

縱向資料分析 作業

410978002 統計三 謝元皓

第一題

(a) Model 1

由下面的表格可以看出，AFT MR 這兩項變數對 RCAF 有顯著的影響，因為這兩個因子之主效果 P 值皆小於 0.05；反之，Age sex 這兩項變數對 RCAF 沒有顯著的影響，因為這兩個因子之主效果 P 值皆大於 0.05。所以可以看出 AFT MR 是造成心導管灼燒術後纖顫復發的兩項因子。

來源	DF	類型 III SS	均方	F 值	Pr > F
Age	1	297.594466	297.594466	2.12	0.1470
AFT	1	4228.117840	4228.117840	30.15	<.0001
sex	1	426.836151	426.836151	3.04	0.0828
MR	1	4129.124353	4129.124353	29.44	<.0001

透過下面參數估計表可以看出參數 B_0 (截距項) AFT(x_2) 和 MR (x_4) 的 P-value 均<0.001，所以有充分證據顯示參數 B_0 x_2 x_4 均顯著不為 0，可以用這三個參數建模。反之，AGE(x_1) 和 sex (x_3) 的 P-value 不顯著，所以不宜取這兩個參數建立模型。

參數	估計值	標準 誤差	t 值	Pr > t
截距	69.68299296	4.76890305	14.61	<.0001
Age	-0.11741548	0.08060455	-1.46	0.1470
AFT	-11.87046995	2.16192681	-5.49	<.0001
sex	-3.29996983	1.89158555	-1.74	0.0828
MR	-7.53986928	1.38957327	-5.43	<.0001

(b) Model 2

由下面的表格可以看出，AFT MR LAD 這三項變數對 RCAF 有顯著的影響，因為這三個因子之主效果 P 值皆小於 0.05；反之，Age sex 這兩項變數對 RCAF 沒有顯著的影響，因為這兩個因子之主效果 P 值皆大於 0.05。所以可以看出 AFT MR LAD 是造成心導管灼燒術後纖顫復發的三項因子。

來源	DF	類型 III SS	均方	F 值	Pr > F
Age	1	124.367849	124.367849	1.02	0.3140
AFT	1	1378.268909	1378.268909	11.30	0.0009
sex	1	838.423593	838.423593	6.87	0.0095
MR	1	1168.826344	1168.826344	9.58	0.0023
LAD	1	3393.549067	3393.549067	27.82	<.0001

透過下面參數估計表可以看出參數 B_0 (截距項) AFT(x_2) MR (x_4) 和 LAD(x_5) sex (x_3) 的 P-value 均 <0.05，所以有充分證據顯示參數 B_0 x_2 x_3 x_4 x_5 均顯著不為 0，可以用這五個參數建模。反之，AGE(x_1) 的 P-value 不顯著，所以不宜取這一個參數建立模型。

參數	估計值	標準 誤差	t 值	Pr > t
截距	94.21976735	6.43563967	14.64	<.0001
Age	-0.07631119	0.07557193	-1.01	0.3140
AFT	-7.36140901	2.18988001	-3.36	0.0009
sex	-4.67523812	1.78319296	-2.62	0.0095
MR	-4.41166545	1.42512817	-3.10	0.0023
LAD	-0.72863912	0.13813756	-5.27	<.0001

(c) Model 3

由下面的表格可以看出，AGE AFT LAS 這三項變數對 RCAF 有顯著的影響，因為這三個因子之主效果 P 值皆小於 0.05；反之，sex MR 這兩項變數對 RCAF 沒有顯著的影響，因為這兩個因子之主效果 P 值皆大於 0.05。所以可以看出 AGE AFT LAS 是造成心導管灼燒術後纖顫復發的三項因子。

來源	DF	類型 III SS	均方	F 值	Pr > F
Age	1	1090.32920	1090.32920	15.77	0.0001
AFT	1	413.30581	413.30581	5.98	0.0155
sex	1	150.60777	150.60777	2.18	0.1417
MR	2	132.47742	66.23871	0.96	0.3856
LAS	1	12666.75897	12666.75897	183.24	<.0001

透過下面參數估計表可以看出參數 B_0 (截距項) AFT Age LAS 的 P-value 均 <0.05，所以有充分證據顯示參數 $B_0 X_1 X_6$ 均顯著不為 0，可以用這四個參數建模。反之，SEX MR 的 P-value 不顯著，所以不宜取這兩個參數建立模型。

參數	估計值		標準 誤差	t 值	Pr > t
截距	80.74924716	B	3.51572727	22.97	<.0001
Age	-0.22796341		0.05739977	-3.97	0.0001
AFT	-3.98251032		1.62871467	-2.45	0.0155
sex 1	-1.98475587	B	1.34464382	-1.48	0.1417
sex 0	0.00000000	B	.	.	.
MR 1	-1.13432504	B	1.78813429	-0.63	0.5267
MR 2	-2.86014019	B	2.16490208	-1.32	0.1882
MR 0	0.00000000	B	.	.	.
LAS	-0.38469056		0.02841860	-13.54	<.0001

(d) Model4

由下面的表格可以看出，AGE AFT LAS 這三項變數對 RCAF 有顯著的影響，因為這三個因子之主效果 P 值皆小於 0.05；反之，sex MR LAD 這三項變數對 RCAF 沒有顯著的影響，因為這三個因子之主效果 P 值皆大於 0.05。所以可以看出 AGE AFT LAS 是造成心導管灼燒術後纖顫復發的三項因子。

來源	DF	類型 III SS	均方	F 值	Pr > F
Age	1	1271.337192	1271.337192	18.63	<.0001
AFT	1	537.066946	537.066946	7.87	0.0056
sex	1	67.299708	67.299708	0.99	0.3221
MR	2	209.011760	104.505880	1.53	0.2191
LAS	1	9664.417998	9664.417998	141.60	<.0001
LAD	1	223.608780	223.608780	3.28	0.0720

