



商業智慧實務

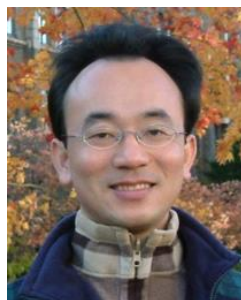
Practices of Business Intelligence

商業智慧的資料探勘 (Data Mining for Business Intelligence)

1032BI05

MI4

Wed, 9,10 (16:10-18:00) (B130)



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/>

2015-03-25



課程大綱 (Syllabus)

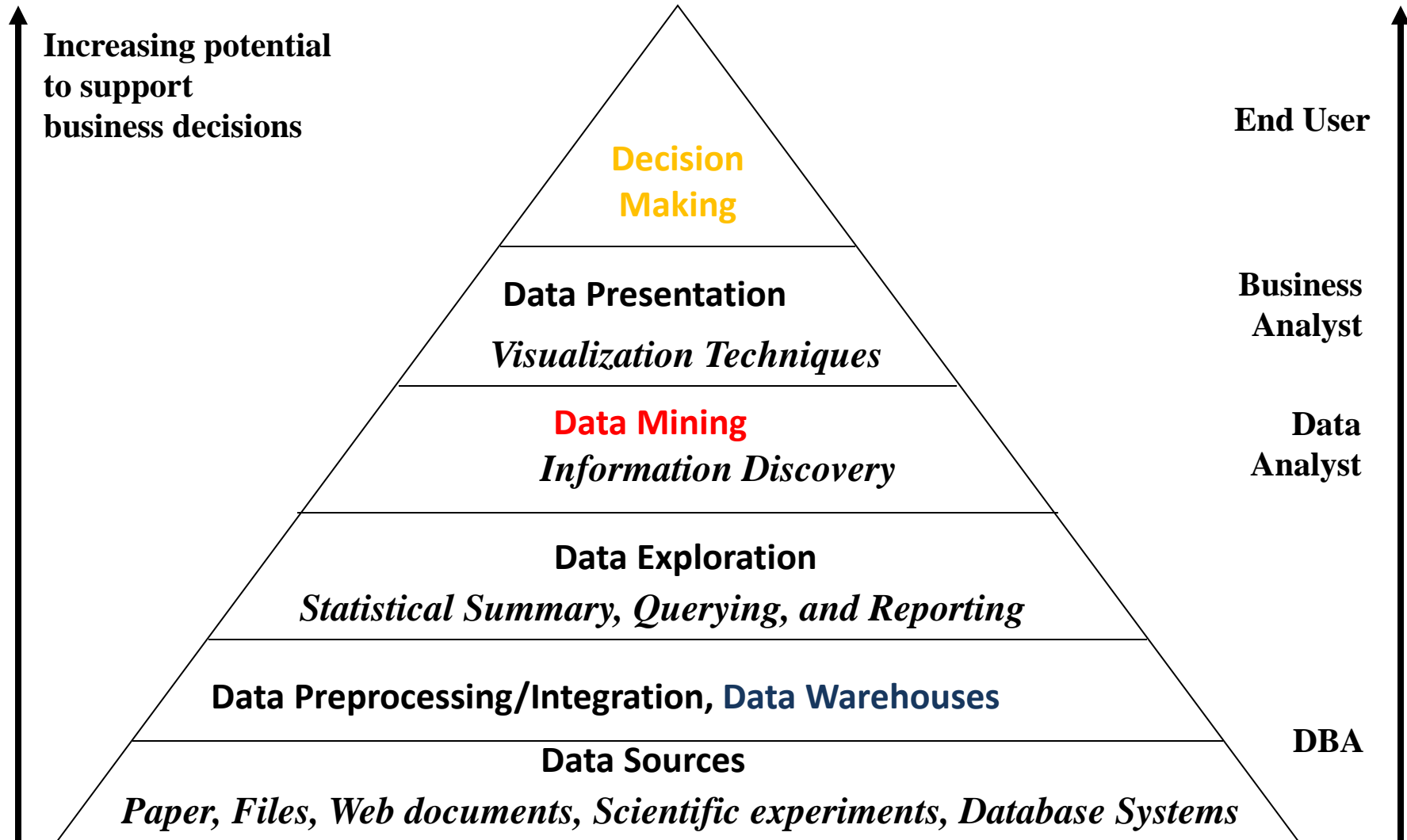
週次 (Week)	日期 (Date)	內容 (Subject/Topics)
1	2015/02/25	商業智慧導論 (Introduction to Business Intelligence)
2	2015/03/04	管理決策支援系統與商業智慧 (Management Decision Support System and Business Intelligence)
3	2015/03/11	企業績效管理 (Business Performance Management)
4	2015/03/18	資料倉儲 (Data Warehousing)
5	2015/03/25	商業智慧的資料探勘 (Data Mining for Business Intelligence)
6	2015/04/01	教學行政觀摩日 (Off-campus study)
7	2015/04/08	商業智慧的資料探勘 (Data Mining for Business Intelligence)
8	2015/04/15	資料科學與巨量資料分析 (Data Science and Big Data Analytics)

課程大綱 (Syllabus)

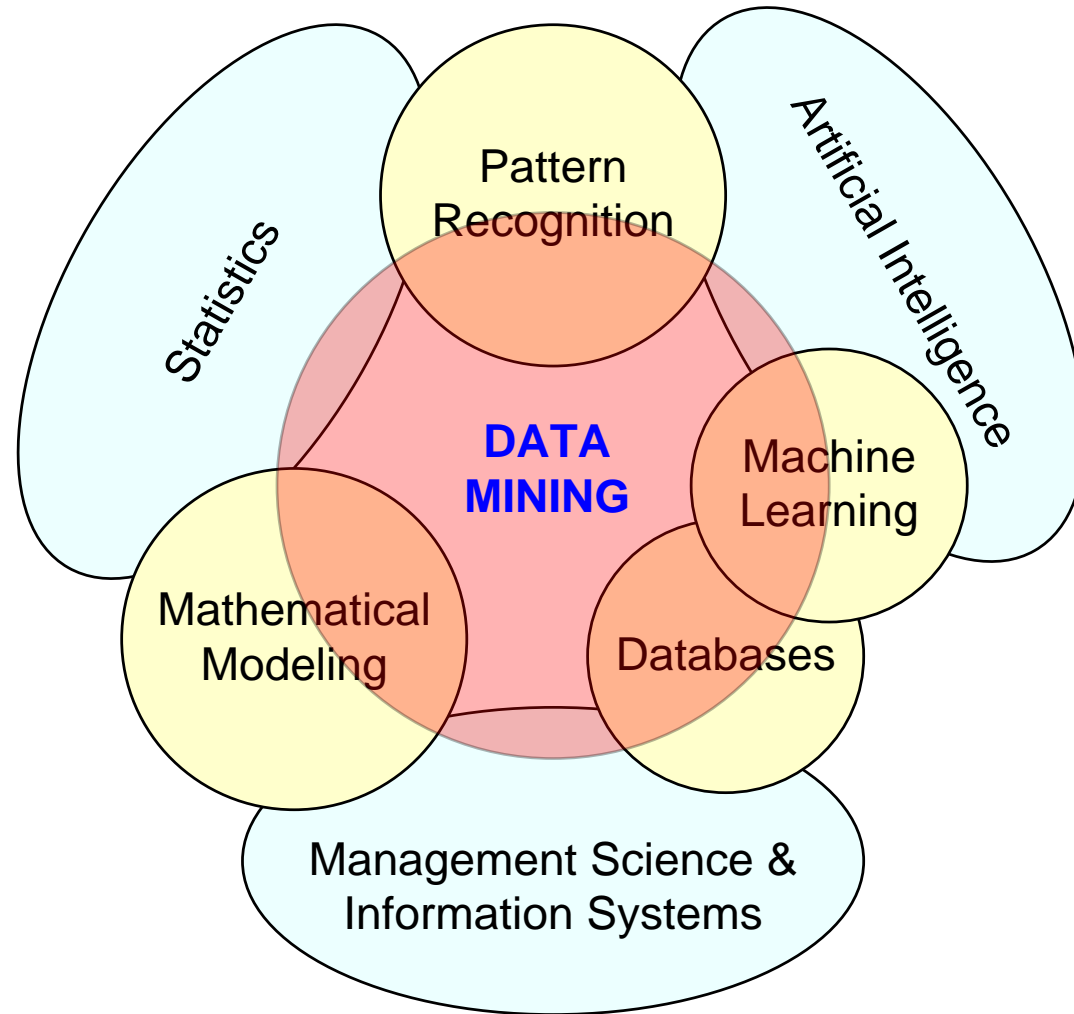
週次	日期	內容 (Subject/Topics)
9	2015/04/22	期中報告 (Midterm Project Presentation)
10	2015/04/29	期中考試週 (Midterm Exam)
11	2015/05/06	文字探勘與網路探勘 (Text and Web Mining)
12	2015/05/13	意見探勘與情感分析 (Opinion Mining and Sentiment Analysis)
13	2015/05/20	社會網路分析 (Social Network Analysis)
14	2015/05/27	期末報告 (Final Project Presentation)
15	2015/06/03	畢業考試週 (Final Exam)

Business Intelligence

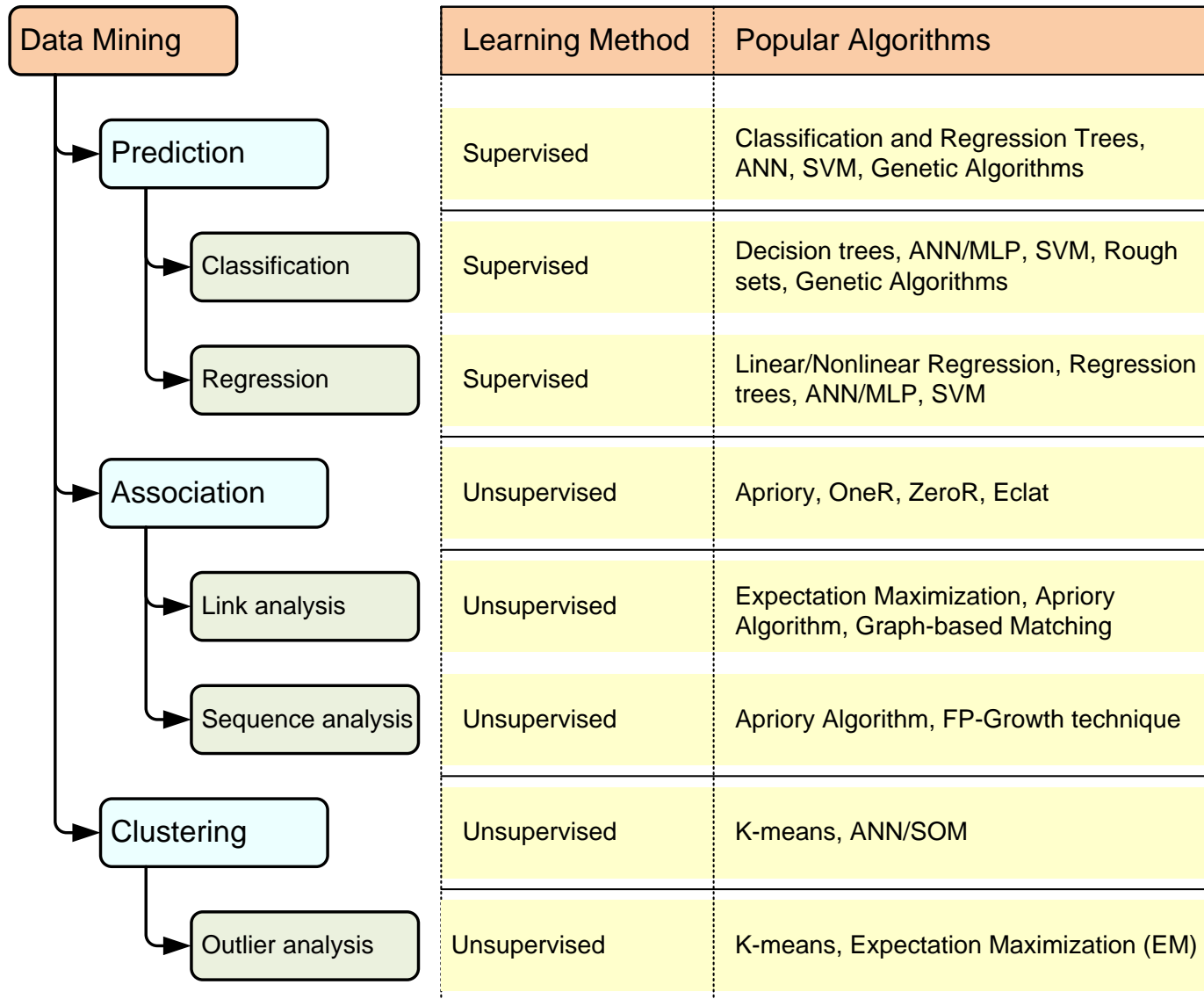
Data Mining, Data Warehouses



Data Mining at the Intersection of Many Disciplines

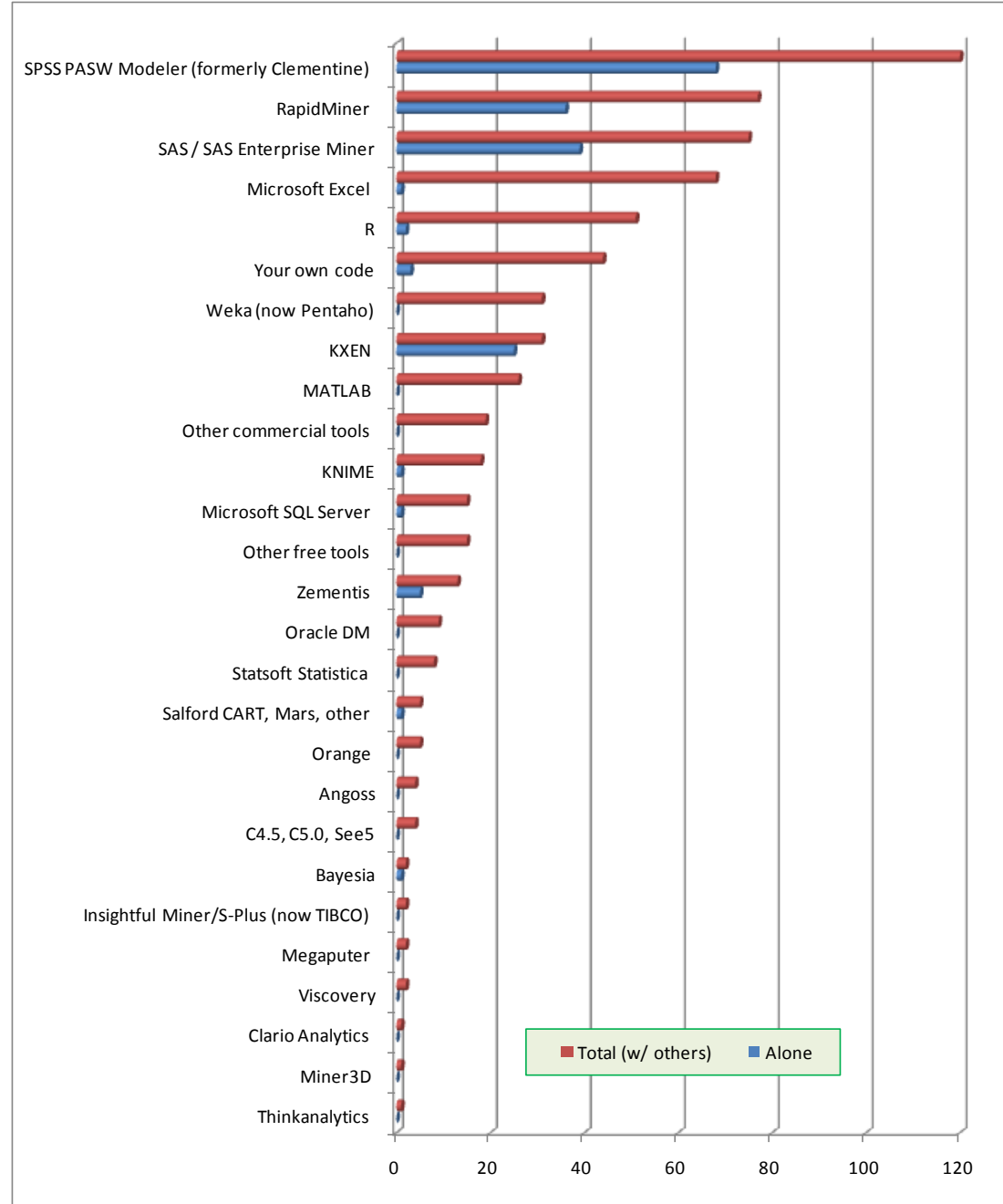


A Taxonomy for Data Mining Tasks



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...



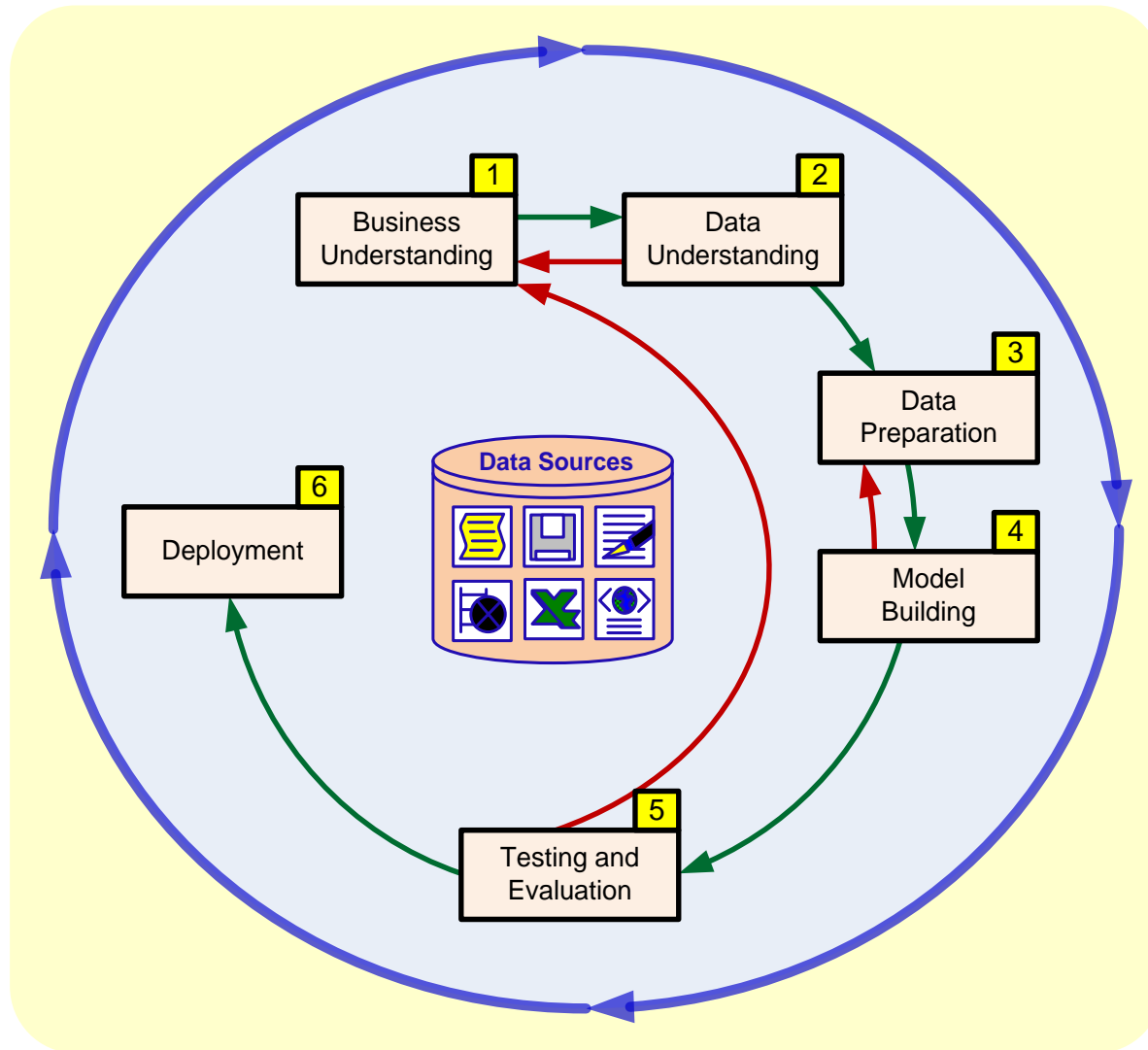
Source: KDNuggets.com, May 2009

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

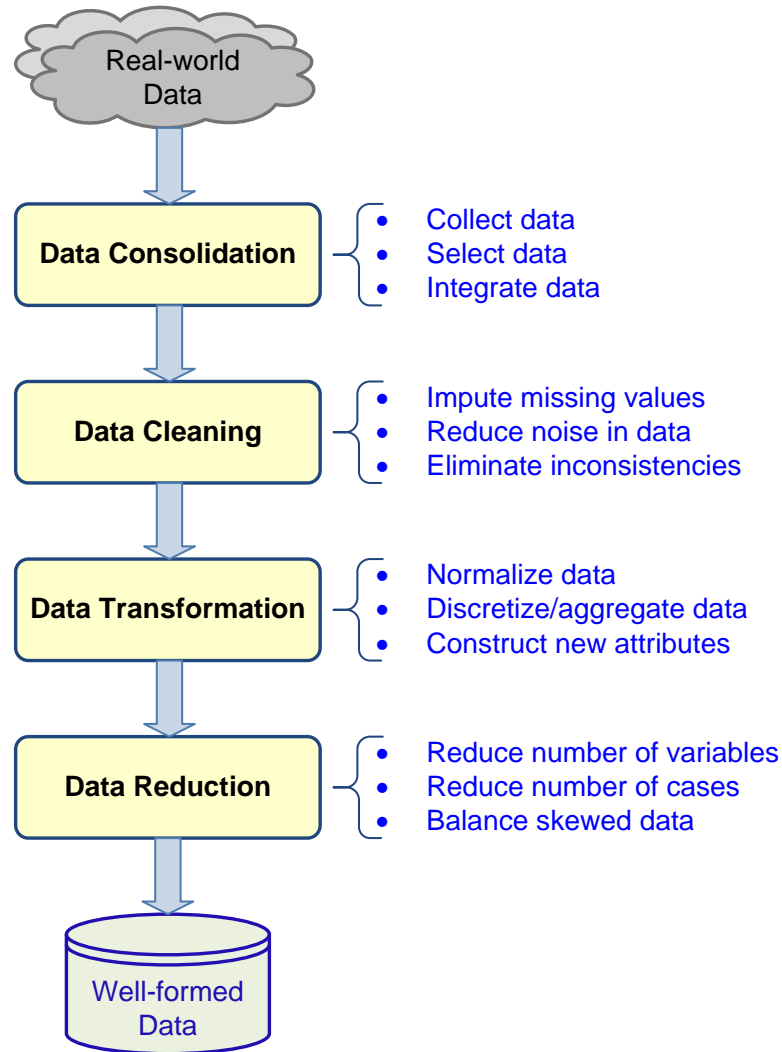
Step 5: Testing and Evaluation

Step 6: Deployment

Accounts for
~85% of total
project time

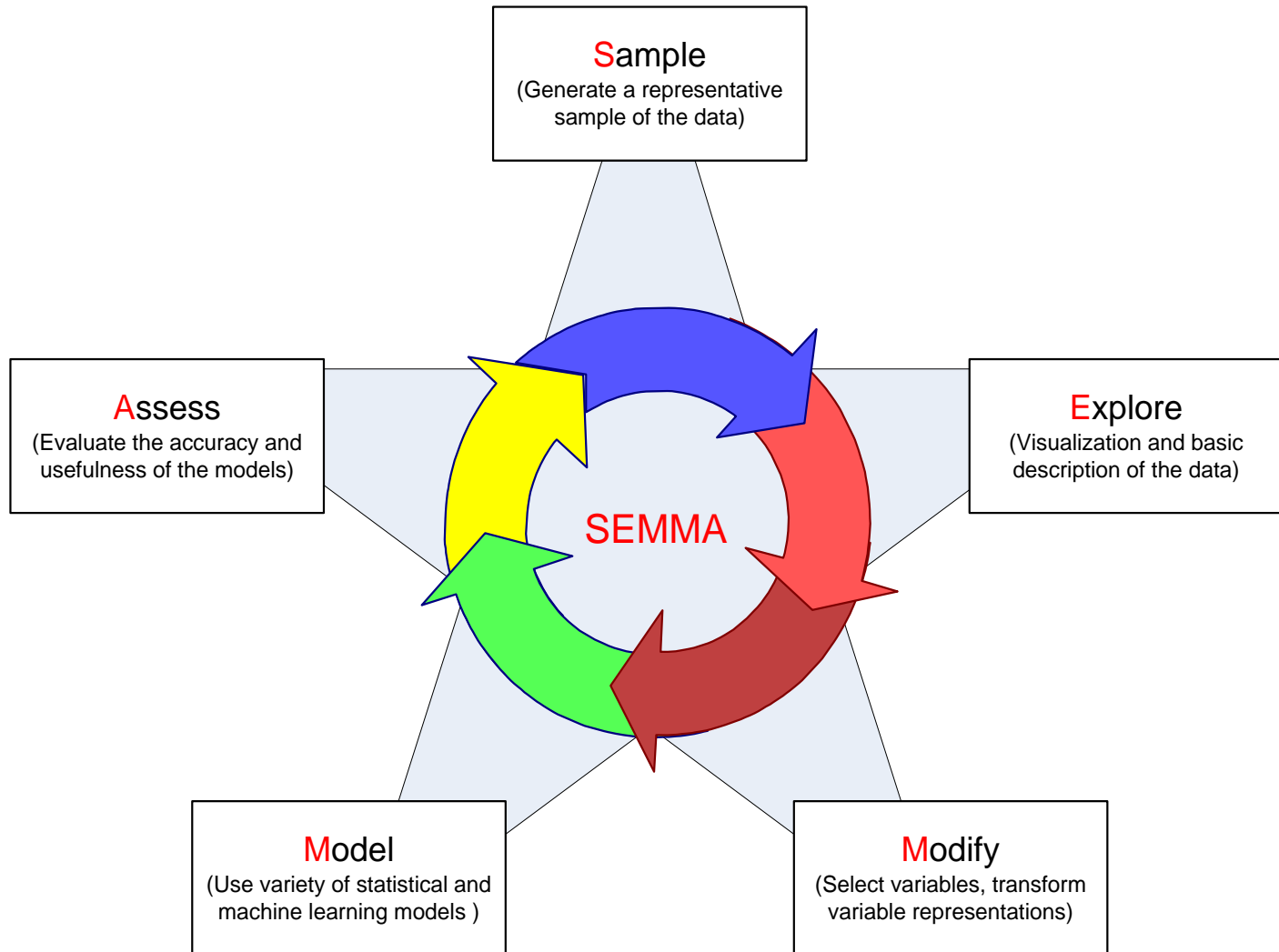
- The process is highly repetitive and experimental (DM: art versus science?)

