

Big Data Mining

巨量資料探勘

個案分析與實作三

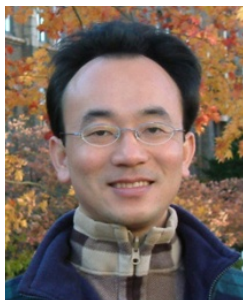
(SAS EM 決策樹、模型評估)：

Case Study 3 (Decision Tree, Model Evaluation using SAS EM)

1072DM08

MI4 (M2244) (2849)

Wed 6, 7 (13:10-15:00) (B206)



Min-Yuh Day

戴敏育

Assistant Professor

專任助理教授

Dept. of Information Management, Tamkang University

淡江大學 資訊管理學系

<http://mail.tku.edu.tw/myday/>

2019/05/01



課程大綱 (Syllabus)

週次 (Week) 日期 (Date) 內容 (Subject/Topics)

- 1 2019/02/20 巨量資料探勘課程介紹
(Course Orientation for Big Data Mining)
- 2 2019/02/27 AI人工智慧與大數據分析
(Artificial Intelligence and Big Data Analytics)
- 3 2019/03/06 分群分析 (Cluster Analysis)
- 4 2019/03/13 個案分析與實作一 (SAS EM 分群分析) :
Case Study 1 (Cluster Analysis - K-Means using SAS EM)
- 5 2019/03/20 關連分析 (Association Analysis)
- 6 2019/03/27 個案分析與實作二 (SAS EM 關連分析) :
Case Study 2 (Association Analysis using SAS EM)
- 7 2019/04/03 教學行政觀摩日 (Off-campus study)
- 8 2019/04/10 分類與預測 (Classification and Prediction)

課程大綱 (Syllabus)

週次 (Week) 日期 (Date) 內容 (Subject/Topics)

9 2019/04/17 期中報告 (Midterm Project Presentation)

10 2019/04/24 期中考試週 (Midterm Exam)

11 2019/05/01 個案分析與實作三 (SAS EM 決策樹、模型評估) :
Case Study 3 (Decision Tree, Model Evaluation using SAS EM)

12 2019/05/08 個案分析與實作四 (SAS EM 迴歸分析、類神經網路) :
Case Study 4 (Regression Analysis, Artificial Neural Network using SAS EM)

13 2019/05/15 機器學習與深度學習
(Machine Learning and Deep Learning)

14 2019/05/22 期末報告 (Final Project Presentation)

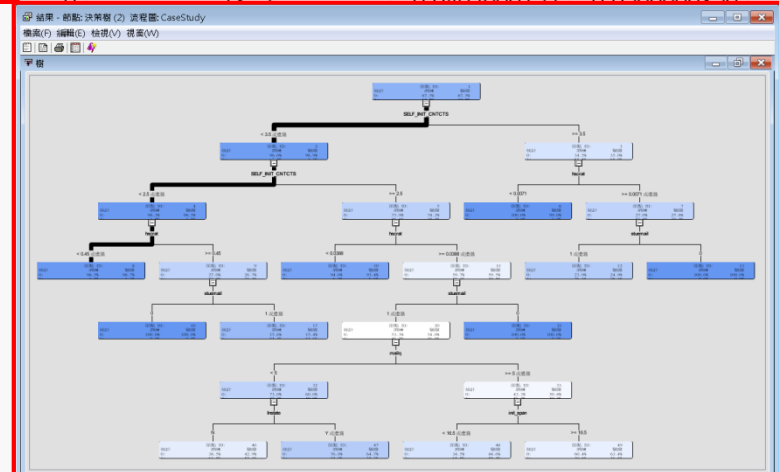
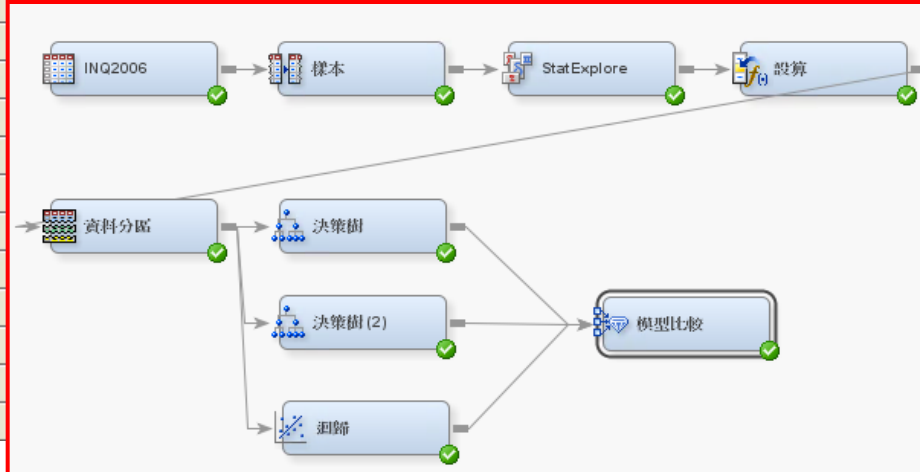
15 2019/05/29 畢業考試週 (Final Exam)

個案分析與實作三 (SAS EM 決策樹、模型評估) :

Case Study 3 (Decision Tree, Model Evaluation using SAS EM)

Enrollment Management

	avg_income	CAMPUS_VISIT	distance	Enroll_Target	hscrat	init_span	Instate	intlrat	int2rat	interest
2855	46007	0	80.107324675	1	0.0970873786	16	Y	0.084962913	0.0566696856	0
2856	48511	0	79.831475811	1	0.149122807	10	Y	0.2142857143	0.0909090909	0
2857	46007	0	80.107324675	1	0.0970873786	9	Y	0.1875	0.0629067245	0
2858	48511	0	79.831475811	1	0.149122807	21	Y	0.084962913	0.0330639236	1
2859	46007	1	80.107324675	1	0.0526315789	7	Y	0.0383403361	0.0824742268	0
2860	46007	0	80.107324675	1	0.0526315789	8	Y	0.0492696752	0.0566696856	0
2861	48511	1	79.831475811	1	0.149122807	14	Y	0.0523532522	0.0909090909	2
2862	48511	1	79.831475811	1	0.149122807	31	Y	0.0320610687	0.0500758725	0
2863	73195	0	79.593944675	1	0.149122807	16	Y	0.0492696752	0.0566696856	0
2864	73195	0	79.593944675	1	0.2105263158	15	Y	0.0492696752	0.0566696856	0
2865	73195	0	79.593944675	1	0.149122807	24	Y	0.084962913	0.0305851064	1
2866	73195	1	79.593944675	1	0.0406091371	23	Y	0.0492696752	0.0566696856	0
2867	62417	0	77.612290966	1	0.0406091371	21	Y	0.0492696752	0.0566696856	0
2868	62417	0	77.612290966	1	0.0406091371	24	Y	0.084962913	0.0566696856	0
2869	62417	0	77.612290966	0	0.0406091371	24	Y	0.052247191	0.0305851064	0
2870	62417	0	77.612290966	0	0.0376516305	4	Y	0	0.0566696856	0
2871	62417	0	77.612290966	0	0	18	Y	0.0402606752	0.0566696856	0



大學招生入學 預測模型

案例情境

- 某大型私立大學董事會今年要求負責招生的教務處與該校的資料分析研究中心共同合作，希望透過分析來找出今年哪些申請入學學生最有可能明年來註冊成為該校新鮮人。
- 董事會希望透過這個專案可以達成以下目標：
 - － 提升大一新鮮人的招生申請
 - － 增加入學的多元性
 - － 增加入學學生的SAT成績
- 從歷年的入學申請經驗，申請入學學生大約超過90,000，最後申請通過且會來註冊的學生數大約在2,400~2,800人。

資料欄位說明

VarID	Name	Model Role	Measurement Level	Description
1	ACADEMIC_INTEREST_1	Rejected	Nominal	Primary academic interest code
2	ACADEMIC_INTEREST_2	Rejected	Nominal	Secondary academic interest code
3	CAMPUS_VISIT	Input	Nominal	Campus visit code
4	CONTACT_CODE1	Rejected	Nominal	First contact code
5	CONTACT_DATE1	Rejected	Nominal	First contact date
6	ETHNICITY	Rejected	Nominal	Ethnicity
7	ENROLL	Target	Binary	1=Enrolled F2004, 0=Not enrolled F2004
8	IRSCHOOL	Rejected	Nominal	High school code
9	INSTATE	Input	Binary	1=In state, 0=Out of state
10	LEVEL_YEAR	Rejected	Unary	Student academic level
11	REFERRAL_CNTCTS	Input	Ordinal	Referral contact count
12	SELF_INIT_CNTCTS	Input	Interval	Self initiated contact count
13	SOLICITED_CNTCTS	Input	Ordinal	Solicited contact count
14	TERRITORY	Input	Nominal	Recruitment area
15	TOTAL_CONTACTS	Input	Interval	Total contact count
16	TRAVEL_INIT_CNTCTS	Input	Ordinal	Travel initiated contact count
17	AVG_INCOME	Input	Interval	Commercial HH income estimate
18	DISTANCE	Input	Interval	Distance from university
19	HSCRAT	Input	Interval	5-year high school enrollment rate
20	INIT_SPAN	Input	Interval	Time from first contact to enrollment date
21	INT1RAT	Input	Interval	5-year primary interest code rate
22	INT2RAT	Input	Interval	5-year secondary interest code rate
23	INTEREST	Input	Ordinal	Number of indicated extracurricular interests
24	MAILQ	Input	Ordinal	Mail qualifying score (1=very interested)
25	PREMIERE	Input	Binary	1=Attended campus recruitment event, 0=Did not
26	SATSCORE	Rejected	Interval	SAT (original) score
27	SEX	Rejected	Binary	Sex
28	STUEMAIL	Input	Binary	1=Have e-mail address, 0=Do not
29	TELECQ	Rejected	Ordinal	Telecounseling qualifying score (1=very interested)

預測模型實機演練

分析目的

依據可能影響申請學生正式註冊入學的考量因素，利用決策樹等分類演算方法建立入學預測模型。

演練重點：

- 資料瀏覽
- 遺失值處理
- 衍生變數產出
- 資料集分割(訓練、驗證、測試資料集)
- 預測模型變數篩選
- 決策樹預測模型建置
- 決策樹模型參數設定與模型解讀
- 跨模型評估方法

SAS Enterprise Miner (SAS EM) Case Study

- SAS EM 資料匯入4步驟
 - Step 1. 新增專案 (New Project)
 - Step 2. 新增資料館 (New / Library)
 - Step 3. 建立資料來源 (Create Data Source)
 - Step 4. 建立流程圖 (Create Diagram)
- SAS EM SEMMA 建模流程

Download

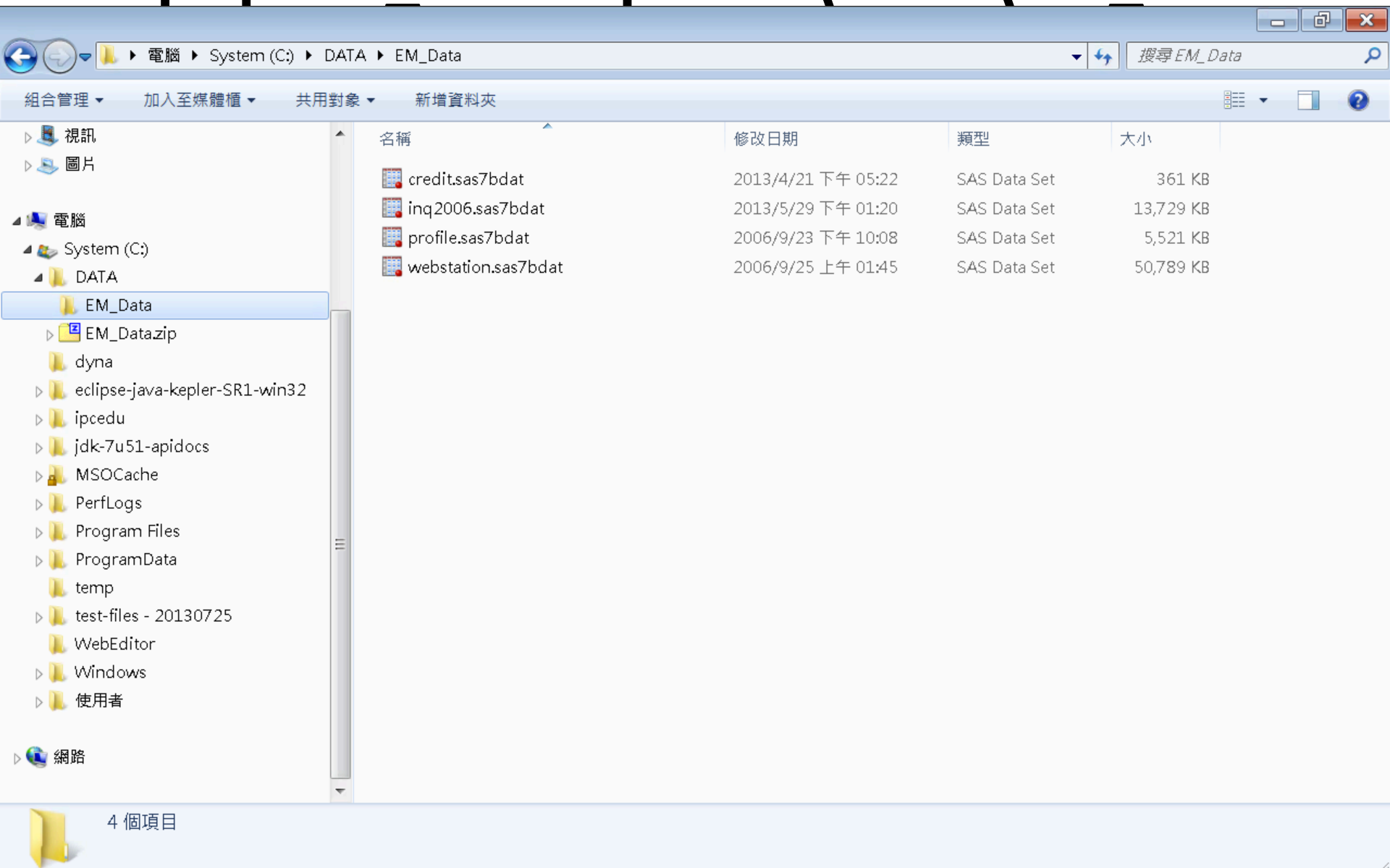
EM_Data.zip

[EM_Data]

(SAS EM Datasets)

http://mail.tku.edu.tw/myday/resources/BDM/Data/EM_Data.zip

Upzip EM_Data.zip to C:\DATA\EM_Data



The screenshot shows a Windows Explorer window with the address bar set to "電腦 > System (C:) > DATA > EM_Data". The search bar contains "搜尋 EM_Data". The left sidebar shows the "電腦" (Computer) section expanded, with "System (C:)" and "DATA" selected. Under "DATA", the "EM_Data" folder is highlighted. The main pane displays a list of four files:

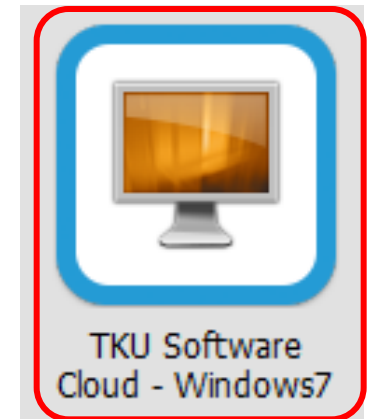
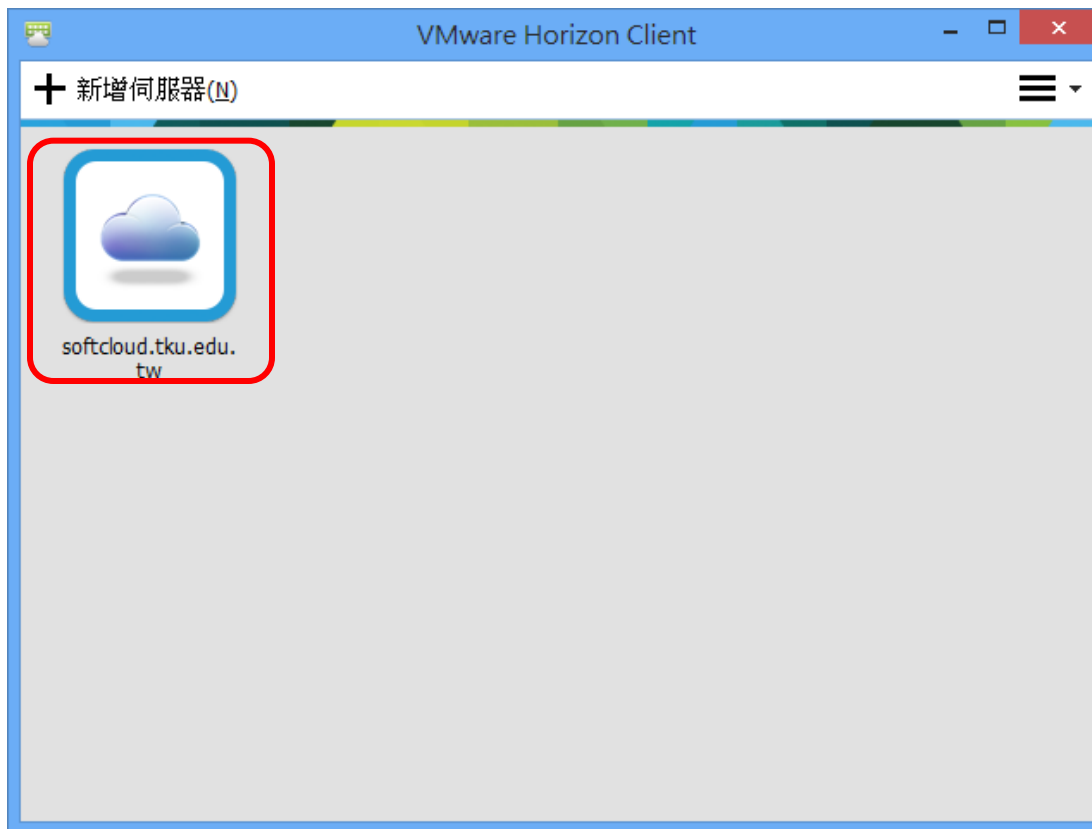
名稱	修改日期	類型	大小
credits.sas7bdat	2013/4/21 下午 05:22	SAS Data Set	361 KB
inq2006.sas7bdat	2013/5/29 下午 01:20	SAS Data Set	13,729 KB
profile.sas7bdat	2006/9/23 下午 10:08	SAS Data Set	5,521 KB
webstation.sas7bdat	2006/9/25 上午 01:45	SAS Data Set	50,789 KB

The status bar at the bottom indicates "4 個項目" (4 items).

VMware Horizon View Client

softcloud.tku.edu.tw

SAS Enterprise Miner



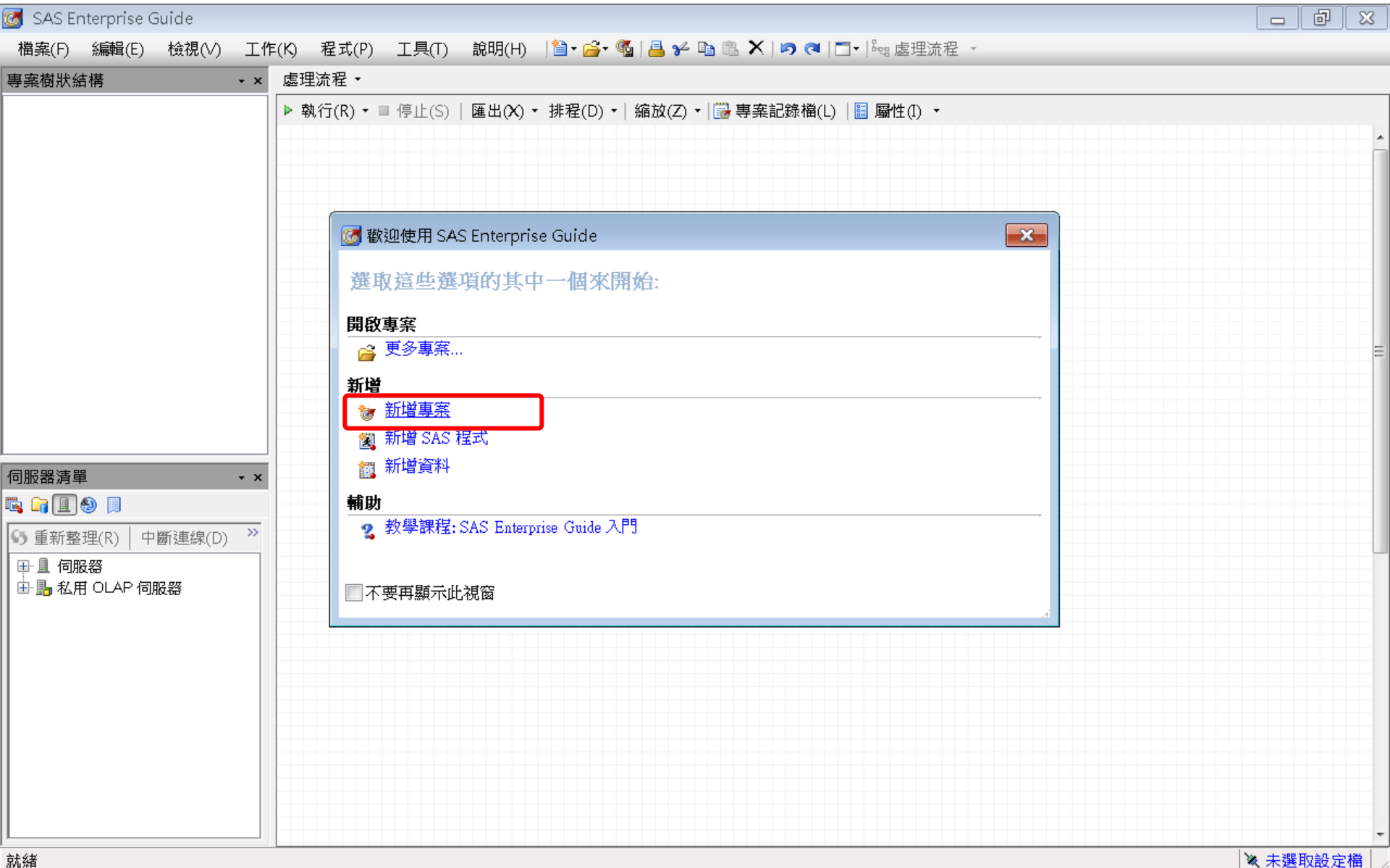
SAS Enterprise Guide (SAS EG)



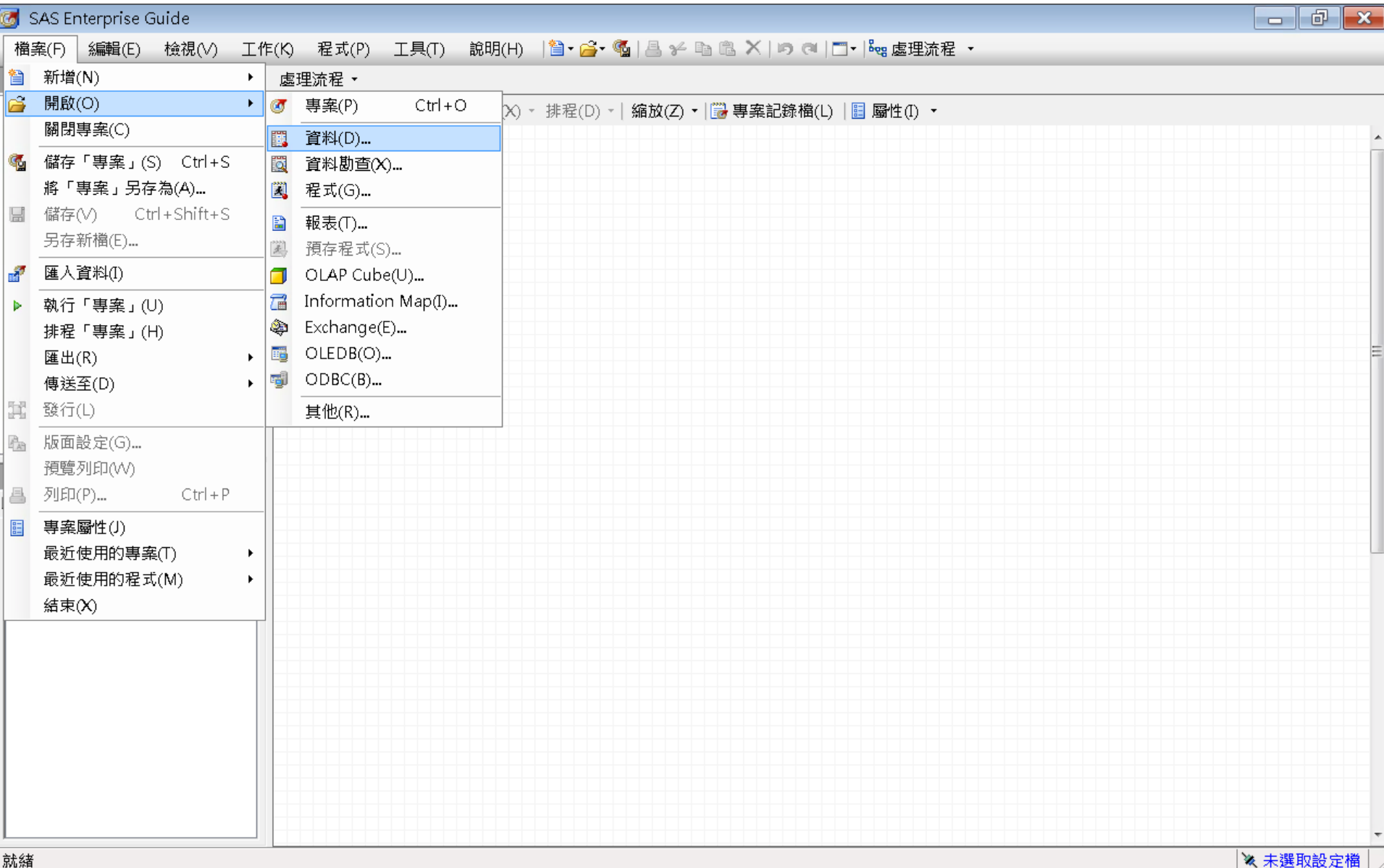
淡江軟體雲使用注意事項

- 登入淡江軟體雲時，會自動產生一部新的雲端主機。登出雲端主機前，請先將檔案儲存於您攜帶的隨身碟、淡江大學網路隨身碟、Google雲端硬碟...等，否則登出後雲端主機的所有資料將會被刪除，並回到您的個人主機畫面。
- 為免雲端主機帳號被冒用，每次使用完請登出。
- 軟體雲相關操作說明，請點選雲端主機桌面上『操作手冊』圖示。
- 軟體雲使用問題請至B212辦公室，或來電26215656-2129 為您服務。
- 請尊重智慧財產權，嚴禁安裝不法軟體，違法者自負相關法律責任。

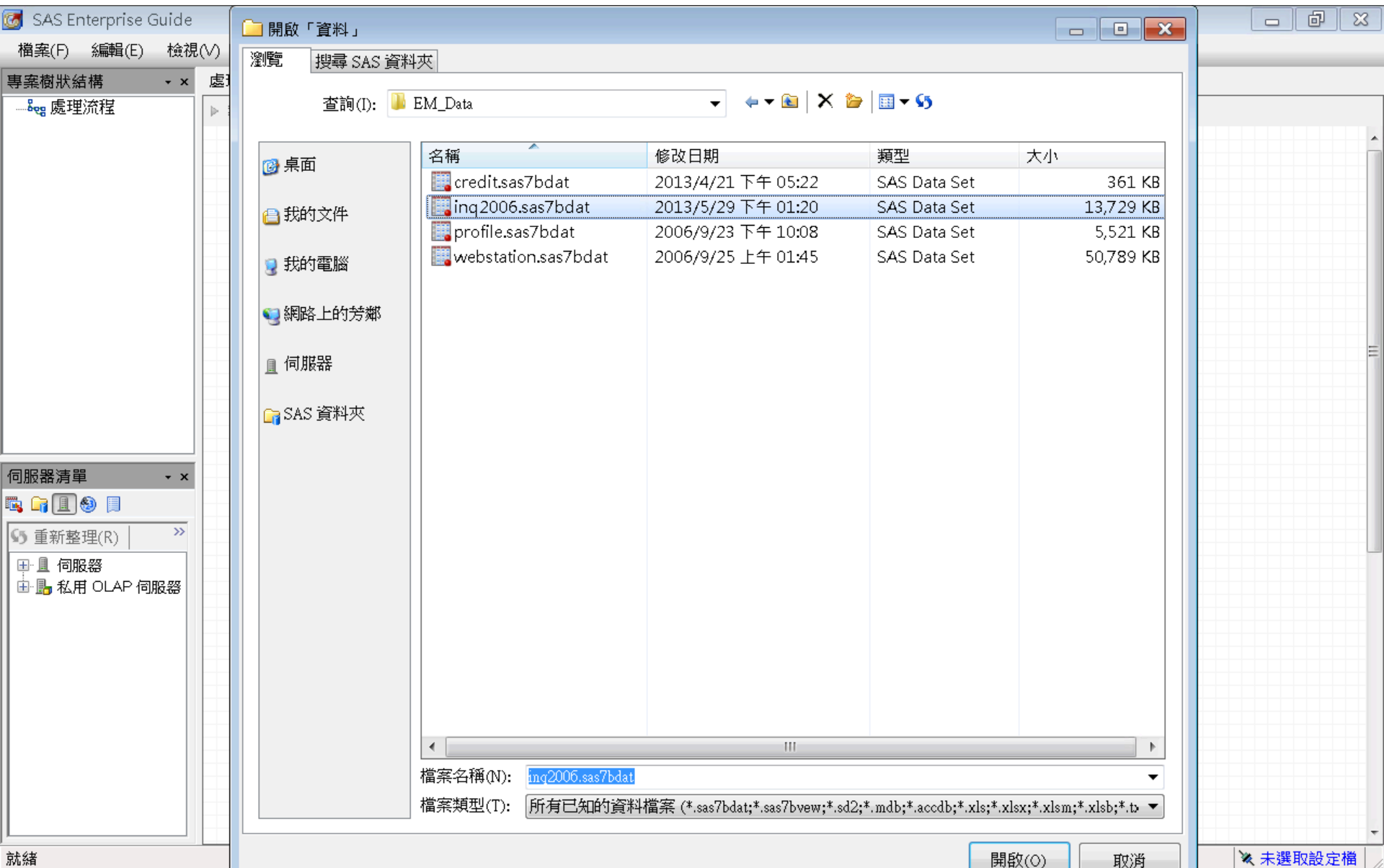
SAS EG New Project



SAS EG Open Data



SAS EG Open **inq2006.sas7bdat**



inq2006.sas7bdat

SAS Enterprise Guide

檔案(F) 編輯(E) 檢視(V) 工作(K) 程式(P) 工具(T) 說明(H) | 處理流程

專案樹狀結構

處理流程
inq2006

伺服器清單

重新整理(R)

伺服器
私用 OLAP 伺服器

inq2006

篩選和排序(L) 查詢產生器(Q) 資料(D) 描述(B) 圖形(G) 分析(Z) 匯出(X) 傳送至(N)

	TERRITORY	Enroll_Target	TOTAL_CONTACTS	SELF_INIT_CNTCTS	TRAVEL_INIT_CNTCTS	SOLICITED_CNTCTS	REFERI
1	N	0	1	0	0	1	
2	N	0	1	0	0	1	
3	N	0	2	2	0	0	
4	N	0	3	2	1	0	
5	N	0	1	1	0	0	
6	N	0	1	0	1	0	
7	N	0	1	0	0	1	
8	N	0	1	1	0	0	
9	N	0	2	2	0	0	
10	N	0	1	1	0	0	
11	N	0	1	0	0	1	
12	N	0	1	0	0	1	
13	N	0	1	1	0	0	
14	N	0	1	1	0	0	
15	N	0	2	2	0	0	
16	N	0	1	0	1	0	
17	N	0	1	0	1	0	
18	N	0	1	0	1	0	
19	N	0	1	0	1	0	
20	N	0	1	1	0	0	
21	N	0	1	1	0	0	
22	N	0	1	1	0	0	
23	N	0	1	1	0	0	
24	N	0	1	1	0	0	
25	N	0	1	1	0	0	
26	N	0	1	1	0	0	
27	N	0	1	1	0	0	
28	N	0	1	1	0	0	
29	N	0	1	1	0	0	
30	N	0	1	1	0	0	

就緒

未選取設定檔

inq2006.sas7bdat

SAS Enterprise Guide

檔案(F) 編輯(E) 檢視(V) 工作(K) 程式(P) 工具(T) 說明(H) | 處理流程

專案樹狀結構

- 處理流程
 - inq2006

伺服器清單

- 重新整理(R)
- 伺服器
- 私用 OLAP 伺服器

篩選和排序(L) 查詢產生器(Q) 資料(D) 描述(B) 圖形(G) 分析(Z) 匯出(X) 傳送至(N)

	REFERRAL_CNTCTS	CAMPUS_VISIT	satscore	mailq	premiere	interest	stuemail	init_sp
1	0	0	.	5	0	0	0	
2	0	0	.	5	0	0	1	
3	0	0	.	5	0	0	0	
4	0	0	.	5	0	0	1	
5	0	0	.	5	0	0	1	
6	0	0	.	2	0	0	1	
7	0	0	.	5	0	0	1	
8	0	0	.	5	0	0	1	
9	0	0	.	5	1	0	0	
10	0	1	.	4	0	0	0	
11	0	0	.	5	0	0	1	
12	0	0	.	5	0	0	1	
13	0	1	.	1	0	0	1	
14	0	0	.	5	0	0	0	
15	0	0	.	4	0	0	0	
16	0	0	.	2	0	0	1	
17	0	0	.	1	0	0	0	
18	0	0	.	5	0	0	0	
19	0	0	.	3	0	0	0	
20	0	0	.	5	0	0	0	
21	0	0	.	5	0	0	0	
22	0	0	.	5	0	1	0	
23	0	0	.	5	0	0	0	
24	0	0	.	5	0	0	0	
25	0	0	.	5	0	0	0	
26	0	0	.	5	0	0	0	
27	0	0	.	5	0	0	0	
28	0	0	.	5	0	0	0	
29	0	0	.	5	0	0	0	
30	0	0	.	5	0	0	0	

