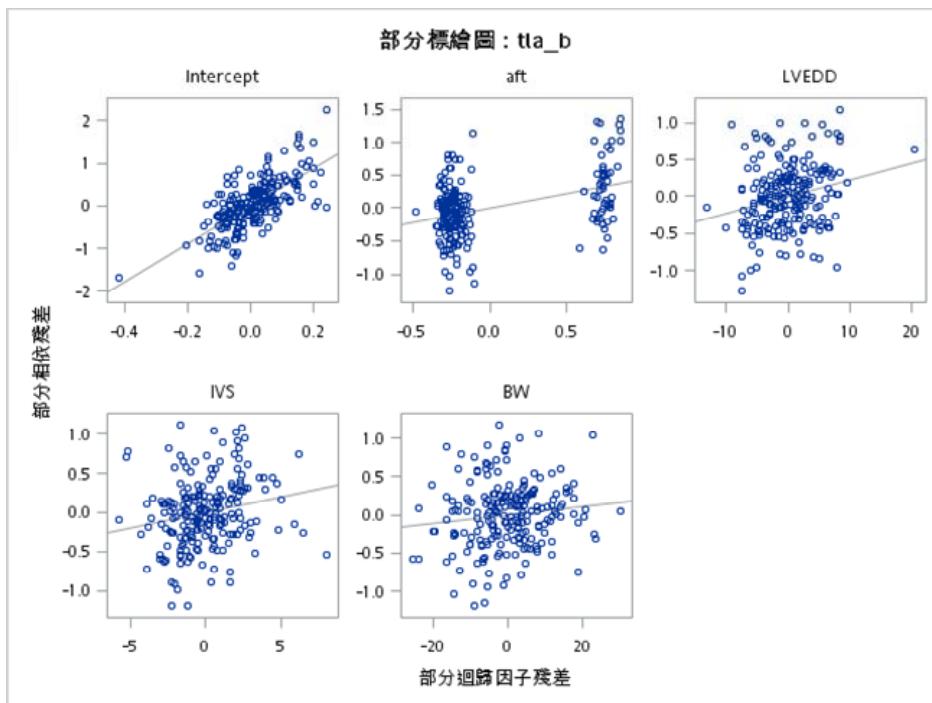
下圖為當變數 aft(心房顫動種類) lvedd(左心室大小) ivs(心室中膈厚度) 已存在模型中，對新預測變數體重的殘差圖，圖型顯示殘差均勻散佈，沒有特定形狀，表示體重不用納入。



從體重的追加變數圖可看出代表迴歸係數估計值的直線斜率呈現出些微的正值，但體重對於配適值的邊際關係是較薄弱的，對於模型中包含變數 aft(心房顫動種類) lvedd(左心室大小) ivs(心室中膈厚度) bw(體重)是否該剔除 bw(體重)之正式 t 檢定，其 P 值等於 0.0796，顯示出支持剔除 bw(體重)的訊息。

參數估計值						
變數	標籤	自由度	參數 估計值	標準 誤差	t 值	Pr > t
Intercept	Intercept	1	4.44994	0.32241	13.80	<.0001
aft	心房顫動種類	1	0.44670	0.06904	6.47	<.0001
LVEDD	左心室大小	1	0.02264	0.00643	3.52	0.0005
IVS	心室中膈厚度	1	0.03957	0.01351	2.93	0.0038
BW	體重	1	0.00536	0.00305	1.76	0.0796

應變數 tla_b 的檢定 1 結果				
來源	自由度	平均值 平方	F 值	Pr > F
分子	1	0.56834	3.10	0.0796
分母	205	0.18315		



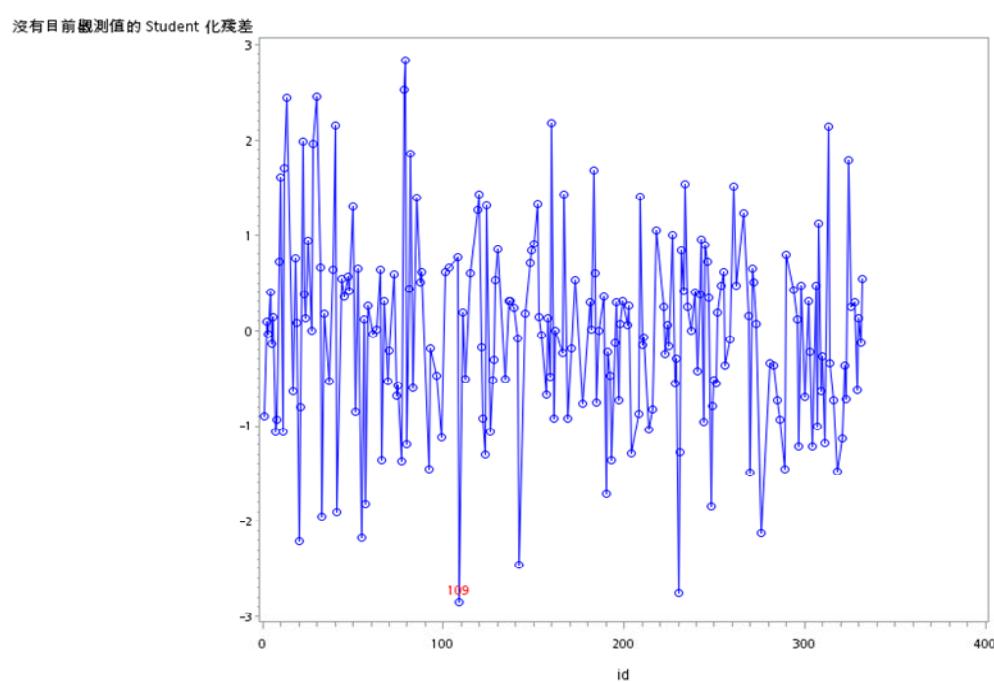
結論是體重不需加入模型中.