Data Preparation – A Critical DM Task



Data Mining Process:

SEMMA

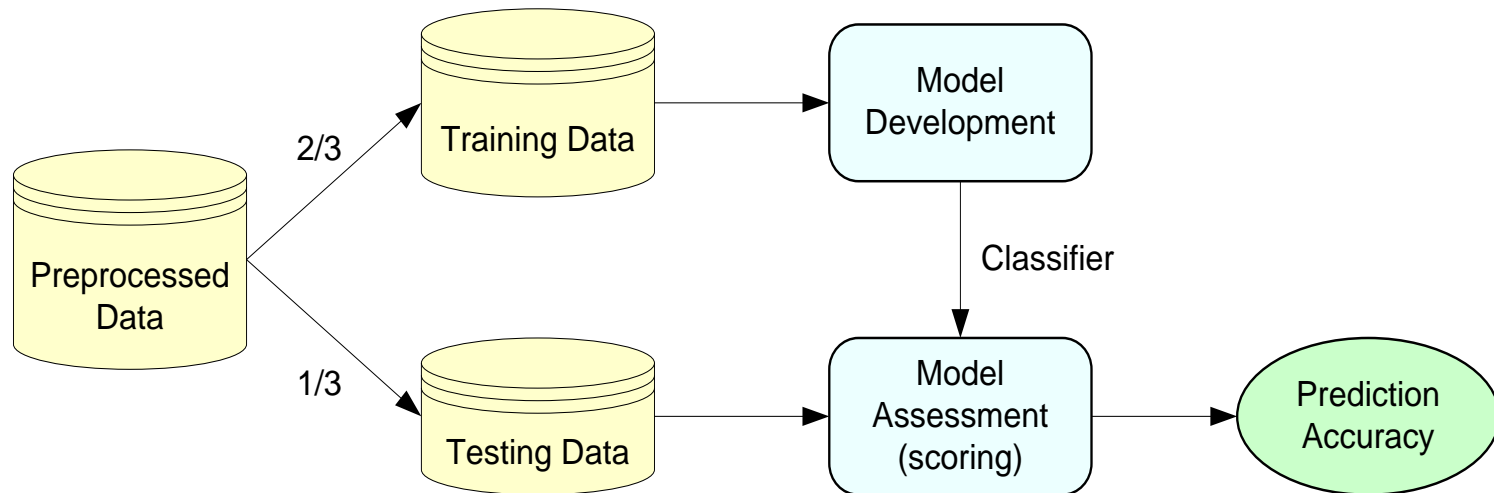


Evaluation

(Accuracy of Classification Model)

Estimation Methodologies for Classification

- **Simple split** (or holdout or test sample estimation)
 - Split the data into 2 mutually exclusive sets training (~70%) and testing (30%)



- For ANN, the data is split into three sub-sets (training [~60%], validation [~20%], testing [~20%])

Estimation Methodologies for Classification

- ***k*-Fold Cross Validation** (rotation estimation)
 - Split the data into k mutually exclusive subsets
 - Use each subset as testing while using the rest of the subsets as training
 - Repeat the experimentation for k times
 - Aggregate the test results for true estimation of prediction accuracy training
- Other estimation methodologies
 - **Leave-one-out, bootstrapping, jackknifing**
 - **Area under the ROC curve**

Assessment Methods for Classification

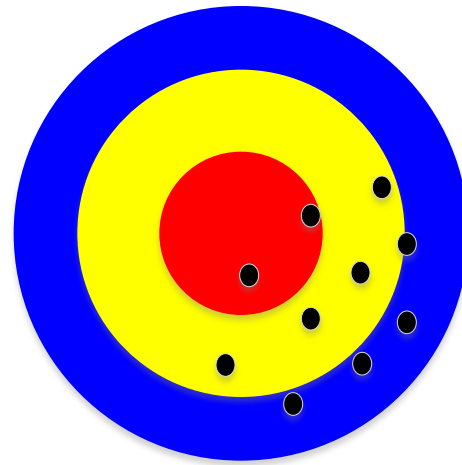
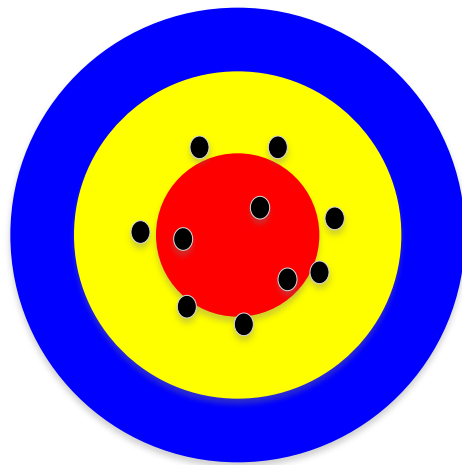
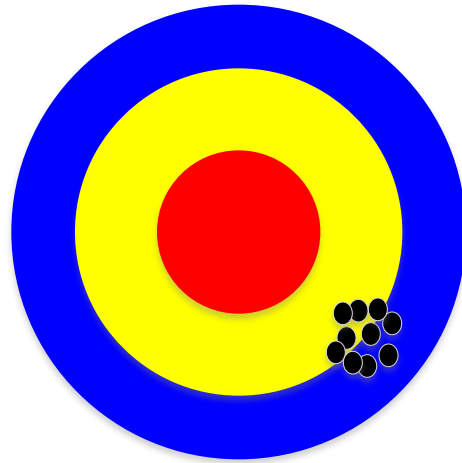
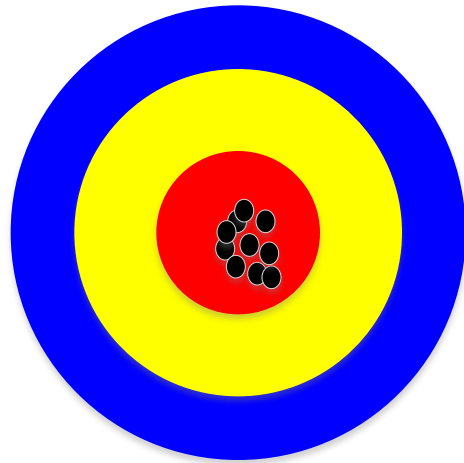
- Predictive accuracy
 - Hit rate
- Speed
 - Model building; predicting
- Robustness
- Scalability
- Interpretability
 - Transparency, explainability

Accuracy

Validity

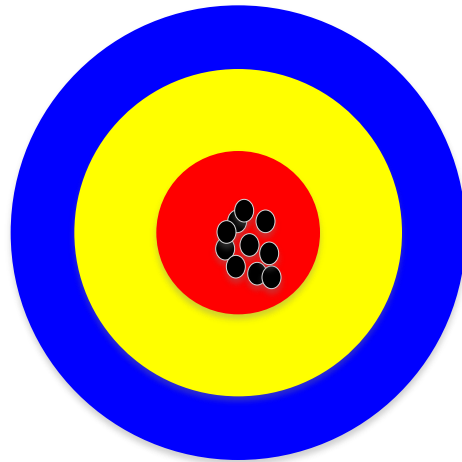
Precision

Reliability



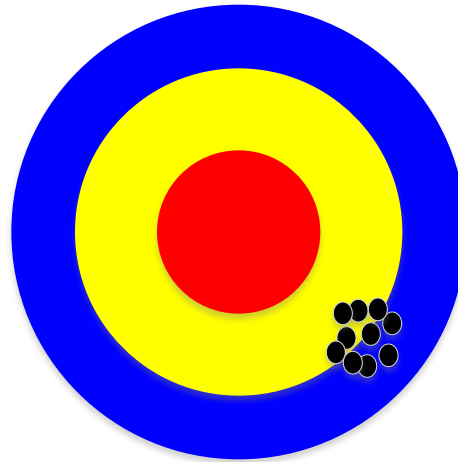
Accuracy vs. Precision

A



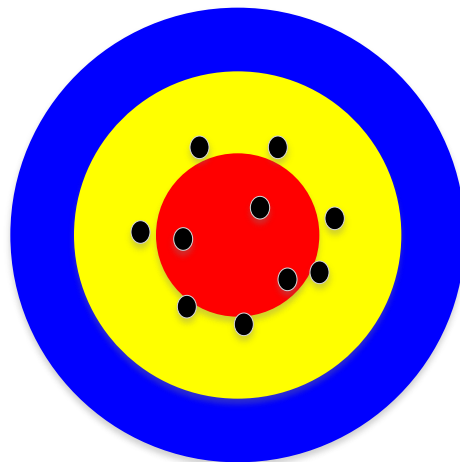
High Accuracy
High Precision

B



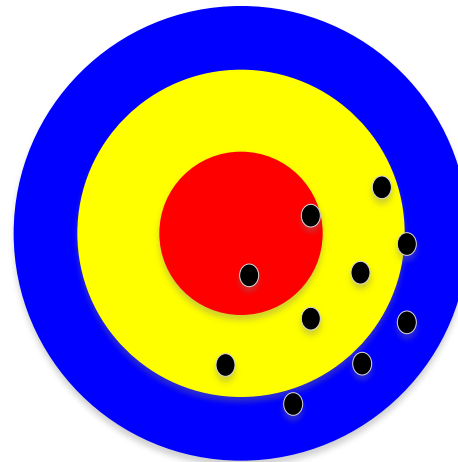
Low Accuracy
High Precision

C



High Accuracy
Low Precision

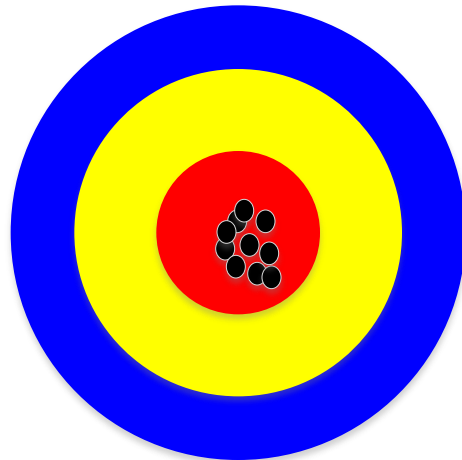
D



Low Accuracy
Low Precision

Accuracy vs. Precision

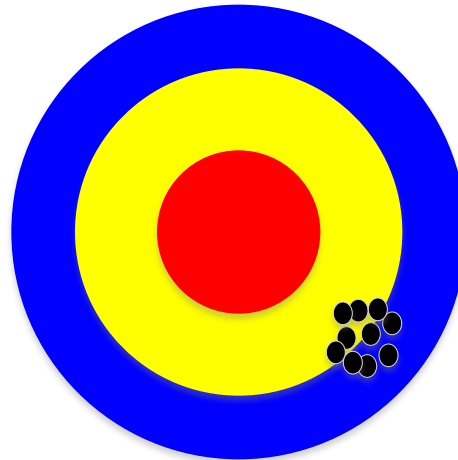
A



**High Accuracy
High Precision**

**High Validity
High Reliability**

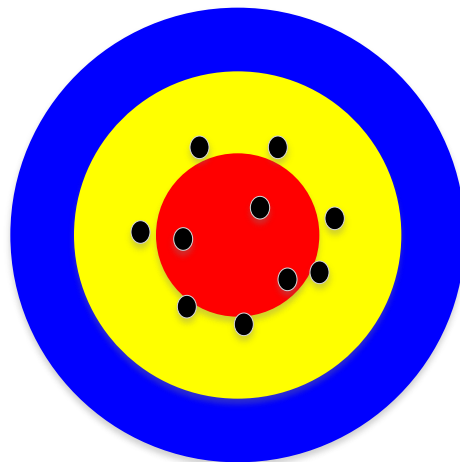
B



**Low Accuracy
High Precision**

**Low Validity
High Reliability**

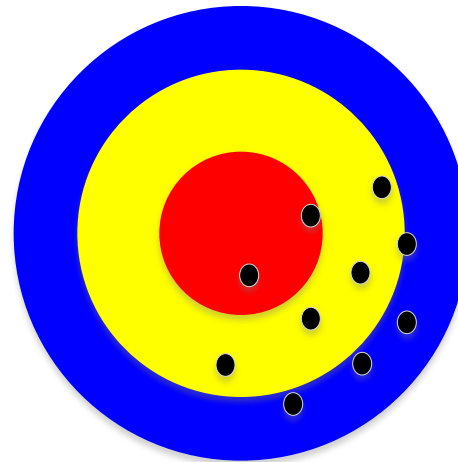
C



**High Accuracy
Low Precision**

**High Validity
Low Reliability**

D

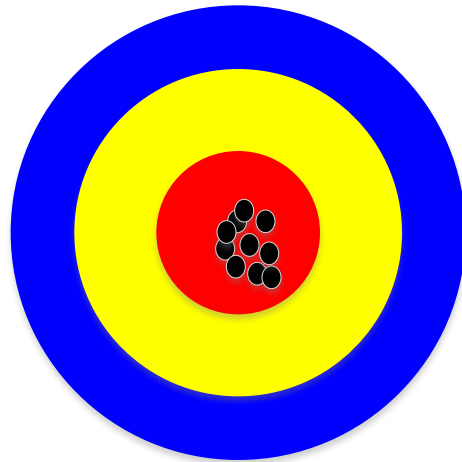


**Low Accuracy
Low Precision**

**Low Validity
Low Reliability**

Accuracy vs. Precision

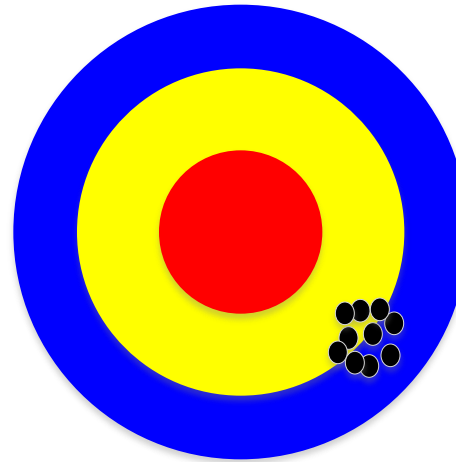
A



High Accuracy
High Precision

High Validity
High Reliability

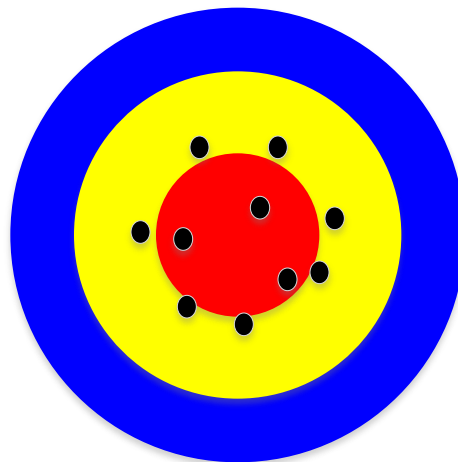
B



Low Accuracy
High Precision

Low Validity
High Reliability

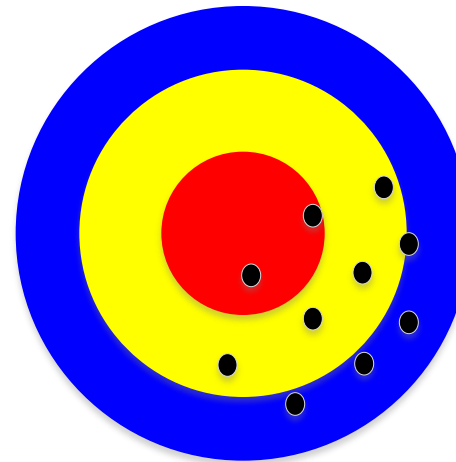
C



High Accuracy
Low Precision

High Validity
Low Reliability

D



Low Accuracy
Low Precision

Low Validity
Low Reliability

Accuracy of Classification Models

- In classification problems, the primary source for accuracy estimation is the **confusion matrix**

		True Class	
		Positive	Negative
Predicted Class	Positive	True Positive Count (TP)	False Positive Count (FP)
	Negative	False Negative Count (FN)	True Negative Count (TN)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

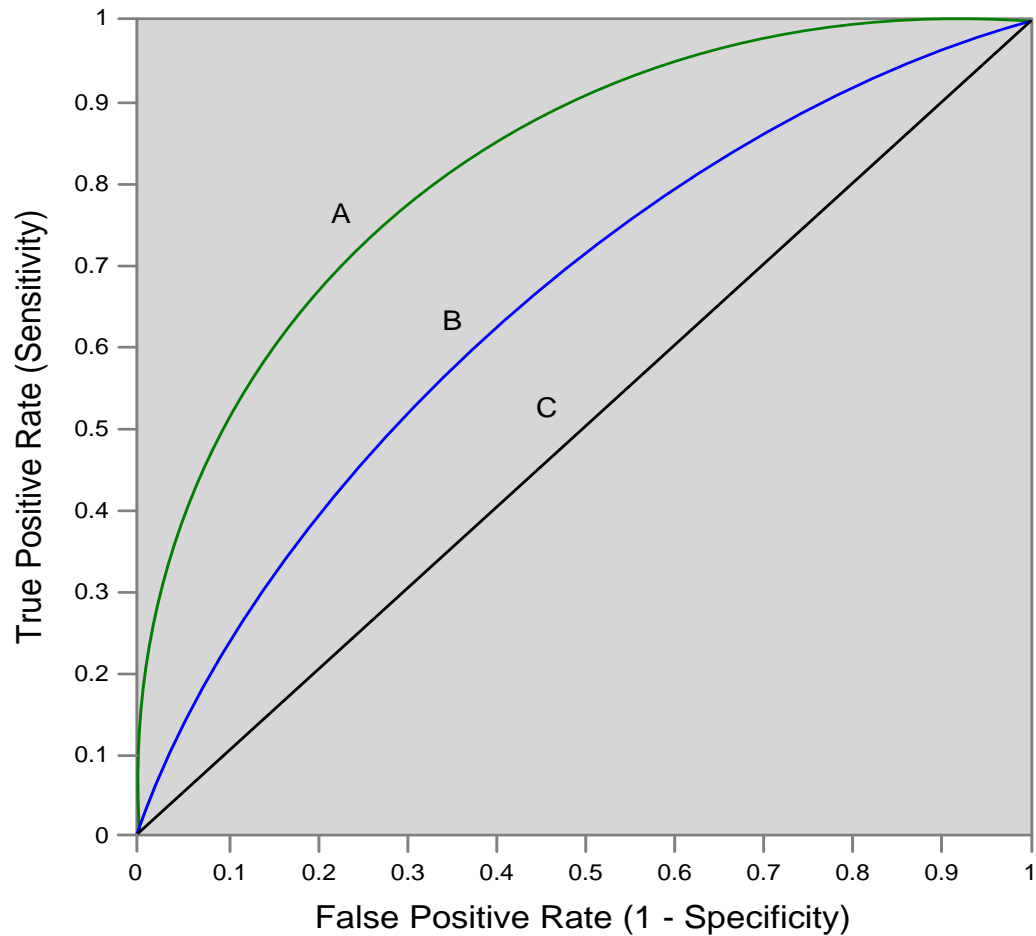
$$True\ Positive\ Rate = \frac{TP}{TP + FN}$$

$$True\ Negative\ Rate = \frac{TN}{TN + FP}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

Estimation Methodologies for Classification – ROC Curve



Sensitivity = True Positive Rate

Specificity = True Negative Rate

		True Class (actual value)		total
		Positive	Negative	
Predictive Class (prediction outcome)	Positive	True Positive (TP)	False Positive (FP)	P'
	Negative	False Negative (FN)	True Negative (TN)	N'
total		P	N	

$$\text{True Positive Rate (Sensitivity)} = \frac{TP}{TP + FN}$$

$$\text{True Negative Rate (Specificity)} = \frac{TN}{TN + FP}$$

$$\text{False Positive Rate} = \frac{FP}{FP + TN}$$

$$\text{False Positive Rate (1 - Specificity)} = \frac{FP}{FP + TN}$$

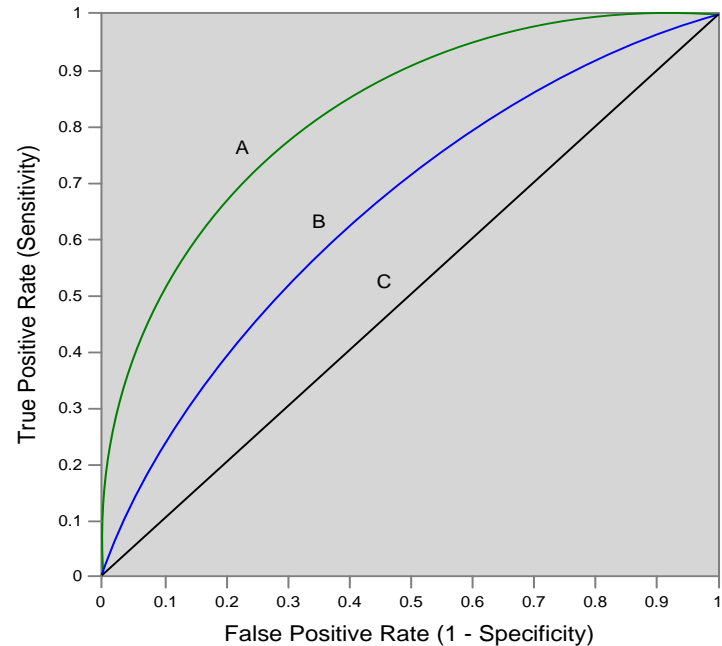
$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{True Positive Rate} = \frac{TP}{TP + FN}$$

$$\text{True Negative Rate} = \frac{TN}{TN + FP}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$



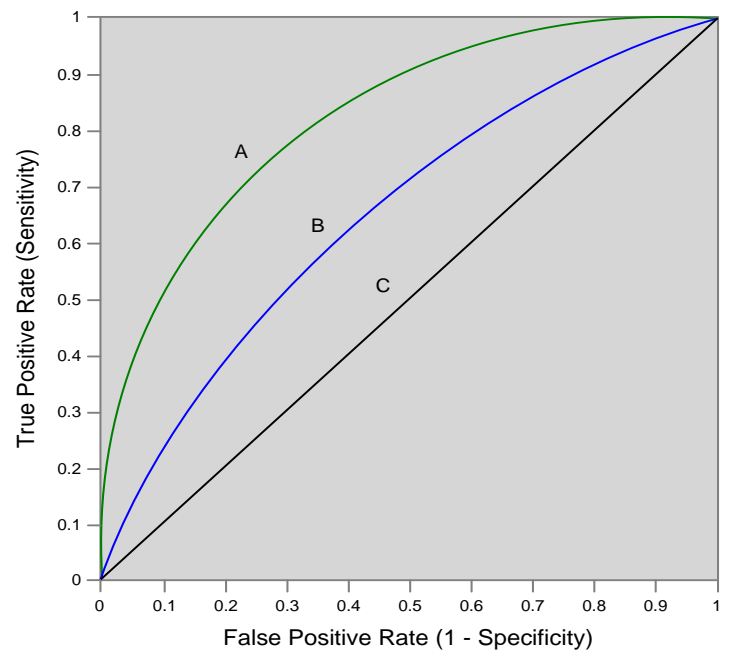
		True Class (actual value)		total
		Positive	Negative	
Predictive Class (prediction outcome)	Positive	True Positive (TP)	False Positive (FP)	P'
	Negative	False Negative (FN)	True Negative (TN)	N'
total		P	N	