就緒 未選取設定檔

inq2006.sas7bdat

SAS Enterprise Guide

檔案(F) 編輯(E) 檢視(V) 工作(K) 程式(P) 工具(T) 說明(H) | 處理流程

專案樹狀結構

- 處理流程
 - inq2006

伺服器清單

- 重新整理(R)
- 伺服器
- 私用 OLAP 伺服器

篩選和排序(L) 查詢產生器(Q) 資料(D) 描述(B) 圖形(G) 分析(Z) 匯出(X) 傳送至(N)

	stuemail	init_span	intlrat	int2rat	hscrat	avg_income	distance	Instate
1	0	4	0.0171827457	0.0203800021	0.0376516305	.	.	N
2	1	19	0.0171827457	0.0203800021	0.0376516305	.	.	N
3	0	27	0.0171827457	0.0203800021	0.0376516305	.	.	N
4	1	34	0.0171827457	0.0203800021	0.0376516305	.	.	N
5	1	45	0.0171827457	0.0203800021	0.0376516305	.	.	N
6	1	46	0.0171827457	0.0203800021	0.0376516305	.	.	N
7	1	29	0.0171827457	0.0203800021	0.0376516305	.	.	N
8	1	27	0.0171827457	0.0203800021	0.0376516305	.	.	N
9	0	36	0.0171827457	0.0203800021	0.0376516305	.	.	N
10	0	34	0.0171827457	0.0203800021	0.0376516305	.	.	N
11	1	30	0.0171827457	0.0203800021	0.0376516305	.	.	N
12	1	30	0.0171827457	0.0203800021	0.0376516305	.	.	N
13	1	41	0.0171827457	0.0203800021	0.0376516305	.	.	N
14	0	58	0.0171827457	0.0203800021	0.0376516305	.	.	N
15	0	27	0.0171827457	0.0203800021	0.0376516305	.	.	N
16	1	32	0.0171827457	0.0203800021	0.0376516305	.	.	N
17	0	35	0.0171827457	0.0203800021	0.0376516305	.	.	N
18	0	34	0.0171827457	0.0203800021	0.0376516305	.	.	N
19	0	72	0.0171827457	0.0203800021	0.0376516305	.	.	N
20	0	61	0.0171827457	0.0203800021	0.0376516305	.	.	N
21	0	59	0.0171827457	0.0203800021	0.0376516305	.	.	N
22	0	58	0.0171827457	0.0203800021	0.0376516305	.	.	N
23	0	58	0.0171827457	0.0203800021	0.0376516305	.	.	N
24	0	58	0.0171827457	0.0203800021	0.0376516305	.	.	N
25	0	56	0.0171827457	0.0203800021	0.0376516305	.	.	N
26	0	56	0.0171827457	0.0203800021	0.0376516305	.	.	N
27	0	56	0.0171827457	0.0203800021	0.0376516305	.	.	N
28	0	55	0.0171827457	0.0203800021	0.0376516305	.	.	N
29	0	55	0.0171827457	0.0203800021	0.0376516305	.	.	N
30	0	55	0.0171827457	0.0203800021	0.0376516305	.	.	N

就緒

未選取設定檔

inq2006.sas7bdat

SAS Enterprise Guide

檔案(F) 編輯(E) 檢視(V) 工作(K) 程式(P) 工具(T) 說明(H) | 處理流程

專案樹狀結構

- 處理流程
 - inq2006

伺服器清單

- 重新整理(R)
- 伺服器
- 私用 OLAP 伺服器

篩選和排序(L) 查詢產生器(Q) 資料(D) 描述(B) 圖形(G) 分析(Z) 匯出(X) 傳送至(N)

	TERRITORY	Enroll_Target	TOTAL_CONTACTS	SELF_INIT_CNTCTS	TRAVEL_INIT_CNTCTS	SOLICITED_CNTCTS	REFERI
91454	5	0	1	0	1	0	
91455	5	0	2	1	1	0	
91456	5	0	1	0	1	0	
91457	5	0	1	0	0	1	
91458	8	0	1	0	1	0	
91459	5	0	2	1	1	0	
91460	5	0	2	0	1	1	
91461	5	0	2	1	1	0	
91462	5	0	6	4	2	0	
91463	5	0	2	1	1	0	
91464	5	0	1	0	1	0	
91465	5	0	1	0	1	0	
91466	5	0	1	0	0	1	
91467	5	0	1	0	1	0	
91468	5	0	3	2	1	0	
91469	5	0	3	1	1	1	
91470	5	1	8	7	0	1	
91471	5	0	1	0	1	0	
91472	5	0	2	1	1	0	
91473	5	0	1	0	0	1	
91474	5	0	1	0	0	1	
91475	5	0	3	2	0	1	
91476	8	0	2	1	1	0	
91477	8	0	2	1	1	0	
91478	8	0	2	1	1	0	
91479	8	0	3	2	1	0	
91480	5	0	1	0	1	0	
91481	8	0	2	1	1	0	
91482	5	0	2	1	1	0	

就緒

未選取設定檔

inq2006.sas7bdat

SAS Enterprise Guide

檔案(F) 編輯(E) 檢視(V) 工作(K) 程式(P) 工具(T) 說明(H) | 處理流程

專案樹狀結構

- 處理流程
 - inq2006

伺服器清單

- 重新整理(R)
- 伺服器
- 私用 OLAP 伺服器

篩選和排序(L) 查詢產生器(Q) | 資料(D) 描述(B) 圖形(G) 分析(Z) 匯出(X) 傳送至(N)

	REFERRAL_CNTCTS	CAMPUS_VISIT	satscore	mailq	premier	interest	stuemai	init_span	intlrat	int2rat
91454	0	0	.	3	0	0	0	23	0.084962913	0.1351351351
91455	0	0	.	2	0	0	1	23	0.0463576159	0.1818181818
91456	0	0	.	1	0	0	0	11	0.0209059233	0.036
91457	0	0	.	5	0	0	1	31	0.0380622837	0.0377358491
91458	0	0	.	3	0	0	0	11	0.052247191	0.0566696856
91459	0	0	.	5	0	0	1	20	0.084962913	0.1111111111
91460	0	0	.	2	0	0	1	29	0.0364640884	0.0566696856
91461	0	0	.	2	0	0	1	21	0.0523532522	0.0611790879
91462	0	0	940	2	0	0	1	46	0.052247191	0.035359116
91463	0	0	.	3	0	0	1	21	0.0364640884	0.0629067245
91464	0	0	.	1	0	0	0	11	0.0492696752	0.0566696856
91465	0	0	.	1	0	0	0	21	0.0523532522	0.035359116
91466	0	0	.	5	0	0	1	31	0.0523532522	0.0566696856
91467	0	0	.	2	0	0	0	23	0.0084745763	0.0566696856
91468	0	0	920	1	0	0	1	23	0.0523532522	0.0629067245
91469	0	0	.	2	0	0	1	21	0.0523532522	0.0892857143
91470	0	0	1200	1	0	0	1	19	0.0523532522	0.0629067245
91471	0	0	.	1	0	0	0	21	0.0492696752	0.0566696856
91472	0	0	.	1	0	1	0	30	0.0383403361	0.0629067245
91473	0	0	.	5	0	0	0	20	0.0492696752	0.0566696856
91474	0	0	.	5	0	0	0	23	0.0492696752	0.0566696856
91475	0	0	.	3	0	0	1	17	0.0731707317	0.0612244898
91476	0	0	.	1	0	0	1	11	0.0492696752	0.0566696856
91477	0	0	.	5	0	0	1	11	0.0492696752	0.0566696856
91478	0	0	670	2	0	0	0	11	0.0523532522	0.039244186
91479	0	0	.	5	0	0	1	12	0.0343065693	0.0566696856
91480	0	0	.	2	0	0	0	23	0.0682926829	0.034591195
91481	0	0	.	2	0	0	1	11	0.0459290188	0.0330639236
91482	0	0	.	3	0	0	1	11	0.0459290188	0.034591195

就緒

未選取設定檔

inq2006.sas7bdat

SAS Enterprise Guide

檔案(F) 編輯(E) 檢視(V) 工作(K) 程式(P) 工具(T) 說明(H) | 處理流程

專案樹狀結構

- 處理流程
 - inq2006

伺服器清單

- 重新整理(R)
- 伺服器
- 私用 OLAP 伺服器

篩選和排序(L) 查詢產生器(Q) 資料(D) 描述(B) 圖形(G) 分析(Z) 匯出(X) 傳送至(N)

	mailq	premier	interest	stuemai	init_span	int1rat	int2rat	hscrat	avg_income	distance	Instate
91454	3	0	0	0	23	0.084962913	0.1351351351	0	28829	532.48271929	Y
91455	2	0	0	1	23	0.0463576159	0.1818181818	0	28829	532.48271929	Y
91456	1	0	0	0	11	0.0209059233	0.036	0	28829	532.48271929	Y
91457	5	0	0	1	31	0.0380622837	0.0377358491	0	28829	532.48271929	Y
91458	3	0	0	0	11	0.052247191	0.0566696856	0	28829	532.48271929	Y
91459	5	0	0	1	20	0.084962913	0.1111111111	0	28829	532.48271929	Y
91460	2	0	0	1	29	0.0364640884	0.0566696856	0.0181818182	28829	532.48271929	Y
91461	2	0	0	1	21	0.0523532522	0.0611790879	0.0181818182	28829	532.48271929	Y
91462	2	0	0	1	46	0.052247191	0.035359116	0.0181818182	28829	532.48271929	Y
91463	3	0	0	1	21	0.0364640884	0.0629067245	0.0181818182	28829	532.48271929	Y
91464	1	0	0	0	11	0.0492696752	0.0566696856	0	28829	532.48271929	Y
91465	1	0	0	0	21	0.0523532522	0.035359116	0	28829	532.48271929	Y
91466	5	0	0	1	31	0.0523532522	0.0566696856	0.0194174757	28829	532.48271929	Y
91467	2	0	0	0	23	0.0084745763	0.0566696856	0.0194174757	28829	532.48271929	Y
91468	1	0	0	1	23	0.0523532522	0.0629067245	0.0194174757	28829	532.48271929	Y
91469	2	0	0	1	21	0.0523532522	0.0892857143	0.0169491525	28829	532.48271929	Y
91470	1	0	0	1	19	0.0523532522	0.0629067245	0.0169491525	28829	532.48271929	Y
91471	1	0	0	0	21	0.0492696752	0.0566696856	0	28829	532.48271929	Y
91472	1	0	1	0	30	0.0383403361	0.0629067245	0.0438596491	.	.	Y
91473	5	0	0	0	20	0.0492696752	0.0566696856	0	.	.	Y
91474	5	0	0	0	23	0.0492696752	0.0566696856	0	.	.	Y
91475	3	0	0	1	17	0.0731707317	0.0612244898	0	.	.	Y
91476	1	0	0	1	11	0.0492696752	0.0566696856	0	.	555.86318982	Y
91477	5	0	0	1	11	0.0492696752	0.0566696856	0	.	555.86318982	Y
91478	2	0	0	0	11	0.0523532522	0.039244186	0	.	555.86318982	Y
91479	5	0	0	1	12	0.0343065693	0.0566696856	0.0294117647	.	555.86318982	Y
91480	2	0	0	0	23	0.0682926829	0.034591195	0.0147058824	.	555.86318982	Y
91481	2	0	0	1	11	0.0459290188	0.0330639236	0	.	555.86318982	Y
91482	3	0	0	1	11	0.0459290188	0.034591195	0.0438596491	.	555.86318982	Y

就緒

未選取設定檔

inq2006.sas7bdat

SAS Enterprise Guide

檔案(F) 編輯(E) 檢視(V) 工作(K) 程式(P) 工具(T) 說明(H) | 處理流程

專案樹狀結構

- 處理流程
 - inq2006

開啟「inq2006」(O)
匯出(X)
篩選和排序(F)...
查詢產生器(Q)...
發行(L)...
傳送至(N)
複製(C)
貼上(P)
刪除(D)
重新命名(M)
關閉資料(S)
屬性(I)

TERRITORY	Enroll_Target	TOTAL_CONTACTS	SELF_INIT_CNTCTS	TRAVEL_INIT_CNTCTS	SOLICITED_CNTCTS	REFERI
91408	0	1	0	1	0	
91469	0	2	1	1	0	
91470	0	1	0	1	0	
91471	0	1	0	0	1	
91472	0	1	0	1	0	
91473	0	2	1	1	0	
91474	0	2	0	1	1	
91475	0	2	1	1	0	
91476	0	6	4	2	0	
91477	0	2	1	1	0	
91478	0	1	0	1	0	
91479	0	1	0	1	0	
91480	0	1	0	0	1	
91481	0	1	0	1	0	
91482	0	1	0	1	0	
91408	0	3	2	1	0	
91469	0	3	1	1	1	
91470	1	8	7	0	1	
91471	0	1	0	1	0	
91472	0	2	1	1	0	
91473	0	1	0	0	1	
91474	0	1	0	0	1	
91475	0	3	2	0	1	
91476	0	2	1	1	0	
91477	0	2	1	1	0	
91478	0	2	1	1	0	
91479	0	3	2	1	0	
91480	0	1	0	1	0	
91481	0	2	1	1	0	
91482	0	2	1	1	0	

伺服器清單

- 重新整理(R)
- 伺服器
- 私用 OLAP 伺服器

就緒

未選取設定檔

inq2006.sas7bdat

SAS Enterprise Guide

檔案(F) 編輯(E) 檢視(V) 工作(K) 程式(P) 工具(T) 說明(H) | 處理流程

專案樹狀結構

- 處理流程
 - inq2006

伺服器清單

- 伺服器
- 私用 OLAP 伺服器

重新整理(R)

篩選和排序(L)

91454 5
91455 5
91456 5
91457 5
91458 8
91459 5
91460 5
91461 5
91462 5
91463 5
91464 5
91465 5
91466 5
91467 5
91468 5
91469 5
91470 5
91471 5
91472 5
91473 5
91474 5
91475 5
91476 8
91477 8
91478 8
91479 8
91480 5
91481 8
91482 5

「inq2006」的屬性

一般
欄
進階
摘要

名稱	類型	長度	格式	輸入格式	標籤
TERRITORY	字元	3	\$3.	\$3.	RECR_CODE
Enroll_Target	數值	8	8.		
TOTAL_CONTACTS	數值	8			
SELF_INIT_CNTCTS	數值	8			
TRAVEL_INIT_CNTCTS	數值	8			
SOLICITED_CNTCTS	數值	8			
REFERRAL_CNTCTS	數值	8			
CAMPUS_VISIT	數值	8			
satscore	數值	8			
mailq	數值	8			
premiere	數值	8			
interest	數值	8			
stuemail	數值	8			
init_span	數值	8			
int1rat	數值	8			
int2rat	數值	8			
hscrat	數值	8			
avg_income	數值	8			
distance	數值	8			
Instate	字元	1			

複製到剪貼簿(B)

其他 (F1)...

確定 取消

就緒

未選取設定檔

inq2006.sas7bdat

篩選和排序

SAS Enterprise Guide

檔案(F) 編輯(E) 檢視(V) 工作(K) 程式(P) 工具(T) 說明(H) | 處理流程

專案樹狀結構

- 處理流程
 - inq2006

伺服器清單

- 重新整理(R)
- 伺服器
- 私用 OLAP 伺服器

篩選和排序(L) | 查詢產生器(Q) | 資料(D) | 描述(B) | 圖形(G) | 分析(Z) | 匯出(X) | 傳送至(N)

	TERRITORY	Enroll_Target	TOTAL_CONTACTS	SELF_INIT_CNTCTS	TRAVEL_INIT_CNTCTS	SOLICITED_CNTCTS	REFERI
1	N	0	1	0	0	1	
2	N	0	1	0	0	1	
3	N	0	2	2	0	0	
4	N	0	3	2	1	0	
5	N	0	1	1	0	0	
6	N	0	1	0	1	0	
7	N	0	1	0	0	1	
8	N	0	1	1	0	0	
9	N	0	2	2	0	0	
10	N	0	1	1	0	0	
11	N	0	1	0	0	1	
12	N	0	1	0	0	1	
13	N	0	1	1	0	0	
14	N	0	1	1	0	0	
15	N	0	2	2	0	0	
16	N	0	1	0	1	0	
17	N	0	1	0	1	0	
18	N	0	1	0	1	0	
19	N	0	1	0	1	0	
20	N	0	1	1	0	0	
21	N	0	1	1	0	0	
22	N	0	1	1	0	0	
23	N	0	1	1	0	0	
24	N	0	1	1	0	0	
25	N	0	1	1	0	0	
26	N	0	1	1	0	0	
27	N	0	1	1	0	0	
28	N	0	1	1	0	0	
29	N	0	1	1	0	0	
30	N	0	1	1	0	0	

就緒

未選取設定檔

inq2006.sas7bdat

篩選和排序

SAS Enterprise Guide

檔案(F) 編輯(E) 檢視(V) 工作(K) 程式(P) 工具(T) 說明(H) | 處理流程

專案樹狀結構

- 處理流程
 - inq2006

伺服器清單

- 伺服器
- 私用 OLAP 伺服器

篩選和排序(L) 查詢產生器(Q) 資料(D) 描述(B) 圖形(G) 分析(Z) 匯出(X) 傳送至(N)

	TERRITORY	Enroll_Target	TOTAL_CONTACTS	SELF_INIT_CNTCTS	TRAVEL_INIT_CNTCTS	SOLICITED_CNTCTS	REFERI
1	N	0	1	0	0	1	
2	N	0	1	0	0	1	
3	N	0	2	2	0	0	
4	N	0	3	2	1	0	
5	N						
6	N						
7	N						
8	N						
9	N						
10	N						
11	N						
12	N						
13	N						
14	N						
15	N						
16	N						
17	N						
18	N						
19	N						
20	N						
21	N						
22	N						
23	N						
24	N						
25	N						
26	N						
27	N						
28	N						
29	N						
30	N						

篩選和排序 - C:\DATA\EM_Data\inq2006.sas7bdat

變數 篩選條件 排序 結果

可用的 (20)(A):

名稱
TERRITORY
Enroll_Target
TOTAL_CONTACTS
SELF_INIT_CNTCTS
TRAVEL_INIT_CN
SOLICITED_CNCT
REFERRAL_CNCT
CAMPUS_VISIT
satscore
mailq
premiere
interest
stuemail
init_span
init_span

選取的 0 / 20 (S):

名稱	類型	標籤
----	----	----

<選取至少一個變數>

☐ 顯示標籤而非變數名稱(L)

顯示預覽(P) 驗證(V) 確定 取消 說明

inq2006.sas7bdat

篩選和排序

SAS Enterprise Guide

檔案(F) 編輯(E) 檢視(V) 工作(K) 程式(P) 工具(T) 說明(H) | 處理流程

專案樹狀結構

- 處理流程
 - inq2006

篩選和排序(L) 查詢產生器(Q) 資料(D) 描述(B) 圖形(G) 分析(Z) 匯出(X) 傳送至(N)

	TERRITORY	Enroll_Target	TOTAL_CONTACTS	SELF_INIT_CNTCTS	TRAVEL_INIT_CNTCTS	SOLICITED_CNTCTS	REFERI
1	N	0	1	0	0	1	
2	N	0	1	0	0	1	
3	N	0	2	2	0	0	
4	N	0	3	2	1	0	
5	N						
6	N						
7	N						
8	N						
9	N						
10	N						
11	N						
12	N						
13	N						
14	N						
15	N						
16	N						
17	N						
18	N						
19	N						
20	N						
21	N						
22	N						
23	N						
24	N						
25	N						
26	N						
27	N						
28	N						
29	N						
30	N						

篩選和排序 - C:\DATA\EM_Data\inq2006.sas7bdat

變數 篩選條件 排序 結果

可用的 (0)(A):

名稱

選取的 20 / 20 (S):

名稱	類型	標籤
avg_income	數值	avg_income
CAMPUS_VISIT	數值	CAMPUS_VISIT
distance	數值	distance
Enroll_Target	數值	Enroll_Target
hscrat	數值	hscrat
init_span	數值	init_span
Instate	字元	Instate
int1rat	數值	int1rat
int2rat	數值	int2rat
interest	數值	interest
mailq	數值	mailq
premiere	數值	premiere
REFERRAL_CNTCTS	數值	REFERRAL_CNTCTS
satscore	數值	satscore
SELF_INIT_CNTCTS	數值	SELF_INIT_CNTCTS

☐ 顯示標籤而非變數名稱(L)

顯示預覽(P) 驗證(V) 確定 取消 說明

伺服器清單

- 伺服器
- 私用 OLAP 伺服器

重新整理(R)

就緒

未選取設定檔

inq2006.sas7bdat

篩選和排序

SAS Enterprise Guide

檔案(F) 編輯(E) 檢視(V) 工作(K) 程式(P) 工具(T) 說明(H) | 處理流程

專案樹狀結構

- 處理流程
 - inq2006

篩選和排序(L) 查詢產生器(Q) | 資料(D) 描述(B) 圖形(G) 分析(Z) 匯出(X) 傳送至(N)

	TERRITORY	Enroll_Target	TOTAL_CONTACTS	SELF_INIT_CNTCTS	TRAVEL_INIT_CNTCTS	SOLICITED_CNTCTS	REFERI
1	N	0	1	0	0	1	
2	N	0	1	0	0	1	
3	N	0	2	2	0	0	
4	N	0	3	2	1	0	
5	N						
6	N						
7	N						
8	N						
9	N						
10	N						
11	N						
12	N						
13	N						
14	N						
15	N						
16	N						
17	N						
18	N						
19	N						
20	N						
21	N						
22	N						
23	N						
24	N						
25	N						
26	N						
27	N						
28	N						
29	N						
30	N						

伺服器清單

- 重新整理(R)
- 伺服器
- 私用 OLAP 伺服器

篩選和排序 - C:\DATA\EM_Data\inq2006.sas7bdat

變數 篩選條件 排序 結果

指定排序:

排序依據(S):

- TERRITORY
- Enroll_Target
- TOTAL_CONTACTS
- SELF_INIT_CNTCTS
- TRAVEL_INIT_CNTCTS
- SOLICITED_CNTCTS
- REFERRAL_CNTCTS
- CAMPUS_VISIT
- satscore
- mailq
- premiere
- interest
- stuemail
- init_span
- intlrat
- int2rat
- hscrat
- avg_income
- distance
- Instate

遞增(A) 遞減(D)

全部清除(C)

顯示預覽(P) 驗證(V) 確定 取消 說明

就緒

未選取設定檔

inq2006.sas7bdat

篩選和排序

SAS Enterprise Guide

檔案(F) 編輯(E) 檢視(V) 工作(W) 程式(P) 工具(T) 說明(H) | 處理流程

專案樹狀結構

- 處理流程
 - inq2006

伺服器清單

- 伺服器
- 私用 OLAP 伺服器

重新整理(R)

篩選和排序 - C:\DATA\EM_Data\inq2006.sas7bdat

變數 篩選條件 排序 結果

指定排序:

排序依據(S):

Enroll_Target

次要鍵(T):

☐ 遞增(A) ☒ 遞減(D)

☐ 遞增(A) ☐ 遞減(D)

☐ 顯示標籤而非變數名稱(L)

全部清除(C)

顯示預覽(P) 驗證(V) 確定 取消 說明

	TERRITORY	Enroll_Target	TOTAL_CONTACTS	SELF_INIT_CNTCTS	TRAVEL_INIT_CNTCTS	SOLICITED_CNTCTS	REFERI
1	N	0	1	0	0	1	
2	N	0	1	0	0	1	
3	N	0	2	2	0	0	
4	N	0	3	2	1	0	
5	N						
6	N						
7	N						
8	N						
9	N						
10	N						
11	N						
12	N						
13	N						
14	N						
15	N						
16	N						
17	N						
18	N						
19	N						
20	N						
21	N						
22	N						
23	N						
24	N						
25	N						
26	N						
27	N						
28	N						
29	N						
30	N						

inq2006.sas7bdat

篩選和排序

SAS Enterprise Guide

檔案(F) 編輯(E) 檢視(V) 工作(K) 程式(P) 工具(T) 說明(H) | 處理流程

專案樹狀結構

- 處理流程
- inq2006
- 篩選和排序

伺服器清單

- 重新整理(R)
- 伺服器
- 私用 OLAP 伺服器

篩選和排序

輸入資料 | 程式碼 | 記錄檔 | 輸出資料

修改工作(Y) | 篩選和排序(L) | 查詢產生器(Q) | 資料(D) | 描述(B) | 圖形(G) | 分析(Z) | 匯出(X) | 傳送至(N) |

	avg_income	CAMPUS_VISIT	distance	Enroll_Target	hscrat	init_span	Instate	intlrat
1	71174	1	1606.4023121	1	0.333333333	19	N	0.0291878173
2	79781	1	1600.0477584	1	1	23	N	0.083333333
3	129811	1	1619.2802604	1	1	13	N	0.0128700129
4	37088	1	769.54495694	1	0.5	7	N	0.0273972603
5	36467	1	227.73504318	1	0.083333333	14	N	0.0171827457
6	36467	0	227.73504318	1	0.083333333	14	N	0.0121163166
7	36467	0	227.73504318	1	0.083333333	18	N	0.0171827457
8	62417	0	77.612290966	1	0.149122807	16	Y	0.0492696752
9	62417	0	77.612290966	1	0.025974026	30	Y	0.0523532522
10	62417	0	77.612290966	1	0.0185185185	20	Y	0.0383403361
11	62417	0	77.612290966	1	0.0294117647	16	Y	0.0492696752
12	63393	1	77.261271794	1	0.0970873786	15	Y	0.0492696752
13	63393	0	77.261271794	1	0.0970873786	18	Y	0.0492696752
14	63393	0	77.261271794	1	0.0970873786	7	Y	0.0492696752
15	59599	0	91.022949602	1	0.166666667	17	Y	0.0492696752
16	59599	0	91.022949602	1	0.0404040404	24	Y	0.1111111111
17	59599	0	91.022949602	1	0.0442477876	25	Y	0.0492696752
18	59599	1	91.022949602	1	0.0404040404	23	Y	0.0364640884
19	59599	1	91.022949602	1	0.0404040404	29	Y	0.084962913
20	59599	1	91.022949602	1	0.0442477876	19	Y	0.0383403361
21	59599	0	91.022949602	1	0.0404040404	24	Y	0.0492696752
22	47706	0	792.32687107	1	0.25	18	N	0.0171827457
23	37851	1	683.72562982	1	1	12	N	0.1004784689
24	49795	1	795.20026761	1	0.25	18	N	0.019969278
25	40546	0	761.1090257	1	0.5	18	N	0.0421052632
26	48698	0	989.825019	1	1	9	N	0.0171827457
27	36500	0	819.16102173	1	0.5	17	N	0.1004784689
28	46091	1	89.59438296	1	0.0404040404	11	Y	0.0343065693

就緒

未選取設定檔

inq2006.sas7bdat

篩選和排序

SAS Enterprise Guide

檔案(F) 編輯(E) 檢視(V) 工作(K) 程式(P) 工具(T) 說明(H) | 處理流程

專案樹狀結構

- 處理流程
 - inq2006
 - 篩選和排序

篩選和排序

輸入資料 | 程式碼 | 記錄檔 | 輸出資料

修改工作(Y) | 篩選和排序(L) | 查詢產生器(Q) | 資料(D) | 描述(B) | 圖形(G) | 分析(Z) | 匯出(X) | 傳送至(N)

	avg_income	CAMPUS_VISIT	distance	Enroll_Target	hscrat	init_span	Instate	intlrat
2855	46007	0	80.107324675	1	0.0970873786	16	Y	0.084962913
2856	48511	0	79.831475811	1	0.149122807	10	Y	0.2142857143
2857	46007	0	80.107324675	1	0.0970873786	9	Y	0.1875
2858	48511	0	79.831475811	1	0.149122807	21	Y	0.084962913
2859	46007	1	80.107324675	1	0.0526315789	7	Y	0.0383403361
2860	46007	0	80.107324675	1	0.0526315789	8	Y	0.0492696752
2861	48511	1	79.831475811	1	0.149122807	14	Y	0.0523532522
2862	48511	1	79.831475811	1	0.149122807	31	Y	0.0320610687
2863	73195	0	79.593944675	1	0.149122807	16	Y	0.0492696752
2864	73195	0	79.593944675	1	0.2105263158	15	Y	0.0492696752
2865	73195	0	79.593944675	1	0.149122807	24	Y	0.084962913
2866	73195	1	79.593944675	1	0.0406091371	23	Y	0.0492696752
2867	62417	0	77.612290966	1	0.0406091371	21	Y	0.0492696752
2868	62417	0	77.612290966	1	0.0406091371	24	Y	0.084962913
2869	62417	0	77.612290966	0	0.0406091371	24	Y	0.052247191
2870	62417	0	77.612290966	0	0.0376516305	4	Y	0
2871	62417	0	77.612290966	0	0	18	Y	0.0492696752
2872	62417	0	77.612290966	0	0.0406091371	16	Y	0.0492696752
2873	.	0	.	0	0.0406091371	14	Y	0.0492696752
2874	62417	0	77.612290966	0	0.0406091371	13	Y	0.0428211587
2875	.	0	.	0	0.0869565217	20	Y	0.0492696752
2876	62417	0	77.612290966	0	0.0376516305	34	Y	0.0523532522
2877	.	0	.	0	0.0406091371	14	Y	0.0492696752
2878	62417	1	77.612290966	0	0.0406091371	26	Y	0.0492696752
2879	62417	0	77.612290966	0	0.0406091371	17	Y	0.0492696752
2880	.	0	.	0	0.2105263158	14	Y	0.0492696752
2881	62417	0	77.612290966	0	0.0406091371	16	Y	0.0492696752
2882	.	0	.	0	0.0317460317	14	Y	0.0492696752

伺服器清單

- 重新整理(R)
- 伺服器
- 私用 OLAP 伺服器

就緒

未選取設定檔

inq2006.sas7bdat

篩選和排序

SAS Enterprise Guide

檔案(F) 編輯(E) 檢視(V) 工作(K) 程式(P) 工具(T) 說明(H) | 處理流程

專案樹狀結構

- 處理流程
 - inq2006
 - 篩選和排序

篩選和排序

輸入資料 | 程式碼 | 記錄檔 | 輸出資料

修改工作(Y) | 篩選和排序(L) | 查詢產生器(O) | 資料(D) | 描述(B) | 圖形(G) | 分析(Z) | 匯出(X) | 傳送至(N)

	int1rat	int2rat	interest	mailq	premiere	REFERRAL_CNTCTS	satscore	SELF_INIT_CNTCTS	SOLICIT
2855	0.084962913	0.0566696856	0	5	1	0	1120	7	
2856	0.2142857143	0.0909090909	0	5	1	0	1340	7	
2857	0.1875	0.0629067245	0	5	1	0	960	5	
2858	0.084962913	0.0330639236	1	1	0	1	1420	2	
2859	0.0383403361	0.0824742268	0	5	0	0	1040	2	
2860	0.0492696752	0.0566696856	0	5	0	0	1160	4	
2861	0.0523532522	0.0909090909	2	5	0	0	1230	5	
2862	0.0320610687	0.0500758725	0	1	0	0	1280	7	
2863	0.0492696752	0.0566696856	0	5	0	0	1090	3	
2864	0.0492696752	0.0566696856	0	5	0	0	1260	4	
2865	0.084962913	0.0305851064	1	1	0	0	1030	3	
2866	0.0492696752	0.0566696856	0	5	0	0	1050	9	
2867	0.0492696752	0.0566696856	0	5	0	0	1100	2	
2868	0.084962913	0.0566696856	0	5	0	1	1270	8	
2869	0.052247191	0.0305851064	0	5	0	0	1030	5	
2870	0	0.0566696856	0	5	0	0	.	0	
2871	0.0492696752	0.0566696856	0	5	1	0	1280	5	
2872	0.0492696752	0.0566696856	0	5	0	0	870	3	
2873	0.0492696752	0.0566696856	0	5	0	0	.	1	
2874	0.0428211587	0.0566696856	0	1	0	0	.	1	
2875	0.0492696752	0.0566696856	0	5	0	0	.	1	
2876	0.0523532522	0.0629067245	0	2	0	1	.	1	
2877	0.0492696752	0.0566696856	0	5	0	0	.	0	
2878	0.0492696752	0.0566696856	0	2	0	0	.	1	
2879	0.0492696752	0.0566696856	0	5	0	0	1010	4	
2880	0.0492696752	0.0566696856	0	5	0	0	.	0	
2881	0.0492696752	0.0566696856	0	5	0	0	1320	1	
2882	0.0492696752	0.0566696856	0	5	0	0	.	0	

伺服器清單

- 重新整理(R)
- 伺服器
- 私用 OLAP 伺服器

就緒

未選取設定檔

inq2006.sas7bdat

篩選和排序

SAS Enterprise Guide

檔案(F) 編輯(E) 檢視(V) 工作(K) 程式(P) 工具(T) 說明(H) | 處理流程

專案樹狀結構

- 處理流程
 - inq2006
 - 篩選和排序

伺服器清單

- 重新整理(R)
- 伺服器
- 私用 OLAP 伺服器

篩選和排序

輸入資料 | 程式碼 | 記錄檔 | 輸出資料

修改工作(Y) | 篩選和排序(L) | 查詢產生器(O) | 資料(D) | 描述(B) | 圖形(G) | 分析(Z) | 匯出(X) | 傳送至(N)

	SELF_INIT_CNTCTS	SOLICITED_CNTCTS	stuemail	TERRITORY	TOTAL_CONTACTS	TRAVEL_INIT_CNTCTS
2855	7	0	1 6		7	0
2856	7	0	1 6		7	0
2857	5	1	1 6		6	0
2858	2	2	1 6		7	2
2859	2	1	1 6		3	0
2860	4	1	1 6		5	0
2861	5	0	1 6		5	0
2862	7	1	1 6		8	0
2863	3	0	1 6		3	0
2864	4	1	1 8		5	0
2865	3	2	1 6		8	3
2866	9	1	1 6		10	0
2867	2	0	1 6		2	0
2868	8	3	1 6		13	1
2869	5	0	1 6		6	1
2870	0	1	1 6		1	0
2871	5	1	1 8		6	0
2872	3	0	1 6		3	0
2873	1	1	1 6		2	0
2874	1	0	0 6		1	0
2875	1	1	0 8		2	0
2876	1	0	0 6		2	0
2877	0	1	0 6		1	0
2878	1	0	1 6		1	0
2879	4	0	1 6		4	0
2880	0	1	0 8		1	0
2881	1	0	0 6		1	0
2882	0	1	0 8		1	0

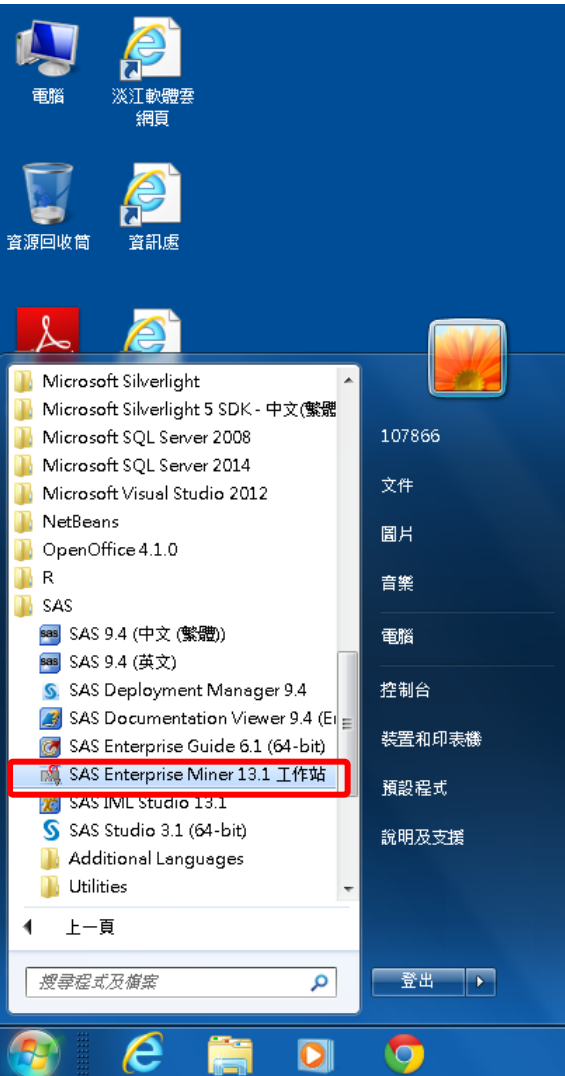
就緒

未選取設定檔

資料欄位說明 (Order by Variable Name)

