



商業智慧實務

Practices of Business Intelligence

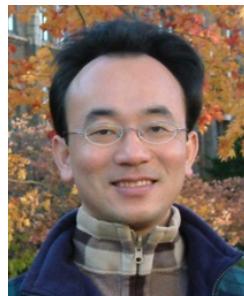
商業智慧的資料探勘

(Data Mining for Business Intelligence)

1022BI05

MI4

Wed, 9,10 (16:10-18:00) (B113)



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/>

2014-03-19



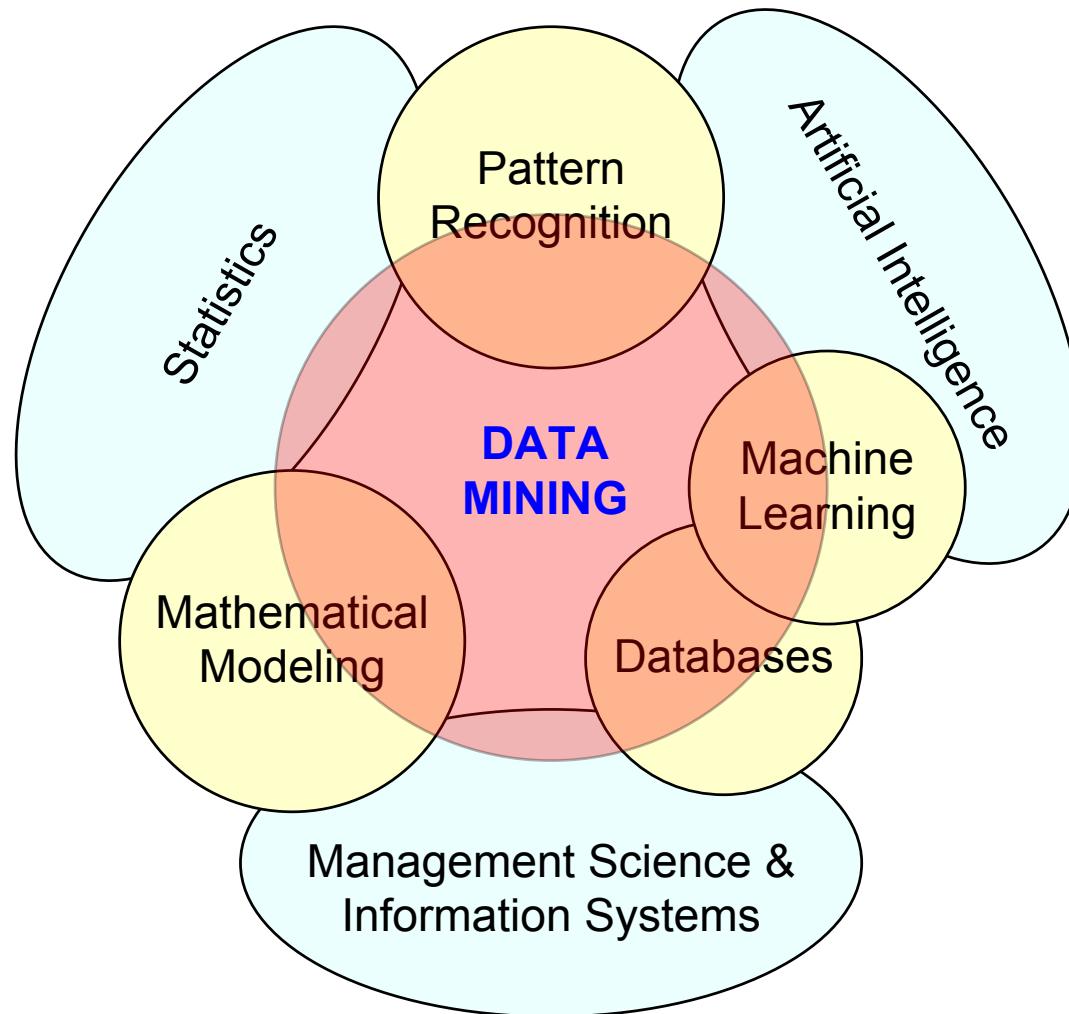
課程大綱 (Syllabus)

週次 (Week)	日期 (Date)	內容 (Subject/Topics)
1	103/02/19	商業智慧導論 (Introduction to Business Intelligence)
2	103/02/26	管理決策支援系統與商業智慧 (Management Decision Support System and Business Intelligence)
3	103/03/05	企業績效管理 (Business Performance Management)
4	103/03/12	資料倉儲 (Data Warehousing)
5	103/03/19	商業智慧的資料探勘 (Data Mining for Business Intelligence)
6	103/03/26	商業智慧的資料探勘 (Data Mining for Business Intelligence)
7	103/04/02	教學行政觀摩日 (Off-campus study)
8	103/04/09	資料科學與巨量資料分析 (Data Science and Big Data Analytics)

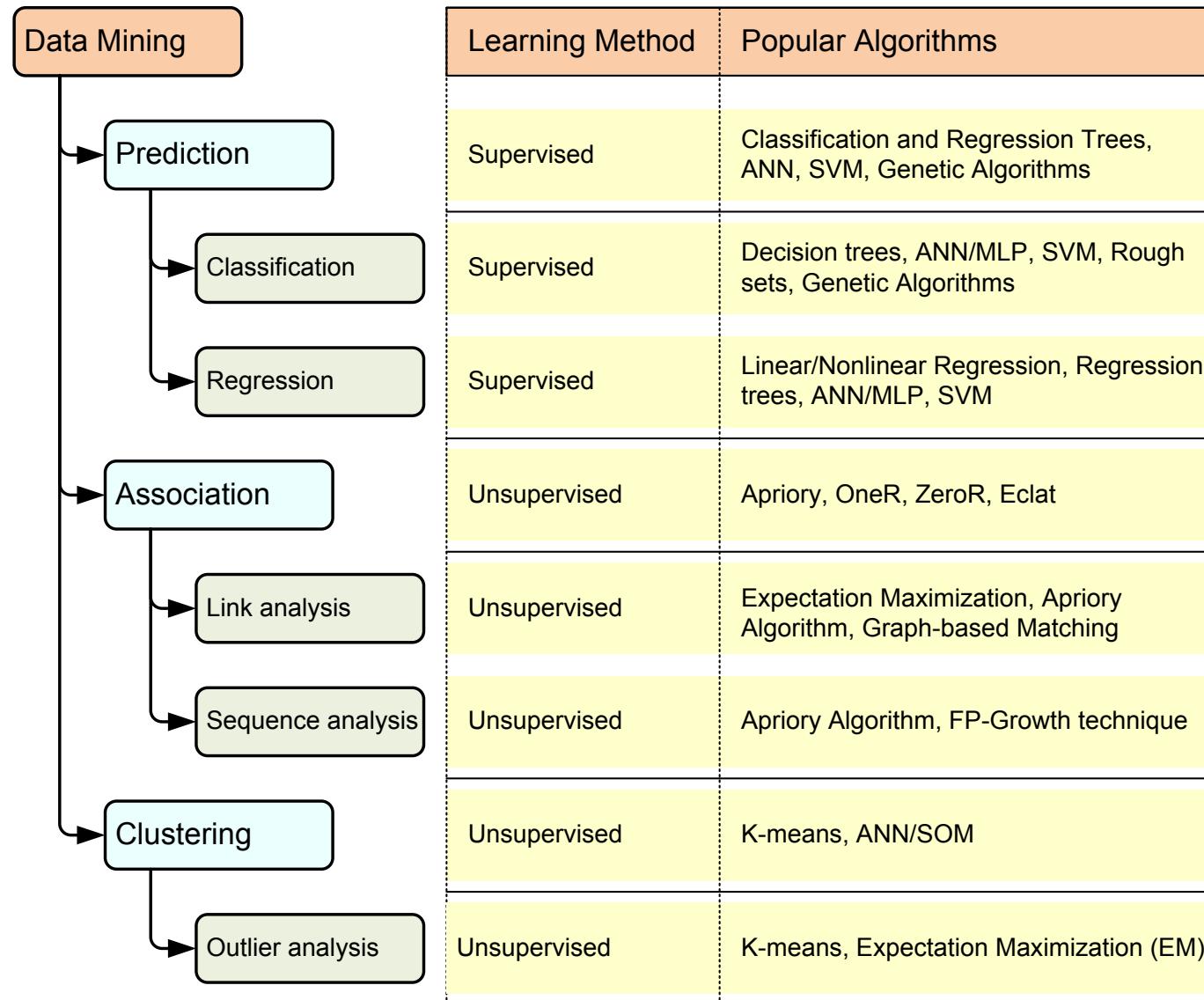
課程大綱 (Syllabus)

週次	日期	內容 (Subject/Topics)
9	103/04/16	期中報告 (Midterm Project Presentation)
10	103/04/23	期中考試週 (Midterm Exam)
11	103/04/30	文字探勘與網路探勘 (Text and Web Mining)
12	103/05/07	意見探勘與情感分析 (Opinion Mining and Sentiment Analysis)
13	103/05/14	社會網路分析 (Social Network Analysis)
14	103/05/21	期末報告 (Final Project Presentation)
15	103/05/28	畢業考試週 (Final Exam)

Data Mining at the Intersection of Many Disciplines

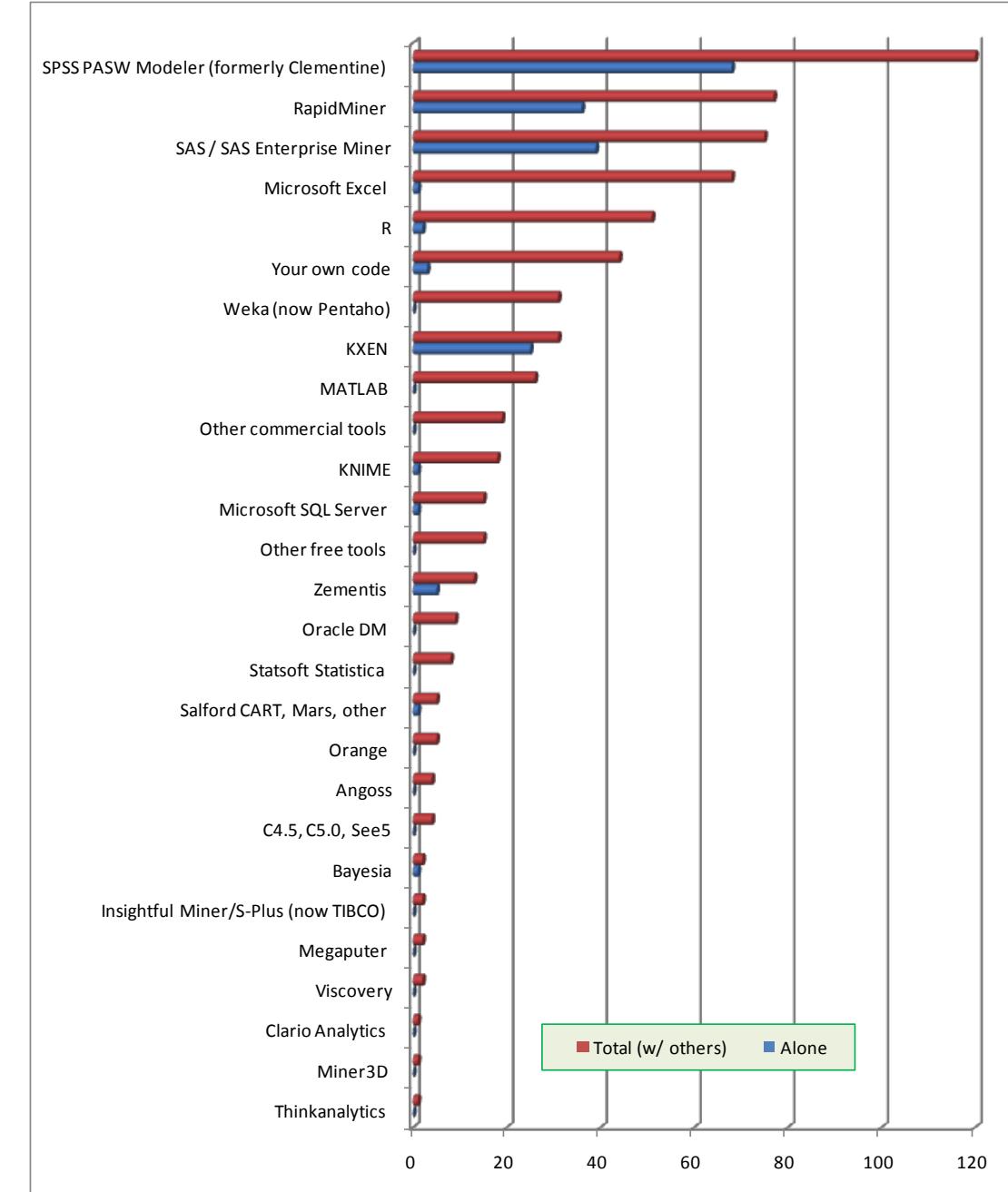


A Taxonomy for Data Mining Tasks



Data Mining Software

- Commercial
 - SPSS - PASW (formerly Clementine)
 - SAS - Enterprise Miner
 - IBM - Intelligent Miner
 - StatSoft – Statistical Data Miner
 - ... many more
- Free and/or Open Source
 - Weka
 - RapidMiner...

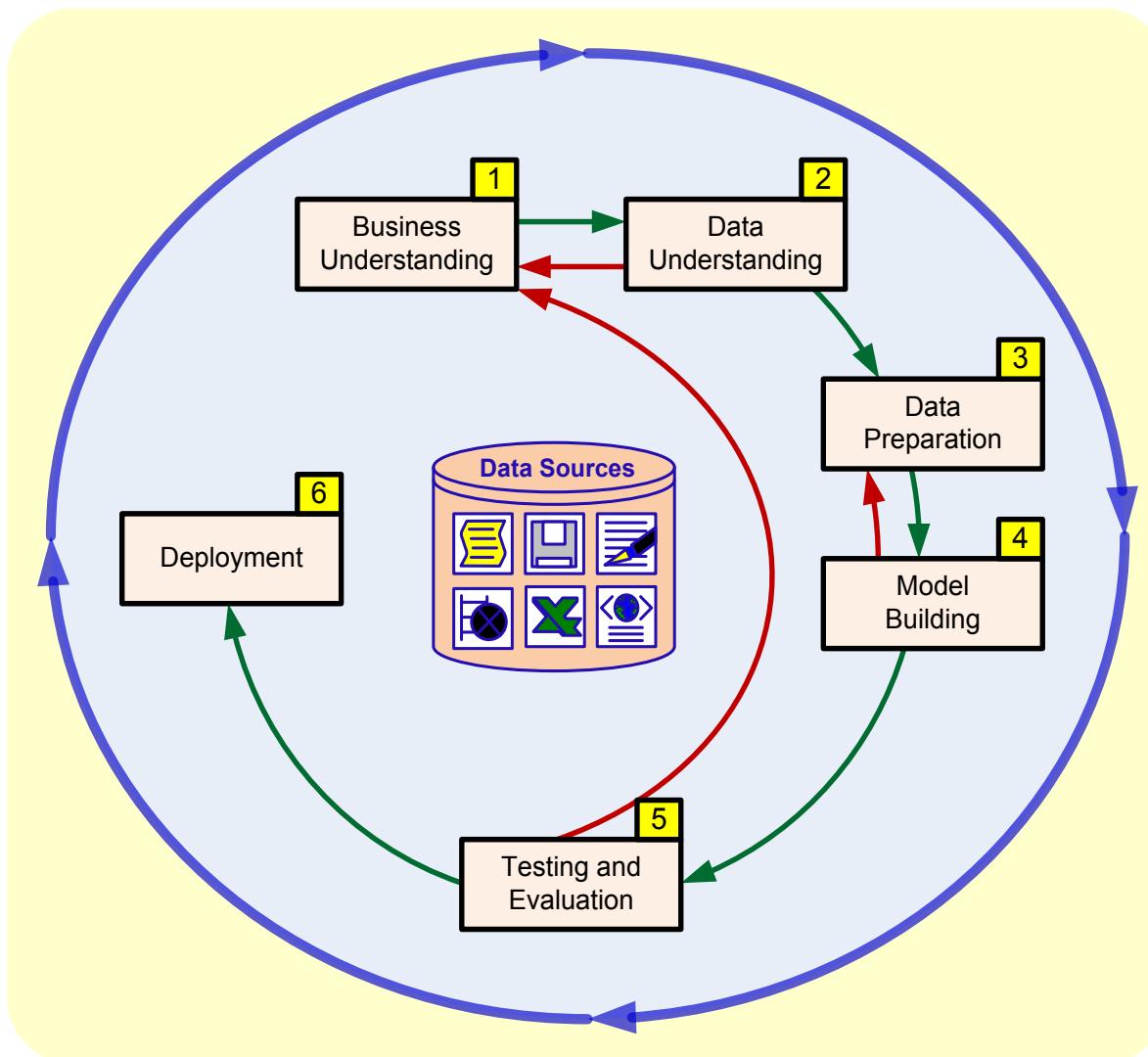


Source: KDNuggets.com, May 2009

Data Mining Process

- A manifestation of best practices
- A systematic way to conduct DM projects
- Different groups has different versions
- Most common standard processes:
 - CRISP-DM
(Cross-Industry Standard Process for Data Mining)
 - SEMMA
(Sample, Explore, Modify, Model, and Assess)
 - KDD
(Knowledge Discovery in Databases)

Data Mining Process: CRISP-DM



Data Mining Process: CRISP-DM

Step 1: Business Understanding

Step 2: Data Understanding

Step 3: Data Preparation (!)

Step 4: Model Building

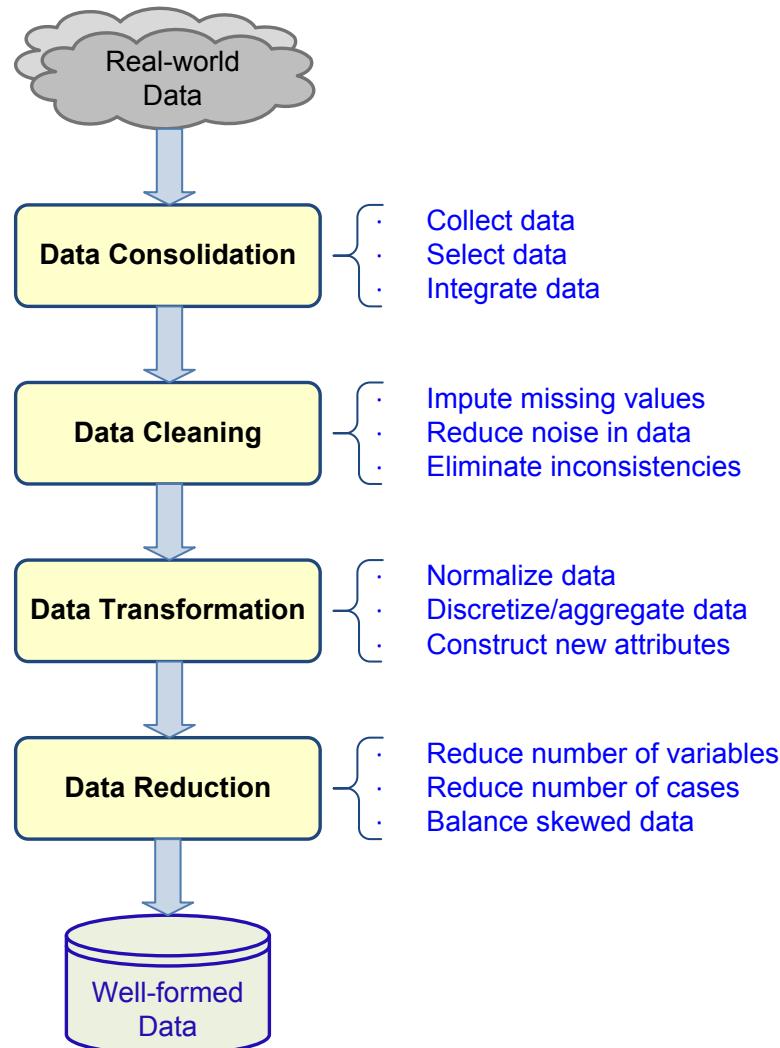
Step 5: Testing and Evaluation

Step 6: Deployment

Accounts for
~85% of total
project time

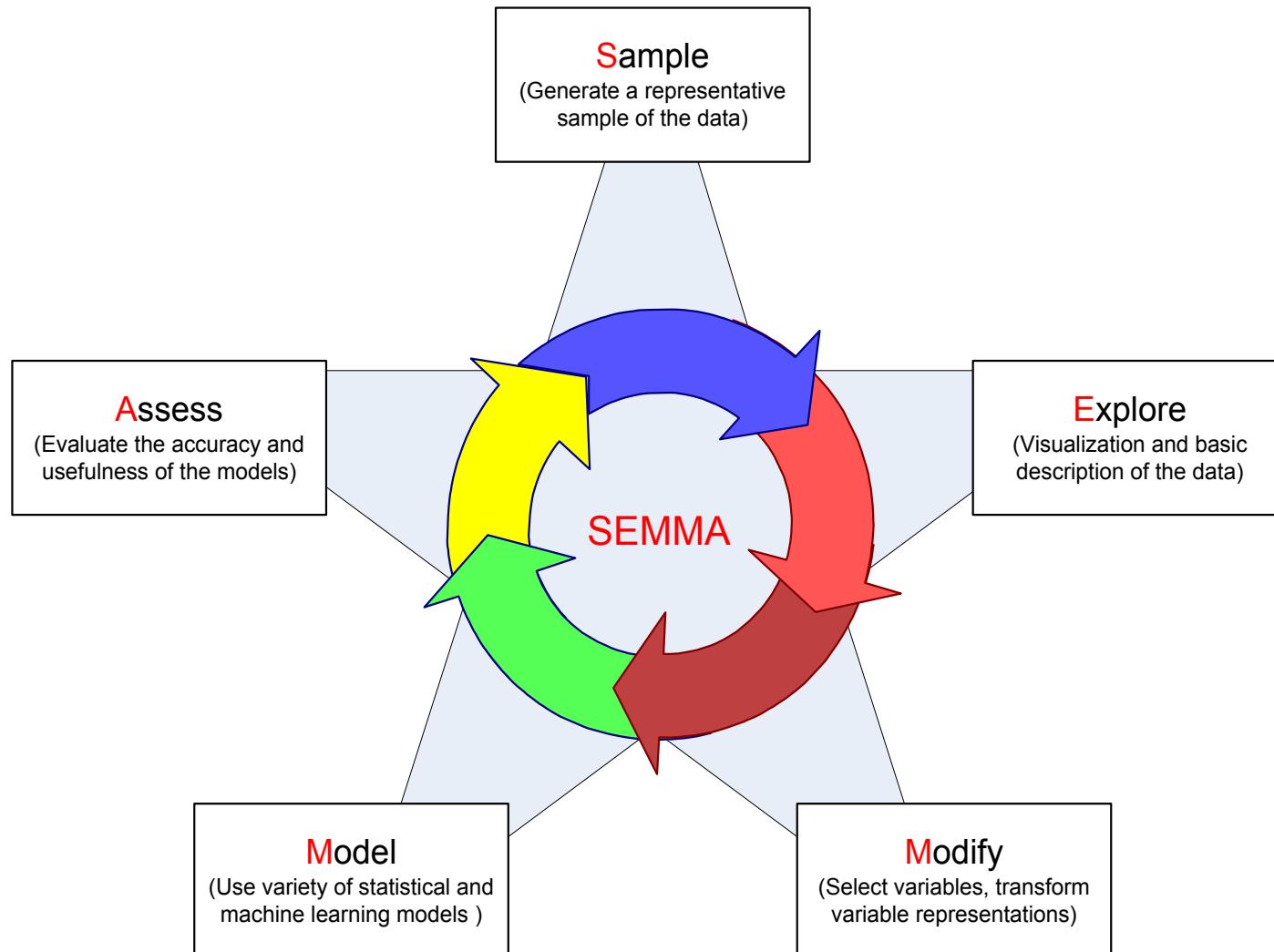
- The process is highly repetitive and experimental (DM: art versus science?)

Data Preparation – A Critical DM Task



Data Mining Process:

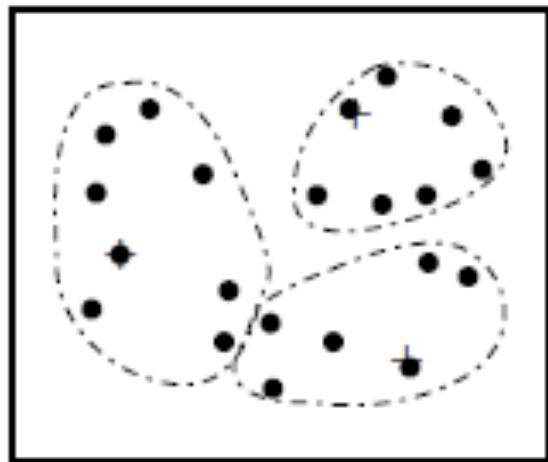
SEMMA



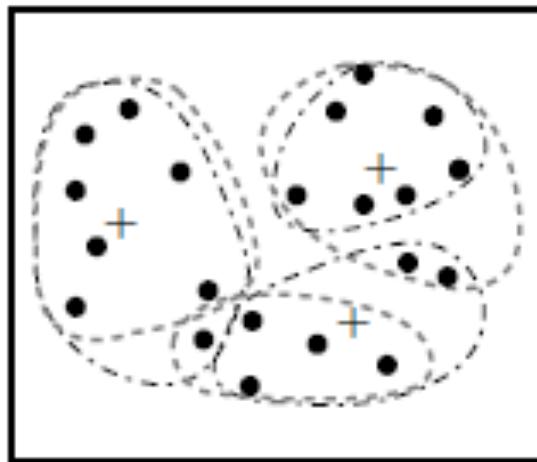
Cluster Analysis

- Used for automatic identification of **natural groupings** of things
- Part of the machine-learning family
- Employ **unsupervised learning**
- Learns the clusters of things from past data, then assigns new instances
- There is not an output variable
- Also known as **segmentation**

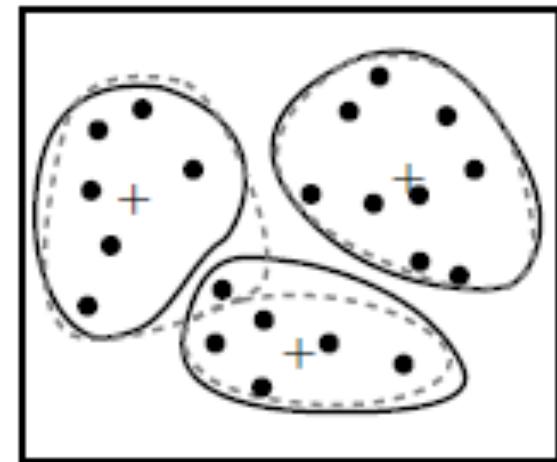
Cluster Analysis



(a)



(b)



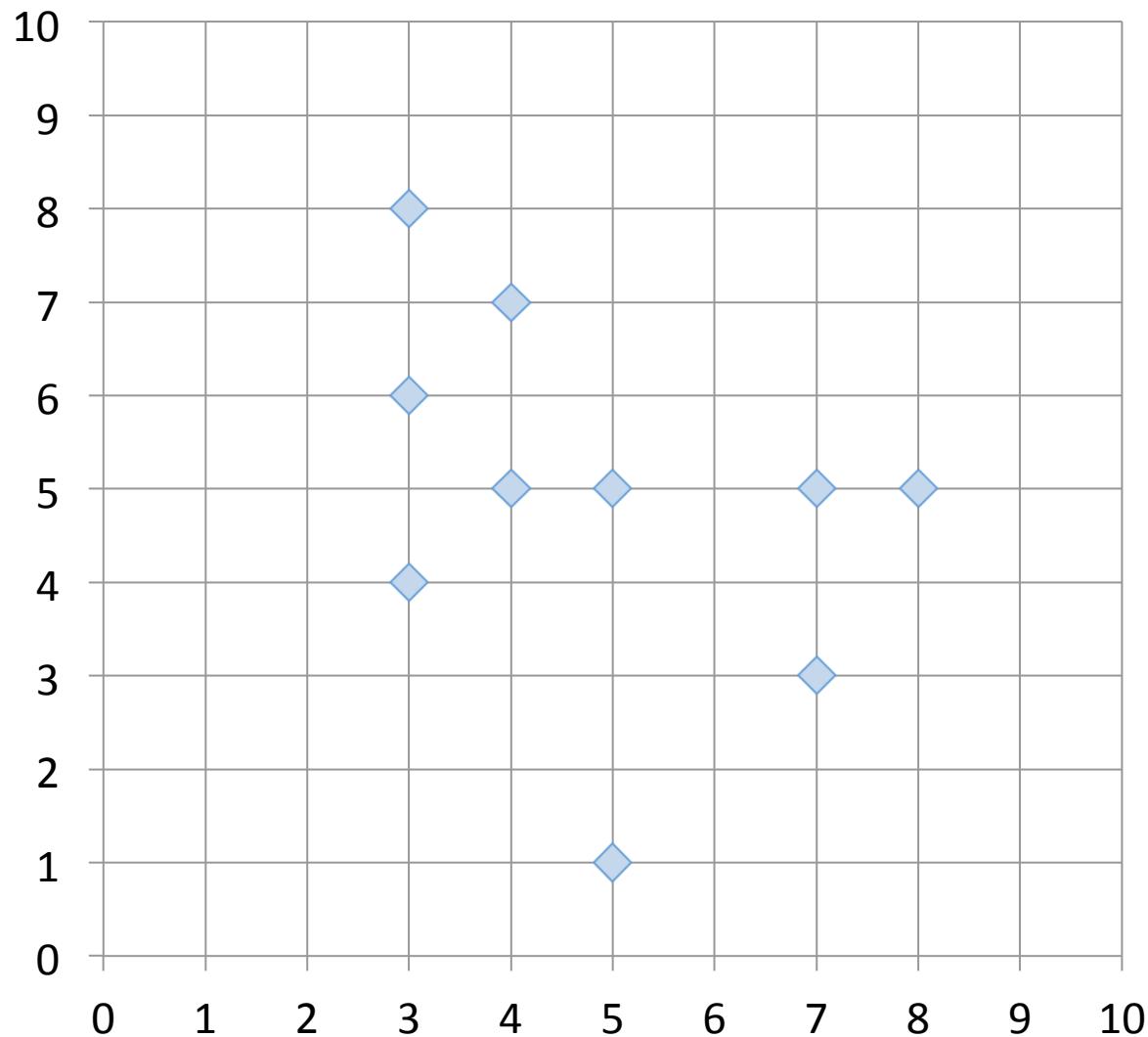
(c)

Clustering of a set of objects based on the *k-means method*.
(The mean of each cluster is marked by a “+”.)

Cluster Analysis

- Clustering results may be used to
 - Identify natural **groupings of customers**
 - Identify rules for assigning new cases to classes for targeting/diagnostic purposes
 - Provide characterization, definition, labeling of populations
 - Decrease the size and complexity of problems for other data mining methods
 - Identify **outliers** in a specific domain (e.g., rare-event detection)

Example of Cluster Analysis



Point	P	P(x,y)
p01	a	(3, 4)
p02	b	(3, 6)
p03	c	(3, 8)
p04	d	(4, 5)
p05	e	(4, 7)
p06	f	(5, 1)
p07	g	(5, 5)
p08	h	(7, 3)
p09	i	(7, 5)
p10	j	(8, 5)

Cluster Analysis for Data Mining

- How many clusters?
 - There is not a “truly optimal” way to calculate it
 - Heuristics are often used
 1. Look at the sparseness of clusters
 2. Number of clusters = $(n/2)^{1/2}$ (n: no of data points)
 3. Use Akaike information criterion (AIC)
 4. Use Bayesian information criterion (BIC)
- Most cluster analysis methods involve the use of a **distance measure** to calculate the closeness between pairs of items
 - Euclidian versus Manhattan (rectilinear) distance

***k*-Means Clustering Algorithm**

- k : pre-determined number of clusters
- Algorithm (**Step 0:** determine value of k)

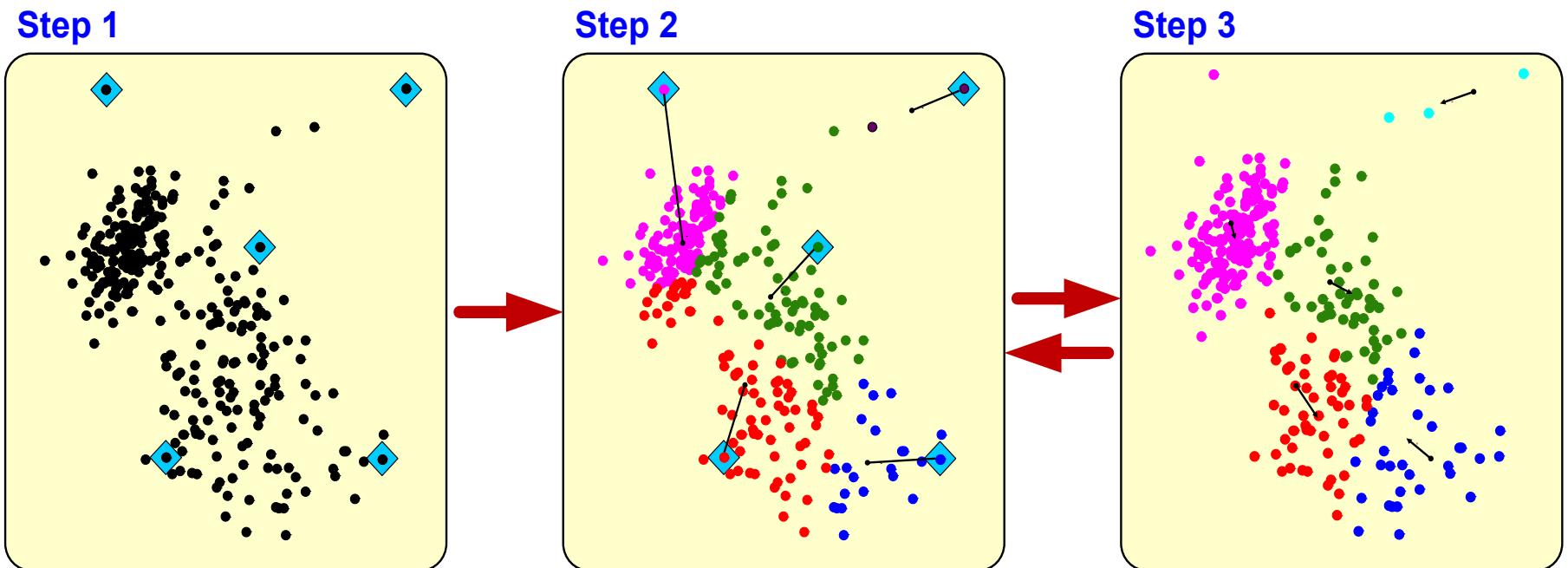
Step 1: Randomly generate k random points as initial cluster centers

Step 2: Assign each point to the nearest cluster center

Step 3: Re-compute the new cluster centers

Repetition step: Repeat steps 2 and 3 until some convergence criterion is met (usually that the assignment of points to clusters becomes stable)

Cluster Analysis for Data Mining - k -Means Clustering Algorithm



Similarity and Dissimilarity Between Objects

- Distances are normally used to measure the similarity or dissimilarity between two data objects
- Some popular ones include: *Minkowski distance*:

$$d(i, j) = \sqrt[q]{(|x_{i1} - x_{j1}|^q + |x_{i2} - x_{j2}|^q + \dots + |x_{ip} - x_{jp}|^q)}$$

where $i = (x_{i1}, x_{i2}, \dots, x_{ip})$ and $j = (x_{j1}, x_{j2}, \dots, x_{jp})$ are two p -dimensional data objects, and q is a positive integer

- If $q = 1$, d is *Manhattan distance*

$$d(i, j) = |x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \dots + |x_{ip} - x_{jp}|$$

Similarity and Dissimilarity Between Objects (Cont.)

