

# 縱向資料分析 hw4

410978002 統計三 謝元皓

(a) Sample covariance matrices:

共變異數矩陣、DF = 525				
	d1	d2	d3	d4
d1	0.1566105378	-.0369364476	-.0376751765	-.0376751765
d2	-.0369364476	0.1542639870	-.0369364476	-.0369364476
d3	-.0376751765	-.0369364476	0.1566105378	-.0376751765
d4	-.0376751765	-.0369364476	-.0376751765	0.1566105378

correlation matrices:

id 1 的估計 V 相關矩陣					
列	欄1	欄2	欄3	欄4	欄5
1	1.0000	0.4392	0.4392	0.4392	0.4392
2	0.4392	1.0000	0.4392	0.4392	0.4392
3	0.4392	0.4392	1.0000	0.4392	0.4392
4	0.4392	0.4392	0.4392	1.0000	0.4392
5	0.4392	0.4392	0.4392	0.4392	1.0000

配適統計值	
-2 殘差對數概度	3194.0
AIC (越小越好)	3198.0
AICC (越小越好)	3198.0
BIC (越小越好)	3203.5

(b) – (1) Homogeneous compound symmetry same in three groups.

模型資訊		配適統計值	
資料集	WORK.DS1A	-2 對數概度	3260.4
應變數	blood	AIC (越小越好)	3318.4
共變異數結構	複合對稱性	AICC (越小越好)	3321.9
監測對象效果	id	BIC (越小越好)	3399.2
估計方法	ML		
殘差變異數方法	設定檔		
固定效果 SE 方法	模型式		
自由度方法	包圍		

– (2) Compound symmetry with unequal variances.

模型資訊		配適統計值	
資料集	WORK.DS1A	-2 對數概度	3014.0
應變數	blood	AIC (越小越好)	3080.0
共變異數結構	異質複合對稱性	AICC (越小越好)	3084.5
監測對象效果	id	BIC (越小越好)	3171.9
估計方法	ML		
殘差變異數方法	設定檔		
固定效果 SE 方法	模型式		
自由度方法	包圍		

– (3) Homogeneous AR(1).

模型資訊		配適統計值	
資料集	WORK.DS1A	-2 對數概度	3259.5
應變數	blood	AIC (越小越好)	3317.5
共變異數結構	自迴歸	AICC (越小越好)	3321.0
監測對象效果	id	BIC (越小越好)	3398.3
估計方法	ML		
殘差變異數方法	設定檔		
固定效果 SE 方法	模型式		
自由度方法	包圍		

– (4) AR(1) with unequal variances.

模型資訊		配適統計值	
資料集	WORK.DS1A	-2 對數概度	3014.0
應變數	blood	AIC (越小越好)	3078.0
共變異數結構	異質自迴歸	AICC (越小越好)	3082.2
監測對象效果	id	BIC (越小越好)	3167.2
估計方法	ML		
殘差變異數方法	設定檔		
固定效果 SE 方法	模型式		
自由度方法	包圍		

– (5) Homogenous Toeplitz structure

模型資訊		配適統計值	
資料集	WORK.DS1A	-2 對數概度	3118.2
應變數	blood	AIC (越小越好)	3184.2
共變異數結構	Toeplitz	AICC (越小越好)	3188.7
監測對象效果	id	BIC (越小越好)	3276.1
估計方法	ML		
殘差變異數方法	設定檔		
固定效果 SE 方法	模型式		
自由度方法	包圍		

– (6) Toeplitz structure with unequal variances.