VarID	Name	Model Role	Measurement Level	Description
1	ACADEMIC_INTEREST_1	Rejected	Nominal	Primary academic interest code
2	ACADEMIC_INTEREST_2	Rejected	Nominal	Secondary academic interest code
3	AVG_INCOME	Input	Interval	Commercial HH income estimate
4	CAMPUS_VISIT	Input	Nominal	Campus visit code
5	CONTACT_CODE1	Rejected	Nominal	First contact code
6	CONTACT_DATE1	Rejected	Nominal	First contact date
7	DISTANCE	Input	Interval	Distance from university
8	ENROLL	Target	Binary	1=Enrolled F2004, 0=Not enrolled F2004
9	ETHNICITY	Rejected	Nominal	Ethnicity
10	HSCRAT	Input	Interval	5-year high school enrollment rate
11	INIT_SPAN	Input	Interval	Time from first contact to enrollment date
12	INSTATE	Input	Binary	1=In state, 0=Out of state
13	INT1RAT	Input	Interval	5-year primary interest code rate
14	INT2RAT	Input	Interval	5-year secondary interest code rate
15	INTEREST	Input	Ordinal	Number of indicated extracurricular interests
16	IRSCHOOL	Rejected	Nominal	High school code
17	LEVEL_YEAR	Rejected	Unary	Student academic level
18	MAILQ	Input	Ordinal	Mail qualifying score (1=very interested)
19	PREMIERE	Input	Binary	1=Attended campus recruitment event, 0=Did not
20	REFERRAL_CNTCTS	Input	Ordinal	Referral contact count
21	SATSCORE	Rejected	Interval	SAT (original) score
22	SELF_INIT_CNTCTS	Input	Interval	Self initiated contact count
23	SEX	Rejected	Binary	Sex
24	SOLICITED_CNTCTS	Input	Ordinal	Solicited contact count
25	STUEMAIL	Input	Binary	1=Have e-mail address, 0=Do not
26	TELECQ	Rejected	Ordinal	Telecounseling qualifying score (1=very interested)
27	TERRITORY	Input	Nominal	Recruitment area
28	TOTAL_CONTACTS	Input	Interval	Total contact count
29	TRAVEL_INIT_CNTCTS	Input	Ordinal	Travel initiated contact count

SAS Enterprise Miner 13.1 (SAS EM)



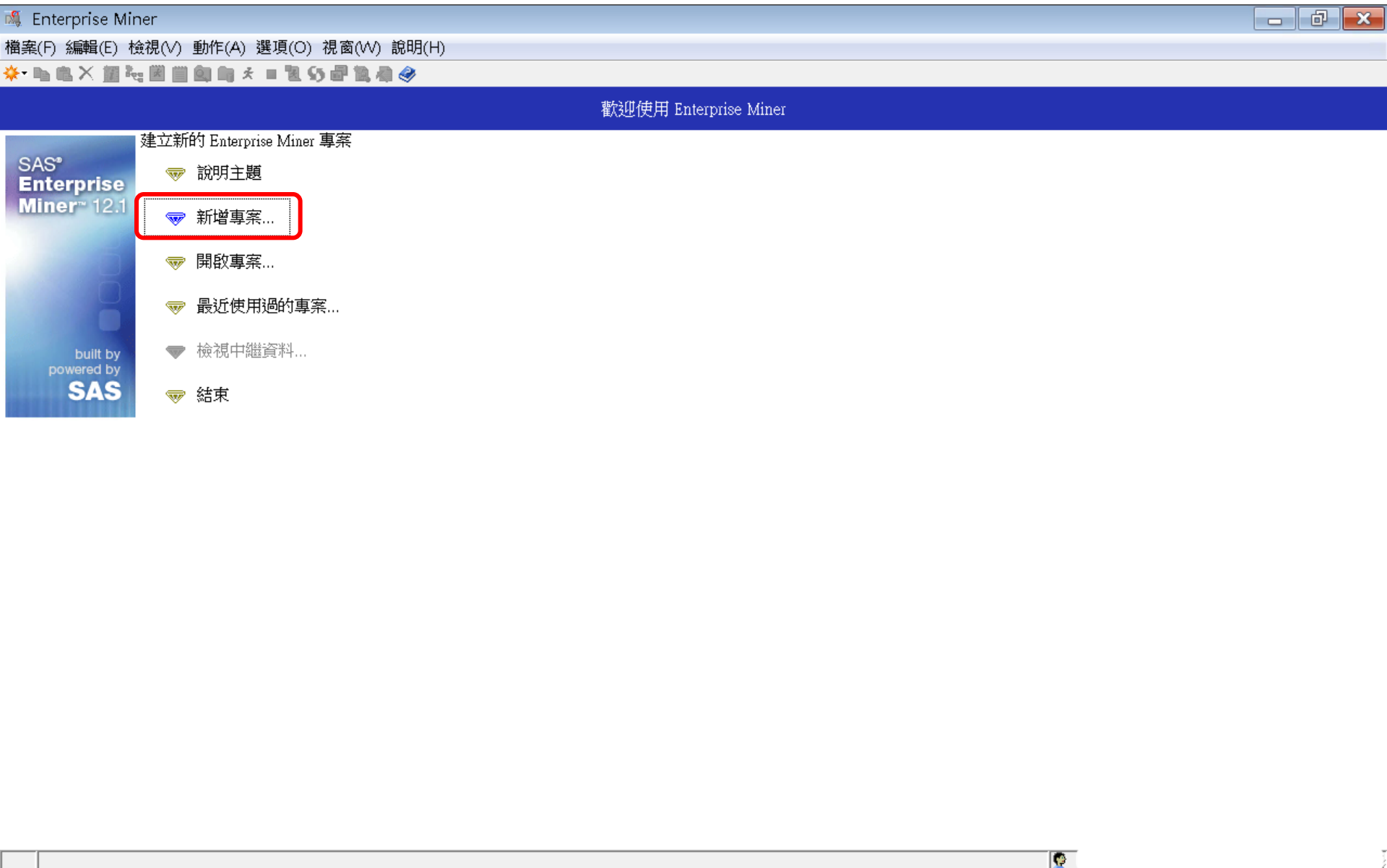
淡江軟體雲使用注意事項

- 登入淡江軟體雲時，會自動產生一部新的雲端主機。登出雲端主機前，請先將檔案儲存於您攜帶的隨身碟、淡江大學網路隨身碟、Google雲端硬碟...等，否則登出後雲端主機的所有資料將會被刪除，並回到您的個人主機畫面。
- 為免雲端主機帳號被冒用，每次使用完請登出。
- 軟體雲相關操作說明，請點選雲端主機桌面上『操作手冊』圖示。
- 軟體雲使用問題請至B212辦公室，或來電26215656-2129 為您服務。
- 請尊重智慧財產權，嚴禁安裝不法軟體，違法者自負相關法律責任。

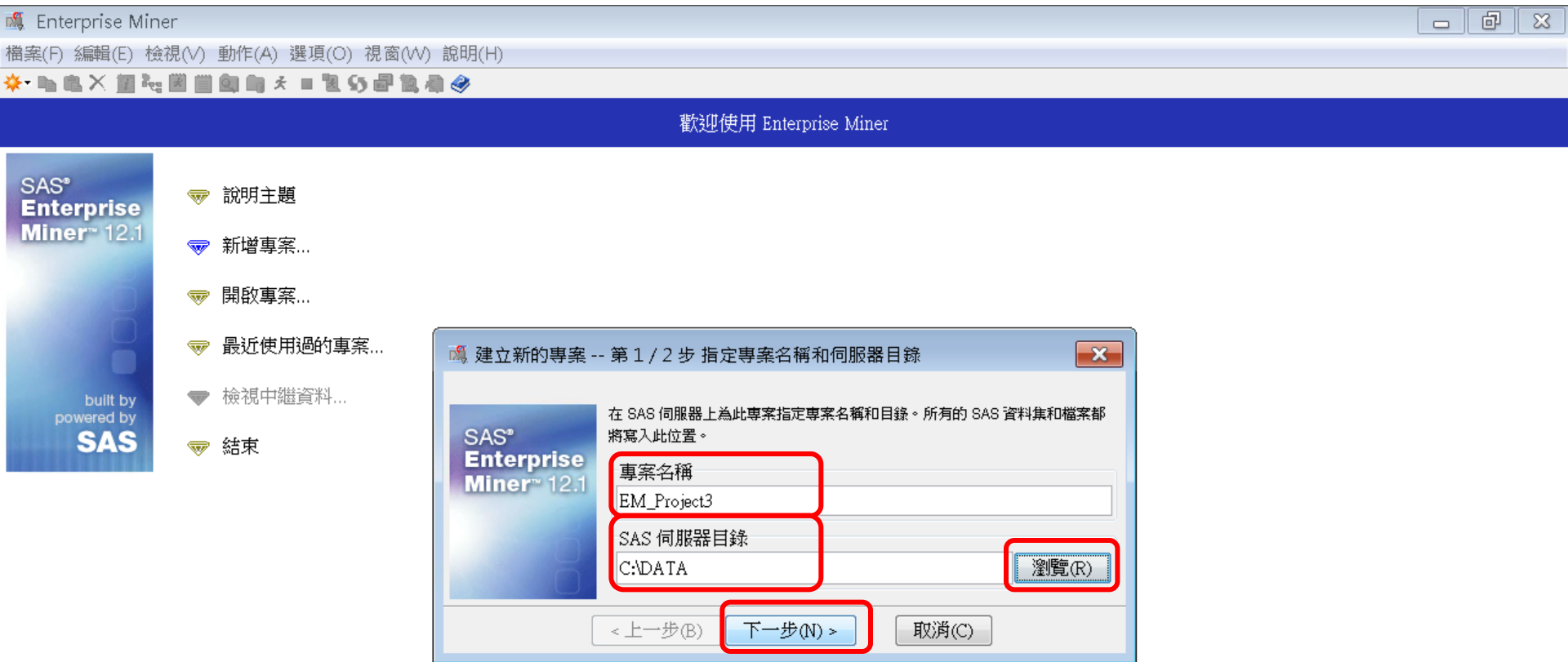
SAS EM 資料匯入4步驟

- Step 1. 新增專案 (New Project)
- Step 2. 新增資料館 (New / Library)
- Step 3. 建立資料來源 (Create Data Source)
- Step 4. 建立流程圖 (Create Diagram)

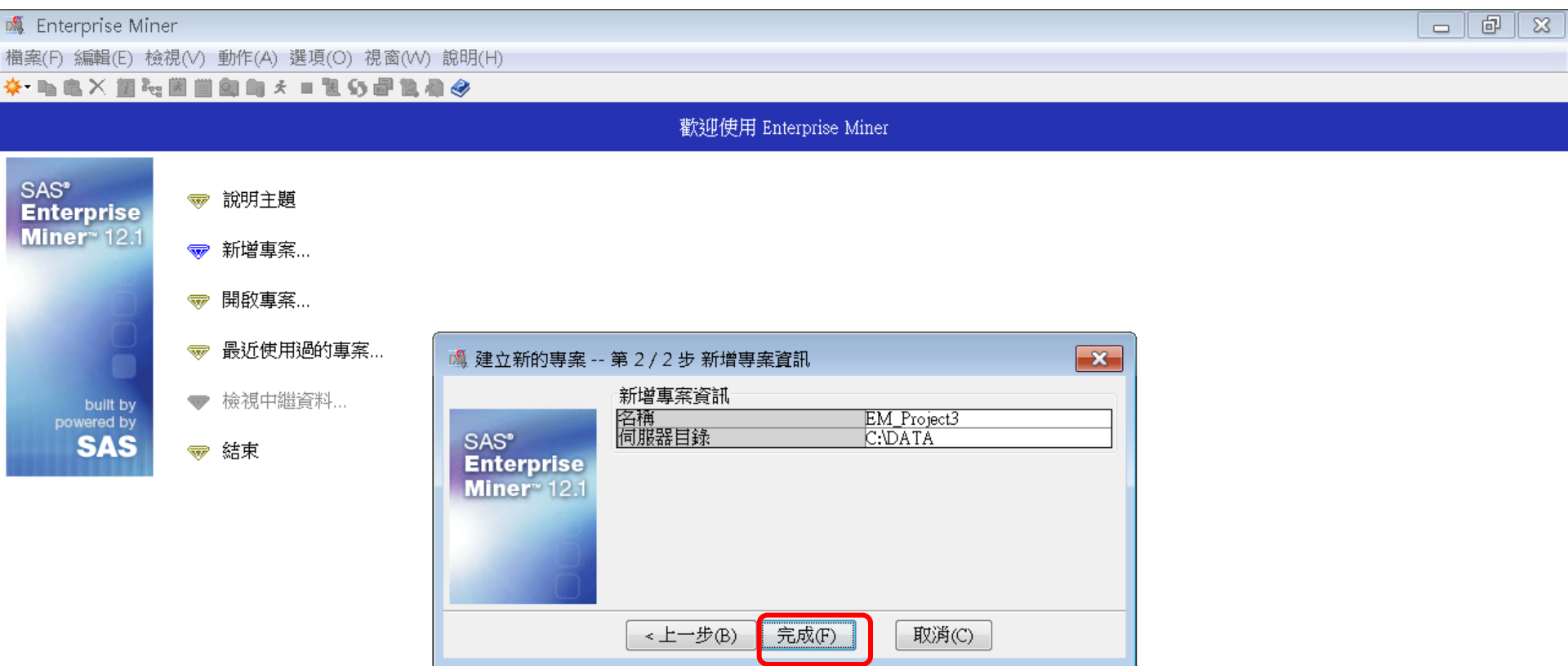
Step 1. 新增專案 (New Project)



Step 1. 新增專案 (New Project)



Step 1. 新增專案 (New Project)



SAS Enterprise Miner (EM_Project3)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3

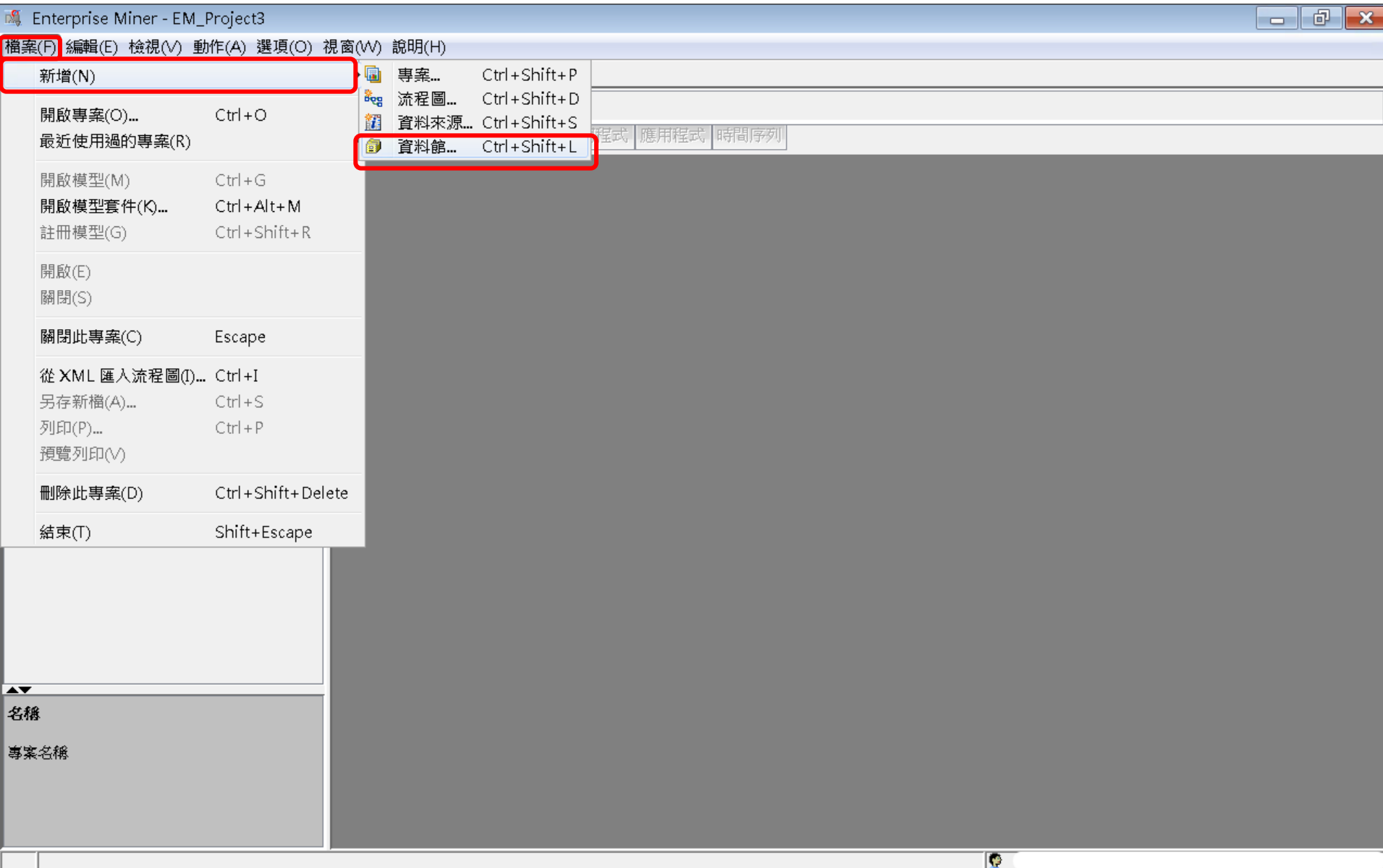
資料來源
流程圖
模型套件

屬性	值
名稱	EM_Project3
專案啟動程式碼	
專案巨集變數	
建立時間	
伺服器	
網格可用	否
路徑	C:\DATA
中繼資料資料夾路徑	
最大並行工作數目	預設

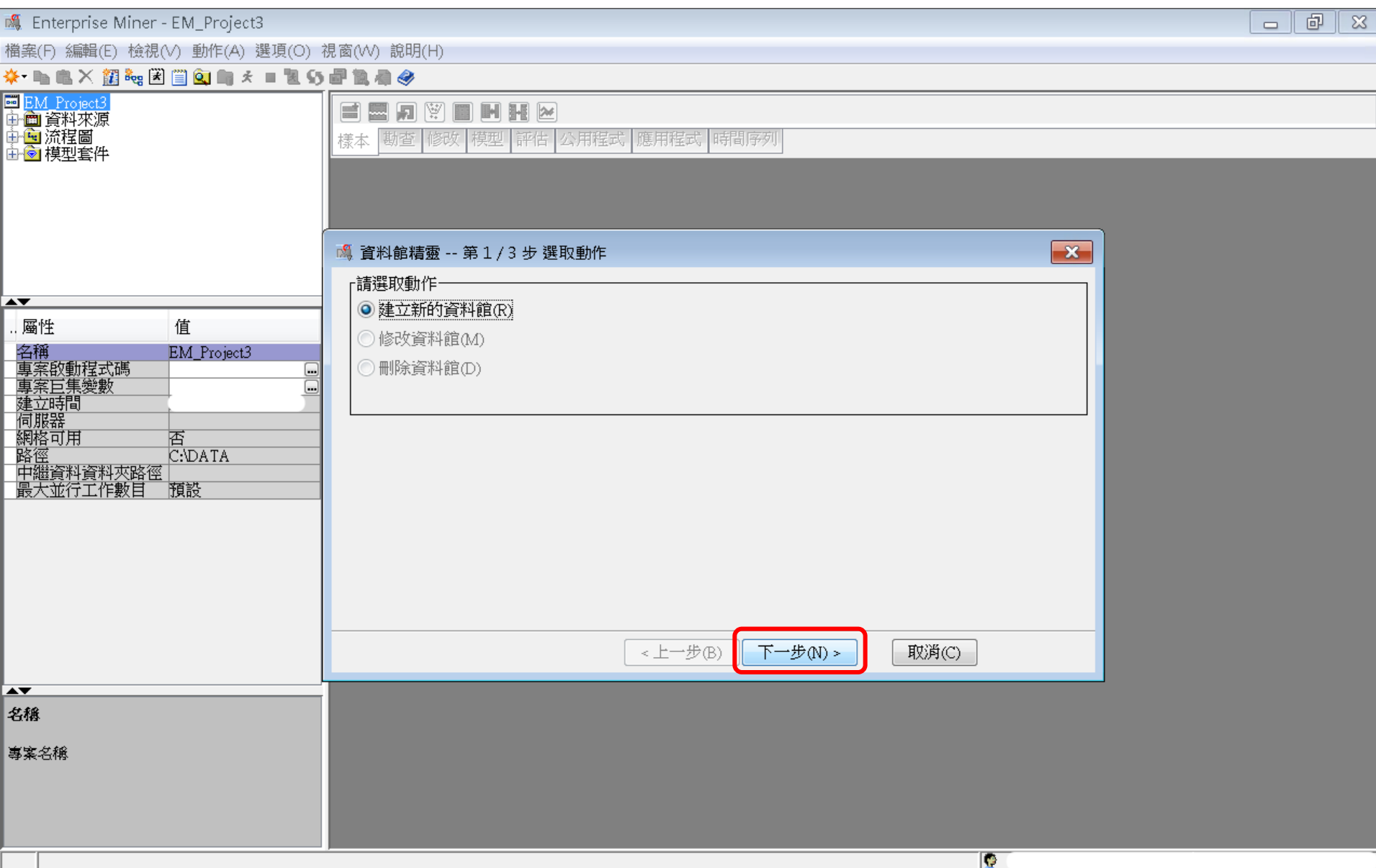
名稱
專案名稱

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

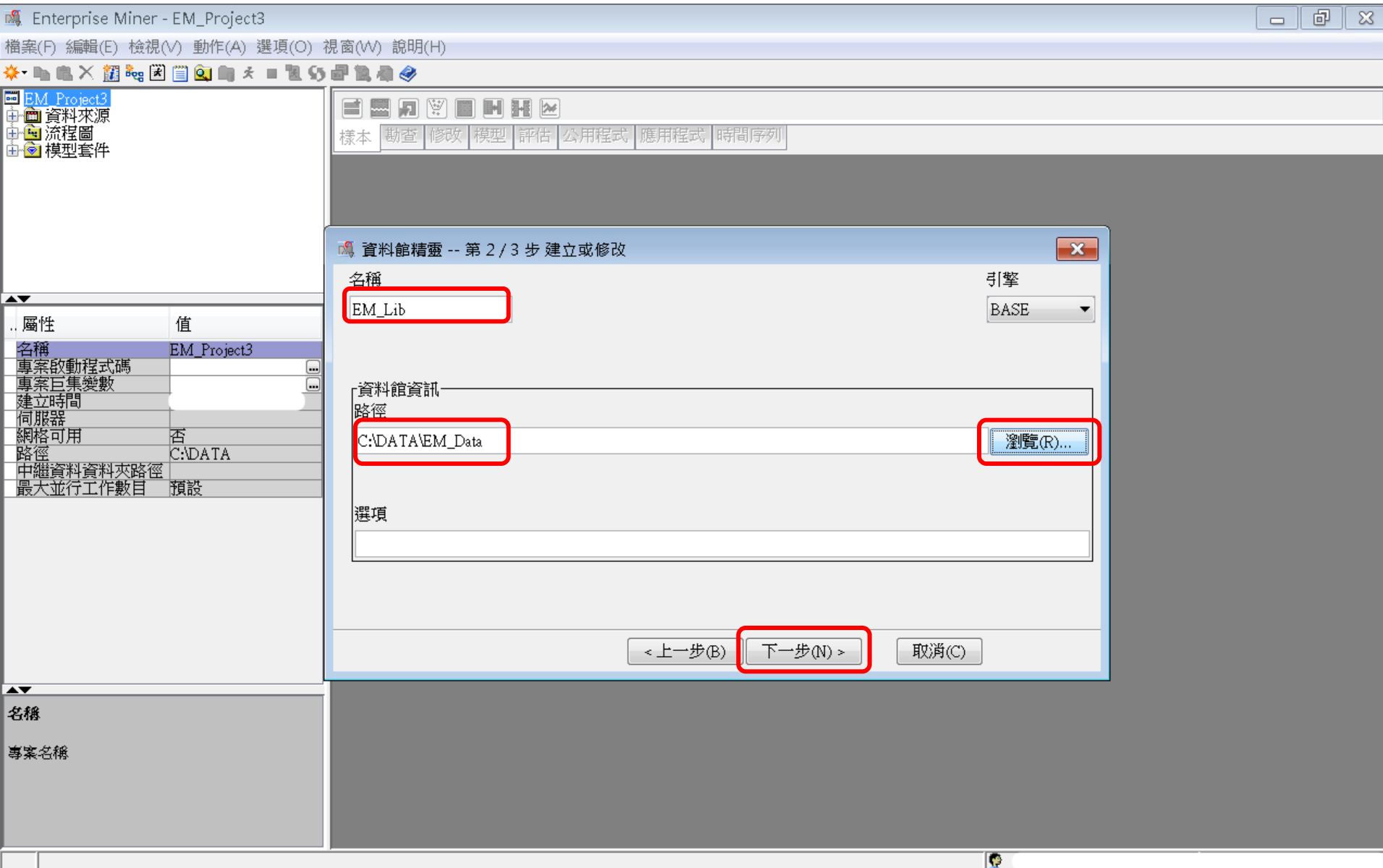
Step 2. 新增資料館 (New / Library)



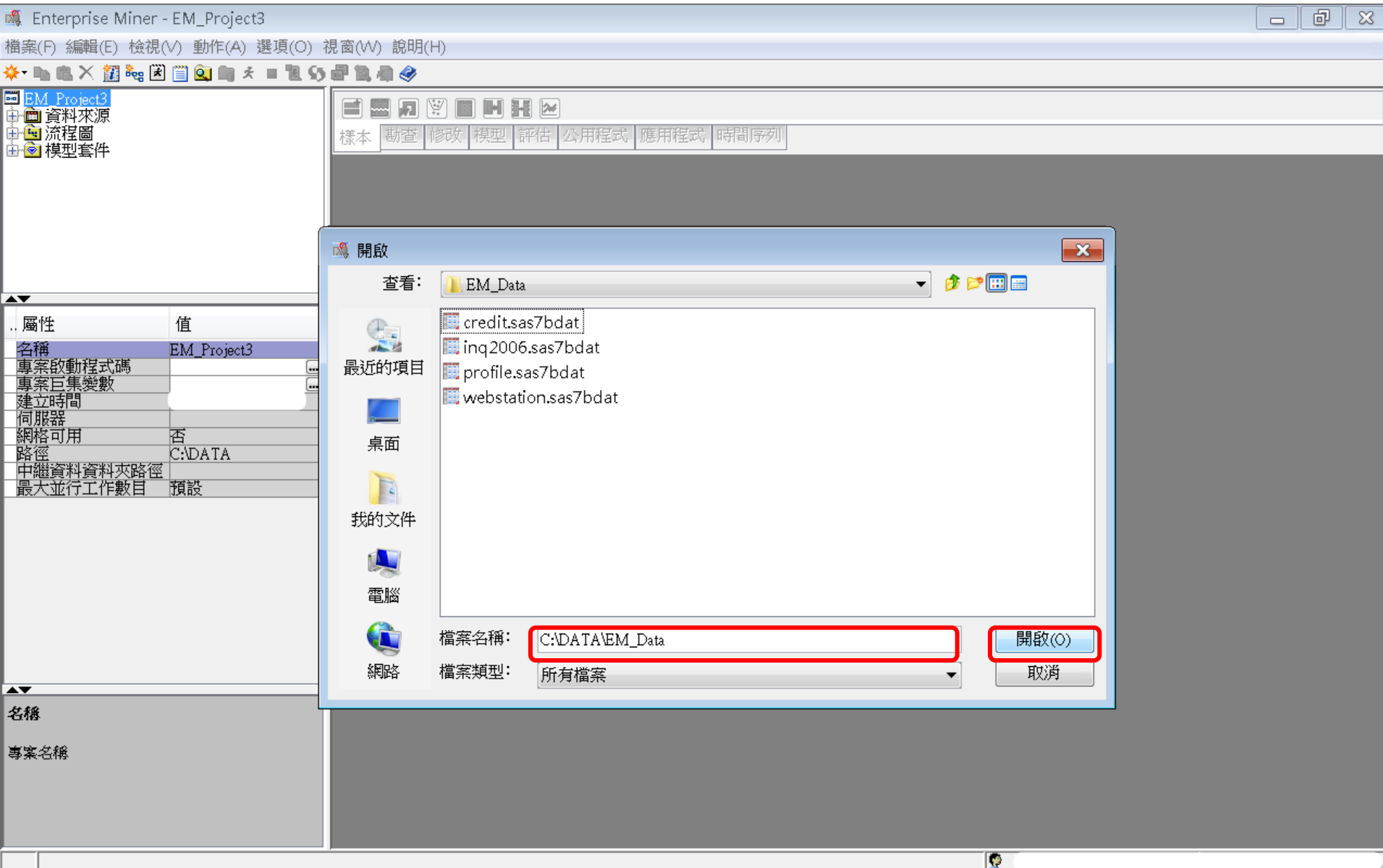
Step 2. 新增資料館 (New / Library)



Step 2. 新增資料館 (New / Library)



Step 2. 新增資料館 (New / Library)



Step 2. 新增資料館 (New / Library)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3

- 資料來源
- 流程圖
- 模型套件

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

資料館精靈 -- 第 3 / 3 步 確認動作

屬性	值
動作	建立新的
名稱	EM_Lib
引擎	BASE
路徑	C:\DATA\EM_Data
選項	

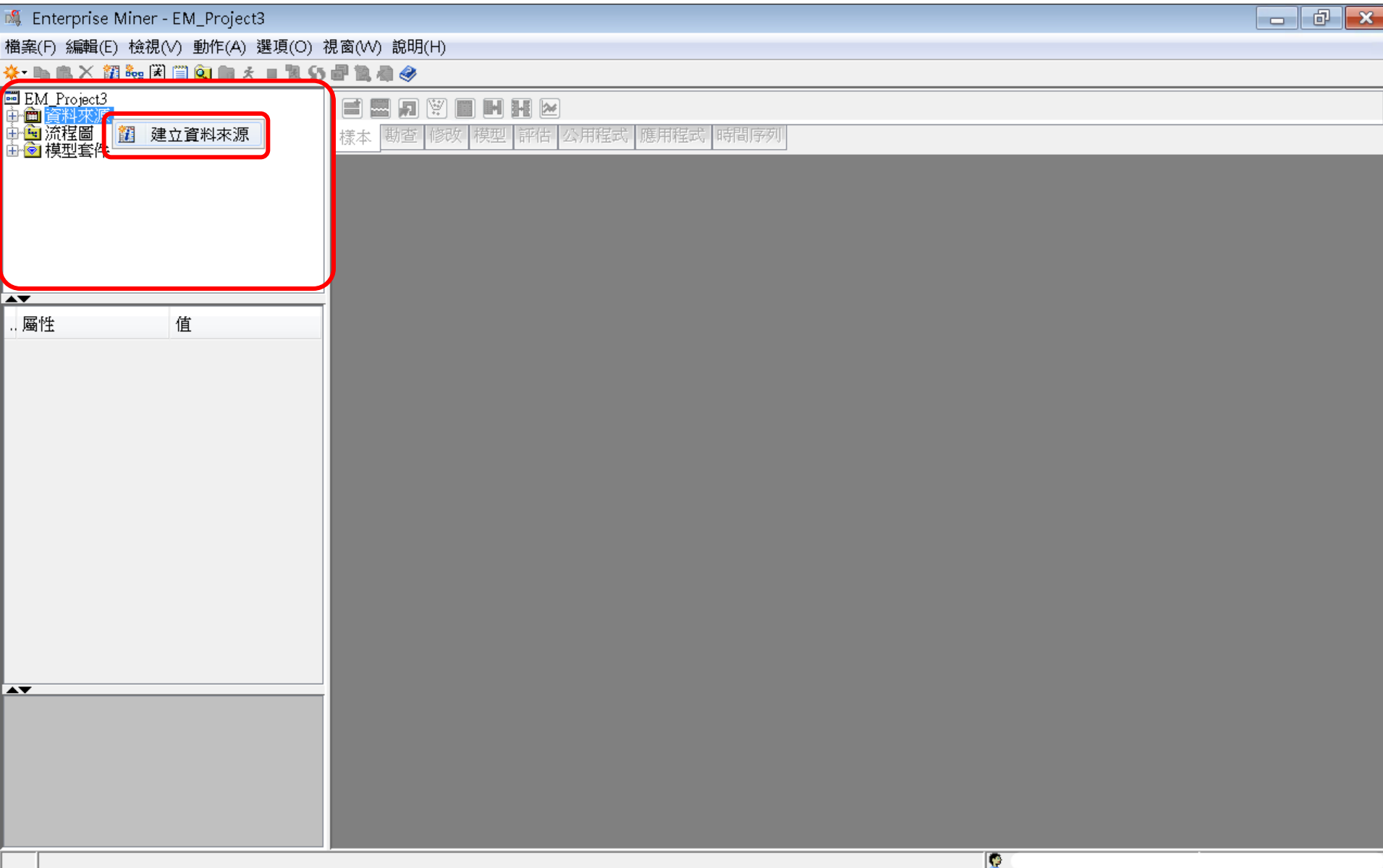
狀態
動作成功!
資料館「EM_Lib」已建立。

< 上一步(B) 完成(F)