- If $q = 2$, d is Euclidean distance:

$$d(i,j) = \sqrt{(|x_{i_1} - x_{j_1}|^2 + |x_{i_2} - x_{j_2}|^2 + \dots + |x_{i_p} - x_{j_p}|^2)}$$

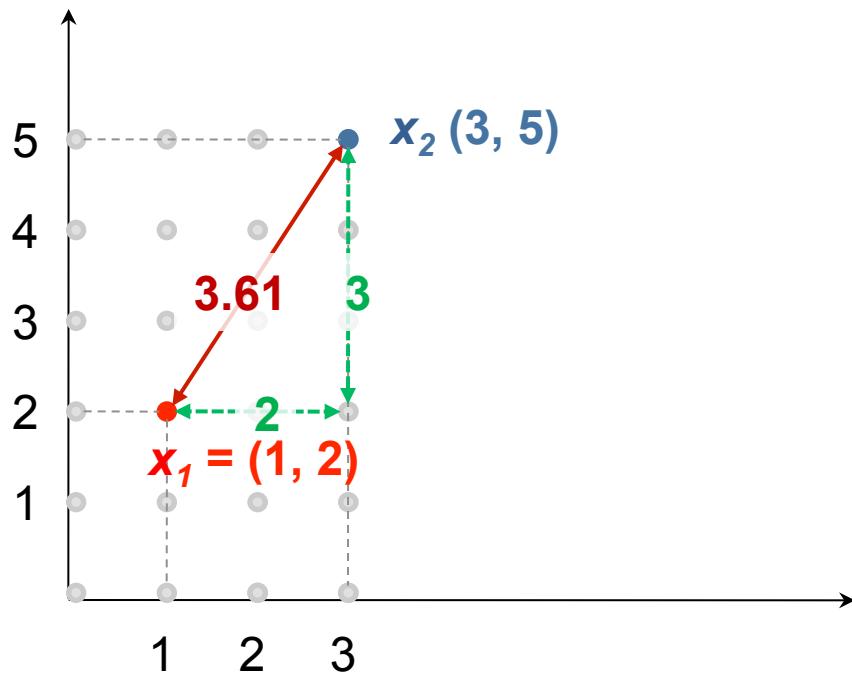
- Properties

- $d(i,j) \geq 0$
- $d(i,i) = 0$
- $d(i,j) = d(j,i)$
- $d(i,j) \leq d(i,k) + d(k,j)$

- Also, one can use weighted distance, parametric Pearson product moment correlation, or other dissimilarity measures

Euclidean distance vs Manhattan distance

- Distance of two point $x_1 = (1, 2)$ and $x_2 (3, 5)$

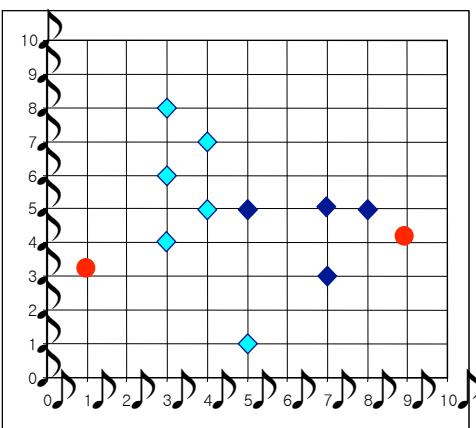


Euclidean distance:
 $= ((3-1)^2 + (5-2)^2)^{1/2}$
 $= (2^2 + 3^2)^{1/2}$
 $= (4 + 9)^{1/2}$
 $= (13)^{1/2}$
 $= 3.61$

Manhattan distance:
 $= (3-1) + (5-2)$
 $= 2 + 3$
 $= 5$

The *K*-Means Clustering Method

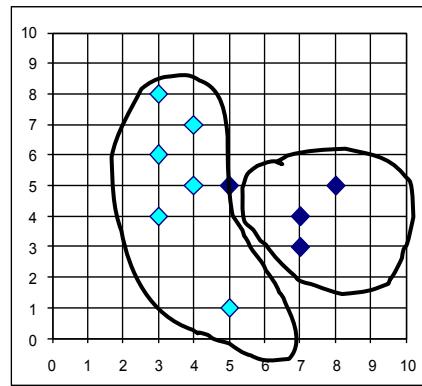
- Example



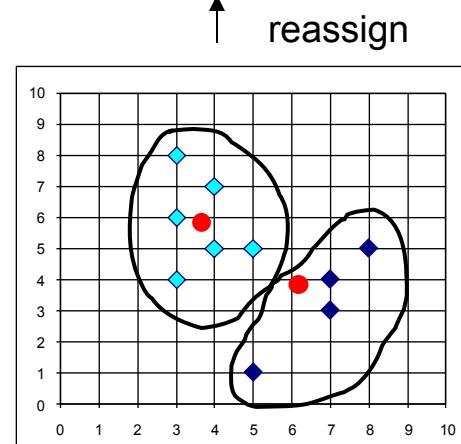
$K=2$

Arbitrarily choose K object as initial cluster center

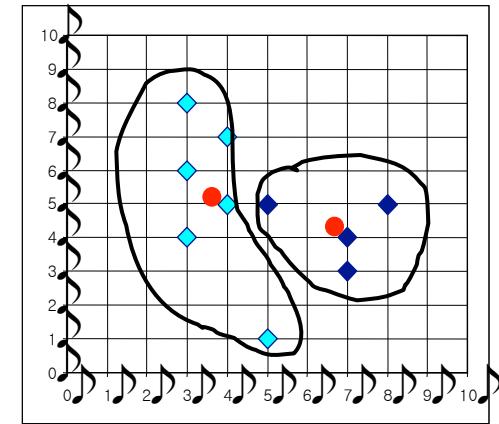
Assign each objects to most similar center



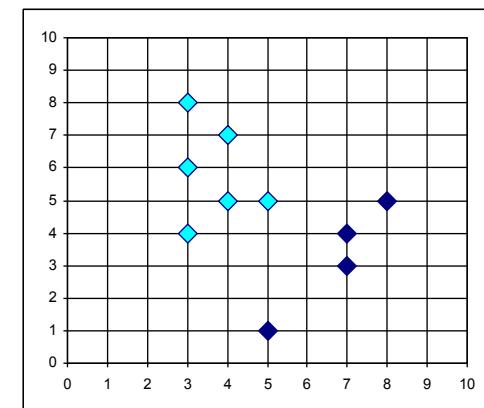
Update the cluster means



Update the cluster means

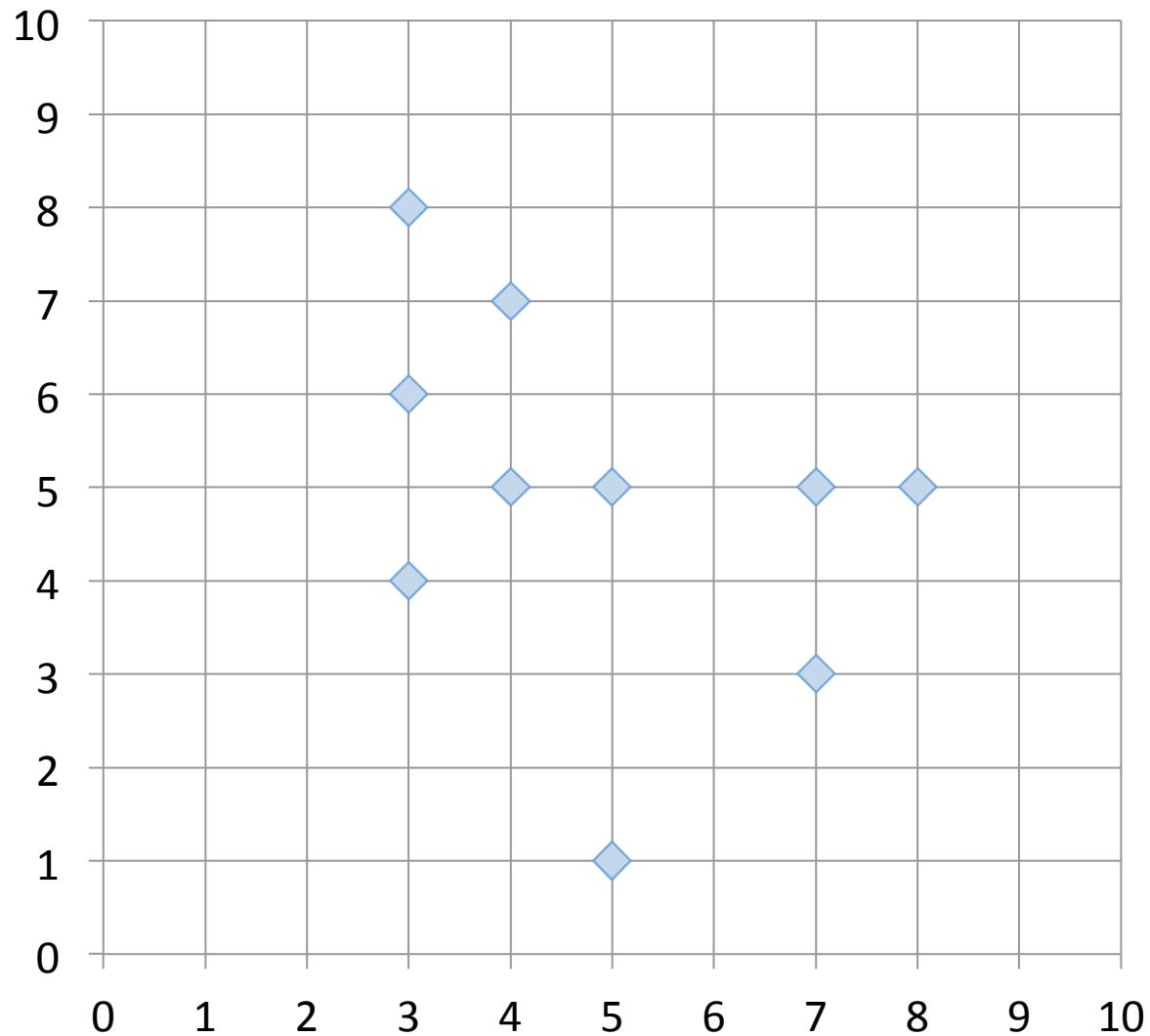


reassign



K-Means Clustering

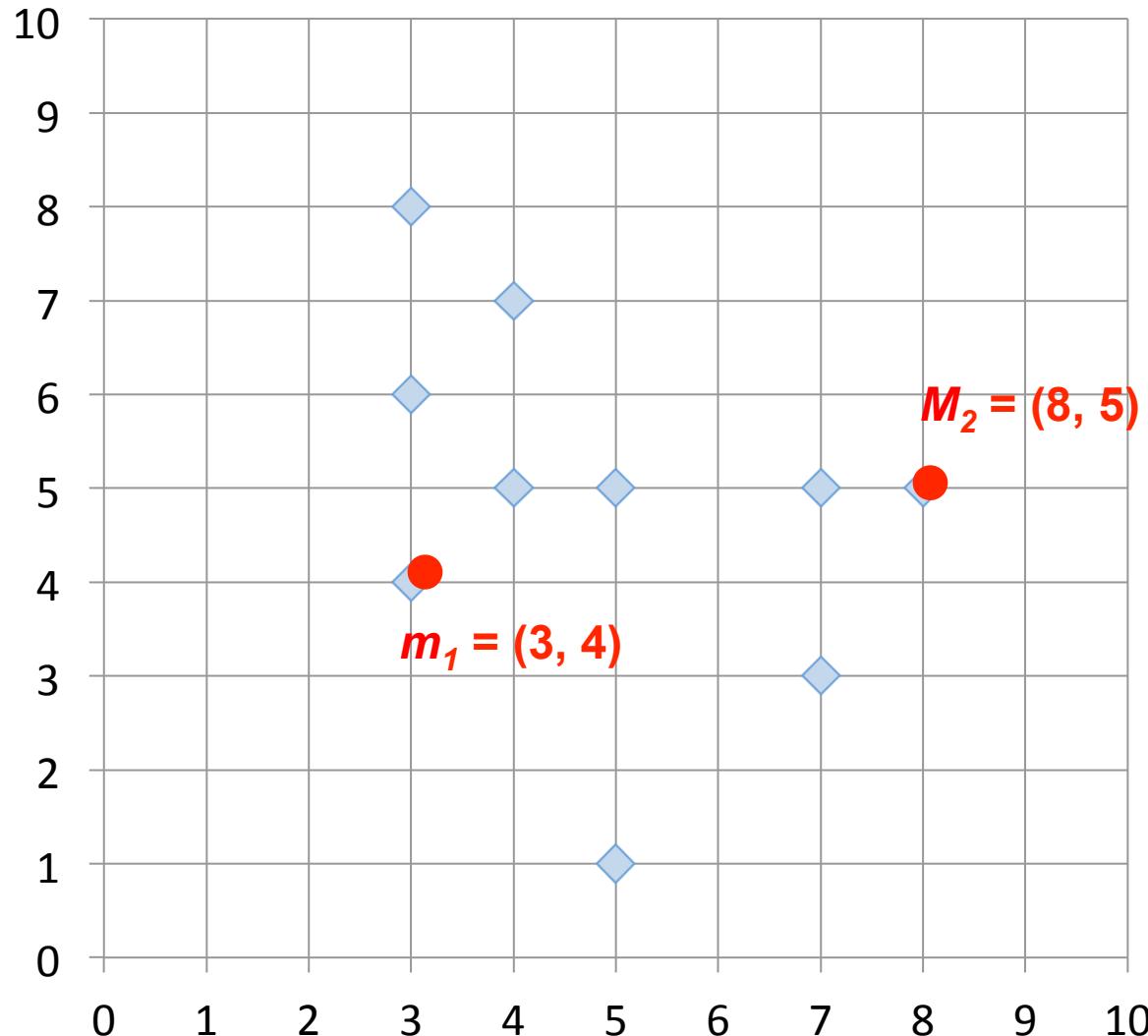
Step by Step



Point	P	P(x,y)
p01	a	(3, 4)
p02	b	(3, 6)
p03	c	(3, 8)
p04	d	(4, 5)
p05	e	(4, 7)
p06	f	(5, 1)
p07	g	(5, 5)
p08	h	(7, 3)
p09	i	(7, 5)
p10	j	(8, 5)

K-Means Clustering

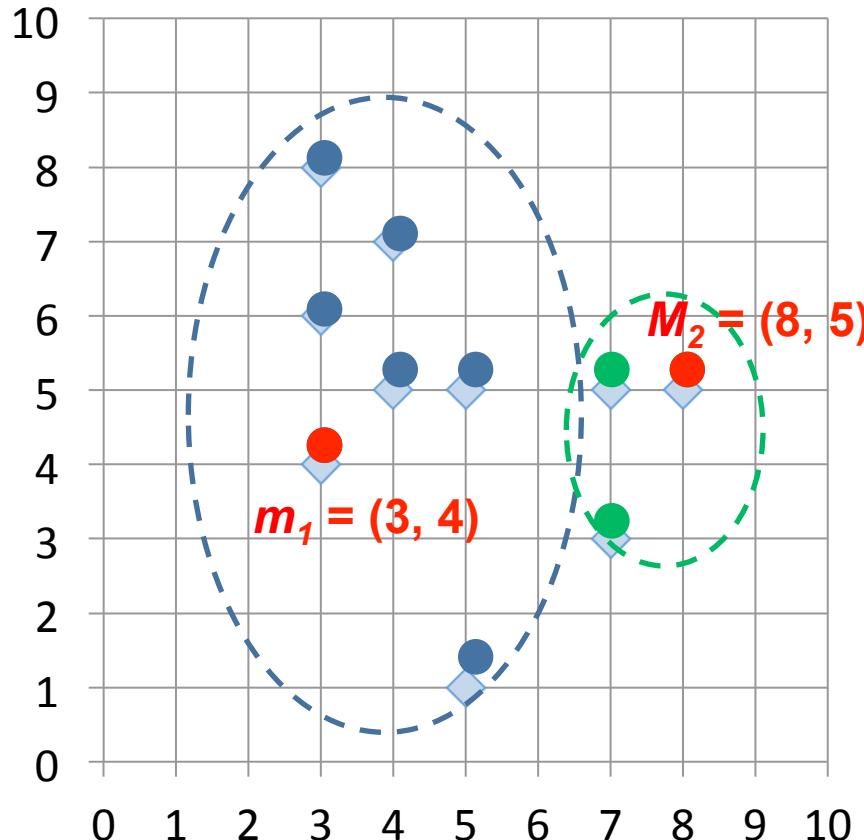
Step 1: K=2, Arbitrarily choose K object as initial cluster center



Initial m_1 (3, 4)
Initial m_2 (8, 5)

Step 2: Compute seed points as the centroids of the clusters of the current partition

Step 3: Assign each objects to most similar center



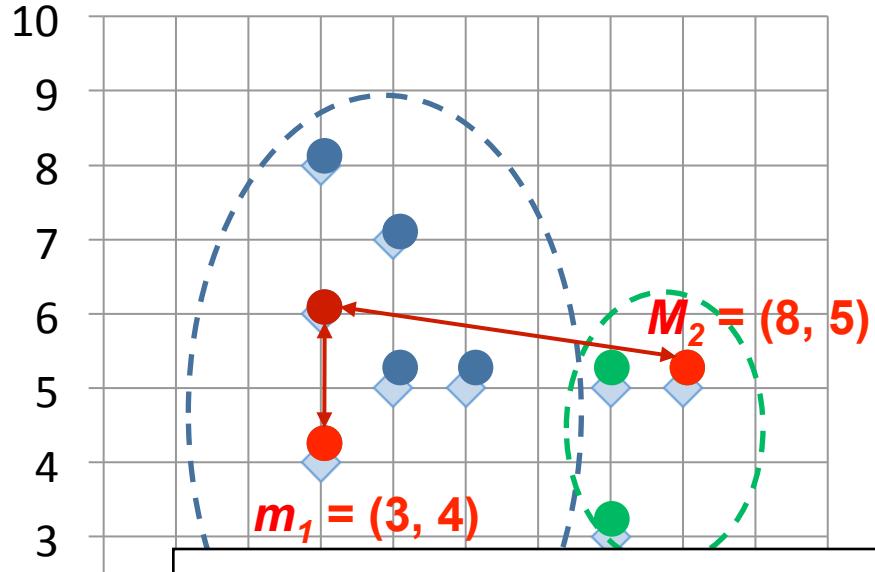
K-Means Clustering

Initial $m_1 (3, 4)$

Initial $m_2 (8, 5)$

Step 2: Compute seed points as the centroids of the clusters of the current partition

Step 3: Assign each objects to most similar center



$$\begin{aligned}
 & \text{Euclidean distance} \\
 & b(3,6) \leftrightarrow m1(3,4) \\
 & = ((3-3)^2 + (6-4)^2)^{1/2} \\
 & = (0^2 + (-2)^2)^{1/2} \\
 & = (0 + 4)^{1/2} \\
 & = (4)^{1/2} \\
 & = 2.00
 \end{aligned}$$

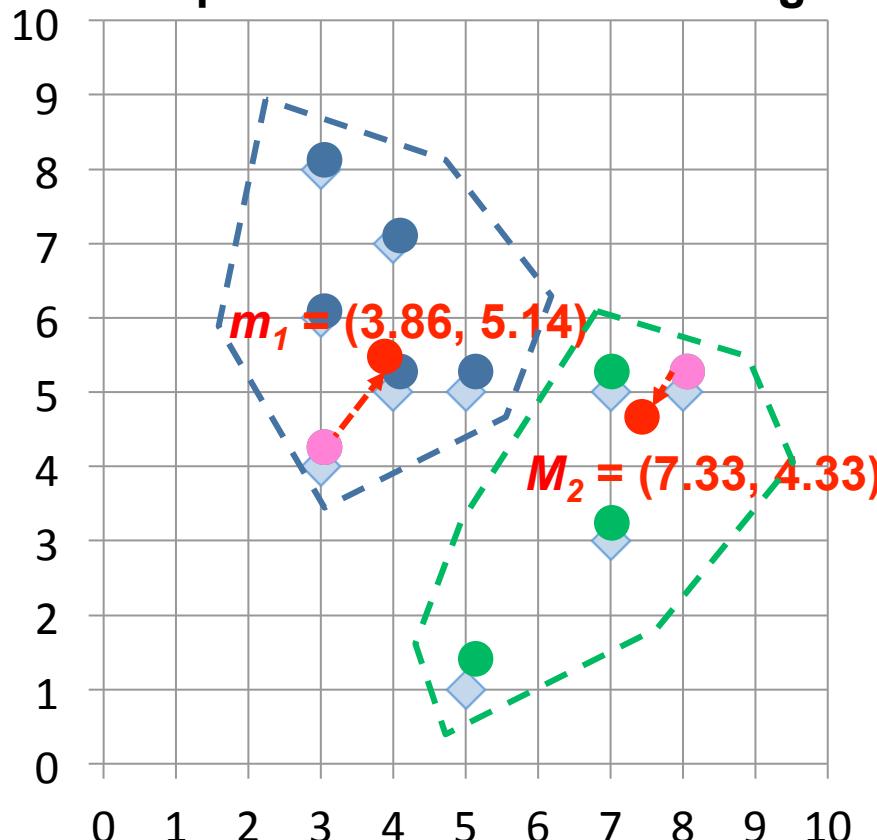
Point	P	P(x,y)	m1 distance	m2 distance	Cluster
p01	a	(3, 4)	0.00	5.10	Cluster1
p02	b	(3, 6)	2.00	5.10	Cluster1
p03	c	(3, 8)	4.00	5.83	Cluster1
p04	d	(4, 5)	1.41	4.00	Cluster1

$$\begin{aligned}
 & \text{Euclidean distance} \\
 & b(3,6) \leftrightarrow m2(8,5) \\
 & = ((8-3)^2 + (5-6)^2)^{1/2} \\
 & = (5^2 + (-1)^2)^{1/2} \\
 & = (25 + 1)^{1/2} \\
 & = (26)^{1/2} \\
 & = 5.10
 \end{aligned}$$

Initial m1 (3, 4)

Initial m2 (8, 5)

**Step 4: Update the cluster means,
Repeat Step 2, 3,
stop when no more new assignment**

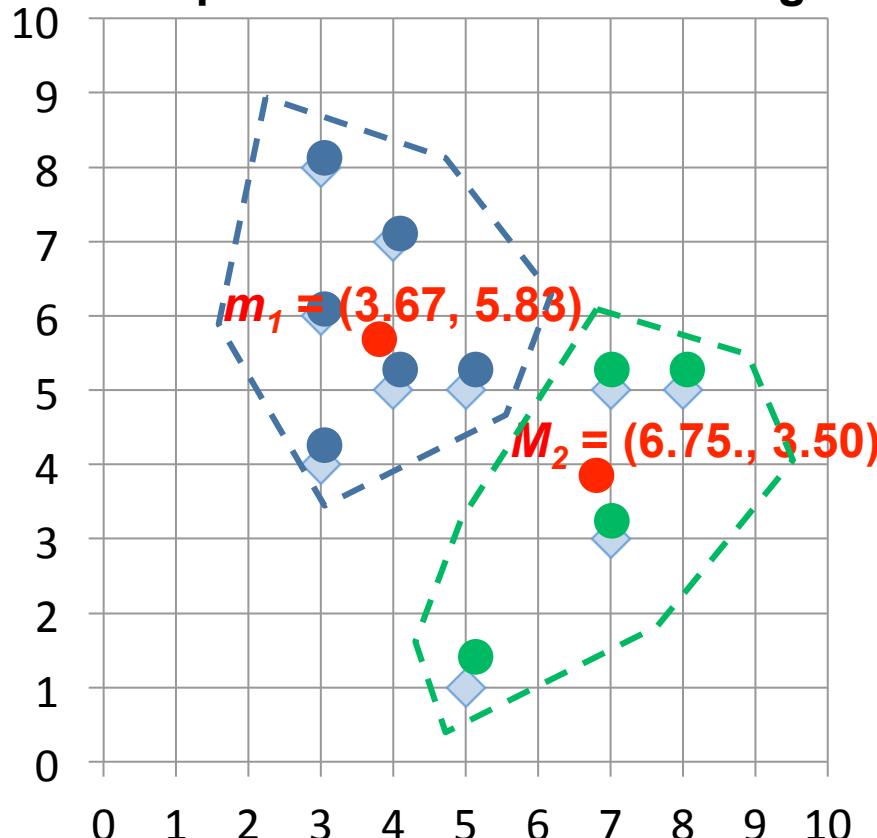


Point	P	P(x,y)	m1 distance	m2 distance	Cluster
p01	a	(3, 4)	1.43	4.34	Cluster1
p02	b	(3, 6)	1.22	4.64	Cluster1
p03	c	(3, 8)	2.99	5.68	Cluster1
p04	d	(4, 5)	0.20	3.40	Cluster1
p05	e	(4, 7)	1.87	4.27	Cluster1
p06	f	(5, 1)	4.29	4.06	Cluster2
p07	g	(5, 5)	1.15	2.42	Cluster1
p08	h	(7, 3)	3.80	1.37	Cluster2
p09	i	(7, 5)	3.14	0.75	Cluster2
p10	j	(8, 5)	4.14	0.95	Cluster2

$$\begin{aligned}m1 &= (3.86, 5.14) \\m2 &= (7.33, 4.33)\end{aligned}$$

K-Means Clustering

**Step 4: Update the cluster means,
Repeat Step 2, 3,
stop when no more new assignment**

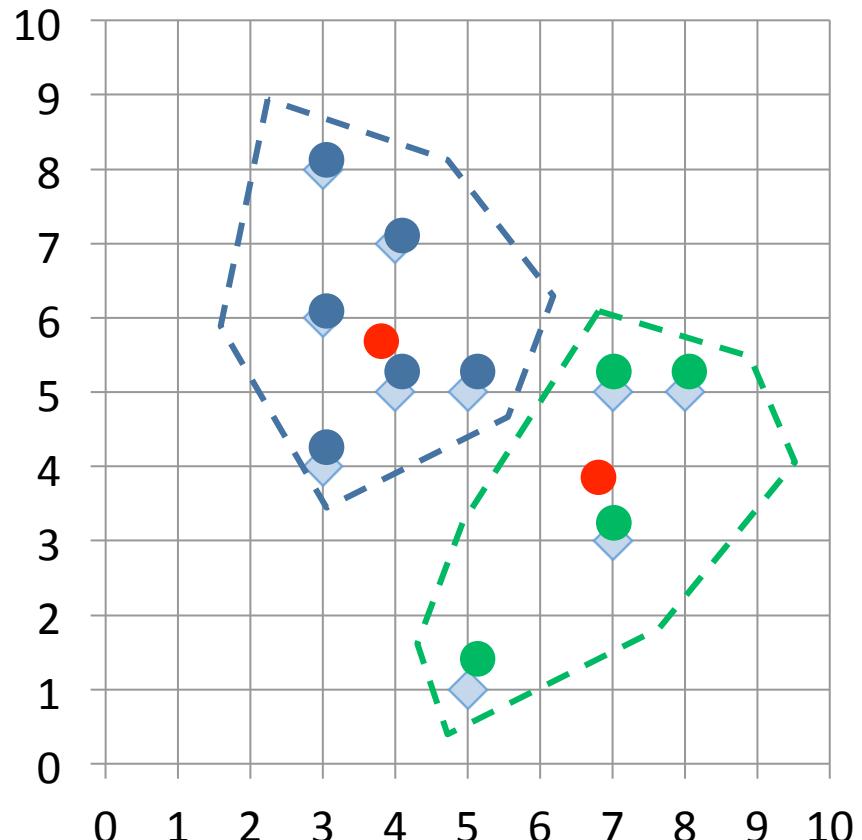


Point	P	P(x,y)	m1 distance	m2 distance	Cluster
p01	a	(3, 4)	1.95	3.78	Cluster1
p02	b	(3, 6)	0.69	4.51	Cluster1
p03	c	(3, 8)	2.27	5.86	Cluster1
p04	d	(4, 5)	0.89	3.13	Cluster1
p05	e	(4, 7)	1.22	4.45	Cluster1
p06	f	(5, 1)	5.01	3.05	Cluster2
p07	g	(5, 5)	1.57	2.30	Cluster1
p08	h	(7, 3)	4.37	0.56	Cluster2
p09	i	(7, 5)	3.43	1.52	Cluster2
p10	j	(8, 5)	4.41	1.95	Cluster2

$$\begin{aligned}m1 &= (3.67, 5.83) \\m2 &= (6.75, 3.50)\end{aligned}$$

K-Means Clustering

stop when no more new assignment



K-Means Clustering

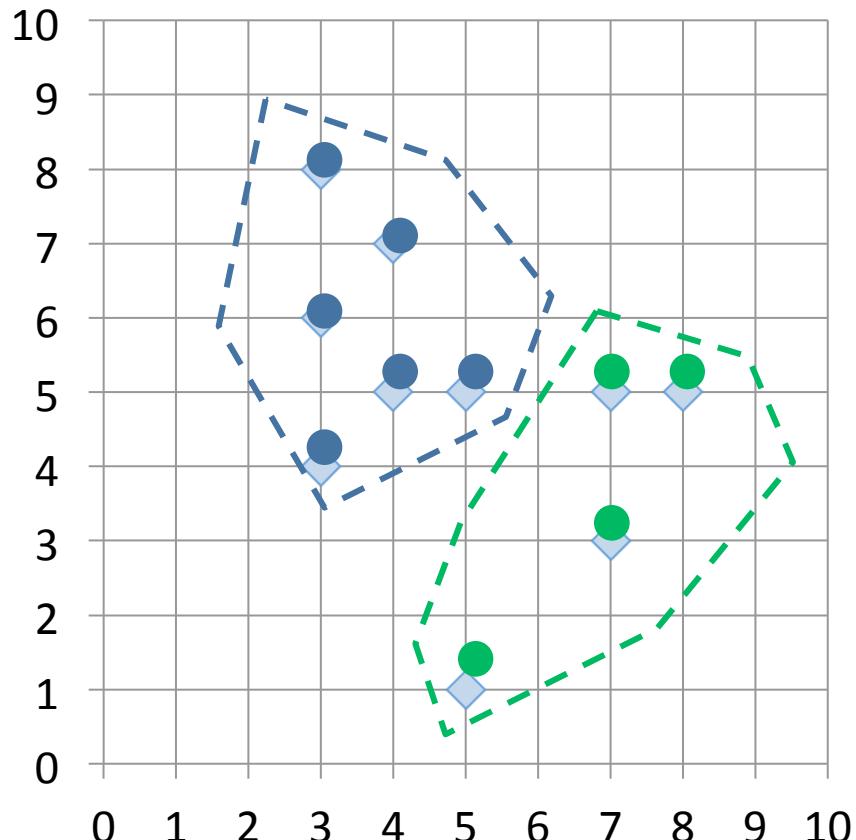
Point	P	P(x,y)	m1 distance	m2 distance	Cluster
p01	a	(3, 4)	1.95	3.78	Cluster1
p02	b	(3, 6)	0.69	4.51	Cluster1
p03	c	(3, 8)	2.27	5.86	Cluster1
p04	d	(4, 5)	0.89	3.13	Cluster1
p05	e	(4, 7)	1.22	4.45	Cluster1
p06	f	(5, 1)	5.01	3.05	Cluster2
p07	g	(5, 5)	1.57	2.30	Cluster1
p08	h	(7, 3)	4.37	0.56	Cluster2
p09	i	(7, 5)	3.43	1.52	Cluster2
p10	j	(8, 5)	4.41	1.95	Cluster2

$$m1 \ (3.67, 5.83)$$

$$m2 \ (6.75, 3.50)$$

K-Means Clustering ($K=2$, two clusters)

stop when no more new assignment



Point	P	P(x,y)	m1 distance	m2 distance	Cluster
p01	a	(3, 4)	1.95	3.78	Cluster1
p02	b	(3, 6)	0.69	4.51	Cluster1
p03	c	(3, 8)	2.27	5.86	Cluster1
p04	d	(4, 5)	0.89	3.13	Cluster1
p05	e	(4, 7)	1.22	4.45	Cluster1
p06	f	(5, 1)	5.01	3.05	Cluster2
p07	g	(5, 5)	1.57	2.30	Cluster1
p08	h	(7, 3)	4.37	0.56	Cluster2
p09	i	(7, 5)	3.43	1.52	Cluster2
p10	j	(8, 5)	4.41	1.95	Cluster2