十三、迴歸分析診斷

1. 離群 Y 觀測值之確認：利用 t 化去點殘差

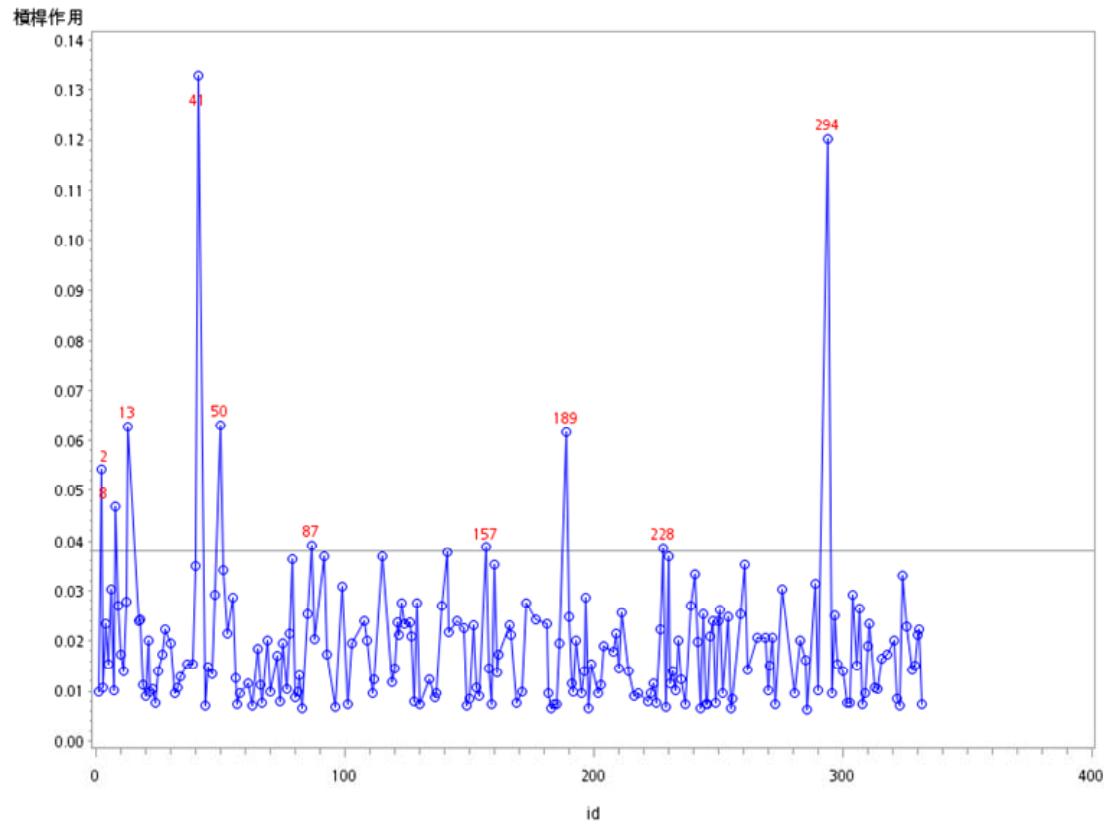
t 化去點殘差絕對值最大的是第 109 個個案 = 2.848($n=193$, $id=109$)

在全族顯著水準 $\alpha=0.1$ 下，採用 Bonferroni 程序進行檢定，我們需要有 $t(1-\alpha/2n; n-p-1)=t(1-0.1/(2*210); 210-4-1)=t(0.999762; 205)$ ，約 3.55。(從 excel 查表可得 $TINV(0.000476; 205)=3.55$)，從正式檢定程序看來第 109 個個案未達顯著水準。沒有離群 Y 觀測值



2. 離群 X 觀測值之確認：利用帽子矩陣槓桿值
 $\text{hat} > (2 * 4 / 210) (\approx 0.038)$, 共有離群 X 觀測值資料十筆.