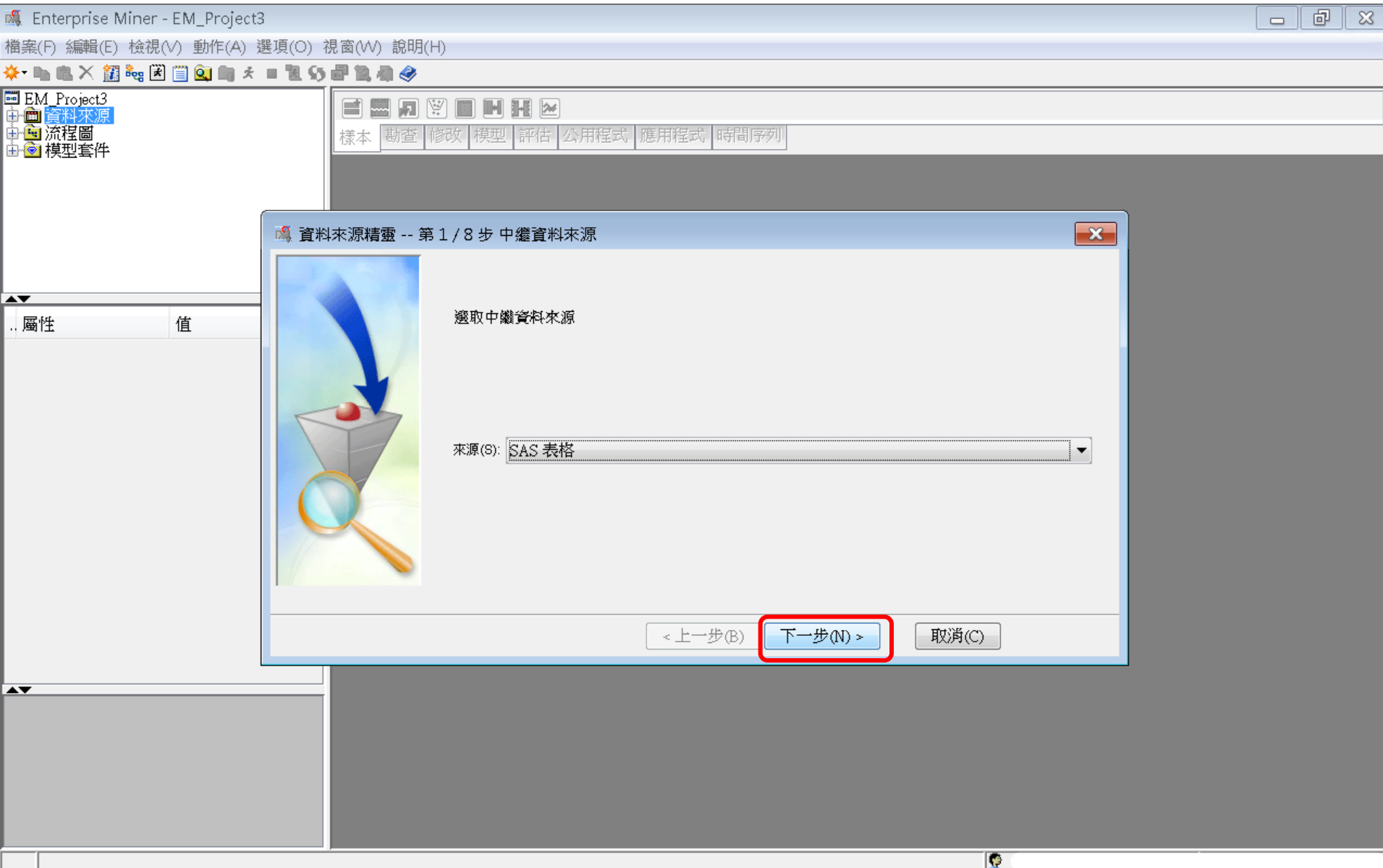
名稱

專案名稱

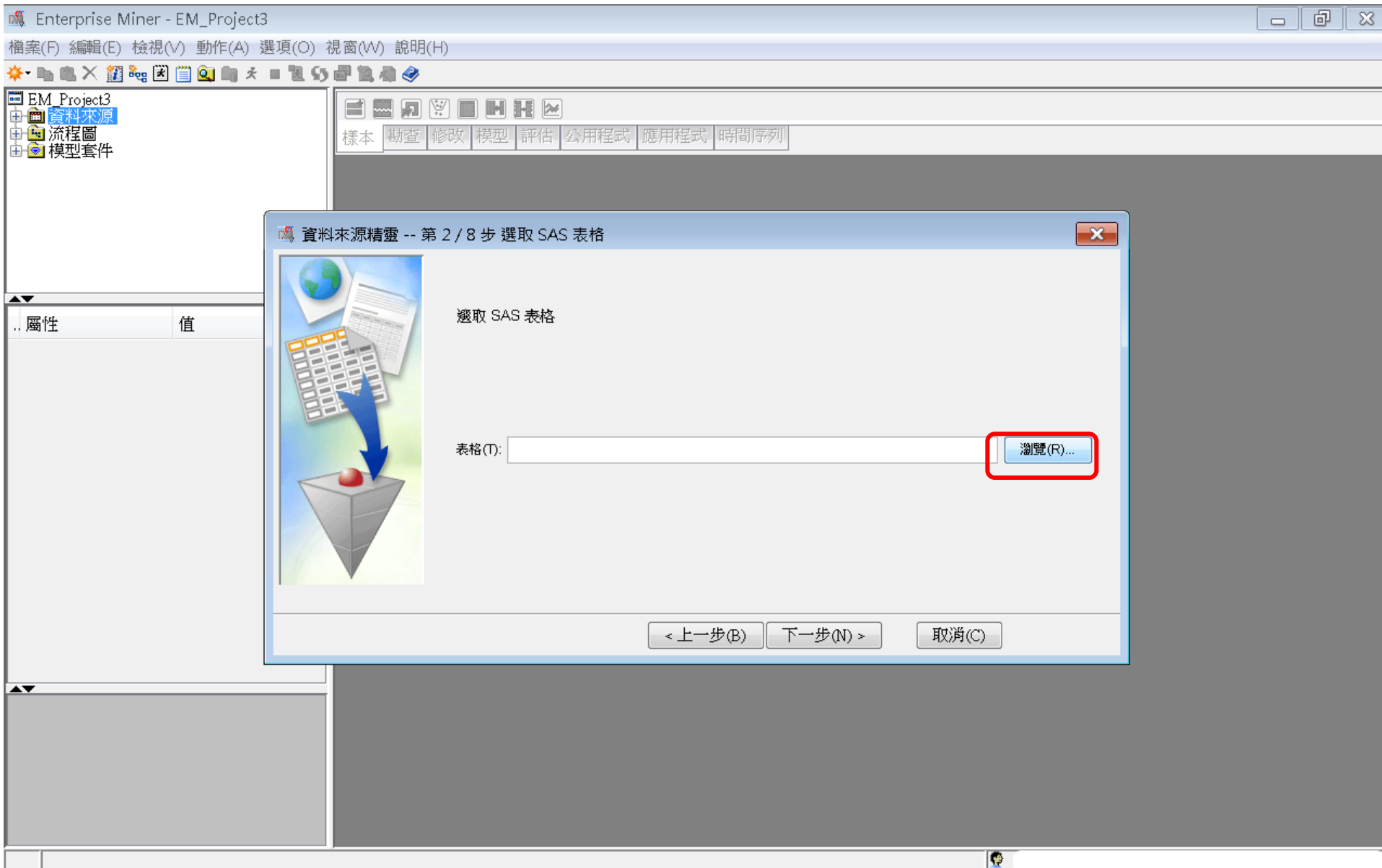
Step 3. 建立資料來源 (Create Data Source)



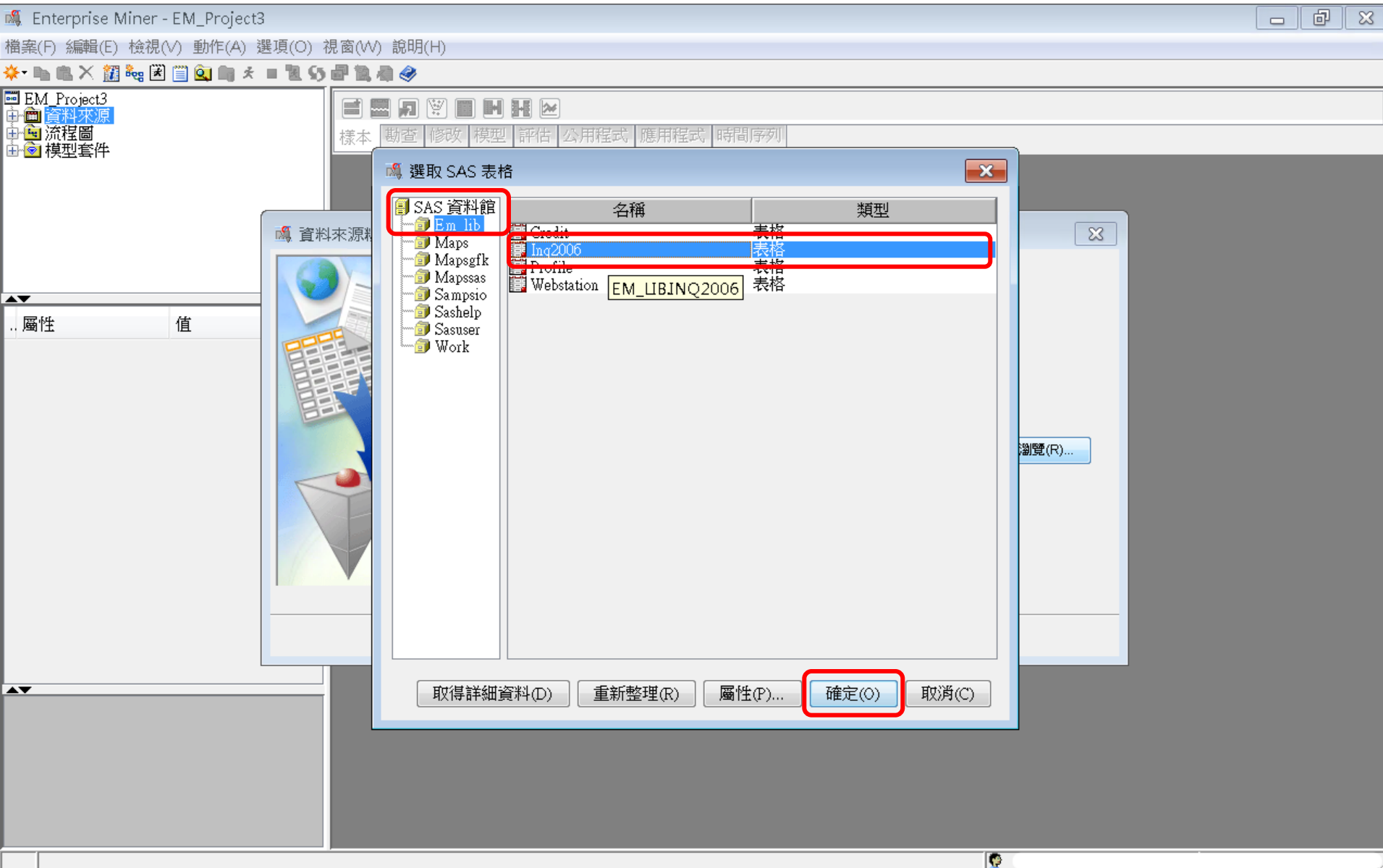
Step 3. 建立資料來源 (Create Data Source)



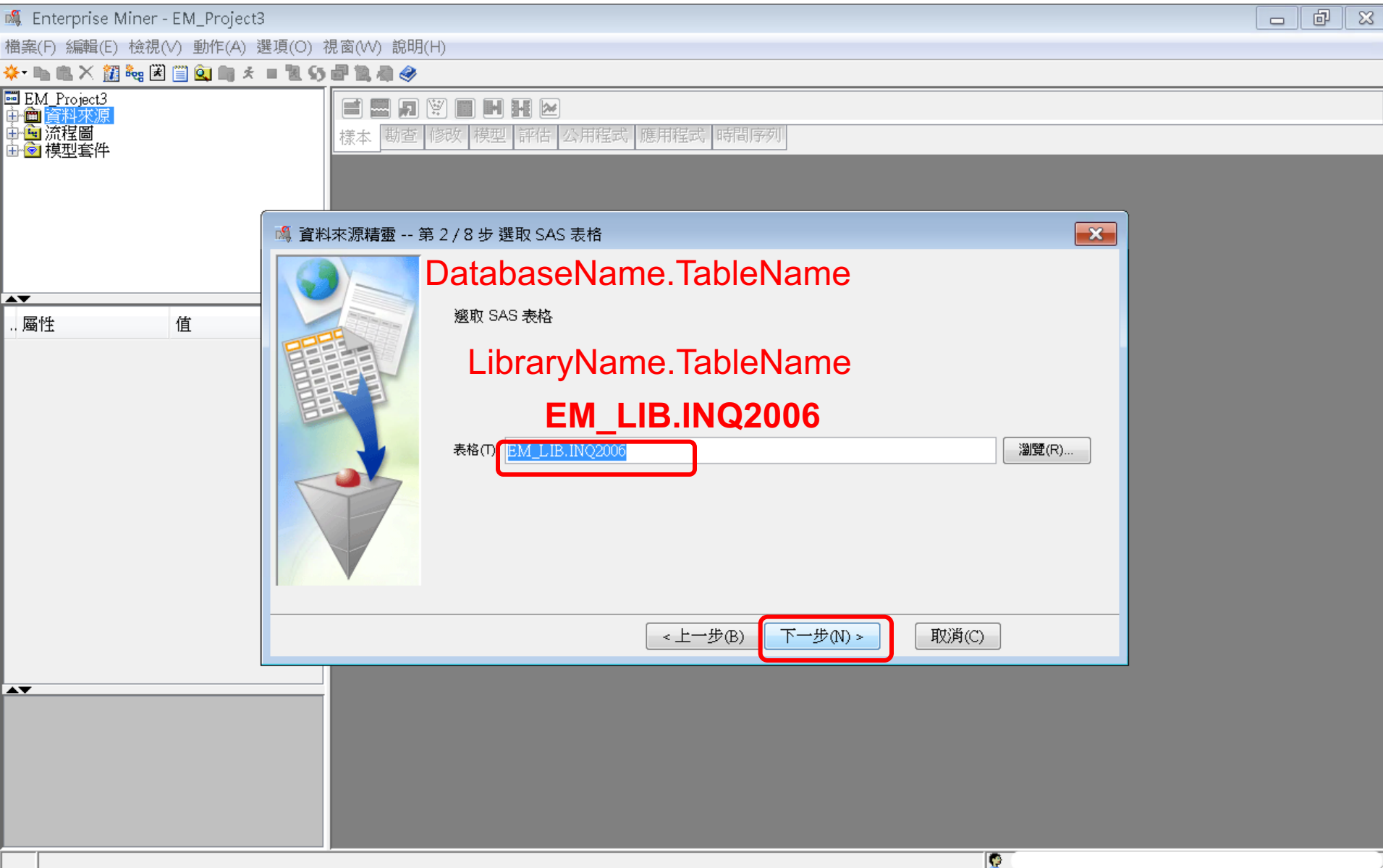
Step 3. 建立資料來源 (Create Data Source)



Step 3. 建立資料來源 (Create Data Source)



Step 3. 建立資料來源 (Create Data Source)



Step 3. 建立資料來源 (Create Data Source)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM Project3

- 資料來源
- 流程圖
- 模型套件

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

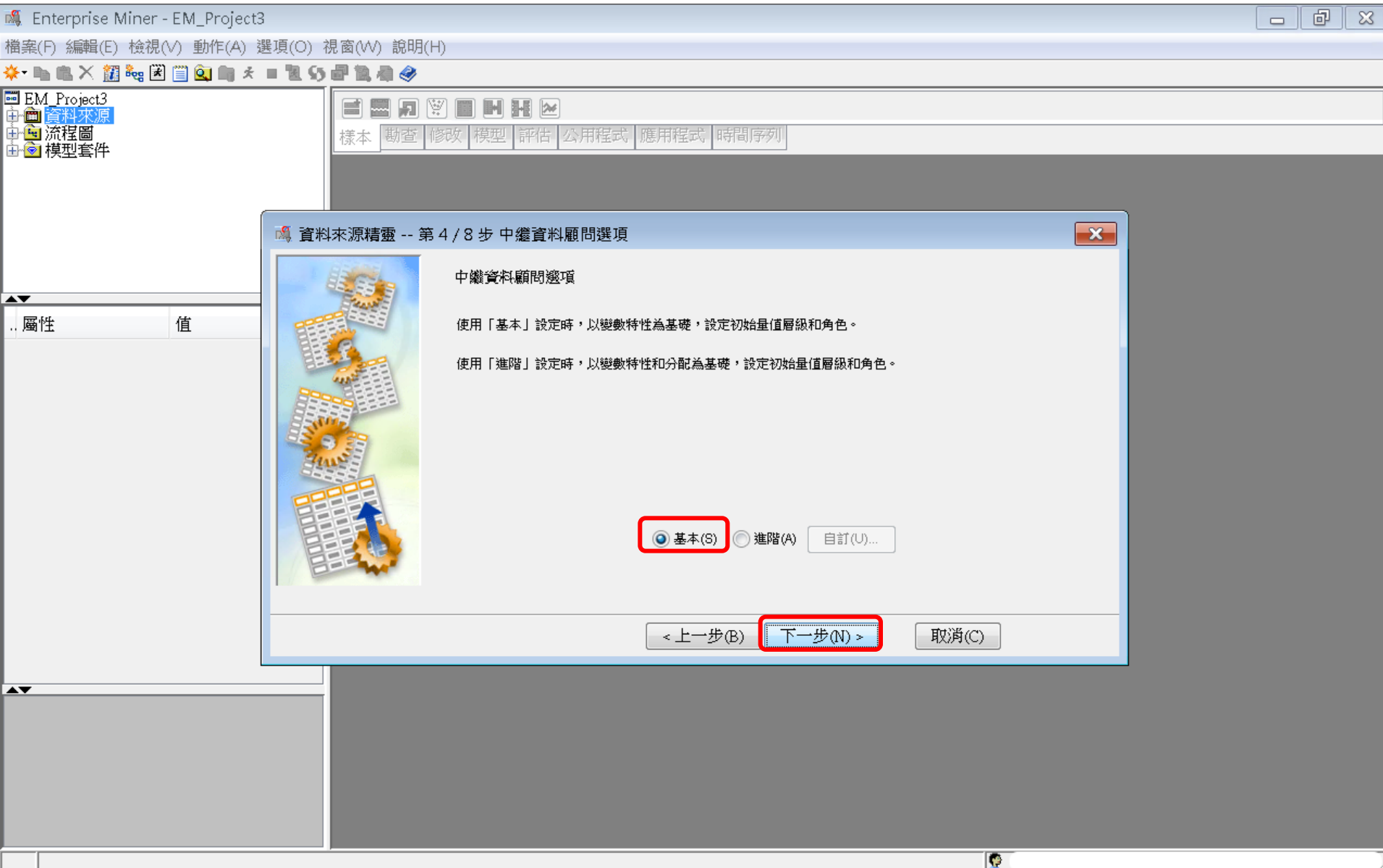
資料來源精靈 -- 第 3 / 8 步 表格資訊

表格屬性

屬性	值
表格名稱	EM_LIB.INQ2006
描述	
成員類型	DATA
資料集類型	DATA
引擎	BASE
變數的數目	20
觀測數目	91482
建立日期	2013年5月29日 下午01時20分20秒
修改日期	2013年5月29日 下午01時20分20秒

< 上一步(B) **下一步(N) >** 取消(C)

Step 3. 建立資料來源 (Create Data Source)



Step 3. 建立資料來源 (Create Data Source)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3
資料來源
流程圖
模型套件

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

資料來源精靈 -- 第 5 / 8 步 欄中繼資料

(無) ☐ 非 等於

欄: ☐ 標籤(A) ☐ 採礦(M) ☐ 基本(I) ☐ 統計(T)

名稱	角色	層級	報表	順序	刪除	下限	上
avg_income	輸入	間隔	否		否	.	
CAMPUS_VI	輸入	間隔	否		否	.	
distance	輸入	間隔	否		否	.	
Enroll_Target	輸入	間隔	否		否	.	
hscrat	輸入	間隔	否		否	.	
init_span	輸入	間隔	否		否	.	
lnstate	輸入	名目	否		否	.	
intlrat	輸入	間隔	否		否	.	
int2rat	輸入	間隔	否		否	.	
interest	輸入	間隔	否		否	.	
mailq	輸入	間隔	否		否	.	
premiere	輸入	間隔	否		否	.	
REFERRAL	輸入	間隔	否		否	.	

Step 3. 建立資料來源 (Create Data Source)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3

- 資料來源
- 流程圖
- 模型套件

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

資料來源精靈 -- 第 5 / 9 步 欄中繼資料

(無) ☐ 非 等於

套用 重新設定

☐ 標籤(A) ☐ 採礦(M) ☐ 基本(I) ☐ 統計(T)

名稱	角色	層級	報表	順序	刪除	下限	上限
avg_income	輸入	間隔	否		否	.	.
CAMPUS_VISIT	輸入	名目	否		否	.	.
distance	輸入	間隔	否		否	.	.
Enroll_Target	目標	二元	否		否	.	.
hscrat	輸入	間隔	否		否	.	.
init_span	輸入	間隔	否		否	.	.
Instate	輸入	二元	否		否	.	.
intlrat	輸入	間隔	否		否	.	.
int2rat	輸入	間隔	否		否	.	.
interest	輸入	間隔	否		否	.	.
mailq	輸入	序數	否		否	.	.
premiere	輸入	二元	否		否	.	.
REFERRAL_CNTCTS	輸入	間隔	否		否	.	.
satscore	輸入	間隔	否		否	.	.
SELF_INIT_CNTCTS	輸入	間隔	否		否	.	.
SOLICITED_CNTCTS	輸入	間隔	否		否	.	.
stuemail	輸入	二元	否		否	.	.
TERRITORY	輸入	名目	否		否	.	.
TOTAL_CONTACTS	輸入	間隔	否		否	.	.
TRAVEL_INIT_CNTCTS	輸入	間隔	否		否	.	.

顯示程式碼(S) 勘查(E) 重新整理摘要(U) < 上一步(B) 下一步(N) > 取消(C)

1. 變數角色調整:

將 Enroll_Target 的變數角色設為 Target

2. 資料型態 (層級) 修改:

將 CAMPUS_VISIT 資料型態改為 Nominal
將 Enroll_Target 資料型態改為 Binary
將 Instate 資料型態改為 Binary
將 Mailq 資料型態改為 Ordinary
將 Premiere 資料型態改為 Binary
將 Stuemail 資料型態改為 Binary
將 TERRITORY 資料型態改為 Nominal

Step 3. 建立資料來源 (Create Data Source)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM Project3

- 資料來源
- 流程圖
- 模型套件

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

資料來源精靈 -- 第 5 / 9 步 欄中繼資料

欄: ☐ 標籤(A) ☐ 採礦(M) ☐ 基本(I) ☐ 統計(T)

名稱	角色	層級	報表	順序	刪除	下限	上限
avg_income	輸入	間隔	否		否	.	.
CAMPUS_VISIT	輸入	名目	否		否	.	.
distance	輸入	間隔	否		否	.	.
Enroll_Target	目標	二元	否		否	.	.
hscrat	輸入	間隔	否		否	.	.
init_span	輸入	間隔	否		否	.	.
lnstate	輸入	二元	否		否	.	.
int1rat	輸入	間隔	否		否	.	.
int2rat	輸入	間隔	否		否	.	.
interest	輸入	間隔	否		否	.	.
mailq	輸入	間隔	否		否	.	.
premiere	輸入	間隔	否		否	.	.
REFERRAL_CNTCTS	輸入	間隔	否		否	.	.
satscore	輸入	間隔	否		否	.	.
SELF_INIT_CNTCTS	輸入	間隔	否		否	.	.
SOLICITED_CNTCTS	輸入	間隔	否		否	.	.
stuemail	輸入	間隔	否		否	.	.
TERRITORY	輸入	名目	否		否	.	.
TOTAL_CONTACTS	輸入	間隔	否		否	.	.
TRAVEL_INIT_CNTCTS	輸入	間隔	否		否	.	.

摘要統計

是否要重新整理摘要統計?

是(Y) 否(N)

顯示程式碼(S) 勘查(E) 重新整理摘要(U) < 上一步(B) 下一步(N) > 取消(C)

Step 3. 建立資料來源 (Create Data Source)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3
資料來源
流程圖
模型套件

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

資料來源精靈 -- 第 5 / 9 步 欄中繼資料

(無) ☐ 非 等於

套用 重新設定

欄: ☐ 標籤(A) ☐ 採礦(M) ☐ 基本(I) ☒ 統計(T)

名稱	角色	層級	報表	順序	刪除	下限	上限	層級數目	遺漏值百分比	最小值	最大值	平均值	標準差	偏態
avg_income	輸入	間隔	否		否	.	.	.	22.87772	4940	200001	47315.33	20608.89	1.258231
CAMPUS_VISIT	輸入	名目	否		否	.	.	3	0	.	.	.	.	.
distance	輸入	間隔	否		否	.	.	.	21.28069	0.417124	4798.899	380.4276	397.9788	2.276541
Enroll_Target	目標	二元	否		否	.	.	2	0	.	.	.	.	.
hscrat	輸入	間隔	否		否	.	.	.	0	0	1	0.037652	0.057399	7.021978
init_span	輸入	間隔	否		否	.	.	.	0	-216	228	19.68616	8.722109	0.758461
lnstate	輸入	二元	否		否	.	.	2	0	.	.	.	.	.
intlrat	輸入	間隔	否		否	.	.	.	0	0	1	0.037091	0.024026	3.496845
intl2rat	輸入	間隔	否		否	.	.	.	0	0	1	0.042896	0.025244	3.215683
interest	輸入	間隔	否		否	.	.	.	0	0	3	0.053639	0.242052	4.817932
mailq	輸入	序數	否		否	.	.	5	0	.	.	.	.	.
premiere	輸入	二元	否		否	.	.	2	0	.	.	.	.	.
REFERRAL_CNTCTS	輸入	間隔	否		否	.	.	.	0	0	5	0.039899	0.226191	7.572001
satscore	輸入	間隔	否		否	.	.	.	70.48272	420	1600	1072.993	180.8305	-0.16183
SELF_INIT_CNTCTS	輸入	間隔	否		否	.	.	.	0	0	56	1.214119	1.666529	2.916263
SOLICITED_CNTCTS	輸入	間隔	否		否	.	.	.	0	0	9	0.547222	0.649426	1.201013
stuemail	輸入	二元	否		否	.	.	2	0	.	.	.	.	.
TERRITORY	輸入	名目	否		否	.	.	11	0.001093	.	.	.	.	.
TOTAL_CONTACTS	輸入	間隔	否		否	.	.	.	0	1	58	2.166098	1.852537	3.062389
TRAVEL_INIT_CNTCTS	輸入	間隔	否		否	.	.	.	0	0	8	0.364815	0.557067	1.418514

顯示程式碼(S) 勘查(E) 重新整理摘要(U) < 上一步(B) 下一步(N) > 取消(C)

Step 3. 建立資料來源 (Create Data Source)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM Project3

- 資料來源
- 流程圖
- 模型套件

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

資料來源精靈 -- 第 6 / 10 步 決策設定

決策處理

是否要根據決策值建置模型?

如果選擇「是」，您可以輸入每個可能決策的成本或利潤、事前機率和成本函數等資訊。將會掃描此資料，取得目標變數的分配。

☒ 否(O) ☐ 是(Y)

< 上一步(B) **下一步(N) >** 取消(C)

Step 3. 建立資料來源 (Create Data Source)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM Project3

- 資料來源
- 流程圖
- 模型套件

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

資料來源精靈 -- 第 7 / 9 步 建立樣本

是否要建立樣本資料集?
☒ 否(O) ☐ 是(Y)

表格資訊

欄	20
列	91482

樣本大小

類型 百分比

百分比 20

列

< 上一步(B) 下一步(N) > 取消(C)

Step 3. 建立資料來源 (Create Data Source)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3

- 資料來源
- 流程圖
- 模型套件

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

資料來源精靈 -- 第 8 / 9 步 資料來源特性

您可以變更名稱和角色，還可以為要建立的資料來源指定母體區段識別碼。

Data Source Attribute Role: Raw

名稱(E): INQ2006

角色(R): 原始

區段(S):

附註(O):

< 上一步(B) 下一步(N) > 取消(C)

Step 3. 建立資料來源 (Create Data Source)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM Project3

- 資料來源
- 流程圖
- 模型套件

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

資料來源精靈 -- 第 9 / 9 步 摘要

中繼資料已完成。

資料館: EM_LIB
資料來源: INQ2006
角色: 原始

角色	層級	計數
輸入	二元	3
輸入	間隔	13
輸入	名目	2
輸入	序數	1
目標	二元	1

< 上一步(B) **完成(F)** 取消(C)

Step 3. 建立資料來源 (Create Data Source)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

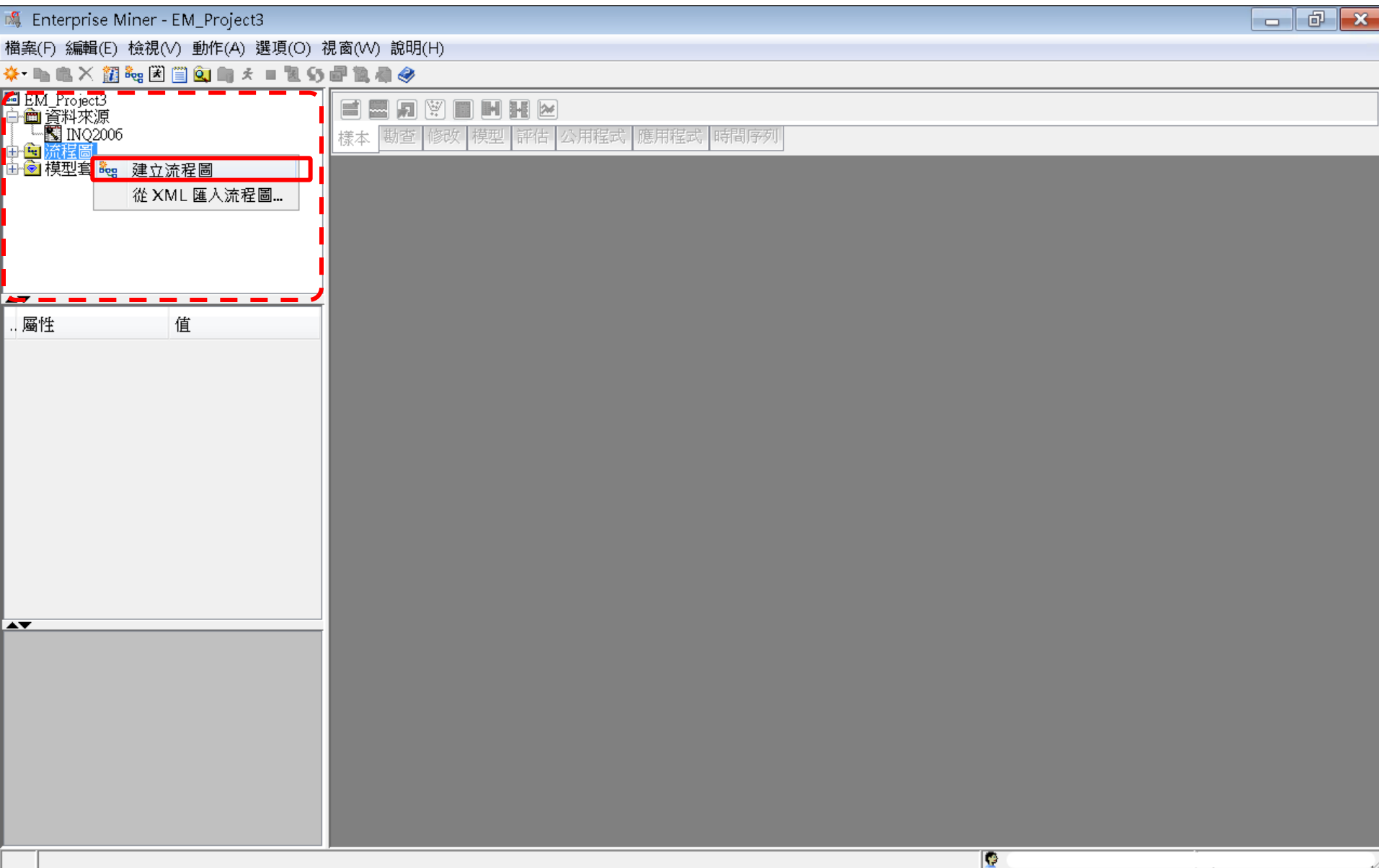
樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

屬性	值
ID	inq
名稱	INQ2006
變數	...
決策	...
角色	原始
附註	...
資料館	EM_LIB
表格	INQ2006
樣本資料集	
大小類型	
樣本大小	
類型	DATA
觀測數目	91482
欄數	20
位元組數	14058496
區段	

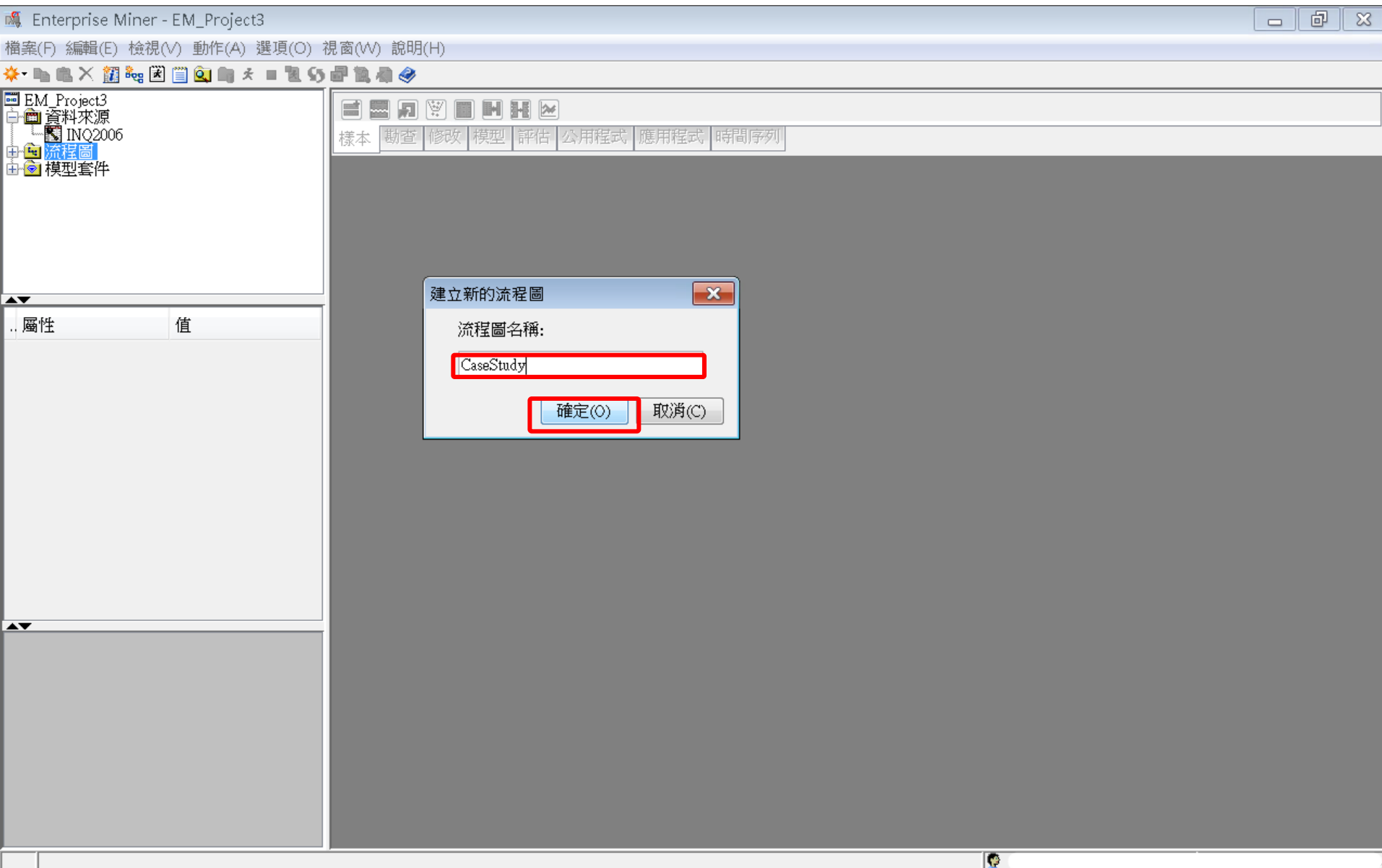
ID

資料來源識別碼。與資料來源關聯的中繼資料表格儲存在 EMDS SAS 資料館中，並使用識別碼作為這些表格名稱的前置碼。

Step 4. 建立流程圖 (Create Diagram)



Step 4. 建立流程圖 (Create Diagram)



Step 4. 建立流程圖 (Create Diagram)

The screenshot displays the Enterprise Miner software interface. The main window is titled "Enterprise Miner - EM_Project3". The left sidebar shows a project tree with "EM_Project3" expanded, containing "資料來源" (Data Sources), "INQ2006", "流程圖" (Flow Diagrams), "CaseStudy", and "模型套件" (Model Kits). The "CaseStudy" flow diagram is selected and highlighted with a red dashed border. Below the project tree is a table showing the properties of the selected flow diagram.