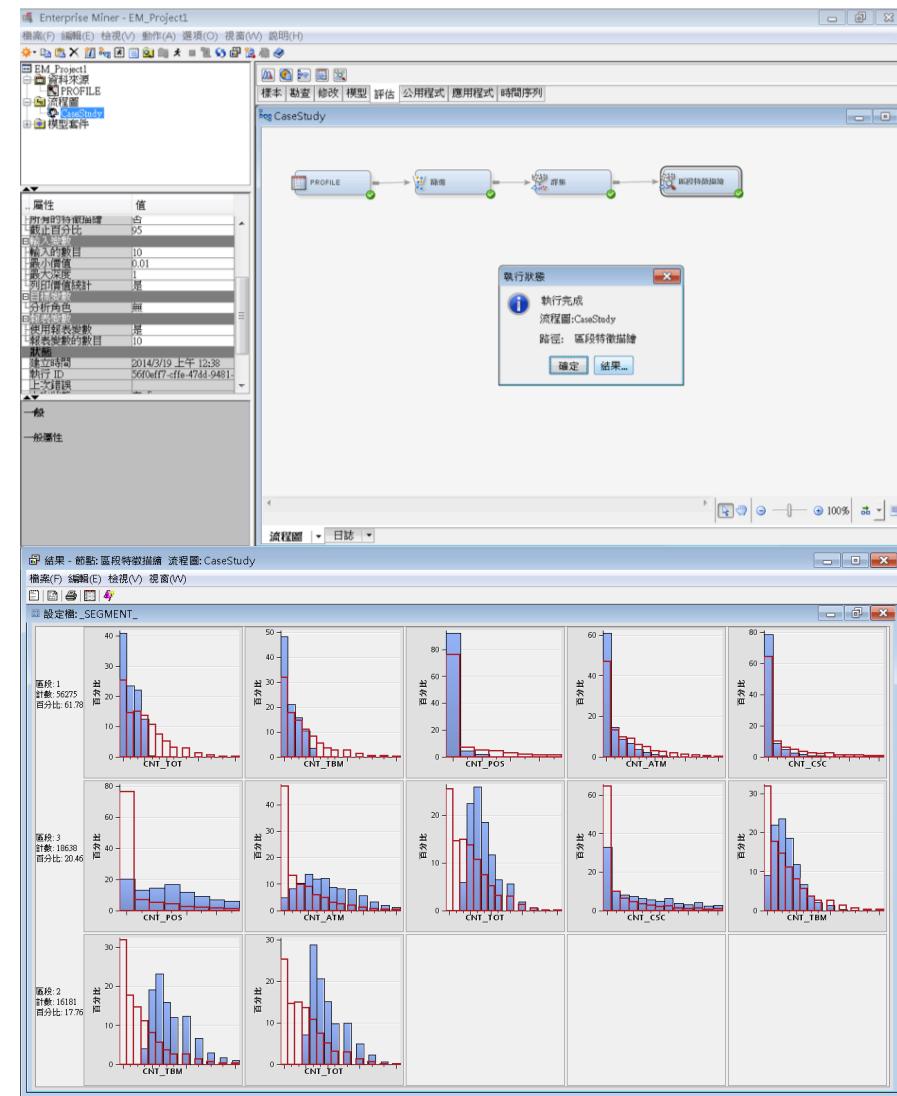
$$m1 \ (3.67, 5.83)$$

$$m2 \ (6.75, 3.50)$$

K-Means Clustering

個案分析與實作一 (SAS EM 分群分析)： Case Study 1 (Cluster Analysis – K-Means using SAS EM) Banking Segmentation

	ID	CNT_TBM	CNT_ATM	CNT_POS	CNT_CSC	CNT_TOT
1	000041360	34.0	3.0	3.0	9.0	49.0
2	000052094	44.0	17.0	5.0	18.0	84.0
3	000057340	122.0	26.0	32.0	36.0	216.0
4	000076885	42.0	3.0	6.0	1.0	52.0
5	000089150	20.0	15.0	2.0	2.0	39.0
6	000094512	83.0	20.0	49.0	3.0	155.0
7	000096396	33.0	9.0	49.0	17.0	108.0
8	000098062	22.0	5.0	9.0	1.0	37.0
9	000122779	21.0	27.0	2.0	1.0	51.0
10	000125838	127.0	3.0	2.0	3.0	135.0
11	000137317	52.0	20.0	2.0	11.0	85.0
12	000147896	101.0	40.0	8.0	1.0	150.0
13	000150206	156.0	52.0	80.0	4.0	292.0
14	000160501	31.0	32.0	42.0	12.0	117.0
15	000176820	107.0	3.0	2.0	1.0	113.0
16	000187784	54.0	14.0	7.0	24.0	99.0
17	000189200	92.0	3.0	2.0	2.0	99.0
18	000198327	44.0	27.0	17.0	3.0	91.0
19	000200237	146.0	3.0	2.0	1.0	152.0
20	000207980	39.0	5.0	2.0	4.0	50.0
21	000227355	180.0	3.0	2.0	10.0	195.0
22	000228317	46.0	14.0	2.0	1.0	63.0
23	000228906	25.0	3.0	2.0	1.0	31.0
24	000229984	15.0	3.0	2.0	1.0	21.0
25	000241022	13.0	3.0	2.0	1.0	19.0
26	000253260	65.0	3.0	2.0	1.0	71.0
27	000257910	35.0	3.0	2.0	1.0	41.0
28	000261717	11.0	5.0	2.0	1.0	19.0
29	000266600	15.0	3.0	2.0	1.0	21.0



行銷客戶分群

案例情境

- ABC銀行的行銷部門想要針對該銀行客戶的使用行為，進行分群分析，以了解現行客戶對本行的往來方式，並進一步提供適宜的行銷接觸模式。
- 該銀行從有效戶(近三個月有交易者)，取出10萬筆樣本資料。
依下列四種交易管道計算交易次數：
 - 傳統臨櫃交易(TBM)
 - 自動櫃員機交易(ATM)
 - 銀行專員服務(POS)
 - 電話客服(CSC)

資料欄位說明

- 資料集名稱：profile.sas7bdat

Name	Model Role	Measurement Level	Description
ID	ID	Nominal	Customer ID
CNT_TBM	Input	Interval	Traditional bank method transaction count
CNT_ATM	Input	Interval	ATM transaction count
CNT_POS	Input	Interval	Point-of-sale transaction count
CNT_CSC	Input	Interval	Customer service transaction count
CNT_TOT	Input	Interval	Total transaction count

EMPL_PROFILE						
	ID	CNT_TBM	CNT_ATM	CNT_POS	CNT_CSC	CNT_TOT
1	0000413634.0	3.0	3.0	9.0	40.0	
2	0000204444.0	17.0	5.0	18.0	94.0	
3	0000122722.0	20.0	2.0	20.0	20.0	
4	0000795142.0	1.0	6.0	1.0	22.0	
5	0000012020.0	15.0	2.0	2.0	30.0	
6	0000412333.0	20.0	49.0	3.0	355.0	
7	0000012022.0	10.0	40.0	1.0	77.0	
8	0000012222.0	5.0	0.0	1.0	7.0	
9	0000122720.0	27.0	2.0	1.0	51.0	
10	00001238127.0	3.0	2.0	3.0	35.0	
11	00001238128.0	20.0	5.0	1.0	86.0	
12	00001238131.0	40.0	8.0	1.0	139.0	
13	000012000196.0	52.0	80.0	4.0	202.0	
14	00001200121.0	32.0	42.0	12.0	117.0	
15	00001200127.0	31.0	2.0	1.0	33.0	
16	00001200128.0	14.0	7.0	2.0	24.0	
17	00001200092.0	10.0	2.0	2.0	30.0	
18	00001237244.0	27.0	17.0	3.0	94.0	
19	00001237246.0	3.0	2.0	1.0	35.0	
20	00001238129.0	7.0	2.0	1.0	20.0	
21	00002175590.0	3.0	2.0	1.0	35.0	
22	00002317145.0	14.0	2.0	1.0	63.0	
23	00002399625.0	3.0	2.0	1.0	31.0	
24	00002399626.0	13.0	3.0	1.0	23.0	
25	00002322110.0	10.0	2.0	1.0	33.0	
26	0000206653.0	3.0	2.0	1.0	71.0	
27	00002791050.0	3.0	2.0	1.0	41.0	
28	00008777110.0	5.0	2.0	1.0	38.0	
29	00008690043.0	10.0	2.0	1.0	31.0	

行銷客戶分群實機演練

- 分析目的

依據各往來交易管道TBM、ATM、POS、CSC進行客戶分群分析。

演練重點:

- 極端值資料處理
- 分群變數選擇
- 衍生變數產出
- 分群參數調整與分群結果解釋

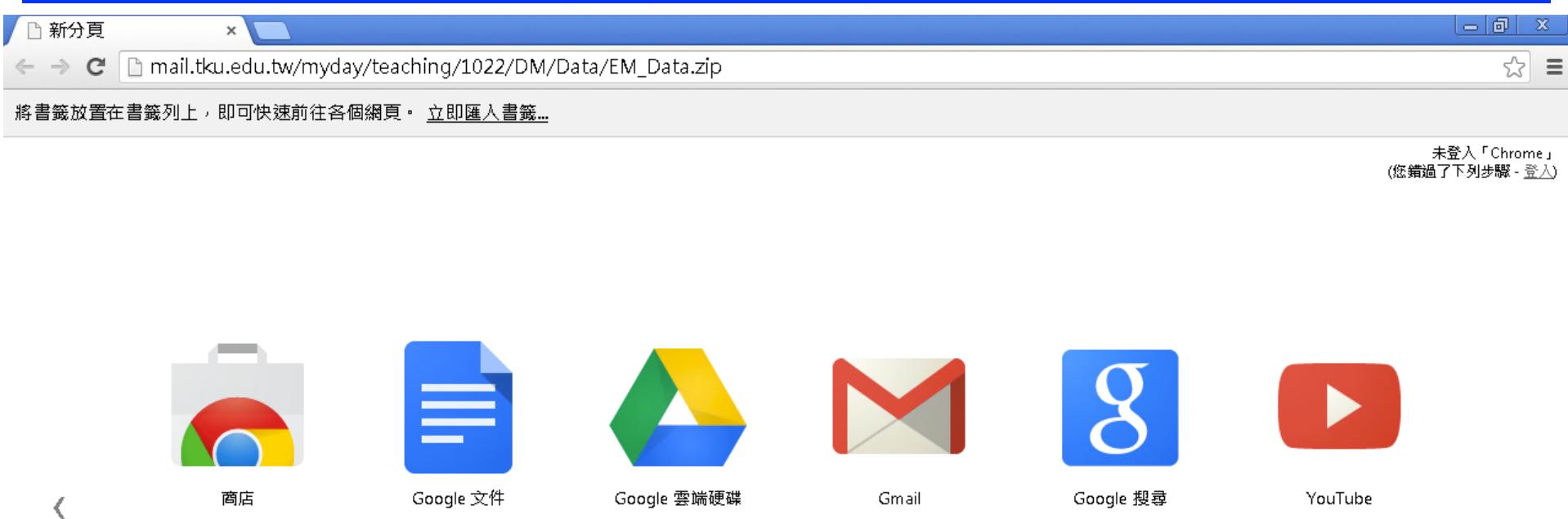
SAS Enterprise Miner (SAS EM)

Case Study

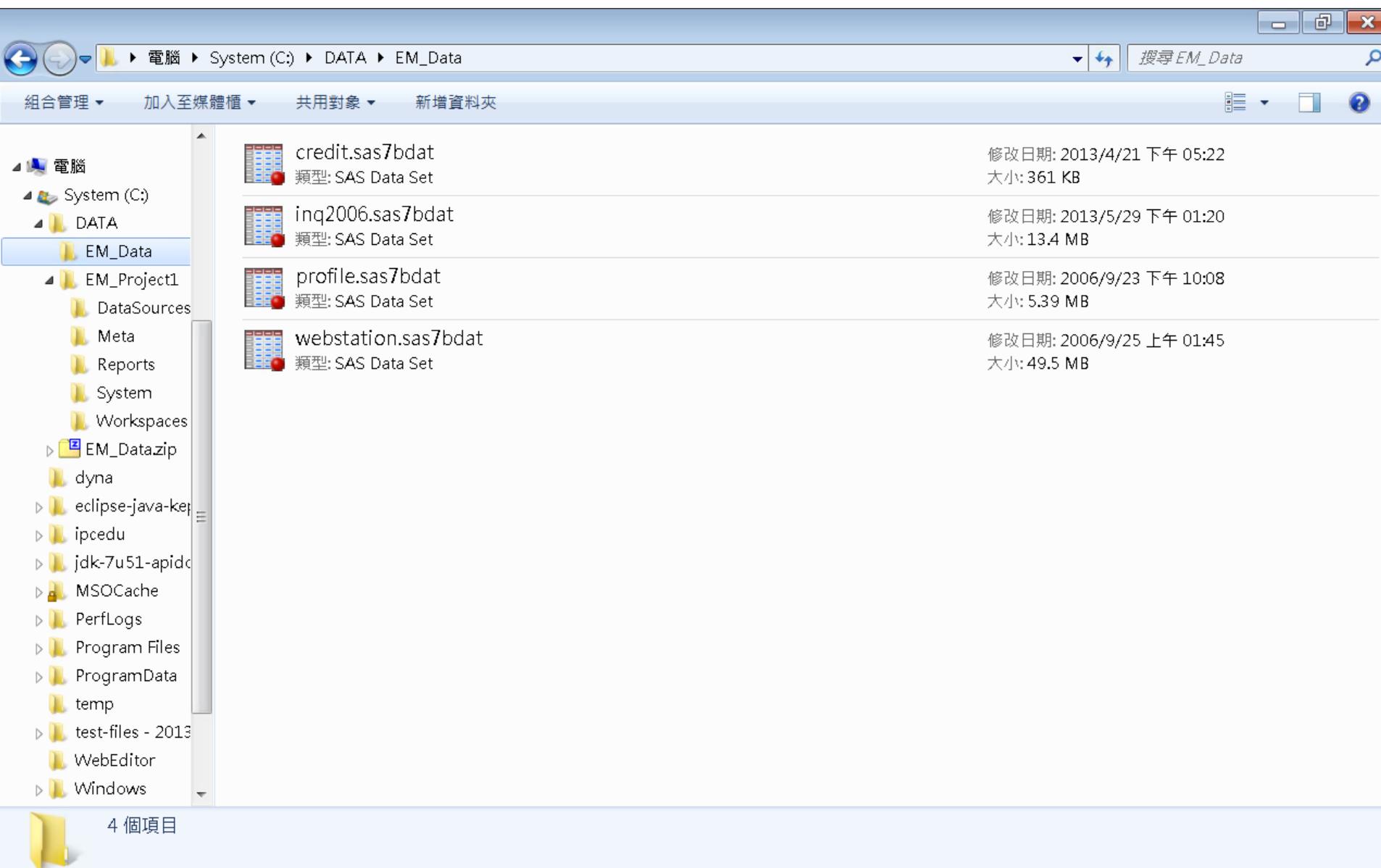
- SAS EM 資料匯入4步驟
 - Step 1. 新增專案 (New Project)
 - Step 2. 新增資料館 (New / Library)
 - Step 3. 建立資料來源 (Create Data Source)
 - Step 4. 建立流程圖 (Create Diagram)
- SAS EM SEMMA 建模流程

Download EM_Data.zip (SAS EM Datasets)

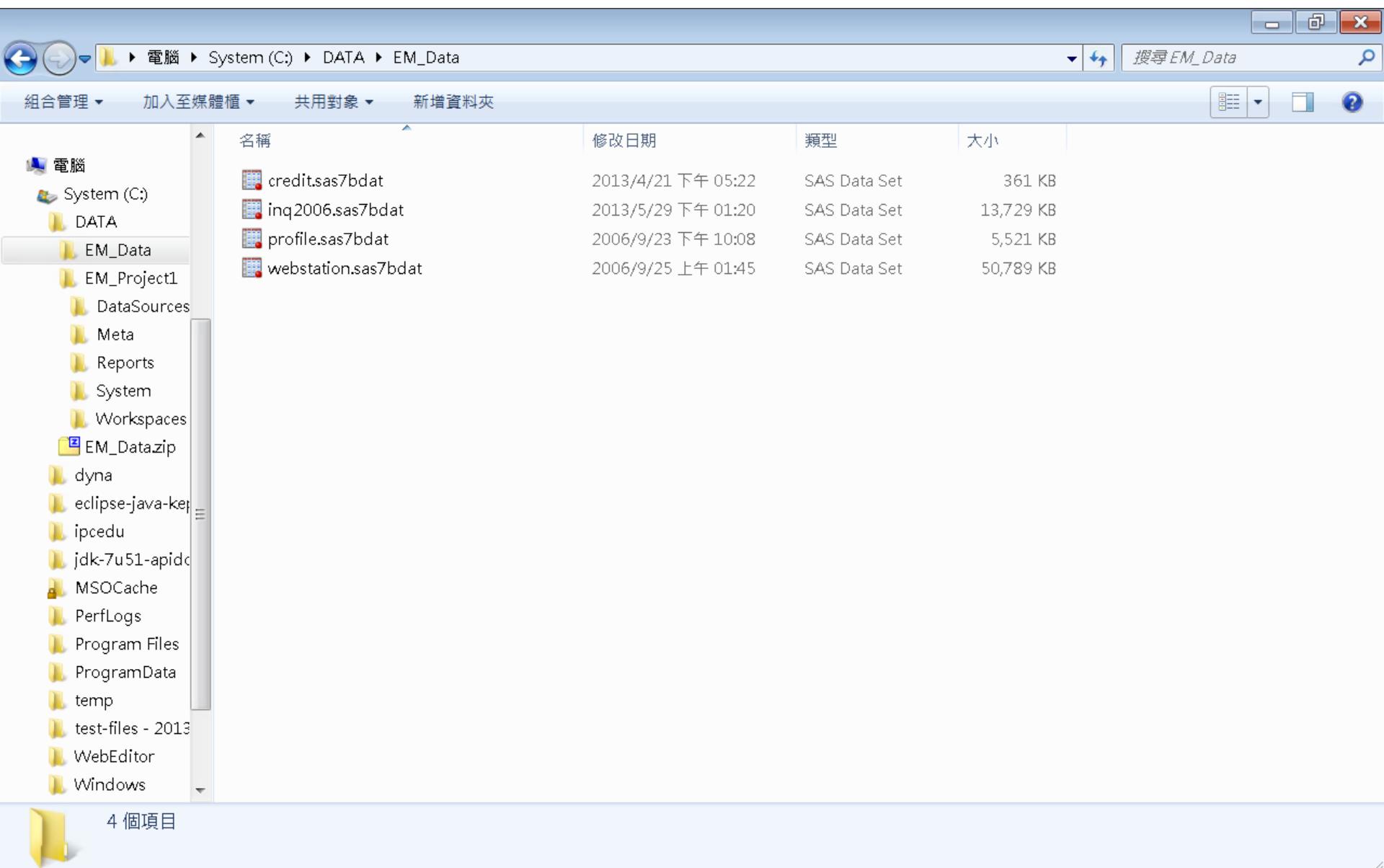
http://mail.tku.edu.tw/myday/teaching/1022/DM/Data/EM_Data.zip



Upzip EM_Data.zip to C:\DATA\EM_Data



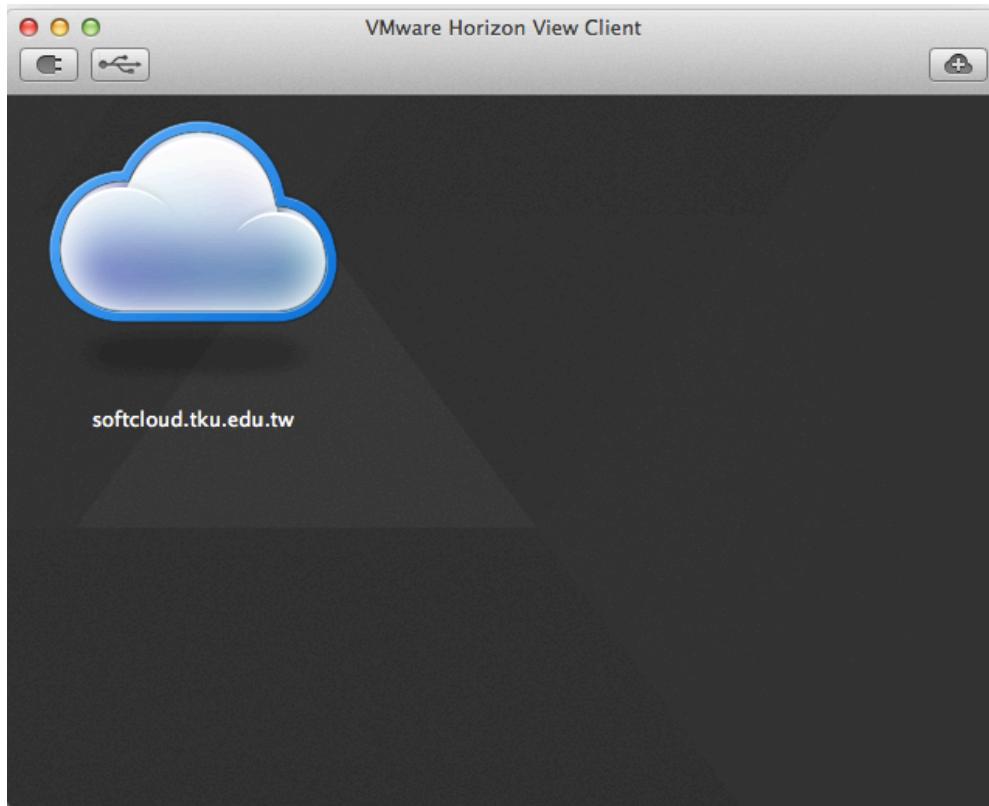
Upzip EM_Data.zip to C:\DATA\EM_Data



VMware Horizon View Client

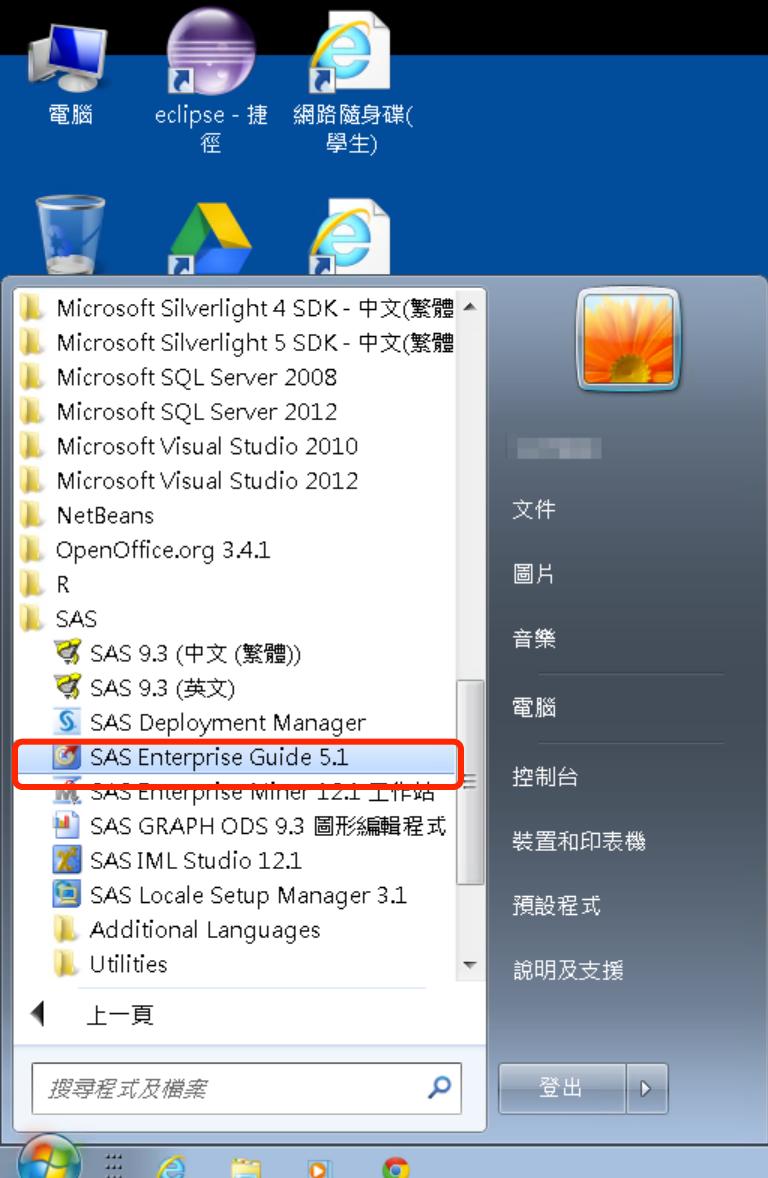
softcloud.tku.edu.tw

SAS Enterprise Miner



SAS Locale Setup Manager
→ English UI

SAS Enterprise Guide 5.1 (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 Enterprise Miner 12.1 (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 Locale Setup Manager 3.1

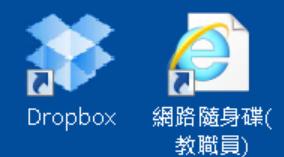


淡江軟體雲使用注意事項

- 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。

淡江軟體雲使用注意事項





淡江軟體雲使用注意事項

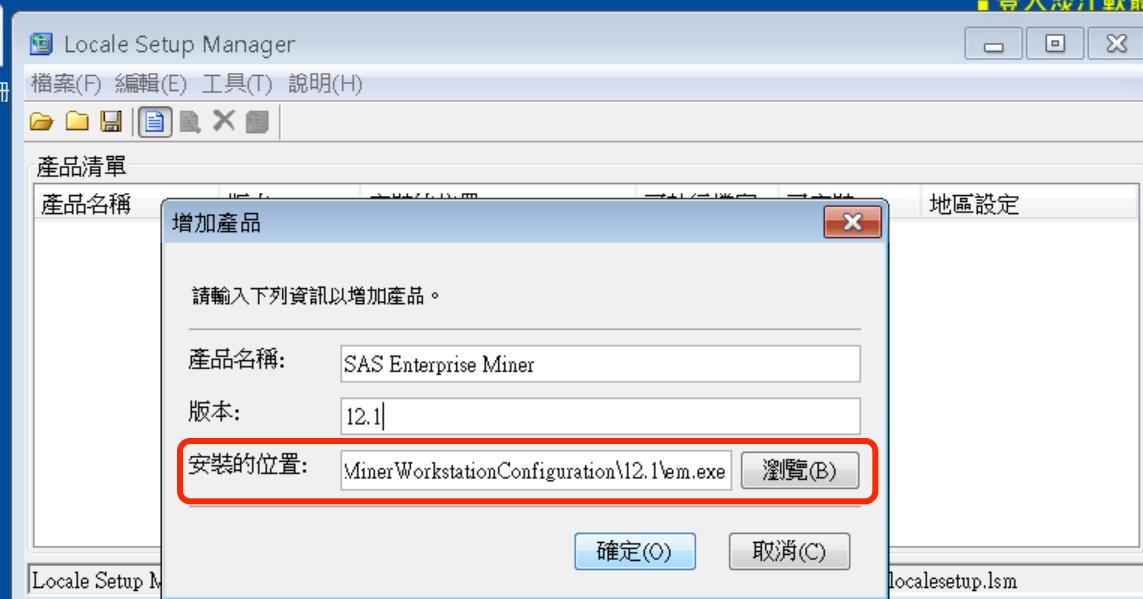
■ 登入淡江軟體雲時，會自動產生一部新
登出雲端主機前，請先將
攜帶的隨身碟、淡江大學
Google雲端硬碟...等，
雲端主機的所有資料將會被
您的個人主機畫面。

帳號被冒用，每次使用完

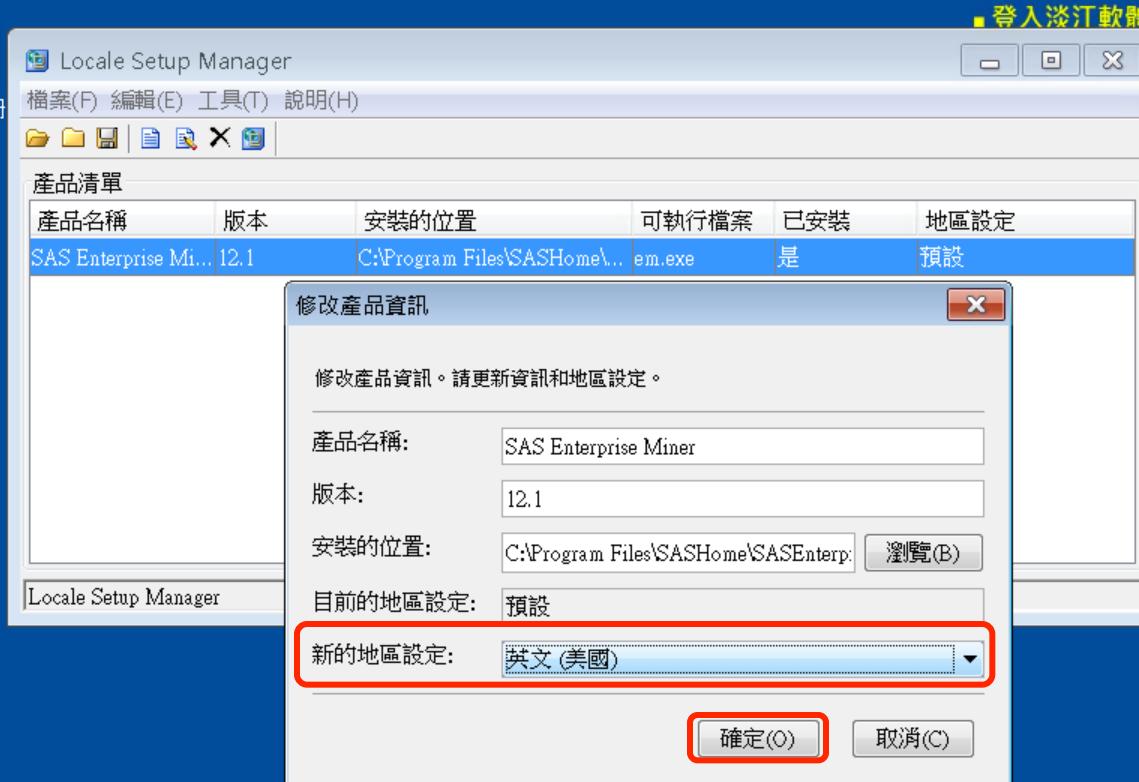
作說明，請點選雲端主機
『手冊』圖示。

請至B212辦公室，或來電
129為您服務。

臺權，嚴禁安裝不法軟
負相關法律責任。



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
登出雲端主機前，請先將
攜帶的隨身碟、淡江大學
Google雲端硬碟...等。
雲端主機的所有資料將會被
您的個人主機畫面。