Obs	id	n	residual	hat	rstudent	dfits	dp	d	yhat
23	87	23	0.21486	0.03917	0.50866	0.10270	0.000014	0.00265	6.64080
36	13	36	1.00564	0.06286	2.44405	0.63297	0.016932	0.09780	6.86836
37	2	37	0.04140	0.05419	0.09872	0.02363	0.000000	0.00014	7.16971
42	8	42	-0.39166	0.04674	-0.93229	-0.20644	0.000226	0.01066	6.79478
122	294	122	0.17344	0.12025	0.42903	0.15861	0.000080	0.00631	7.17503
145	157	145	-0.28444	0.03870	-0.67356	-0.13514	0.000042	0.00458	6.84188
165	228	165	-0.23523	0.03860	-0.55680	-0.11157	0.000020	0.00312	6.94343
166	41	166	-0.75927	0.13271	-1.90749	-0.74616	0.031716	0.13743	7.46747
196	50	196	0.54107	0.06311	1.30175	0.33785	0.001572	0.02844	6.73904
209	189	209	0.14893	0.06171	0.35669	0.09148	0.000009	0.00210	6.25419

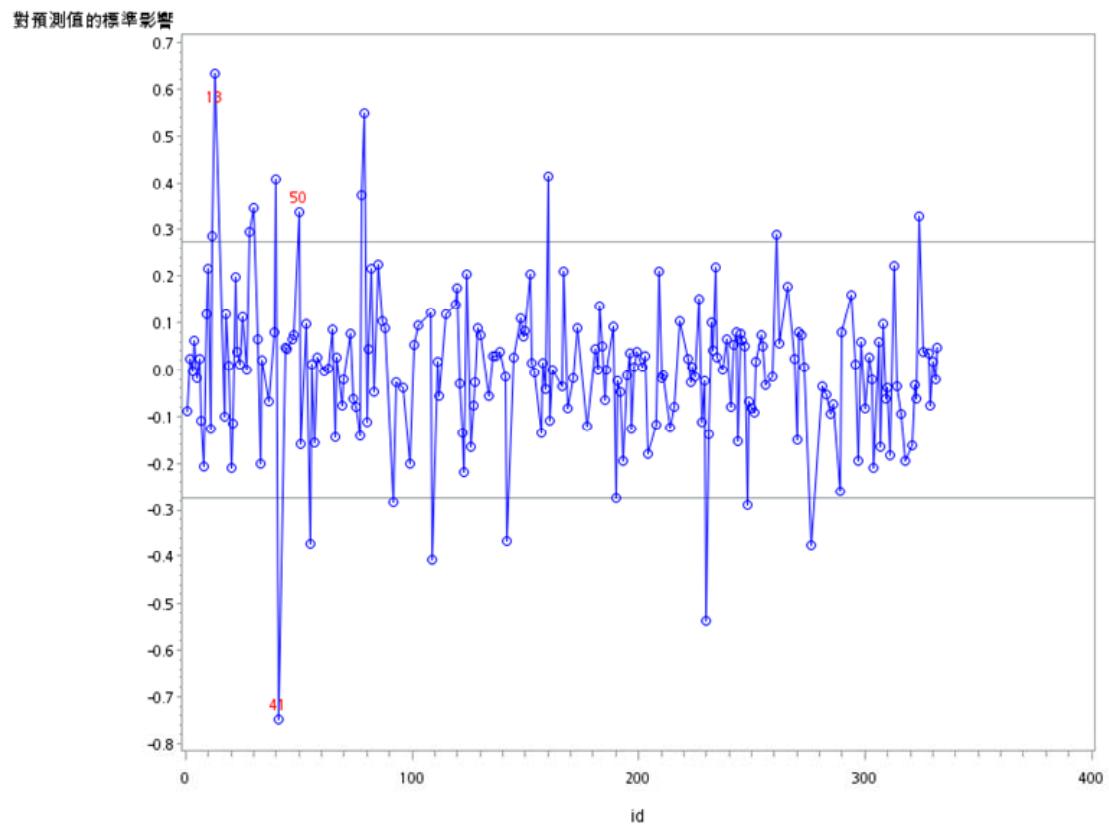


3. 辨識影響個案: 對單一配適值之影響 : *DFFITS*

