

統計套裝軟體

Yi-Ting Hwang

Department of Statistics

National Taipei University

課程進度

主題	說明	周次
SAS 簡介	SAS 基本概念	1-2
中秋節	補課: 期中考	3
BASE	基本 SAS 分析	4
描述性統計	描述性統計分析步驟	5
DATA MANIPULATION	資料分析, 運算, 計算新變數、 擷取資料、檔案管理 資料模擬	6-9
SAS 圖	統計圖表	10-12
推論性統計	連續、離散變數推論性統計分析 工具	13-14
報告程序	整理報表	15-16
開國紀念日	1/14 補課	17
SAS macro	SAS 巨集	18-19

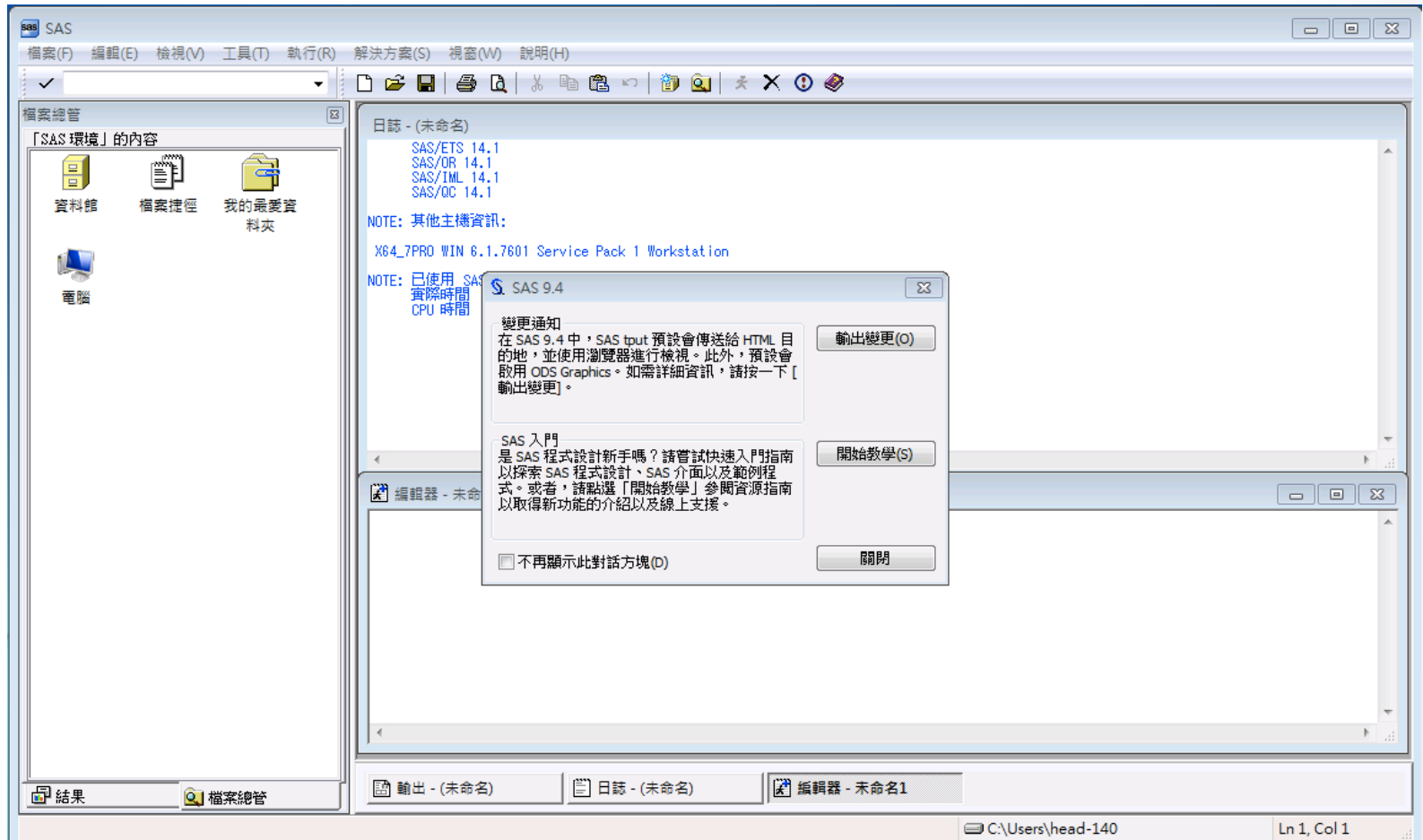
學習統計軟體的好處

- 資料合併與整理
- 統計模擬
- 敘述統計量
- 統計圖
- 統計推論

SAS 產品

- Base SAS
 - SAS procedures
 - SAS language
- SAS/GRAPH
- SAS/SQL
- SAS/STAT

SAS 視窗



SAS 視窗種類

- Editor – SAS 程式
 - Data 程序
 - Procedure 程序
- Log – 提供檔案執行過程
 - 觀察值的數目
 - 提供 DATA 遺失值的個數
 - 程式輸入錯誤的資訊
- Output -- 提供 PROC 執行結果
 - HTML
 - Listing
- Graphics – 提供高畫質圖檔
- Explorer – 檔案總管
- Help – SAS 幫手