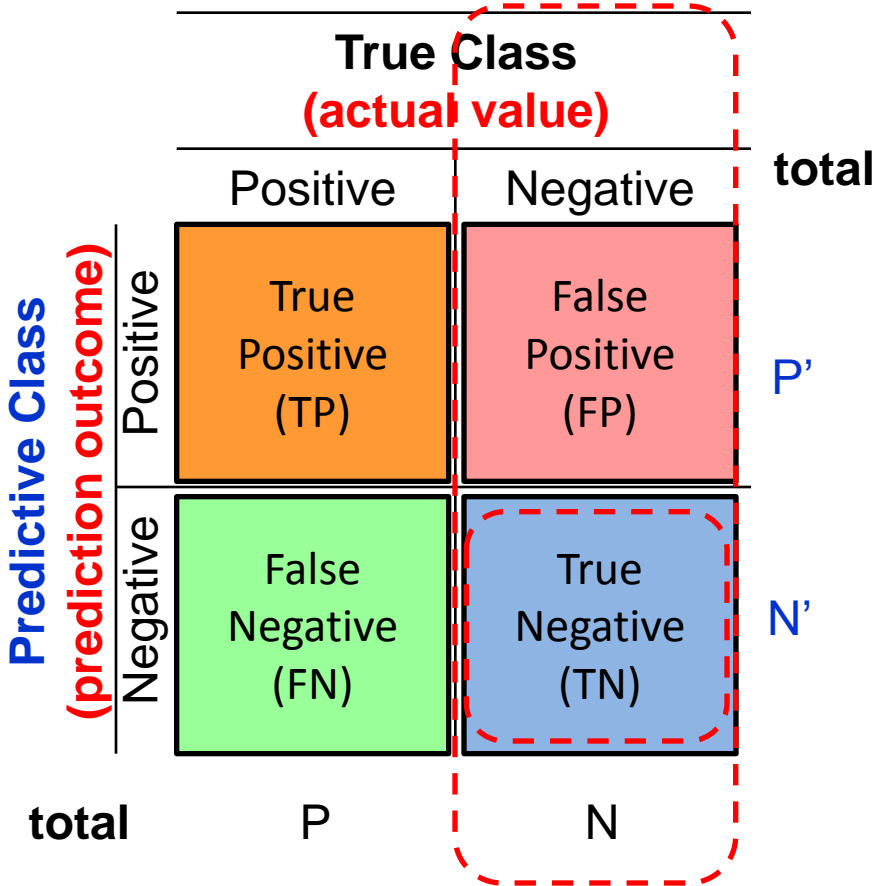
$$\text{True Positive Rate} = \frac{TP}{TP + FN}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{True Positive Rate (Sensitivity)} = \frac{TP}{TP + FN}$$

- Sensitivity**
- = True Positive Rate
- = Recall
- = Hit rate
- = $TP / (TP + FN)$



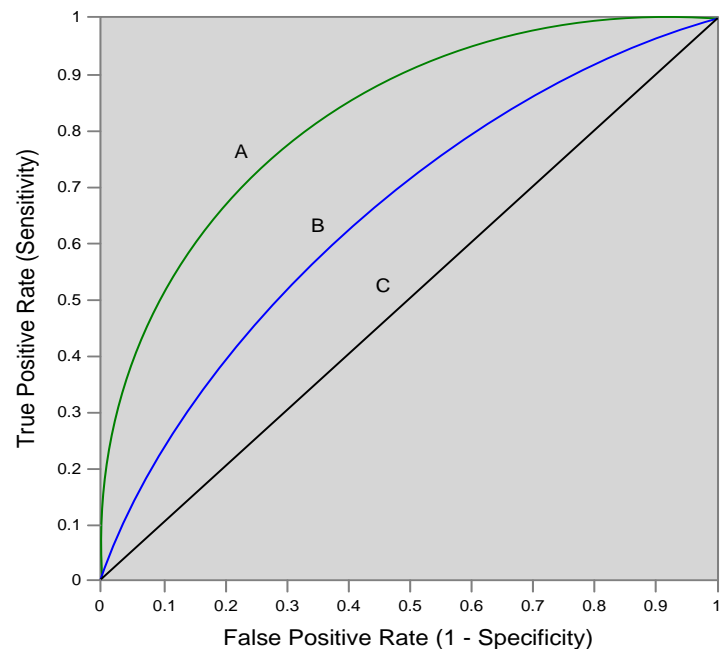


$$\text{True Negative Rate} = \frac{TN}{TN + FP}$$

Specificity
 = True Negative Rate
 = TN / N
 = $TN / (TN + FP)$

$$\text{True Negative Rate (Specificity)} = \frac{TN}{TN + FP}$$

$$\text{False Positive Rate (1 - Specificity)} = \frac{FP}{FP + TN}$$



		True Class (actual value)		total
		Positive	Negative	
Predictive Class (prediction outcome)	Positive	True Positive (TP)	False Positive (FP)	P'
	Negative	False Negative (FN)	True Negative (TN)	N'
total		P	N	

Precision

= Positive Predictive Value (PPV)

$$Precision = \frac{TP}{TP + FP}$$

Recall

= True Positive Rate (TPR)

= Sensitivity

= Hit Rate

$$Recall = \frac{TP}{TP + FN}$$

F1 score (F-score)(F-measure)

is the harmonic mean of precision and recall

$$= 2TP / (P + P')$$

$$= 2TP / (2TP + FP + FN)$$

$$F = 2 * \frac{precision * recall}{precision + recall}$$

A		
63 (TP)	28 (FP)	91
37 (FN)	72 (TN)	109
100	100	200

Recall

= True Positive Rate (TPR)
 = Sensitivity
 = Hit Rate
 = $TP / (TP + FN)$

Specificity

= True Negative Rate
 = TN / N
 = $TN / (TN + FP)$

$$TPR = 0.63$$

$$Recall = \frac{TP}{TP + FN}$$

$$True\ Negative\ Rate\ (Specificity) = \frac{TN}{TN + FP}$$

$$FPR = 0.28$$

$$False\ Positive\ Rate\ (1 - Specificity) = \frac{FP}{FP + TN}$$

$$PPV = 0.69$$

$$= 63 / (63 + 28)$$

$$= 63 / 91$$

$$Precision = \frac{TP}{TP + FP}$$

Precision

= Positive Predictive Value (PPV)

$$F1 = 0.66$$

$$= 2 * (0.63 * 0.69) / (0.63 + 0.69)$$

$$= (2 * 63) / (100 + 91)$$

$$= (0.63 + 0.69) / 2 = 1.32 / 2 = 0.66$$

$$F = 2 * \frac{precision * recall}{precision + recall}$$

F1 score (F-score) (F-measure)

is the harmonic mean of precision and recall

$$= 2TP / (P + P')$$

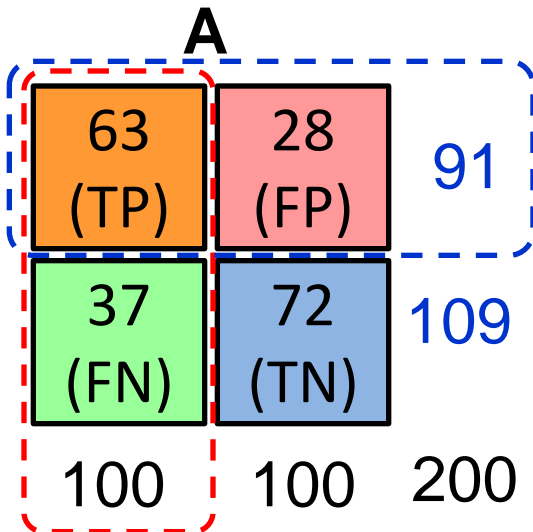
$$= 2TP / (2TP + FP + FN)$$

$$ACC = 0.68$$

$$= (63 + 72) / 200$$

$$= 135 / 200 = 0.675$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$



$$\text{TPR} = 0.63$$

$$\text{FPR} = 0.28$$

$$\begin{aligned} \text{PPV} &= 0.69 \\ &= 63 / (63 + 28) \\ &= 63 / 91 \end{aligned}$$

$$\text{F1} = 0.66$$

$$= 2 * (0.63 * 0.69) / (0.63 + 0.69)$$

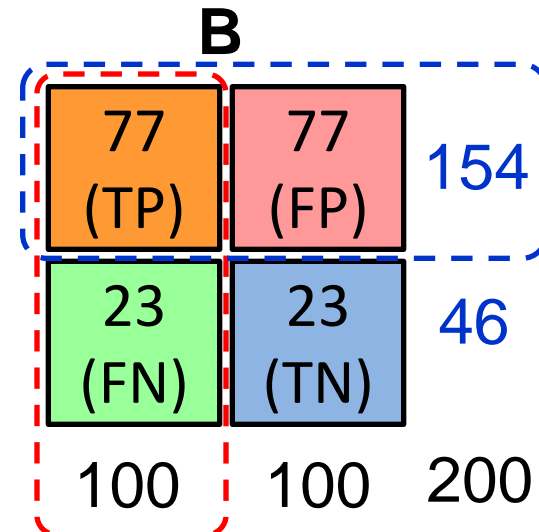
$$= (2 * 63) / (100 + 91)$$

$$= (0.63 + 0.69) / 2 = 1.32 / 2 = 0.66$$

$$\text{ACC} = 0.68$$

$$= (63 + 72) / 200$$

$$= 135 / 200 = 0.675$$



$$\text{TPR} = 0.77$$

$$\text{FPR} = 0.77$$

$$\text{PPV} = 0.50$$

$$\text{F1} = 0.61$$

$$\text{ACC} = 0.50$$

Recall

= True Positive Rate (TPR)

= Sensitivity

= Hit Rate

$$\text{Recall} = \frac{TP}{TP + FN}$$

Precision

= Positive Predictive Value (PPV)

$$\text{Precision} = \frac{TP}{TP + FP}$$

C

24 (TP)	88 (FP)	112
76 (FN)	12 (TN)	88
100	100	200

TPR = 0.24

FPR = 0.88

PPV = 0.21

F1 = 0.22

ACC = 0.18

C'

76 (TP)	12 (FP)	88
24 (FN)	88 (TN)	112
100	100	200

TPR = 0.76

FPR = 0.12

PPV = 0.86

F1 = 0.81

ACC = 0.82

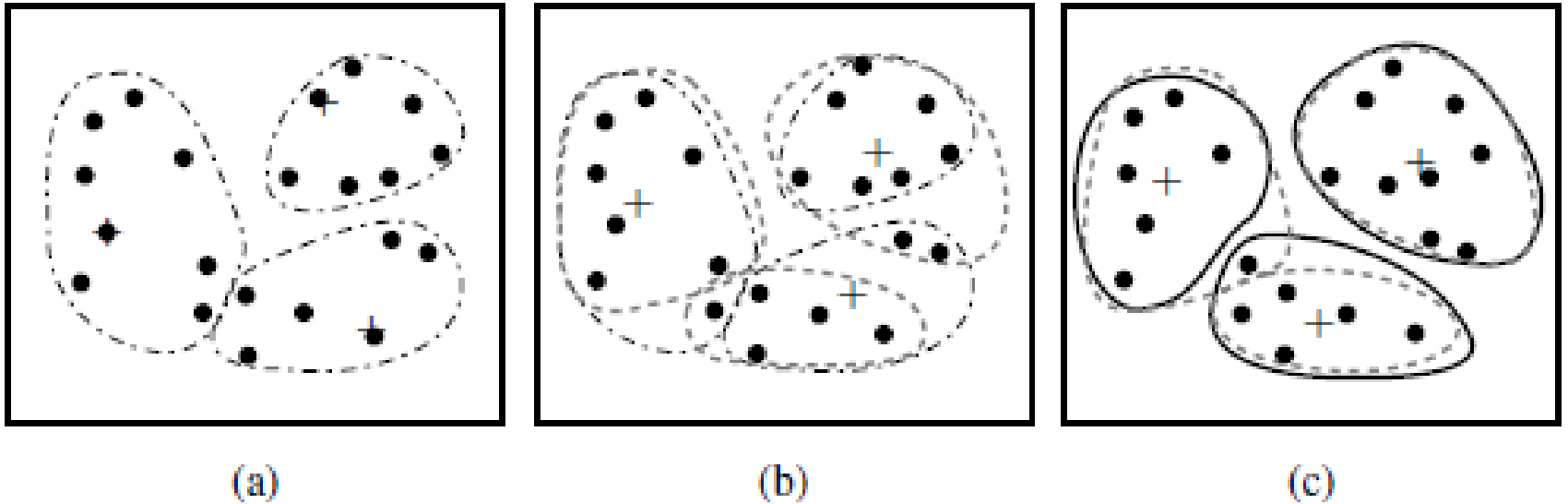
Recall
 = True Positive Rate (TPR) $Recall = \frac{TP}{TP + FN}$
 = Sensitivity
 = Hit Rate

Precision
 = Positive Predictive Value (PPV) $Precision = \frac{TP}{TP + FP}$

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

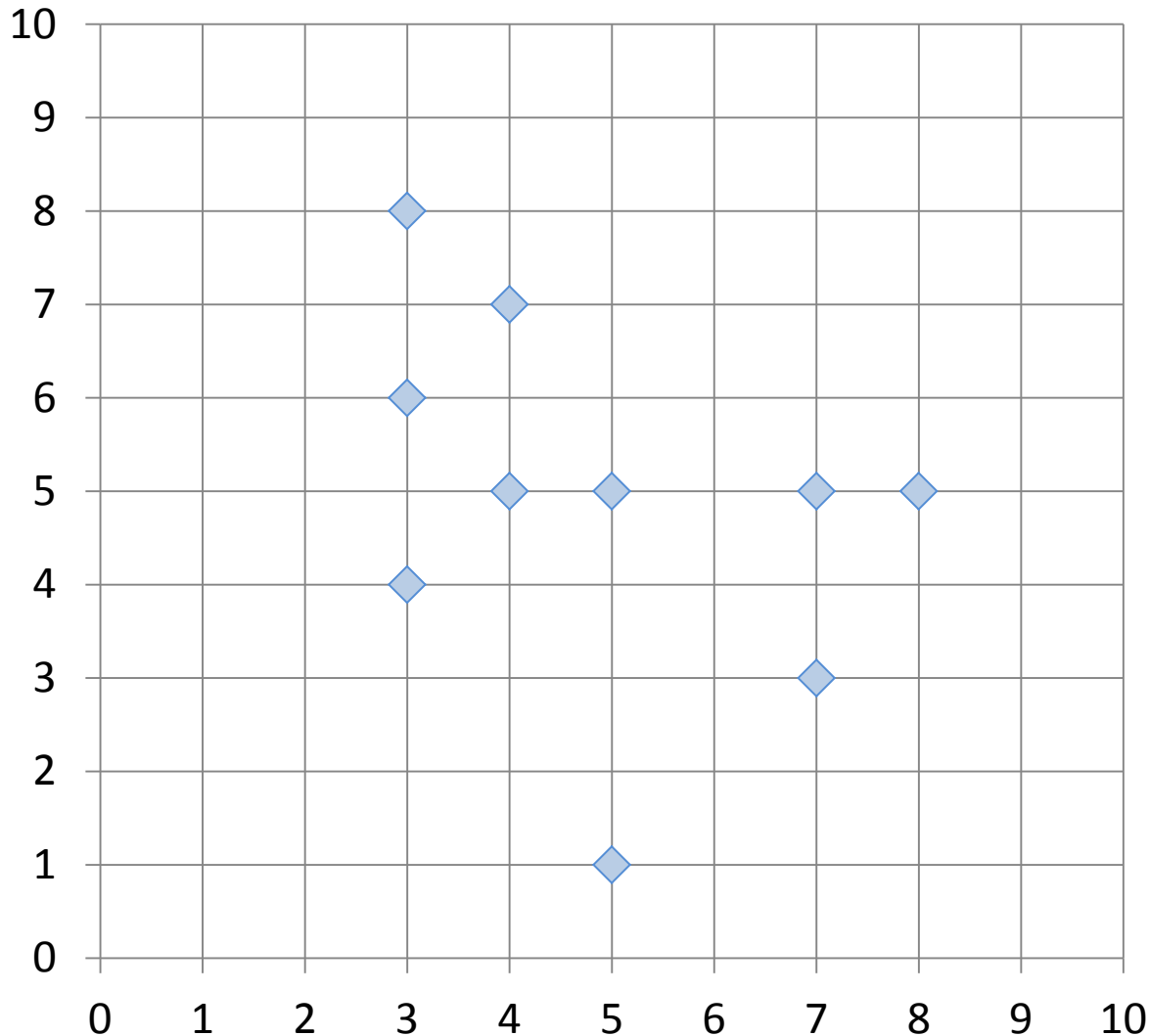


Clustering of a set of objects based on the *k-means method*.
(The mean of each cluster is marked by a “+”.)

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)

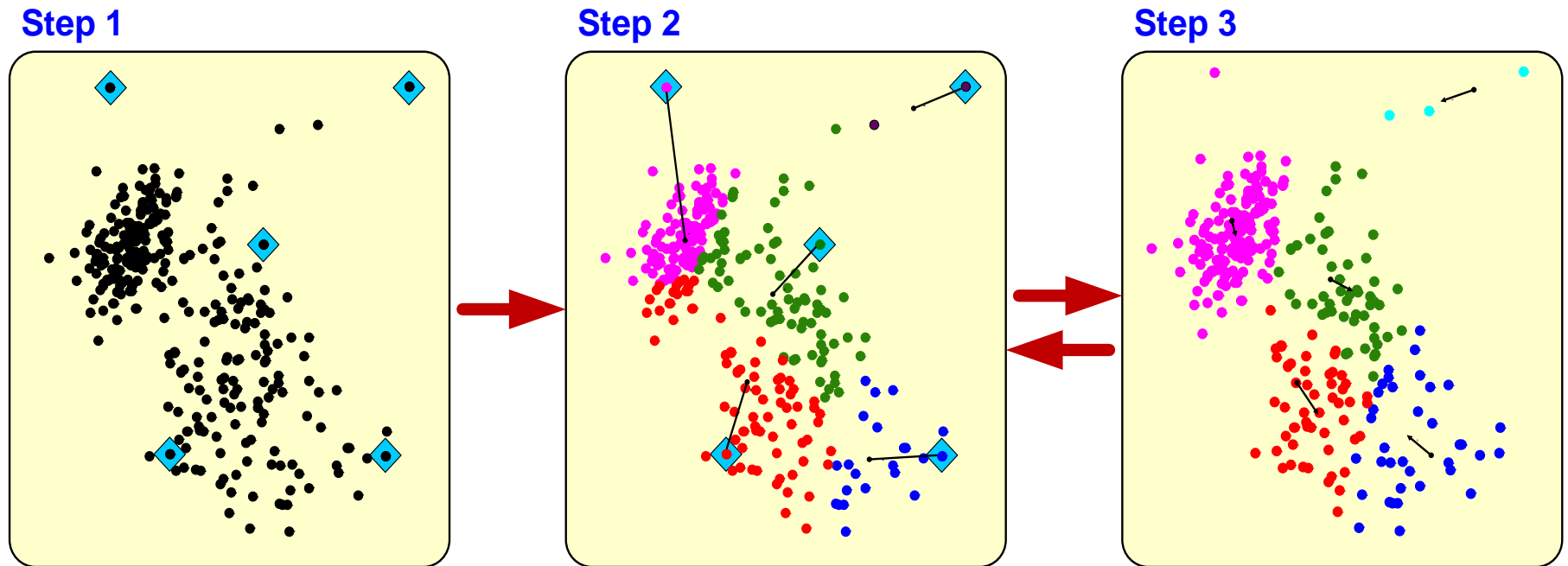
Step 1: Randomly generate k random points as initial cluster centers

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*:

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

where $i = (x_{i_1}, x_{i_2}, \dots, x_{i_p})$ and $j = (x_{j_1}, x_{j_2}, \dots, x_{j_p})$ are two p -dimensional data objects, and q is a positive integer

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

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

Similarity and Dissimilarity Between Objects (Cont.)

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

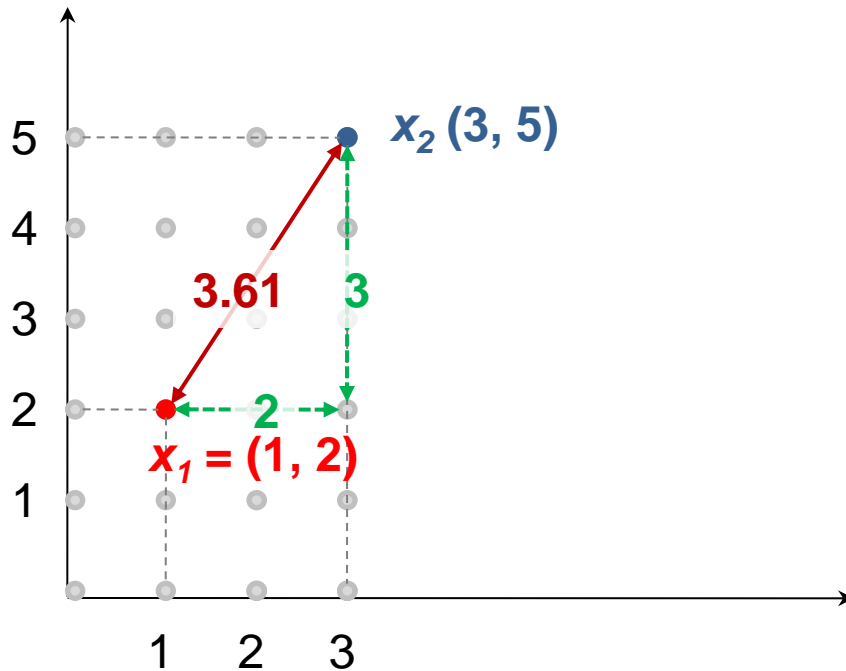
$$d(i, j) = \sqrt{(|x_{i_1} - x_{j_1}|^2 + |x_{i_2} - x_{j_2}|^2 + \dots + |x_{i_p} - x_{j_p}|^2)}$$

– 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:

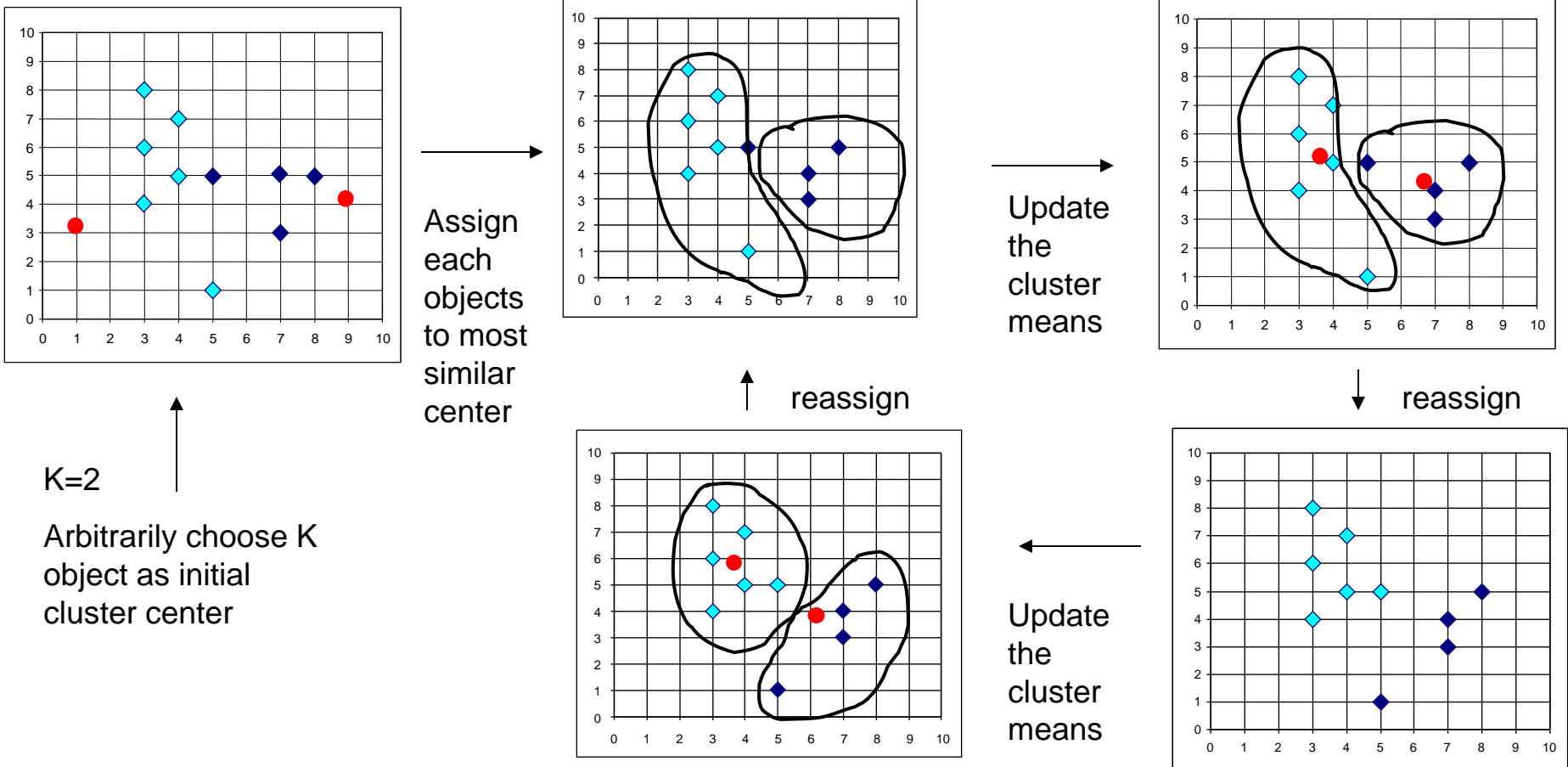
$$\begin{aligned} &= ((3-1)^2 + (5-2)^2)^{1/2} \\ &= (2^2 + 3^2)^{1/2} \\ &= (4 + 9)^{1/2} \\ &= (13)^{1/2} \\ &= 3.61 \end{aligned}$$