$$|DFFITS| > 2 * \sqrt{4/210} (\approx 0.276)$$

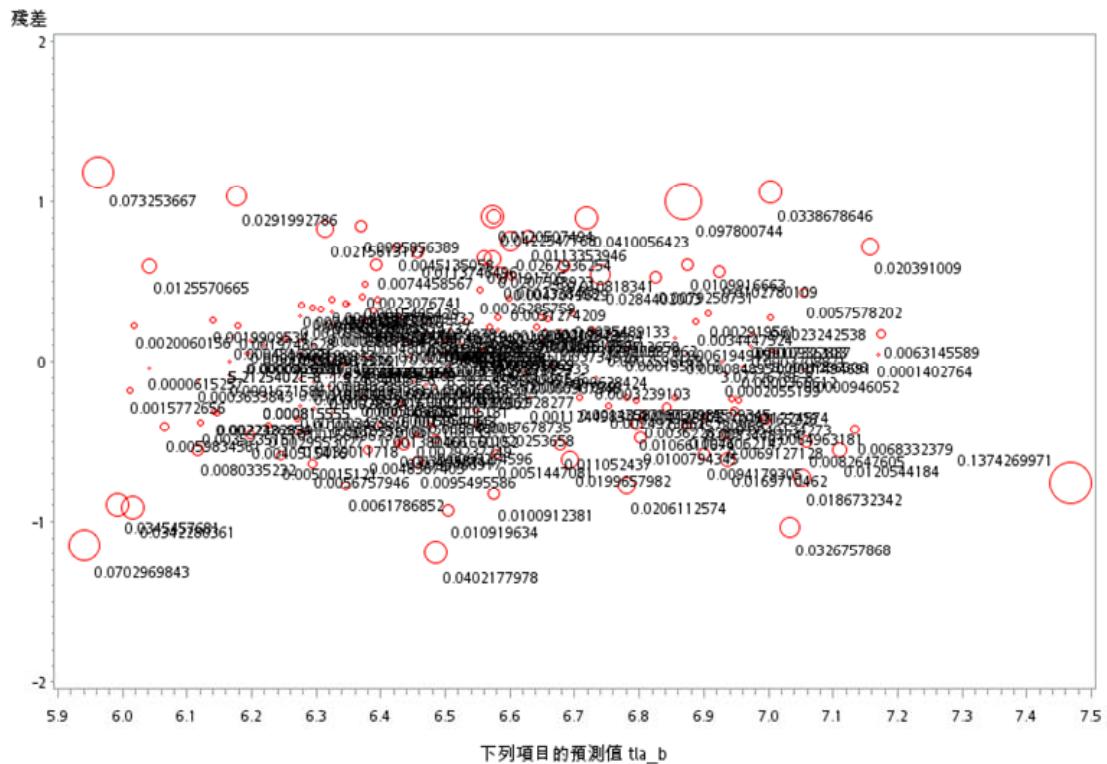
離群 X 觀測值中, 共三筆資料有影響

Obs	id	n	residual	hat	rstudent	dffits	dp	d	yhat
36	13	36	1.00564	0.06286	2.44405	0.63297	0.016932	0.09780	6.86836
166	41	166	-0.75927	0.13271	-1.90749	-0.74616	0.031716	0.13743	7.46747
196	50	196	0.54107	0.06311	1.30175	0.33785	0.001572	0.02844	6.73904



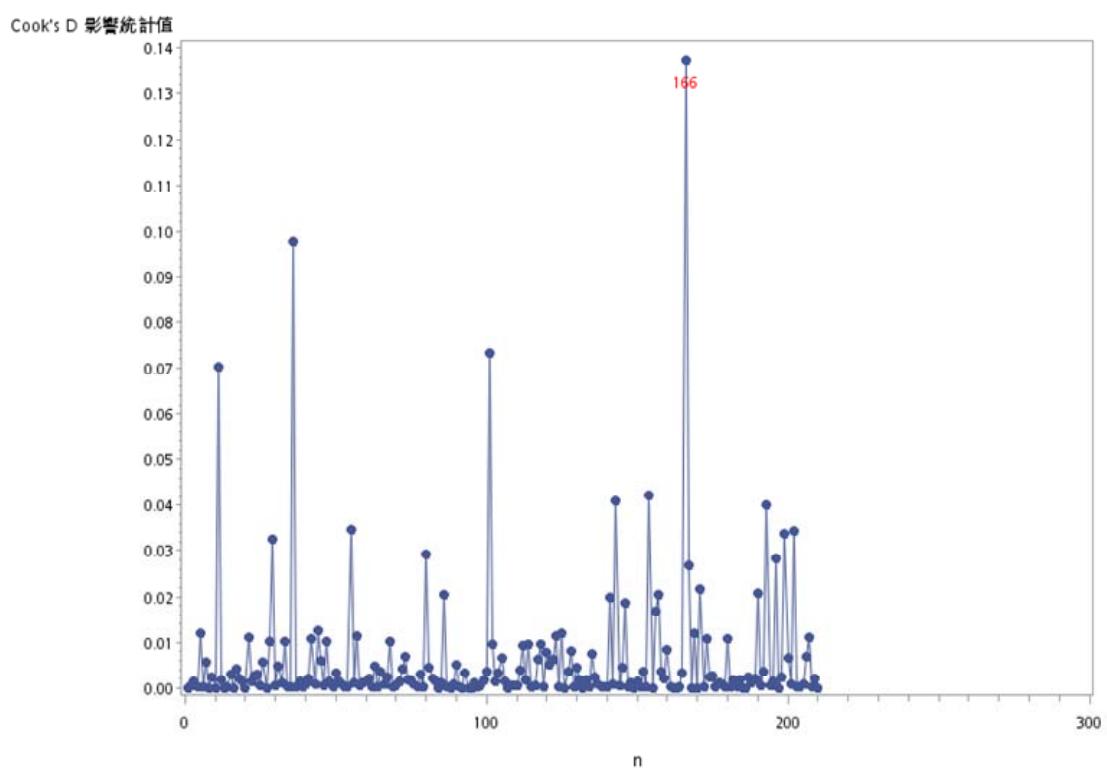
4. 辨識影響個案:對所有配適值之影響—Cook 距離

proportional influence plot



最大 cookd 是 N=166 (id=41), 有 0.137426997, 相對於 F 分配的位置在第 3.17 個百分位數值，對於配適值的影響不大

index influence plot



5. 對迴歸係數之影響：*DFBETAS* 量數

對 DFB_intercept 有影響的個案

Obs	id	DFFITS	HatDiagonal	cooksd	DFB_Intercept	DFB_aft	DFB_LVEDD	DFB_ivs
122	294	0.1586	0.1202	0.00631	-0.1459	-0.0356	0.1492	0.0197
166	41	-0.7462	0.1327	0.13743	0.2199	-0.2067	0.1062	-0.6898