SAS 簡介 資料

資料來源

- 問卷收集
- 病例資料
- 次級資料
 - 政府統計
 - 信用卡刷卡記錄
 - 購買頻率
 - 選課資料
 - 健保資料庫

範例 --老人生活狀況調查問卷

- 請問您今年幾歲？(Q1)
 - 1. 50~54歲 2. 55~59歲 3. 60~64歲 4. 65~69歲 5. 70~74歲 6. 75~79歲 7. 80~84歲 8. 85歲以上
- 請問您是否具有榮民、榮眷或原住民的身分？(Q2)
 - 1. 原住民 2. 榮民 3. 榮眷 4. 一般民眾
- 請問您目前的婚姻狀況？(Q3)
 - 1. 有配偶或同居 2. 喪偶 3. 離婚或分居 4. 未婚
- 總共養育過幾個兒子？(若未婚指收養幾個兒子)(Q4)
- 請問您覺得自己目前的健康狀況如何？(Q5)
 - 1. 很好 2. 還算好 3. 普通 4. 不太好 5. 很不好 6. 很難說 7. 拒答

範例 – 大專生對百貨公司及大型購物中心之消費情境問卷

1. 請問你是否去過大型購物中心 (Shopping Mall) 或百貨公司？ ☐1. 有。(請續達第二題) ☐2. 沒有。未曾去過大型購物中心或百貨公司的原因為何？ 【單選，請跳答第九題】 ☐ (1) 交通不便 ☐ (2) 沒有消費能力 ☐ (3) 沒有消費需求

2. 你最常去的大型購物中心或百貨公司為哪幾間？【可複選，最多三項】 ☐1. 新光三越 ☐2. SOGO ☐3. 中興百貨 ☐4. 遠東百貨 ☐5. 明德春天 ☐6. 衣蝶 ☐7. 明曜百貨 ☐8. 來來百貨 ☐9. 先施百貨 ☐10. 高島 ☐11. 大亞百貨 ☐12. 微風廣場 ☐13. 京華城 ☐其他_____

3. 你常去的原因為何？【可複選，最多三項】

☐1. 交通便利 ☐2. 服務完善 ☐3. 種類齊全 ☐4. 環境舒適
☐5. 會員優惠 ☐6. 動線流暢 ☐7. 贈品豐富 ☐8. 符合個人需求
☐9. 朋友相陪 ☐10. 無聊閒晃 ☐11. 商圈大吸引人 ☐12. 產品符合時尚 ☐13. 特賣活動與折扣多
☐14. 其他_____

範例 – 學生體適能調查

編號	姓名	出生年月日		身高	體重	身體質量指數	坐姿體前彎	一分鐘仰臥起坐
		年	月					
1	A	70	12	161	56.5	21.80	17	28
2	B	71	1	159	48	18.99	31	36
3	C	71	10	171	60	50.52	48	27
4	D	71	6	169	63	22.06	26	27
5	E	71	7	159	50	19.78	34	49
6	F	70	9	153	52	22.21	43	22
7	G	71	2	156	55	22.60	27	31
8	H	71	5	158.5	51	20.30	34	23

範例 – 心肌梗塞列聯表

	心肌梗塞		
	死亡	發病	沒有
安慰劑	18	171	10845
阿司匹靈	5	99	10933

範例 – 學生自覺健康

健康編號	家族病史	過去病史	是否有氣喘的病史	睡眠時間	早餐習慣	每週運動3次	吸菸習慣
0001			否，從沒發生過	不足7~8小時	偶而	沒有	不吸菸
0002			否，從沒發生過	不足7~8小時	每天吃	沒有	不吸菸
0003			否，從沒發生過	不足7~8小時	每天吃	沒有	不吸菸
0004			否，從沒發生過	不足7~8小時	偶而	沒有	不吸菸
0005			否，從沒發生過	不足7~8小時	每天吃	沒有	不吸菸
0006	心臟病.中風. 糖尿病.高血壓.蠶豆症	海鮮類過敏	否，從沒發生過	不足7~8小時	每天吃	沒有	不吸菸