屬性	值
ID	EMWS1
名稱	CaseStudy
狀態	開啟
附註	
歷程記錄	
編碼	ms-950 Traditional Chi
資料表示法	WINDOWS_32
原生作業系統	是

Below the table, the ID "ID" is displayed, followed by a description: "流程圖識別碼。此識別碼和 SAS 資料館名稱對應，SAS 資料館名稱是用於識別此流程圖內容在伺服器上的實體位置。"

The main workspace is titled "CaseStudy" and is currently empty. The bottom status bar shows "流程圖「CaseStudy」已開啟" (Flow Diagram "CaseStudy" is open).

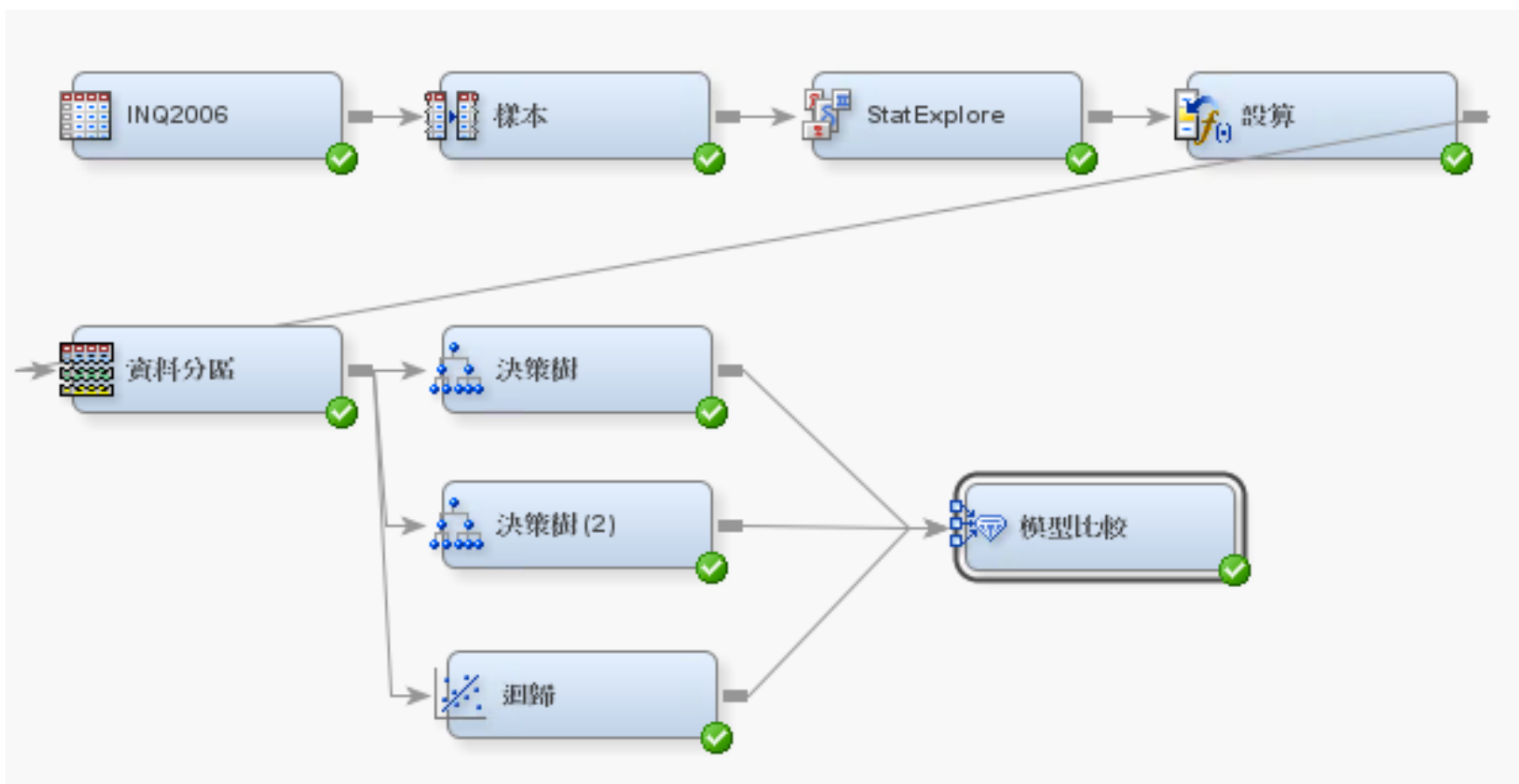
SAS Enterprise Miner (SAS EM) Case Study

- SAS EM 資料匯入4步驟

- Step 1. 新增專案 (New Project)
- Step 2. 新增資料館 (New / Library)
- Step 3. 建立資料來源 (Create Data Source)
- Step 4. 建立流程圖 (Create Diagram)

- SAS EM SEMMA 建模流程

案例情境模型流程



樣本資料匯入 (Sample)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3
資料來源
INQ2006
加註圖
CaseStudy
模型套件

屬性	值
一般	
節點 ID	Ids
匯入的資料	...
匯出的資料	...
附註	...
訓練	
輸出類型	檢視
角色	原始
重新執行	否
彙總	否
刪除對應變數	是
變數	
變數	...
決策	...
重新整理中繼資料	...
顧問	基本
進階選項	...
資料	
資料選取	資料來源
樣本	預設
樣本選項	...
資料來源	
資料來源	INQ2006
資料來源屬性	...
新增表格	
表格名稱	...
驗證驗證	嚴格

CaseStudy

樣本 INQ2006

流程圖 日誌

流程圖「CaseStudy」已開啟

EM_LIB.INQ2006

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3

- 資料來源
 - INQ2006
- 流程圖
- CaseStudy
- 模型套件

屬性 值

一般

節點 ID Ids

匯入的資料

匯出的資料

附註

訓練

輸出類型 檢視

角色 原始

重新執行 否

彙總 否

刪除對應變數 是

變數

決策

重新整理中繼資料

顧問 基本

進階選項

資料

資料選取 資料來源

樣本 預設

樣本選項

資料來源

資料來源 INQ2006

資料來源屬性

新增表格

表格名稱

變數命途

匯入的資料

節點匯入的表格集

CaseStudy

INQ2006

匯入的資料 - INQ2006

連接埠	來源	表格
DATA		EM_LIB.INQ2006

瀏覽(B)...

EM_LIB.INQ2006

	RECR_CODE	Enroll_Target	TOTAL_CONTACTS	SELF_INIT_CNI
1	5	1.0	4.0	4.0
2	8	1.0	6.0	6.0
3	5	1.0	4.0	4.0
4	5	1.0	3.0	3.0
5	5	1.0	10.0	8.0
6	2	1.0	5.0	4.0
7	7	1.0	7.0	6.0
8	5	1.0	16.0	14.0
9	5	1.0	7.0	6.0
10	6	1.0	14.0	10.0
11	8	1.0	11.0	6.0
12	6	1.0	9.0	7.0
13	6	1.0	13.0	10.0
14	8	1.0	8.0	6.0
15	6	1.0	5.0	4.0
16	6	1.0	4.0	3.0
17	6	1.0	6.0	6.0
18	6	1.0	8.0	8.0
19	6	1.0	12.0	11.0
20	6	1.0	7.0	7.0
21	6	1.0	10.0	9.0
22	5	1.0	11.0	9.0
23	5	1.0	11.0	8.0
24	6	1.0	4.0	3.0
25	8	1.0	9.0	9.0
26	6	1.0	6.0	6.0

流程圖 日誌

流程圖「CaseStudy」已開啟

樣本 (Sample)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3
資料來源
INQ2006
流程圖
CaseStudy
模型套件

屬性	值
一般	
節點 ID	Smpl
匯入的資料	
匯出的資料	
附註	
訓練	
變數	
輸出類型	資料
抽樣方法	預設
隨機種子	12345
大小	
類型	百分比
觀測	.
百分比	10.0
Alpha	0.01
PValue	0.01
群集方法	隨機
分層	
準則	比例
忽略小分層	否
最小分層大小	5
以層級為基礎的選項	
層級選取	事件
層級比例	100.0
樣本比例	50.0
過度抽樣	
調整左數	否

一般
一般屬性

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

CaseStudy

INQ2006 樣本

流程图 日誌

流程图「CaseStudy」已開啟

樣本 (Sample)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3
資料來源
INQ2006
流程圖
CaseStudy
模型套件

屬性	值
一般	
節點 ID	Smpl
匯入的資料	
匯出的資料	
附註	
訓練	
變數	
輸出類型	資料
抽樣方法	預設
隨機種子	12345
大小	
類型	百分比
觀測	
百分比	10.0
Alpha	0.01
PValue	0.01
群集方法	隨機
分層	
準則	比例
忽略小分層	否
最小分層大小	5
以層級為基礎的選擇	
層級選取	事件
層級比例	100.0
樣本比例	50.0
過度抽樣	
調整左數	否

一般
一般屬性

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

CaseStudy

INQ2006 → 樣本

→

流程图 日誌

流程图「CaseStudy」已開啟

樣本 (Sample)

分層 準則 — 以層級為基礎；樣本比例 12.5

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3
資料來源
INQ2006
流程圖
CaseStudy
模型套件

屬性 值

變數	
輸出類型	資料
抽樣方法	預設
隨機種子	12345
大小	
類型	百分比
觀測	
百分比	100.0
Alpha	0.01
PValue	0.01
群集方法	隨機
分層	
準則	以層級為基礎
忽略小分層	否
最小分層大小	5
以層級為基礎的選項	
層級選取	事件
層級比例	100.0
樣本比例	12.5
調整次數	否
根據計數	否
排除遺漏層級	否
報表	
間隔目標	是
類別目標	是

樣本比例

指定樣本的那個部分包含選取的層級。

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

CaseStudy

INQ2006 → 樣本

大小	
類型	百分比
觀測	
百分比	100.0
Alpha	0.01
PValue	0.01
群集方法	隨機
分層	
準則	以層級為基礎
忽略小分層	否
最小分層大小	5
以層級為基礎的選項	
層級選取	事件
層級比例	100.0
樣本比例	12.5

流程圖 日誌

流程圖「CaseStudy」已開啟

樣本 (Sample)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3
資料來源
INQ2006
流程圖
CaseStudy
模型套件

屬性	值
變數	
輸出類型	資料
抽樣方法	預設
隨機種子	12345
大小	
類型	百分比
觀測	
百分比	100.0
Alpha	0.01
PValue	0.01
群集方法	隨機
分層	
準則	以層級為基礎
忽略小分層	否
最小分層大小	5
以層級為基礎的選項	
層級選取	事件
層級比例	100.0
樣本比例	12.5
過度抽樣	
調整次數	否
根據計數	否
排除遺漏層級	否
報表	
間隔目標	是
類別目標	是

樣本比例

指定樣本的哪個部分包含選取的層級。

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

CaseStudy

INQ2006 → 樣本

- 編輯變數...
- 更新
- 執行
- 建立模型套件...
- 結果...
- 將路徑匯出作為 SAS 程式
- 剪下
- 複製(C)
- 刪除
- 重新命名
- 全選
- 選取節點
- 連接節點
- 中斷連接節點

流程圖 日誌

100%

「CaseStudy」已開啟

樣本 (Sample)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3

- 資料來源
 - INQ2006
- 流程圖
- 模型套件

屬性	值
變數	
輸出類型	資料
抽樣方法	預設
隨機種子	12345
大小	
類型	百分比
觀測	.
百分比	100.0
Alpha	0.01
PValue	0.01
群集方法	隨機
分層	
準則	以層級為基礎
忽略小分層	否
最小分層大小	5
以層級為基礎的選項	
層級選取	事件
層級比例	100.0
樣本比例	12.5
過度抽樣	
調整次數	否
根據計數	否
排除遺漏層級	否
報表	
間隔目標	是
類別目標	是

樣本比例

指定樣本的哪個部分包含選取的層級。

樣本 調查 修改 模型 評估 公用程式 應用程式 時間序列

CaseStudy

INQ2006 → 樣本

執行狀態

執行完成
流程圖: CaseStudy
路徑: 樣本

確定 結果...

流程圖 日誌

執行完成

樣本 (Sample) 結果

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3

- 資料來源
 - INQ2006
- 流程圖
 - CaseStudy
- 模型套件

屬性 值

變數	資料
輸出類型	預設
抽樣方法	12345
隨機種子	
大小	
類型	百分比
觀測	
百分比	100.0
Alpha	0.01
PValue	0.01
群集方法	隨機
分層	
準則	以層級為基礎
忽略小分層	否
最小分層大小	5
以層級為基礎的選項	
層級選取	事件
層級比例	100.0
樣本比例	12.5
過度抽樣	
調整次數	否
根據次數	否
排除遺漏層級	否
報表	
間隔目標	是
類別目標	是

樣本比例

指定樣本的哪個部分包含選取的層級。

執行完成

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

CaseStudy

INQ2006 樣本

結果 - 節點: 樣本 流程圖: CaseStudy

檔案(F) 編輯(E) 檢視(V) 視窗(W)

輸出

```
40 *-----*
41
42
43
44 類別目標的摘要統計
45 (最多列印 500 個觀測)
46
47 資料=DATA
48
49 變數      數值  格式化值  次數計數  百分比  標籤
50
51 Enroll_Target  0    0      88614    96.8650
52 Enroll_Target  1    1      2868     3.1350
53
54
55 資料=SAMPLE
56
57 變數      數值  格式化值  次數計數  百分比  標籤
58
59 Enroll_Target  0    0      20076    87.5
60 Enroll_Target  1    1      2868     12.5
61
```

勘查 — StatExplore (摘要統計)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3
資料來源
INQ2006
流程圖
CaseStudy
模型套件

屬性 值

變數	
輸出類型	資料
抽樣方法	預設
隨機種子	12345
大小	
類型	百分比
觀測	
百分比	100.0
Alpha	0.01
PValue	0.01
群集方法	隨機
分層	
準則	以層級為基礎
忽略小分層	否
最小分層大小	5
以層級為基礎的選項	
層級選取	事件
層級比例	100.0
樣本比例	12.5
過度抽樣	
調整次數	否
根據計數	否
排除遺漏層級	否
報表	
間隔目標	是
類別目標	是

樣本比例

指定樣本的哪個部分包含選取的層級。

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

StatExplore
產生摘要和關聯統計。

INQ2006 樣本

執行完成

勘查 — StatExplore (摘要統計)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3
資料來源
INQ2006
流程圖
CaseStudy
模型套件

屬性	值
一般	
節點 ID	Stat
匯入的資料	
匯出的資料	
附註	
訓練	
變數	
資料	
觀測數目	100000
驗證	否
測試	否
標準報表	
間隔分配	是
類別分配	是
層級摘要	是
使用區段變數	否
交叉資料表	
變數選取	
隱藏拒絕變數	是
選取的變數數目	1000
卡方統計	
卡方	是
間隔變數	否
等距箱數的數目	5
相關統計	
相關	是
Pearson 相關	是
一般屬性	

CaseStudy

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

INQ2006 → 樣本 → Stat Explore

流程图 日誌

執行完成

勘查 — StatExplore (摘要統計)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3
資料來源
INQ2006
流程圖
CaseStudy
模型套件

屬性	值
一般	
節點 ID	Stat
匯入的資料	
匯出的資料	
附註	
訓練	
變數	
資料	
觀測數目	100000
驗證	否
測試	否
標準報表	
間隔分配	是
類別分配	是
層級摘要	是
使用區段變數	否
交叉資料表	
變數選取	
隱藏拒絕變數	是
選取的變數數目	1000
卡方統計	
卡方	是
間隔變數	否
等距箱數的數目	5
相關統計	
相關	是
Pearson 相關	是

一般
一般屬性

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

CaseStudy

INQ2006 → 樣本 → Stat Explore

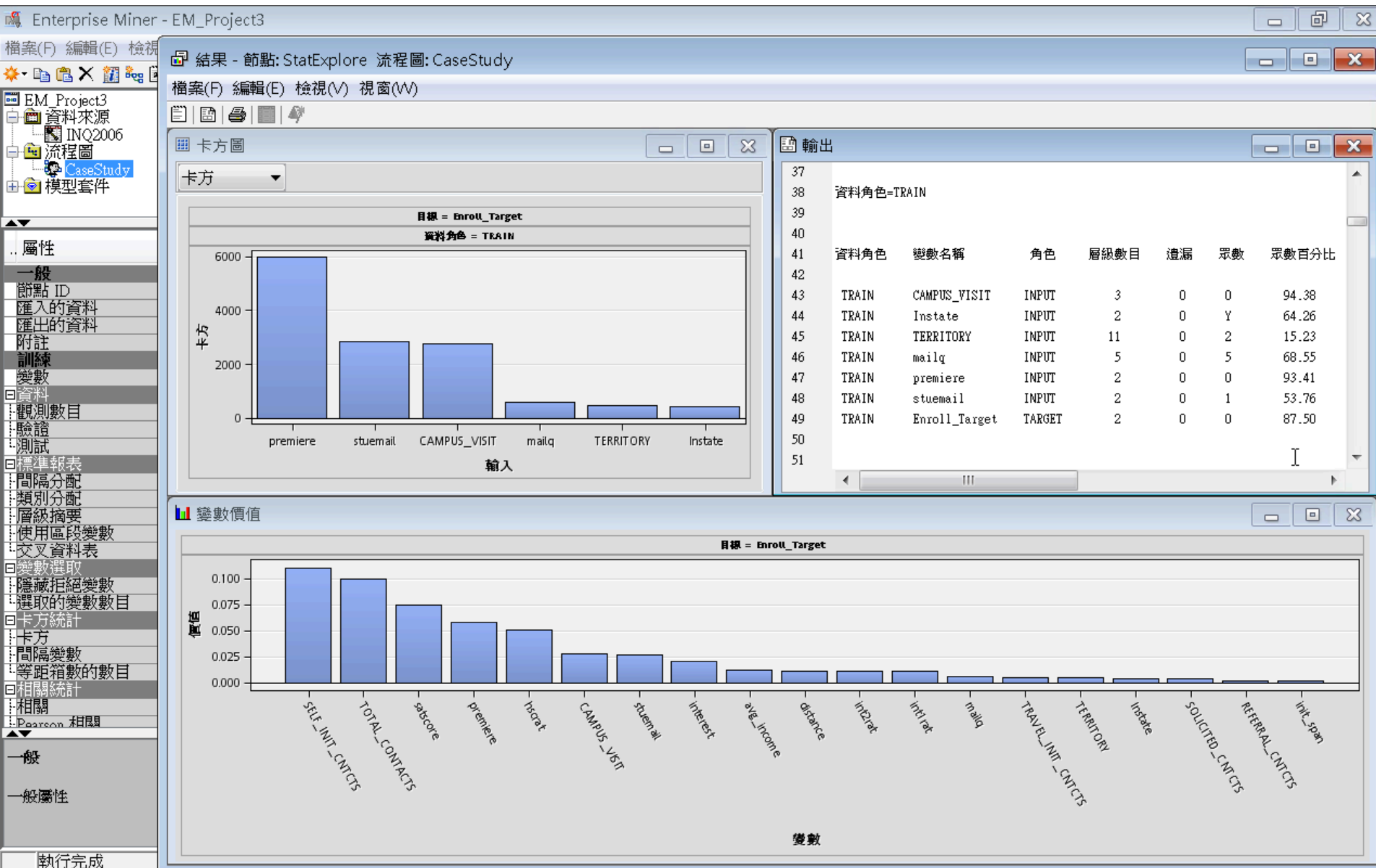
- 編輯變數...
- 更新
- 執行
- 建立模型套件...
- 結果...
- 將路徑匯出作為 SAS 程式
- 剪下
- 複製(C)
- 刪除
- 重新命名
- 全選
- 選取節點
- 連接節點
- 中斷連接節點

100%

流程圖 日誌

執行完成

勘查—StatExplore (摘要統計) 結果



勘查 — StatExplore (摘要統計) 結果輸出

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 視窗(W)

結果 - 節點: StatExplore 流程圖: CaseStudy

輸出

52

53 類別目標和區段變數的分配

54 (最多列印 500 個觀測)

55

56 資料角色=TRAIN

57

58 資料角色 變數名稱 角色 層級 次數計數 百分比

59

60 TRAIN Enroll_Target TARGET 0 20076 87.5

61 TRAIN Enroll_Target TARGET 1 2868 12.5

62

63

64

65 間隔變數摘要統計

66 (最多列印 500 個觀測)

67

68 資料角色=TRAIN

69

70 變數 角色 平均值 標準差 非遺漏值 遺漏 最小值 中位數 最大值 偏態 峰度

71

72 REFERRAL_CNTCTS INPUT 0.043323 0.242836 22944 0 0 0 5 7.690076 84.49707

73 SELF_INIT_CNTCTS INPUT 1.629184 2.190758 22944 0 0 1 21 2.134981 5.276025

74 SOLICITED_CNTCTS INPUT 0.554088 0.671876 22944 0 0 0 9 1.370244 4.391323

75 TOTAL_CONTACTS INPUT 2.607174 2.438973 22944 0 1 2 28 2.32783 6.579614

76 TRAVEL_INIT_CNTCTS INPUT 0.380535 0.57859 22944 0 0 0 5 1.463831 2.492621

77 avg_income INPUT 48502.21 21137.54 18086 4858 8495 43440 200001 1.194938 1.689842

78 distance INPUT 366.4313 393.3769 18469 4475 0.417124 175.1839 3882.192 2.373977 9.85169

79 hscrat INPUT 0.047536 0.085484 22944 0 0 0.037652 1 6.958368 67.02765

80 init_span INPUT 19.63385 8.781231 22944 0 3 19 228 1.412374 14.85753

81 intlrat INPUT 0.038747 0.02684 22944 0 0 0.045929 1 6.452302 166.1262

82 int2rat INPUT 0.04464 0.028403 22944 0 0 0.05667 1 5.343147 103.8113

83 interest INPUT 0.075793 0.286499 22944 0 0 0 3 4.023848 17.70172

84 satscore INPUT 1099.063 172.11 8326 14618 420 1110 1600 -0.23215 -0.09686

85

屬性

一般

節點 ID

匯入的資料

匯出的資料

附註

訓練

變數

資料

觀測數目

驗證

測試

標準報表

間隔分配

類別分配

層級摘要

使用區段變數

交叉資料表

變數選取

隱藏拒絕變數

選取的變數數目

卡方統計

卡方

間隔變數

等距箱數的數目

相關統計

相關

Pearson 相關

一般

一般屬性

執行完成

修改—設算(Impute)—遺失值處理

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3
資料來源
INQ2006
流程圖
CaseStudy
模型套件

屬性 值

一般	值
節點 ID	Stat
匯入的資料	
匯出的資料	
附註	
訓練變數	
資料	
觀測數目	100000
驗證	否
測試	否
標準報表	
間隔分配	是
類別分配	是
層級摘要	是
使用區段變數	否
交叉資料表	
變數選取	
隱藏拒絕變數	是
選取的變數數目	1000
卡方統計	
卡方	是
間隔變數	否
等距箱數的數目	5
相關統計	
相關	是
Pearson 相關	是

一般屬性

執行完成

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

設算
設算遺漏值

INQ2006 → 樣本 → Stat Explore

遺失值處理 Impute

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

流程圖 日誌

修改—設算(Impute)—遺失值處理

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3

- 資料來源
 - INQ2006
- 流程圖
 - CaseStudy
- 模型套件

屬性	值
一般	
節點 ID	Impt
匯入的資料	
匯出的資料	
附註	
訓練	
變數	
非遺漏變數	否
缺少截止值	50.0
類別變數	
預設輸入方法	計數
預設目標方法	無
正規化值	是
間隔變數	
預設輸入方法	平均值
預設目標方法	無
預設常數值	
預設字元值	
預設數值	
方法選項	
隨機種子	12345
調整參數	
樹設算	
評分	
隱藏原始變數	是
指標變數	
類型	無
來源	設算變數
角色	拒絕

CaseStudy

INQ2006 → 樣本 → StatExplore → 設算

遺失值處理 Impute

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

流程圖 日誌

執行完成

設算(Impute)－遺失值處理

設定 間隔變數/預設輸入方法→樹 (Tree)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3
資料來源
INQ2006
流程圖
CaseStudy
模型套件

屬性	值
一般	
節點 ID	Impt
匯入的資料	
匯出的資料	
附註	
訓練	
變數	
非遺漏變數	否
缺少截止值	50.0
類別變數	
預設輸入方法	計數
預設目標方法	無
正規化值	是
間隔變數	
預設輸入方法	樹
預設目標方法	無
預設常數值	
預設字元值	
預設數值	
方法選項	
隨機種子	12345
調整參數	
樹設算	
評分	
隱藏原始變數	是
指標變數	
類型	無
來源	設算變數
角色	拒絕

匯出的資料

節點匯出的表格集

執行完成

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

CaseStudy

INQ2006 → 樣本 → Stat Explore → 設算

遺失值處理
設定 間隔變數/預設輸入方法
將預設 平均值 改為 樹 (Tree) 補值

編輯變數...
更新
執行
建立模型套件...
結果...
將路徑匯出作為 SAS 程式
剪下
複製(C)
刪除
重新命名
全選
選取節點
連接節點
中斷連接節點

流程圖 日誌

設算(Impute)－遺失值處理

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3

- 資料來源
 - INQ2006
- 流程圖
 - CaseStudy
- 模型套件

屬性	值
一般	
節點 ID	Impt
匯入的資料	
匯出的資料	
附註	
訓練	
變數	
非遺漏變數	否
缺少截止值	50.0
類別變數	
預設輸入方法	計數
預設目標方法	無
正規化值	是
間隔變數	
預設輸入方法	樹
預設目標方法	無
預設常數值	
預設字元值	
預設數值	
方法選項	
隨機種子	12345
調整參數	
樹設算	
評分	
隱藏原始變數	是
指標變數	
類型	無
來源	設算變數
角色	拒絕

匯出的資料

節點匯出的表格集。

執行完成

CaseStudy

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

INQ2006 → 樣本 → StatExplore → 設算

確認

是否要執行此路徑?
流程圖:CaseStudy
路徑: 設算

是(Y) 否(N)

流程圖 日誌

設算(Impute)－遺失值處理

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3
資料來源
INQ2006
流程圖
CaseStudy
模型套件

屬性	值
一般	
節點 ID	Impt
匯入的資料	
匯出的資料	
附註	
訓練	
變數	
非遺漏變數	否
缺少截止值	50.0
類別變數	
預設輸入方法	計數
預設目標方法	無
正規化值	是
間隔變數	
預設輸入方法	樹
預設目標方法	無
預設常數值	
預設字元值	
預設數值	
方法選項	
隨機種子	12345
調整參數	
樹設算	
評分	
隱藏原始變數	是
指標變數	
類型	無
來源	設算變數
角色	拒絕

匯出的資料

節點匯出的表格集

執行完成

CaseStudy

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

INQ2006 → 樣本 → Stat Explore → 設算

執行狀態

執行完成
流程圖: CaseStudy
路徑: 設算

確定 結果...

流程圖 日誌

修改—設算(Impute)—結果

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3

- 資料來源
 - INQ2006
- 流程圖
 - CaseStudy
- 模型套件

屬性 值

一般	
節點 ID	Impt
匯入的資料	
匯出的資料	
附註	
訓練	
變數	
非遺漏變數	否
缺少截止值	50.0
類別變數	
預設輸入方法	計數
預設目標方法	無
正規化值	是
間隔變數	
預設輸入方法	樹
預設目標方法	無
預設常數值	
預設字元值	
預設數值	
方法選項	
隨機種子	12345
調整參數	
樹設算	
評分	
隱藏原始變數	是
指標變數	
類型	無
來源	設算變數
角色	拒絕

匯出的資料

節點匯出的表格集。

執行完成

CaseStudy

INQ2006 → 樣本 → StatExplore → 設算

結果 - 節點: 設算 流程圖: CaseStudy

檔案(F) 編輯(E) 檢視(V) 視窗(W)

設算摘要

變數名稱	設算方法	設算的變數	設算值	角色	量值層級	標籤	「TRAIN」的遺漏數目
avg_income	TREE	IMP_avg_income		.INPUT	INTERVAL		4858
distance	TREE	IMP_distance		.INPUT	INTERVAL		4475

輸出

```
38 * 報表輸出
39 *-----*
40
41
42
43 設算摘要
44 觀測數目
45
46 變數名稱    設算方法    設算的變數    設算值    角色    量值層級    標籤    「TRAIN」的遺漏數目
47
48 avg_income   TREE        IMP_avg_income    .    INPUT   INTERVAL    .    4858
49 distance     TREE        IMP_distance      .    INPUT   INTERVAL    .    4475
50
```

設算(Impute)－匯出的資料－瀏覽

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3

- 資料來源
 - INQ2006
- 流程圖
- CaseStudy
- 模型套件

屬性 值

一般

節點 ID Impt

匯入的資料

匯出的資料

附註

訓練變數

非遺漏變數 否

缺少截止值 50.0

類別變數

預設輸入方法 計數

預設目標方法 無

正規化值 是

間隔變數

預設輸入方法 樹

預設目標方法 無

預設常數值

預設字元值

預設數值

方法選項

隨機種子 12345

調整參數

樹設算

評分

隱藏原始變數 是

指標變數

類型 無

來源 設算變數

角色 拒絕

匯出的資料

節點匯出的表格集。

執行完成

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

CaseStudy

INQ2006 → 樣本 → StatExplore → 設算

匯出的資料 - 設算

連接埠	表格	角色	資料存在
TRAIN	EMWS1.Impt_TRAIN	訓練	是
VALIDATE	EMWS1.Impt_VALIDATE	驗證	否
TEST	EMWS1.Impt_TEST	測試	否

瀏覽(B)... 勘查(X)... 屬性(P)... 確定(O)

流程圖 日誌

設算(Impute)－匯出的資料－瀏覽

遺失值處理結果

EMWS1Impt_TRAIN													
	interest	stuemail	init_span	int1rat	int2rat	hscrat	avg_income	distance	Instate	_dataobs_	Imputed avg_income	警告	Imputed distance
1	0.0	1.0	45.0	0.017182745659566852	0.020380002122916886	0.0376516305204012	.	.	N	5.0	60264.7689946277	.	1529.69814898739
2	0.0	1.0	29.0	0.017182745659566852	0.020380002122916886	0.0376516305204012	.	.	N	7.0	60264.7689946277	.	1529.69814898739
3	0.0	1.0	30.0	0.017182745659566852	0.020380002122916886	0.0376516305204012	.	.	N	12.0	60264.7689946277	.	1529.69814898739
4	0.0	1.0	41.0	0.017182745659566852	0.020380002122916886	0.0376516305204012	.	.	N	13.0	60264.7689946277	.	1529.69814898739
5	0.0	0.0	34.0	0.017182745659566852	0.020380002122916886	0.0376516305204012	.	.	N	18.0	60264.7689946277	.	1529.69814898739
6	0.0	0.0	72.0	0.017182745659566852	0.020380002122916886	0.0376516305204012	.	.	N	19.0	60264.7689946277	.	1529.69814898739
7	0.0	0.0	56.0	0.017182745659566852	0.020380002122916886	0.0376516305204012	.	.	N	27.0	60264.7689946277	.	1529.69814898739
8	0.0	0.0	55.0	0.017182745659566852	0.020380002122916886	0.0376516305204012	.	.	N	28.0	60264.7689946277	.	1529.69814898739
9	0.0	1.0	50.0	0.017182745659566852	0.020380002122916886	0.0376516305204012	.	.	N	32.0	60264.7689946277	.	1529.69814898739
10	0.0	1.0	47.0	0.017182745659566852	0.020380002122916886	0.0376516305204012	.	.	N	36.0	60264.7689946277	.	1529.69814898739
11	0.0	0.0	46.0	0.017182745659566852	0.020380002122916886	0.0376516305204012	.	.	N	38.0	60264.7689946277	.	1529.69814898739
12	0.0	0.0	45.0	0.017182745659566852	0.020380002122916886	0.0376516305204012	.	.	N	40.0	60264.7689946277	.	1529.69814898739
13	0.0	1.0	38.0	0.017182745659566852	0.020380002122916886	0.0376516305204012	.	.	N	44.0	60264.7689946277	.	1529.69814898739
14	0.0	0.0	35.0	0.017182745659566852	0.020380002122916886	0.0376516305204012	.	.	N	49.0	60264.7689946277	.	1529.69814898739
15	0.0	0.0	34.0	0.017182745659566852	0.020380002122916886	0.0376516305204012	.	.	N	51.0	60264.7689946277	.	1529.69814898739
16	0.0	0.0	29.0	0.017182745659566852	0.020380002122916886	0.0376516305204012	.	.	N	65.0	60264.7689946277	.	1529.69814898739
17	0.0	1.0	25.0	0.017182745659566852	0.020380002122916886	0.0376516305204012	.	.	N	84.0	60264.7689946277	.	1529.69814898739
18	0.0	1.0	23.0	0.017182745659566852	0.020380002122916886	0.0376516305204012	.	.	N	90.0	60264.7689946277	.	1529.69814898739
19	0.0	0.0	23.0	0.017182745659566852	0.020380002122916886	0.0376516305204012	.	.	N	91.0	60264.7689946277	.	1529.69814898739
20	0.0	1.0	23.0	0.017182745659566852	0.020380002122916886	0.0376516305204012	.	.	N	95.0	60264.7689946277	.	1529.69814898739
21	0.0	1.0	22.0	0.017182745659566852	0.020380002122916886	0.0376516305204012	.	.	N	99.0	60264.7689946277	.	1529.69814898739
22	0.0	1.0	22.0	0.017182745659566852	0.020380002122916886	0.0376516305204012	.	.	N	101.0	60264.7689946277	.	1529.69814898739
23	0.0	1.0	22.0	0.017182745659566852	0.020380002122916886	0.0376516305204012	.	.	N	107.0	60264.7689946277	.	1529.69814898739
24	0.0	1.0	21.0	0.017182745659566852	0.020380002122916886	0.0376516305204012	.	.	N	110.0	60264.7689946277	.	1529.69814898739
25	0.0	1.0	11.0	0.017182745659566852	0.020380002122916886	0.0376516305204012	.	.	N	120.0	60264.7689946277	.	1529.69814898739
26	0.0	1.0	20.0	0.017182745659566852	0.020380002122916886	0.0376516305204012	.	.	N	124.0	60264.7689946277	.	1529.69814898739
27	0.0	1.0	20.0	0.017182745659566852	0.020380002122916886	0.0376516305204012	.	.	N	135.0	60264.7689946277	.	1529.69814898739
28	0.0	1.0	20.0	0.017182745659566852	0.020380002122916886	0.0376516305204012	.	.	N	137.0	60264.7689946277	.	1529.69814898739

設算(Impute)－匯出的資料

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3
資料來源
INQ2006
流程圖
CaseStudy
模型套件

屬性 值

一般	值
節點 ID	Impt
匯入的資料	
匯出的資料	
附註	
訓練	
變數	
非遺漏變數	否
缺少截止值	50.0
類別變數	
預設輸入方法	計數
預設目標方法	無
正規化值	是
間隔變數	
預設輸入方法	樹
預設目標方法	無
預設常數值	
預設字元值	
預設數值	
方法選項	
隨機種子	12345
調整參數	
樹設算	
評分	
隱藏原始變數	是
指標變數	
類型	無
來源	設算變數
角色	拒絕

匯出的資料

節點匯出的表格集

執行完成

CaseStudy

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

INQ2006 樣本 Stat Explore 設算

匯出的資料 - 設算

連接埠	表格	角色	資料存在
TRAIN	EMWS1.Impt_TRAIN	訓練	是
VALIDATE	EMWS1.Impt_VALIDATE	驗證	否
TEST	EMWS1.Impt_TEST	測試	否

瀏覽(B)... 勘查(X)... 屬性(P)... 確定(O)

流程圖 日誌

樣本－資料分區 (Data Partition)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3

- 資料來源
 - INQ2006
- 流程圖
 - CaseStudy
- 模型套件

屬性	值
一般	
節點 ID	Impt
匯入的資料	
匯出的資料	
附註	
訓練	
變數	
非遺漏變數	否
缺少截止值	50.0
類別變數	
預設輸入方法	計數
預設目標方法	無
正規化值	是
間隔變數	
預設輸入方法	樹
預設目標方法	無
預設常數值	
預設字元值	
預設數值	
方法選項	
隨機種子	12345
調整參數	
樹計算	
評分	
隱藏原始變數	是
指標變數	
類型	無
來源	計算變數
角色	拒絕

匯出的資料

節點匯出的表格集。

執行完成

樣本 資料分區

將資料分區到個別的表格中。

INQ2006 → 樣本 → StatExplore → 計算

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

100%

流程圖 日誌

樣本－資料分區 (Data Partition)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3
資料來源
INQ2006
流程圖
CaseStudy
模型套件

屬性	值
一般	
節點 ID	Part
匯入的資料	
匯出的資料	
附註	
訓練	
變數	
輸出類型	資料
分區方法	預設
隨機種子	12345
資料集配置	
訓練	40.0
驗證	30.0
測試	30.0
報表	
間隔目標	是
類別目標	是
狀態	

一般
一般屬性

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

CaseStudy

INQ2006 → 樣本 → Stat Explore → 計算 → 資料分區

執行完成

資料分區 (Data Partition)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3
資料來源
INQ2006
流程圖
CaseStudy
模型套件

屬性 值

一般	值
節點 ID	Part
匯入的資料	
匯出的資料	
附註	
訓練	
變數	
輸出類型	資料
分區方法	預設
隨機種子	12345
資料集配置	
訓練	70.0
驗證	30.0
測試	0.0
報表	
間隔目標	是
類別目標	是
狀態	

訓練
指定訓練資料集的配置。預設值是 40%。

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

CaseStudy

INQ2006 → 樣本 → Stat Explore → 計算 → 資料分區

資料集配置

訓練	70.0
驗證	30.0
測試	0.0

流程图 日誌

執行完成

資料分區 (Data Partition)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3
資料來源
INQ2006
流程圖
CaseStudy
模型套件

屬性	值
一般	
節點 ID	Part
匯入的資料	
匯出的資料	
附註	
訓練	
變數	
輸出類型	資料
分區方法	預設
隨機種子	12345
資料集配置	
訓練	70.0
驗證	30.0
測試	0.0
報表	
間隔目標	是
類別目標	是
狀態	
一般	
一般屬性	

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

CaseStudy

INQ2006 → 樣本 → Stat Explore → 計算 → 資料分區

編輯變數...