Manhattan distance:

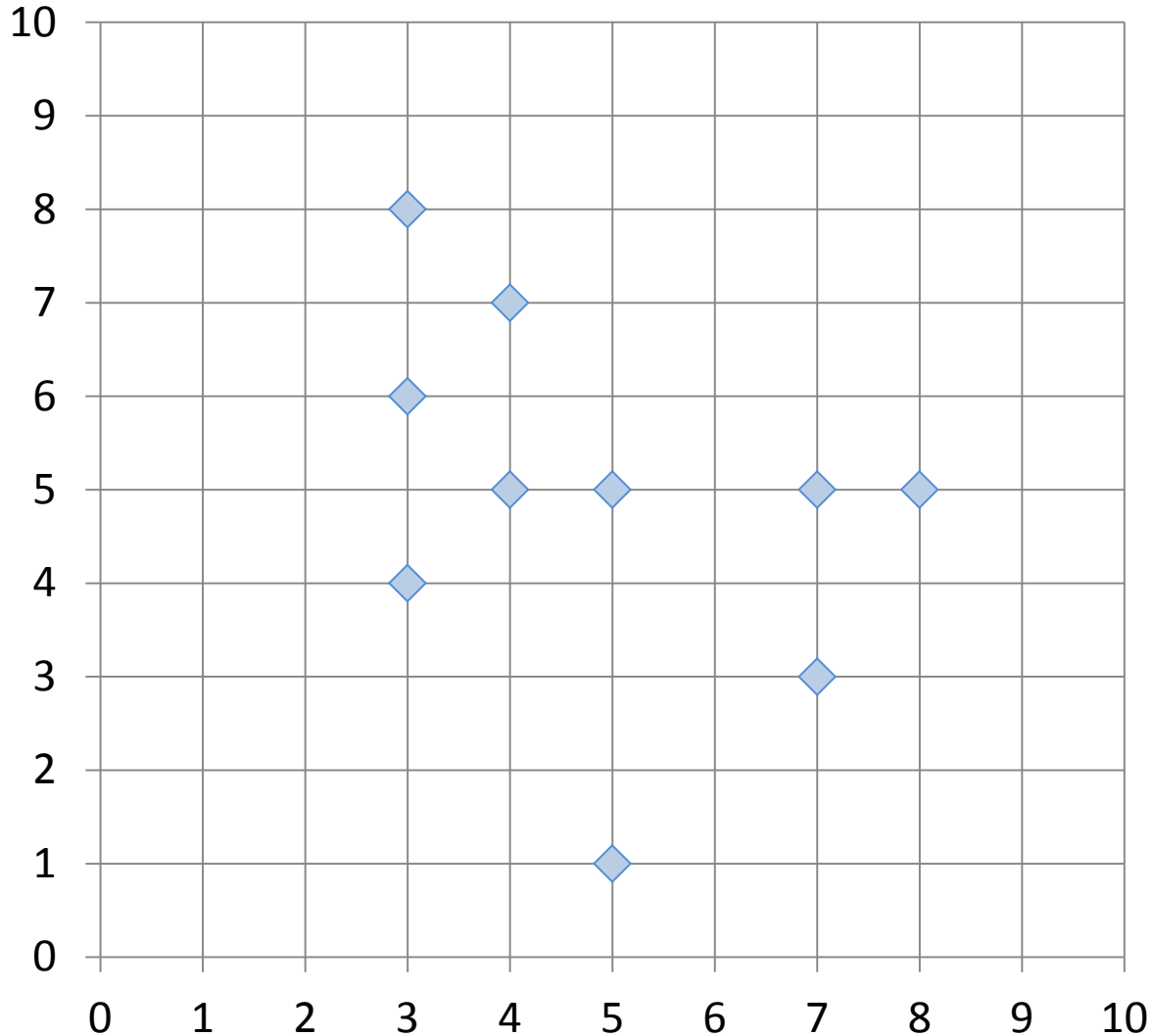
$$\begin{aligned} &= (3-1) + (5-2) \\ &= 2 + 3 \\ &= 5 \end{aligned}$$

The *K-Means* Clustering Method

- Example



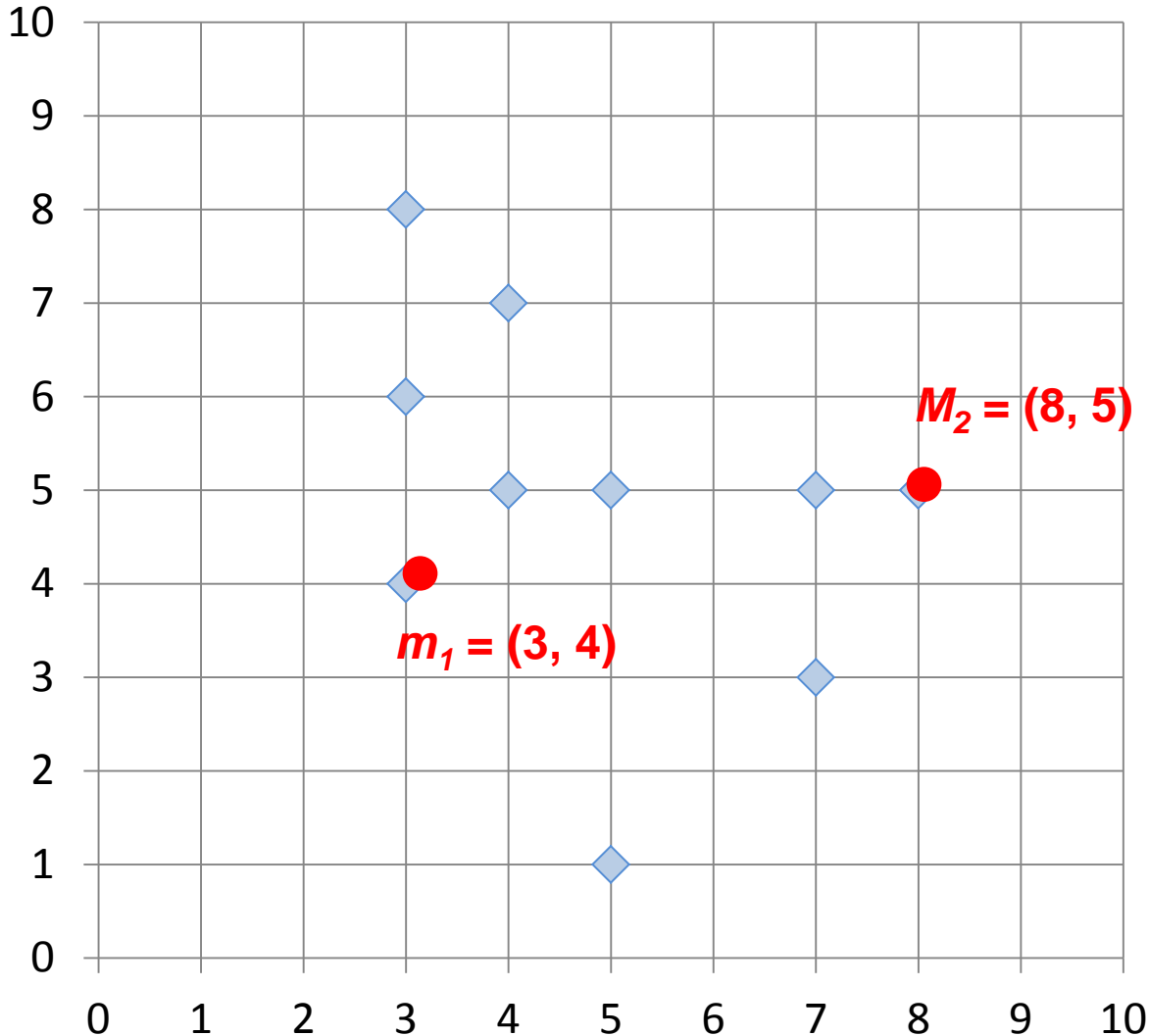
K-Means Clustering Step by Step



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

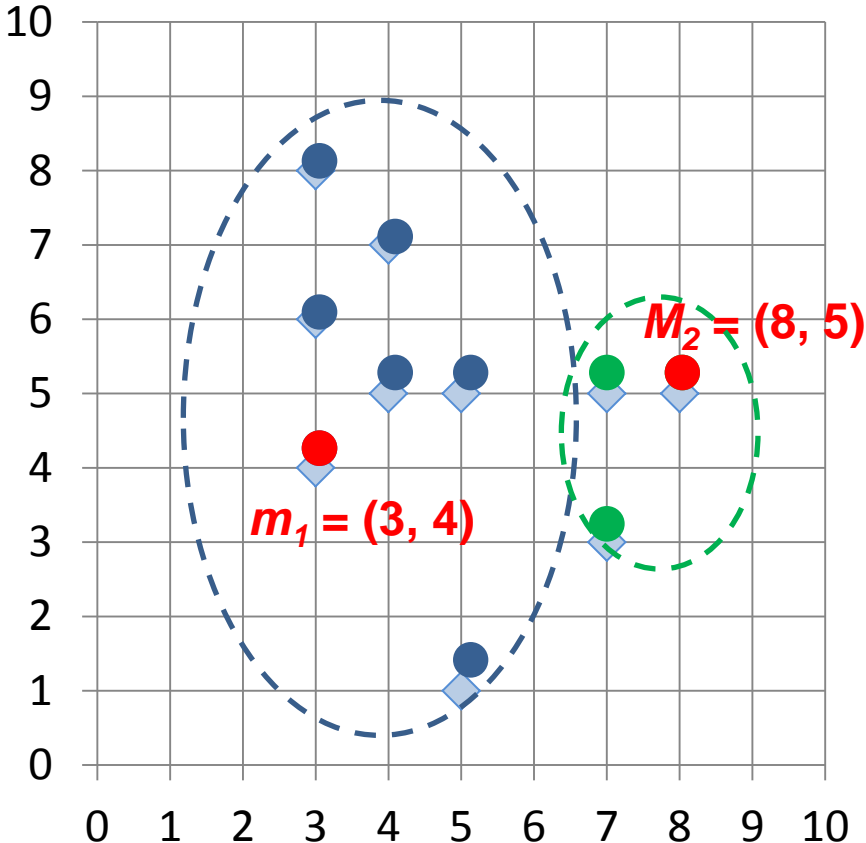


Point	P	P(x,y)
p01	a	(3, 4)
p02	b	(3, 6)
p03	c	(3, 8)
p04	d	(4, 5)
p05	e	(4, 7)
p06	f	(5, 1)
p07	g	(5, 5)
p08	h	(7, 3)
p09	i	(7, 5)
p10	j	(8, 5)

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



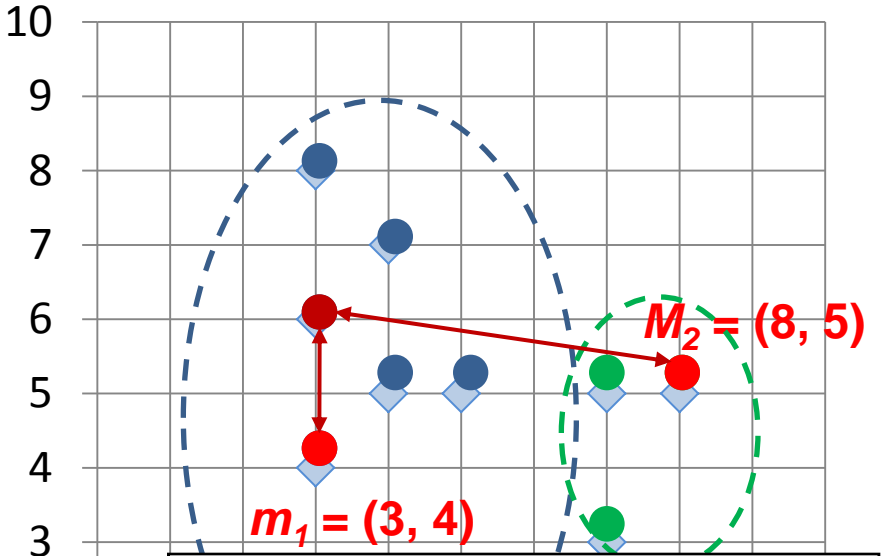
Point	P	P(x,y)	m1 distance	m2 distance	Cluster
p01	a	(3, 4)	0.00	5.10	Cluster1
p02	b	(3, 6)	2.00	5.10	Cluster1
p03	c	(3, 8)	4.00	5.83	Cluster1
p04	d	(4, 5)	1.41	4.00	Cluster1
p05	e	(4, 7)	3.16	4.47	Cluster1
p06	f	(5, 1)	3.61	5.00	Cluster1
p07	g	(5, 5)	2.24	3.00	Cluster1
p08	h	(7, 3)	4.12	2.24	Cluster2
p09	i	(7, 5)	4.12	1.00	Cluster2
p10	j	(8, 5)	5.10	0.00	Cluster2

Initial m1 (3, 4)
Initial m2 (8, 5)

K-Means Clustering

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

Step 3: Assign each objects to most similar center



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

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

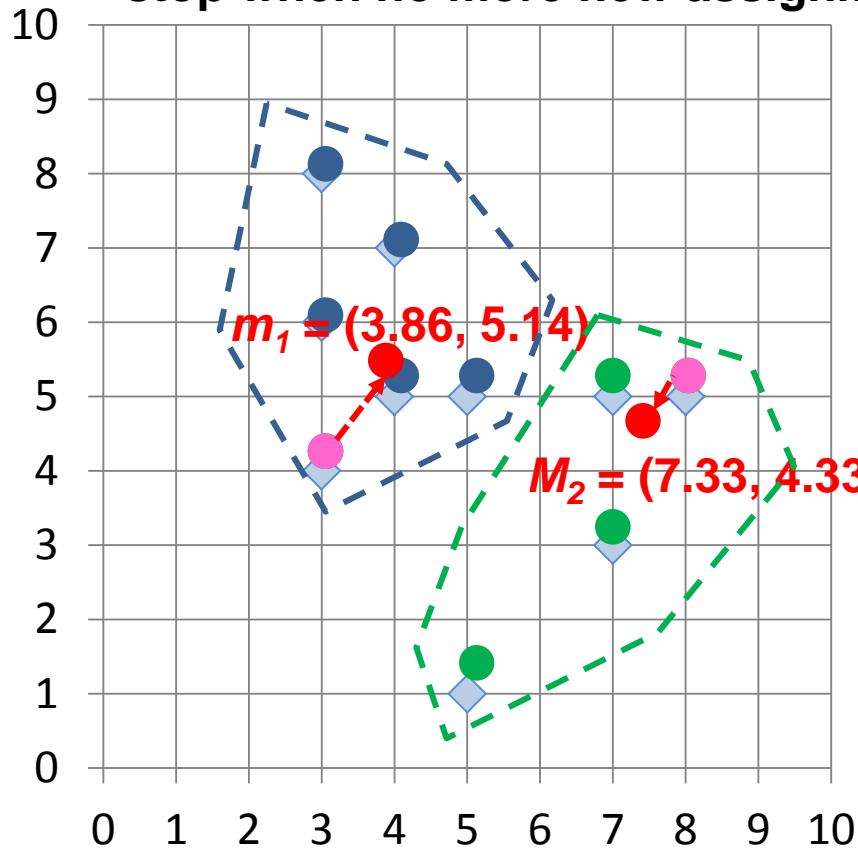
Euclidean distance
 $b(3,6) \leftrightarrow m2(8,5)$
 $= ((8-3)^2 + (5-6)^2)^{1/2}$
 $= (5^2 + (-1)^2)^{1/2}$
 $= (25 + 1)^{1/2}$
 $= (26)^{1/2}$
 $= 5.10$

Initial m1 (3, 4)
 Initial m2 (8, 5)

K-1

Cluster1
 Cluster1
 Cluster1
 Cluster1
 Cluster1
 Cluster2
 Cluster2
 Cluster2

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



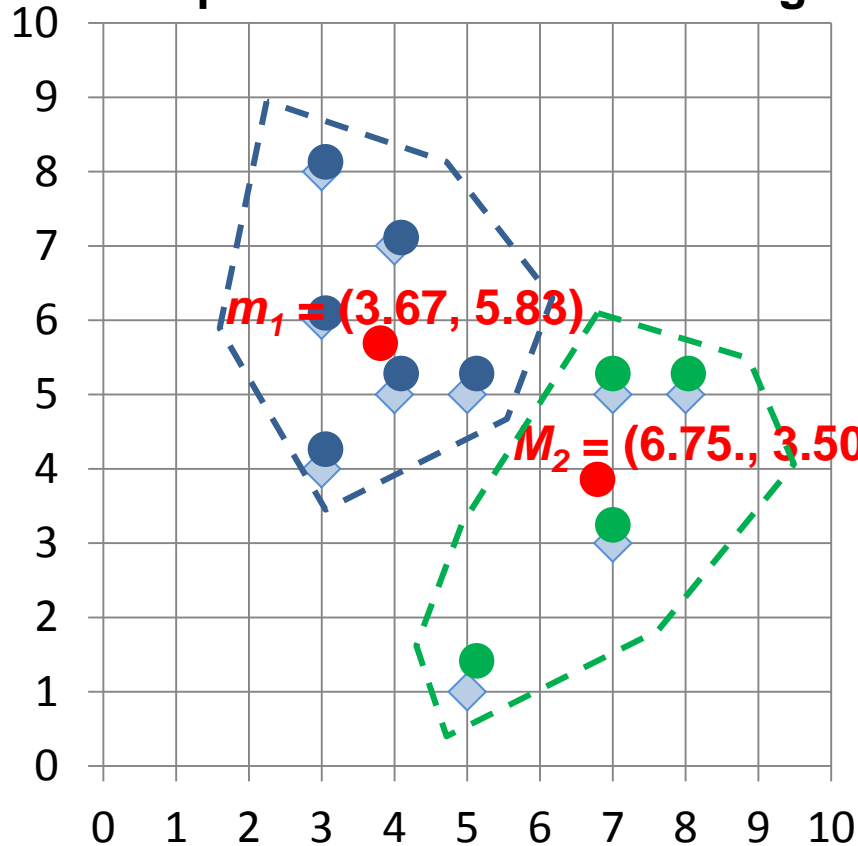
Point	P	P(x,y)	m1 distance	m2 distance	Cluster
p01	a	(3, 4)	1.43	4.34	Cluster1
p02	b	(3, 6)	1.22	4.64	Cluster1
p03	c	(3, 8)	2.99	5.68	Cluster1
p04	d	(4, 5)	0.20	3.40	Cluster1
p05	e	(4, 7)	1.87	4.27	Cluster1
p06	f	(5, 1)	4.29	4.06	Cluster2
p07	g	(5, 5)	1.15	2.42	Cluster1
p08	h	(7, 3)	3.80	1.37	Cluster2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

m1 (3.86, 5.14)

m2 (7.33, 4.33)

***K-Means* Clustering**

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



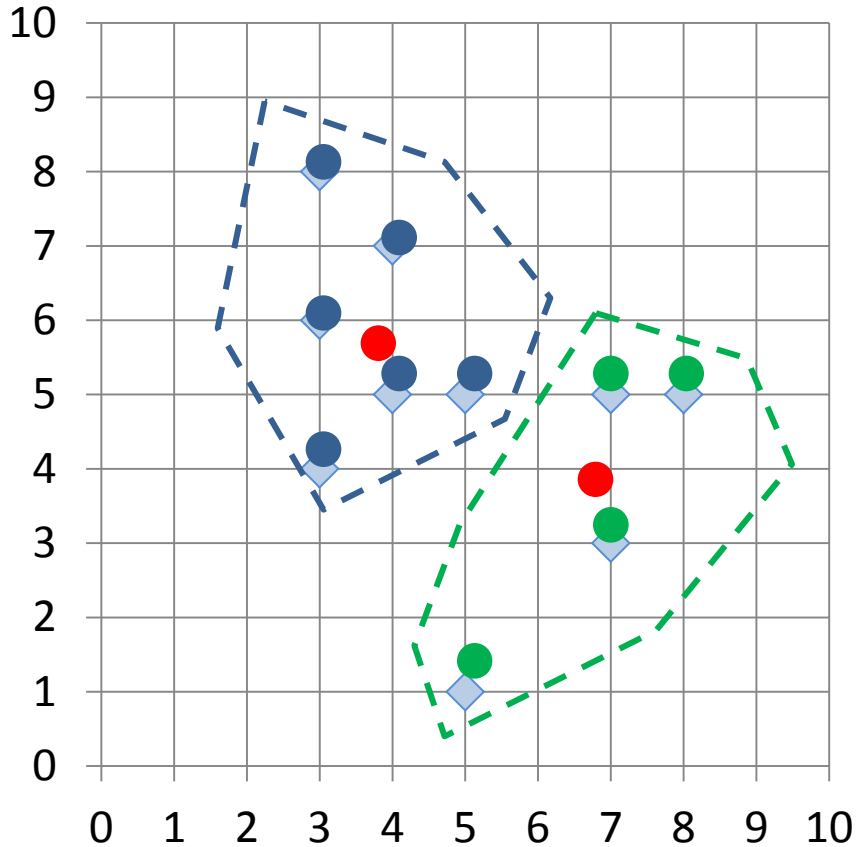
Point	P	P(x,y)	m1 distance	m2 distance	Cluster
p01	a	(3, 4)	1.95	3.78	Cluster1
p02	b	(3, 6)	0.69	4.51	Cluster1
p03	c	(3, 8)	2.27	5.86	Cluster1
p04	d	(4, 5)	0.89	3.13	Cluster1
p05	e	(4, 7)	1.22	4.45	Cluster1
p06	f	(5, 1)	5.01	3.05	Cluster2
p07	g	(5, 5)	1.57	2.30	Cluster1
p08	h	(7, 3)	4.37	0.56	Cluster2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

m1 (3.67, 5.83)

m2 (6.75, 3.50)

***K-Means* Clustering**

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

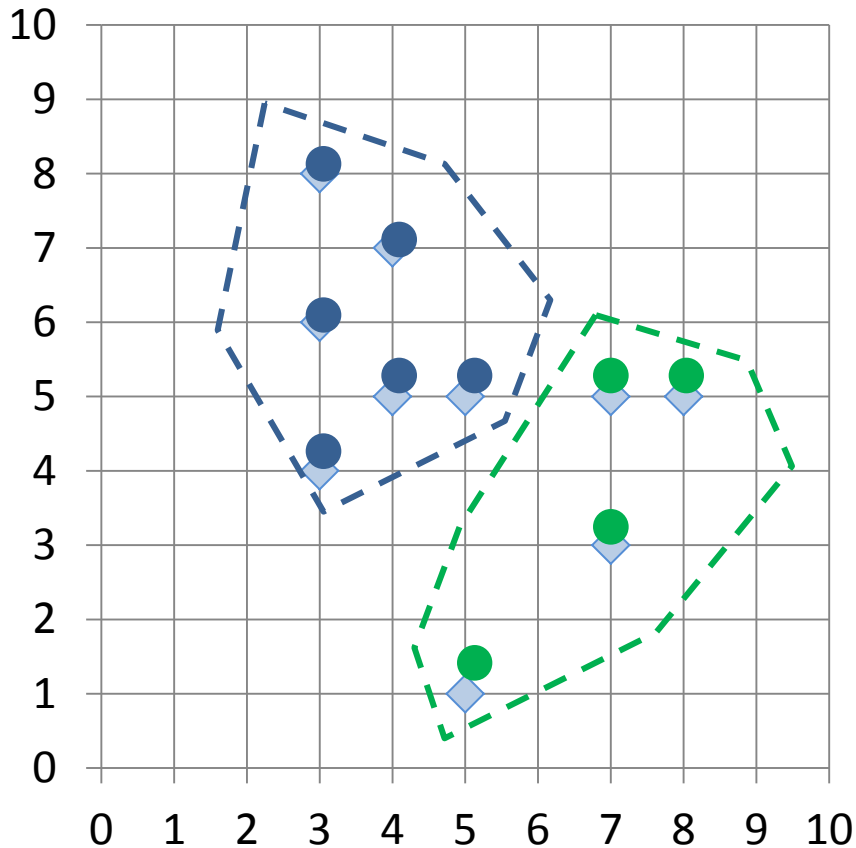
m1 (3.67, 5.83)

m2 (6.75, 3.50)

K-Means Clustering

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

K-Means Clustering

m1 (3.67, 5.83)