帳號被冒用，每次使用完

作說明，請點選雲端主機
『手冊』圖示。

請至B212辦公室，或來電
129為您服務。

盜權，嚴禁安裝不法軟
負相關法律責任。

淡江軟體雲使用注意事項

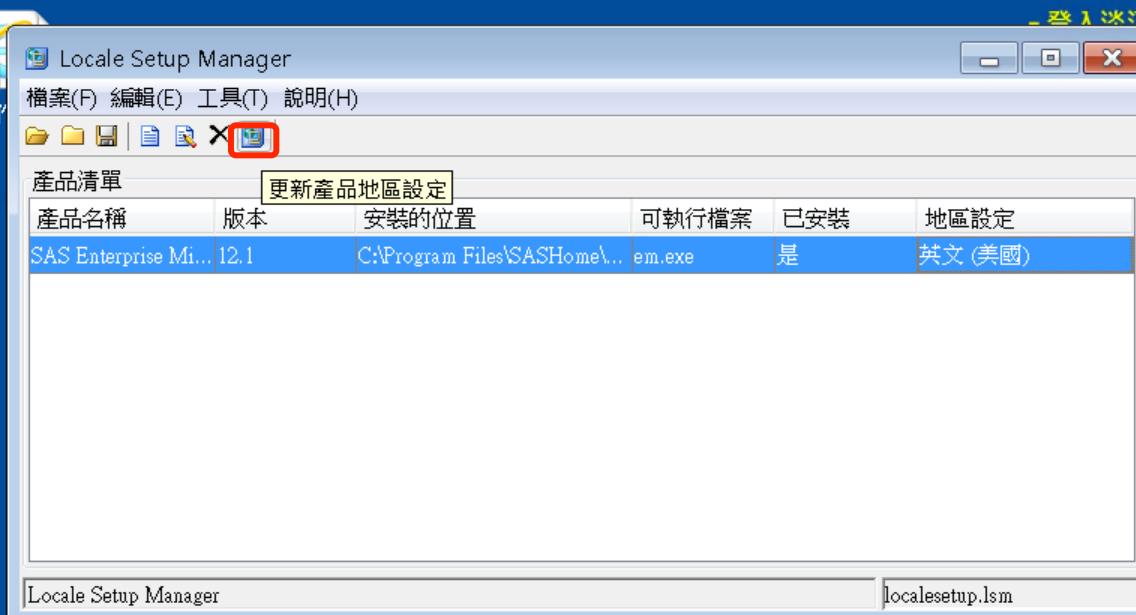
登入淡江軟體雲時，會自動產生一部新機。登出雲端主機前，請先將屬於您攜帶的隨身碟、淡江大學碟、Google雲端硬碟...等，之後雲端主機的所有資料將會被回到您的個人主機畫面。

主機帳號被冒用，每次使用完

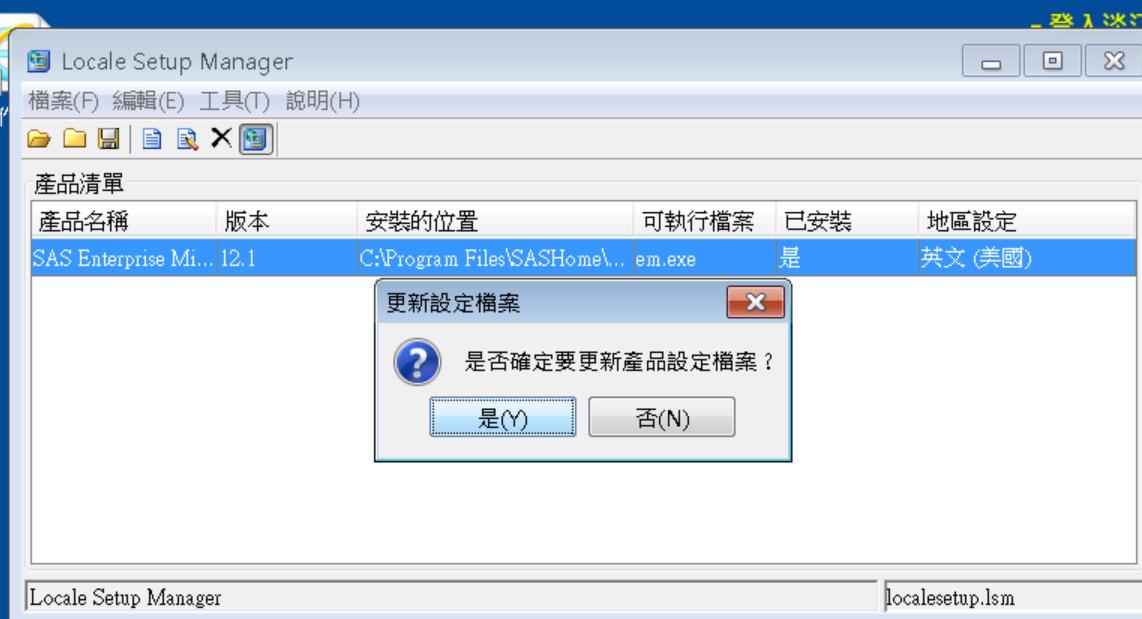
請操作說明，請點選雲端主機『操作手冊』圖示。

用問題請至B212辦公室，或來電56-2129為您服務。

慧財產權，嚴禁安裝不法軟
者自負相關法律責任。



淡江軟體雲使用注意事項



登入淡江軟體雲時，會自動產生一部新機。登出雲端主機前，請先將於您攜帶的隨身碟、淡江大學碟、Google雲端硬碟...等，之後雲端主機的所有資料將會被回到您的個人主機畫面。

主機帳號被冒用，每次使用完

請操作說明，請點選雲端主機「操作手冊」圖示。

用問題請至B212辦公室，或來電56-2129為您服務。

慧財產權，嚴禁安裝不法軟
者自負相關法律責任。

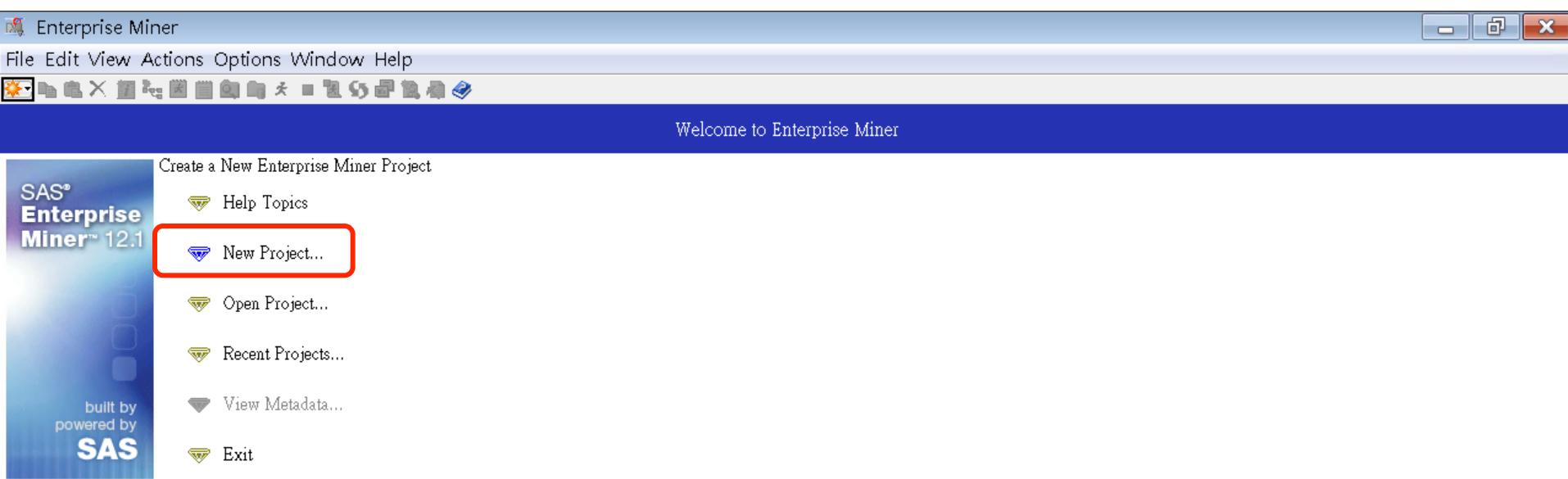
SAS Enterprise Miner 12.1 (SAS EM)



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 Enterprise Miner (SAS EM)



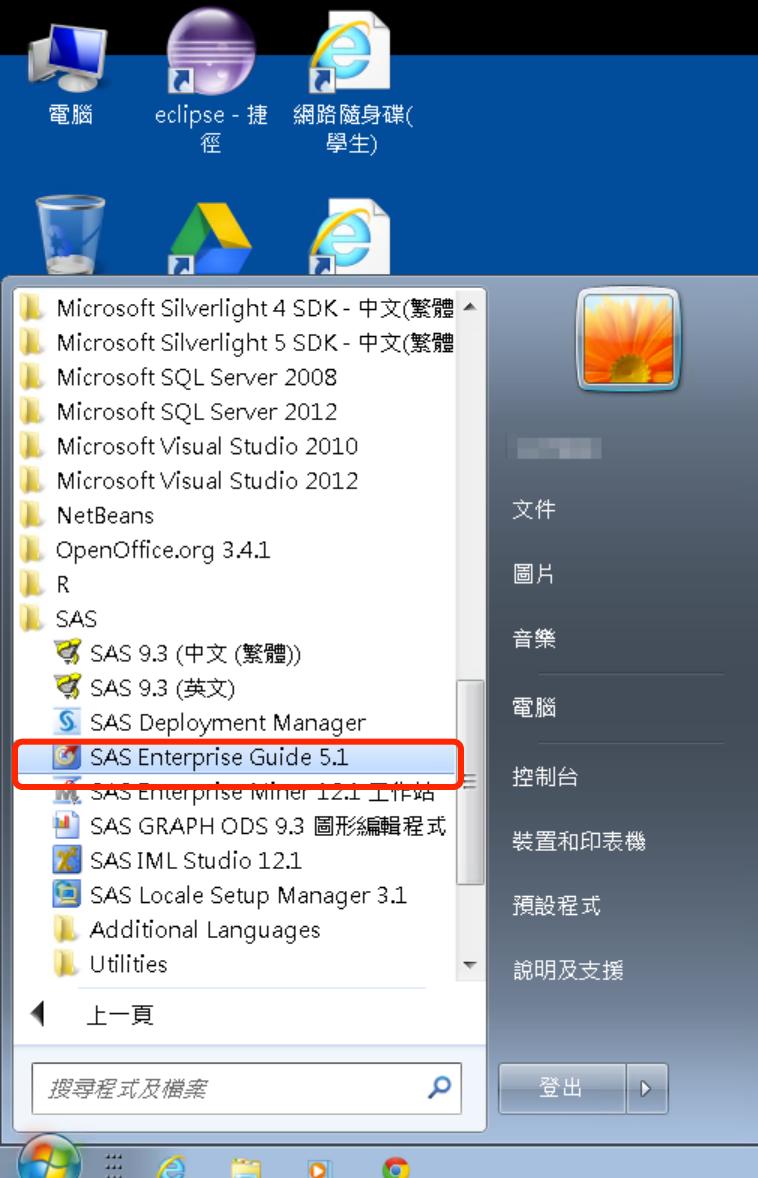
SAS Enterprise Miner (SAS EM)
English UI

SAS Enterprise Guide 5.1

(SAS EG)

Open SAS .sas7bdat File
Export to Excel .xlsx File
Import Excel .xlsx File
Export to SAS .sas7bdat File

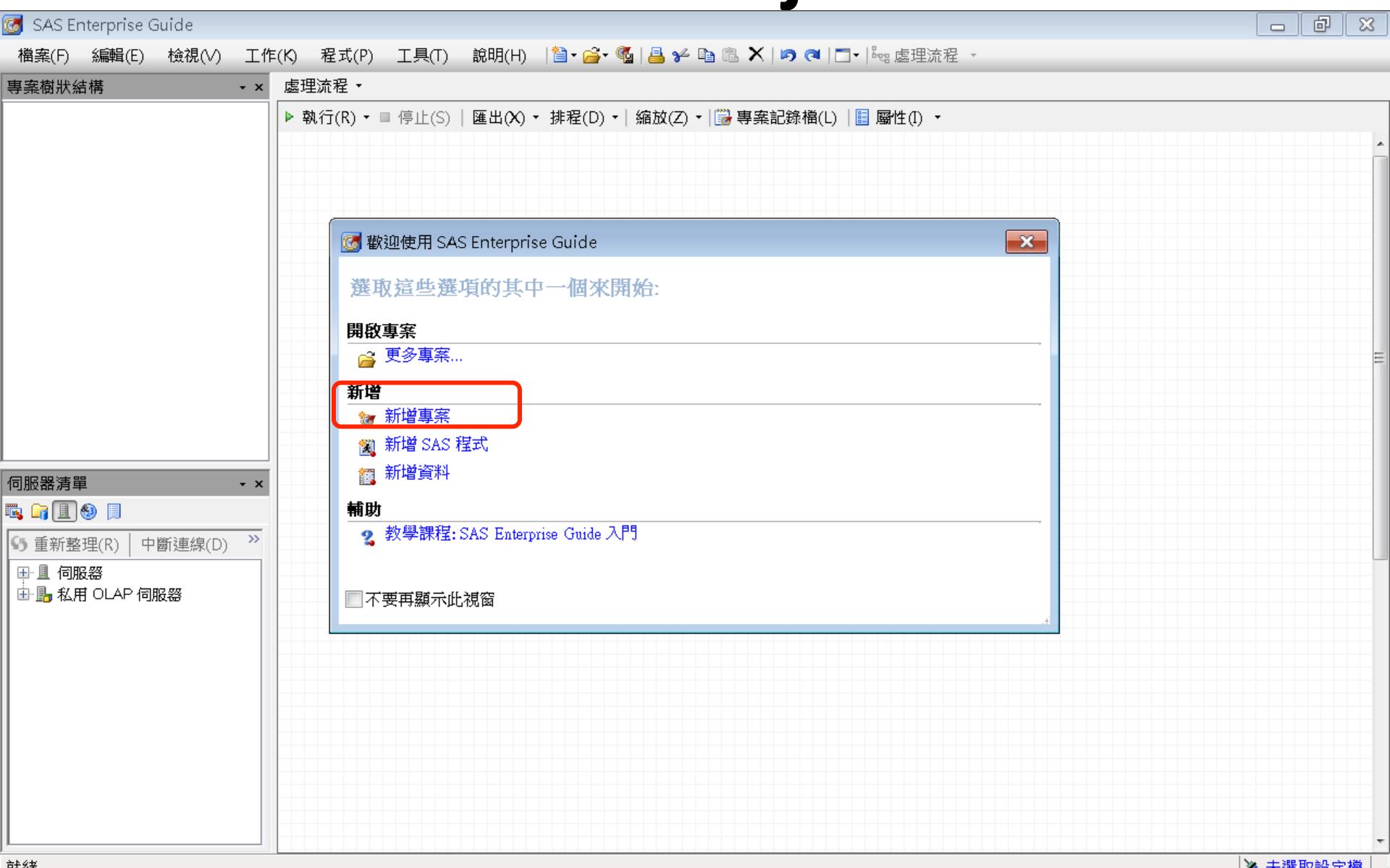
SAS Enterprise Guide 5.1 (SAS EG)



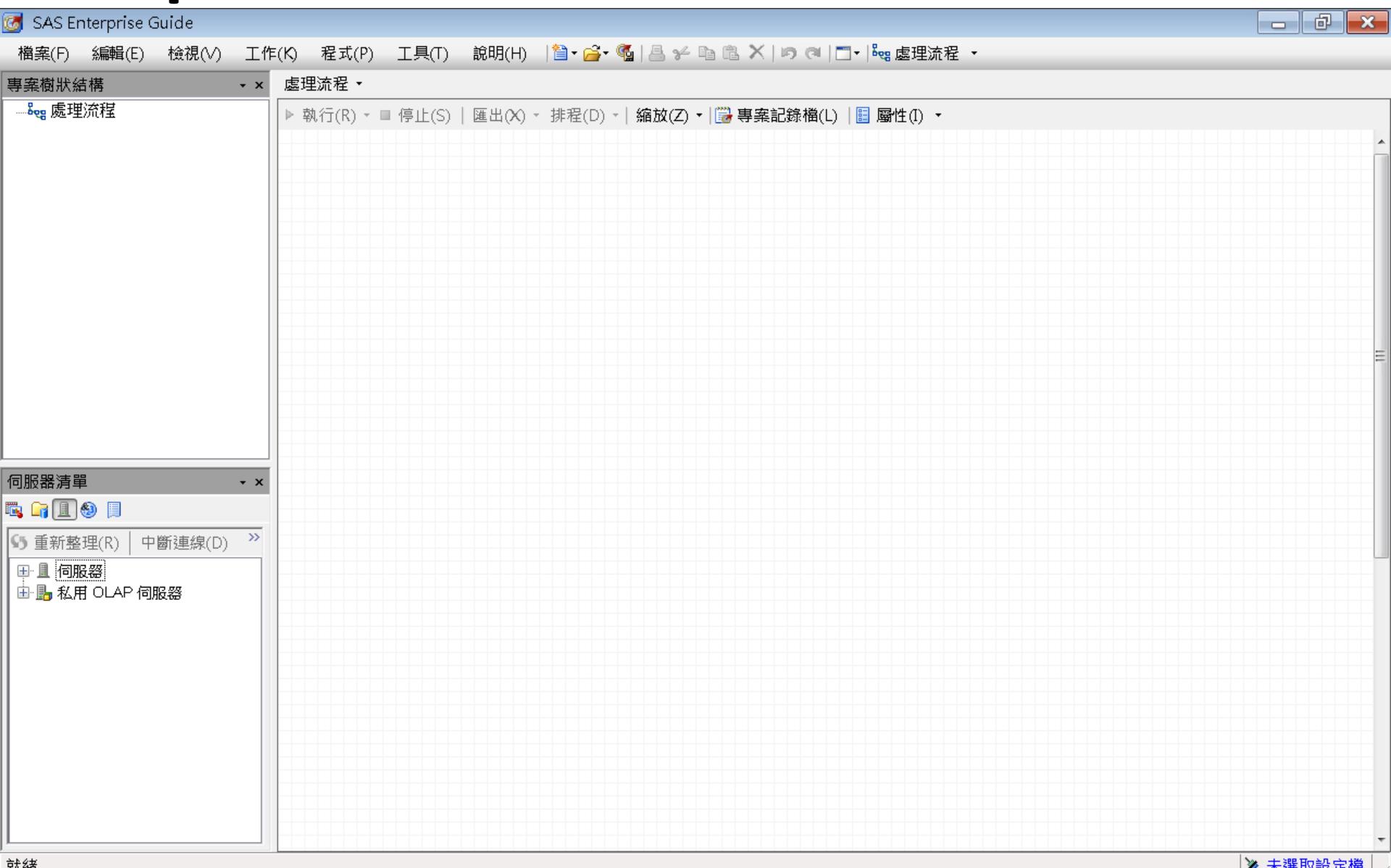
淡江軟體雲使用注意事項

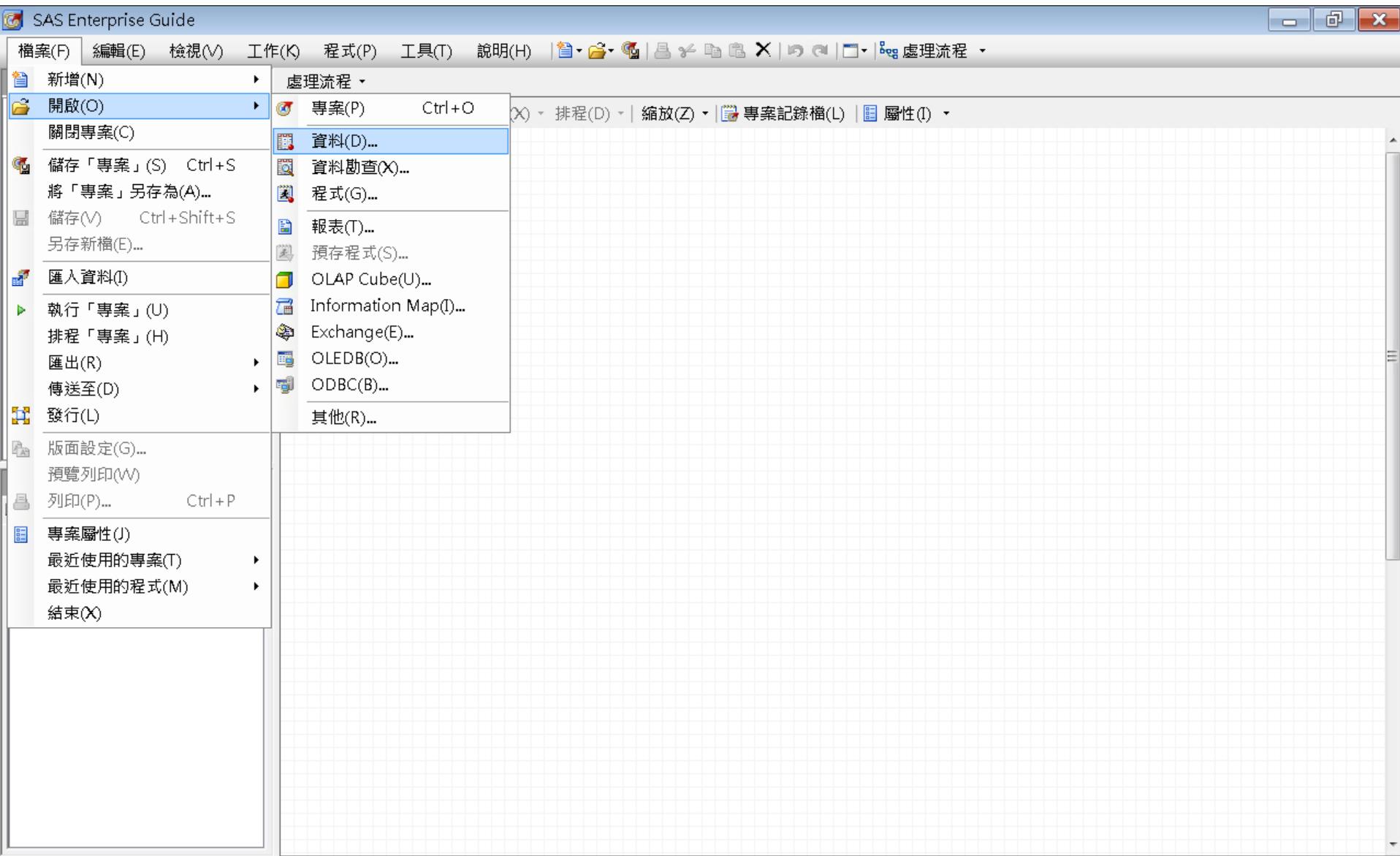
- 登入淡江軟體雲時，會自動產生一部新的雲端主機。登出雲端主機前，請先將檔案儲存於您攜帶的隨身碟、淡江大學網路隨身碟、Google雲端硬碟...等，否則登出後雲端主機的所有資料將會被刪除，並回到您的個人主機畫面。
- 為免雲端主機帳號被冒用，每次使用完請登出。
- 軟體雲相關操作說明，請點選雲端主機桌面上『操作手冊』圖示。
- 軟體雲使用問題請至B212辦公室，或來電26215656-2129為您服務。
- 請尊重智慧財產權，嚴禁安裝不法軟體，違法者自負相關法律責任。

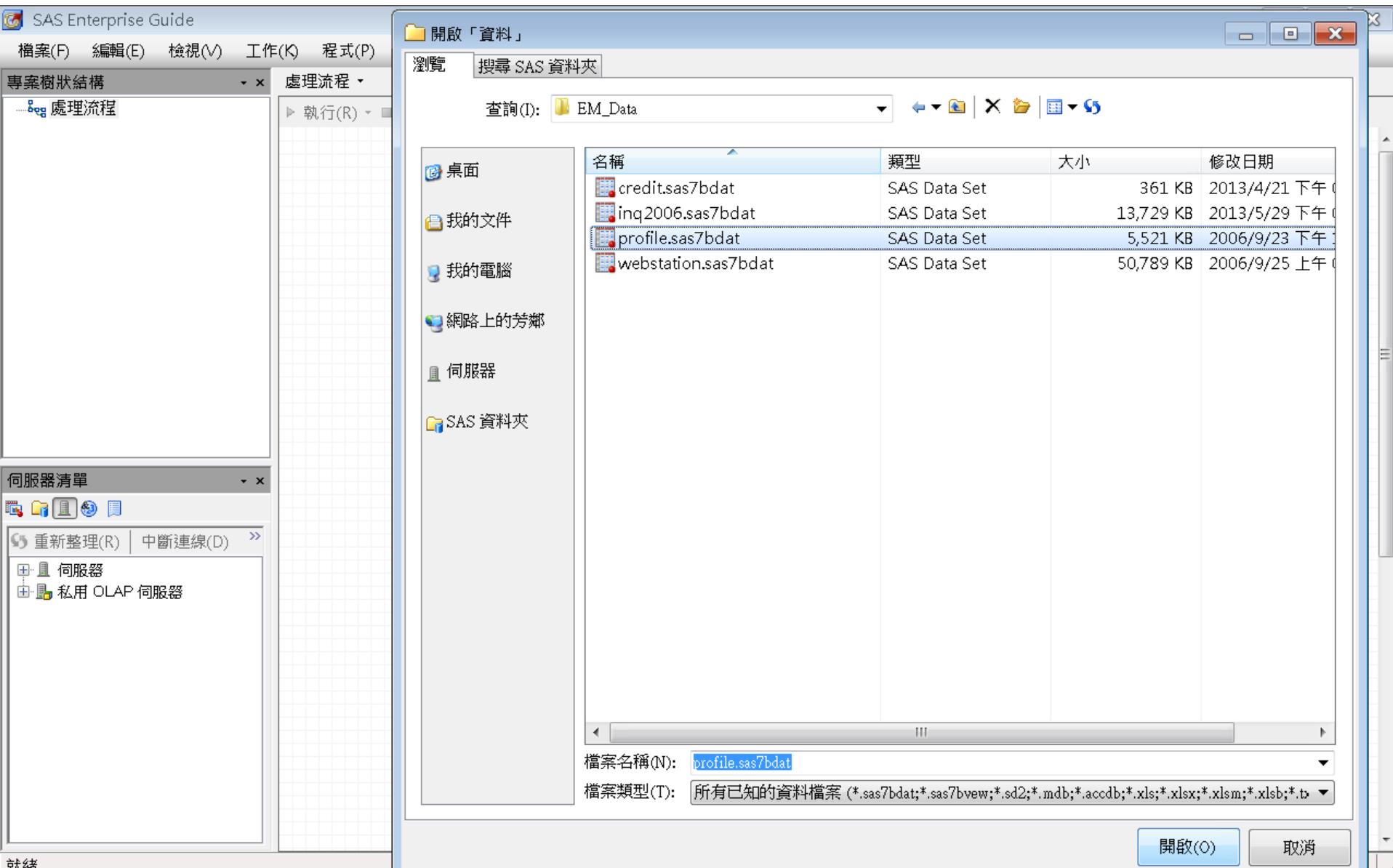
New Project



Open SAS Data File: Profile.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(X) | 處理流程

專案樹狀結構

- 處理流程
 - profile

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)

ID	CNT_TBM	CNT_ATM	CNT_POS	CNT_CSC	CNT_TOT
1 000041360	34	3	3	9	49
2 000052094	44	17	5	18	84
3 000057340	122	26	32	36	216
4 000076885	42	3	6	1	52
5 000089150	20	15	2	2	39
6 000094512	83	20	49	3	155
7 000096396	33	9	49	17	108
8 000098062	22	5	9	1	37
9 000122779	21	27	2	1	51
10 000125838	127	3	2	3	135
11 000137317	52	20	2	11	85
12 000147896	101	40	8	1	150
13 000150206	156	52	80	4	292
14 000160501	31	32	42	12	117
15 000176820	107	3	2	1	113
16 000187784	54	14	7	24	99
17 000189200	92	3	2	2	99
18 000198327	44	27	17	3	91
19 000200237	146	3	2	1	152
20 000207980	39	5	2	4	50
21 000227355	180	3	2	10	195
22 000228317	46	14	2	1	63
23 000228906	25	3	2	1	31
24 000229984	15	3	2	1	21
25 000241022	13	3	2	1	19
26 000253260	65	3	2	1	71
27 000257910	35	3	2	1	41
28 000261717	11	5	2	1	19
29 000266600	15	3	2	1	21
30 000269176	107	78	22	7	214
31 000273510	110	15	6	21	152

伺服器清單

重新整理(R) 中斷連線(D)

伺服器

本機

資料館

- EC100001
- MAPS
- MAPSGFK
- MAPSSAS
- SASHHELP
- SASUSER
- WORK

檔案

私用 OLAP 伺服器

就緒

未選取設定檔

SAS Enterprise Guide

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

專案樹狀結構 profile

處理流程 profile

開啟「profile」(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)

ID	CNT_TBM	CNT_ATM	CNT_POS	CNT_CSC	CNT_TOT
000041360	34	3	3	9	49
000052094	44	17	5	18	84
000057340	122	26	32	36	216
000076885	42	3	6	1	52
000089150	20	15	2	2	39
000094512	83	20	49	3	155
000096396	33	9	49	17	108
000098062	22	5	9	1	37
000122779	21	27	2	1	51
000125838	127	3	2	3	135
000137317	52	20	2	11	85
000147896	101	40	8	1	150
000150206	156	52	80	4	292
000160501	31	32	42	12	117
15	000176820	107	3	2	113
16	000187784	54	14	7	99
17	000189200	92	3	2	99
18	000198327	44	27	17	91
19	000200237	146	3	2	152
20	000207980	39	5	2	50
21	000227355	180	3	2	195
22	000228317	46	14	2	63
23	000228906	25	3	2	31
24	000229984	15	3	2	21
25	000241022	13	3	2	19
26	000253260	65	3	2	71
27	000257910	35	3	2	41
28	000261717	11	5	2	19
29	000266600	15	3	2	21
30	000269176	107	78	22	214
31	000273510	110	15	6	152

伺服器清單

重新整理(R) 中斷連線(D)

伺服器
本機
資料館
EC100001
MAPS
MAPSGFK
MAPSSAS
SASHHELP
SASUSER
WORK
檔案
私用 OLAP 伺服器

就緒 未選取設定檔

SAS Enterprise Guide

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

專案樹狀結構

- 處理流程
- profile

伺服器清單

- 重新整理(R) 中斷連
- 伺服器
- 本機
- 資料館
 - EC10000
 - MAPS
 - MAPSGF
 - MAPSSA
 - SASHHELP
 - SASUSER
 - WORK
- 檔案
- 私用 OLAP 伺服器

「profile」的屬性

一般

檔案屬性

檔案名稱: C:\DATA\EM_Data\profile.sas7bdat [變更\(C\)...](#)

建立日期: 2006年9月23日下午 06:07

上次修改日期: 2006年9月23日下午 06:07

資料類型: SAS 資料集

標籤:

其他 (F1)...