更新

執行

建立模型套件...

結果...

將路徑匯出作為 SAS 程式

剪下

複製(C)

刪除

重新命名

全選

選取節點

連接節點

中斷連接節點

100%

流程圖 日誌

執行完成

資料分區 (Data Partition)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3
資料來源
INQ2006
流程圖
CaseStudy
模型套件

屬性	值
一般	
節點 ID	Part
匯入的資料	
匯出的資料	
附註	
訓練	
變數	
輸出類型	資料
分區方法	預設
隨機種子	12345
資料集配置	
訓練	70.0
驗證	30.0
測試	0.0
報表	
間隔目標	是
類別目標	是
狀態	

一般
一般屬性

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

CaseStudy

INQ2006 → 樣本 → Stat Explore → 計算 → 資料分區

確認

是否要執行此路徑?
流程圖: CaseStudy
路徑: 資料分區

是(Y) 否(N)

100%

流程圖 日誌

執行完成

資料分區 (Data Partition)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3
資料來源
INQ2006
流程圖
CaseStudy
模型套件

屬性	值
一般	
節點 ID	Part
匯入的資料	
匯出的資料	
附註	
訓練	
變數	
輸出類型	資料
分區方法	預設
隨機種子	12345
資料集配置	
訓練	70.0
驗證	30.0
測試	0.0
報表	
間隔目標	是
類別目標	是
狀態	

一般
一般屬性

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

CaseStudy

INQ2006 → 樣本 → Stat Explore → 計算 → 資料分區

執行狀態

執行完成
流程圖: CaseStudy
路徑: 資料分區

確定 結果...

流程图 日誌

執行完成

樣本－資料分區 (Data Partition) 結果

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3

- 資料來源
 - INQ2006
- 流程圖
 - CaseStudy
- 模型套件

屬性 值

一般	
節點 ID	Part
匯入的資料	
匯出的資料	
附註	
訓練	
變數	
輸出類型	資料
分區方法	預設
隨機種子	12345
資料集配置	
訓練	70.0
驗證	30.0
測試	0.0
報表	
間隔目標	是
類別目標	是
狀態	

執行完成

樣本 調查 修改 模型 評估 公用程式 應用程式 時間序列

CaseStudy

INQ2006 → 樣本 → Stat Explore → 設算 → 資料分區

結果 - 節點: 資料分區 流程圖: CaseStudy

檔案(F) 編輯(E) 檢視(V) 視窗(W)

輸出

51	資料=DATA					
52						
53	變數	數值	格式化值	次數計數	百分比	標籤
54						
55	Enroll_Target	0	0	20076	87.5	
56	Enroll_Target	1	1	2868	12.5	
57						
58						
59	資料=TRAIN					
60						
61	變數	數值	格式化值	次數計數	百分比	標籤
62						
63	Enroll_Target	0	0	14052	87.4969	
64	Enroll_Target	1	1	2008	12.5031	
65						
66						
67	資料=VALIDATE					
68						
69	變數	數值	格式化值	次數計數	百分比	標籤
70						
71	Enroll_Target	0	0	6024	87.5073	
72	Enroll_Target	1	1	860	12.4927	

決策樹 (Decision Tree)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3
資料來源
INQ2006
流程圖
CaseStudy
模型套件

屬性 值

一般	值
節點 ID	Part
匯入的資料	
匯出的資料	
附註	
訓練	
變數	
輸出類型	資料
分區方法	預設
隨機種子	12345
資料集配置	
訓練	70.0
驗證	30.0
測試	0.0
報表	
間隔目標	是
類別目標	是
狀態	

一般屬性

執行完成

樣本 查詢 修改 模型 評估 公用程式 應用程式 時間序列

決策樹
經驗樹代表透過套用一系列的簡單規則建立資料區段。

INQ2006 → 樣本 → Stat Explore → 計算 → 資料分區

流程图 日誌

決策樹 (Decision Tree)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3
資料來源
INQ2006
流程圖
CaseStudy
模型套件

屬性	值
一般	
節點 ID	Tree
匯入的資料	
匯出的資料	
附註	
訓練	
變數	
互動式	
使用凍結樹	否
使用多重目標	否
精確度	4
分割規則	
間隔準則	ProbF
名目準則	ProbChisq
序數準則	熵
顯著水準	0.2
遺漏值	搜尋時使用
使用輸入一次	否
最大分支	2
最大深度	6
最小類別大小	5
分割精確度	4
節點	
分葉大小	5
規則數目	5
替代規則數目	0
分割大小	.
分割搜尋	
使用決策	否
使用事前機率	否
徹底	5000
一般	

樣本 調查 修改 模型 評估 公用程式 應用程式 時間序列

CaseStudy

INQ2006 → 樣本 → Stat Explore → 設算

資料分區 → 決策樹

流程图 日誌

執行完成

互動式決策樹 (Interactive Decision Tree)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3
資料來源
INQ2006
流程圖
CaseStudy
模型套件

屬性 值
一般
節點 ID Tree
匯入的資料
匯出的資料
附註
訓練
變數
互動式
使用凍結樹 否
使用多重目標 否
精確度 4
分割規則
間隔準則 ProbF
名目準則 ProbChisq
序數準則 熵
顯著水準 0.2
遺漏值 搜尋時使用
使用輸入一次 否
最大分支 2
最大深度 6
最小類別大小 5
分割精確度 4
節點
分葉大小 5
規則數目 5
替代規則數目 0
分割大小 .
分割搜尋
使用決策 否
使用事前機率 否

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

CaseStudy

INQ2006 → 樣本 → Stat Explore → 設算

資料分區 → 決策樹

互動式決策樹 - EMWS1.TREE_BROWSETREE[EMWS1.EM_TREE]

檔案 編輯 檢視 動作 視窗

樹檢視

統計	訓練	驗證
0:	87.50%	87.51%
1:	12.50%	12.49%
計數:	16060	6884

分割計數	路徑中的節點 ...	分割變數	值	統計	訓練	驗證
0	1 (已選取)		所有的案例或...	計數	16060	6884
				預測	0	0
				目標是下列項目的百...	87.50%	87.51%
				目標是下列項目的百...	12.50%	12.49%
				% 正確的預測	87.50%	87.51%

執行完成

互動式決策樹 (Interactive Decision Tree)

分割節點 (Split Node)

互動式決策樹 - EMWS1.TREE_BROWSETREE[EMWS1.EM_TREE]

檔案 編輯 檢視 動作 視窗

樹檢視

統計 訓練 驗證
0: 87.50% 87.51%
1: 12.50% 12.49%
計數: 16060 6884

樹節點(Node) 按右鍵 / 分割節點 (Split Node)
互動式 (手動) 建立決策樹

圖形屬性...
列印
檢視
工具
節點統計
分割節點...
訓練節點
修剪節點
切換目標...
複製下階
貼上下階
貼上儲存的樹...
復原修剪

分割節點...

分割計數	路徑中的節點 ID	分割變數	值
0	1 (已選取)		所有的案例或遺漏值

統計	訓練	驗證
計數	16060	6884
預測	0	0
目標是下列項目的百分比 = 0	87.50%	87.51%
目標是下列項目的百分比 = 1	12.50%	12.49%
% 正確的預測	87.50%	87.51%

互動式決策樹 (Interactive Decision Tree)

分割節點 (Split Node)

互動式決策樹 - EMWS1.TREE_BROWSETREE[EMWS1.EM_TREE]

檔案 編輯 檢視 動作 視窗

樹檢視

統計	訓練	驗證
0: 87.50%	87.51%	
1: 12.50%	12.49%	
計數: 16060	6884	

分割節點「1」

目標變數: Enroll_Target

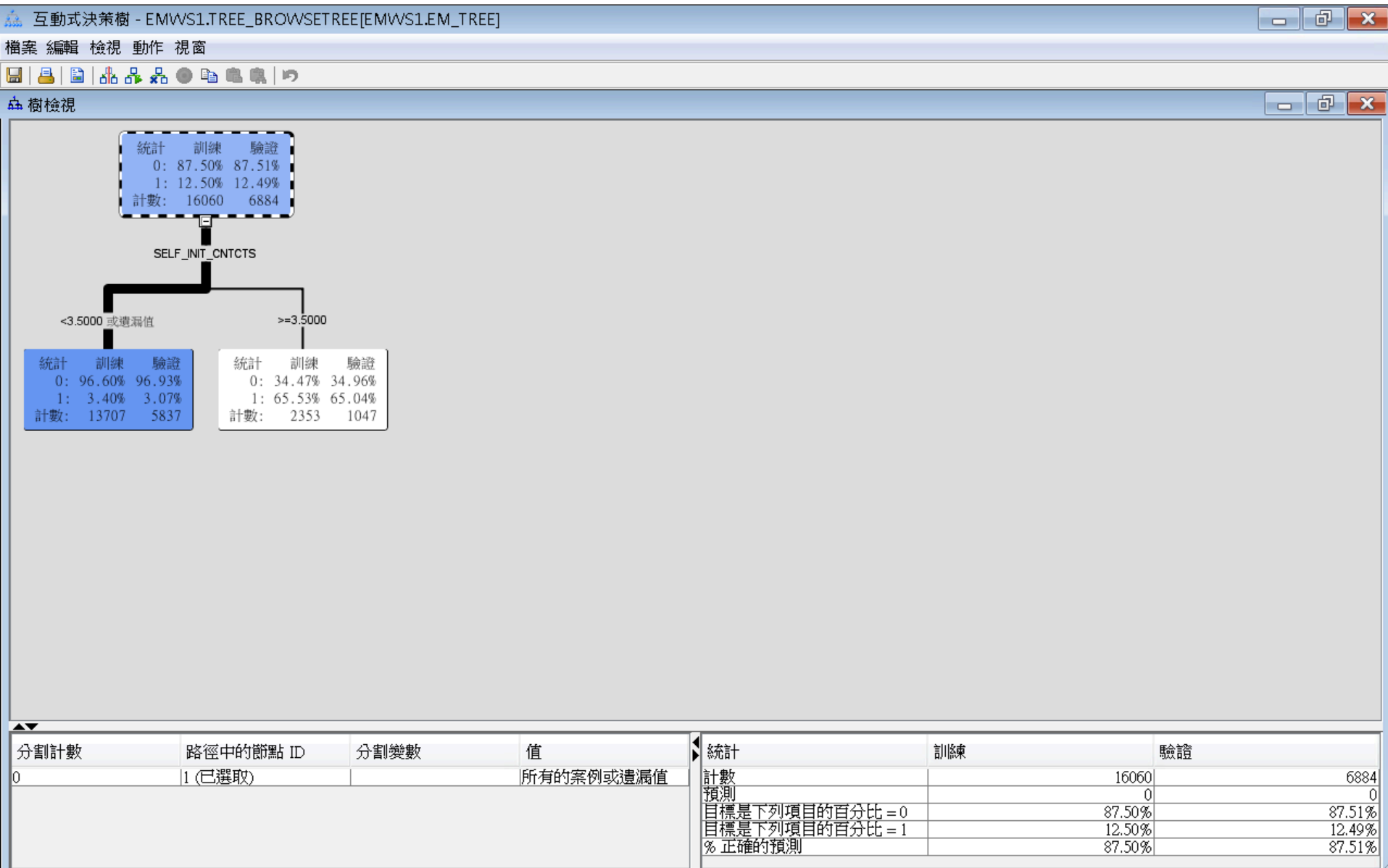
變數	變數描述	-Log(p)	分支
SELF_INIT_CNTCTS	SELF_INIT_CNTCTS	1539.7668	2
TOTAL_CONTACTS	TOTAL_CONTACTS	1297.371	2
premiere	premiere	890.4942	2
hscrat	hscrat	583.0456	2
stuemail	stuemail	430.8993	2
CAMPUS_VISIT	CAMPUS_VISIT	410.9033	2
interest	interest	319.8554	2
IMP_avg_income	Imputed avg_income	95.1729	2
int1rat	int1rat	94.7048	2
IMP_distance	Imputed distance	93.0931	2
int2rat	int2rat	92.7231	2
mailq	mailq	67.6788	2
Instate	Instate	66.4243	2
TERRITORY	RECR_CODE	60.1217	2
TRAVEL_INIT_CNTCTS	TRAVEL_INIT_CNTCTS	47.6199	2
SOLICITED_CNTCTS	SOLICITED_CNTCTS	33.8226	2

編輯規則...

確定
取消
套用
重新整理

分割計數	路徑中的節點 ID	分割變數	值	統計	訓練	驗證
0	1 (已選取)		所有的案例或遺漏值	計數	16060	6884
				預測	0	0
				目標是下列項目的百分比 = 0	87.50%	87.51%
				目標是下列項目的百分比 = 1	12.50%	12.49%
				% 正確的預測	87.50%	87.51%

互動式決策樹 (Interactive Decision Tree)



互動式決策樹 (Interactive Decision Tree)

互動式決策樹 - EMWS1.TREE_BROWSETREE[EMWS1.EM_TREE]

檔案 編輯 檢視 動作 視窗

樹檢視

統計 訓練 驗證

0: 87.50% 87.51%

1: 12.50% 12.49%

計數: 16060 6884

SELF_INIT_CNTCTS

<3.5000 或遺漏值

>=3.5000

統計 訓練 驗證

0: 96.60% 96.93%

1: 3.40% 3.07%

計數: 13707 5837

統計 訓練 驗證

0: 34.47% 34.96%

1: 65.53% 65.04%

計數: 2353 1047

圖形屬性...

列印

檢視

工具

節點統計

分割節點...

訓練節點

修剪節點

切換目標...

複製下階

貼上下階

貼上儲存的樹...

復原修剪

分割變數	值	統計	訓練	驗證
SELF_INIT_CNTCTS	<3.5000 或遺漏值	計數	13707	5837
		預測	0	0
		目標是下列項目的百分比 = 0	96.60%	96.93%
		目標是下列項目的百分比 = 1	3.40%	3.07%
		% 正確的預測	96.60%	96.93%

互動式決策樹 (Interactive Decision Tree)

互動式決策樹 - EMWS1.TREE_BROWSETREE[EMWS1.EM_TREE]

檔案 編輯 檢視 動作 視窗

樹檢視

統計 訓練 驗證
0: 87.50% 87.51%
1: 12.50% 12.49%
計數: 16060 6884

SELF_INIT_CNTCTS

<3.5000 或遺漏值

統計 訓練 驗證
0: 96.60% 96.93%
1: 3.40% 3.07%
計數: 13707 5837

>=3.5000

統計 訓練 驗證
0: 34.47% 34.96%
1: 65.53% 65.04%
計數: 2353 1047

分割節點「3」

目標變數: Enroll_Target

變數	變數描述	-Log(p)	分支
SELF_INIT_CNTCTS	SELF_INIT_CNTCTS	341.7195	2
hscrat	hscrat	205.5885	2
TOTAL_CONTACTS	TOTAL_CONTACTS	182.2041	2
stuemail	stuemail	120.7655	2
interest	interest	99.947	2
init_span	init_span	56.6105	2
CAMPUS_VISIT	CAMPUS_VISIT	56.0912	2
IMP_avg_income	Imputed avg_income	39.354	2
premiere	premiere	24.9827	2
intlrat	intlrat	22.8795	2
IMP_distance	Imputed distance	17.8275	2
int2rat	int2rat	15.7647	2
Instate	Instate	11.7578	2
SOLICITED_CNTCTS	SOLICITED_CNTCTS	10.9002	2
TERRITORY	RECR_CODE	10.845	2
mailq	mailq	4.2802	2
REFERRAL_CNTCTS	REFERRAL_CNTCTS	3.1983	2
TRAVEL_INIT_CNTCTS	TRAVEL_INIT_CNT...	1.4705	2

編輯規則...

確定 取消 套用 重新整理

分割計數	路徑中的節點 ID	分割變數	值	統計	訓練	驗證
0	1	SELF_INIT_CNTCTS	<3.5000 或遺漏值	計數	13707	5837
1	3 (已選取)			預測	0	0
				目標是下列項目的百分比 = 0	96.60%	96.93%
				目標是下列項目的百分比 = 1	3.40%	3.07%
				% 正確的預測	96.60%	96.93%

互動式決策樹 (Interactive Decision Tree)

互動式決策樹 - EMWS1.TREE_BROWSETREE[EMWS1.EM_TREE]

檔案 編輯 檢視 動作 視窗

樹檢視

The diagram shows a decision tree structure. The root node is a blue box with the following statistics:

統計	訓練	驗證
0: 87.50%	87.51%	
1: 12.50%	12.49%	
計數: 16060	6884	

The root node splits on the variable SELF_INIT_CNTCTS with a threshold of 3.5000. The left branch is labeled "<3.5000 或遺漏值" and the right branch is labeled ">=3.5000".

The left branch leads to a blue box with the following statistics:

統計	訓練	驗證
0: 96.60%	96.93%	
1: 3.40%	3.07%	
計數: 13707	5837	

The right branch leads to a dashed box with the following statistics:

統計	訓練	驗證
0: 34.47%	34.96%	
1: 65.53%	65.04%	
計數: 2353	1047	

The left branch further splits on the variable SELF_INIT_CNTCTS with a threshold of 2.5000. The left sub-branch is labeled "<2.5000 或遺漏值" and the right sub-branch is labeled ">=2.5000".

The left sub-branch leads to a blue box with the following statistics:

統計	訓練	驗證
0: 98.25%	98.52%	
1: 1.75%	1.48%	
計數: 12777	5456	

The right sub-branch leads to a blue box with the following statistics:

統計	訓練	驗證
0: 73.87%	74.28%	
1: 26.13%	25.72%	
計數: 930	381	

A context menu is open over the dashed box, with the "分割節點..." (Split Node...) option highlighted in red. The menu options are:

- 圖形屬性...
- 列印
- 檢視
- 工具
- 節點統計
- 分割節點...**
- 訓練節點
- 修剪節點
- 切換目標...
- 複製下階
- 貼上下階
- 貼上儲存的樹...
- 復原修剪

At the bottom left, a table shows the split statistics for the selected node:

分割計數	路徑中的節點 ID	分割變數
0	1	SELF_INIT_CN
1	4 (已選取)	

At the bottom right, a table shows the statistics for the selected node:

統計	訓練	驗證
計數	2353	1047
預測	1	1
目標是下列項目的百分比 = 0	34.47%	34.96%
目標是下列項目的百分比 = 1	65.53%	65.04%
% 正確的預測	65.53%	65.04%

互動式決策樹 (Interactive Decision Tree)

互動式決策樹 - EMWS1.TREE_BROWSER[EMWS1.EM_TREE]

檔案 編輯 檢視 動作 視窗

樹檢視

統計 訓練 驗證
0: 87.50% 87.51%
1: 12.50% 12.49%
計數: 16060 6884

SELF_INIT_CNTCTS

<3.5000 或遺漏值

統計 訓練 驗證
0: 96.60% 96.93%
1: 3.40% 3.07%
計數: 13707 5837

SELF_INIT_CNTCTS

<2.5000 或遺漏值

統計 訓練 驗證
0: 98.25% 98.52%
1: 1.75% 1.48%
計數: 12777 5456

>=3.5000

統計 訓練 驗證
0: 34.47% 34.96%
1: 65.53% 65.04%
計數: 2353 1047

>=2.5000

統計 訓練 驗證
0: 73.87% 74.28%
1: 26.13% 25.72%
計數: 930 381

分割節點「4」

目標變數: Enroll_Target

變數	變數描述	-Log(p)	分支
hscrat	hscrat	110.5617	2
stuemail	stuemail	55.0412	2
premiere	premiere	38.7475	2
SELF_INIT_CNTCTS	SELF_INIT_CNTCTS	33.1178	2
TOTAL_CONTACTS	TOTAL_CONTACTS	25.3574	2
CAMPUS_VISIT	CAMPUS_VISIT	9.9815	2
int2rat	int2rat	9.8966	2
intlrat	intlrat	8.3302	2
init_span	init_span	5.388	2
IMP_avg_income	Imputed avg_income	3.3842	2
IMP_distance	Imputed distance	3.3834	2
interest	interest	3.084	2
Instate	Instate	1.5891	2
TRAVEL_INIT_CNTCTS	TRAVEL_INIT_CNTCTS	1.247	2
mailq	mailq	0.8813	2
REFERRAL_CNTCTS	REFERRAL_CNTCTS	0.8805	2
SOLICITED_CNTCTS	SOLICITED_CNTCTS	0.1816	2
TERRITORY	RECR_CODE	0.0732	2

編輯規則...

確定 取消 套用 重新整理

分割計數	路徑中的節點 ID	分割變數	值	統計	訓練	驗證
0	1	SELF_INIT_CNTCTS	>=3.5000	計數	2353	1047
1	4 (已選取)			預測	1	1
				目標是下列項目的百分比 = 0	34.47%	34.96%
				目標是下列項目的百分比 = 1	65.53%	65.04%
				% 正確的預測	65.53%	65.04%

互動式決策樹 (Interactive Decision Tree)

互動式決策樹 - EMWS1.TREE_BROWSETREE[EMWS1.EM_TREE]

檔案 編輯 檢視 動作 視窗

樹檢視

統計 訓練 驗證
0: 87.50% 87.51%
1: 12.50% 12.49%
計數: 16060 6884

SELF_INIT_CNTCTS

<3.5000 或遺漏值

統計 訓練 驗證
0: 96.60% 96.93%
1: 3.40% 3.07%
計數: 13707 5837

SELF_INIT_CNTCTS

<2.5000 或遺漏值

統計 訓練 驗證
0: 98.25% 98.52%
1: 1.75% 1.48%
計數: 12777 5456

>=3.5000

統計 訓練 驗證
0: 34.47% 34.96%
1: 65.53% 65.04%
計數: 2353 1047

>=2.5000

統計 訓練 驗證
0: 73.87% 74.28%
1: 26.13% 25.72%
計數: 930 381

hscrat - 間隔分割規則

目標變數: Enroll_Target

將遺漏值指派到

☒ 特定的分支 2

☐ 個別的遺漏值分支

☐ 所有的分支

分支

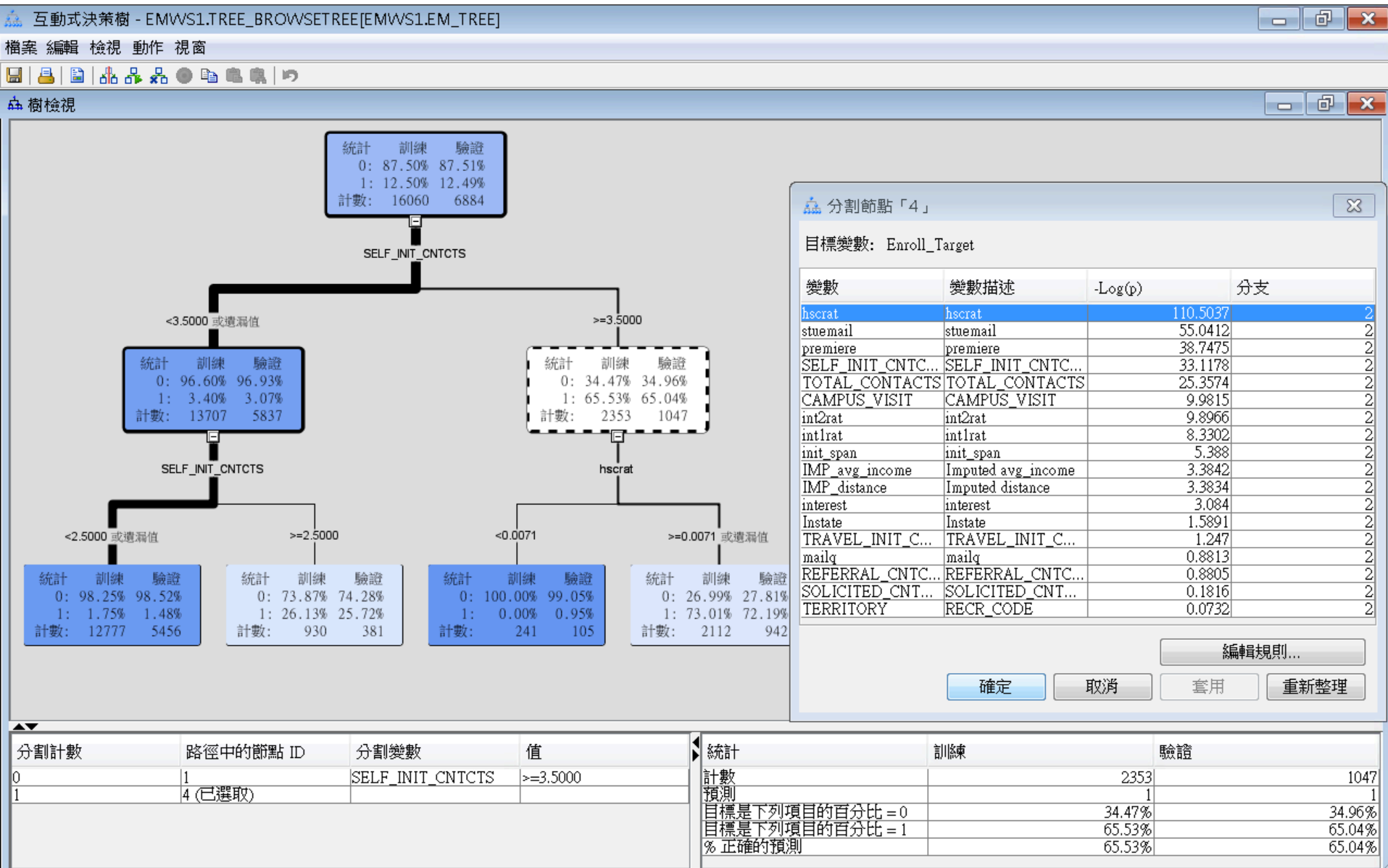
分支	分割點
1	< 0.0071
2	>= 0.0071

新增分割點: 增加分支 移除分支

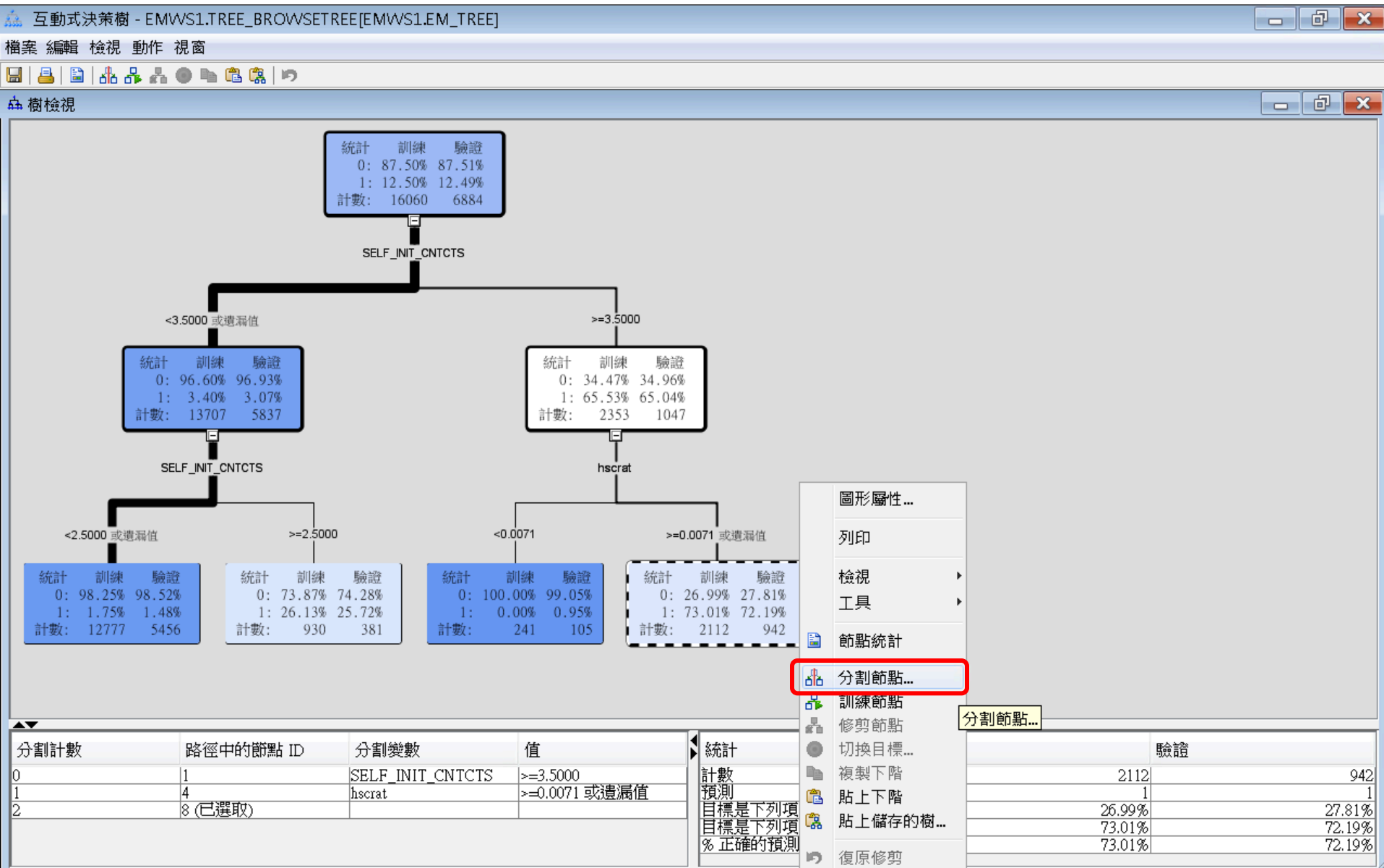
確定 取消 套用 重設

分割計數	路徑中的節點 ID	分割變數	值	統計	訓練	驗證
0	1	SELF_INIT_CNTCTS	>=3.5000	計數	2353	1047
1	4 (已選取)			預測	1	1
				目標是下列項目的百分比 = 0	34.47%	34.96%
				目標是下列項目的百分比 = 1	65.53%	65.04%
				% 正確的預測	65.53%	65.04%

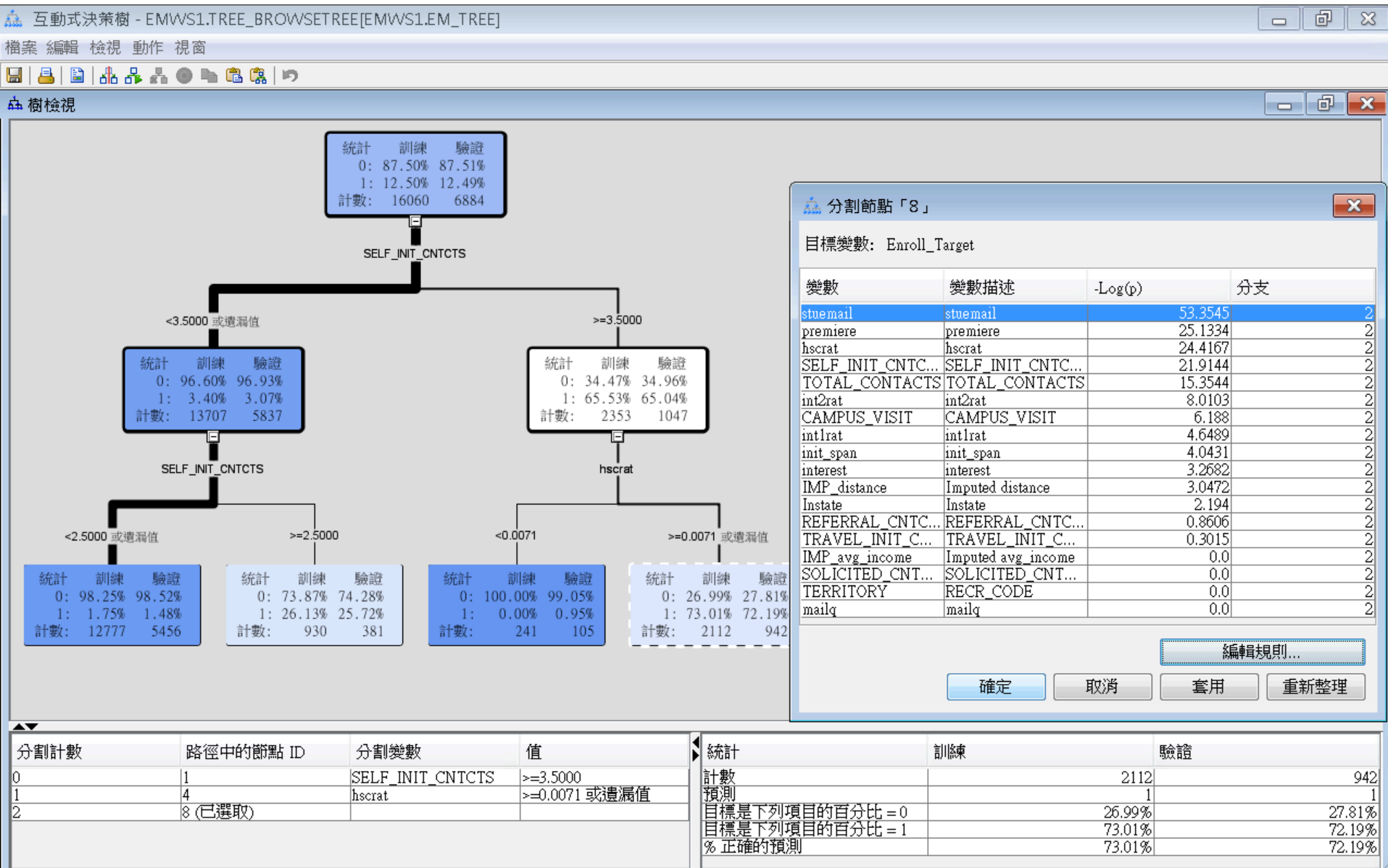
互動式決策樹 (Interactive Decision Tree)



互動式決策樹 (Interactive Decision Tree)



互動式決策樹 (Interactive Decision Tree)



互動式決策樹 (Interactive Decision Tree)

修剪節點 (Prune Node)

互動式決策樹 - EMWS1.TREE_BROWSETREE[EMWS1.EM_TREE]

檔案 編輯 檢視 動作 視窗

樹檢視

統計 訓練 驗證
0: 96.60% 96.93%
1: 3.40% 3.07%
計數: 13707 5837

統計 訓練 驗證
0: 34.47% 34.96%
1: 65.53% 65.04%
計數: 2353 1047

SELF_INIT_CNTCTS

統計 訓練 驗證
0: 98.25% 98.52%
1: 1.75% 1.48%
計數: 12777 5456

統計 訓練 驗證
0: 73.87% 74.28%
1: 26.13% 25.72%
計數: 930 381

統計 訓練 驗證
0: 100.00% 99.05%
1: 0.00% 0.95%
計數: 241 105

統計 訓練 驗證
0: 26.99% 27.81%
1: 73.01% 72.19%
計數: 2112 942

hscrat

stuemail

1.0000 或遺漏值

統計 訓練 驗證
0: 23.89% 24.86%
1: 76.11% 75.14%
計數: 2026 905

統計 訓練 驗證
0: 1.00% 1.00%
1: 99.00% 99.00%
計數: 1 1

圖形屬性...