透過下面參數估計表可以看出參數 B₀(截距項) AFT Age LAS 的 P-value 均 <0.05，所以有充分證據顯示參數 B₀ X₁ X₂ X₆ 均顯著不為 0，可以用這四個參數建模。反之，SEX MR LAD 的 P-value 不顯著，所以不宜取這三個參數建立模型。

參數	估計值		標準 誤差	t 值	Pr > t
截距	73.75560247	B	5.20882860	14.16	<.0001
Age	-0.25368502		0.05877781	-4.32	<.0001
AFT	-4.65941815		1.66098784	-2.81	0.0056
sex 1	-1.36907288	B	1.37869739	-0.99	0.3221
sex 0	0.00000000	B	.	.	.
MR 1	-1.27331285	B	1.77840410	-0.72	0.4749
MR 2	-3.59776738	B	2.18937375	-1.64	0.1021
MR 0	0.00000000	B	.	.	.
LAS	-0.42403156		0.03563357	-11.90	<.0001
LAD	0.23852628		0.13177736	1.81	0.0720

第二題

比較 model1 跟 model2 之間的差異。

- (1) 觀察兩模型的 F 檢定後可看出，顯著性 p 直均遠小於 0.05，所以得到兩組模型中的所有參數均有顯著的差異。
- (2) 由下表可觀察到，在 model 2 中，R square 的值大於 model 1，所以得知 model 2 模型的解釋能力比 model 1 好。
- (3) Model 1 的 C.V. 微大於 model 2 的 C.V.，得知 Model 1 的離散程度稍微比 Model 2 來的大。
- (4) Model2 的 MSE 稍微比 Model1 的 MSE 來的小，所以相比之下 Model2 的估計量比 Model1 來的準確。

Model 1

來源	DF	平方和	均方	F 值	Pr > F
模型	5	13576.32046	2715.26409	19.41	<.0001
誤差	178	24902.23780	139.90021		
已校正的總計	183	38478.55826			

R 平方	變異係數	根 MSE	LAEF 平均值
0.352828	23.55654	11.82794	50.21087

Model 2

來源	DF	平方和	均方	F 值	Pr > F
模型	6	16802.27021	2800.37837	22.87	<.0001
誤差	177	21676.28805	122.46490		
已校正的總計	183	38478.55826			

R 平方	變異係數	根 MSE	LAEF 平均值
0.436666	22.03982	11.06639	50.21087

第三題

比較 model1 跟 model3 之間的差異。

- (1) 觀察兩模型的 F 檢定後可看出，顯著性 p 直均遠小於 0.05，所以得到兩組模型中的所有參數均有顯著的差異。
- (2) 由下表可觀察到，在 model 3 中，R square 的值大於 model 1，所以得知 model 3 模型的解釋能力比 model 1 好。
- (3) Model 1 的 C.V. 大於 model 2 的 C.V.，得知 Model 1 的離散程度比 Model 3 來的大。
- (4) Model3 的 MSE 比 Model1 的 MSE 來的小，所以相比之下 Model3 的估計量比 Model1 來的準確。

Model 1

來源	DF	平方和	均方	F 值	Pr > F
模型	5	13576.32046	2715.26409	19.41	<.0001
誤差	178	24902.23780	139.90021		
已校正的總計	183	38478.55826			

R 平方	變異係數	根 MSE	LAEF 平均值
0.352828	23.55654	11.82794	50.21087

Model 3

來源	DF	平方和	均方	F 值	Pr > F
模型	6	26243.07943	4373.84657	63.27	<.0001
誤差	177	12235.47883	69.12700		
已校正的總計	183	38478.55826			

R 平方	變異係數	根 MSE	LAEF 平均值
0.682018	16.55870	8.314265	50.21087

第四題

比較 model1 跟 model4 之間的差異。

- (1) 觀察兩模型的 F 檢定後可看出，顯著性 p 直均遠小於 0.05，所以得到兩組模型中的所有參數均有顯著的差異。
- (2) 由下表可觀察到，在 model 4 中，R square 的值大於 model 1，所以得知 model 4 模型的解釋能力比 model 1 好。
- (3) Model 1 的 C.V. 大於 model 4 的 C.V.，得知 Model 1 的離散程度比 Model 4 來的大。
- (4) Model 4 的 MSE 比 Model1 的 MSE 來的小，所以相比之下 Model4 的估計量比 Model1 來的不偏且集中。

Model 1

來源	DF	平方和	均方	F 值	Pr > F
模型	5	13576.32046	2715.26409	19.41	<.0001
誤差	178	24902.23780	139.90021		
已校正的總計	183	38478.55826			

R 平方	變異係數	根 MSE	LAEF 平均值
0.352828	23.55654	11.82794	50.21087

Model 4

來源	DF	平方和	均方	F 值	Pr > F
模型	7	26466.68821	3780.95546	55.40	<.0001
誤差	176	12011.87005	68.24926		
已校正的總計	183	38478.55826			

R 平方	變異係數	根 MSE	LAEF 平均值
0.687830	16.45323	8.261311	50.21087

第五題

根據分析上面四組模型後，我會選擇模型四。因為其各項數居均顯示其模型估計得比其他三組精確；再者，模型四含有的變數較多，較能顯示出其真實的樣貌，所以我會推薦模型四去做估計。

第六題

根據上述分析，可得到每組模型均有不顯著的參數值。我建議將這些參數值透過像是標準化之類的統計方法進行調整後再訪入模型中。如此一來，模型中的參數才會更具代表性及解釋力。