對 DFB_aft 有影響的個案

Obs	id	DFFITS	HatDiagonal	cooksd	DFB_Intercept	DFB_aft	DFB_LVEDD	DFB_ivs
36	13	0.6330	0.0629	0.09780	-0.1342	0.2937	0.3554	-0.4327
166	41	-0.7462	0.1327	0.13743	0.2199	-0.2067	0.1062	-0.6898

對 DFB_lvedd 有影響的個案

Obs	id	DFFITS	HatDiagonal	cooksd	DFB_Intercept	DFB_aft	DFB_LVEDD	DFB_ivs
36	13	0.6330	0.0629	0.097801	-0.1342	0.2937	0.3554	-0.4327
122	294	0.1586	0.1202	0.006315	-0.1459	-0.0356	0.1492	0.0197

對 DFB_ivs 有影響的個案

Obs	id	DFFITS	HatDiagonal	cooksd	DFB_Intercept	DFB_aft	DFB_LVEDD	DFB_ivs
36	13	0.6330	0.0629	0.09780	-0.1342	0.2937	0.3554	-0.4327
42	8	-0.2064	0.0467	0.01066	0.0779	0.0509	0.0025	-0.1906
166	41	-0.7462	0.1327	0.13743	0.2199	-0.2067	0.1062	-0.6898
196	50	0.3379	0.0631	0.02844	-0.0554	-0.0663	-0.0854	0.3174

id=8,13,41,50,294 共五個個案有影響

6. 對推論的影響

原配適迴歸函數:

$$4.52452 + 0.45330 * \text{aft} + 0.02622 * \text{lvedd} + 0.04912 * \text{ivs}$$

排除五個影響個案後的配適迴歸函數:

$$4.52975 + 0.44927 * \text{aft} + 0.02265 * \text{lvedd} + 0.06328 * \text{ivs}$$

兩者差異不大.

直接檢驗這五個個案在預測變數範圍內對於配適迴歸函數推論的影響,先計算所有函數的配適值,再計算排除這五個個案後的配適值,計算兩個配適值間絕對值的平均差異,得到平均差異為 0.4%.而且所有的差異值都小於 5%,在此直接證據下可以知道這五個個案對於推論的影響很小,不需要採取矯正.

分析變動 : df				
N	平均值	標準差	最小值	最大值
210	0.0041327	0.0035497	7.4834065E-6	0.0217919

dfyesno	次數	百分比	累計 次數	累計 百分比
1	210	100.00	210	100.00

十四、多重共線性診斷

變異數膨脹因子: VIF 與平均值都在 1 左右,應該沒有嚴重的多重共線性問題.

參數估計值								
變動	標籤	自由度	參數 估計值	標準 誤差	t 值	Pr > t	標準化 估計值	變異數 膨脹
Intercept	Intercept	1	4.52452	0.32125	14.08	<.0001	0	0
aft	心房顫動種類	1	0.45330	0.06929	6.54	<.0001	0.38128	1.01516
LVEDD	左心室大小	1	0.02622	0.00614	4.27	<.0001	0.25073	1.02940
IVS	心室中隔厚度	1	0.04912	0.01243	3.95	0.0001	0.23109	1.02259

分析變動 : VarianceInflation 變異數 膨脹				
N	平均值	標準差	最小值	最大值
3	1.02223804	0.0071220	1.0151569	1.0293964

十五、交互作用項

- il=lvedd* ivs;
- al=aft*lvedd;
- ai=aft*ivs;
- model tla_b=aft lvedd ivs ai

自變數兩兩乘積檢驗有無交互作用項，我們發現 ai=aft*ivs(心房顫動種類*心室中隔厚度)是存在的，將新的方程式重新跑一遍可得以下結果：

變異數分析						
來源	自由度	平方和	平均值 平方	F 值	Pr > F	
模型	4	18.64414	4.66103	26.07	<.0001	
誤差	205	36.65747	0.17882			
配適不足	136	26.39429	0.19408	1.30	0.1099	
純誤差	69	10.26318	0.14874			
已校正的總計	209	55.30161				