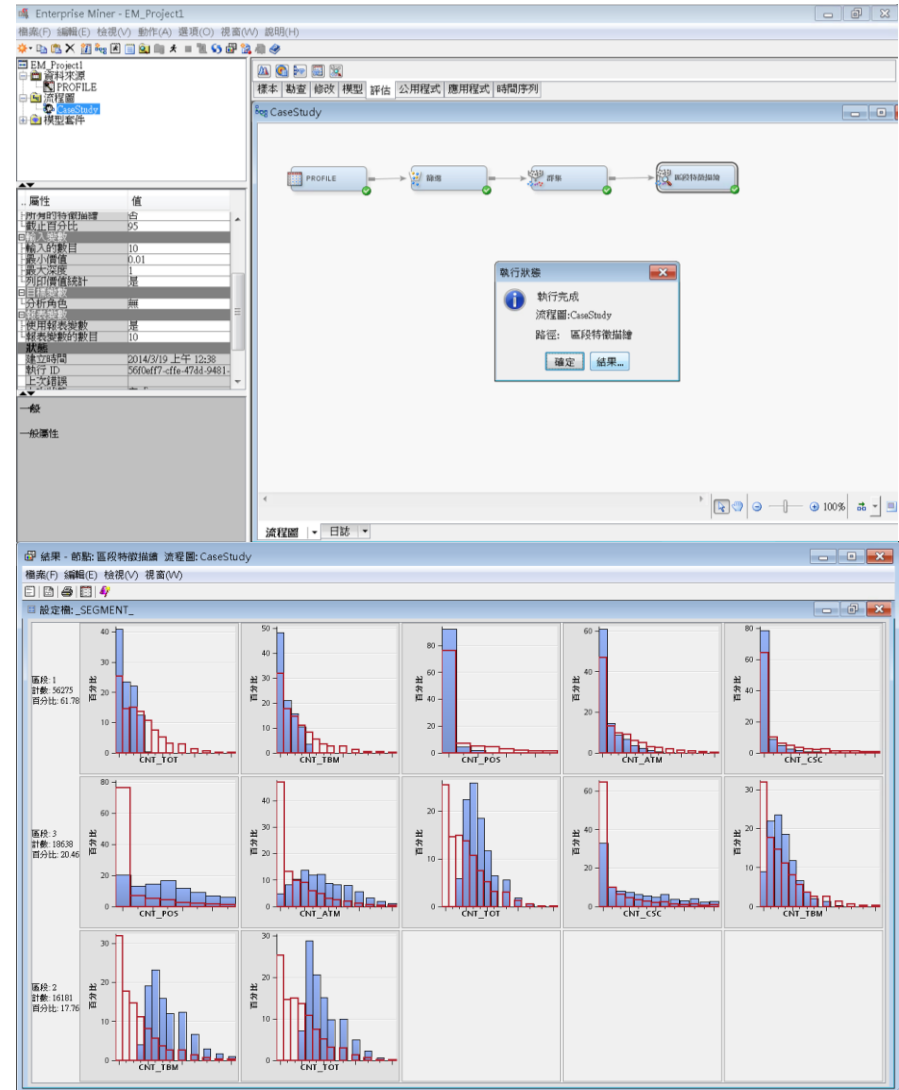
m2 (6.75, 3.50)

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

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	35.0	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

ID	CNT_TBM	CNT_ATM	CNT_POS	CNT_CSC	CNT_TOT
1	3.0	3.0	3.0	3.0	12.0
2	17.0	5.0	1.0	1.0	24.0
3	26.0	5.0	3.0	3.0	37.0
4	3.0	6.0	1.0	5.0	15.0
5	15.0	5.0	5.0	3.0	28.0
6	20.0	4.0	3.0	3.0	30.0
7	9.0	4.0	1.0	3.0	17.0
8	5.0	6.0	1.0	1.0	13.0
9	25.0	5.0	1.0	3.0	34.0
10	3.0	3.0	3.0	3.0	12.0
11	20.0	5.0	1.0	3.0	29.0
12	4.0	6.0	1.0	3.0	14.0
13	5.0	4.0	1.0	3.0	13.0
14	3.0	4.0	1.0	1.0	9.0
15	3.0	5.0	1.0	1.0	10.0
16	14.0	7.0	3.0	3.0	27.0
17	3.0	3.0	3.0	3.0	12.0
18	25.0	1.0	3.0	3.0	32.0
19	3.0	3.0	1.0	3.0	10.0
20	5.0	3.0	1.0	3.0	12.0
21	3.0	3.0	3.0	3.0	12.0
22	14.0	3.0	1.0	3.0	21.0
23	3.0	3.0	1.0	1.0	8.0
24	3.0	3.0	1.0	1.0	8.0
25	3.0	3.0	1.0	1.0	8.0
26	3.0	3.0	1.0	1.0	8.0
27	3.0	3.0	1.0	1.0	8.0
28	5.0	3.0	1.0	1.0	10.0
29	3.0	3.0	1.0	1.0	8.0

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

新分頁 x

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

將書籤放置在書籤列上，即可快速前往各個網頁。立即匯入書籤...

未登入「Chrome」
(您錯過了下列步驟 - 登入)

商店 Google 文件 Google 雲端硬碟 Gmail Google 搜尋 YouTube

chrome 最常造訪 應用程式 最近開關的分頁 線上應用程式商店

EM_Data.zip 顯示所有下載...

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 folder tree with "EM_Data" selected. The main pane displays four SAS Data Set files:

File Name	Type	Modification Date	Size
credit.sas7bdat	SAS Data Set	2013/4/21 下午 05:22	361 KB
inq2006.sas7bdat	SAS Data Set	2013/5/29 下午 01:20	13.4 MB
profile.sas7bdat	SAS Data Set	2006/9/23 下午 10:08	5.39 MB
webstation.sas7bdat	SAS Data Set	2006/9/25 上午 01:45	49.5 MB

At the bottom of the window, a status bar indicates "4 個項目".

Upzip EM_Data.zip to C:\DATA\EM_Data

電腦 > System (C:) > DATA > EM_Data

搜尋 EM_Data

組合管理 加入至媒體種 共用對象 新增資料夾

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

電腦

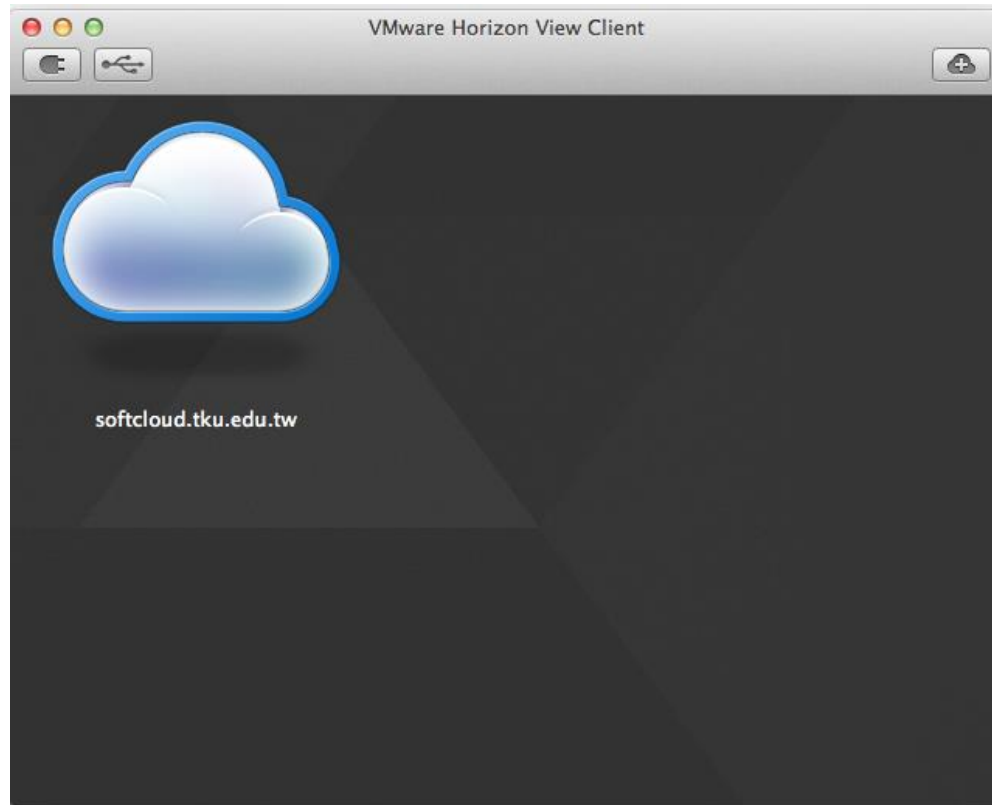
- System (C:)
- DATA
- EM_Data
- EM_Project1
- DataSources
- Meta
- Reports
- System
- Workspaces
- EM_Data.zip
- dyna
- eclipse-java-ke...
- ipcedu
- jdk-7u51-apidc...
- MSOCache
- PerfLogs
- Program Files
- ProgramData
- temp
- test-files - 2013
- WebEditor
- Windows

4 個項目

VMware Horizon View Client

softcloud.tku.edu.tw

SAS Enterprise Miner



SAS Locale Setup Manager

→ English UI

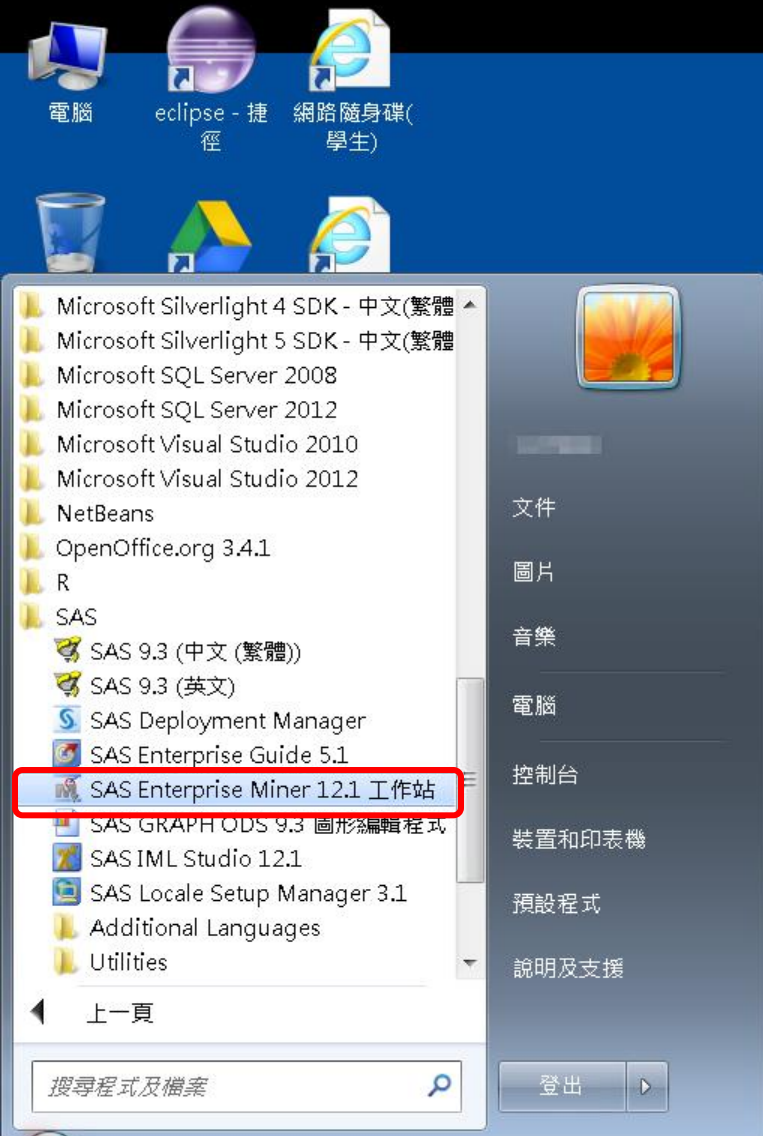
SAS Enterprise Guide 5.1 (SAS EG)



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

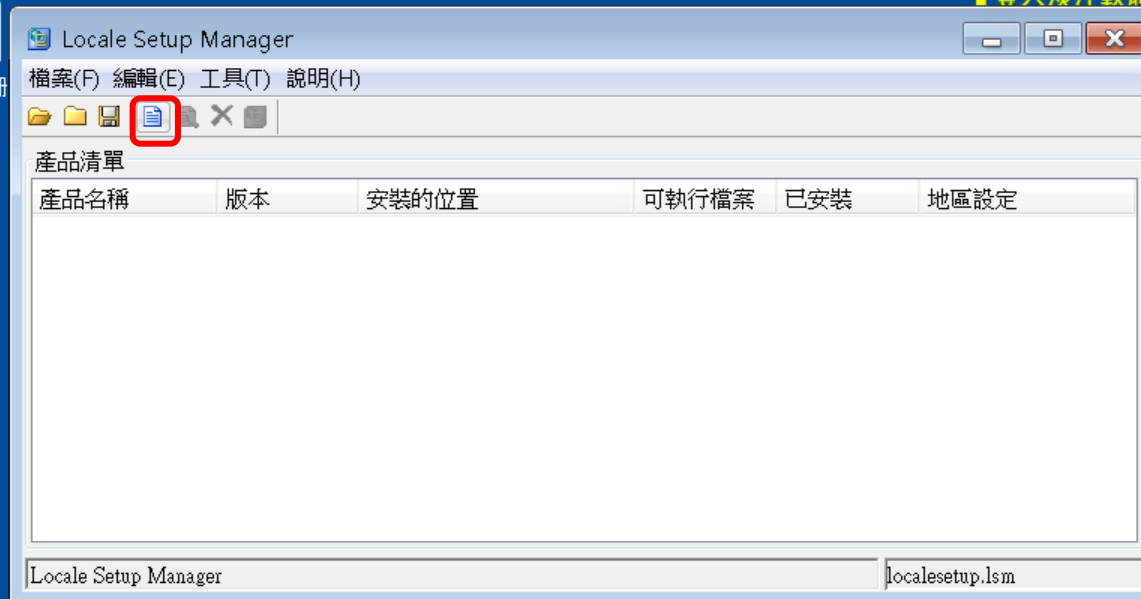


淡江軟體雲使用注意事項

- 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

- 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。
■ 產權，嚴禁安裝不法軟負相關法律責任。



C:\Program Files\SASHome\SASEnterpriseMinerWorkstationConfiguration\12.1\em.exe

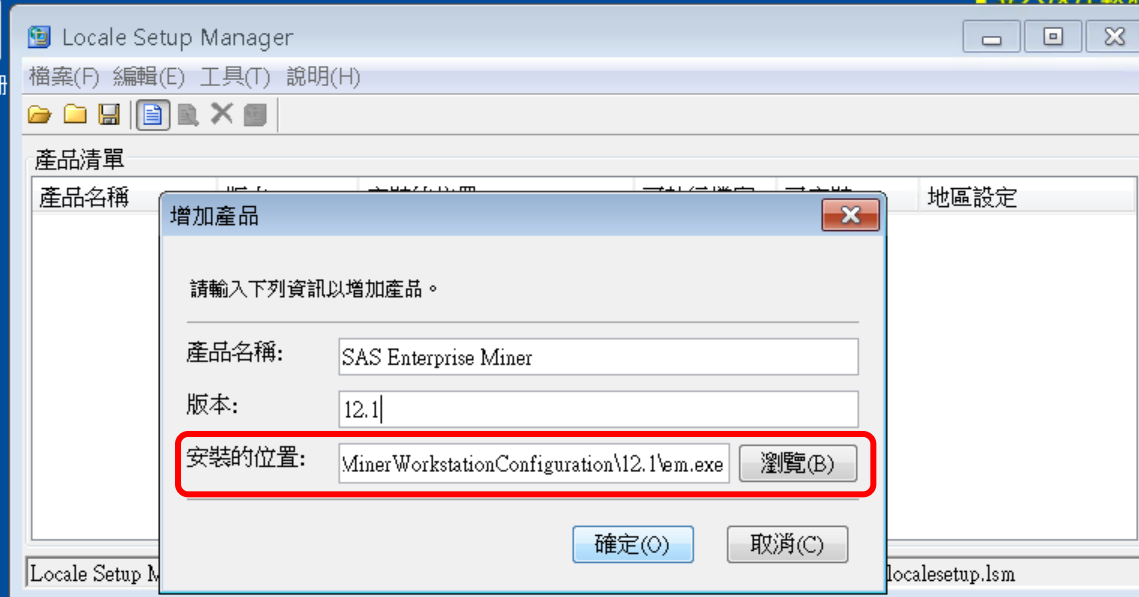
淡江軟體雲使用注意事項

- 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。

產權，嚴禁安裝不法軟負相關法律責任。



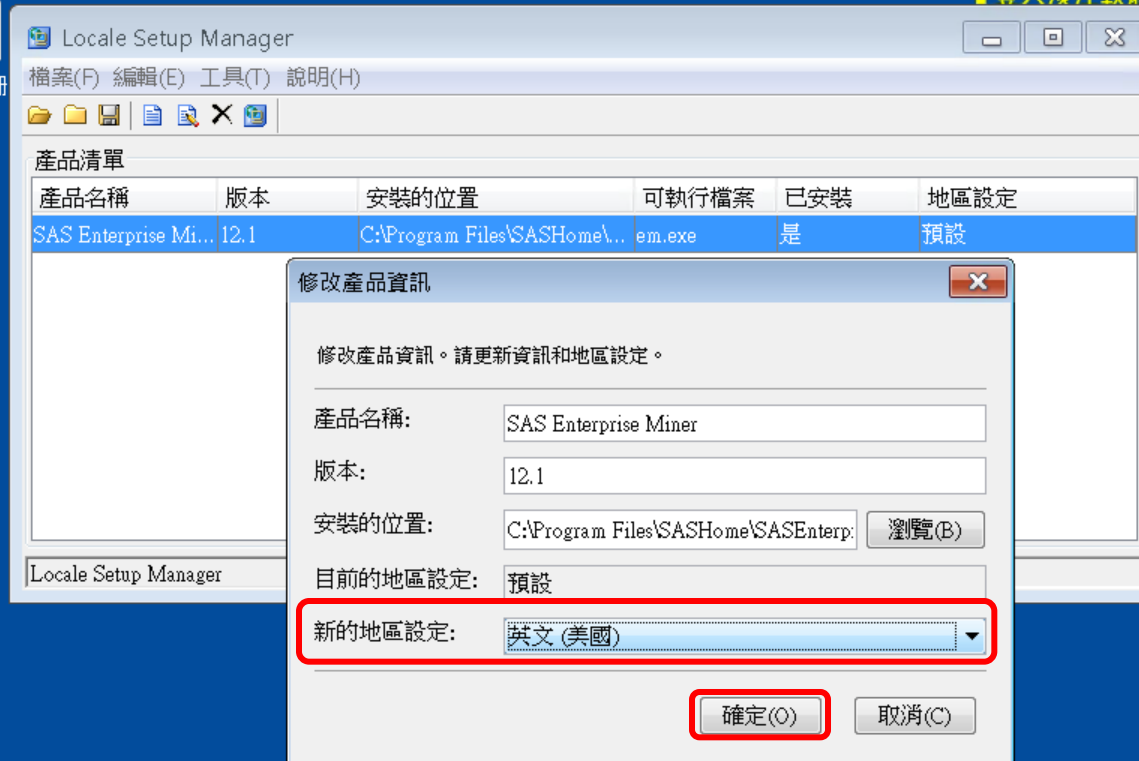
淡江軟體雲使用注意事項

- 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。

■ 產權，嚴禁安裝不法軟體，負相關法律責任。



淡江軟體雲使用注意事項

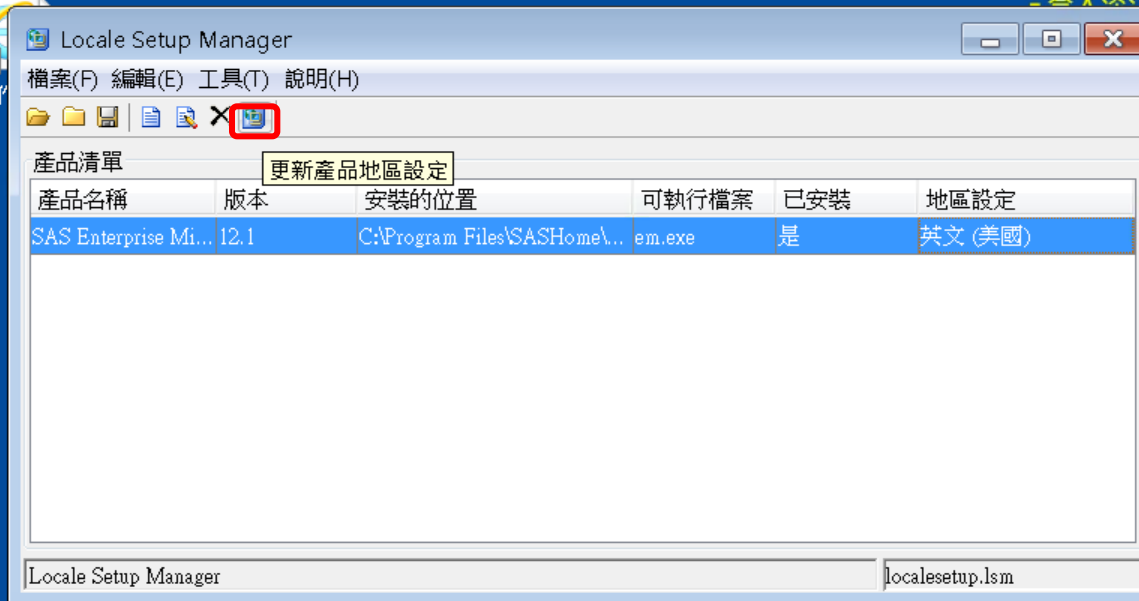
登入淡江軟體雲時，會自動產生一部新主機。登出雲端主機前，請先將資料備份於您攜帶的隨身碟、淡江大學網路隨身碟、Google雲端硬碟...等，登出後雲端主機的所有資料將會被清除，請回到您的個人主機畫面。

若您的主機帳號被冒用，每次使用完

請閱讀相關操作說明，請點選雲端主機畫面中的『操作手冊』圖示。

如有任何使用問題請至B212辦公室，或來電02-2656-2129 為您服務。

本系統保留智慧財產權，嚴禁安裝不法軟體，使用者自負相關法律責任。



淡江軟體雲使用注意事項

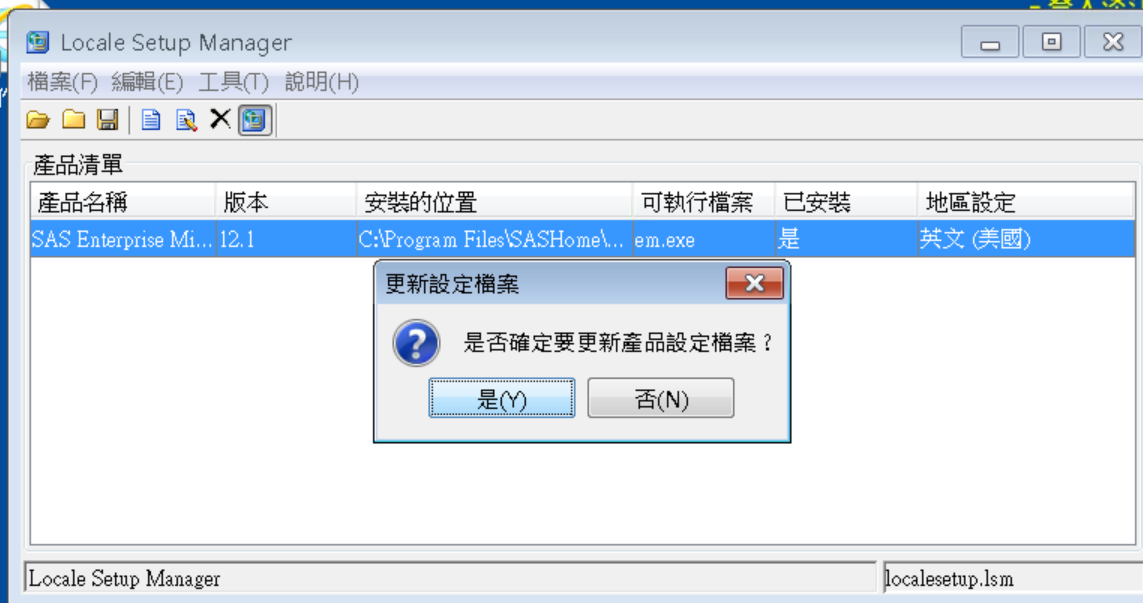
登入淡江軟體雲時，會自動產生一部新主機。登出雲端主機前，請先將位於您攜帶的隨身碟、淡江大學網路隨身碟、Google雲端硬碟...等，從雲端主機的所有資料將會被刪除，請回到您的個人主機畫面。

主機帳號被冒用，每次使用完

請參閱操作說明，請點選雲端主機「操作手冊」圖示。

如有任何問題請至B212辦公室,或來電02-2656-2129 為您服務。

請注意財產權，嚴禁安裝不法軟體，否則使用者自負相關法律責任。



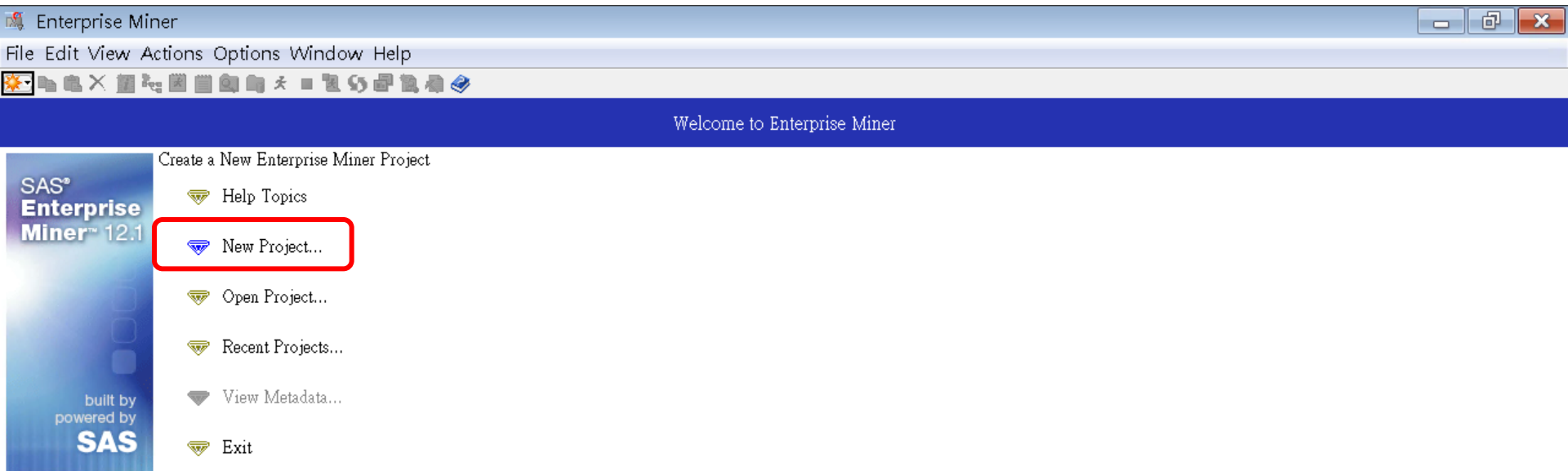
SAS Enterprise Miner 12.1 (SAS EM)



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

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

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

歡迎使用 SAS Enterprise Guide

選取這些選項的其中一個來開始:

開啟專案

- 更多專案...