確定 取消

30	000269176	107	78	22	7	214
31	000273510	110	15	6	21	152

未選取設定檔

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 profile

伺服器清單

重新整理(R) 中斷連

伺服器 本機 資料館 EC10000 MAPS MAPSGF MAPSSA SASHELP SASUSER WORK 檔案 私用 OLAP 伺服器

顯示資料項目的欄屬性 (名稱、類型、長度、格式、輸入格式及標籤)。您可以使用特定欄的 [屬性] 視窗來修改該欄的這些屬性。

「profile」的屬性

一般 欄 進階 摘要

名稱	類型	長度	格式	輸入格式	標籤
ID	字元	9			
CNT_TBM	數值	8			
CNT_ATM	數值	8			
CNT_POS	數值	8			
CNT_CSC	數值	8			
CNT_TOT	數值	8			

複製到剪貼簿(B)

確定 取消

就緒 未選取設定檔

SAS Enterprise Guide

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

專案樹狀結構 profile

處理流程 profile 開啟「profile」(O) 決定篩選和排序(L) 決定查詢產生器(Q) 決定資料(D) 決定描述(B) 決定圖形(G) 決定分析(Z) 決定匯出(X) 決定傳送至(N)

ID	CNT_TBM	CNT_ATM	CNT_POS	CNT_CSC	CNT_TOT
1000041360	34	3	3	9	49
匯出(X)	匯出「profile」(X)...	17	5	18	84
	匯出「profile」作為專案中的步驟(S)...	26	32	36	216
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)					
X 刪除(D)					
重新命名(M)					
關閉資料(S)					
屬性(I)					
伺服器清單					
重新整理(R)	中斷連線(D)				
伺服器					
本機					
資料館					
EC100001					
MAPS					
MAPSGFK					
MAPSSAS					
SASHHELP					
SASUSER					
WORK					
檔案					
私用 OLAP 伺服器					

就緒 未選取設定檔

匯出

儲存位置(I): EM_Data

名稱

	修改日期	類型
credit.sas7bdat	2013/4/21 下午 05:22	SAS 資料檔案 (V7 完整名稱) (*.sas7bdat)
inq2006.sas7bdat	2013/5/29 下午 01:20	SAS 資料檔案 (V7 完整名稱) (*.sas7bdat)
profile.sas7bdat	2006/9/23 下午 10:08	SAS 資料檔案 (V7 完整名稱) (*.sas7bdat)
webstation.sas7bdat	2006/9/25 上午 01:45	SAS 資料檔案 (V7 完整名稱) (*.sas7bdat)

檔案名稱(N): profile

檔案類型(T): SAS 資料檔案 (V7 完整名稱) (*.sas7bdat)

儲存(S) 取消

送至(N)

CSC	CNT_TOT
9	49
18	84
36	216
1	52
2	39
3	155
17	108
1	37
1	51
3	135
11	85
1	150
4	292
12	117
1	113
24	99
2	99
3	91
1	152
4	50
10	195
1	63
1	31
1	21
1	19
1	71
1	41
1	19
1	21
7	214
21	152

未選取設定檔

SAS Enterprise Guide

檔案(F) 編輯(E) 檢視(V) 工作(W) 程式(P) 工具

專案樹狀結構

- 處理流程
 - profile

匯出

儲存位置(I): EM_Data

名稱

	修改日期	類型
credit.sas7bdat	2013/4/21 下午 05:22	SAS 資料檔案 (V7 完整名稱) (*.sas7bdat)
inq2006.sas7bdat	2013/5/29 下午 01:20	SAS 資料檔案 (V6) (*.sd2)
profile.sas7bdat	2006/9/23 下午 10:08	Microsoft Excel 活頁簿 (*.xlsx)
webstation.sas7bdat	2006/9/25 上午 01:45	Microsoft Access 2002-2003 資料庫 (*.mdb)

桌面
我的文件
我的電腦
網路上的芳鄰
伺服器

重新整理(R) | 中斷連線(D)

伺服器清單

伺服器

- 本機
 - 資料館
 - EC100001
 - MAPS
 - MAPSGFK
 - MAPSSAS
 - SASHHELP
 - SASUSER
 - WORK
 - 檔案

私用 OLAP 伺服器

儲存(S) 取消

檔案名稱(N):

檔案類型(T): SAS 資料檔案 (V7 完整名稱) (*.sas7bdat)

SAS Enterprise Guide

檔案(F) 編輯(E) 檢視(V) 工作(W) 程式(P) 工具

專案樹狀結構

處理流程 profile

伺服器清單

重新整理(R) 中斷連線(D)

伺服器 本機 資料館 EC100001 MAPS MAPSGFK MAPSSAS SASHELP SASUSER WORK 檔案 私用 OLAP 伺服器

就緒

匯出

儲存位置(I): EM_Data

名稱

修改日期

類型

這個資料夾是空的。

1 000041360
2 000052094
3 000057340
4 000076885
5 000089150
6 000094512
7 000096396
8 000098062
9 000122779
10 000125838
11 000137317
12 000147896
13 000150206
14 000160501
15 000176820
16 000187784
17 000189200
18 000198327
19 000200237
20 000207980
21 000227355
22 000228317
23 000228906
24 000229984
25 000241022
26 000253260
27 000257910
28 000261717
29 000266600
30 000269176
31 000273510

檔案名稱(N): profile_excel

檔案類型(T): Microsoft Excel 活頁簿 (*.xlsx)

儲存(S) 取消

Export SAS .sas7bdat to Excel .xlsx File

SAS Enterprise Guide

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

專案樹狀結構

處理流程 profile

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

ID	CNT_TBM	CNT_ATM	CNT_POS	CNT_CSC	CNT_TOT
1 000041360	34	3	3	9	49
2 000052094	44	17	5	18	84
3 000057340	122	26	32	36	216
4 000076885	42	3	6	1	52
5 000089150	20	15	2	2	39
6 000094512	83	20	49	3	155
7 000096396	33	9	49	17	108
8 000098062	22	5	9	1	37
9 000122779	21	27	2	1	51
10 000125838	127	3	2	3	135
11 000137317	50	20	2	11	85
12 000147896				1	150
13 000150206			80	4	292
14 000160501			42	12	117
15 000176820			2	1	113
16 000187784	54	14	7	24	99
17 000189200	92	3	2	2	99
18 000198327	44	27	17	3	91
19 000200237	146	3	2	1	152
20 000207980	39	5	2	4	50
21 000227355	180	3	2	10	195
22 000228317	46	14	2	1	63
23 000228906	25	3	2	1	31
24 000229984	15	3	2	1	21
25 000241022	13	3	2	1	19
26 000253260	65	3	2	1	71
27 000257910	35	3	2	1	41
28 000261717	11	5	2	1	19
29 000266600	15	3	2	1	21
30 000269176	107	78	22	7	214
31 000273510	110	15	6	21	152

Enterprise Guide 正在匯出「profile」...

profile.sas7bdat SAS Data Set 5.39 MB

profile_excelexcel.xlsx Microsoft Excel 工作表 1.45 MB

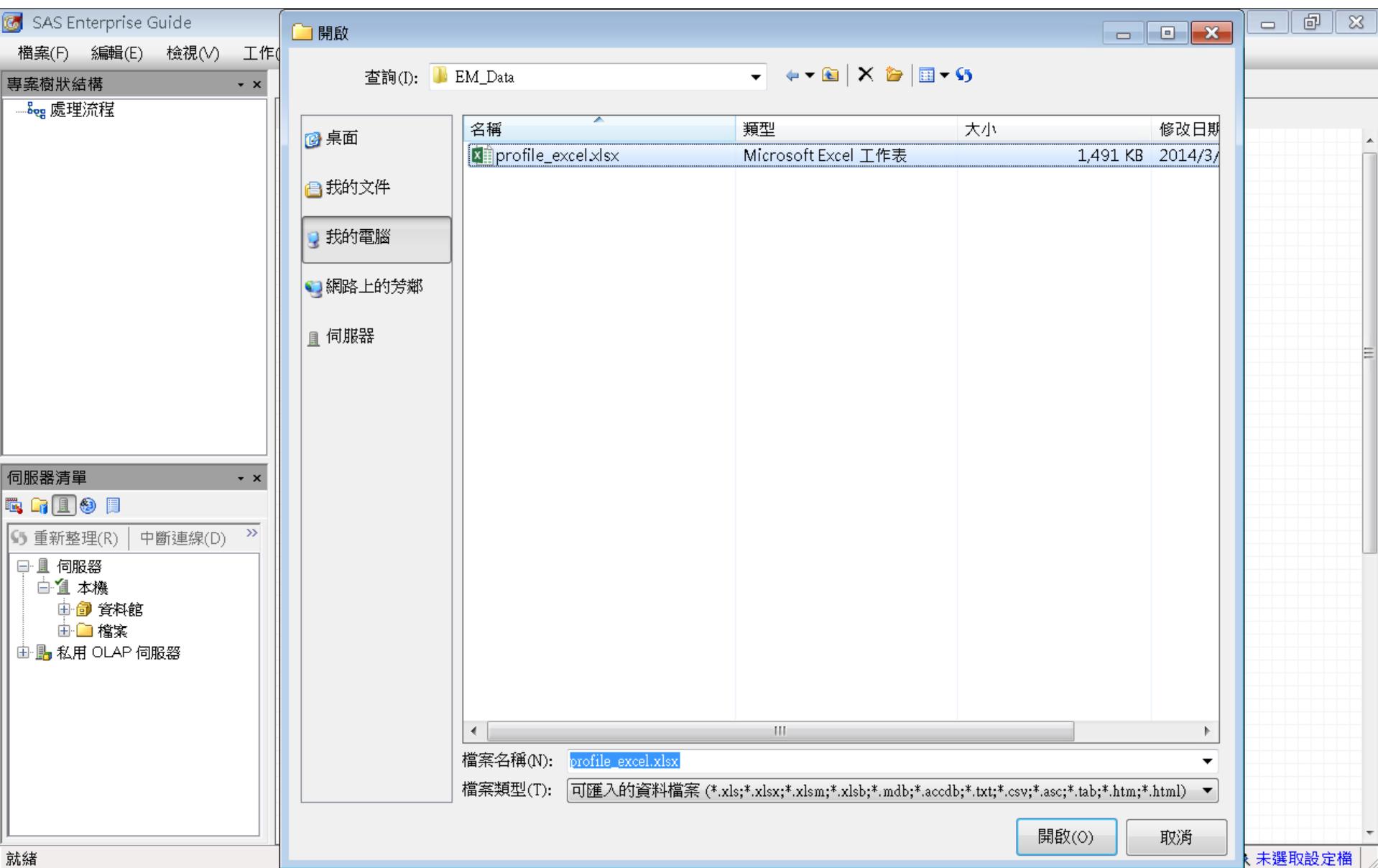
(詳細資料) 正在執行: 匯出工作 - 正在執行...

未選取設定檔

Import Excel File to SAS EG



Import Excel File to SAS EG



SAS Enterprise Guide

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

專案樹狀結構

處理流程

profile_excel.xlsx
匯入資料 (profile_excel.xls)

處理流程

執行(R) | 停止(S) | 匯出(C)

profile_e... | **匯入資料 (profile_...**

從 profile_excel.xlsx 匯入資料

1 / 4 指定資料

對於其他需要使用 SAS 資料檔案以進行資料分析和報表的工作，「匯入資料」精靈可將非 SAS 資料轉換成 SAS 資料檔案。

來源資料檔案

位置(L): 本機檔案系統
檔案路徑(P): C:\DATA\EM_Data\profile_excel.xlsx
資料類型(T): Excel 活頁簿

輸出 SAS 資料集

SAS 伺服器(S): Local
資料館(R): WORK
資料集(A): profile_excel

瀏覽(W)...

<上一步(B) | 下一步(N) > | 完成(F) | 取消 | 說明

就緒

未選取設定檔

SAS Enterprise Guide

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

專案樹狀結構

處理流程

執行(R) 停止(S) 決策(X) 檔案(Y) 檔案(Z) 檔案(E) 檔案(F) 檔案(G) 檔案(H) 檔案(I) 檔案(J) 檔案(K) 檔案(L) 檔案(M) 檔案(N) 檔案(O) 檔案(P) 檔案(Q) 檔案(R) 檔案(S) 檔案(T) 檔案(U) 檔案(V) 檔案(W) 檔案(X) 檔案(Y) 檔案(Z)

profile_excel.xlsx

匯入資料 (profile_excel.xls)

處理流程

匯入資料 (profile_excel.xls)

profile_e... 決策(X) 檔案(Y) 檔案(Z) 檔案(E) 檔案(F) 檔案(G) 檔案(H) 檔案(I) 檔案(J) 檔案(K) 檔案(L) 檔案(M) 檔案(N) 檔案(O) 檔案(P) 檔案(Q) 檔案(R) 檔案(S) 檔案(U) 檔案(V) 檔案(W) 檔案(X) 檔案(Y) 檔案(Z)

從 profile_excel.xlsx 決策(X) 檔案(Y) 檔案(Z) 檔案(E) 檔案(F) 檔案(G) 檔案(H) 檔案(I) 檔案(J) 檔案(K) 檔案(L) 檔案(M) 檔案(N) 檔案(O) 檔案(P) 檔案(Q) 檔案(R) 檔案(S) 檔案(U) 檔案(V) 檔案(W) 檔案(X) 檔案(Y) 檔案(Z)

2 / 4 選取資料來源

選取範圍

使用工作表(W)
profile

範圍的第一列包含欄位名稱(M)

重新命名欄以符合 SAS 命名規則(R)。

使用工作表內特定範圍的儲存格(S)

左上方儲存格(T):

右下方儲存格(L):

必要時展開列範圍(X)

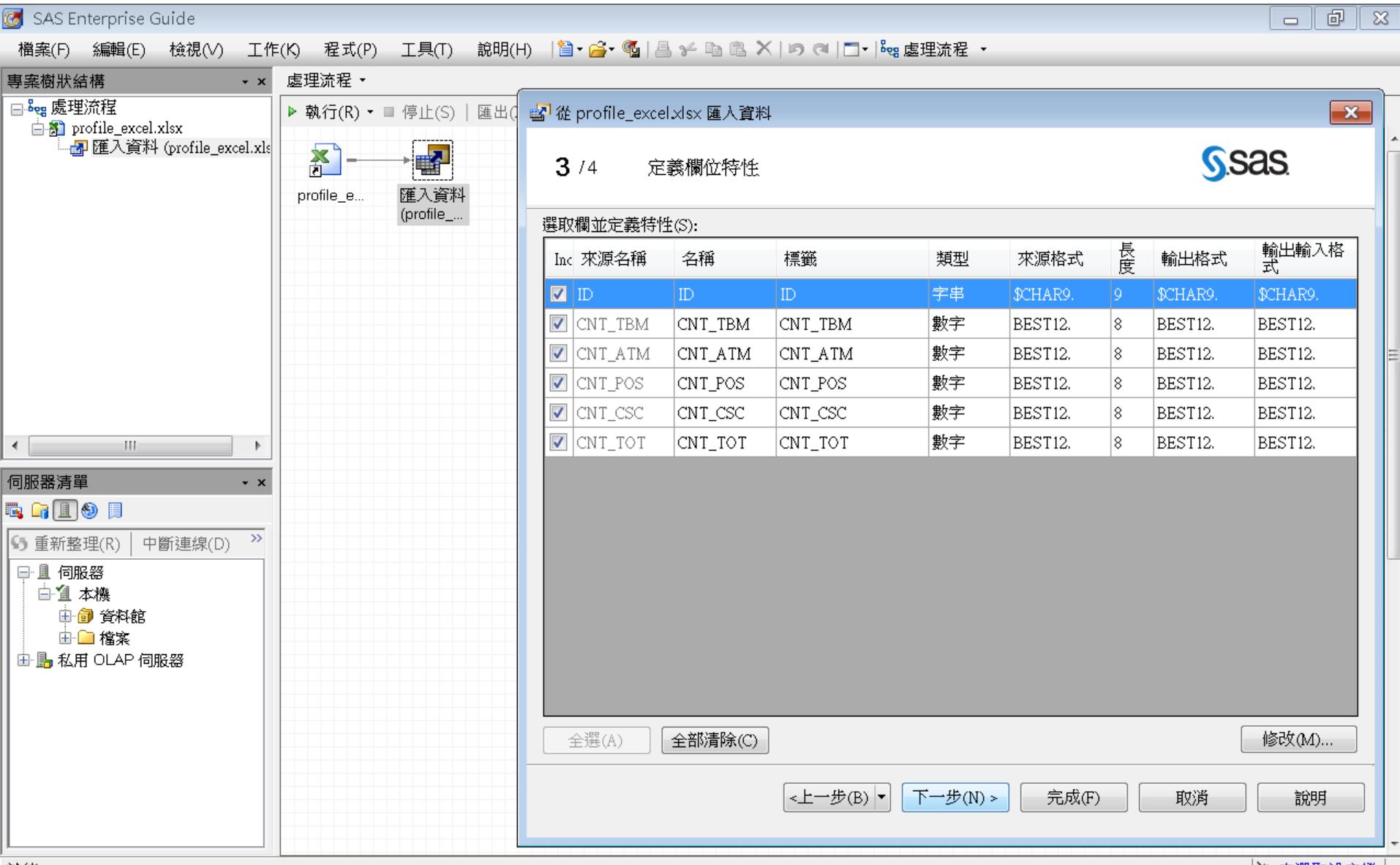
重設範圍(A)

使用預先定義的命名範圍(P)

上一步(B) 下一步(N) > 完成(F) 取消 說明

就緒

未選取設定檔



SAS Enterprise Guide

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

專案樹狀結構

- 處理流程
 - profile_excel.xlsx
 - 匯入資料 (profile_excel.xls)

處理流程

執行(R) 停止(S) | 汇出(C)

從 profile_excel.xlsx 匯入資料

4 / 4 進階選項

將資料內嵌在產生的 SAS 程式碼中(E)。

盡可能使 SAS/ACCESS Interface to PC Files 匯入資料(P)。

從文字型資料檔案移除可能導致傳輸錯誤的字元(R)。

<上一步(B) 下一步(N) > 完成(F) 取消 說明

伺服器清單

重新整理(R) 中斷連線(D)

伺服器

- 本機
- 資料館
- 檔案

私用 OLAP 伺服器

SAS Enterprise Guide

檔案(F) 編輯(E) 檢視(V) 工作(W) 程式(P) 工具(T) 說明(H) | 檔案(F) 資料(B) 命令(C) 組合(G) 檔案(E) X | 檔案(F) 資料(B) 命令(C) 組合(G) 檔案(E) | 處理流程

專案樹狀結構

- 處理流程
 - profile_excel.xlsx
 - 匯入資料 (profile_excel.xlsx)

處理流程

執行(R) | 停止(S) | 匯出(X) | 排程(D) | 縮放(Z) | 專案記錄檔(L) | 屬性(I)

```
graph LR; A[profile_e...]; A --> B[匯入資料  
(profile_...  
匯入的資料)]; B --> C[從  
profile_ex  
celxlsx  
匯入的資料]
```

伺服器清單

重新整理(R) | 中斷連線(D) >>

- 伺服器
 - 本機
 - 資料館
 - 檔案
- 私用 OLAP 伺服器

就緒

未選取設定檔

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 | 處理流程

專案樹狀結構

- 處理流程
- profile_excel.xlsx
- 匯入資料 (profile_excel.xls)

匯入資料 (profile_excel.xlsx[profile])

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) |

ID	CNT_TBM	CNT_ATM	CNT_POS	CNT_CSC	CNT_TOT
1	000041360	34	3	3	49
2	000052094	44	17	5	84
3	000057340	122	26	32	216
4	000076885	42	3	6	52
5	000089150	20	15	2	39
6	000094512	83	20	49	155
7	000096396	33	9	49	108
8	000098062	22	5	9	37
9	000122779	21	27	2	51
10	000125838	127	3	2	135
11	000137317	52	20	2	85
12	000147896	101	40	8	150
13	000150206	156	52	80	292
14	000160501	31	32	42	117
15	000176820	107	3	2	113
16	000187784	54	14	7	99
17	000189200	92	3	2	99
18	000198327	44	27	17	91
19	000200237	146	3	2	152
20	000207980	39	5	2	50
21	000227355	180	3	2	195
22	000228317	46	14	2	63
23	000228906	25	3	2	31
24	000229984	15	3	2	21
25	000241022	13	3	2	19
26	000253260	65	3	2	71
27	000257910	35	3	2	41
28	000261717	11	5	2	19
29	000266600	15	3	2	21

就緒 未選取設定檔

SAS Enterprise Guide

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

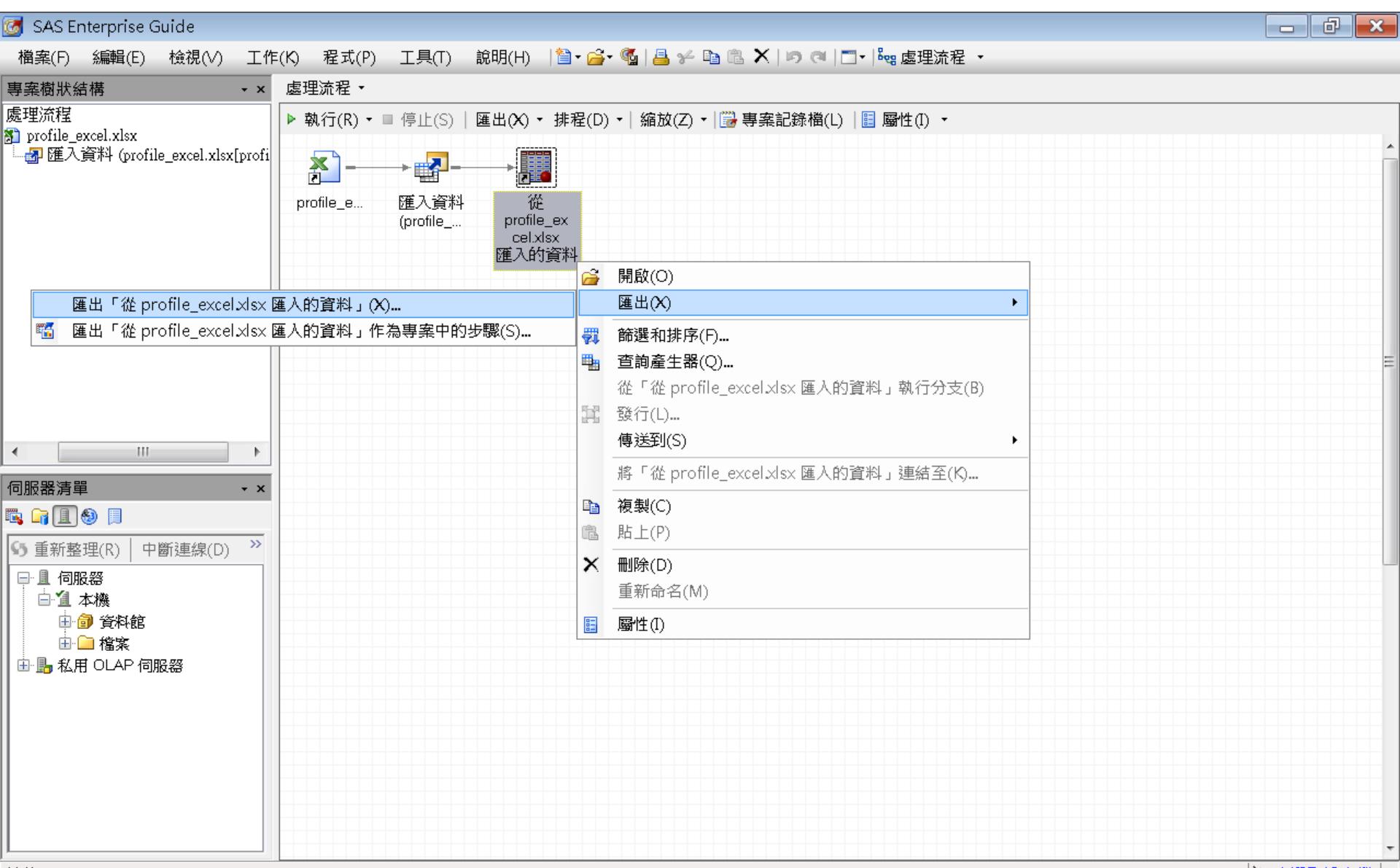
匯入資料 (profile_excel.xlsx[profile])

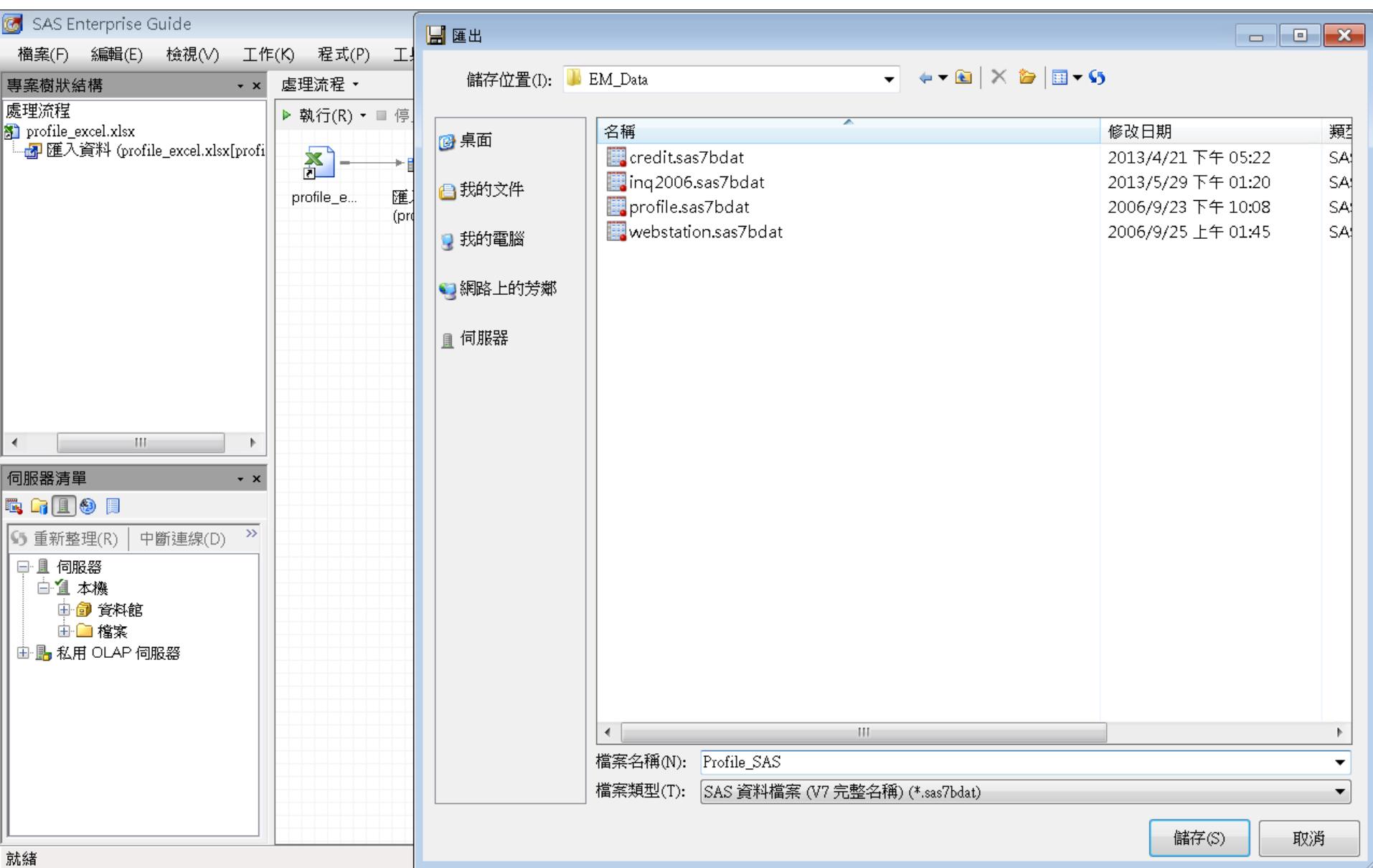
程式碼 記錄檔 輸出資料

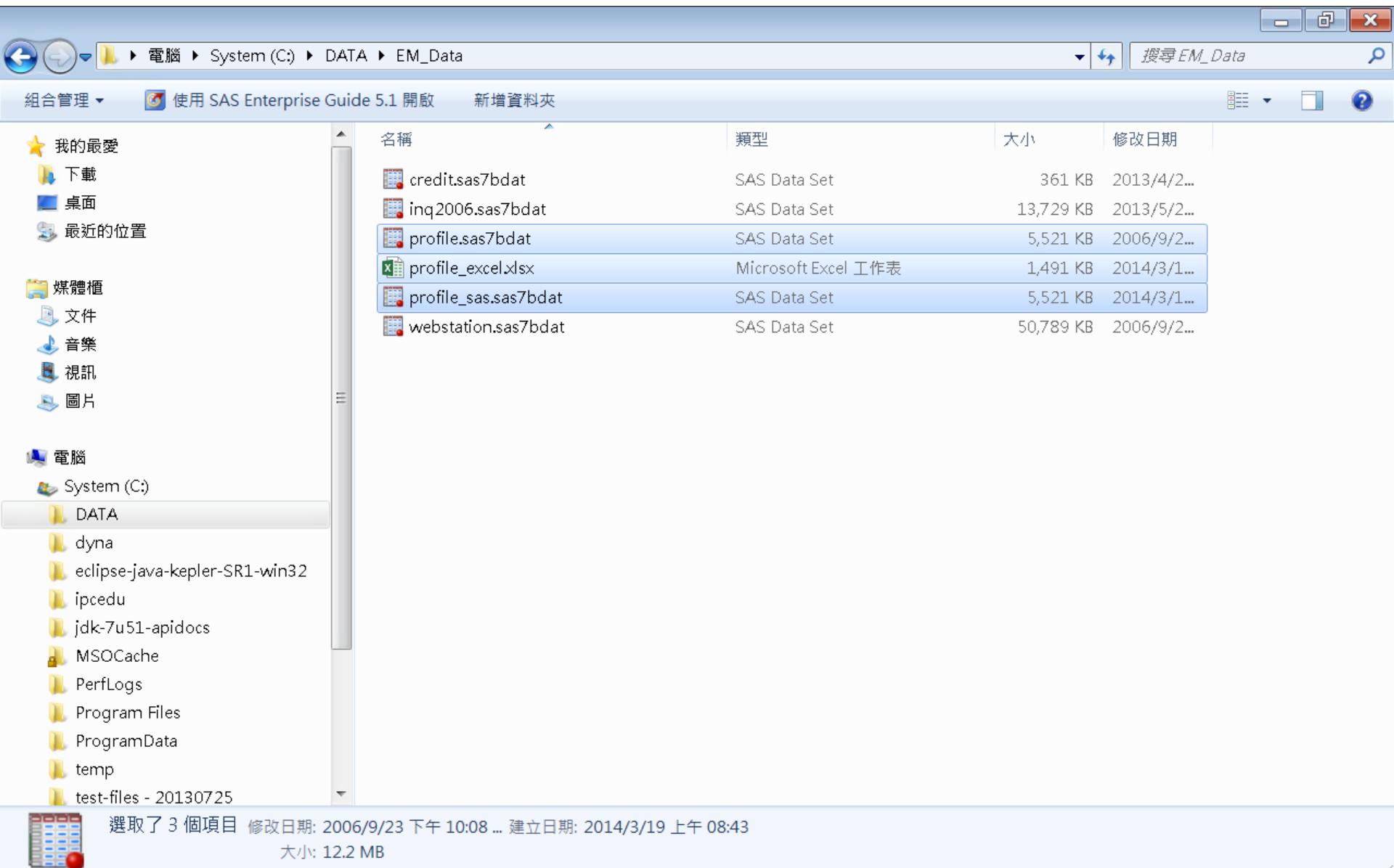
修改工作(Y) 篩選和排序(L) 查詢產生器(Q) 資料(D) 描述(B) 圖形(G) 分析(Z) 匯出(X) 傳送至(N)

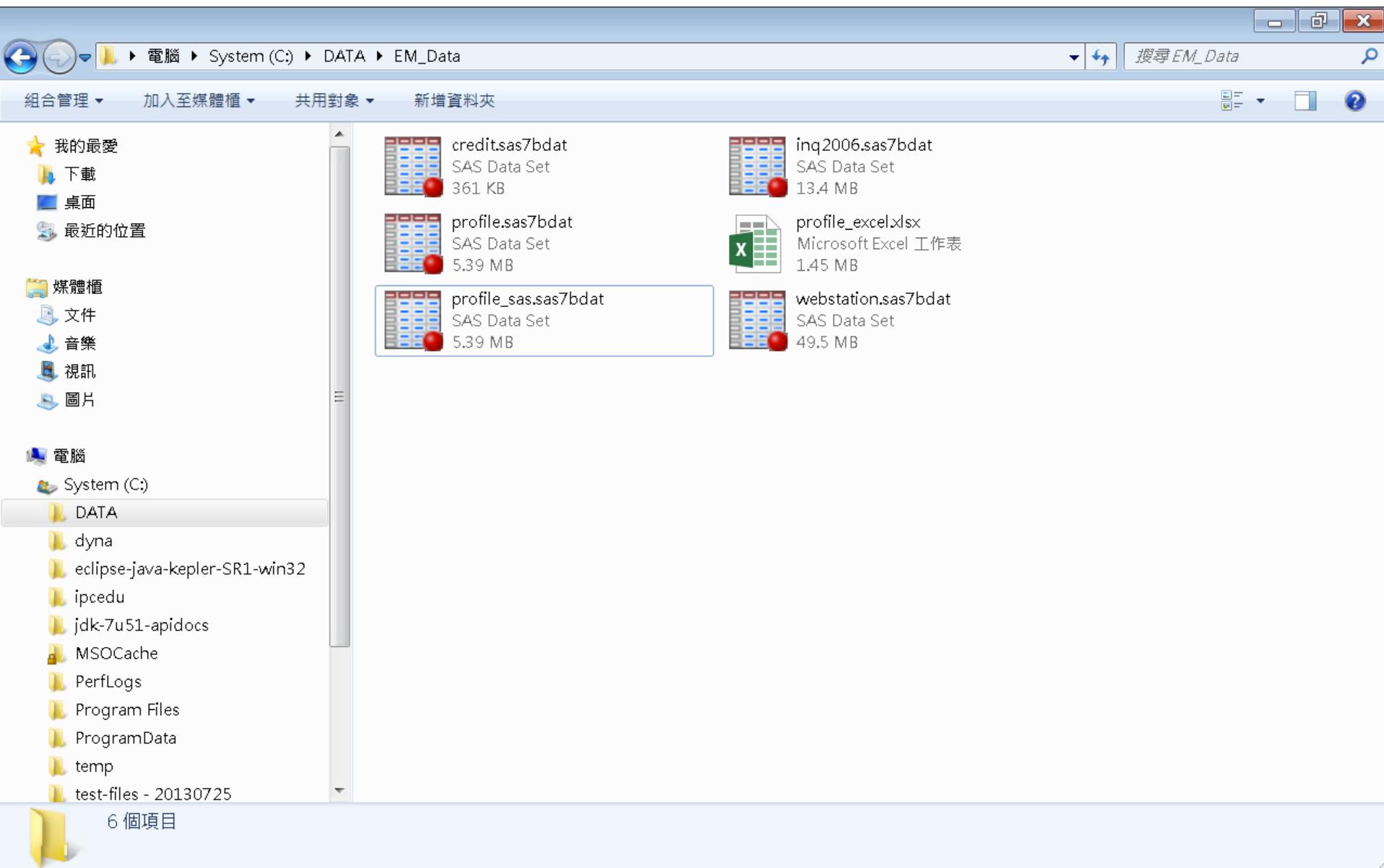
ID	CNT_TBM	CNT_ATM	CNT_POS	CNT_CSC	CNT_TOT
99972	875475072	81	37	2	121
99973	875478517	56	15	10	84
99974	875481914	49	28	91	255
99975	875484829	233	10	5	249
99976	875488178	104	3	2	110
99977	875505005	27	47	37	164
99978	875508051	28	3	2	34
99979	875511290	13	5	2	21
99980	875514158	54	8	18	84
99981	875528729	71	4	14	90
99982	875536451	109	3	2	125
99983	875555772	26	22	2	51
99984	875568109	124	15	2	153
99985	875577833	115	80	9	209
99986	875589010	48	3	82	141
99987	875593202	143	9	2	158
99988	875594835	51	3	2	57
99989	875599149	58	15	27	117
99990	875606621	25	3	2	31
99991	875619290	103	6	2	121
99992	875625622	10	4	2	17
99993	875625635	41	3	2	47
99994	875628606	20	14	23	58
99995	875636141	132	3	2	148
99996	875638608	24	6	2	33
99997	875656523	79	9	2	97
99998	875663288	44	111	57	221
99999	875670226	10	5	2	18
100000	875674475	128	18	20	167