模型資訊		最後一次迭代的共變異數參數值		
資料集	WORK.DS1A	共變異數參數	監測對象	估計值
應變數	blood	Var(1)	id	1.46E-9
共變異數結構	異質 Toeplitz	Var(2)	id	4.7209
監測對象效果	id	Var(3)	id	25.0334
估計方法	ML	Var(4)	id	44.4954
殘差變異數方法	設定檔	Var(5)	id	72.5495
固定效果 SE 方法	模型式	Var(6)	id	0.000631
自由度方法	包圍	TOEPH(1)	id	1.0000
		TOEPH(2)	id	1.0000
		TOEPH(3)	id	0.9135
		TOEPH(4)	id	0.5617
		TOEPH(5)	id	1.0000
		Residual		10.0236

– (7) A common completely unstructured covariance matrix

模型資訊	
資料集	WORK.DS1A
應變數	blood
共變異數結構	非結構化
監測對象效果	id
估計方法	ML
殘差變異數方法	設定檔
固定效果 SE 方法	模型式
自由度方法	包圍

配適統計值	
-2 對數概度	2998.8
AIC (越小越好)	3092.8
AICC (越小越好)	3102.3
BIC (越小越好)	3223.9

結論:因為在異質自回歸中有最小的 AIC，故最好的共變異數結構為 AR(1) with unequal variances.

(C)

### Model 1:

模型資訊	
資料集	WORK.DS1A
應變數	blood
共變異數結構	非結構化
監測對象效果	id
估計方法	REML
殘差變異數方法	設定檔
固定效果 SE 方法	模型式
自由度方法	包圍

配適統計值	
-2 殘差對數概度	3194.0
AIC (越小越好)	3198.0
AICC (越小越好)	3198.0
BIC (越小越好)	3203.5

Null 模型概度比檢定		
DF	卡方	Pr > ChiSq
1	121.46	<.0001

共變異數參數估計值					
共變異數參數	監測對象	估計值	標準誤差	Z 值	Pr > Z
UN(1,1)	id	17.0690	2.9478	5.79	<.0001
Residual		21.7953	1.5746	13.84	<.0001

Model 2:

模型資訊	
資料集	WORK.DS1C
應變數	blood
共變異數結構	非結構化
監測對象效果	id
估計方法	ML
殘差變異數方法	設定檔
固定效果 SE 方法	模型式
自由度方法	包圍

配適統計值	
-2 對數概度	3029.2
AIC (越小越好)	3061.2
AICC (越小越好)	3062.3
BIC (越小越好)	3105.8

Null 模型概度比檢定		
DF	卡方	Pr > ChiSq
3	366.67	<.0001

共變異數參數估計值					
共變異數參數	監測對象	估計值	標準誤差	Z 值	Pr Z
UN(1,1)	id	2.2985	1.2371	1.86	0.0316
UN(2,1)	id	-0.01233	0.3478	-0.04	0.9717
UN(2,2)	id	1.0978	0.1815	6.05	<.0001
Residual		9.7272	0.8143	11.95	<.0001

Model 3:

模型資訊	
資料集	WORK.DS1C
應變數	blood
共變異數結構	非結構化
監測對象效果	id
估計方法	ML
殘差變異數方法	設定檔
固定效果 SE 方法	模型式
自由度方法	包圍

配適統計值	
-2 對數概度	3017.1
AIC (越小越好)	3067.1
AICC (越小越好)	3069.7
BIC (越小越好)	3136.7

Null 模型概度比檢定		
DF	卡方	Pr > ChiSq
6	376.23	<.0001

共變異數參數估計值					
共變異數參數	監測對象	估計值	標準 誤差	Z 值	Pr Z
UN(1,1)	id	6.3670	7.9365	0.80	0.2112
UN(2,1)	id	1.6382	1.3555	1.21	0.2268
UN(2,2)	id	1.5595	0.3645	4.28	<.0001
UN(3,1)	id	-1.4141	2.1708	-0.65	0.5148
UN(3,2)	id	-0.5811	0.4231	-1.37	0.1696
UN(3,3)	id	0.5307	0.6587	0.81	0.2102
Residual		8.9063	0.9203	9.68	<.0001