新增

- 新增專案**
- 新增 SAS 程式
- 新增資料

輔助

- 教學課程: SAS Enterprise Guide 入門

不要再顯示此視窗

伺服器清單

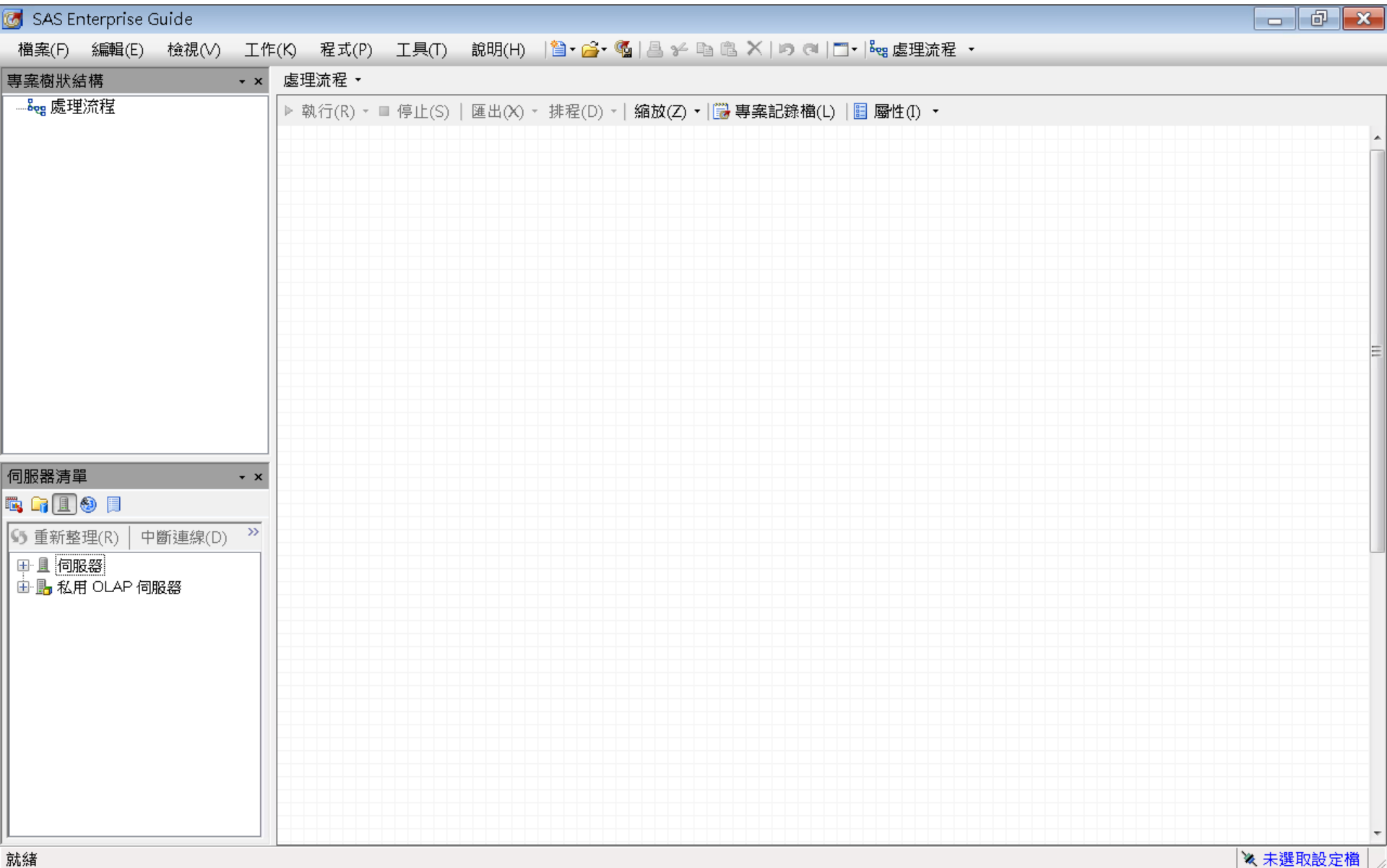
重新整理(R) | 中斷連線(D)

- 伺服器
- 私有 OLAP 伺服器

就緒

未選取設定檔

Open SAS Data File: Profile.sas7bdat



- 新增(N) ▾
- 開啟(O) ▾
- 關閉專案(C)
- 儲存「專案」(S) Ctrl+S
- 將「專案」另存為(A)...
- 儲存(V) Ctrl+Shift+S
- 另存新檔(E)...
- 匯入資料(I)
- 執行「專案」(U)
- 排程「專案」(H)
- 匯出(R) ▾
- 傳送至(D) ▾
- 發行(L)
- 版面設定(G)...
- 預覽列印(W)
- 列印(P)...
- Ctrl+P
- 專案屬性(J)
- 最近使用的專案(T) ▾
- 最近使用的程式(M) ▾
- 結束(X)

- 處理流程 ▾
- 專案(P) Ctrl+O
- 資料(D)...
- 資料勘查(X)...
- 程式(G)...
- 報表(T)...
- 預存程式(S)...
- OLAP Cube(U)...
- Information Map(I)...
- Exchange(E)...
- OLEDB(O)...
- ODBC(B)...
- 其他(R)...

專案樹狀結構

- 處理流程

- 伺服器
- 私有 OLAP 伺服器

開啟「資料」

瀏覽 搜尋 SAS 資料夾

查詢(I): EM_Data

名稱	類型	大小	修改日期
credit.sas7bdat	SAS Data Set	361 KB	2013/4/21 下午 0
inq2006.sas7bdat	SAS Data Set	13,729 KB	2013/5/29 下午 0
profile.sas7bdat	SAS Data Set	5,521 KB	2006/9/23 下午 3
webstation.sas7bdat	SAS Data Set	50,789 KB	2006/9/25 上午 0

檔案名稱(N): profile.sas7bdat

檔案類型(T): 所有已知的資料檔案 (*.sas7bdat;*.sas7bview;*.sd2;*.mdb;*.accdb;*.xls;*.xlsx;*.xlsm;*.xlsb;*.t

開啟(O) 取消

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

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

處理流程

- 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
000176820	107	3	2	1	113
000187784	54	14	7	24	99
000189200	92	3	2	2	99
000198327	44	27	17	3	91
000200237	146	3	2	1	152
000207980	39	5	2	4	50
000227355	180	3	2	10	195
000228317	46	14	2	1	63
000228906	25	3	2	1	31
000229984	15	3	2	1	21
000241022	13	3	2	1	19
000253260	65	3	2	1	71
000257910	35	3	2	1	41
000261717	11	5	2	1	19
000266600	15	3	2	1	21
000269176	107	78	22	7	214
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
 - SASHELP
 - SASUSER
 - WORK
 - 檔案
 - 私用 OLAP 伺服器

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

「profile」的屬性

一般

profile

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

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

「profile」的屬性

一般
欄
進階
摘要

欄

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)

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。

其他 (F1)...

確定 取消

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

專案樹狀結構 profile

處理流程 profile

- 開啟「profile」(O)
- 匯出(X)
 - 匯出「profile」(X)...
 - 匯出「profile」作為專案中的步驟(S)...
- 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
000041260	34	3	3	9	49
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
000176820	107	3	2	1	113
000187784	54	14	7	24	99
000189200	92	3	2	2	99
000198327	44	27	17	3	91
000200237	146	3	2	1	152
000207980	39	5	2	4	50
000227355	180	3	2	10	195
000228317	46	14	2	1	63
000228906	25	3	2	1	31
000229984	15	3	2	1	21
000241022	13	3	2	1	19
000253260	65	3	2	1	71
000257910	35	3	2	1	41
000261717	11	5	2	1	19
000266600	15	3	2	1	21
000269176	107	78	22	7	214
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
 - SASHELP
 - SASUSER
 - WORK
 - 檔案
 - 私用 OLAP 伺服器

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
inq2006.sas7bdat	2013/5/29 下午 01:20	SAS
profile.sas7bdat	2006/9/23 下午 10:08	SAS
webstation.sas7bdat	2006/9/25 上午 01:45	SAS

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

專案樹狀結構

- 處理流程
 - profile

profile

篩選和排序(L)

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

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
 - 私有 OLAP 伺服器

儲存位置(I): EM_Data

- 桌面
- 我的文件
- 我的電腦
- 網路上的芳鄰
- 伺服器

名稱	修改日期	類型
credit.sas7bdat	2013/4/21 下午 05:22	SAS
inq2006.sas7bdat	2013/5/29 下午 01:20	SAS
profile.sas7bdat	2006/9/23 下午 10:08	SAS
webstation.sas7bdat	2006/9/25 上午 01:45	SAS

檔案名稱(N):

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

儲存(S) 取消

SAS Enterprise Guide

檔案(F) 編輯(E) 檢視(V) 工作(K) 程式(P) 工具

專案樹狀結構

- 處理流程
 - profile

篩選和排序(L)

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

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
 - 私人 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。

檔案名稱(N): profile_excel

檔案類型(T): Microsoft Excel 活頁簿 (*.xlsx)

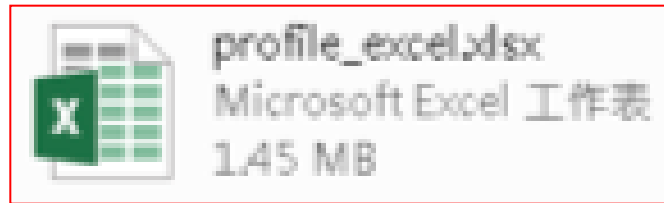
儲存(S) 取消

Export SAS .sas7bdat to to Excel .xlsx File

The screenshot displays the SAS Enterprise Guide interface. The main window shows a data table with columns: ID, CNT_TBM, CNT_ATM, CNT_POS, CNT_CSC, and CNT_TOT. A dialog box titled "Enterprise Guide" is overlaid on the table, stating "正在匯出「profile」..." (Exporting 'profile'...). On the right side, a file explorer shows the output files: "profile.sas7bdat SAS Data Set 5.39 MB" and "profile_excel.xlsx Microsoft Excel 工作表 1.45 MB". A red arrow points from the SAS file to the Excel file. The status bar at the bottom indicates "正在執行: 匯出工作 - 正在執行..." (Running: Export job - running...).

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	50	20	2	11	85
12	001147896		8	1	150
13	001150206		80	4	292
14	001160501		42	12	117
15	001176820		2	1	113
16	001187784	54	14	7	99
17	001189200	92	3	2	99
18	001198327	44	27	17	91
19	00200237	146	3	2	152
20	00207980	39	5	2	50
21	00227355	180	3	2	195
22	00228317	46	14	2	63
23	00228906	25	3	2	31
24	00229984	15	3	2	21
25	00241022	13	3	2	19
26	00253260	65	3	2	71
27	00257910	35	3	2	41
28	00261717	11	5	2	19
29	00266600	15	3	2	21
30	00269176	107	78	22	214
31	00273510	110	15	6	152

Import Excel File to SAS EG



Import Excel File to SAS EG

The screenshot displays the SAS Enterprise Guide interface. On the left, the '專案樹狀結構' (Project Tree) shows a '處理流程' (Process) and a '伺服器清單' (Server List) containing '本機' (Local) and '私用 OLAP 伺服器' (Private OLAP Server). The main window is a file explorer titled '開啟' (Open) showing the 'EM_Data' folder. A table lists the contents:

名稱	類型	大小	修改日期
profile_excel.xlsx	Microsoft Excel 工作表	1,491 KB	2014/3/

At the bottom of the file explorer, the '檔案名稱(N):' field contains 'profile_excel.xlsx' and the '檔案類型(T):' dropdown is set to '可匯入的資料檔案 (*.xls;*.xlsx;*.xlsm;*.xlsb;*.mdb;*.accdb;*.txt;*.csv;*.asc;*.tab;*.htm;*.html)'. The status bar at the bottom left shows '就緒' (Ready) and the bottom right shows '未選取設定檔' (No profile selected).

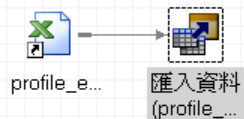
專案樹狀結構

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(O)



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

伺服器清單

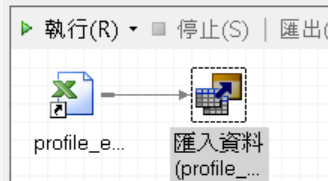
重新整理(R) 中斷連線(D)

伺服器

- 本機
 - 資料館
 - 檔案
- 私用 OLAP 伺服器

處理流程

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



重新整理(R) | 中斷連線(D)

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

從 profile_excel.xlsx 匯入資料

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)

使用預先定義的命名範圍(P)

專案樹狀結構

處理流程

- 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(O)

profile_e... → 匯入資料 (profile_...)

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

從 profile_excel.xlsx 匯入資料

3 / 4 定義欄位特性



選取欄並定義特性(S):

Inc	來源名稱	名稱	標籤	類型	來源格式	長度	輸出格式	輸出輸入格式
<input checked="" type="checkbox"/>	ID	ID	ID	字串	\$CHAR9.	9	\$CHAR9.	\$CHAR9.
<input checked="" type="checkbox"/>	CNT_TBM	CNT_TBM	CNT_TBM	數字	BEST12.	8	BEST12.	BEST12.
<input checked="" type="checkbox"/>	CNT_ATM	CNT_ATM	CNT_ATM	數字	BEST12.	8	BEST12.	BEST12.
<input checked="" type="checkbox"/>	CNT_POS	CNT_POS	CNT_POS	數字	BEST12.	8	BEST12.	BEST12.
<input checked="" type="checkbox"/>	CNT_CSC	CNT_CSC	CNT_CSC	數字	BEST12.	8	BEST12.	BEST12.
<input checked="" type="checkbox"/>	CNT_TOT	CNT_TOT	CNT_TOT	數字	BEST12.	8	BEST12.	BEST12.

全選(A) 全部清除(C) 修改(M)...

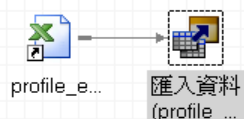
<上一步(B) 下一步(N) > 完成(F) 取消 說明

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(O)



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

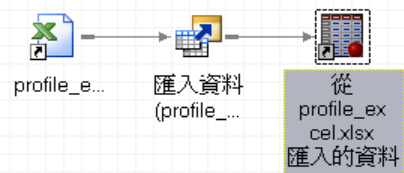
重新整理(R) | 中斷連線(D)

- 伺服器
- 本機
 - 資料館
 - 檔案
- 私有 OLAP 伺服器

專案樹狀結構

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

處理流程



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
 - 私有 OLAP 伺服器

專案樹狀結構

處理流程

- profile_excel.xlsx
 - 匯入資料 (profile_excel.xls)

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

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
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

專案樹狀結構

匯入資料 (profile_excel.xlsx[profile])

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

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_TOI
99972	875475072	81	37	2	1	121
99973	875478517	56	15	10	3	84
99974	875481914	49	28	91	87	255
99975	875484829	233	10	5	1	249
99976	875488178	104	3	2	1	110
99977	875505005	27	47	37	53	164
99978	875508051	28	3	2	1	34
99979	875511290	13	5	2	1	21
99980	875514158	54	8	18	4	84
99981	875528729	71	4	14	1	90
99982	875536451	109	3	2	11	125
99983	875555772	26	22	2	1	51
99984	875568109	124	15	2	12	153
99985	875577833	115	80	9	5	209
99986	875589010	48	3	82	8	141
99987	875593202	143	9	2	4	158
99988	875594835	51	3	2	1	57
99989	875599149	58	15	27	17	117
99990	875606621	25	3	2	1	31
99991	875619290	103	6	2	10	121
99992	875625622	10	4	2	1	17
99993	875625635	41	3	2	1	47
99994	875628606	20	14	23	1	58
99995	875636141	132	3	2	11	148
99996	875638608	24	6	2	1	33
99997	875656523	79	9	2	7	97
99998	875663288	44	111	57	9	221
99999	875670226	10	5	2	1	18
100000	875674475	128	18	20	1	167

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
 私有 OLAP 伺服器

Export Excel .xlsx File to SAS .sas7bdat File

The screenshot displays the SAS Enterprise Guide interface. The main workspace shows a workflow with three steps: 'profile_e...', '匯入資料 (profile_...', and '從 profile_excel.xlsx 匯入的資料'. A context menu is open over the third step, with the '匯出(X)' option selected. A secondary menu is also visible, showing options like '匯出「從 profile_excel.xlsx 匯入的資料」(X)...' and '匯出「從 profile_excel.xlsx 匯入的資料」作為專案中的步驟(S)...'. The left sidebar contains a '專案樹狀結構' (Project Structure) pane and a '伺服器清單' (Server List) pane. The status bar at the bottom indicates '就緒' (Ready) and '未選取設定檔' (No profile selected).

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_excel.xlsx

匯入資料 (profile_excel.xlsx[profi

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

profile_e... 匯入資料 (profile_... 從 profile_excel.xlsx 匯入的資料

匯出「從 profile_excel.xlsx 匯入的資料」(X)...

匯出「從 profile_excel.xlsx 匯入的資料」作為專案中的步驟(S)...

開啟(O)

匯出(X)

篩選和排序(F)...

查詢產生器(Q)...

從「從 profile_excel.xlsx 匯入的資料」執行分支(B)

發行(L)...

傳送到(S)

將「從 profile_excel.xlsx 匯入的資料」連結至(K)...

複製(C)

貼上(P)

刪除(D)

重新命名(M)

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

檔案

私有 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...

專案樹狀結構

處理流程

處理流程

- profile_excel.xlsx
 - 匯入資料 (profile_excel.xlsx[profi

執行(R) 停...

profile_e... 匯 (pr

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

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	SA...
inq2006.sas7bdat	2013/5/29 下午 01:20	SA...
profile.sas7bdat	2006/9/23 下午 10:08	SA...
webstation.sas7bdat	2006/9/25 上午 01:45	SA...

檔案名稱(N): Profile_SAS

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

儲存(S) 取消

電腦 > System (C:) > DATA > EM_Data

搜尋 EM_Data

組合管理 使用 SAS Enterprise Guide 5.1 開啟 新增資料夾

我的最愛
下載
桌面
最近的位置
媒體櫃
文件
音樂
視訊
圖片
電腦
System (C:)
DATA
dyna
eclipse-java-kepler-SR1-win32
ipcedu
jdk-7u51-apidocs
MSOCache
PerfLogs
Program Files
ProgramData
temp
test-files - 20130725

名稱	類型	大小	修改日期
credits.sas7bdat	SAS Data Set	361 KB	2013/4/2...
inq2006.sas7bdat	SAS Data Set	13,729 KB	2013/5/2...
profile.sas7bdat	SAS Data Set	5,521 KB	2006/9/2...
profile_excel.xlsx	Microsoft Excel 工作表	1,491 KB	2014/3/1...
profile_sas.sas7bdat	SAS Data Set	5,521 KB	2014/3/1...
webstation.sas7bdat	SAS Data Set	50,789 KB	2006/9/2...

選取了 3 個項目 修改日期: 2006/9/23 下午 10:08 ... 建立日期: 2014/3/19 上午 08:43
大小: 12.2 MB

Windows Explorer window showing the directory path: 電腦 > System (C:) > DATA > EM_Data. The search bar contains "搜尋 EM_Data".