根 MSE	0.42287	R 平方	0.3371
應變平均值	6.53150	調整 R 平方	0.3242
變異係數	6.47428		

參數估計值														
變動	標籤	自由度	參數 估計值	標準 誤差	t 值	Pr > t	第一型 SS	第二型 SS	標準化 估計值	平方 偏	平方 偏	變異數 膨脹	95% 信賴界限	
Intercept	Intercept	1	4.37839	0.31994	13.69	<.0001	8958.69839	33.48906	0	.	.	0	3.74759	5.00918
aft	心房顫動種類	1	1.38999	0.33513	4.15	<.0001	9.98225	3.07616	1.16914	0.18051	0.07742	24.57300	0.72925	2.05074
LVEDD	左心室大小	1	0.02334	0.00612	3.82	0.0002	4.31675	2.60391	0.22321	0.09525	0.06632	1.05814	0.01128	0.03540
IVS	心室中隔厚度	1	0.07316	0.01484	4.93	<.0001	2.88801	4.34424	0.34416	0.07043	0.10595	1.50784	0.04389	0.10242
ai		1	-0.07612	0.02667	-2.85	0.0048	1.45713	1.45713	-0.81718	0.03823	0.03823	25.34419	-0.12870	-0.02355

可以發現調整 R 平方變高了，但是多重共線性的問題也跑出來(變異數膨脹因子變高了). 跟原先估計值的共變異數相比，迴歸係數估計的變異性擴大了.

- $S^2\{aft\}=0.1123118744, S\{aft\}=0.335$
- $S^2\{ivs\}=0.0002202848, S\{ivs\}=0.0148$

估計值的共變異數							
變動	標籤	Intercept	aft	LVEDD	IVS	ai	
Intercept	Intercept	0.1023608502	-0.015341577	-0.001627842	-0.001746208	0.0013651449	
aft	心房顫動種類	-0.015341577	0.1123118744	-0.000371604	0.0027150182	-0.008750256	
LVEDD	左心室大小	-0.001627842	-0.000371604	0.000037403	-0.000018123	0.0000268789	
IVS	心室中隔厚度	-0.001746208	0.0027150182	-0.000018123	0.0002202848	-0.000224527	
ai		0.0013651449	-0.008750256	0.0000268789	-0.000224527	0.0007111127	

於是我們將變數置中再跑一遍

參數估計值														
變動	自由度	參數 估計值	標準 誤差	t 值	Pr > t	第一型 SS	第二型 SS	標準化 估計值	相關第一型	平方 偏 倚	相關第二型	平方 偏 倚	變異數 膨脹	95% 信賴界限
Intercept	1	0.00566	0.02925	0.19	0.8467	0	0.00670	0	0	-0.05200 0.06333
cx1	1	0.47011	0.06837	6.88	<.0001	9.98225	8.45415	0.39541	0.18051	0.18741	0.97777	1.02274	0.33531	0.60491
cx2	1	0.02334	0.00612	3.82	0.0002	4.31675	2.60391	0.22321	0.09525	0.06632	0.94506	1.05814	0.01128	0.03540
cx3	1	0.05431	0.01236	4.39	<.0001	2.88801	3.45368	0.25549	0.07043	0.08610	0.95678	1.04518	0.02994	0.07867
cx1x3	1	-0.07612	0.02667	-2.85	0.0048	1.45713	1.45713	-0.16638	0.03823	0.03823	0.95184	1.05059	-0.12870	-0.02355

降低多重共線性後，迴歸係數估計的變異性與原來相似

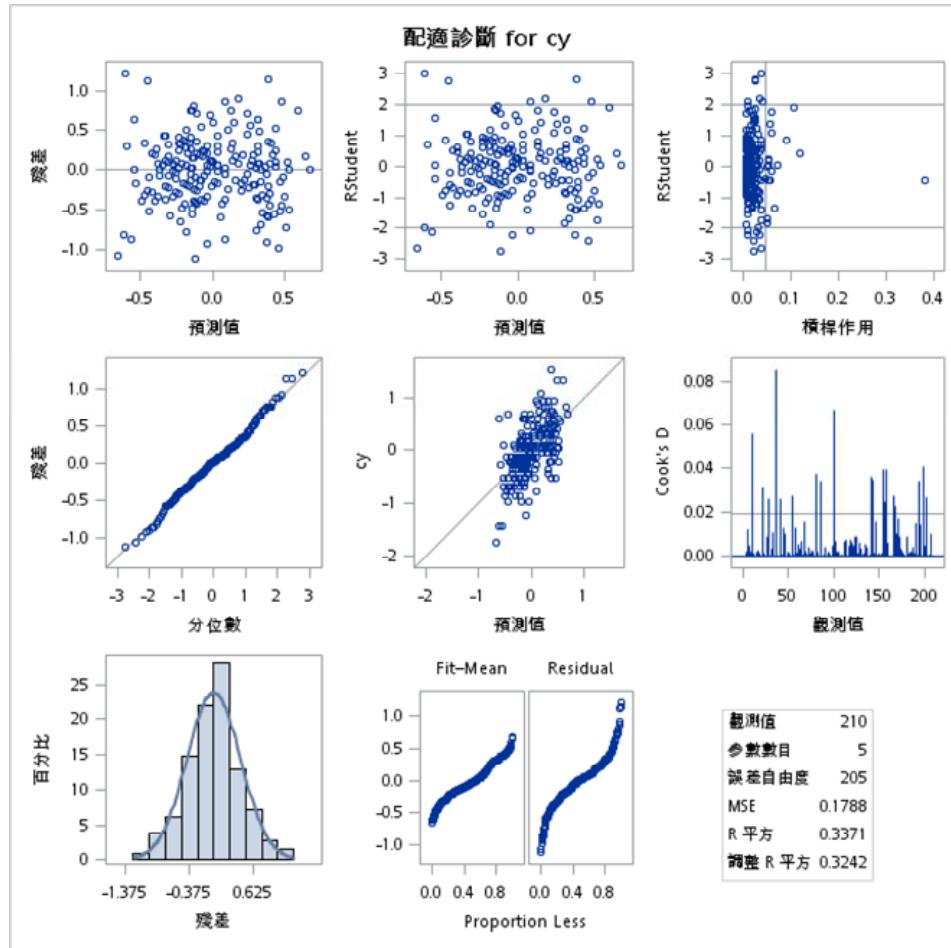
- $S^2\{aft\}=0.0046744619, S\{aft\}=0.068$
- $S^2\{lvedd\}=0.000037403, S\{lvedd\}=0.0061$
- $S^2\{ivs\}=0.0001526923, S\{ivs\}=0.012$