Export Excel .xlsx File to SAS .sas7bdat File









Profile_Excel.xlsx



profile_excel.xlsx
Microsoft Excel 工作表
1.45 MB

profile_excel.xlsx - Excel

檔案 常用 插入 版面配置 公式 資料 校閱 檢視 小組

貼上 剪貼簿

字型 對齊方式 數值 樣式 儲存格 編輯

A1 : fx ID

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	C
1	ID	CNT_TBM	CNT_ATM	CNT_POS	CNT_CSC	CNT_TOT									
2	000041360	34	3	3	9	49									
3	000052094	44	17	5	18	84									
4	000057340	122	26	32	36	216									
5	000076885	42	3	6	1	52									
6	000089150	20	15	2	2	39									
7	000094512	83	20	49	3	155									
8	000096396	33	9	49	17	108									
9	000098062	22	5	9	1	37									
10	000122779	21	27	2	1	51									
11	000125838	127	3	2	3	135									
12	000137317	52	20	2	11	85									
13	000147896	101	40	8	1	150									
14	000150206	156	52	80	4	292									
15	000160501	31	32	42	12	117									
16	000176820	107	3	2	1	113									
17	000187784	54	14	7	24	99									
18	000189200	92	3	2	2	99									
19	000198327	44	27	17	3	91									
20	000200237	146	3	2	1	152									
21	000207000	20	5	2	1	50									

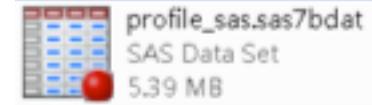
profile +

就緒

100%

80

Profile_SAS.sas7bdat



SAS Enterprise Guide

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

專案樹狀結構 profile_sas

處理流程 profile_sas

伺服器清單 重新整理(R) 中斷連線(D)

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)

ID	CNT_TBM	CNT_ATM	CNT_POS	CNT_CSC	CNT_TOT
1	34	3	3	9	49
2	44	17	5	18	84
3	122	26	32	36	216
4	42	3	6	1	52
5	20	15	2	2	39
6	83	20	49	3	155
7	33	9	49	17	108
8	22	5	9	1	37
9	21	27	2	1	51
10	127	3	2	3	135
11	52	20	2	11	85
12	101	40	8	1	150
13	156	52	80	4	292
14	31	32	42	12	117
15	107	3	2	1	113
16	54	14	7	24	99
17	92	3	2	2	99
18	44	27	17	3	91
19	146	3	2	1	152
20	39	5	2	4	50
21	180	3	2	10	195
22	46	14	2	1	63
23	25	3	2	1	31
24	15	3	2	1	21
25	13	3	2	1	19
26	65	3	2	1	71
27	35	3	2	1	41
28	11	5	2	1	19
29	15	3	2	1	21
30	107	78	22	7	214
31	110	15	6	21	152

未選取設定檔

SAS Enterprise Miner 12.1 (SAS EM)



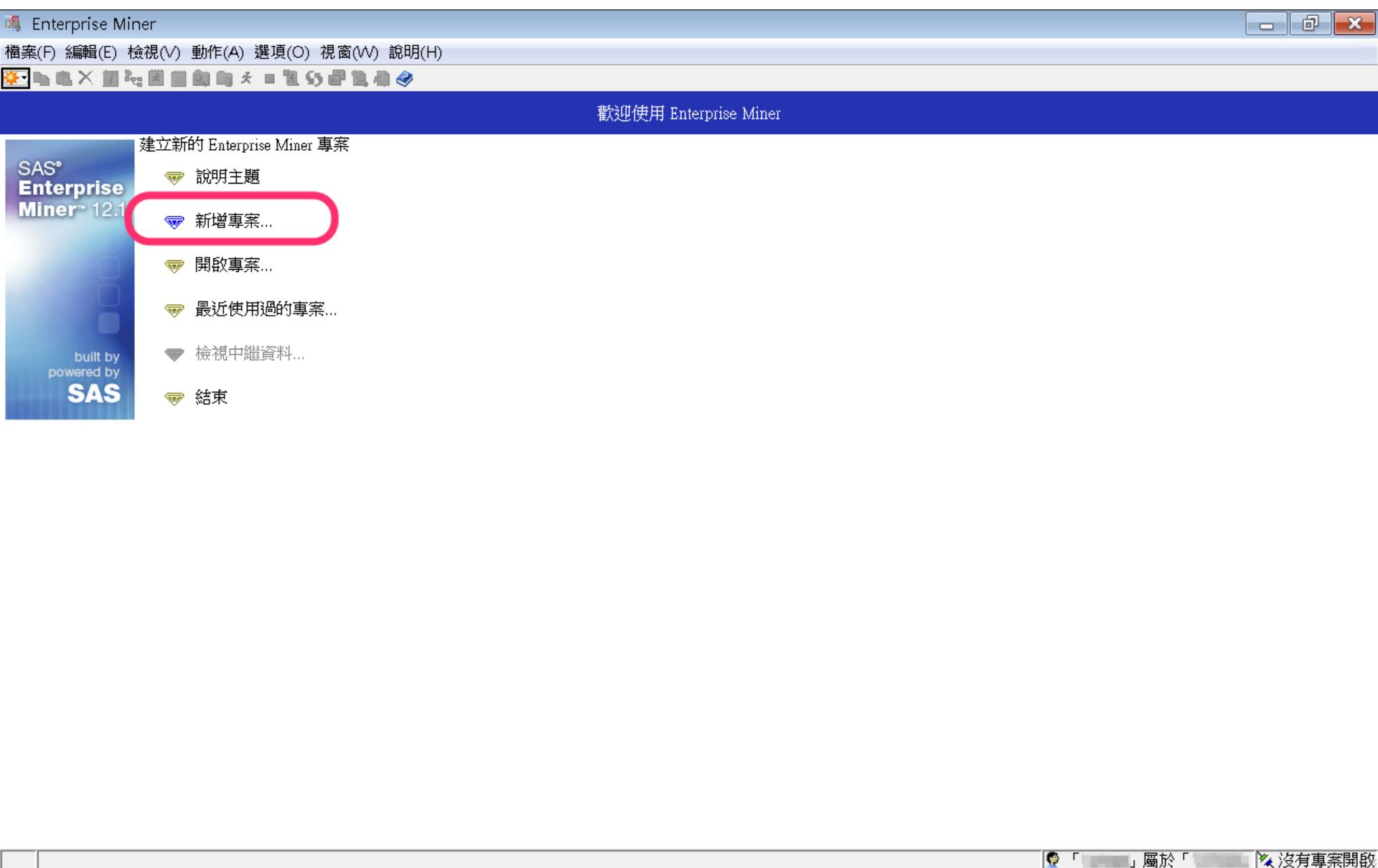
淡江軟體雲使用注意事項

- 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。

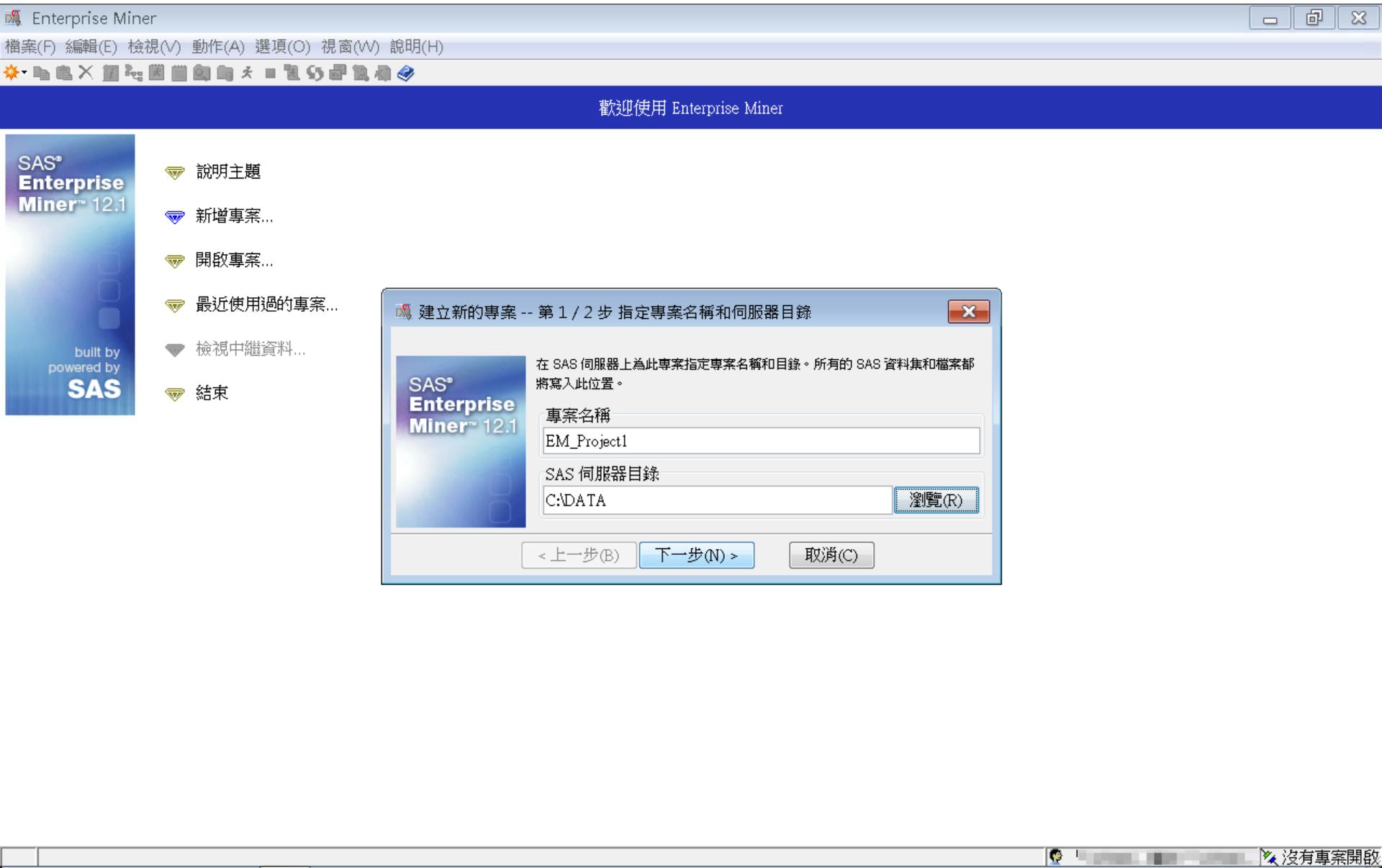
SAS EM 資料匯入4步驟

- 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_Project1)

Enterprise Miner - EM_Project1

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

EM Project1
資料來源
流程圖
模型套件

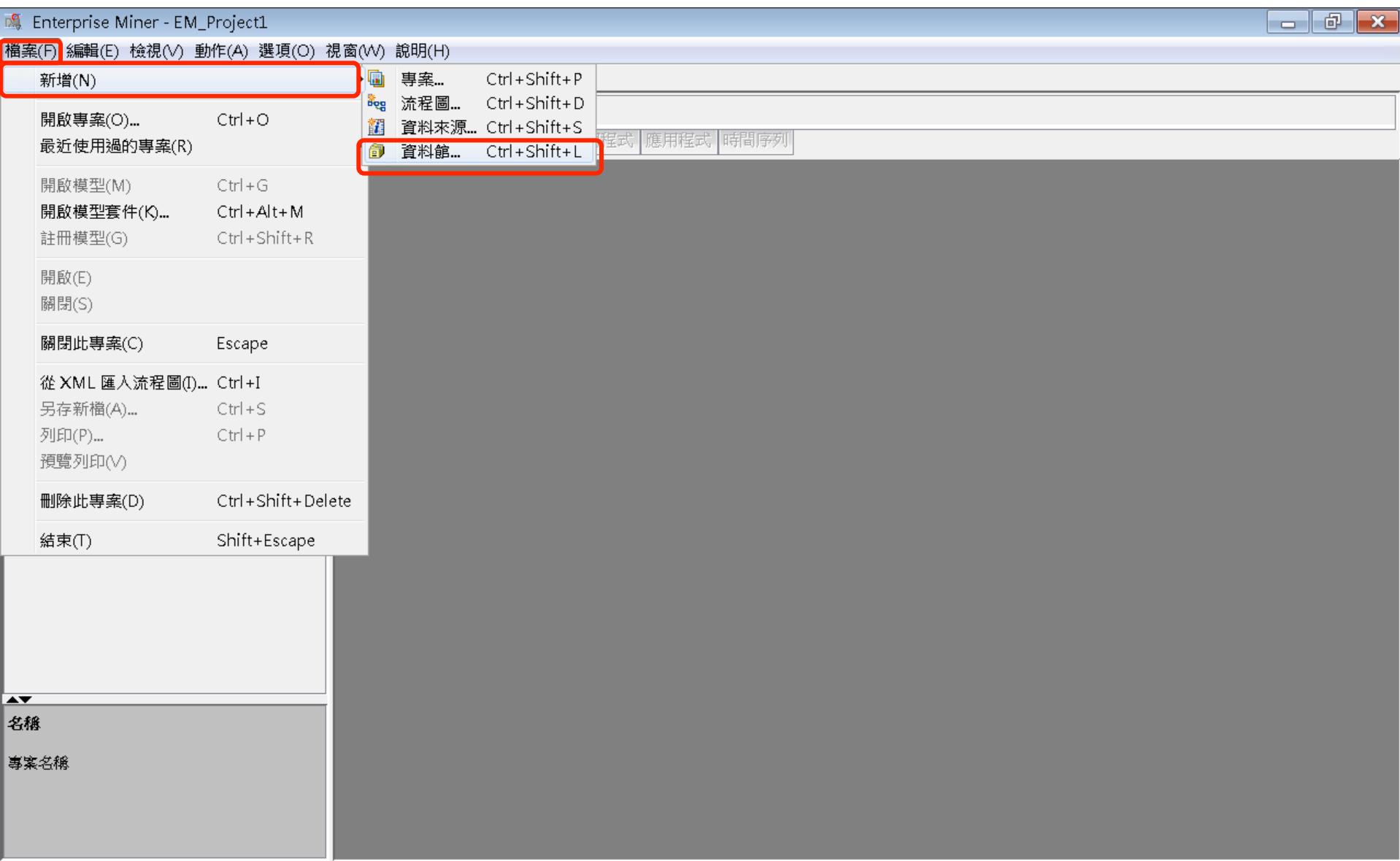
樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

.. 屬性	值
名稱	EM_Project1
專案啟動程式碼	...
專案巨集變數	...
建立時間	2014/3/18 下午 10:38
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

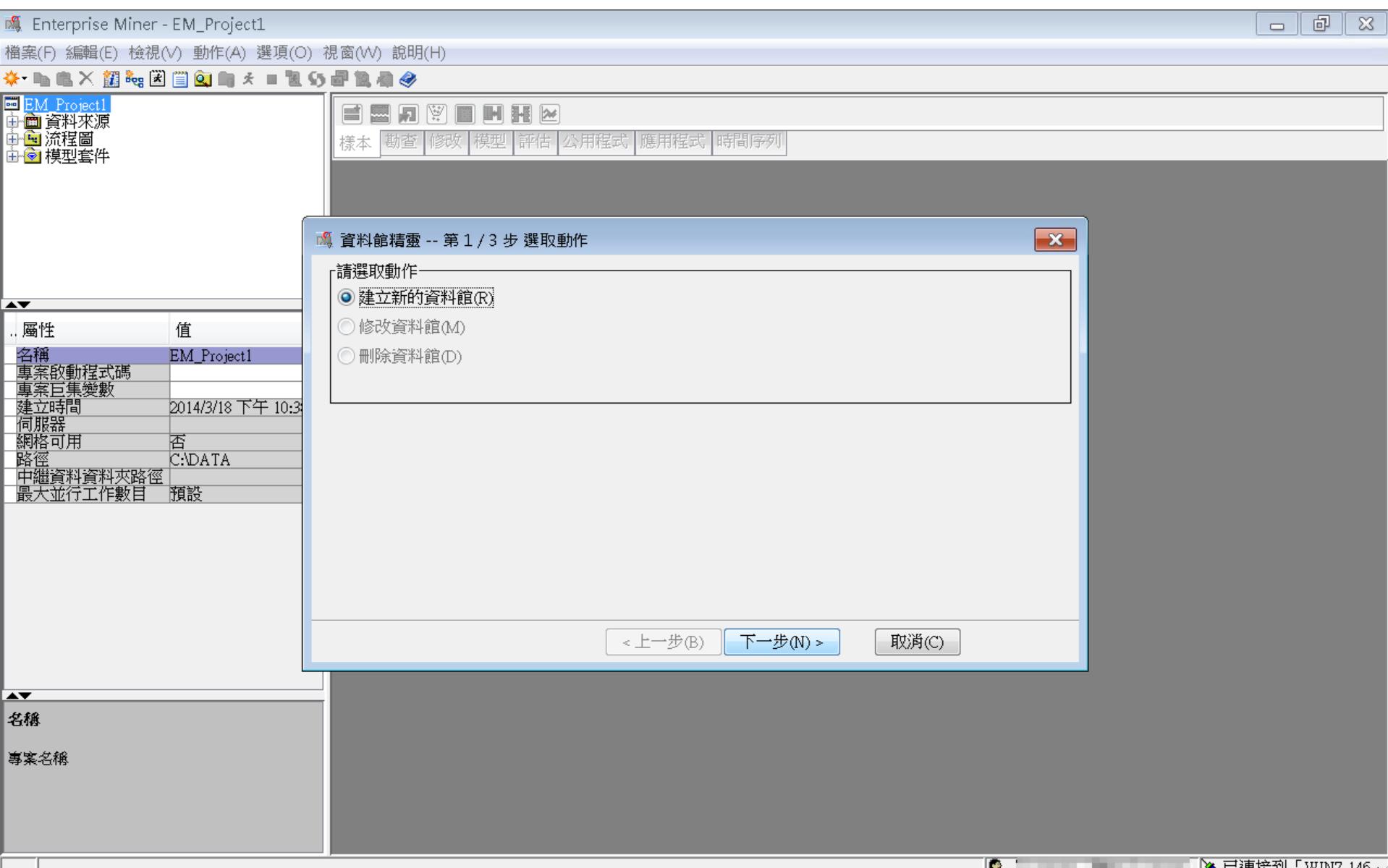
名稱
專案名稱

已連接到「WIN7-146」

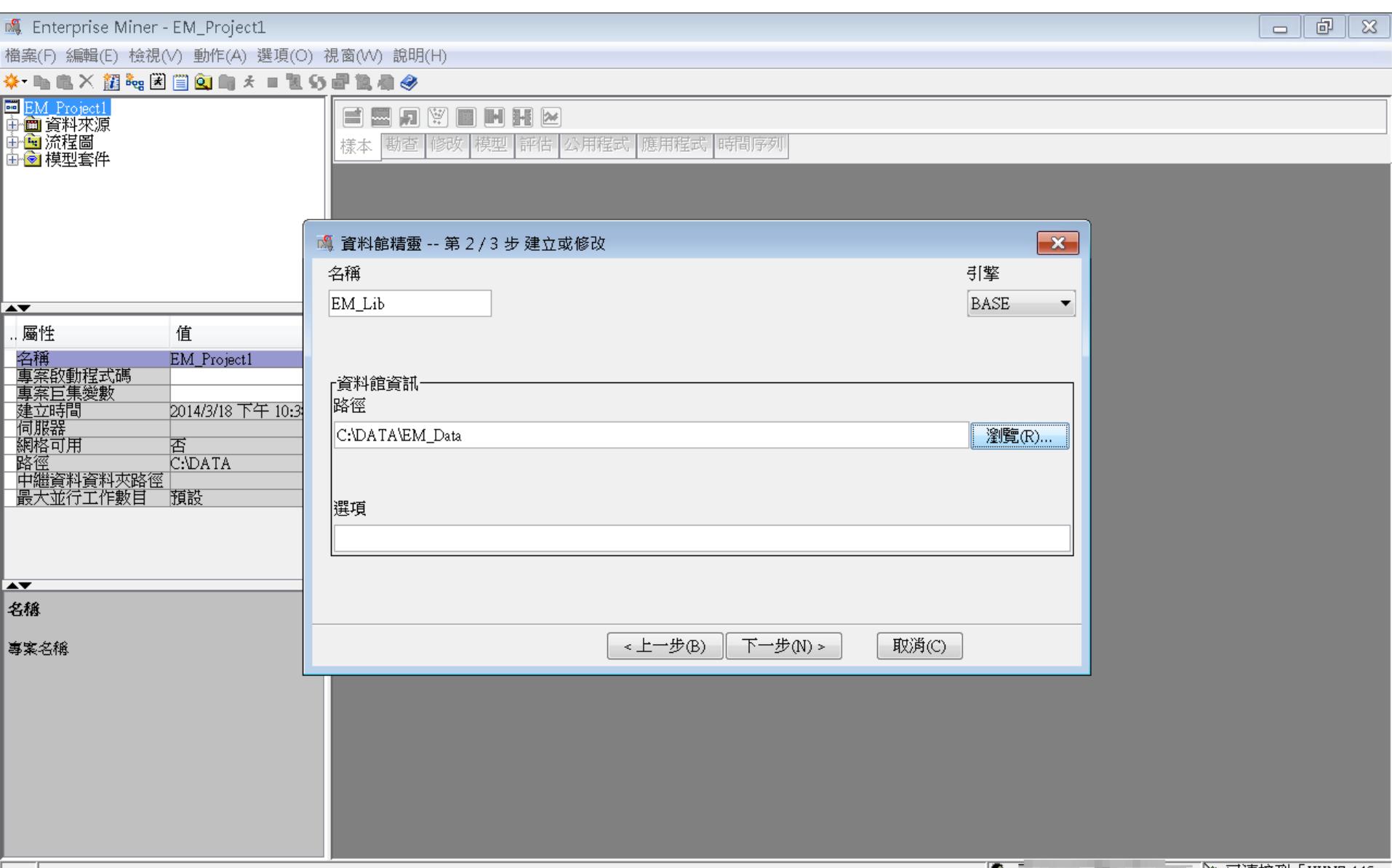
Step 2. 新增資料館 (New / Library)



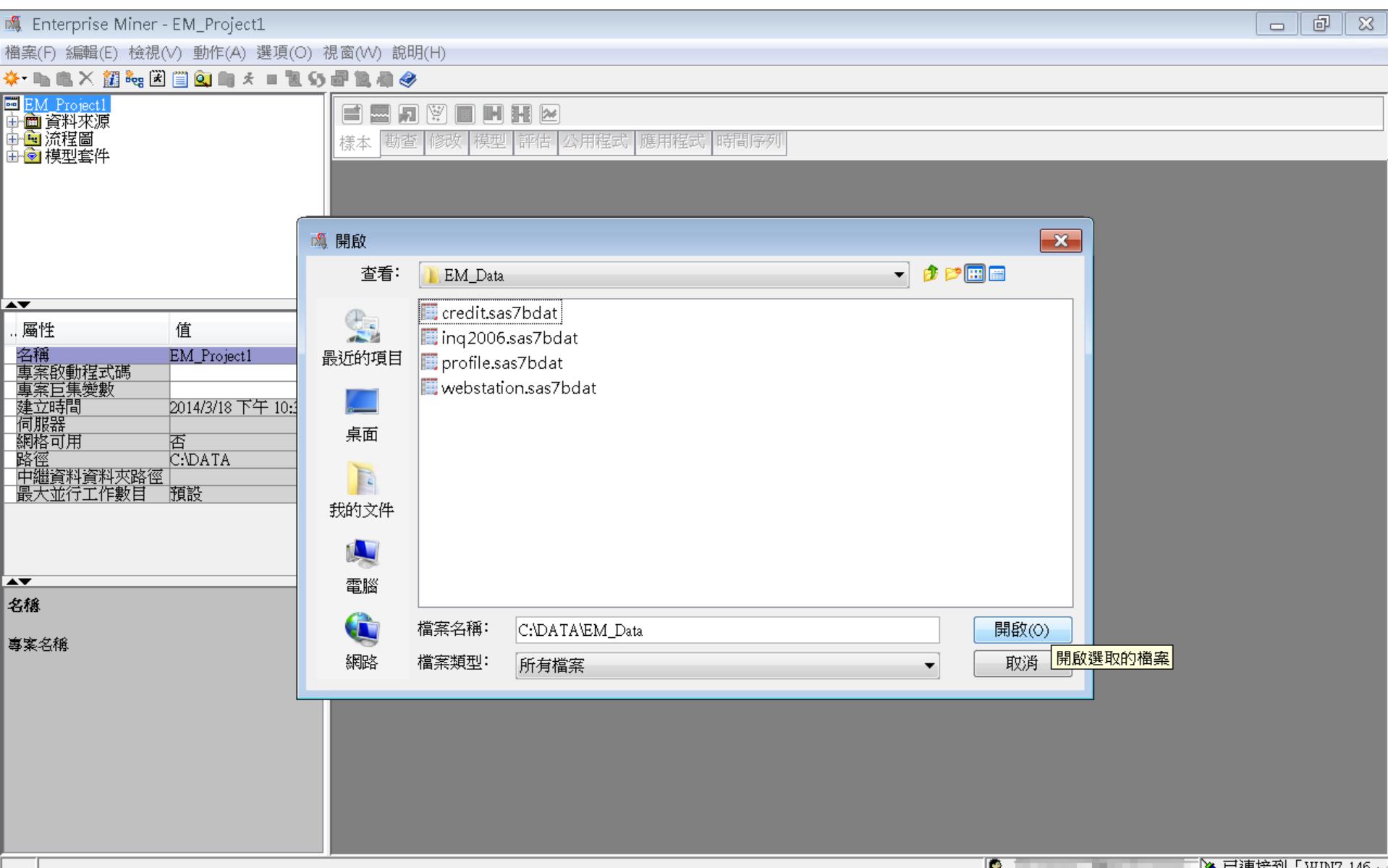
Step 2. 新增資料館 (New / Library)



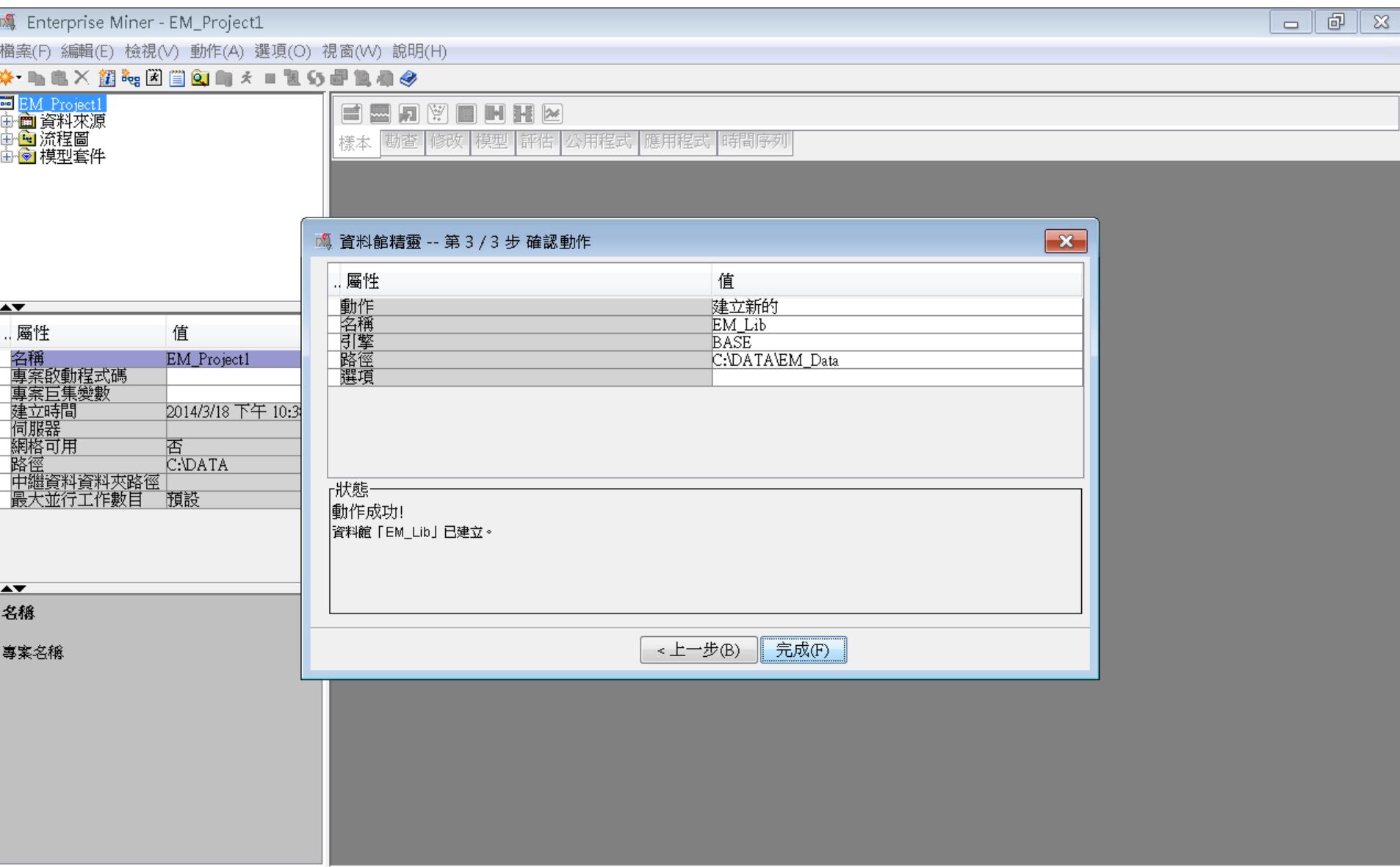
Step 2. 新增資料館 (New / Library)



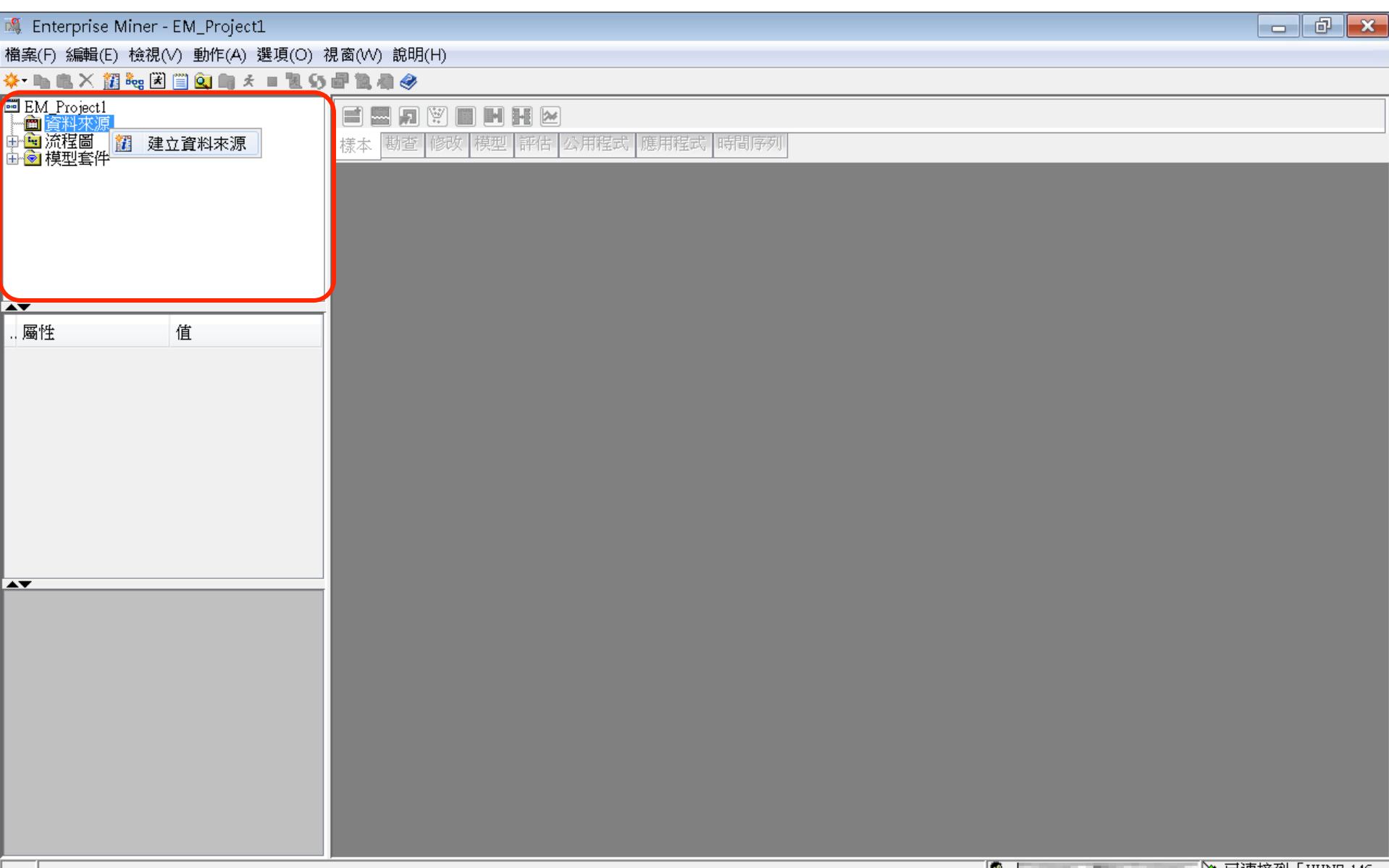
Step 2. 新增資料館 (New / Library)



Step 2. 新增資料館 (New / Library)