Navigation pane (left):

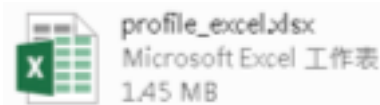
- 我的最愛
- 下載
- 桌面
- 最近的位置
- 媒體櫃
 - 文件
 - 音樂
 - 視訊
 - 圖片
- 電腦
 - System (C:)
 - DATA (selected)
 - dyna
 - eclipse-java-kepler-SR1-win32
 - ipcedu
 - jdk-7u51-apidocs
 - MSOCache
 - PerfLogs
 - Program Files
 - ProgramData
 - temp
 - test-files - 20130725

Main pane (right) showing files:

File Name	Type	Size
credit.sas7bdat	SAS Data Set	361 KB
profile.sas7bdat	SAS Data Set	5.39 MB
profile_sas.sas7bdat	SAS Data Set	5.39 MB
inq2006.sas7bdat	SAS Data Set	13.4 MB
profile_excel.xlsx	Microsoft Excel 工作表	1.45 MB
webstation.sas7bdat	SAS Data Set	49.5 MB

Bottom status bar: 6 個項目

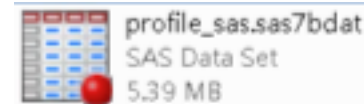
Profile_Excel.xlsx



Microsoft Excel interface showing the ribbon (File, Home, Insert, Layout, Formulas, Data, Review, View, Help) and the spreadsheet data. The spreadsheet has columns A through C and rows 1 through 21. The data is as follows:

	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	4	50									

Profile_SAS.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
profile_sas

伺服器清單

重新整理(R) | 中斷連線(D)

伺服器
私有 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	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

就緒

未選取設定檔

SAS Enterprise Miner 12.1 (SAS EM)



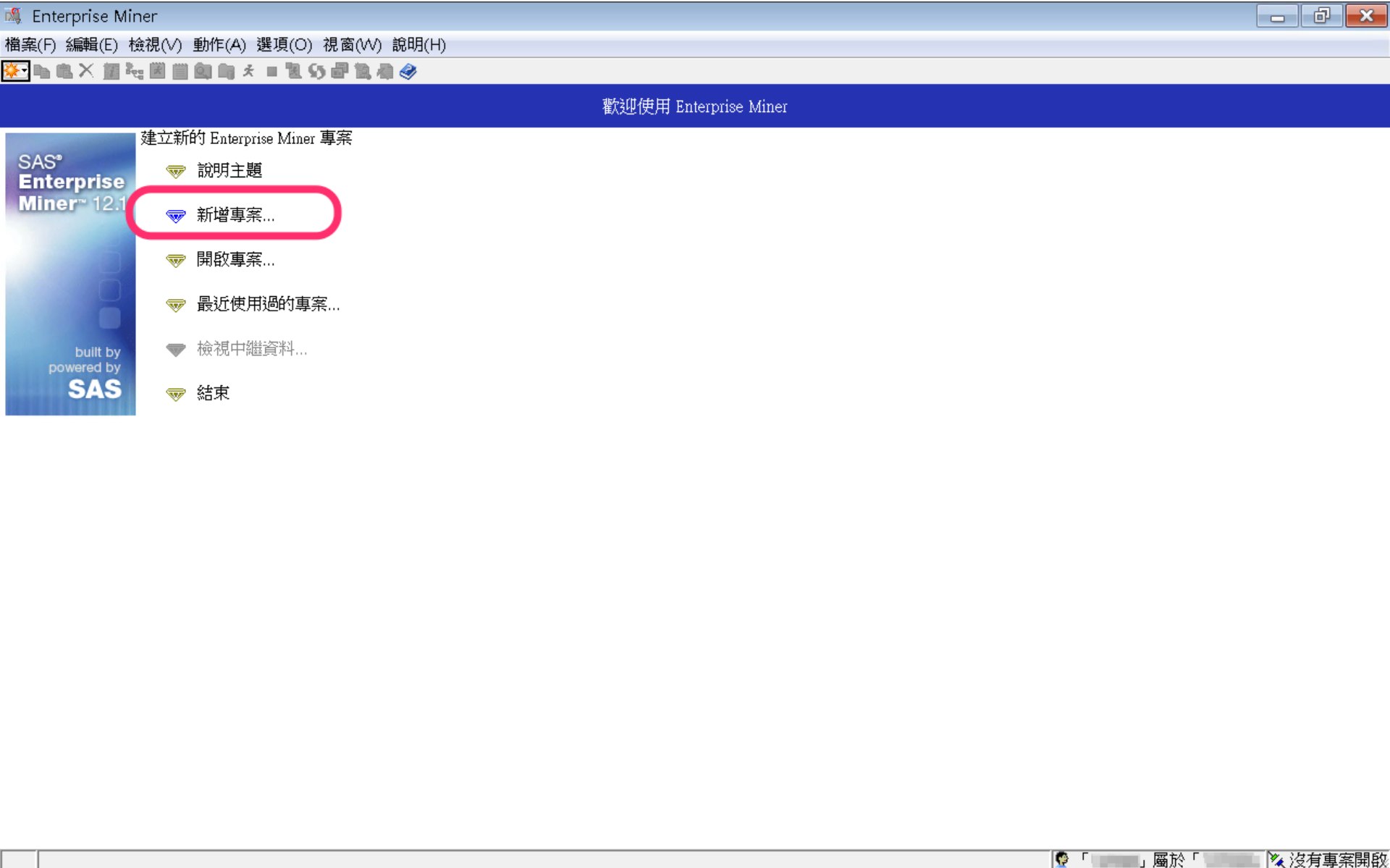
淡江軟體雲使用注意事項

- 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)

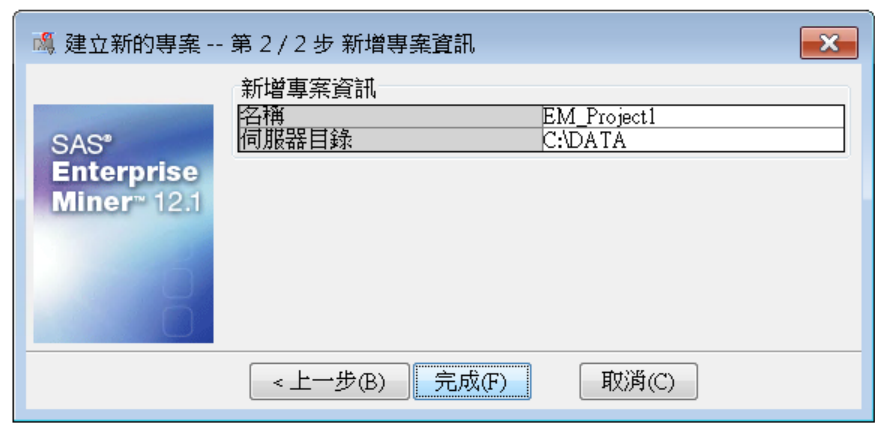
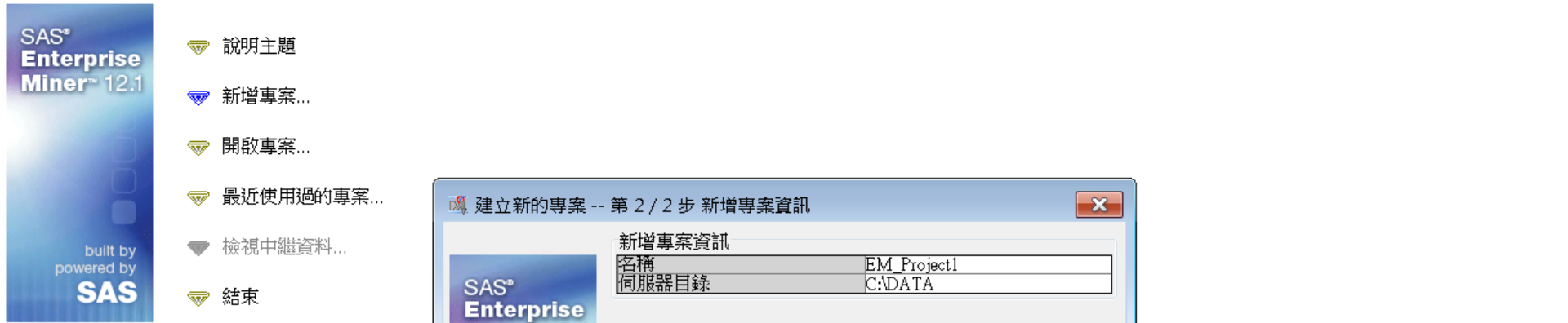
The screenshot displays the SAS Enterprise Miner 12.1 application window. The title bar reads "Enterprise Miner" and the menu bar includes "檔案(F)", "編輯(E)", "檢視(V)", "動作(A)", "選項(O)", "視窗(W)", and "說明(H)". The main window has a blue header with the text "歡迎使用 Enterprise Miner". On the left side, there is a vertical sidebar with the SAS Enterprise Miner 12.1 logo and a list of options: "說明主題", "新增專案...", "開啟專案...", "最近使用過的專案...", "檢視中繼資料...", and "結束".

In the center, a dialog box titled "建立新的專案 -- 第 1 / 2 步 指定專案名稱和伺服器目錄" is open. The dialog contains the following text and fields:

- Text: "在 SAS 伺服器上為此專案指定專案名稱和目錄。所有的 SAS 資料集和檔案都將寫入此位置。"
- Field: "專案名稱" with the value "EM_Project1".
- Field: "SAS 伺服器目錄" with the value "C:\DATA" and a "瀏覽(B)" button.
- Buttons: "< 上一步(B)", "下一步(N) >", and "取消(C)".

The taskbar at the bottom right shows the system tray with the text "沒有專案開啟".

Step 1. 新增專案 (New Project)



SAS Enterprise Miner (EM_Project1)

The screenshot displays the SAS Enterprise Miner interface for a project named 'EM_Project1'. The window title is 'Enterprise Miner - EM_Project1'. The menu bar includes '檔案(F)', '編輯(E)', '檢視(V)', '動作(A)', '選項(O)', '視窗(W)', and '說明(H)'. The toolbar contains various icons for file operations and project management. The left sidebar shows a tree view with 'EM_Project1' expanded, containing '資料來源', '流程圖', and '模型套件'. Below this is a table of project attributes.

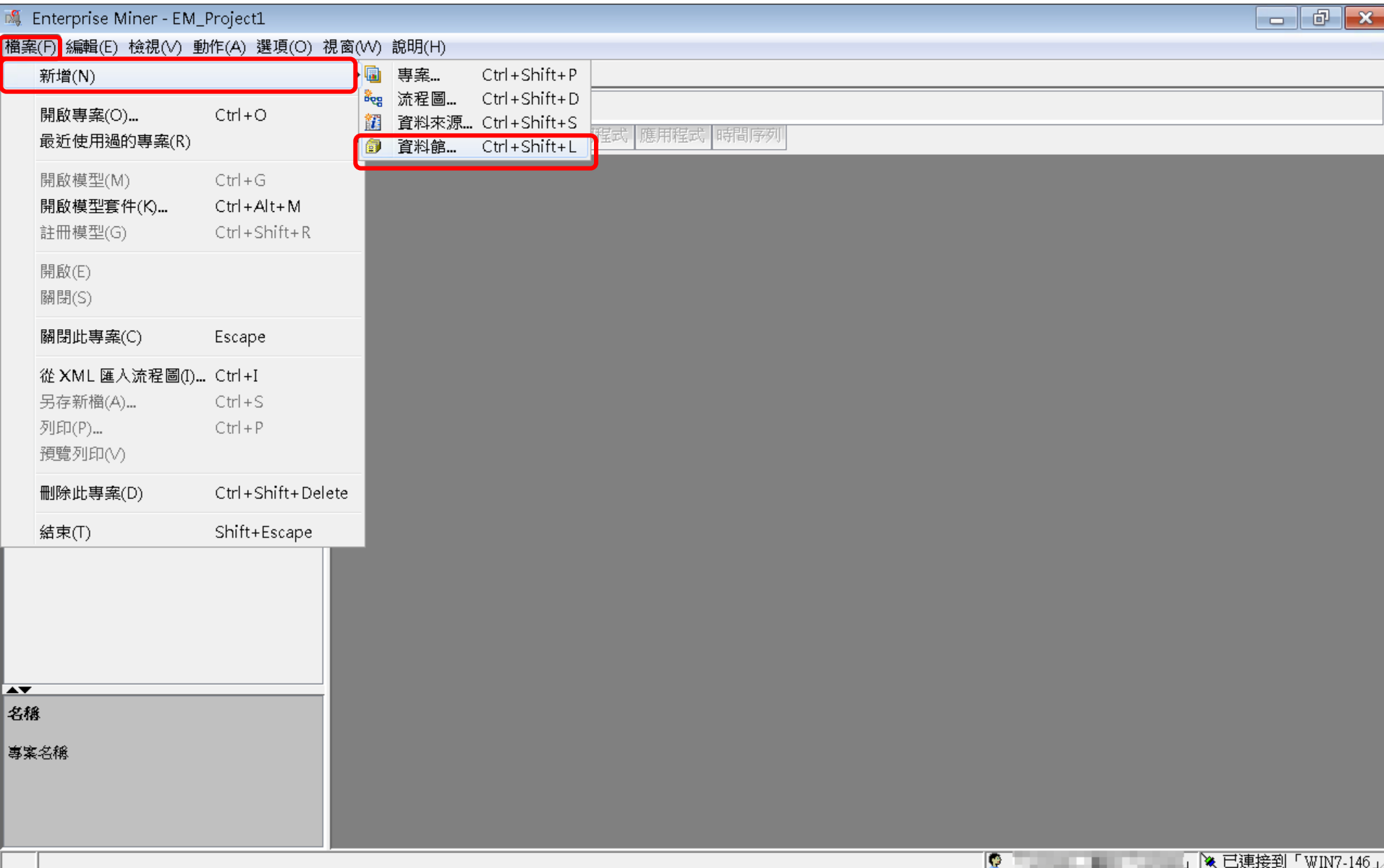
.. 屬性	值
名稱	EM_Project1
專案啟動程式碼	
專案巨集變數	
建立時間	2014/3/18 下午 10:38
伺服器	
網格可用	否
路徑	C:\DATA
中繼資料資料夾路徑	
最大並行工作數目	預設

Below the table is a section for naming the project:

名稱
專案名稱

The main workspace area is currently empty, with a toolbar at the top containing icons for '樣本', '動查', '修改', '模型', '評估', '公用程式', '應用程式', and '時間序列'.

Step 2. 新增資料館 (New / Library)



Step 2. 新增資料館 (New / Library)

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 調查 修改 模型 評估 公用程式 應用程式 時間序列

資料館精靈 -- 第 1 / 3 步 選取動作

請選取動作

- 建立新的資料館(R)
- 修改資料館(M)
- 刪除資料館(D)

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 10:3
伺服器	
網路可用	否
路徑	C:\DATA
中繼資料資料夾路徑	
最大並行工作數目	預設

< 上一步(B) 下一步(N) > 取消(C)

已連接到「WIN7-146」

Step 2. 新增資料館 (New / Library)

The screenshot shows the Enterprise Miner interface with the '資料館精靈' (Library Wizard) dialog box open. The dialog is titled '資料館精靈 -- 第 2 / 3 步 建立或修改' (Library Wizard -- Step 2 / 3 Create or Modify). It contains the following fields and controls:

- 名稱 (Name):** A text box containing 'EM_Lib'.
- 引擎 (Engine):** A dropdown menu set to 'BASE'.
- 資料館資訊 (Library Information):**
 - 路徑 (Path):** A text box containing 'C:\DATA\EM_Data' with a '瀏覽(R)...' (Browse...) button to its right.
 - 選項 (Options):** An empty text box.
- Navigation:** Buttons for '< 上一步(B)' (Previous Step), '下一步(N) >' (Next Step), and '取消(C)' (Cancel).

In the background, the Enterprise Miner main window is visible, showing a project named 'EM_Project1' and a table of properties:

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 10:3
伺服器	
網絡可用	否
路徑	C:\DATA
中繼資料資料夾路徑	
最大並行工作數目	預設

Step 2. 新增資料館 (New / Library)

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 調查 修改 模型 評估 公用程式 應用程式 時間序列

開啟

查看: EM_Data

最近的项目

- credit.sas7bdat
- inq2006.sas7bdat
- profile.sas7bdat
- webstation.sas7bdat

桌面

我的文件

電腦

網路

檔案名稱: C:\DATA\EM_Data

檔案類型: 所有檔案

開啟(O)

取消 開啟選取的檔案

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 10:30
伺服器	
網絡可用	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)

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 動查 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	

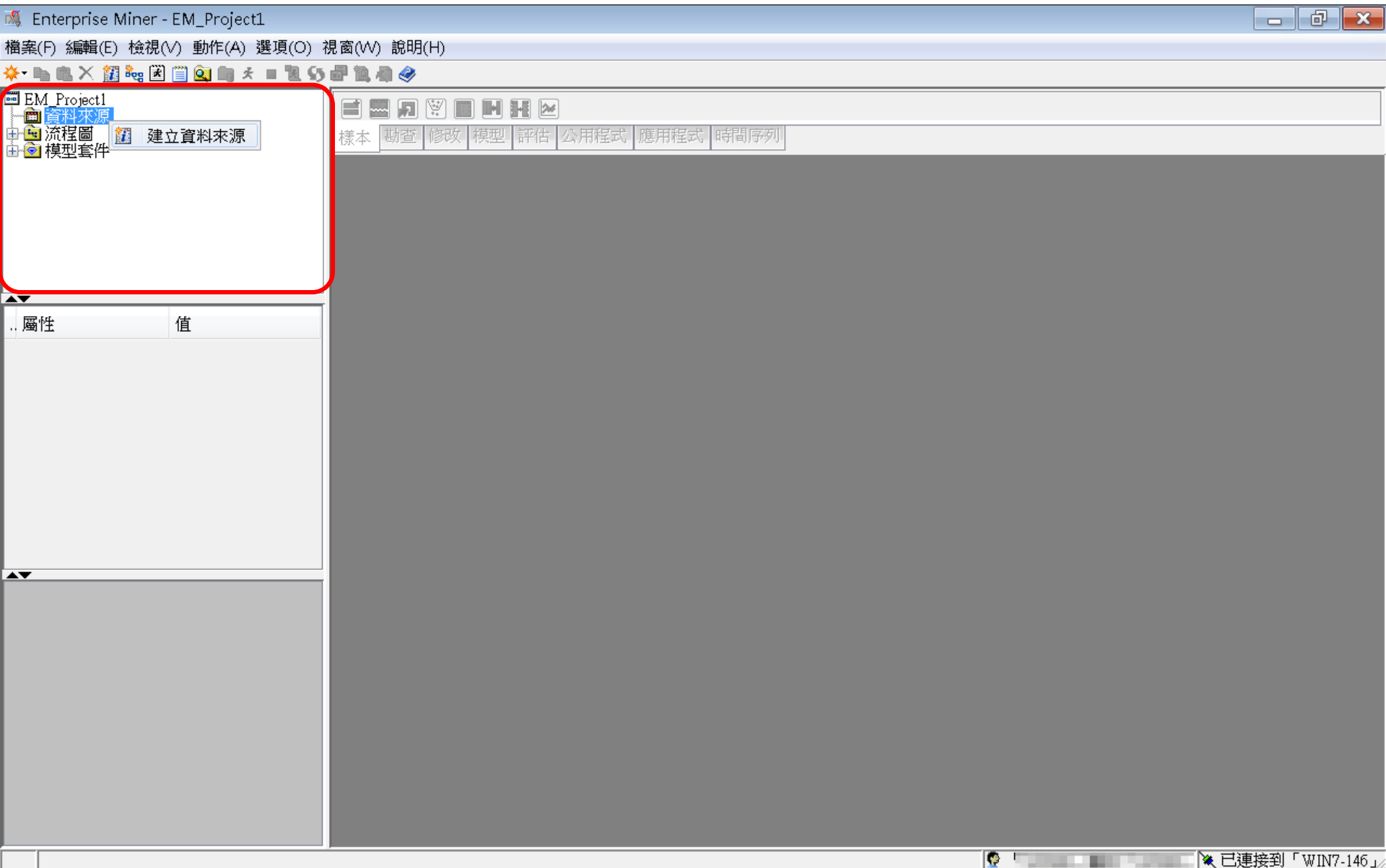
狀態
動作成功!
資料館「EM_Lib」已建立。

< 上一步(B) 完成(F)

名稱
專案名稱

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 1 / 8 步 中繼資料來源

選取中繼資料來源

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 選擇 SAS 表格

選擇 SAS 表格

表格(T): 瀏覽(R)...

< 上一步(B) 下一步(N) > 取消(C)

.. 屬性 值

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 調查 修改 模型 評估 公用程式 應用程式 時間序列

選擇 SAS 表格

SAS 資料館	名稱	引擎	路徑
Em_lib	Em_lib	BASE	C:\DATA\EM_Data
Maps	Maps	V9	C:\Program Files\SASHome\SASFo...
Mapsgfk	Mapsgfk	V9	C:\Program Files\SASHome\SASFo...
Mapssas	Mapssas	V9	C:\Program Files\SASHome\SASFo...
Sampsio	Sampsio	V9	C:\Program Files\SASHome\SASFo...
Sashelp	Sashelp	V9	C:\Program Files\SASHome\SASFo...
Sasuser	Sasuser	V9	C:\Users\..... Documents\My SA...
Work	Work	V9	C:\Users\..... AppData\Local\Te...

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

瀏覽(R)...

資料來源

屬性 值

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 調查 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):

選取 SAS 表格

名稱	類型
Credit	表格
Inq2006	表格
Profile	表格
Webstation	表格

SAS 資料館

- Em lib
- Maps
- Mapsgfk
- Mapssas
- Sampsio
- Sashelp
- Sasuser
- Work

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 選擇 SAS 表格

選擇 SAS 表格

EM_LIB.PROFILE

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)



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 動查 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」

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 調查 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	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) 調查(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)

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 調查 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 / 8 步 摘要

中繼資料已完成。

資料館: EM_LIB
資料來源: PROFILE
角色: 原始

角色	層級	計數
ID	名目	1
輸入	間隔	5

< 上一步(B) 完成(F) 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
 - PROFILE**
- 流程圖
- 模型套件

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	

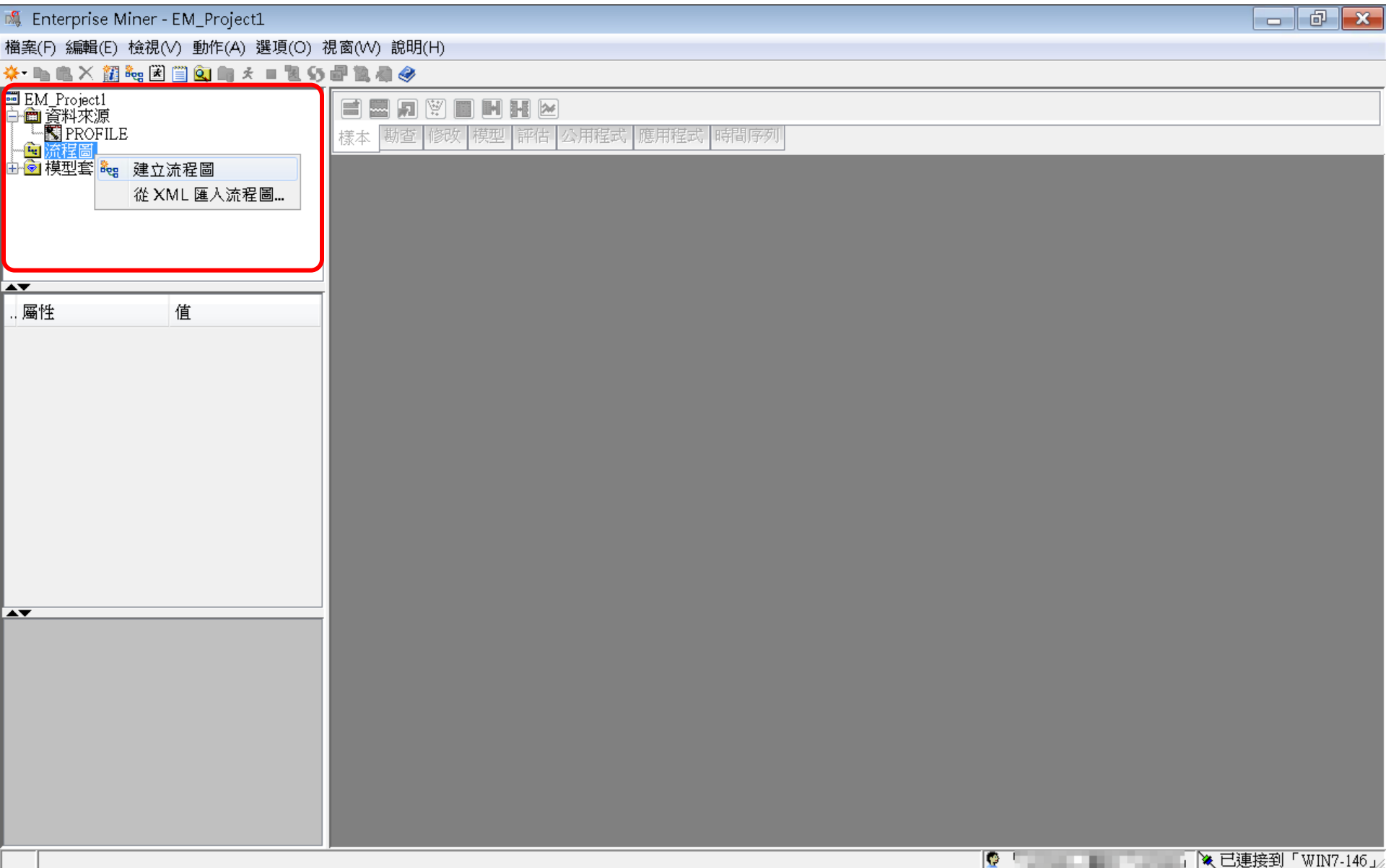
ID

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。

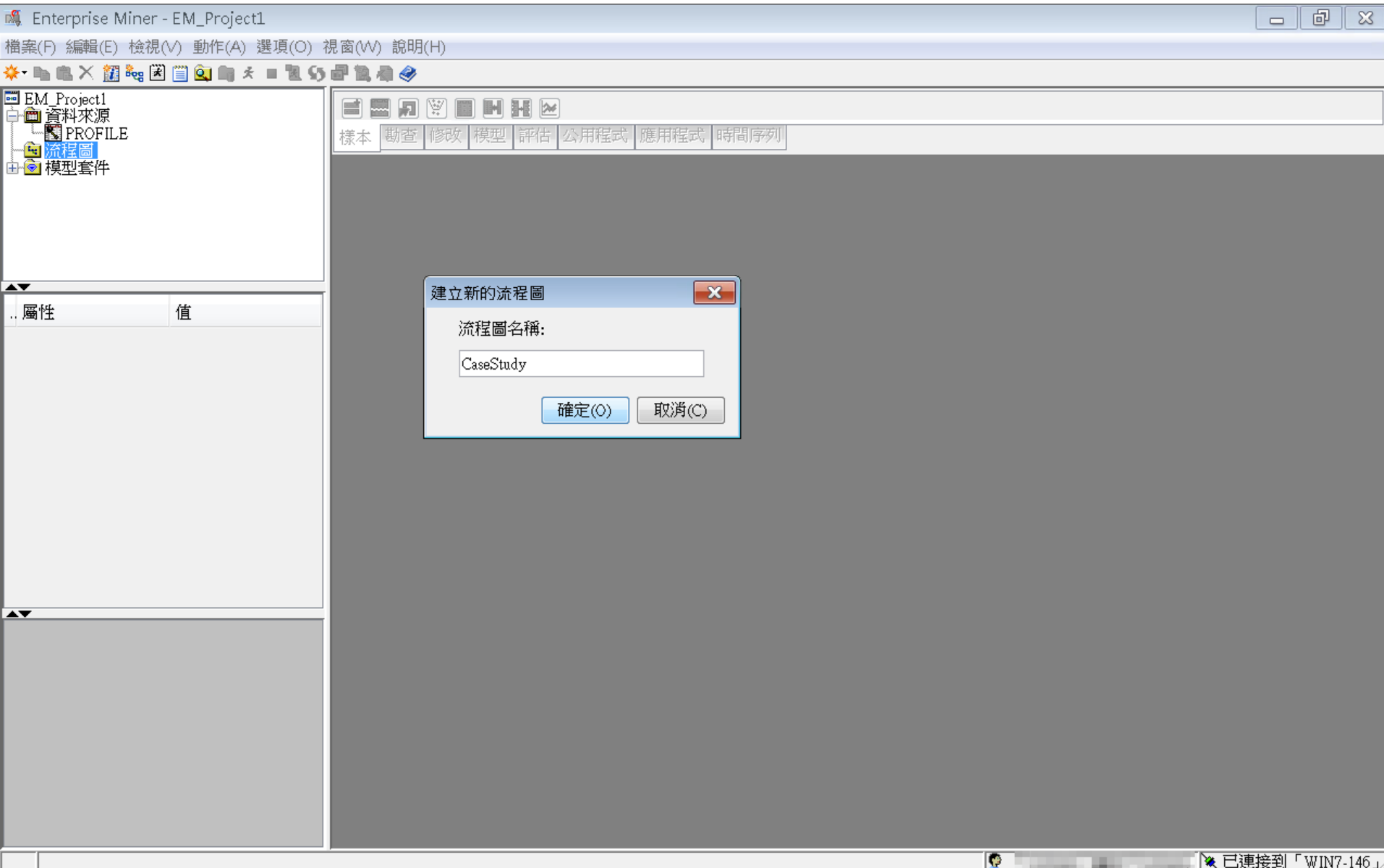
樣本 動查 修改 模型 評估 公用程式 應用程式 時間序列

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)

The screenshot displays the SAS Enterprise Miner interface. The main window is titled 'Enterprise Miner - EM_Project1'. The menu bar includes '檔案(F)', '編輯(E)', '檢視(V)', '動作(A)', '選項(O)', '視窗(W)', and '說明(H)'. The toolbar contains various icons for file operations and editing. The left sidebar shows a project tree with 'EM_Project1' expanded, containing '資料來源', 'PROFILE', '流程圖', 'CaseStudy', and '模型套件'. Below the tree is a table of properties for the selected 'CaseStudy' object.