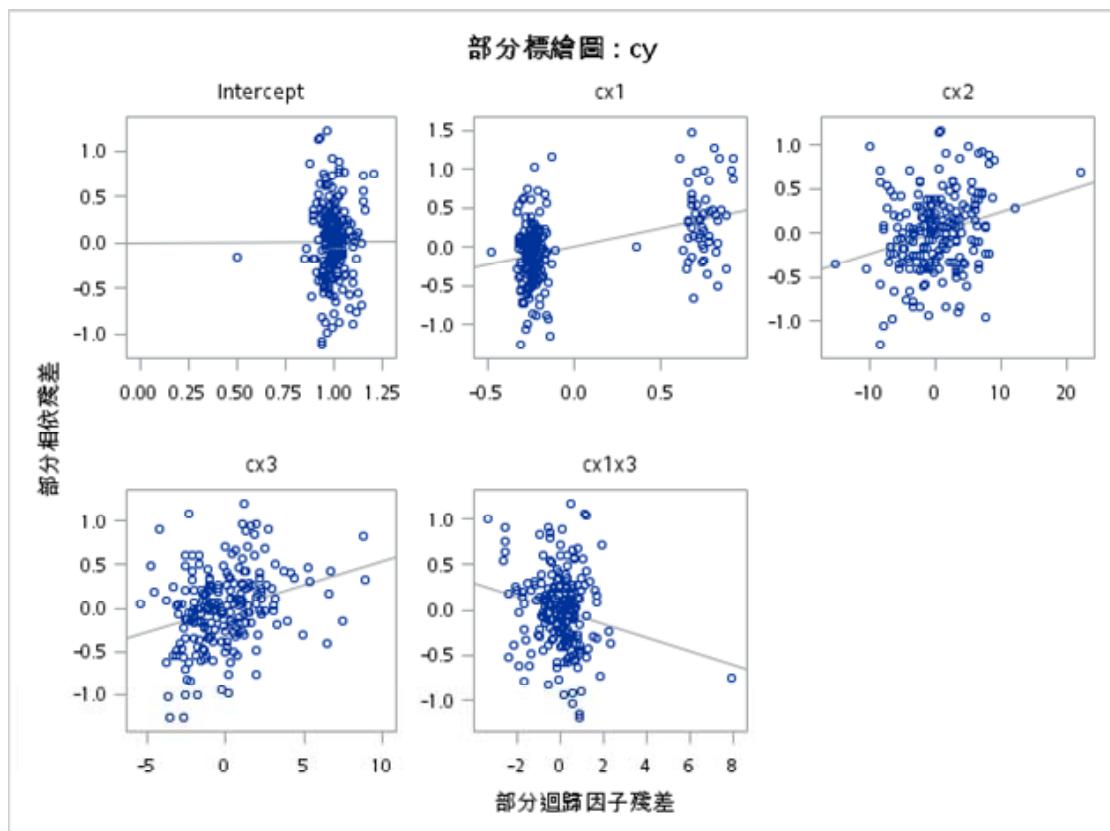
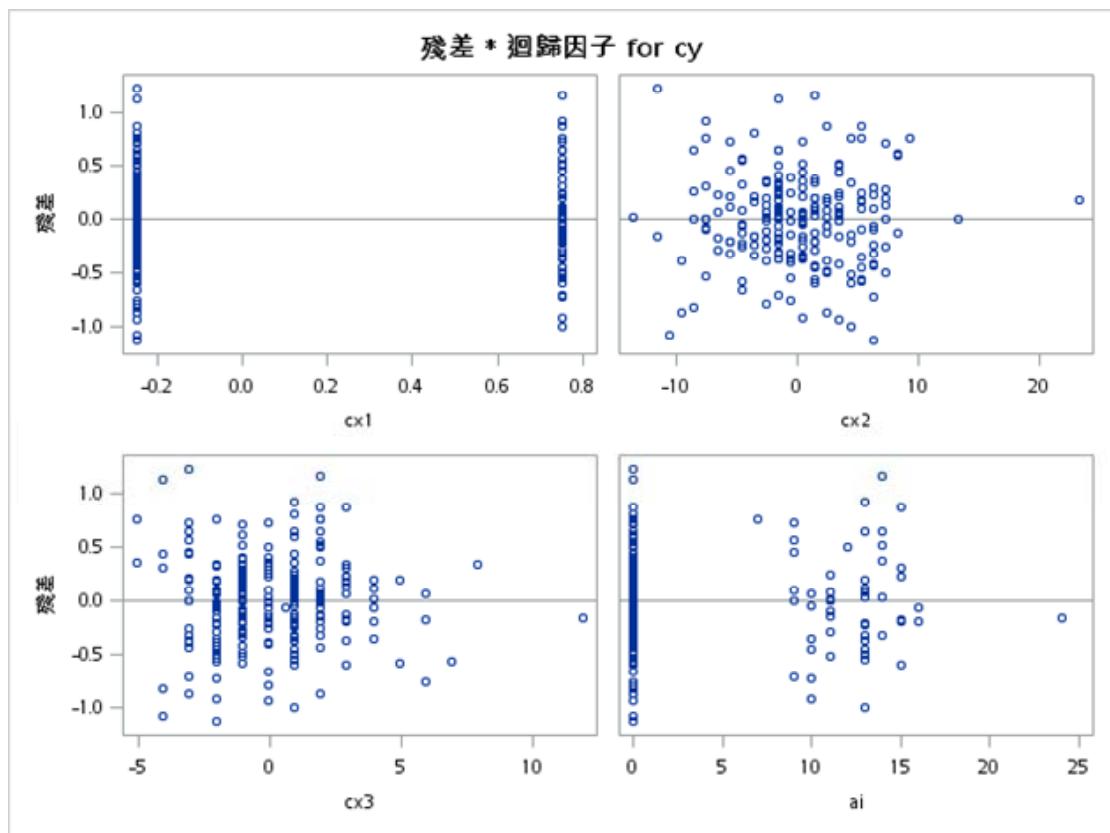
估計值的共變異數						
變動	Intercept	cx1	cx2	cx3	cx1x3	
Intercept	0.0008554419	0.0000116731	-1.998915E-6	3.6025193E-6	-0.000052884	
cx1	0.0000116731	0.0046744619	-0.000046791	-0.000037101	-0.000156966	
cx2	-1.998915E-6	-0.000046791	0.000037403	-0.000011467	0.0000268789	
cx3	3.6025193E-6	-0.000037101	-0.000011467	0.0001526923	-0.000048442	
cx1x3	-0.000052884	-0.000156966	0.0000268789	-0.000048442	0.0007111127	