列印

檢視

工具

節點統計

分割節點...

訓練節點

修剪節點

切換目標...

複製下階

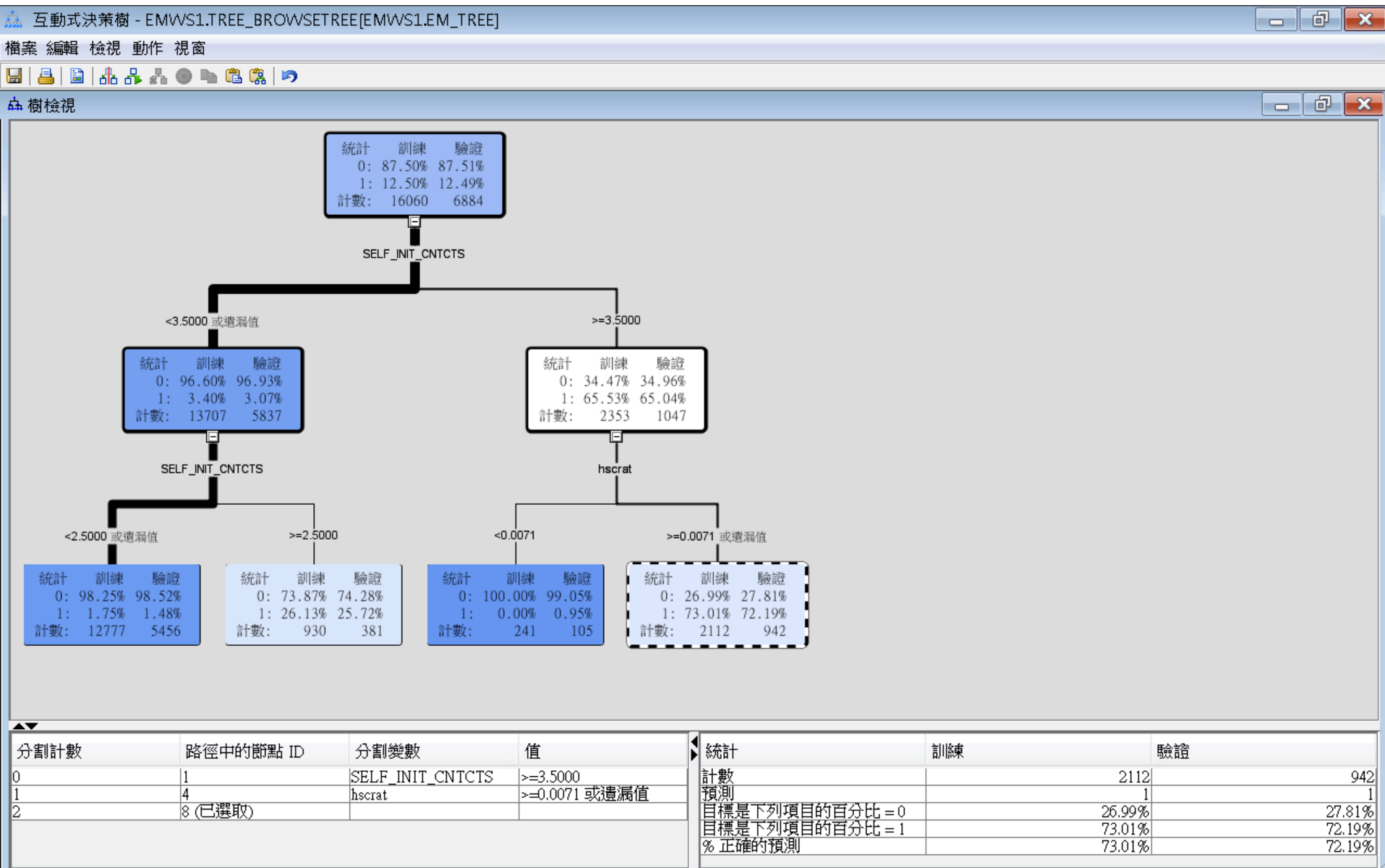
貼上下階

貼上儲存的樹...

復原修剪

分割計數	路徑中的節點 ID	分割變數	值	統計	訓練	驗證
0	1	SELF_INIT_CNTCTS	>=3.5000	計數	2112	942
1	4	hscrat	>=0.0071 或遺漏值	預測	1	1
2	8 (已選取)			目標是下列項目	26.99%	27.81%
				目標是下列項目	73.01%	72.19%
				% 正確的預測	73.01%	72.19%

互動式決策樹 (Interactive Decision Tree)



互動式決策樹 (Interactive Decision Tree)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3

- 資料來源
 - INQ2006
- 流程圖
- CaseStudy
- 模型套件

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

CaseStudy

INQ2006 → 樣本 → Stat Explore → 設算

資料分區 → 決策樹

編輯變數...

更新

執行

建立模型套件...

結果...

將路徑匯出作為 SAS 程式

剪下

複製(C)

刪除

重新命名

全選

選取節點

連接節點

中斷連接節點

100%

流程圖 日誌

執行完成

屬性	值
一般	
節點 ID	Tree
匯入的資料	
匯出的資料	
附註	
訓練	
變數	
互動式	
使用凍結樹	否
使用多重目標	否
精確度	4
分割規則	
閾值規則	ProbF
名目規則	ProbChisq
序數規則	熵
顯著水準	0.2
遺漏值	搜尋時使用
使用輸入一次	否
最大分支	2
最大深度	6
最小類別大小	5
分割精確度	4
節點	
分葉大小	5
規則數目	5
替代規則數目	0
分割大小	.
分割搜尋	
使用決策	否
使用事前機率	否

互動式決策樹 (Interactive Decision Tree)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3
資料來源
INQ2006
流程圖
CaseStudy
模型套件

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

CaseStudy

INQ2006 → 樣本 → Stat Explore → 計算

資料分區 → 決策樹

執行狀態

執行完成
流程圖: CaseStudy
路徑: 決策樹

確定 結果...

屬性	值
一般	
節點 ID	Tree
匯入的資料	
匯出的資料	
附註	
訓練	
變數	
互動式	
使用凍結樹	否
使用多重目標	否
精確度	4
分割規則	
間隔規則	ProbF
名目規則	ProbChisq
序數規則	熵
顯著水準	0.2
遺漏值	搜尋時使用
使用輸入一次	否
最大分支	2
最大深度	6
最小類別大小	5
分割精確度	4
節點	
分葉大小	5
規則數目	5
替代規則數目	0
分割大小	.
分割搜尋	
使用決策	否
使用事前機率	否
一般	

執行完成

互動式決策樹 (Interactive Decision Tree)

結果



決策樹 (Decision Tree)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3
資料來源
INQ2006
流程圖
CaseStudy
模型套件

樣本 調查 修改 模型 評估 公用程式 應用程式 時間序列

CaseStudy

INQ2006 → 樣本 → StatExplore → 設算

資料分區 → 決策樹 → 決策樹 (2)

屬性 值

一般	值
節點 ID	Tree2
匯入的資料	
匯出的資料	
附註	
訓練	
變數	
互動式	
使用凍結樹	否
使用多重目標	否
精確度	4
分割規則	
間隔準則	ProbF
名目準則	ProbChisq
序數準則	熵
顯著水準	0.2
遺漏值	搜尋時使用
使用輸入一次	否
最大分支	2
最大深度	6
最小類別大小	5
分割精確度	4
節點	
分葉大小	5
規則數目	5
替代規則數目	0
分割大小	.
分割搜尋	
使用決策	否
使用事前機率	否

執行完成

決策樹 (Decision Tree) — 執行 (全自動)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3

- 資料來源
 - INQ2006
- 流程圖
- CaseStudy
- 模型套件

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

CaseStudy

INQ2006 → 樣本 → Stat Explore → 計算

資料分區 → 決策樹 → 決策樹 (2)

編輯變數...
更新
執行
建立模型套件...
結果...
將路徑匯出作為 SAS 程式

剪下
複製(C)
刪除
重新命名
全選
選取節點
連接節點
中斷連接節點

100%

流程圖 日誌

執行完成

屬性	值
一般	
節點 ID	Tree2
匯入的資料	
匯出的資料	
附註	
訓練	
變數	
互動式	
使用凍結樹	否
使用多重目標	否
精確度	4
分割規則	
閾值規則	ProbF
名目規則	ProbChisq
序數規則	熵
顯著水準	0.2
遺漏值	搜尋時使用
使用輸入一次	否
最大分支	2
最大深度	6
最小類別大小	5
分割精確度	4
節點	
分葉大小	5
規則數目	5
替代規則數目	0
分割大小	.
分割搜尋	
使用決策	否
使用事前機率	否
一般	

決策樹 (Decision Tree)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3
資料來源
INQ2006
流程圖
CaseStudy
模型套件

屬性	值
一般	
節點 ID	Tree2
匯入的資料	
匯出的資料	
附註	
訓練	
變數	
互動式	
使用凍結樹	否
使用多重目標	否
精確度	4
分割規則	
間隔準則	ProbF
名目準則	ProbChisq
序數準則	熵
顯著水準	0.2
遺漏值	搜尋時使用
使用輸入一次	否
最大分支	2
最大深度	6
最小類別大小	5
分割精確度	4
節點	
分葉大小	5
規則數目	5
替代規則數目	0
分割大小	.
分割搜尋	
使用決策	否
使用事前機率	否

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

CaseStudy

INQ2006 → 樣本 → Stat Explore → 試算

資料分區 → 決策樹 → 決策樹 (2)

確認

是否要執行此路徑?
流程圖: CaseStudy
路徑: 決策樹 (2)

是(Y) 否(N)

執行完成

決策樹(自動)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3

- 資料來源
 - INQ2006
- 流程圖
- CaseStudy
- 模型套件

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

CaseStudy

INQ2006 → 樣本 → Stat Explore → 設算

資料分區 → 決策樹 → 決策樹(2)

執行狀態

執行完成
流程圖: CaseStudy
路徑: 決策樹 (2)

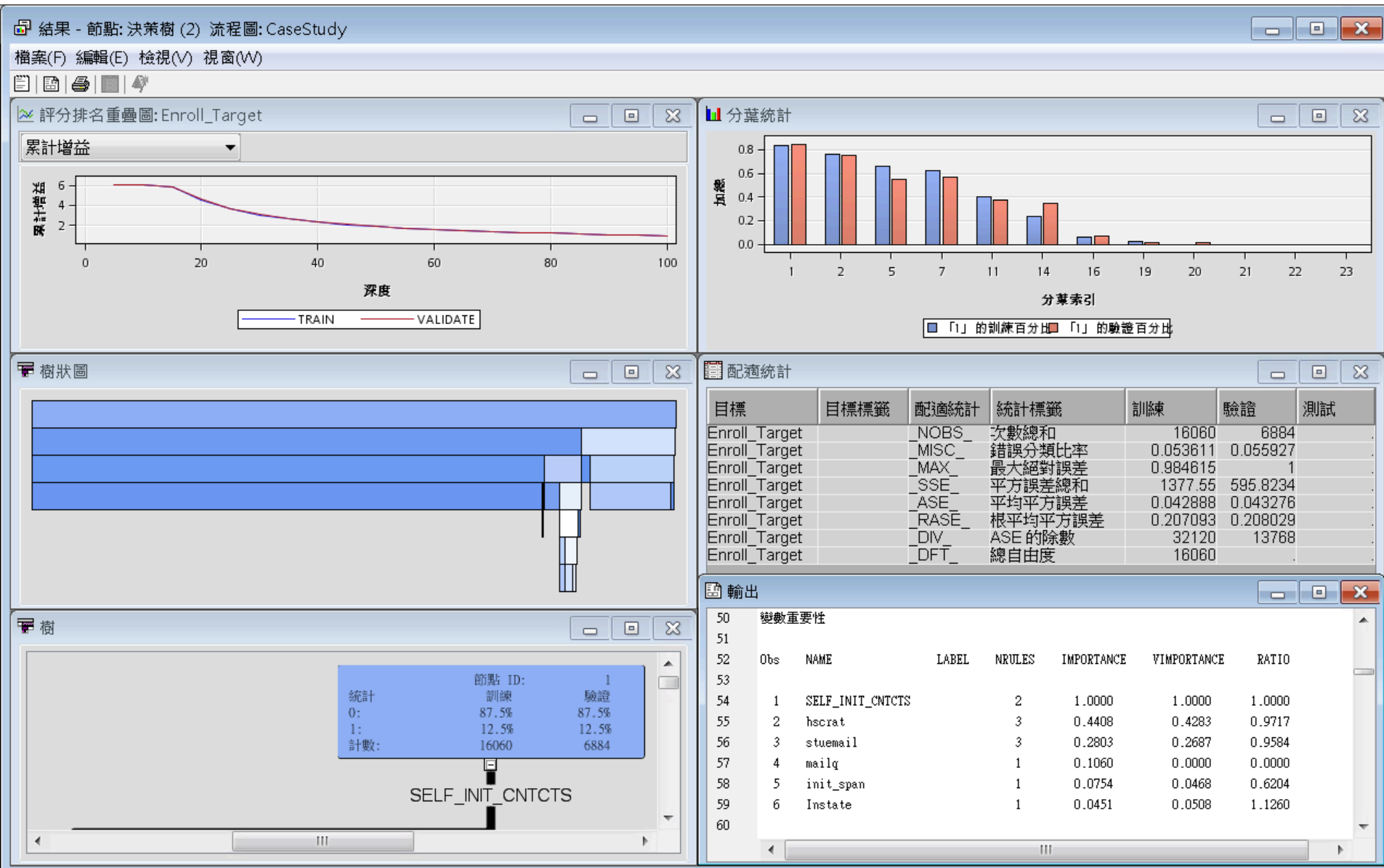
確定 結果...

屬性 值

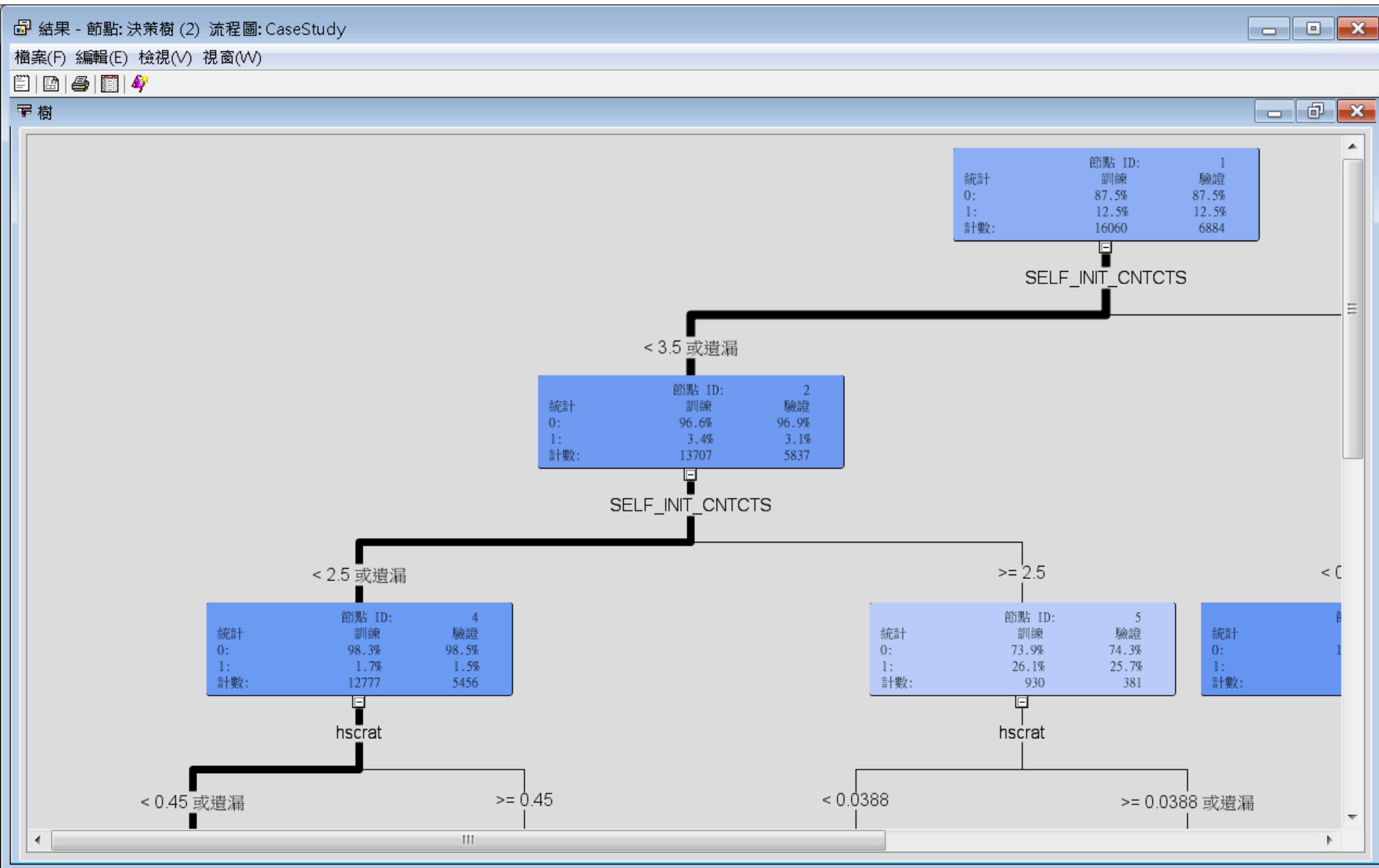
屬性	值
一般	
節點 ID	Tree2
匯入的資料	
匯出的資料	
附註	
訓練	
變數	
互動式	
使用凍結樹	否
使用多重目標	否
精確度	4
分割規則	
間隔準則	ProbF
名目準則	ProbChisq
序數準則	熵
顯著水準	0.2
遺漏值	搜尋時使用
使用輸入一次	否
最大分支	2
最大深度	6
最小類別大小	5
分割精確度	4
節點	
分葉大小	5
規則數目	5
替代規則數目	0
分割大小	.
分割搜尋	
使用決策	否
使用事前機率	否
一般	