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

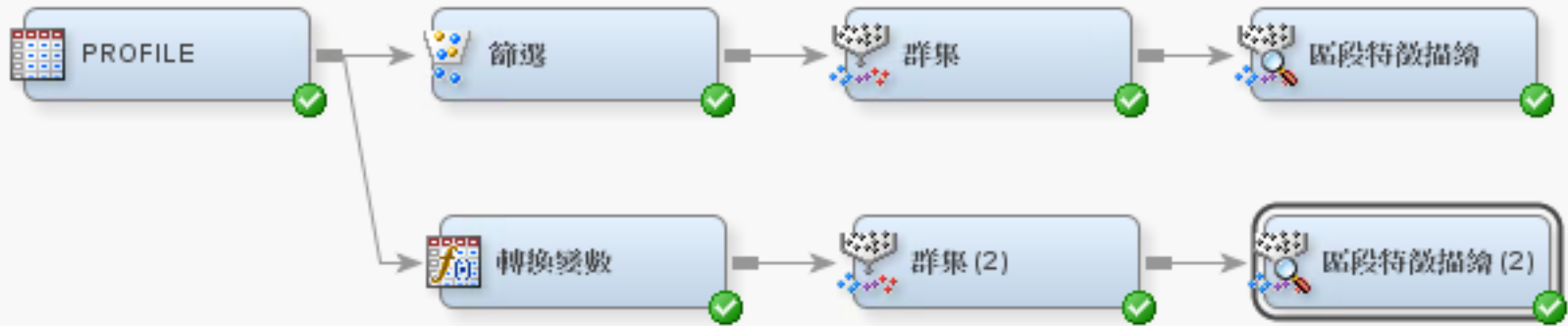
Below the table, there is a section for 'ID' with a description: '流程圖識別碼。此識別碼和 SAS 資料館名稱對應，SAS 資料館名稱是用於識別此流程圖內容在伺服器上的實體位置。'

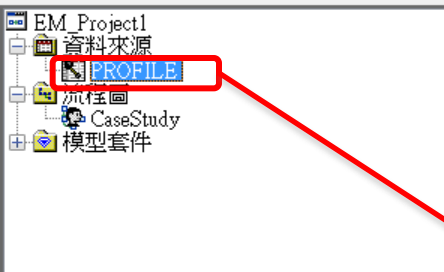
The main workspace is titled 'CaseStudy' and contains a blank area for creating a flowchart. The toolbar above the workspace includes icons for '樣本', '勘查', '修改', '模型', '評估', '公用程式', '應用程式', and '時間序列'. The status bar at the bottom shows '流程圖' and '日誌' dropdown menus, and a system tray icon indicating '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

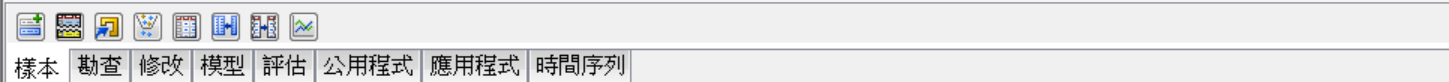




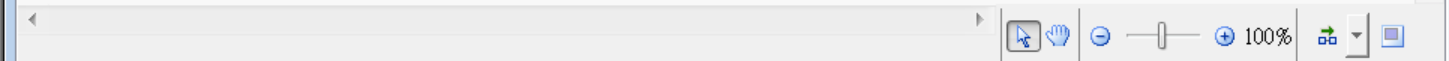
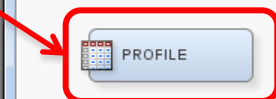
屬性	值
ID	profile
名稱	PROFILE
變數	
決策	
角色	原始
附註	
資料館	EM_LIB
表格	PROFILE
樣本資料集	
大小類型	
樣本大小	
類型	DATA
觀測數目	100000
欄數	6
位元組數	5653504
區段	

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。



CaseStudy



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



EM_Project1
 資料來源
 PROFILE
 流程圖
 CaseStudy
 模型套件

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。



樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

篩選

根據指定的準則將資料從觀測中移除。



流程圖 | 日誌



EM_Project1

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

CaseStudy

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
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

一般

一般屬性

Navigation and zoom controls: back, forward, zoom in, zoom out, 100%, and refresh icons.

流程图 | 日誌



EM_Project1
 資料來源
 PROFILE
 流程圖
 CaseStudy
 模型套件

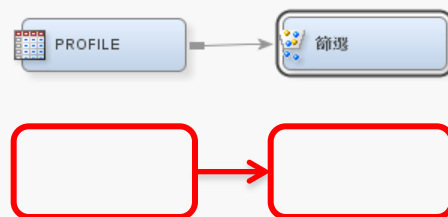
屬性	值
一般	
節點 ID	Filter
匯入的資料	
匯出的資料	
附註	
訓練	
匯出表格	已篩選
篩選表格	訓練資料
分配資料集	是
類別變數	
類別變數	
預設篩選方法	稀有值 (百分比)
保留遺漏值	是
正規化值	是
最小次數截止值	1
最小百分比截止值	0.01

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



流程圖 | 日誌



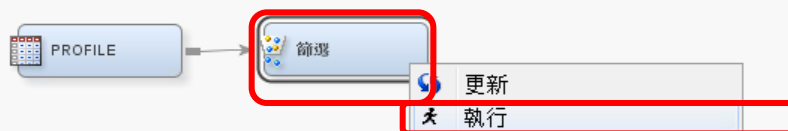
EM_Project1

- 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
更新最佳值	否

樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

CaseStudy



- 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

流程图 | 日誌 | 100%

- 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

執行狀態

- 執行完成
- 流程圖:CaseStudy
- 路徑: 篩選

確定 結果...

屬性	值
----	---

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
更新異值層級	否

執行完成

篩選一間隔變數 ...

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

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
更新量值層級	否
狀態	
一般	
一般屬性	

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

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

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

PROFILE → 篩選

互動式間隔篩選

CNT_ATM

COUNT

VALUE

≤ 100

套用篩選(F) 清除篩選(L)

欄: 標籤(A) 採礦(M) 基本(I) 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)

篩選一匯入的資料... [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

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)

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

匯入的資料

節點匯入的表格集

執行完成

流程圖 日誌

篩選一匯入的資料

EMWS1Ids_DATA

檔案(F) 檢視(V) 動作(A) 視窗(W)

樣本屬性

屬性	值
列	未知
欄	6
資料庫	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+	0.775194	000041360
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
1	00041360	34	3	3	9	49
2	00052094	44	17	5	18	84
3	00057340	122	26	32	36	216
4	00076885	42	3	6	1	52
5	00089150	20	15	2	2	39
6	00094512	83	20	49	3	155
7	00096396	33	9	49	17	108
8	00098062	22	5	9	1	37
9	00122779	21	27	2	1	51
10	00125838	127	3	2	3	135
11	00137317	52	20	2	11	85
12	00147896	101	40	8	1	150
13	00150206	156	52	80	4	292
14	00160501	31	32	42	12	117
15	00176820	107	3	2	1	113
16	00187784	54	14	7	24	99
17	00189200	92	3	2	2	99
18	00198327	44	27	17	3	91

篩選一匯出的資料... [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

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(X)... 屬性(P)... 確定(O)

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

節點匯出的表格彙集

執行完成

流程圖 日誌

篩選一匯出的資料

EMWS1.Filter_TRAIN

檔案(F) 檢視(V) 動作(A) 視窗(W)

樣本屬性

屬性	值
列	未知
欄	6
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				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
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	000160501	31	32	42	12	117
14	000176820	107	3	2	1	113
15	000187784	54	14	7	24	99
16	000189200	92	3	2	2	99
17	000198327	44	27	17	3	91
18	000200237	146	3	2	1	152

勘查 (Explore) — 群集 (Cluster)

The screenshot displays the Enterprise Miner software interface. The main window shows a workflow diagram with a 'PROFILE' node connected to a '篩選' (Filter) node. The 'Explore' (勘查) step is highlighted with a red box in the toolbar. A tooltip for the 'Cluster' (群集) icon indicates it is used for segmenting data for observation.

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

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
節點匯出的表格案。

執行完成

樣本 修改 模型 評估 公用程式 應用程式 時間序列

群集
執行分段資料庫可以使用的觀測群集。

樣本 勘查 修改 模型 評估 公用程式

100%

流程圖 日誌



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



- 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

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

流程圖 日誌

執行完成



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

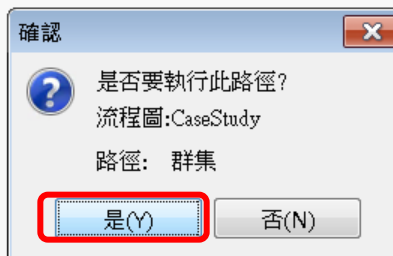
一般屬性

執行完成

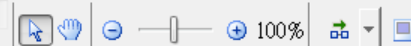


樣本 勘查 修改 模型 評估 公用程式 應用程式 時間序列

CaseStudy



流程圖 | 日誌





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



Navigation and zoom controls: back, forward, home, zoom slider (100%), refresh, and window icons.

流程圖 | 日誌

正在執行「群集」



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

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



執行狀態



執行完成

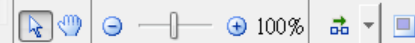
流程圖:CaseStudy

路徑: 群集

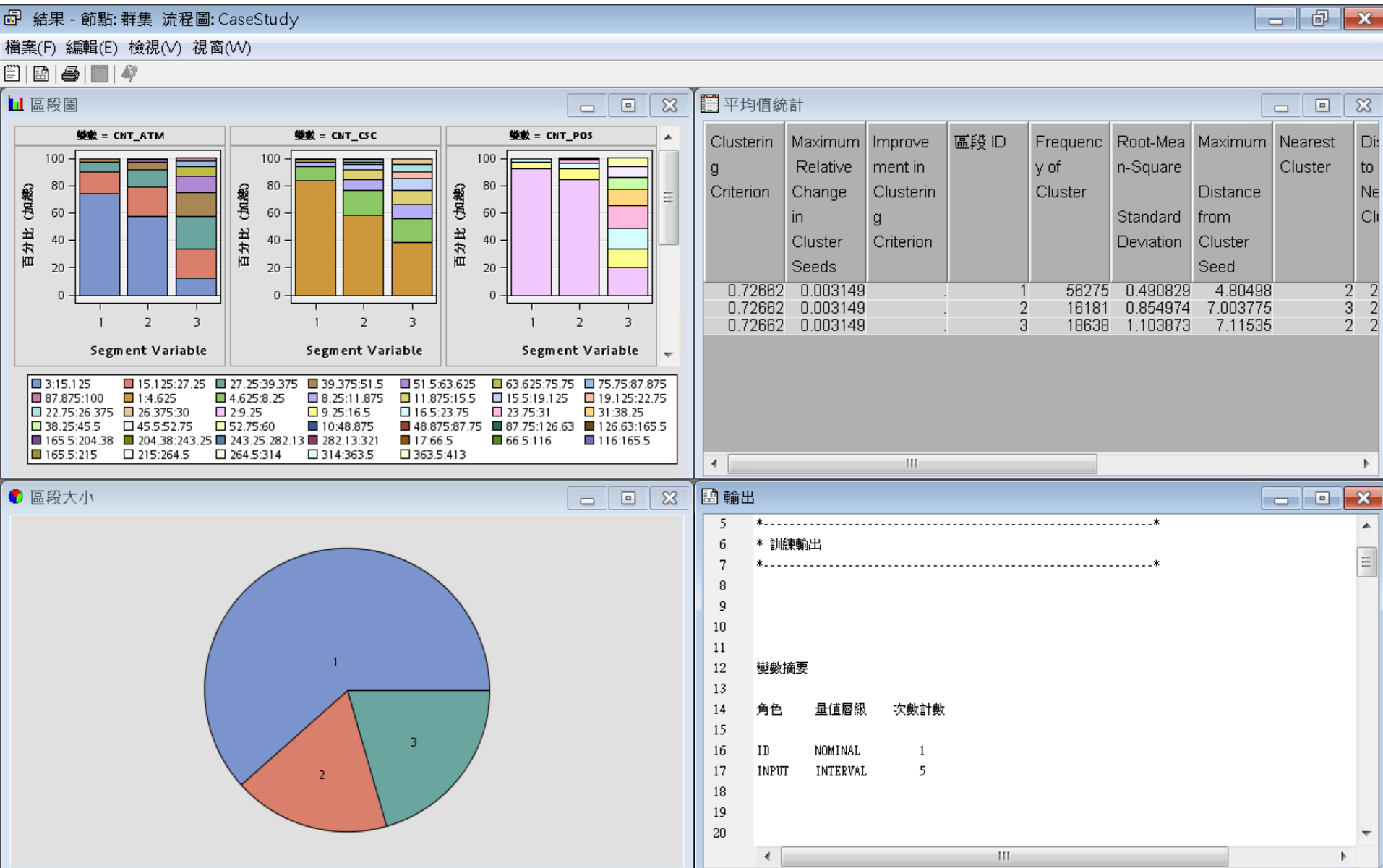
確定

結果...

流程圖 | 日誌



群集 (Cluster) 結果



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

The screenshot displays the Enterprise Miner software interface. The main window shows a workflow diagram with three steps: 'PROFILE', '篩選' (Filter), and '群集' (Cluster). The 'Assess' (評估) step is highlighted in red. Below the workflow, a toolbar contains icons for '樣本' (Sample), '勘查' (Explore), '修改' (Modify), '模型' (Model), and '評估' (Assess). The '評估' icon is circled in red. A tooltip for the '評估' icon reads: '區段特徵描繪 (Segment Profile) 使用「區段特徵描繪」節點檢查分段或群集的资料，並識別母體中資料區段不同的因子。' (Segment Profile uses the 'Segment Profile' node to check segmented or clustered data, and identify factors that differentiate data segments in the population.)

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

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

區段特徵描繪 (Segment Profile)

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

執行完成

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

The screenshot displays the Enterprise Miner interface. The main window shows a workflow diagram with four steps: PROFILE, 篩選 (Filter), 群集 (Cluster), and 區段特徵描繪 (Segment Profile). The '區段特徵描繪' step is highlighted with a red box, and a red arrow points to it from a red box around the 'Assess' tab in the top menu. The left sidebar shows a tree view of the project structure and a properties table for the selected step.

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

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

The screenshot shows the Enterprise Miner interface. The main window displays a workflow diagram with four steps: PROFILE, 篩選, 群集, and 區段特徵描繪. The '區段特徵描繪' step is highlighted with a red box, and a context menu is open over it, with the '執行' (Execute) option also highlighted with a red box.

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 → 群集 → 區段特徵描繪

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

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

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

The screenshot displays the Enterprise Miner interface. On the left, a tree view shows the project structure: EM_Project1, 資料來源, PROFILE, 流程圖, CaseStudy, and 模型套件. Below this is a properties table for the selected '區段特徵描繪' node.

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	

The main workspace shows a workflow diagram with four steps: PROFILE, 篩選, 群集, and 區段特徵描繪. The '區段特徵描繪' step is highlighted with a green border. The status bar at the bottom indicates '正在執行「區段特徵描繪」'.

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

The screenshot displays the Enterprise Miner interface. On the left, a tree view shows the project structure: EM_Project1, 資料來源 (Data Sources), PROFILE, 流程圖 (Flowchart), CaseStudy, and 模型套件 (Model Kits). Below the tree is a table of properties for the selected node.

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	

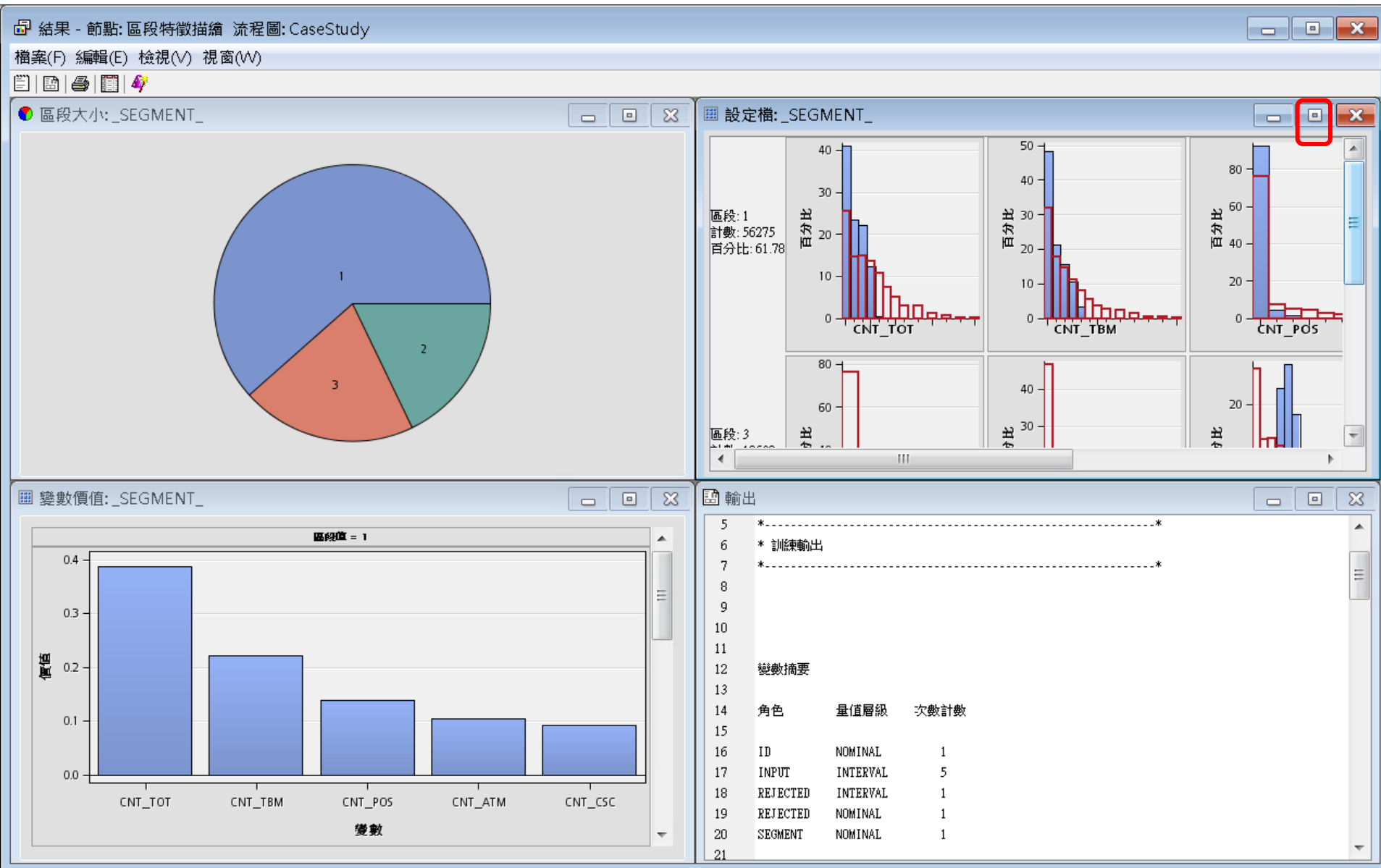
The main workspace shows a workflow diagram with four steps: PROFILE, 篩選 (Filter), 群集 (Cluster), and 區段特徵描繪 (Segment Profile). A dialog box titled "執行狀態" (Execution Status) is open, indicating that the execution is complete. The dialog box contains the following text:

執行狀態
執行完成
流程圖: CaseStudy
路徑: 區段特徵描繪

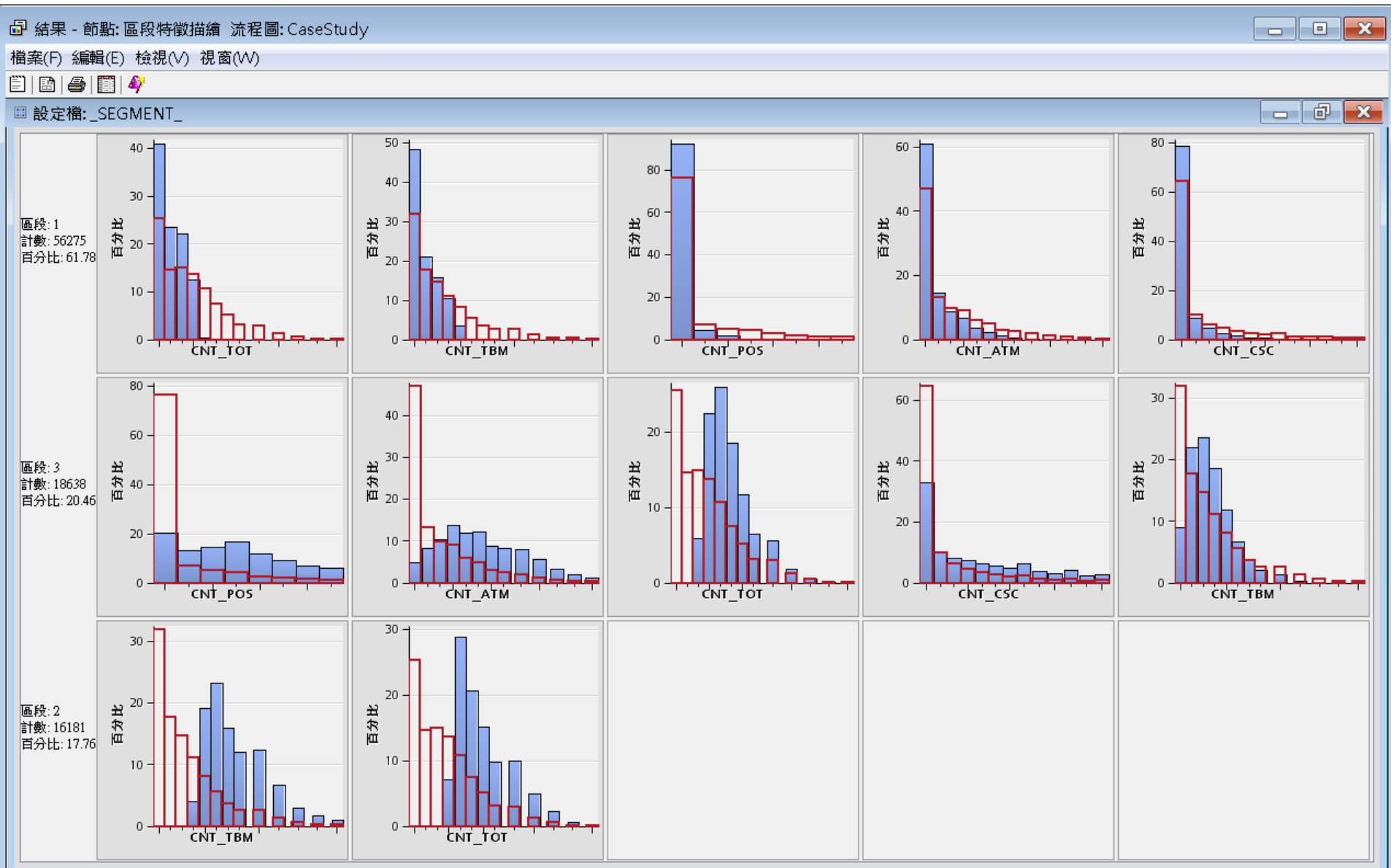
Buttons: 確定 (OK), 結果... (Results...)

The status bar at the bottom of the window shows "執行完成" (Execution Complete).

區段特徵描繪 (Segment Profile)



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