最終我們得到的方程式是

$$0.00566 + 0.47011 * cx1 + 0.02334 * cx2 + 0.05431 * cx3 - 0.07612 * cx1 * cx3$$

從標準化估計值可知心房顫動種類有最大的影響力.





殘差

殘差屬於常態分配

常態性檢定				
檢定	統計值		p 值	
Shapiro-Wilk	W	0.993023	Pr < W	0.4244
Kolmogorov-Smirnov	D	0.041501	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.075743	Pr > W-Sq	0.2387
Anderson-Darling	A-Sq	0.468487	Pr > A-Sq	0.2498

殘差符合恆常性

Breusch-Pagan test

變異數分析					
來源	自由度	平方和	平均值 平方	F 值	Pr > F
模型	4	0.46263	0.11566	1.63	0.1680
誤差	205	14.54730	0.07096		
已校正的總計	209	15.00992			

十六、結論

無論是心室或心房，皆會因長期壓力及容積過度負荷而擴大。當高血壓、嚴重瓣膜逆流升高心室灌流壓力，會使心房壁張力增加而導致腔室擴張。同時，左心房壓力增加也會造成心房動作電位縮短，因而使心房顫動容易被誘發；另一方面，實驗已經證明持續的心房顫動會改變電流傳導以及影響左心房的結構重塑，也就是說，心房顫動與左心房大小的關係複雜，想要建立彼此的因果關係較為困難，因影響左心房的結構改變因素不僅僅是心律不整，而是可能與潛在的心臟病態生理學有關，所以過去許多有關左心房大小的研究都會排除心房顫動的病人。此次迴歸期末報告，嘗試以因心房顫動接受電生理手術病患的非侵襲性檢查資料探討與左心房大小(預測值)有關的預測因子。我們得到的迴歸方程式與造成左心房變大的生理機制相符：一旦心房顫動變成慢性的，左心房的結構被重塑，會使得心房顫動日益嚴重，亦即就心因心房顫動接受電生理手術病患而言，最主要影響左心房大小的變數是心房顫動種類。從結果可知，偶發性心房顫動的病人其左心室越大，心室中膈越厚 (代表心室灌流壓力大)，左心房越大。慢性心房顫動的病人左心房較偶發性心房顫動的病人大(由截距可得)，且與左心室大小呈正相關；但由於交互作用的影響，心室中膈厚度的影響減弱。由以上結果，建議有心房顫動的病人應及早治療，避免變成慢性心房顫動，與左心房擴大彼此影響，造成治療上的困難及增加病患發生腦血管梗塞中風機率。