執行完成

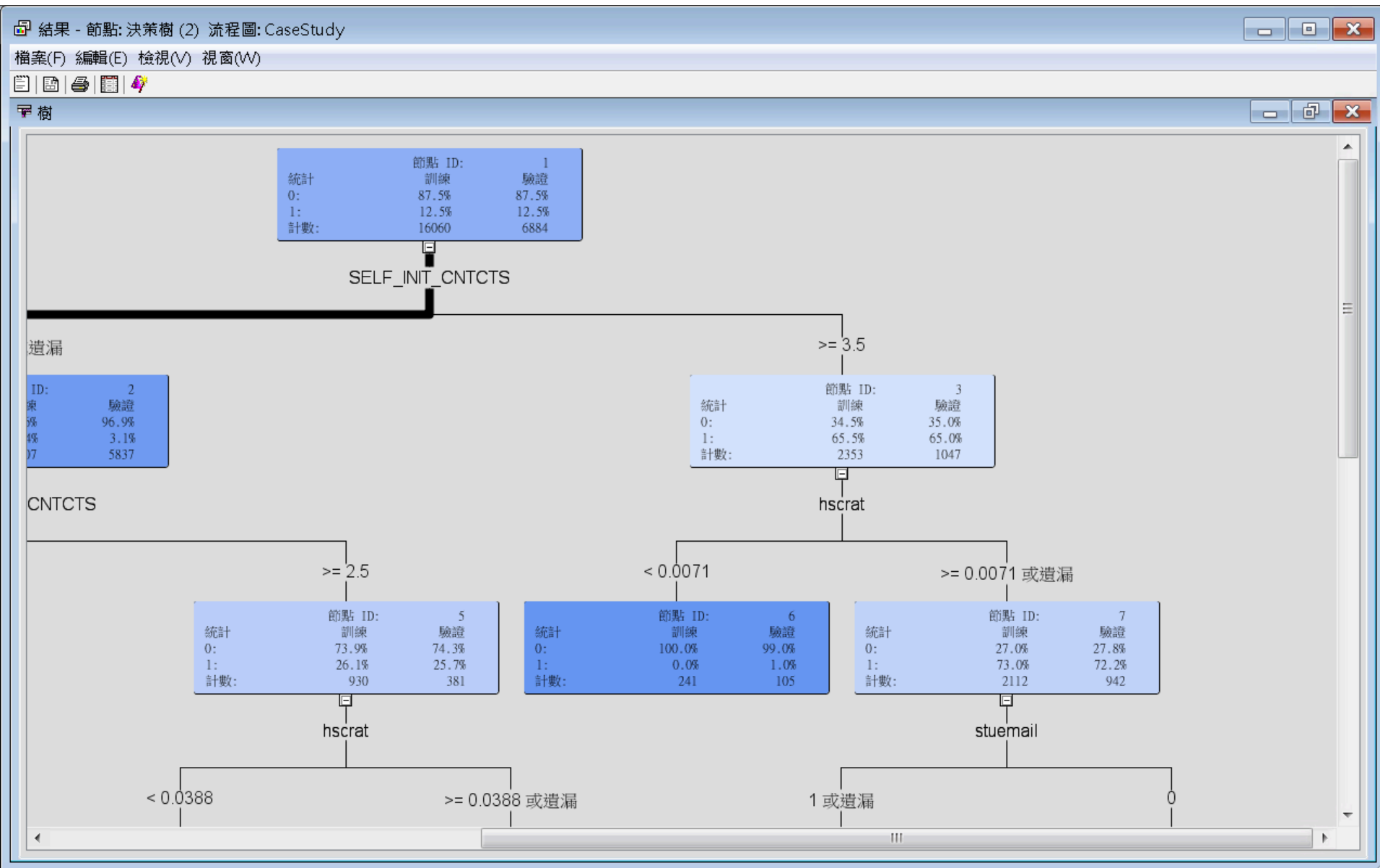
決策樹(自動)結果



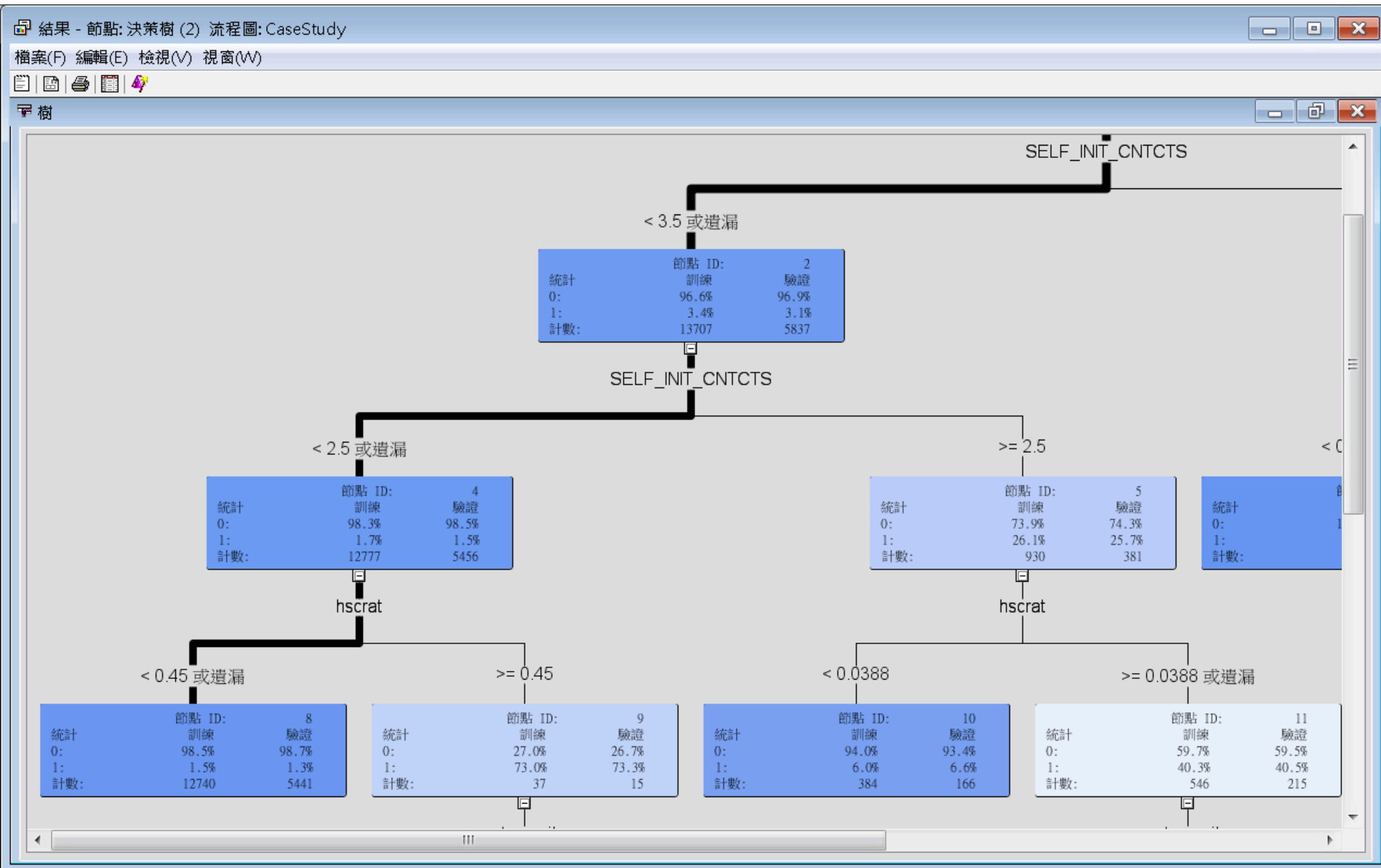
決策樹(自動)結果



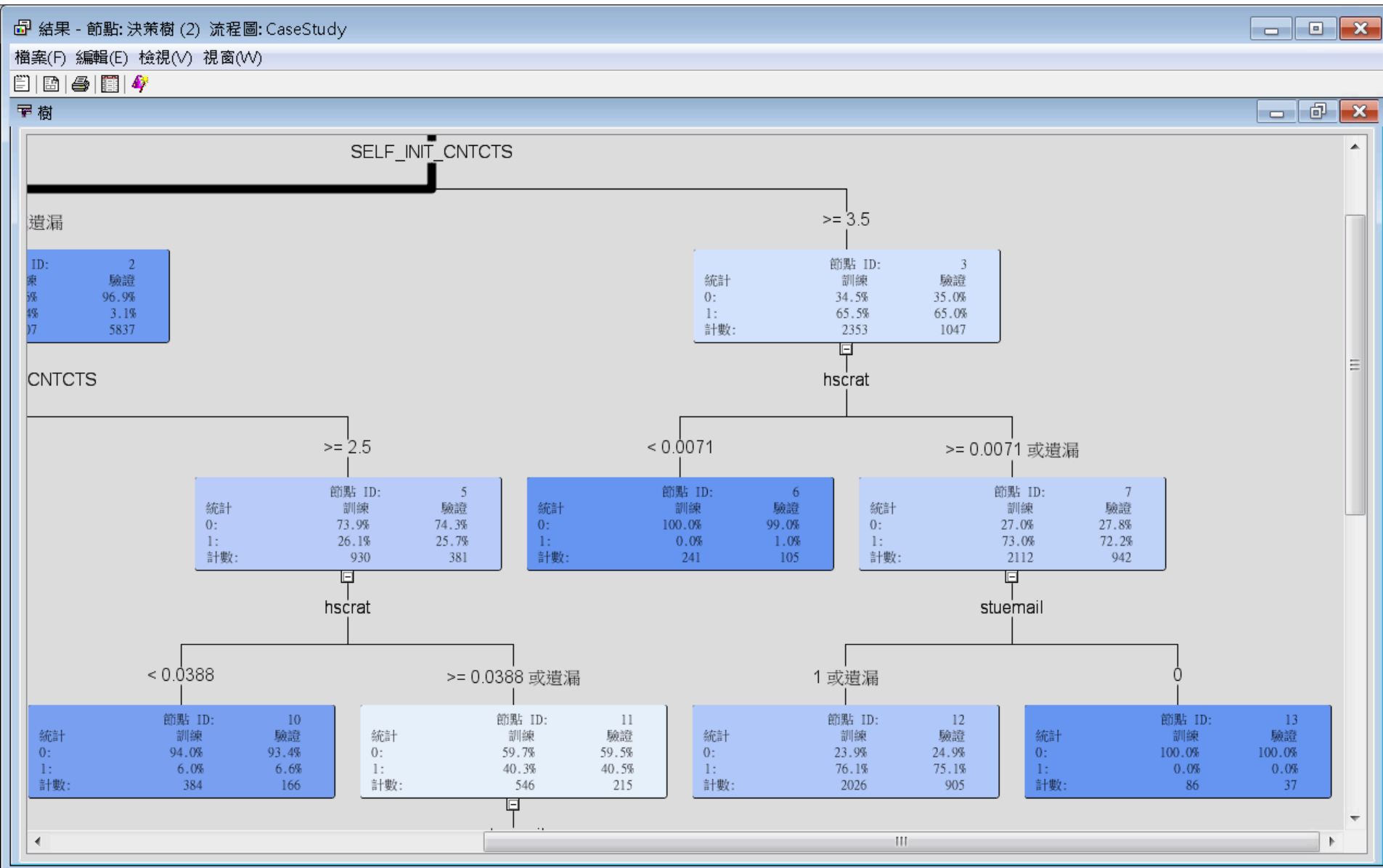
決策樹(自動)結果



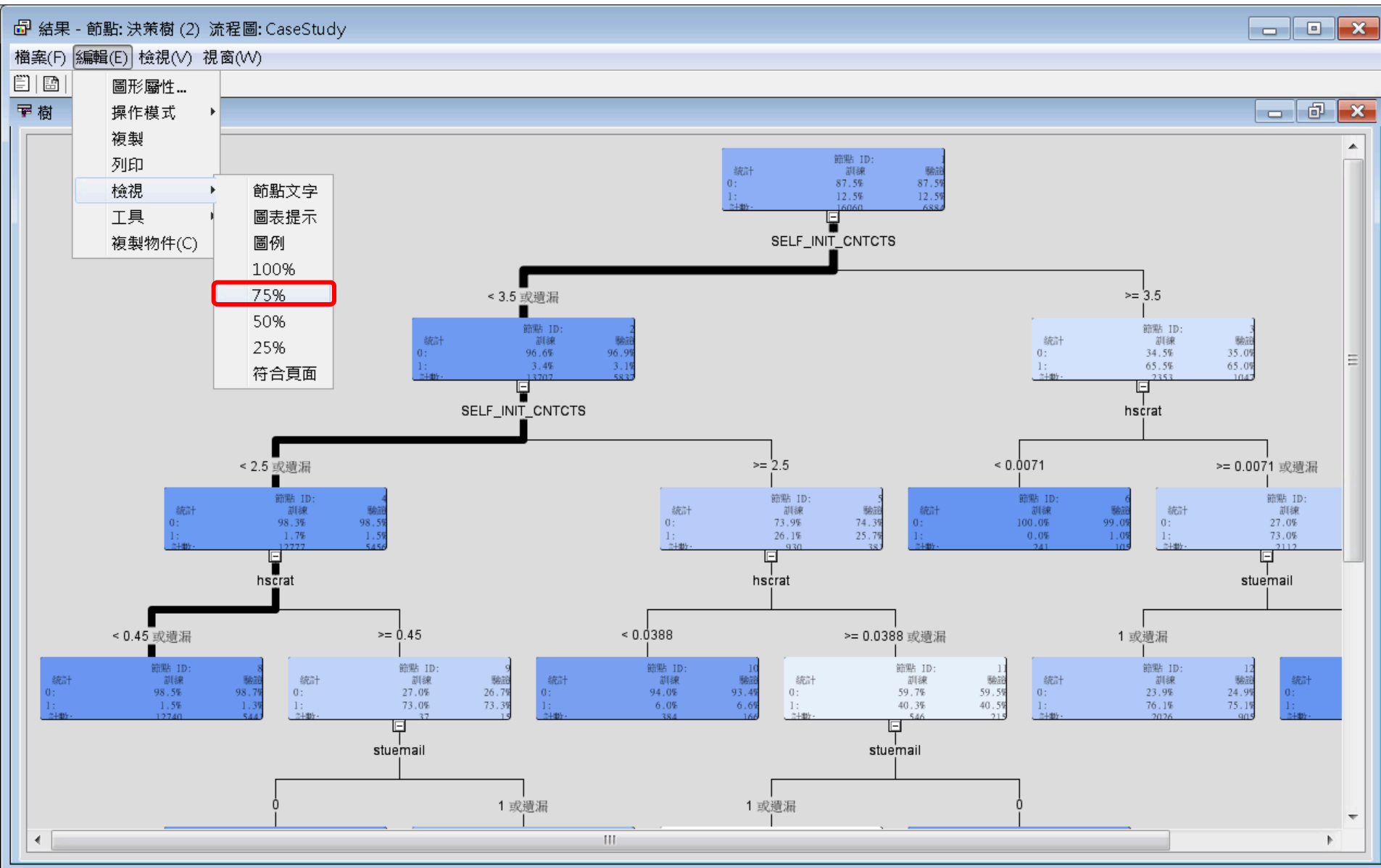
決策樹(自動)結果



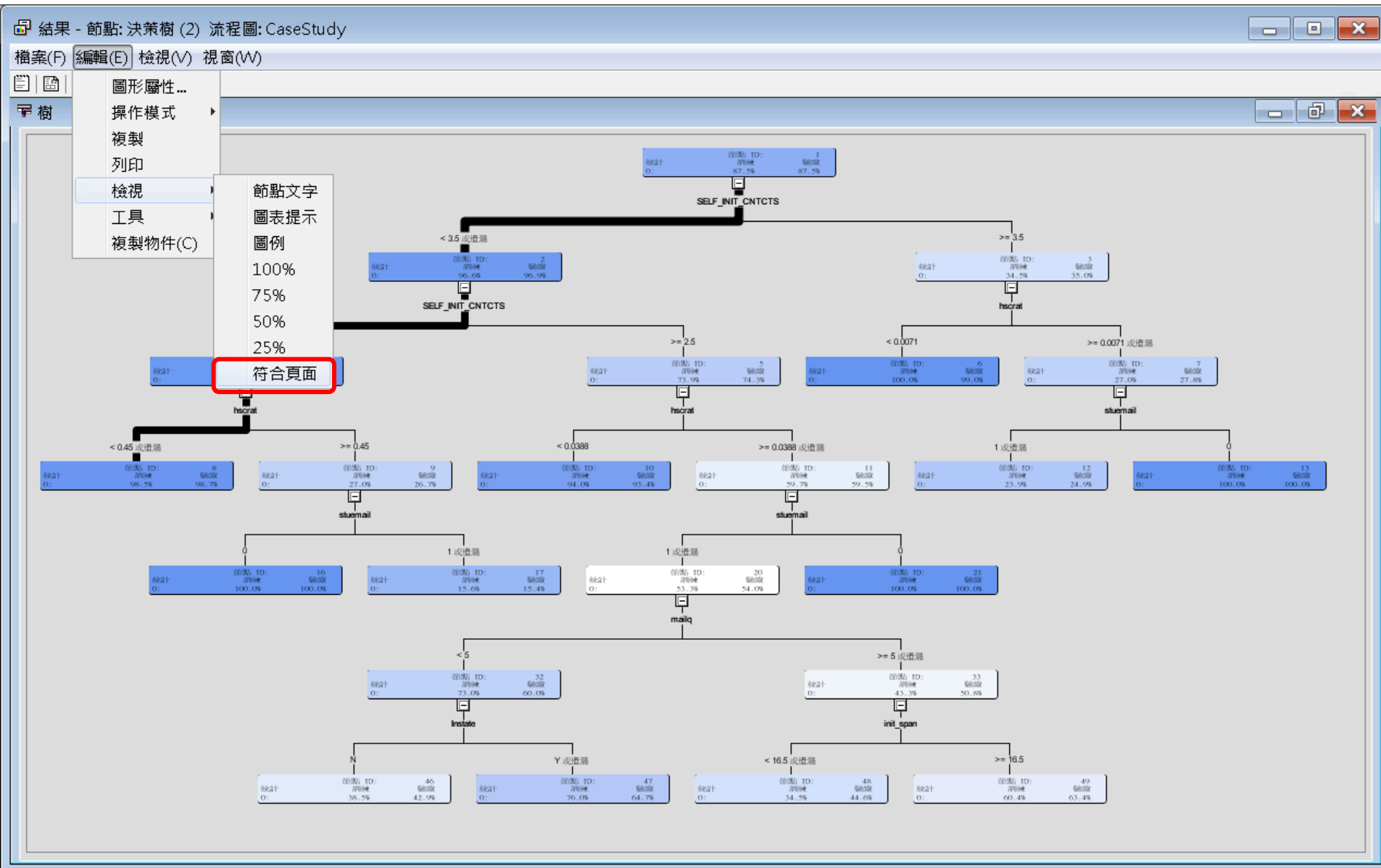
決策樹(自動)結果



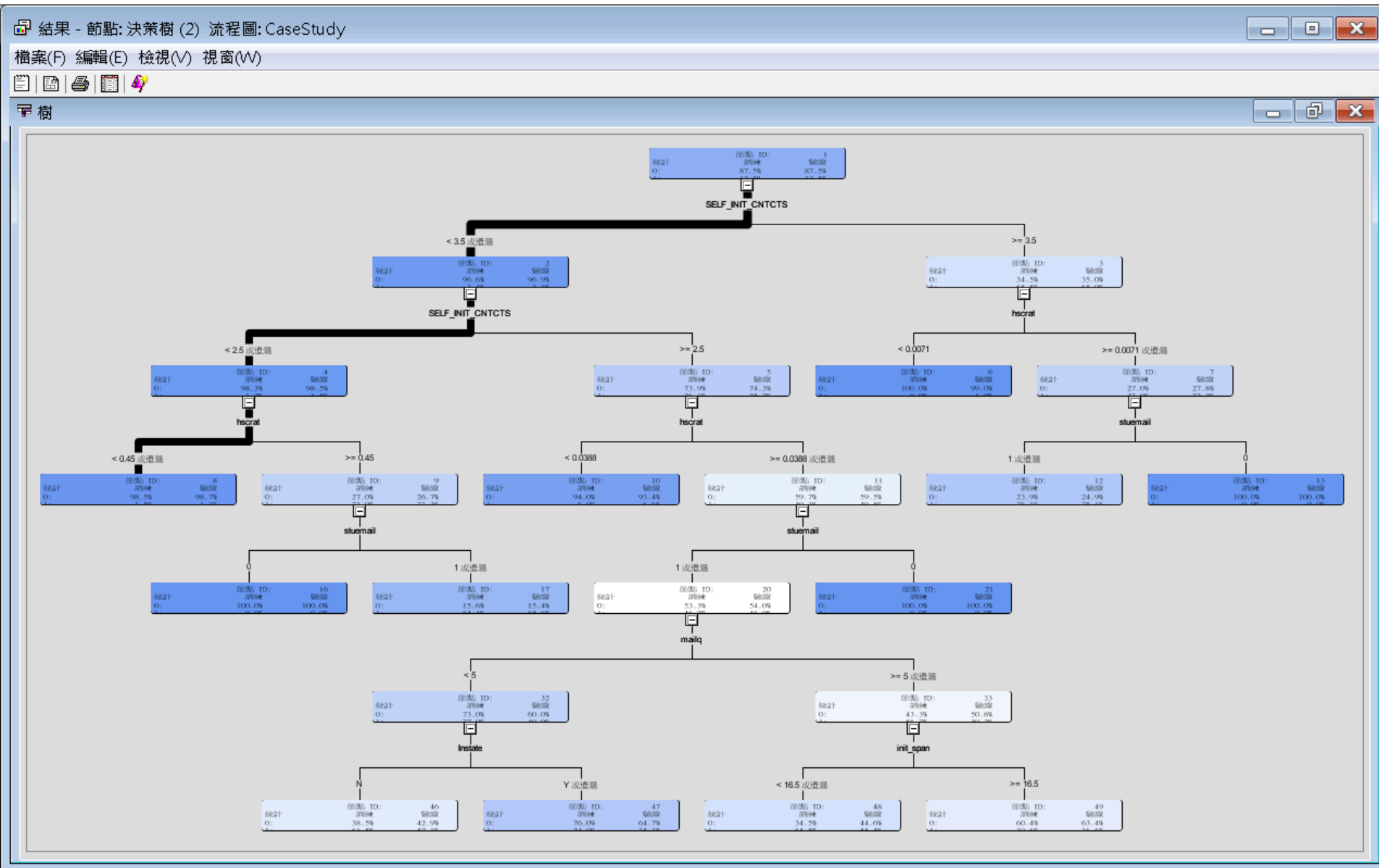
決策樹(自動)結果



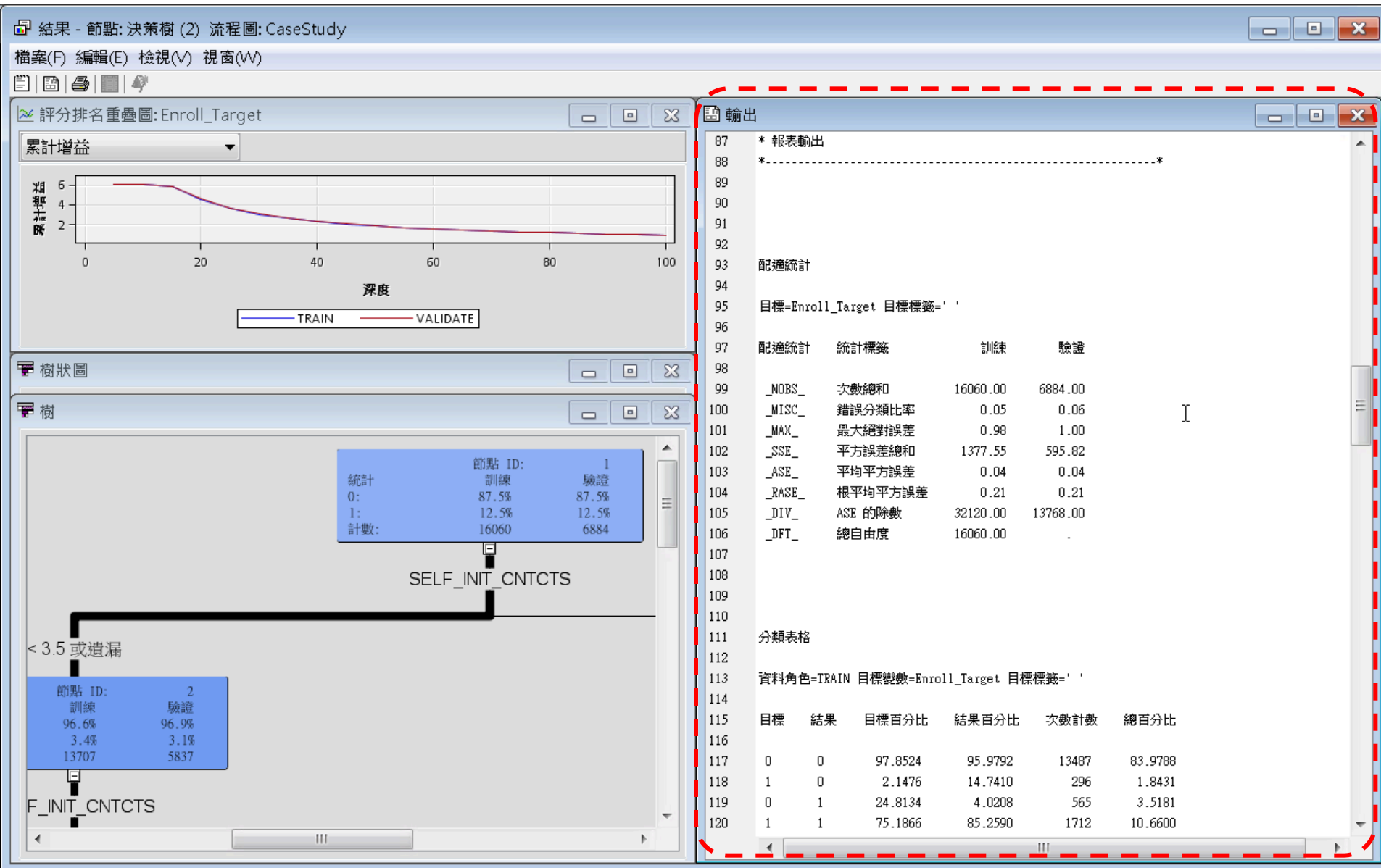
決策樹 (自動) 結果



決策樹(自動)結果



決策樹(自動)結果



模型－迴歸 (Regression)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3

- 資料來源
 - INQ2006
- 流程圖
- CaseStudy
- 模型套件

屬性 值

一般	值
節點 ID	Tree2
匯入的資料	
匯出的資料	
附註	
訓練	
變數	
互動式	
使用凍結樹	否
使用多重目標	否
精確度	4
分割規則	
間隔準則	ProbF
名目準則	ProbChisq
序數準則	熵
顯著水準	0.2
遺漏值	搜尋時使用
使用輸入一次	否
最大分支	2
最大深度	6
最小類別大小	5
分割精確度	4
節點	
分葉大小	5
規則數目	5
替代規則數目	0
分割大小	.
分割搜尋	
使用決策	否
使用事前機率	否

CaseStudy

樣本 勘查 修改 模型 評估 公用程式 應用 迴歸

配適線性和羅吉斯迴歸模型。

INQ2006 → 樣本 → Stat Explore → 設算

資料分區 → 決策樹 → 決策樹 (1)

執行完成

模型－迴歸 (Regression)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3
資料來源
INQ2006
流程圖
CaseStudy
模型套件

屬性 值

一般

節點 ID Reg

匯入的資料

匯出的資料

附註

訓練

變數

方程式

主要效果 是

雙因子交互作用 否

多項式條件 否

多項式次數 2

使用者條件 否

條件編輯器

類別目標

迴歸類型 羅吉斯迴歸

連結函數 對數優劣比

模型選項

隱藏截距 否

輸入編碼 偏差

模型選取

選取模型 無

選取準則 預設

使用選取預設 是

選取選項

最佳化選項

技術 預設

預設最佳化 是

最大迭代 0

最大函數呼叫次數 0

一般

執行完成

CaseStudy

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

INQ2006 樣本 Stat Explore 設算

資料分區 決策樹 決策樹(2) 迴歸

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

流程图 日誌

模型－迴歸 (Regression)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3

- 資料來源
 - INQ2006
- 流程圖
- CaseStudy
- 模型套件

屬性 值

一般	值
節點 ID	Reg
匯入的資料	
匯出的資料	
附註	
訓練變數	
方程式	
主要效果	是
雙因子交互作用	否
多項式條件	否
多項式次數	2
使用者條件	否
條件編輯器	
類別目標	
迴歸類型	羅吉斯迴歸
連結函數	對數優劣比
模型選項	
隱藏截距	否
輸入編碼	偏差
模型選取	
選取模型	無
選取準則	預設
使用選取預設	是
選取選項	
最佳化選項	
技術	預設
預設最佳化	是
最大迭代	0
最大函數呼叫次數	0

CaseStudy

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

INQ2006 → 樣本 → Stat Explore → 計算

資料分區 → 決策樹 → 決策樹 (2) → 迴歸

編輯變數...
更新
執行
建立模型套件...
結果...
將路徑匯出作為 SAS 程式
剪下
複製(C)
刪除
重新命名
全選
選取節點
連接節點
中斷連接節點

100%

執行完成

模型－迴歸 (Regression)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3
資料來源
 INQ2006
 流程圖
 CaseStudy
 模型套件

屬性 值

一般	值
節點 ID	Reg
匯入的資料	
匯出的資料	
附註	
訓練	
變數	
方程式	
主要效果	是
雙因子交互作用	否
多項式條件	否
多項式次數	2
使用者條件	否
條件編輯器	
類別目標	
迴歸類型	羅吉斯迴歸
連結函數	對數優劣比
模型選項	
隱藏截距	否
輸入編碼	偏差
模型選取	
選取模型	無
選取準則	預設
使用選取預設	是
選取選項	
最佳化選項	
技術	預設
預設最佳化	是
最大迭代	0
最大函數呼叫次數	0

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

CaseStudy

INQ2006 → 樣本 → StatExplore → 設算

資料分區 → 決策樹 → 決策樹(2) → 迴歸

確認

是否要執行此路徑?
流程圖: CaseStudy
路徑: 迴歸

是(Y) 否(N)

執行完成

模型－迴歸 (Regression)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3

- 資料來源
 - INQ2006
- 流程圖
- CaseStudy
- 模型套件

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

CaseStudy

INQ2006 → 樣本 → Stat Explore → 計算

資料分區 → 決策樹 → 決策樹(2) → 迴歸

執行狀態

執行完成
流程圖: CaseStudy
路徑: 迴歸

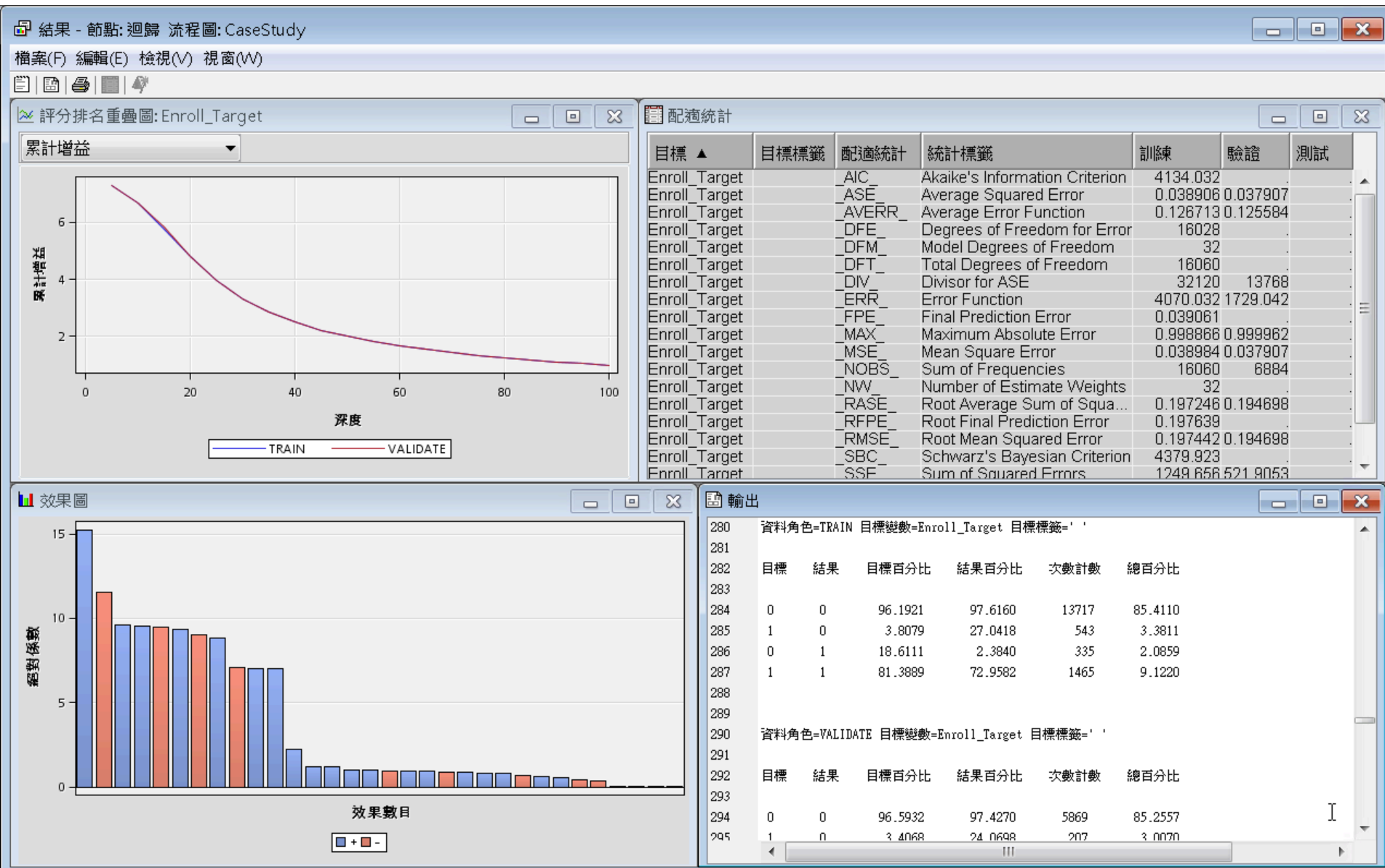
確定 結果...

屬性 值

屬性	值
一般	
節點 ID	Reg
匯入的資料	
匯出的資料	
附註	
訓練	
變數	
方程式	
主要效果	是
雙因子交互作用	否
多項式條件	否
多項式次數	2
使用者條件	否
條件編輯器	
類別目標	
迴歸類型	羅吉斯迴歸
連結函數	對數優劣比
模型選項	
隱藏截距	否
輸入編碼	偏差
模型選取	
選取模型	無
選取準則	預設
使用選取預設	是
選取選項	
最佳化選項	
技術	預設
預設最佳化	是
最大迭代	0
最大函數呼叫次數	0

執行完成

迴歸 (Regression) 結果



評估－模型比較 (Model Comparison)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3

- 資料來源
 - INQ2006
- 流程圖
- CaseStudy
- 模型套件

屬性 值

一般	值
節點 ID	Reg
匯入的資料	
匯出的資料	
附註	
訓練變數	
方程式	
主要效果	是
雙因子交互作用	否
多項式條件	否
多項式次數	2
使用者條件	否
條件編輯器	
類別目標	
迴歸類型	羅吉斯迴歸
連結函數	對數優劣比
模型選項	
隱藏截距	否
輸入編碼	偏差
模型選取	
選取模型	無
選取準則	預設
使用選取預設	是
選取選項	
最佳化選項	
技術	預設
預設最佳化	是
最大迭代	0
最大函數呼叫次數	0

CaseStudy

模型比較

比較來自先前建模節點的模型和預測。

INQ2006 → 樣本 → Stat Explore → 設算

資料分區 → 決策樹 → 決策樹(2) → 迴歸

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

執行完成

評估－模型比較 (Model Comparison)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3

- 資料來源
 - INQ2006
- 流程圖
- CaseStudy
- 模型套件

屬性	值
一般	
節點 ID	MdlComp
匯入的資料	
匯出的資料	
附註	
訓練	
變數	
評估報表	
等距箱數的數目	20
ROC 圖	是
重新計算	否
模型選取	
選取資料	預設
選取統計	預設
網格選取統計	預設
選取表格	訓練
選取深度	10
評分	
選取編輯器	
報表	
選取的模型	
目標	
模型節點	
模型描述	
選取準則	
狀態	

CaseStudy

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

INQ2006 → 樣本 → Stat Explore → 設算

資料分區 → 決策樹 → 模型比較

資料分區 → 決策樹(2) → 模型比較

資料分區 → 迴歸 → 模型比較

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

執行完成

評估－模型比較 (Model Comparison)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3

- 資料來源
 - INQ2006
- 流程圖
- CaseStudy
- 模型套件

屬性	值
一般	
節點 ID	MdlComp
匯入的資料	
匯出的資料	
附註	
訓練	
變數	
評估報表	
等距箱數的數目	20
ROC 圖	是
重新計算	否
模型選取	
選取資料	預設
選取統計	預設
網格選取統計	預設
選取表格	訓練
選取深度	10
評分	
選取編輯器	
報表	
選取的模型	
目標	
模型節點	
模型描述	
選取準則	
狀態	

CaseStudy

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

INQ2006 → 樣本 → Stat Explore → 計算

資料分區 → 決策樹 → 模型比較

資料分區 → 決策樹(2) → 模型比較

資料分區 → 迴歸 → 模型比較

模型比較

編輯變數...

更新

執行

建立模型套件...

結果...

將路徑匯出作為 SAS 程式

剪下

複製(C)

刪除

重新命名

全選

選取節點

連接節點

中斷連接節點

100%

執行完成

評估－模型比較 (Model Comparison)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3
資料來源
INQ2006
流程圖
CaseStudy
模型套件

屬性 值

一般	值
節點 ID	MdlComp
匯入的資料	
匯出的資料	
附註	
訓練	
變數	
評估報表	
等距箱數的數目	20
ROC 圖	是
重新計算	否
模型選取	
選取資料	預設
選取統計	預設
網格選取統計	預設
選取表格	訓練
選取深度	10
評分	
選取編輯器	
報表	
選取的模型	
目標	
模型節點	
模型描述	
選取準則	
狀態	

樣本 調查 修改 模型 評估 公用程式 應用程式 時間序列

CaseStudy

INQ2006 → 樣本 → Stat Explore → 設算

資料分區 → 決策樹 → 模型比較

資料分區 → 決策樹(2) → 模型比較

資料分區 → 迴歸 → 模型比較

確認

是否要執行此路徑?
流程圖: CaseStudy
路徑: 模型比較

是(Y) 否(N)

執行完成

評估－模型比較 (Model Comparison)

Enterprise Miner - EM_Project3

檔案(F) 編輯(E) 檢視(V) 動作(A) 選項(O) 視窗(W) 說明(H)

EM_Project3

- 資料來源
 - INQ2006
- 流程圖
 - CaseStudy
- 模型套件

屬性	值
一般	
節點 ID	MdlComp
匯入的資料	
匯出的資料	
附註	
訓練	
變數	
評估報表	
等距箱數的數目	20
ROC 圖	是
重新計算	否
模型選取	
選取資料	預設
選取統計	預設
網格選取統計	預設
選取表格	訓練
選取深度	10
評分	
選取編輯器	
報表	
選取的模型	
目標	Enroll_Target
模型節點	Reg
模型描述	迴歸
選取準則	有效: 錯誤分類比率
狀態	

CaseStudy

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

```
graph LR; INQ2006 --> 樣本; 樣本 --> StatExplore; StatExplore --> 設算; 資料分區 --> 決策樹; 資料分區 --> 決策樹2[決策樹 (2)]; 資料分區 --> 迴歸; 決策樹 --> 模型比較; 決策樹2 --> 模型比較; 迴歸 --> 模型比較; 設算 --> 模型比較;
```

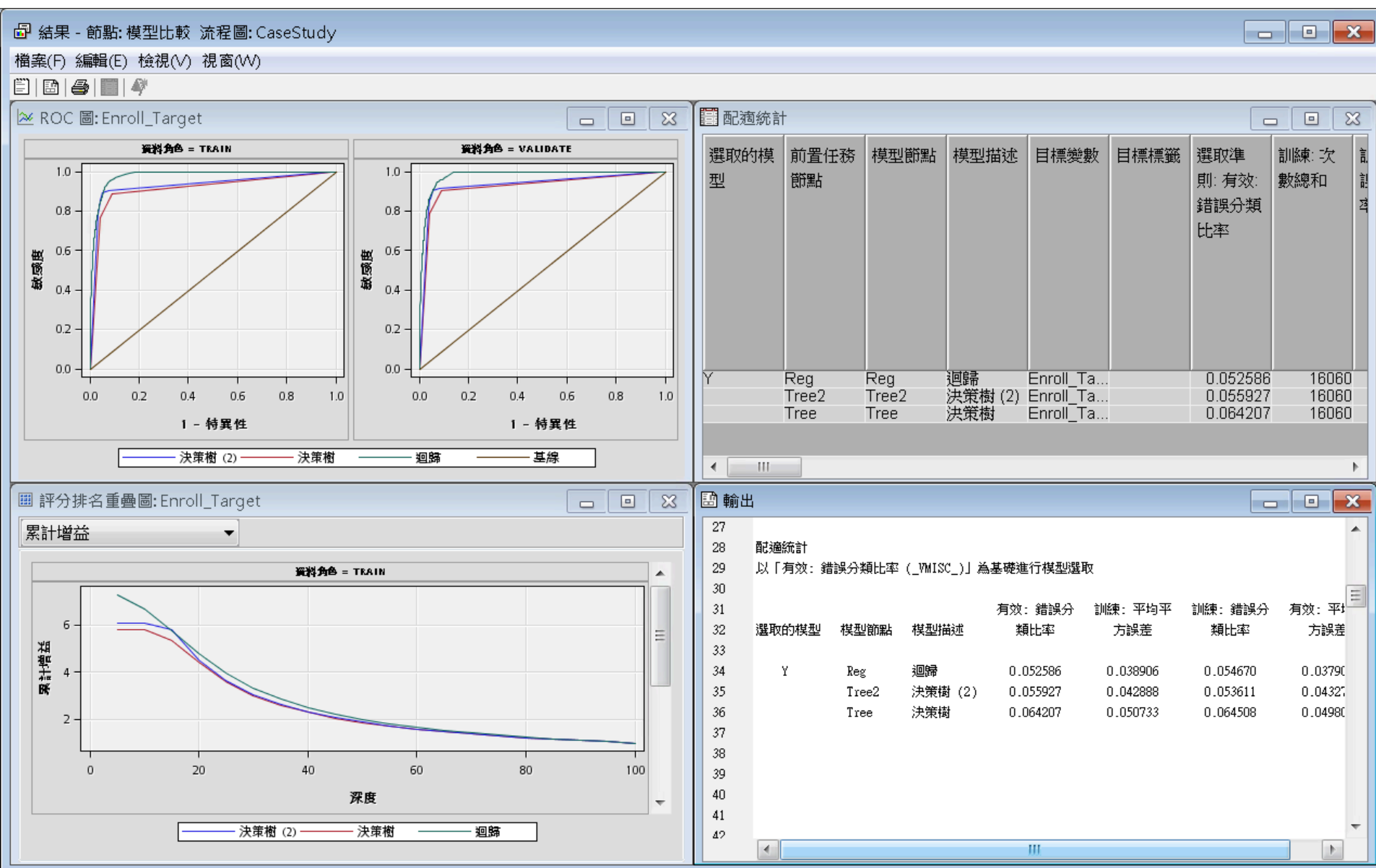
執行狀態

執行完成
流程圖: CaseStudy
路徑: 模型比較

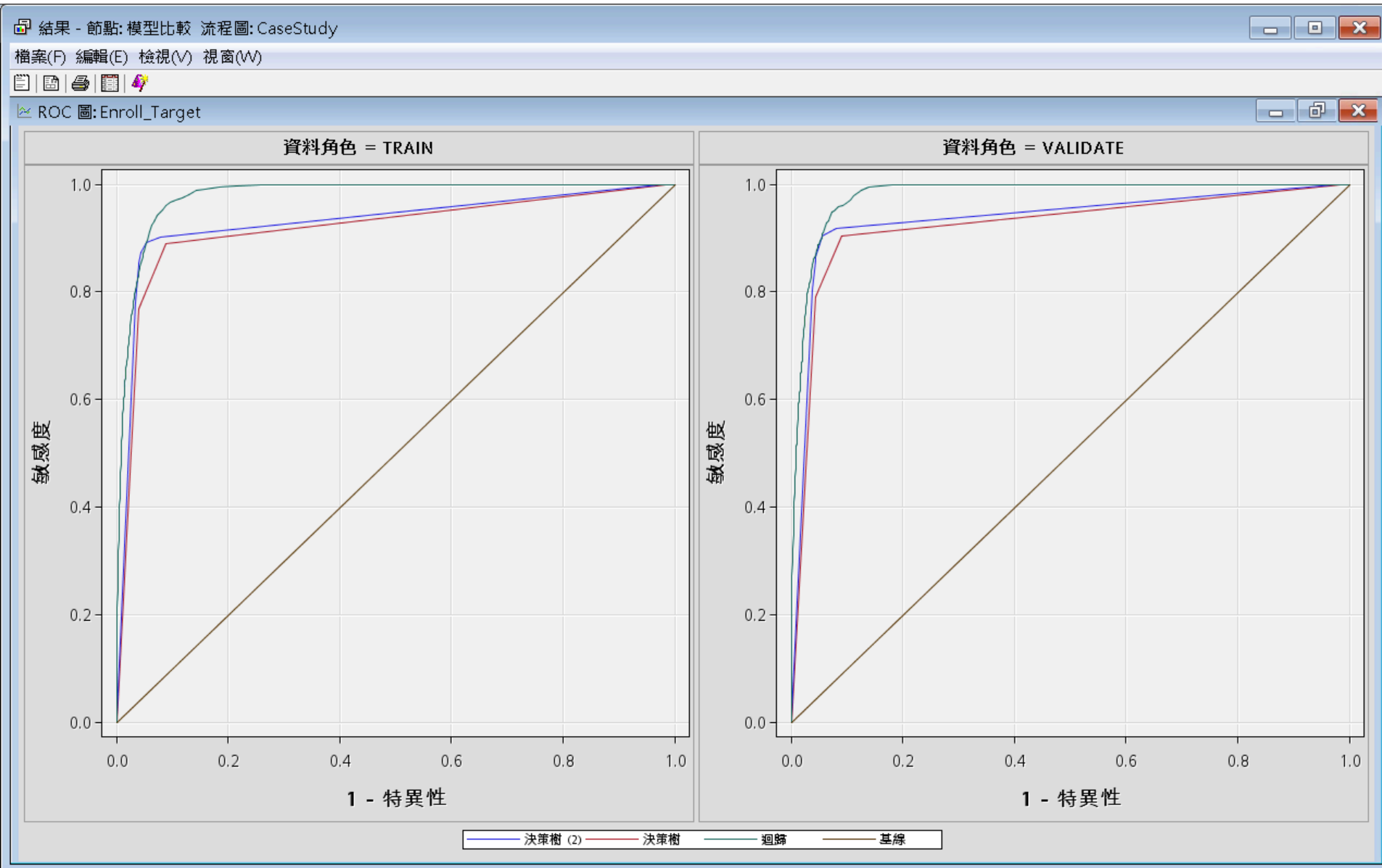
確定 結果...

執行完成

跨模型比較(Model Comparison) 結果



ROC



Reference

- 資料採礦運用：以SAS Enterprise Miner為工具，李淑娟，2015，SAS賽仕電腦軟體
- Jim Georges, Jeff Thompson and Chip Wells, Applied Analytics Using SAS Enterprise Miner, SAS, 2010
- SAS Enterprise Miner Course Notes, 2014, SAS
- SAS Enterprise Miner Training Course, 2014, SAS
- SAS Enterprise Guide Training Course, 2014, SAS