藉由 **AIC test** 可以看出，**Model 2 AIC** 估計值最小，故在選用模型時用 **model 2** 最好。

#### **(d)**

首先，多做一行 **time\_0**，而此行代表的意義是：當測試時間在一開始(week 0)時，則 **time\_0 = 1**，其他階段 **time\_0 = 0**。

觀測值	id	age	gender	time	blood	treat	a	g1	g2	time_4	time_0
1	1	0	1	0	31.8	1	0	1	0	4	1
2	1	0	1	2	31.6	1	0	1	0	4	0
3	1	0	1	4	39.9	1	0	1	0	4	0
4	1	0	1	6	40.5	1	0	1	0	6	0
5	1	0	1	8	48.3	1	0	1	0	8	0
6	2	0	0	0	24.5	1	0	1	0	4	1
7	2	0	0	2	28.1	1	0	1	0	4	0
8	2	0	0	4	30.6	1	0	1	0	4	0
9	2	0	0	6	34.8	1	0	1	0	6	0
10	2	0	0	8	35.3	1	0	1	0	8	0

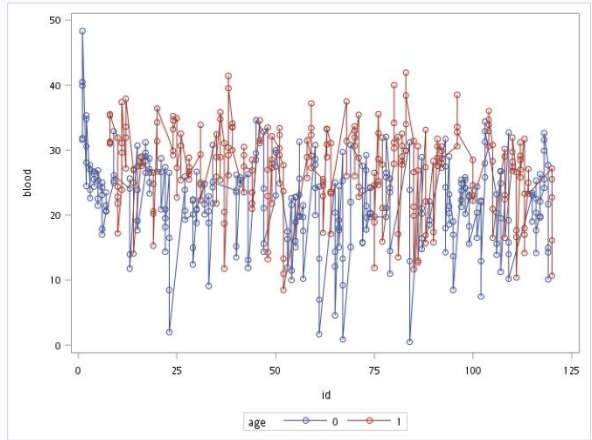
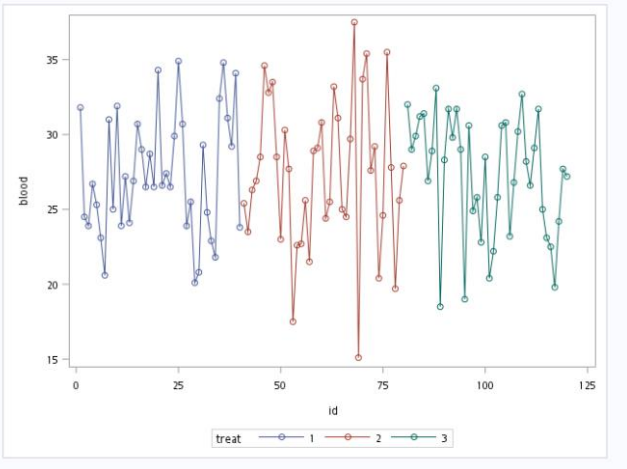
再來用此 model 做 likelihood test，以下表格:

共變異數參數估計值					
共變異數參數	監測對象	估計值	標準 誤差	Z 值	Pr Z
CS	id	17.0690	2.9478	5.79	<.0001
Residual		21.7953	1.5746	13.84	<.0001

配適統計值	
-2 殘差對數概度	3189.8
AIC (越小越好)	3193.8
AICC (越小越好)	3193.8
BIC (越小越好)	3199.4

Null 模型概度比檢定		
DF	卡方	Pr > ChiSq
1	121.46	<.0001

由以下兩張圖可以看出，blood lead level 值並沒有因為 treat 和 age 分組而固定，固由圖形可看出其為完全隨機設計。





H0: mean lead level at baseline for subjects to be the same in each group for each age group

Ha : mean lead level at baseline for subjects to be different in each group for each age group

(e)

Likelihood test :