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_Project1

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

EM Project1
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 1 / 8 步 中繼資料來源

选取中繼資料來源

來源(S): SAS 表格

< 上一步(B) 下一步(N) > 取消(C)

已連接到「WIN7-146」

Step 3. 建立資料來源 (Create Data Source)

Enterprise Miner - EM_Project1

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

EM Project1
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 2 / 8 步 選取 SAS 表格

選取 SAS 表格

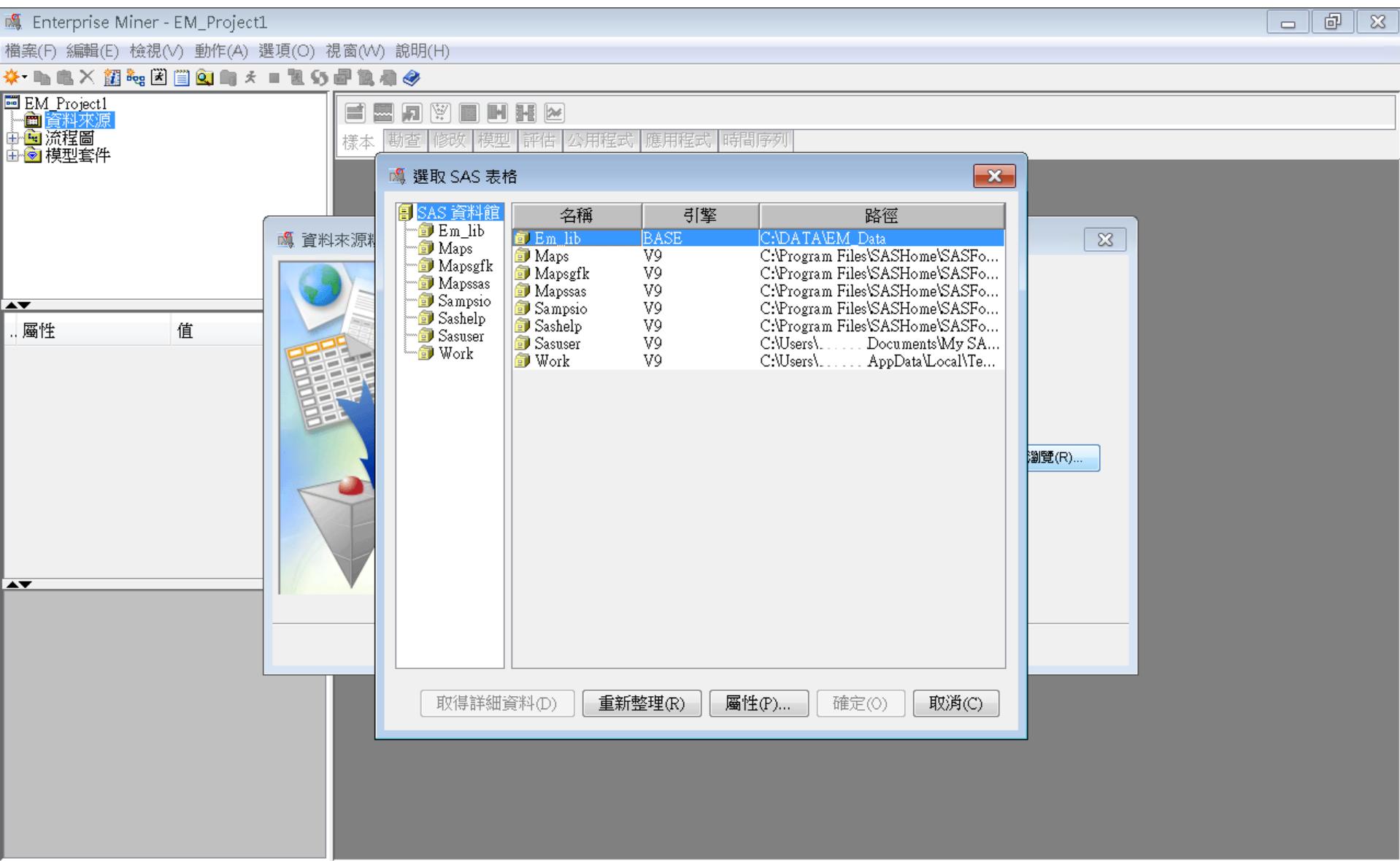
表格(T): [瀏覽\(R\)...](#)

< 上一步(B) 下一步(N) > 取消(C)

已連接到「WIN7-146」



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_Project1

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

EM Project1
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 2 / 8 步 選取 SAS 表格

選取 SAS 表格

選取 SAS 表格

名稱	類型
Credit	表格
Inq2006	表格
Profile	表格
Webstation	表格

取得詳細資料(D) 重新整理(R)... 屬性(P)... 確定(O) 取消(C)

WIN7-146

Step 3. 建立資料來源 (Create Data Source)

Enterprise Miner - EM_Project1

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

EM Project1
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 2 / 8 步 選取 SAS 表格

EM_LIB.PROFILE

表格(T): [EM_LIB.PROFILE](#) [瀏覽\(R\)...](#)

< 上一步(B) [下一步\(N\) >](#) 取消(C)

已連接到「WIN7-146」



Step 3. 建立資料來源 (Create Data Source)

Enterprise Miner - EM_Project1

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

EM Project1
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.PROFILE
描述	
成員類型	DATA
資料集類型	DATA
引擎	BASE
變數的數目	6
觀測數目	100000
建立日期	2006年9月23日下午06時07分59秒
修改日期	2006年9月23日下午06時07分59秒

< 上一步(B) 下一步(N) > 取消(C)

已連接到「WIN7-146」

Step 3. 建立資料來源 (Create Data Source)

Enterprise Miner - EM_Project1

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

EM Project1
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 4 / 8 步 中繼資料顧問選項

中繼資料顧問選項

使用「基本」設定時，以變數特性為基礎，設定初始量值層級和角色。

使用「進階」設定時，以變數特性和分配為基礎，設定初始量值層級和角色。

基本(S) 進階(A) 自訂(U)...

< 上一步(B) **下一步(N) >** 取消(C)

已連接到「WIN7-146」

Step 3. 建立資料來源 (Create Data Source)

Enterprise Miner - EM_Project1

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

EM Project1
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 欄中變資料

(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
CNT_ATM	輸入	間隔	否		否	.	.
CNT_CSC	輸入	間隔	否		否	.	.
CNT_POS	輸入	間隔	否		否	.	.
CNT_TBM	輸入	間隔	否		否	.	.
CNT_TOT	輸入	間隔	否		否	.	.
ID	ID	名目	否		否	.	.

顯示程式碼(S) 勘查(E) 計算摘要(O) < 上一步(B) 下一步(N) > 取消(C)

已連接到「WIN7-146」 101

Step 3. 建立資料來源 (Create Data Source)

Enterprise Miner - EM_Project1

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

EM Project1
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 欄中變資料

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
CNT_ATM	輸入	間隔	否		否	.	.
CNT_CSC	輸入	間隔	否		否	.	.
CNT_POS	輸入				否	.	.
CNT_TBM	輸入				否	.	.
CNT_TOT	輸入				否	.	.
ID	ID				否	.	.

摘要統計

是否要計算摘要統計?

是(Y) 否(N)

顯示程式碼(S) 勘查(E) 計算摘要(O) < 上一步(B) 下一步(N) > 取消(C)

已連接到「WIN7-146」

Step 3. 建立資料來源 (Create Data Source)

Enterprise Miner - EM_Project1

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

EM Project1
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 欄中總資料

(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	最小值	最大值	平均值	標準差	偏態	峰度
CNT_ATM	3	628	19.49971	20.8561	2.357293	15.33353
CNT_CSC	1	607	6.68411	12.12856	6.236494	108.4633
CNT_POS	2	345	11.9233	20.73384	3.343805	15.43941
CNT_TBM	10	14934	68.13696	101.1542	53.05219	5925.782
CNT_TOT	17	15225	106.2441	113.3704	39.2061	3976.09
ID						

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)

已連接到「WIN7-146」

Step 3. 建立資料來源 (Create Data Source)

Enterprise Miner - EM_Project1

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

EM Project1
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 / 8 步 建立樣本

是否要建立樣本資料集?
 否(O) 是(Y)

表格資訊
欄 6
列 100000

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)

已連接到「WIN7-146」

Step 3. 建立資料來源 (Create Data Source)

Enterprise Miner - EM_Project1

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

EM Project1
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 / 8 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。

名稱(E): PROFILE

角色(R): **原始**

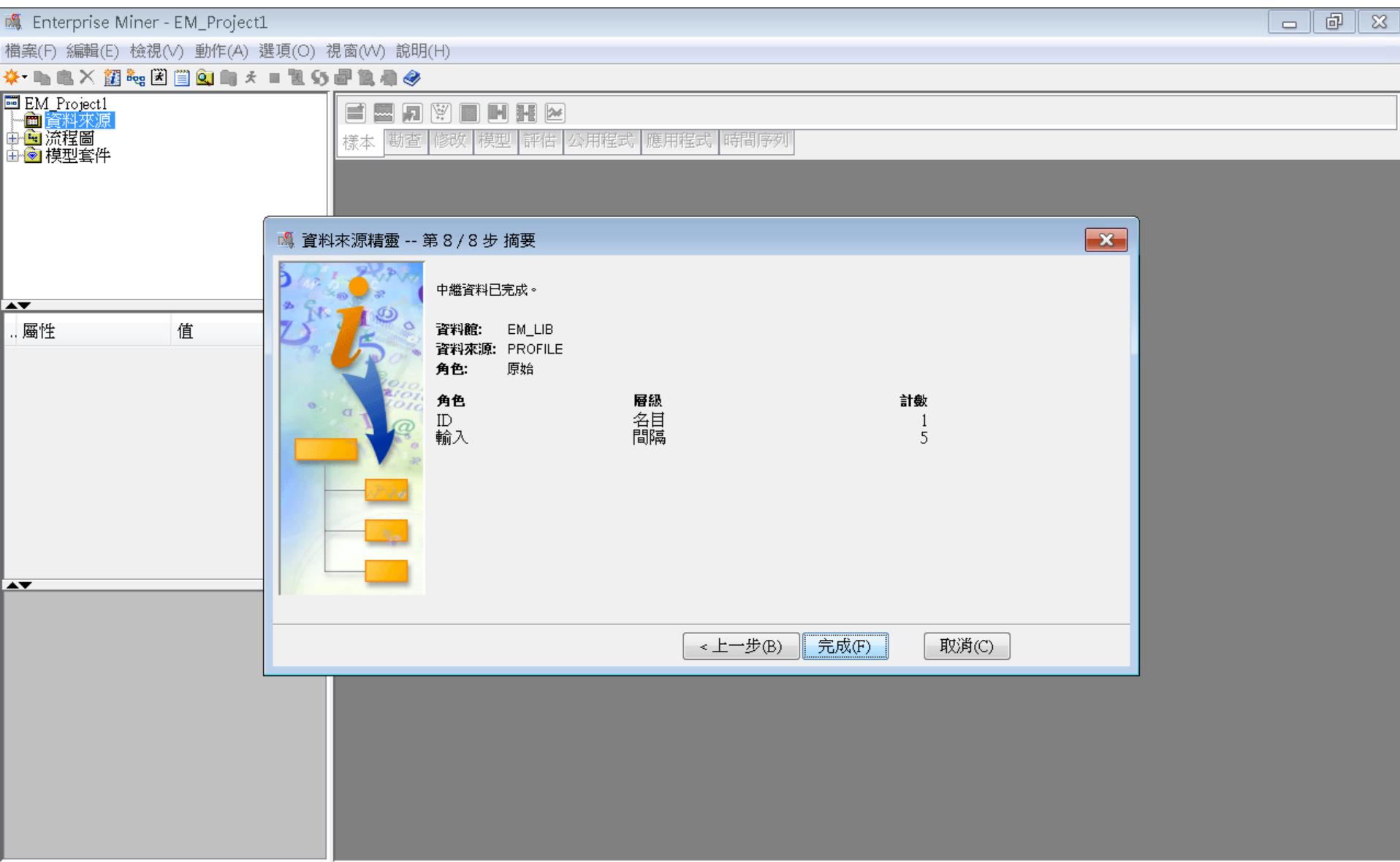
區段(S):

附註(O):

< 上一步(B) 下一步(N) > 取消(C)

已連接到「WIN7-146」

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_Project1

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

EM_Project1
資料來源 PROFILE
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

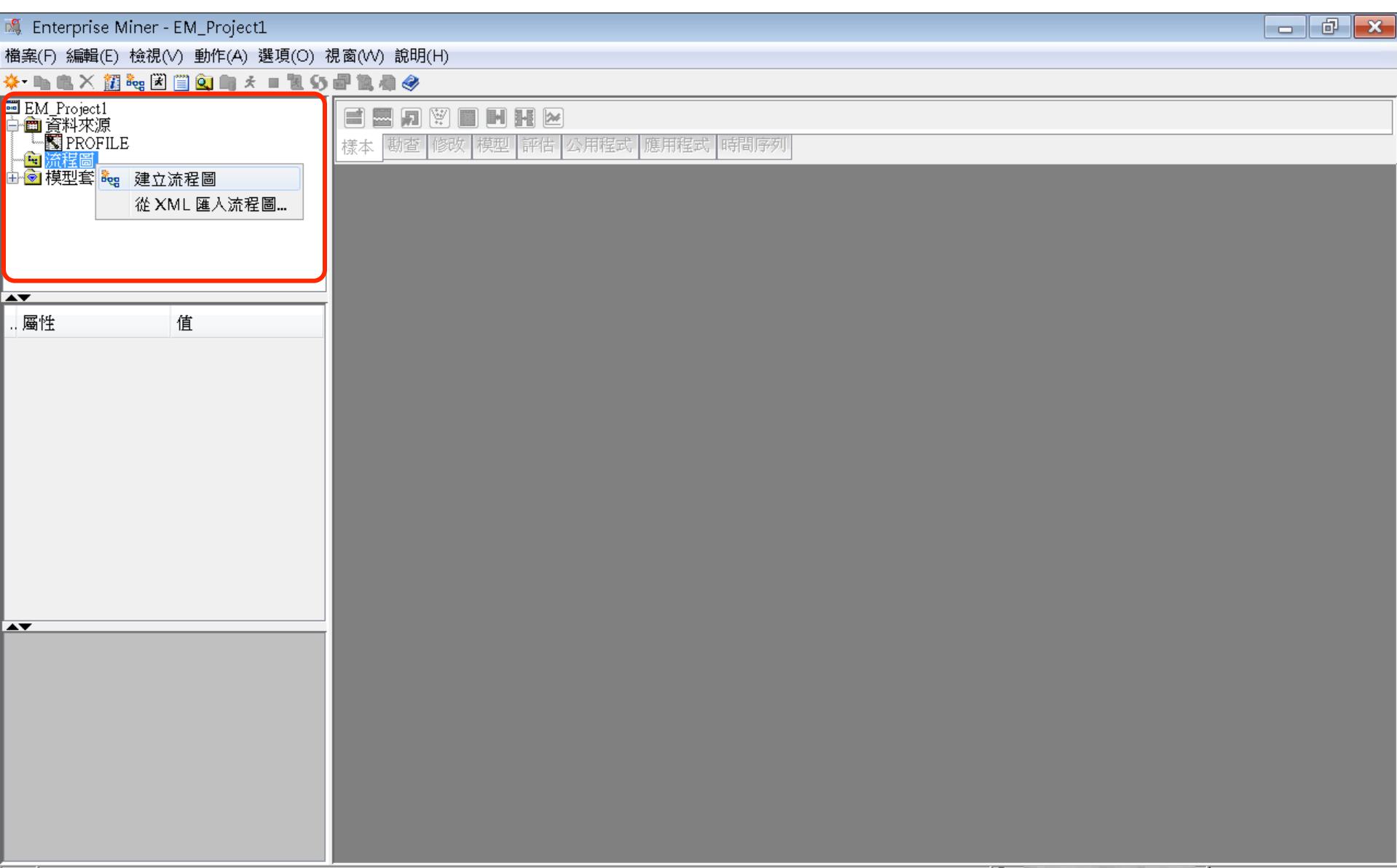
.. 屬性	值
ID	profile
名稱	PROFILE
變數	
決策	
角色	原始
附註	
資料館	EM_LIB
表格	PROFILE
樣本資料集	
大小類型	
樣本大小	
類型	DATA
觀測數目	100000
欄數	6
位元組數	5653504
區段	

ID

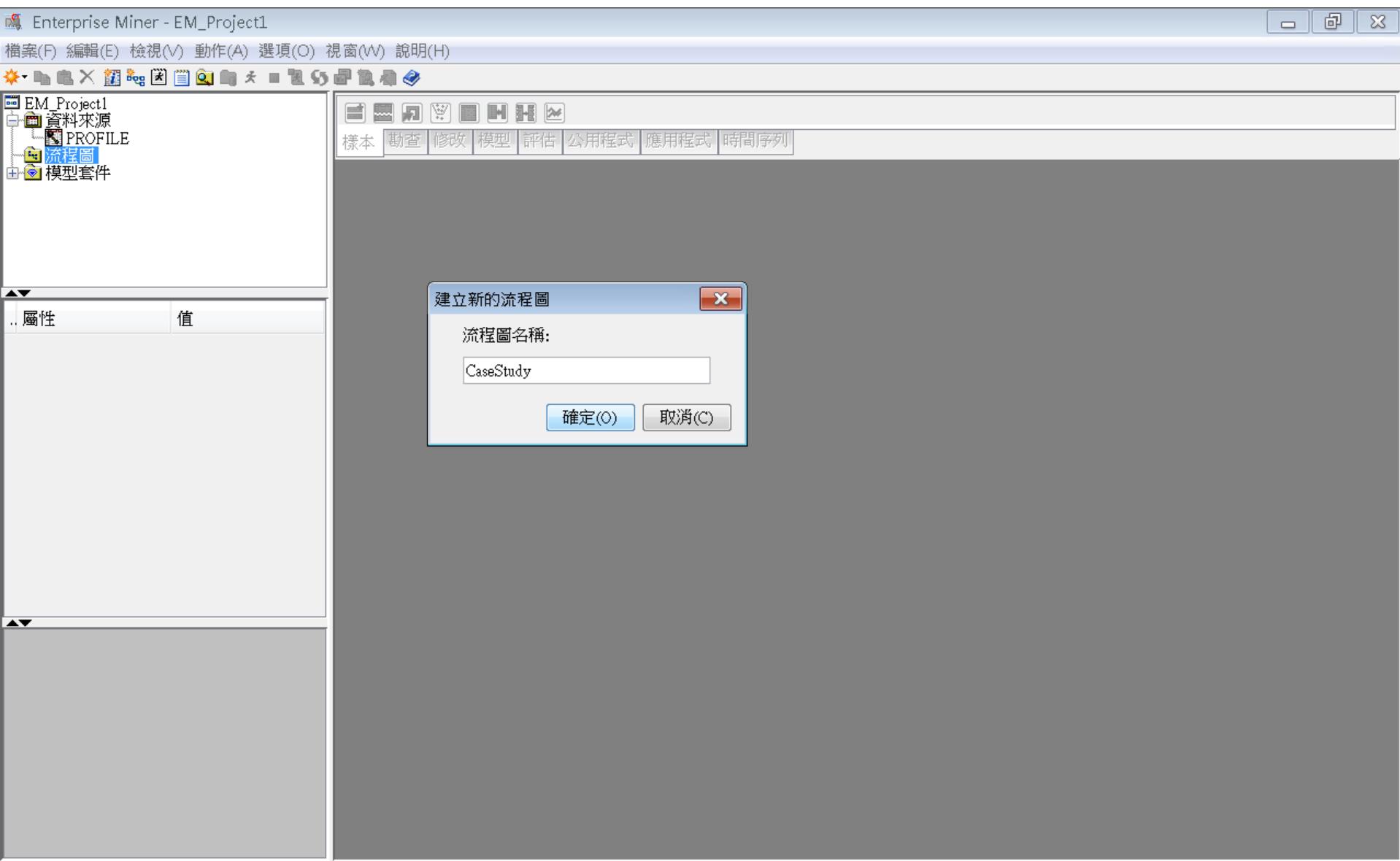
資料來源識別碼。與資料來源關聯的中繼資料表格儲存在 EMDS SAS 資料館中，並使用識別碼作為這些表格名稱的前置碼。

已連接到「WIN7-146」

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)

Enterprise Miner - EM_Project1

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

EM_Project1
資料來源
PROFILE
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

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

ID
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。

流程圖 | 日誌

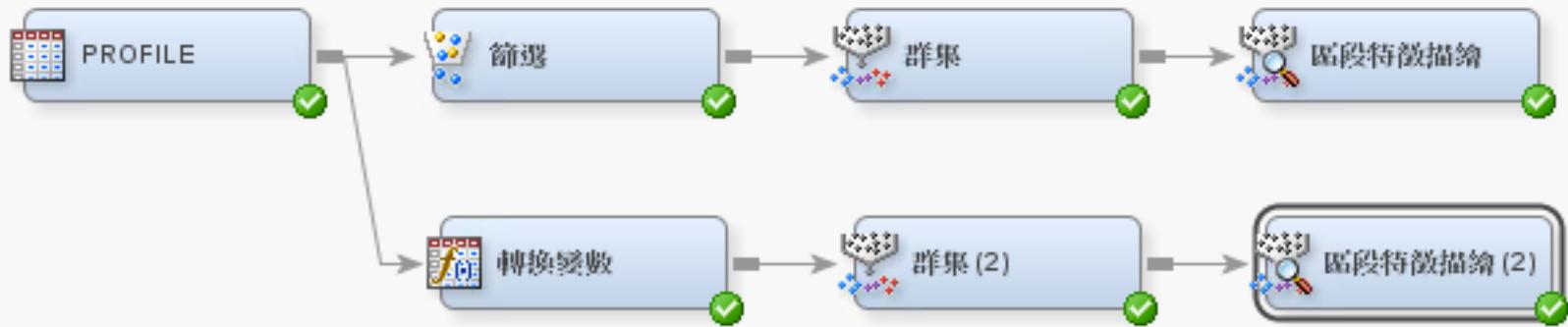
已連接到「WIN7-146」

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



Enterprise Miner - EM_Project1

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

EM_Project1
資料來源
PROFILE
流任務
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

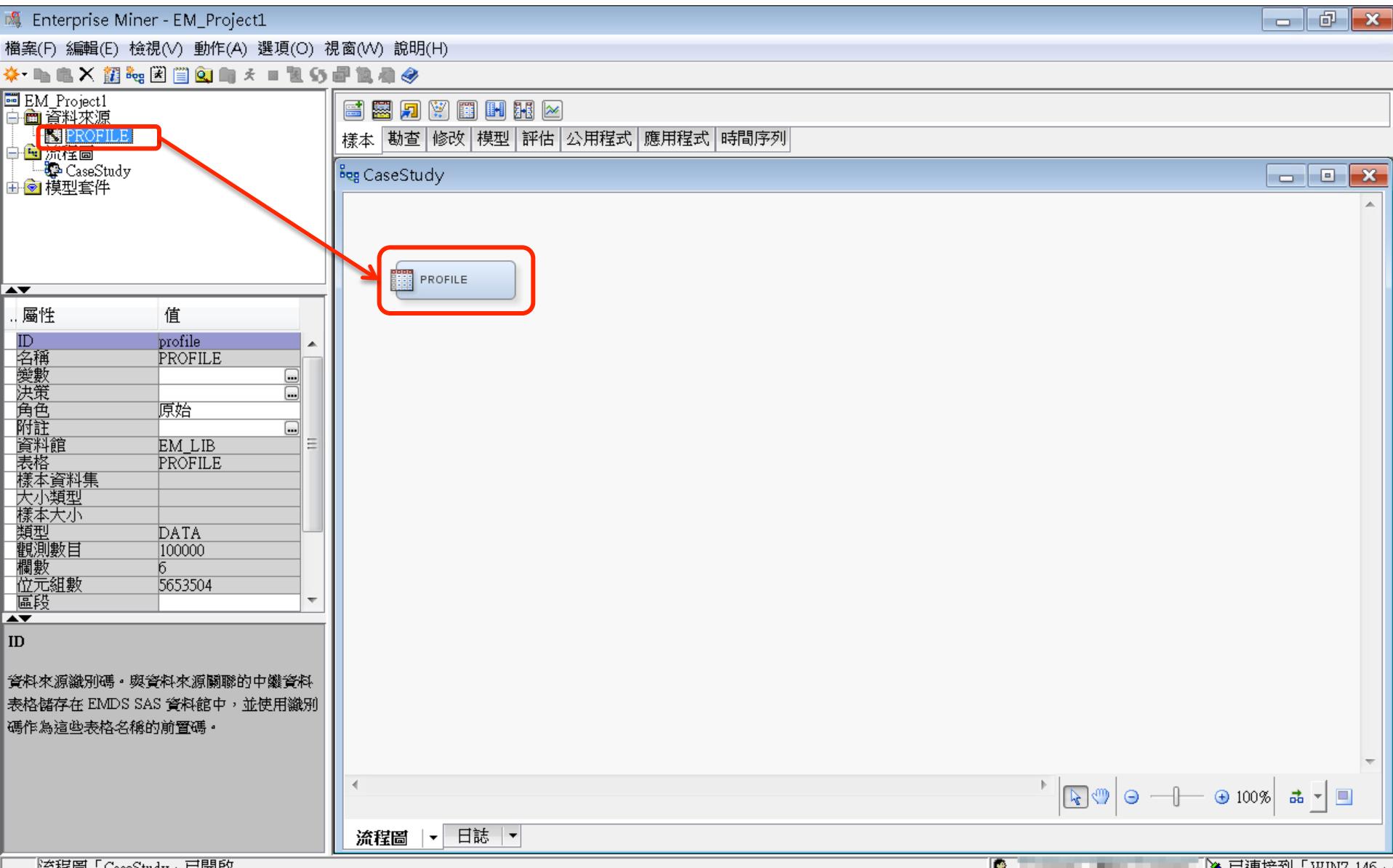
.. 屬性 值

ID	profile
名稱	PROFILE
變數	
決策	
角色	原始
附註	
資料館	EM_LIB
表格	PROFILE
樣本資料集	
大小類型	
樣本大小	
類型	DATA
觀測數目	100000
欄數	6
位元組數	5653504
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。

流程圖 | 日誌



EM_Lib.Profile

	ID	CNT_TBM	CNT_ATM	CNT_POS	CNT_CSC	CNT_TOT
1	000041360	34.0	3.0	3.0	9.0	49.0
2	000052094	44.0	17.0	5.0	18.0	84.0
3	000057340	122.0	26.0	32.0	36.0	216.0
4	000076885	42.0	3.0	6.0	1.0	52.0
5	000089150	20.0	15.0	2.0	2.0	39.0
6	000094512	83.0	20.0	49.0	3.0	155.0
7	000096396	33.0	9.0	49.0	17.0	108.0
8	000098062	22.0	5.0	9.0	1.0	37.0
9	000122779	21.0	27.0	2.0	1.0	51.0
10	000125838	127.0	3.0	2.0	3.0	135.0
11	000137317	52.0	20.0	2.0	11.0	85.0
12	000147896	101.0	40.0	8.0	1.0	150.0
13	000150206	156.0	52.0	80.0	4.0	292.0
14	000160501	31.0	32.0	42.0	12.0	117.0
15	000176820	107.0	3.0	2.0	1.0	113.0
16	000187784	54.0	14.0	7.0	24.0	99.0
17	000189200	92.0	3.0	2.0	2.0	99.0
18	000198327	44.0	27.0	17.0	3.0	91.0
19	000200237	146.0	3.0	2.0	1.0	152.0
20	000207980	39.0	5.0	2.0	4.0	50.0
21	000227355	180.0	3.0	2.0	10.0	195.0
22	000228317	46.0	14.0	2.0	1.0	63.0
23	000228906	25.0	3.0	2.0	1.0	31.0
24	000229984	15.0	3.0	2.0	1.0	21.0
25	000241022	13.0	3.0	2.0	1.0	19.0
26	000253260	65.0	3.0	2.0	1.0	71.0
27	000257910	35.0	3.0	2.0	1.0	41.0
28	000261717	11.0	5.0	2.0	1.0	19.0
29	000266600	15.0	3.0	2.0	1.0	21.0

Enterprise Miner - EM_Project1

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

EM_Project1
資料來源 PROFILE
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
篩選
CaseStudy 根據指定的準則將資料從觀測中移除。

.. 屬性 值

ID	profile
名稱	PROFILE
變數	[...]
決策	[...]
角色	原始
附註	[...]
資料館	EM_LIB
表格	PROFILE
樣本資料集	
大小類型	
樣本大小	
類型	DATA
觀測數目	100000
欄數	6
位元組數	5653504
區段	

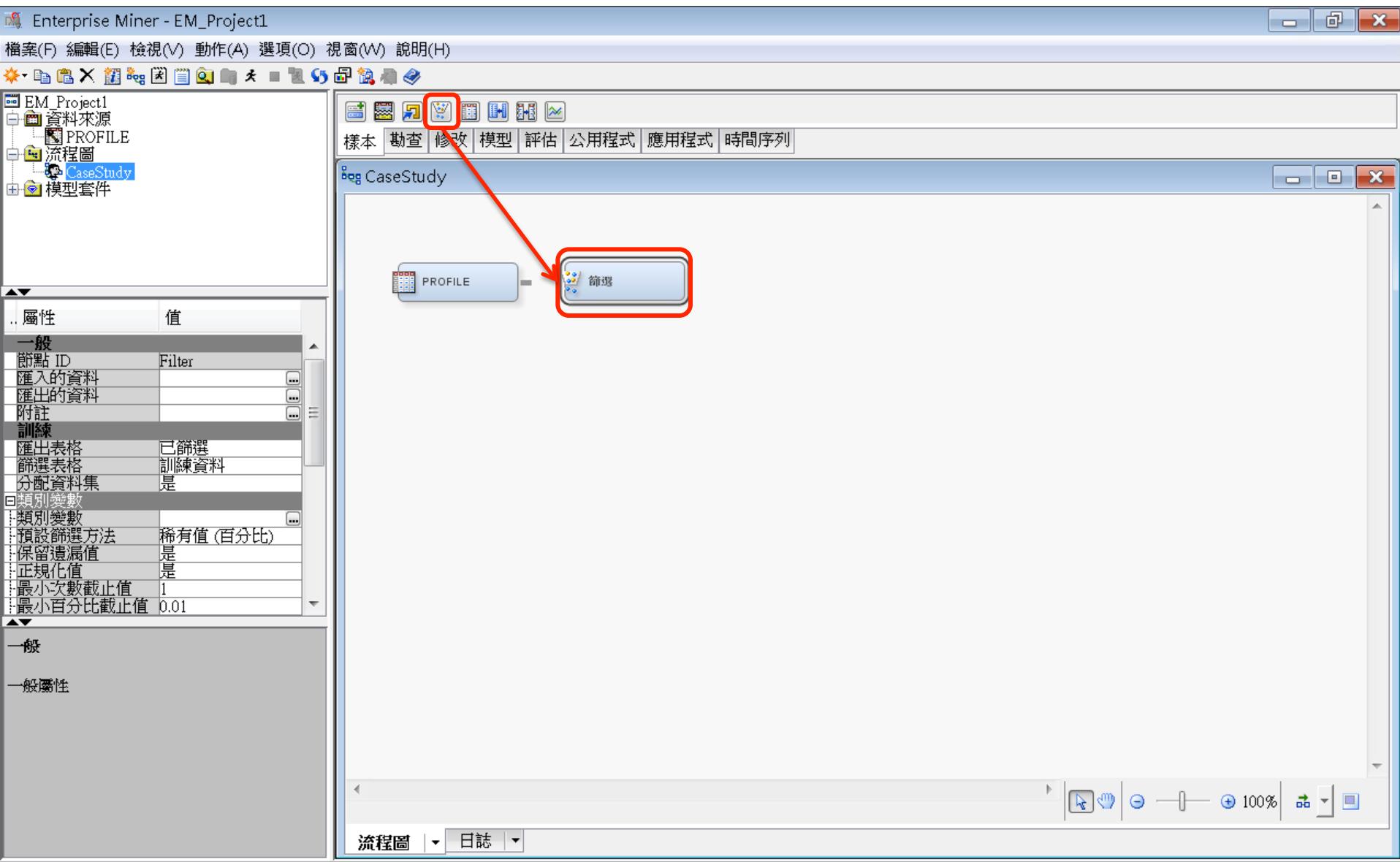
ID

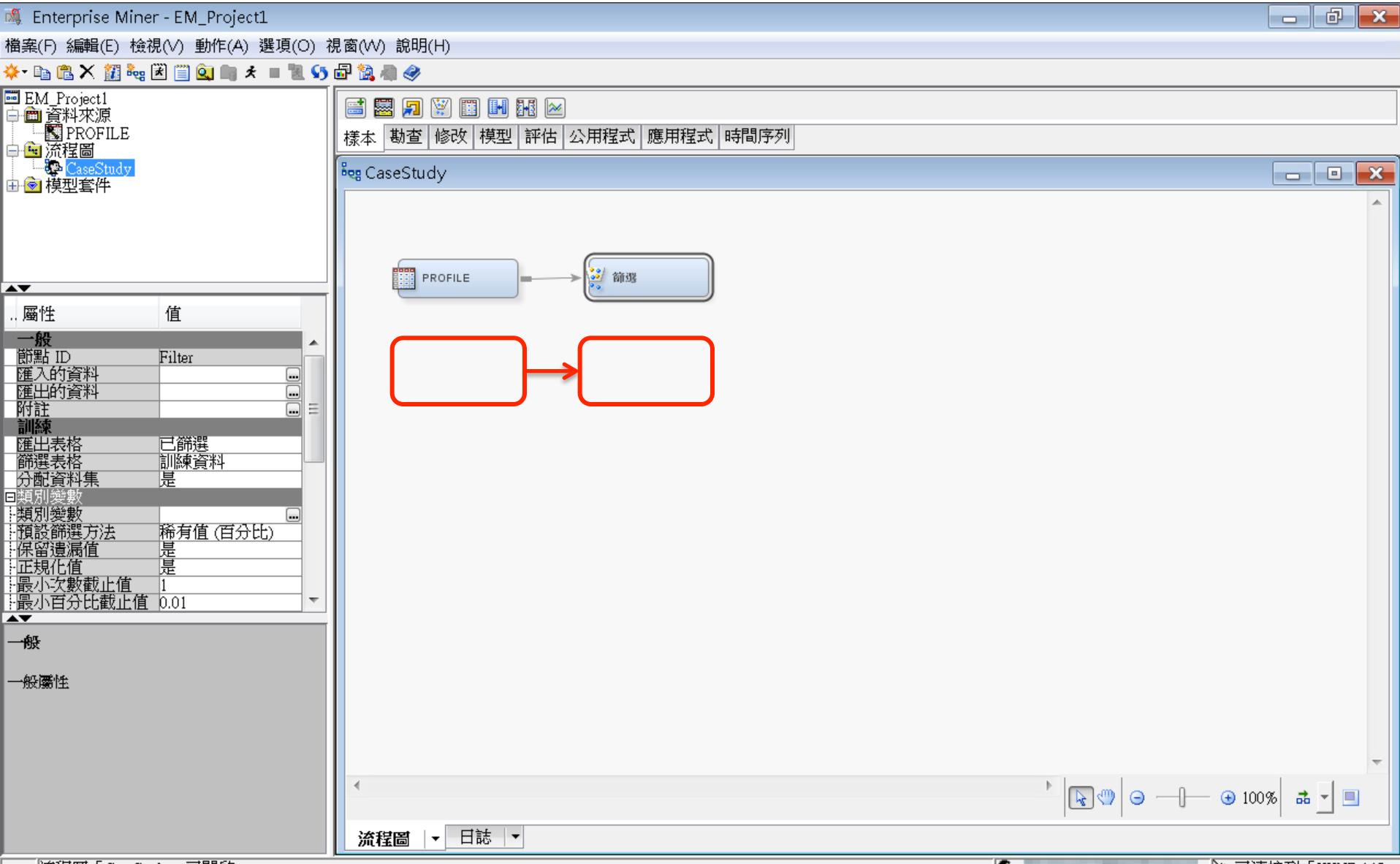
資料來源識別碼。與資料來源關聯的中繼資料表格儲存在 EMDS SAS 資料館中，並使用識別碼作為這些表格名稱的前置碼。

流程圖 | 日誌 |

已連接到「WIN7-146」

The screenshot shows the SAS Enterprise Miner interface with the project 'EM_Project1' open. A context menu is displayed over the 'Delete' icon in the toolbar, with the tooltip '根據指定的準則將資料從觀測中移除。' (Delete according to specified criteria). The left pane shows the project tree with nodes like 'EM_Project1', '資料來源 PROFILE', '流程圖 CaseStudy', and '模型套件'. The right pane displays the properties for the selected 'CaseStudy' node, specifically the 'PROFILE' dataset. The properties table includes fields such as ID, Name, Variables, Decisions, Role, Notes, Library, Table, Sample Dataset, Size Type, Sample Size, Type, Observations, Columns, Bits per Byte, and Segments. The bottom status bar indicates '流程圖「CaseStudy」已開啟' (Diagram 'CaseStudy' is open) and '已連接到「WIN7-146」' (Connected to 'WIN7-146').