配適統計值	
-2 殘差對數概度	2943.9
AIC (越小越好)	2973.9
AICC (越小越好)	2974.9
BIC (越小越好)	3015.7

Null 模型概度比檢定		
DF	卡方	Pr > ChiSq
14	367.38	<.0001

共變異數參數估計值					
共變異數參數	監測對象	估計值	標準誤差	Z 值	Pr Z
UN(1,1)	id	15.1533	2.1688	6.99	<.0001
UN(2,1)	id	11.0212	2.6041	4.23	<.0001
UN(2,2)	id	35.8110	5.0230	7.13	<.0001
UN(3,1)	id	14.5240	3.1678	4.58	<.0001
UN(3,2)	id	33.1015	5.3243	6.22	<.0001
UN(3,3)	id	53.9487	7.3692	7.32	<.0001
UN(4,1)	id	17.4215	3.9506	4.41	<.0001
UN(4,2)	id	42.2339	6.6974	6.31	<.0001
UN(4,3)	id	57.8305	8.4356	6.86	<.0001
UN(4,4)	id	81.7347	11.2757	7.25	<.0001
UN(5,1)	id	3.0936	1.4481	2.14	0.0327
UN(5,2)	id	2.5325	2.0990	1.21	0.2276
UN(5,3)	id	0.3171	2.5365	0.13	0.9005
UN(5,4)	id	4.2009	3.2289	1.30	0.1933
UN(5,5)	id	12.0255	3.7124	3.24	<.0001

固定效果的類型 3 檢定				
效果	分子自由度	分母自由度	F 值	Pr > F
time	4	116	6.36	0.0001
g1	1	116	7.66	0.0066
g2	1	116	0.60	0.4391
age	1	116	87.59	<.0001
g1*time	4	116	2.02	0.0969
g2*time	4	116	1.08	0.3681

因為  $G^2 < \chi^2$ , 我們不會拒絕  $H_0$ .

(f)

$H_0$ : model (2) to have no dependence on age in either the "slope" terms.

$H_a$ : model (2) to have dependence on age in either the "slope" terms.

由以下表格估計可看出，

估計值					
標籤	估計值	標準 誤差	DF	t 值	Pr >  t
the slope(t1-t0) between age=0 and age=1 g1	0.9933	0.3479	119	2.86	0.0051
the slope(t1-t0) between age=0 and age=1 g2	0.6558	0.3482	119	1.88	0.0621
the slope(t1-t0) between age=0 and age=1 g0	0.7639	0.3473	119	2.20	0.0297

由此可知在  $g = 0$ (安慰劑)中 年齡差異會影響斜率的不同

由此可知在  $g = 1$ (低劑量)中 年齡差異會影響斜率的不同

由此可知在  $g = 2$ (高劑量)中 年齡差異不會影響斜率的不同

(g)

$H_0$ : 在相同時間點上，高劑量和低劑量的平均鉛含量變化模式相同。

$H_A$ : 在相同時間點上，高劑量和低劑量的平均鉛含量變化模式不同

估計值					
標籤	估計值	標準 誤差	DF	t 值	Pr >  t
g1-g2	2.1039	1.0569	116	1.99	0.0489

由上表可知，**reject  $h_0$** ，得出結論:在固定時間下，使用高劑量和低劑量所造成的平均鉛含量差異明顯。

(h)

估計值					
標籤	估計值	標準 誤差	DF	t 值	Pr >  t
the mean lead level at week 4 for treatment group 2	Non-est	.	.	.	.
the mean lead level at week 8 for treatment average of group2 and and 3	Non-est	.	.	.	.
the mean lead level at week 0 for treatment average of group2 and and 3	Non-est	.	.	.	.
the mean lead level at week 8 for treatment group 1	Non-est	.	.	.	.
the mean lead level at week 0 for treatment group 1	Non-est	.	.	.	.
the mean lead level at week 8 for treatment between group 1 and average of group2 and and 3	-1.0520	0.5284	116	-1.99	0.0489