Enterprise Miner - EM_Project1

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

EM_Project1
資料來源 PROFILE
流程圖 CaseStudy
模型套件

CaseStudy

.. 屬性 值

一般

節點 ID	Filter
匯入的資料	[...]
匯出的資料	[...]
附註	[...]

訓練

匯出表格	已篩選
節選表格	訓練資料
分配資料集	是

類別變數

類別變數	
預設篩選方法	稀有值(百分比)
保留遺漏值	是
正規化值	是
最小次數截止值	1
最小百分比截止值	0.01
最大層級數目截止值	25

間隔變數

間隔變數	
預設篩選方法	平均值的標準差
保留遺漏值	是
調整參數	

評分

建立評分程式碼	是
重新量值結構	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

PROFILE 篩選

執行

更新

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

流程圖「CaseStudy」已開啟

「」屬於「」 已連接到「」

118

Enterprise Miner - EM_Project1

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

EM_Project1
資料來源
PROFILE
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

PROFILE → 篩選

... 屬性 值

一般

節點 ID	Filter
匯入的資料	...
匯出的資料	...
附註	...
訓練	
匯出表格	已篩選
篩選表格	訓練資料
分配資料集	是
類別變數	
類別變數	預設篩選方法 稀有值(百分比)
保留遺漏值	是
正規化值	是
最小次數截止值	1
最小百分比截止值	0.01
最大層級數目截止值	25
間隔變數	
間隔變數	預設篩選方法 平均值的標準差
保留遺漏值	是
調整參數	...
評分	
建立評分程式碼	是
重新量值圖紙	否

一般

一般屬性

執行完成

執行狀態

執行完成
流程圖: CaseStudy
路徑: 篩選

確定 結果...

流程圖 | 日誌 |

119

篩選—間隔變數 ...

Enterprise Miner - EM_Project1

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

EM_Project1
資料來源
PROFILE
流程圖
CaseStudy
模型套件

.. 屬性 值

一般
節點 ID Filter
匯入的資料
匯出的資料
附註
訓練
匯出表格 已篩選
篩選表格 訓練資料
分配資料集 是
類別變數
類別變數
預設篩選方法 稀有值(百分比)
保留遺漏值 是
正規化值 是
最小次數截止值 1
最小百分比截止值 0.01
最大層級數目截止值 25
間隔變數
間隔變數
預設篩選方法 平均值的標準差
保留遺漏值 是
調整參數
評分
建立評分程式碼 是
更新量值層級 否
狀態

流程圖 檢視

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

CaseStudy

間隔變數
間隔變數
預設篩選方法 平均值的標準差
保留遺漏值 是
調整參數

執行完成

Enterprise Miner - EM_Project1

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

EM Project1
資料來源 PROFILE
流程圖 CaseStudy
模型套件

... 屬性 值
 一般
 篩點 ID Filter
 準入的資料
 準出的資料
 附註
 訓練
 準出表格 已篩選
 篩選表格 訓練資料
 分配資料集 是
 類別變數
 預設篩選方法 稀有值(百分比)
 保留遺漏值 是
 正規化值 是
 最小次數截止值 1
 最小百分比截止值 0.01
 最大層級數目截止值 25
 間隔變數
 間隔變數 平均值的標準差
 預設篩選方法 使用者指定
 保留遺漏值 是
 調整參數
 評分
 建立評分程式碼 是
 更新量值層級 否
 狀態
 一般
 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

CaseStudy

互動式間隔篩選

CNT_ATM

COUNT

VALUE

套用篩選(F) 清除篩選(L)

欄:	<input type="checkbox"/> 標籤(A)	<input type="checkbox"/> 採礦(M)	<input type="checkbox"/> 基本(I)	<input type="checkbox"/> 統計(T)			
名稱	報表	篩選方法	保留遺漏值	篩選下限	篩選上限	角色	層級
CNT_ATM	否	使用者指定	預設	0	100	輸入	間隔
CNT_CSC	否	使用者指定	預設	0	30	輸入	間隔
CNT_POS	否	使用者指定	預設	0	60	輸入	間隔
CNT_TBM	否	使用者指定	預設	0	321	輸入	間隔
CNT_TOT	否	使用者指定	預設	0	441	輸入	間隔

重新整理摘要(R) 確定(O) 取消(C)

篩選—匯入的資料... [勘查]

Enterprise Miner - EM_Project1

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

EM_Project1
資料來源
PROFILE
流程圖
CaseStudy
模型套件

.. 屬性 值

一般
節點 ID Filter
匯入的資料 ...

匯出的資料
附註
訓練
匯出表格 已篩選
篩選表格 訓練資料
分配資料集 是

類別變數
預設篩選方法 稀有值(百分比)
保留遺漏值 是
正規化值 是
最小次數截止值 1
最小百分比截止值 0.01
最大層級數目截止值 25

間隔變數
預設篩選方法 平均值的標準差
保留遺漏值 是
調整參數

評分
建立評分程式碼 是
更新量值層級 否
狀態

匯入的資料
節點匯入的表格集。

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

CaseStudy

PROFILE → 篩選

匯入的資料 - 篩選

連接埠	來源	表格	角色	資料存在
DATA		EMWS1.Ids_DATA	原始	是
VALIDATE				否
TEST				否

瀏覽(B)... 勘查(X)... 屬性(P)... 確定(O)

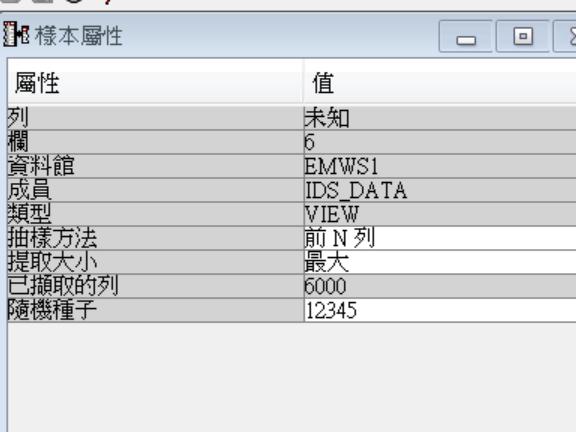
執行完成

100%

篩選 - 匯入的資料

勘查 - EMWS1Ids_DATA

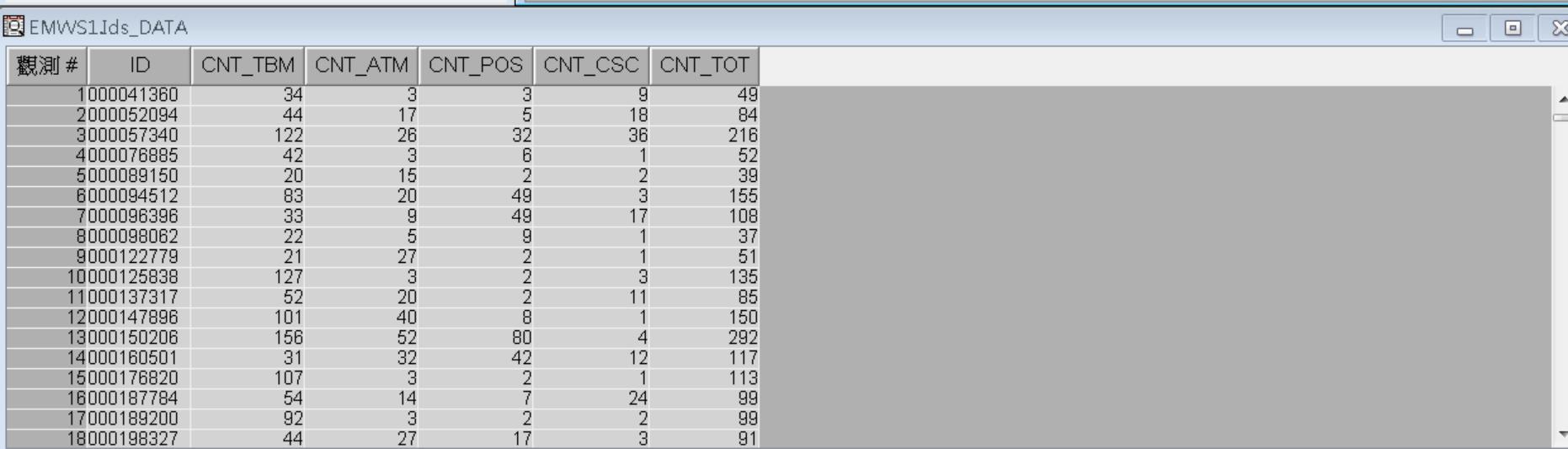
檔案(F) 檢視(V) 動作(A) 視窗(W)

 樣本屬性

屬性	值
列欄	未知
資料館	EMWS1
成員	IDS_DATA
類型	VIEW
抽樣方法	前 N 列
提取大小	最大
已擷取的列	6000
隨機種子	12345

 樣本統計資料

觀測 #	變數名稱	標籤	類型	遺漏值...	最小值	最大值	平均值	層級數目	眾數百...	眾數
1	ID		CLASS	0	.	.	128+	0.775194000041360	.	.
2	CNT_ATM		VAR	0	3	250	19.02667	.	.	.
3	CNT_CSC		VAR	0	1	328	6.6545	.	.	.
4	CNT_POS		VAR	0	2	212	11.83617	.	.	.
5	CNT_TBM		VAR	0	10	2741	68.1285	.	.	.
6	CNT_TOT		VAR	0	17	2747	105.6458	.	.	.

 EMWS1Ids_DATA

觀測 #	ID	CNT_TBM	CNT_ATM	CNT_POS	CNT_CSC	CNT_TOT
1000041360		34	3	3	9	49
2000052094		44	17	5	18	84
3000057340		122	26	32	36	218
4000076885		42	3	6	1	52
5000089150		20	15	2	2	39
6000094512		83	20	49	3	155
7000096396		33	9	49	17	108
8000098062		22	5	9	1	37
9000122779		21	27	2	1	51
10000125838		127	3	2	3	135
11000137317		52	20	2	11	85
12000147896		101	40	8	1	150
13000150206		156	52	80	4	292
14000160501		31	32	42	12	117
15000176820		107	3	2	1	113
16000187784		54	14	7	24	99
17000189200		92	3	2	2	99
18000198327		44	27	17	3	91

篩選 - 匯出的資料 ... [勘查]

Enterprise Miner - EM_Project1

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

EM_Project1
資料來源
PROFILE
流程圖
CaseStudy
模型套件

.. 屬性 值

一般

節點 ID	Filter
匯入的資料	
匯出的資料	
附註	
訓練	已篩選
匯出表格	訓練資料
分配資料集	是
目類別變數	
類別變數	
預設篩選方法	稀有值(百分比)
保留遺漏值	是
正規化值	是
最小次數截止值	1
最小百分比截止值	0.01
最大層級數目截止值	25
間隔變數	
間隔變數	
預設篩選方法	平均值的標準差
保留遺漏值	是
調整參數	
評分	
建立評分程式碼	是
更新量值層級	否
狀態	

匯出的資料

節點匯出的表格集。

流程圖 | 日誌 |

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

CaseStudy

PROFILE 篩選

匯出的資料 - 篩選

連接埠	表格	角色	資料存在
TRAIN	EMWS1.Filter_TRAIN	訓練	是
VALIDATE	EMWS1.Filter_VALIDATE	驗證	否
TEST	EMWS1.Filter_TEST	測試	否

瀏覽(B)... 勘查(O)... 屬性(P)... 確定(O)

執行完成

篩選 - 匯出的資料

勘查 - EMWS1.Filter_TRAIN

檔案(F) 檢視(V) 動作(A) 視窗(W)

樣本屬性

屬性	值
列欄	未知
資料館	EMWS1
成員	FILTER_TRAIN
類型	VIEW
抽樣方法	前 N 列
提取大小	最大
已擷取的列	60000
隨機種子	12345

樣本統計資料

觀測 #	變數名稱	標籤	類型	遺漏值...	最小值	最大值	平均值	層級數目	眾數百...	眾數
1	ID		CLASS	0	0	128+	0.775194	000041360	.	.
2	CNT_ATM		VAR	0	3	82	17.48542	.	.	.
3	CNT_CSC		VAR	0	1	43	5.217783	.	.	.
4	CNT_POS		VAR	0	2	74	9.300717	.	.	.
5	CNT_TBM		VAR	0	10	371	62.84018	.	.	.
6	CNT_TOT		VAR	0	17	445	94.8441	.	.	.

EMWS1.Filter_TRAIN

觀測 #	ID	CNT_TBM	CNT_ATM	CNT_POS	CNT_CSC	CNT_TOT
1000041360		34	3	3	9	49
2000052094		44	17	5	18	84
3000057340		122	26	32	36	218
4000076885		42	3	6	1	52
5000089150		20	15	2	2	39
6000094512		83	20	49	3	155
7000096396		33	9	49	17	108
8000098062		22	5	9	1	37
9000122779		21	27	2	1	51
10000125838		127	3	2	3	135
11000137317		52	20	2	11	85
12000147896		101	40	8	1	150
13000160501		31	32	42	12	117
14000176820		107	3	2	1	113
15000187784		54	14	7	24	99
16000189200		92	3	2	2	99
17000198327		44	27	17	3	91
18000200237		146	3	2	1	152

勘查 (Explore) — 群集 (Cluster)

Enterprise Miner - EM_Project1

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

EM_Project1
資料來源 PROFILE
流程圖 CaseStudy
模型套件

.. 屬性 值

一般

節點 ID	Filter
匯入的資料	
匯出的資料	
附註	
訓練	
匯出表格	已篩選
篩選表格	訓練資料
分配資料集	是
類別變數	
類別變數	
預設篩選方法	稀有值(百分比)
保留遺漏值	是
正規化值	是
最小次數截止值	1
最小百分比截止值	0.01
最大層級數目截止值	25
間隔變數	
間隔變數	
預設篩選方法	平均值的標準差
保留遺漏值	是
調整參數	
評分	
建立評分程式碼	是
更新量值層級	否
狀態	

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 時間序列

群集

執行分段資料庫可以使用的觀測群集。

PROFILE → 篩選

流程圖 | 日誌 |

126

Enterprise Miner - EM_Project1

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

EM_Project1
資料來源
PROFILE
流程圖
CaseStudy
模型套件

.. 屬性 值
一般
節點 ID Clus
匯入的資料
匯出的資料
附註
訓練
變數
群集變數角色 區段
內部標準化 標準化
群集的數目
指定方法 自動
最大的群集數目 10
選取準則
群集方法 Ward
初步最大值 50
最小值 2
最終最大值 20
CCC 截止值 3
類別變數編碼
序數編碼 排名
名目編碼 GLM
初始群集種子
種子初始化方法 預設
最小半徑 0.0
在訓練期間漂移 否
訓練選項
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

PROFILE → 篩選 → 群集

群集 (右鍵)
執行

流程圖 | 日誌 |

The screenshot shows the SAS Enterprise Miner interface. On the left is a tree view of the project structure. In the center, a flow diagram titled 'CaseStudy' shows three nodes: 'PROFILE', '篩選' (Filter), and '群集' (Cluster). A red arrow points from the '樣本' (Sample) button in the toolbar to the '群集' node. A context menu is open over the '群集' node, with the '執行' (Execute) option highlighted by a red box. The '執行' option is part of a larger menu that includes '更新' (Update), '建立模型套件...' (Create Model Script...), '結果...' (Results...), '將路徑匯出作為 SAS 程式' (Export Path as SAS Program), '剪下' (Cut), '複製(C)' (Copy), '刪除' (Delete), '重新命名' (Rename), '全選' (Select All), '選取節點' (Select Node), '連接節點' (Connect Node), and '中斷連接節點' (Break Connection Node).

Enterprise Miner - EM_Project1

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

EM_Project1

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

CaseStudy

... 屬性 值

一般

節點 ID	Clus
匯入的資料	[...]
匯出的資料	[...]
附註	[...]

訓練

變數	[...]
群集變數角色	區段
內部標準化	標準化
自群集的數目	[...]
指定方法	自動
最大的群集數目	10
選取準則	[...]
群集方法	Ward
初步最大值	50
最小值	2
最終最大值	20
CCC 截止值	3
類別變數編碼	[...]
序數編碼	排名
名目編碼	GLM
初始群集種子	[...]
種子初始化方法	預設
最小半徑	0.0
在訓練期間漂移	否
訓練選項	[...]

一般

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

PROFILE → 篩選 → 群集

確認

是否要執行此路徑?
流程圖:CaseStudy
路徑: 群集

是(Y) 否(N)

流程圖 | 日誌

Enterprise Miner - EM_Project1

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

EM_Project1
資料來源
PROFILE
流程圖
CaseStudy
模型套件

.. 屬性 值

一般

節點 ID	Clus
匯入的資料	[...]
匯出的資料	[...]
附註	[...]
訓練	
變數	[...]
群集變數角色	區段
內部標準化	標準化
群集的數目	[...]
指定方法	自動
最大的群集數目	10
選取準則	[...]
群集方法	Ward
初步最大值	50
最小值	2
最終最大值	20
CCC 截止值	3
類別變數編碼	[...]
序數編碼	排名
名目編碼	GLM
初始群集種子	[...]
種子初始化方法	預設
最小半徑	0.0
在訓練期間漂移	否
訓練選項	[...]

流程圖 | 日誌 |

正在執行「群集」

Enterprise Miner - EM_Project1

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

EM_Project1
資料來源
PROFILE
流程圖
CaseStudy
模型套件

.. 屬性 值

一般

節點 ID Clus

匯入的資料

匯出的資料

附註

訓練

變數

群集變數角色 區段

內部標準化 標準化

群集的數目

指定方法 自動

最大的群集數目 10

擷取準則

群集方法 Ward

初步最大值 50

最小值 2

最終最大值 20

CCC 截止值 3

類別變數編碼

序數編碼 排名

名目編碼 GLM

初始群集種子

種子初始化方法 預設

最小半徑 0.0

在訓練期間漂移 否

訓練選項

一般

一般屬性

流程圖 | 日誌 |

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

CaseStudy

PROFILE → 篩選 → 群集

執行狀態

執行完成
流程圖: CaseStudy
路徑: 群集

確定 結果...

The screenshot shows the SAS Enterprise Miner interface with a project named 'EM_Project1'. On the left, there's a tree view of the project structure. A properties panel on the left displays settings for a 'Cluster' node, including the number of clusters (10), method (Ward), and various cutoff values. The main workspace shows a flow diagram with three nodes: 'PROFILE', '篩選' (Filter), and '群集' (Cluster). A results dialog box is open, indicating the execution completed successfully for the 'CaseStudy' flowchart, specifically for the 'Cluster' path. The 'Results...' button in the dialog is highlighted with a red rectangle.

群集 (Cluster) 結果

結果 - 節點: 群集 流程圖: CaseStudy

檔案(F) 編輯(E) 檢視(V) 視窗(W)

區段圖

變數 = CNT_ATM	變數 = CNT_CSC	變數 = CNT_POS
Segment Variable 1: 15.125, 87.875, 22.75, 38.25, 165.5, 165.5	Segment Variable 2: 15.125, 87.875, 22.75, 38.25, 165.5, 165.5	Segment Variable 3: 15.125, 87.875, 22.75, 38.25, 165.5, 165.5

平均值統計

Clusterin g Criterion	Maximum Relative Change in Cluster Seeds	Improve ment in Clusterin g Criterion	區段 ID	Frequenc y of Cluster	Root-Mea n-Square Standard Deviation	Maximum Distance from Cluster Seed	Nearest Cluster	Di to Ne Cli
0.72662	0.003149	.	1	56275	0.490829	4.80498	2	2
0.72662	0.003149	.	2	16181	0.854974	7.003775	3	2
0.72662	0.003149	.	3	18638	1.103873	7.11535	2	2

區段大小

輸出

```

5   *
6   * 訓練輸出
7   *
8
9
10
11
12 變數摘要
13
14 角色    量值層級    次數計數
15
16 ID      NOMINAL      1
17 INPUT    INTERVAL      5
18
19
20

```

評估 (Assess) — 區段特徵描繪(Segment Profile)

Enterprise Miner - EM_Project1

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

EM_Project1
資料來源
PROFILE
流程圖
CaseStudy
模型套件

.. 屬性 值

一般

節點 ID	Clus
匯入的資料	
匯出的資料	
附註	
訓練	
變數	
群集變數角色	區段
內部標準化	標準化
群集的數目	
指定方法	自動
最大的群集數目	10
選取準則	
群集方法	Ward
初步最大值	50
最小值	2
最終最大值	20
CCC 截止值	3
類別變數編碼	
序數編碼	排名
名目編碼	GLM
初始群集種子	
種子初始化方法	預設
最小半徑	0.0
在訓練期間漂移	否
訓練選項	

流程圖 | 日誌 |

樣本 勘查 修改 模型 **評估**

區段特徵描繪 (Segment Profile)

說明

使用「區段特徵描繪」節點檢查分段或群集的資料，並識別母體中資料區段不同的因子。

```
graph LR; PROFILE[PROFILE] --> FILTER[篩選]; FILTER --> CLUSTER[群集]
```

執行完成

評估 (Assess) – 區段特徵描繪 (Segment Profile)

Enterprise Miner - EM_Project1

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

EM_Project1
資料來源
PROFILE
流程圖
CaseStudy
模型套件

... 屬性 值

一般
節點 ID Prof
匯入的資料
匯出的資料
附註
訓練
變數
一般
中間點數目 8
所有的特徵描繪 否
截止百分比 95
輸入變數
輸入的數目 10
最小價值 0.01
最大深度 1
列印價值統計 是
目標變數
分析角色 無
報表變數
使用報表變數 是
報表變數的數目 10
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

The screenshot shows the Microsoft SQL Server Data Mining Enterprise Miner interface. The main window displays a workflow titled 'CaseStudy' with the following sequence of nodes:

```
graph LR; A[PROFILE] --> B[篩選]; B --> C[群集]; C --> D[區段特徵描繪];
```

A red arrow points from the '評估 (Assess)' tab in the ribbon to the '區段特徵描繪' (Segment Profile) node in the workflow. A red box highlights this node. On the left, the properties pane for the 'Segment Profile' node is open, showing settings like '節點 ID' (Node ID) set to 'Prof'. The bottom status bar indicates '執行完成' (Execution completed).

評估 (Assess) – 區段特徵描繪(Segment Profile)

Enterprise Miner - EM_Project1

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

EM_Project1
資料來源
PROFILE
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

The screenshot shows the SAS Enterprise Miner interface with the following details:

- Workflow Diagram:** A sequence of nodes: PROFILE → 篩選 (Filter) → 群聚 (Cluster) → **區段特徵描繪 (Segment Profile)**. The last node is highlighted with a red box.
- Properties Panel (Left):**

屬性	值
一般	
節點 ID	Prof
匯入的資料	
匯出的資料	
附註	
訓練	
變數	
一般	
中間點數目	8
所有的特徵描繪	否
截止百分比	95
輸入變數	
輸入的數目	10
最小價值	0.01
最大深度	1
列印價值統計	是
目標變數	
分析角色	無
報表變數	
使用報表變數	是
報表變數的數目	10
狀態	
- Properties Panel (Right):**
 - 編輯變數...
 - 執行** (highlighted with a red box)
 - 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
- Toolbars:** Standard toolbar icons (File, Edit, View, etc.)
- Bottom Bar:** 流程圖 (Diagram), 日誌 (Log), 100% zoom, and other interface controls.

評估 (Assess) – 區段特徵描繪 (Segment Profile)

Enterprise Miner - EM_Project1

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

EM_Project1
資料來源
PROFILE
流程圖
CaseStudy
模型套件

.. 屬性 值

一般

節點 ID Prof

匯入的資料

匯出的資料

附註

訓練

變數

一般

中間點數目 8

所有的特徵描繪 否

截止百分比 95

輸入變數

輸入的數目 10

最小價值 0.01

最大深度 1

列印價值統計 是

目標變數

分析角色 無

報表變數

使用報表變數 是

報表變數的數目 10

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

```
graph LR; PROFILE[PROFILE] --> FILTRE[Filtration]; FILTRE --> CLUSTER[Clustering]; CLUSTER --> PROFILE_PROFILE[Segment Profile]
```

正在執行「區段特徵描繪」

評估 (Assess) – 區段特徵描繪(Segment Profile)

Enterprise Miner - EM_Project1

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

EM_Project1
資料來源
PROFILE
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

.. 屬性 值

一般

節點 ID	Prof
匯入的資料	[...]
匯出的資料	[...]
附註	
訓練	
變數	[...]
一般	
中間點數目	8
所有的特徵描繪	否
截止百分比	95
輸入變數	
輸入的數目	10
最小價值	0.01
最大深度	1
列印價值統計	是
目標變數	
分析角色	無
報表變數	
使用報表變數	是
報表變數的數目	10
狀態	

一般

一般屬性

執行完成

執行狀態

執行完成
流程圖:CaseStudy
路徑: 區段特徵描繪

確定 結果...

流程圖 日誌

```
graph LR; PROFILE[PROFILE] --> FILTER[篩選]; FILTER --> CLUSTER[群聚]; CLUSTER --> PROFILE[Segment Profile]
```

區段特徵描繪 (Segment Profile)

結果 - 節點: 區段特徵描繪 流程圖: CaseStudy

檔案(F) 編輯(E) 檢視(V) 視窗(W)

區段大小: _SEGMENT_

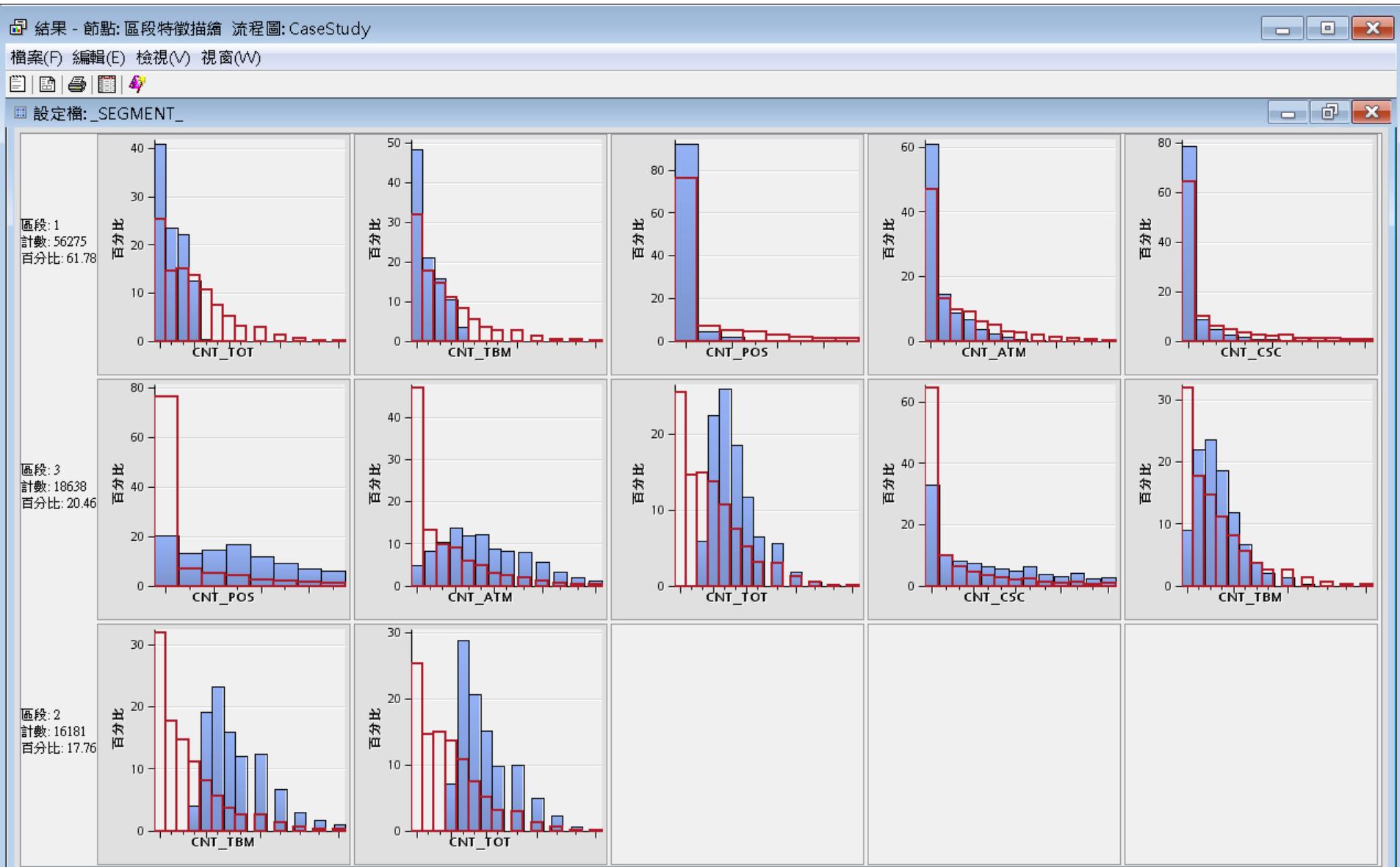
設定檔: _SEGMENT_

變數價值: _SEGMENT_

輸出

```
5 *
6 * 訓練輸出
7 *
8
9
10
11
12 變數摘要
13
14 角色      量值層級    次數計數
15
16 ID        NOMINAL     1
17 INPUT     INTERVAL     5
18 REJECTED  INTERVAL     1
19 REJECTED  NOMINAL     1
20 SEGMENT   NOMINAL     1
21
```

區段特徵描繪 (Segment Profile)



References

- Efraim Turban, Ramesh Sharda, Dursun Delen,
Decision Support and Business Intelligence Systems,
Ninth Edition, 2011, Pearson.
- Jiawei Han and Micheline Kamber,
Data Mining: Concepts and Techniques, Second Edition,
2006, Elsevier
- 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