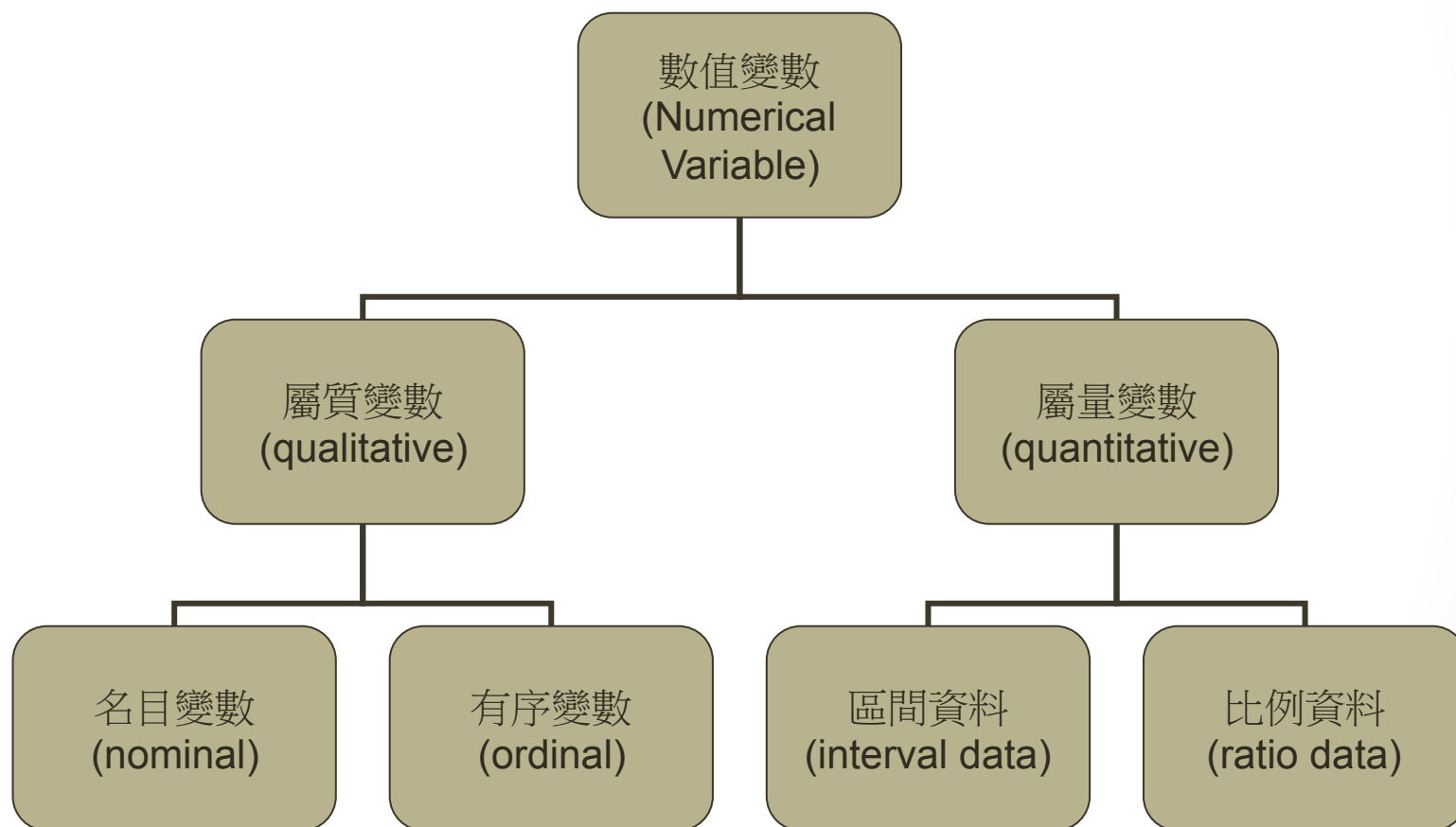
資料類型

- 數值資料 (系統性)
 - 離散資料
 - 計數資料
 - 類別資料
 - 有序資料 (ordinal data)
 - 名目資料 (nominal data)
 - 連續資料
 - 區間資料 (interval data)
 - 比例資料 (ratio data)
- 字串資料
 - 過敏病史, 地址, 電話等

範例

- 年齡
 - 以年為單位
 - 分 (0, 10], (11, 20], (20, 30], (30, 40], 40以上
- 種族、婚姻狀況、性別
- 健康狀況
 - 0 - 100
 - 健康量表
 - 四個類別 (極佳,好,普通,差)
- 身高、體重、血壓

數值資料類型摘要



SAS 簡介

SAS 變數

SAS 變數名稱

- 名稱
 - 最多只能有 32 字元
 - 可以以大寫或小寫方式輸入
- 第一個字元必須為英文字母或底線
- 名稱不可包含
 - 空格
 - 除底線外的特殊字元
- 變數名稱不可為 SAS 自動產生的變數
 - `_NUMERIC_`, `_CHARACTER_`, `_ALL_`, `_N_`, `_ERROR_`

SAS 變數包含訊息

屬性	可能數值	預設值
名稱 (name)	任何有效的名稱	無
型態 (type)	數值 (numeric) 字串 (character)	數值
長度 (length)	2 to 8 bytes	8 bytes
格式 (format)		數值: BEST12. 字串: \$w.
標籤 (label)	最多 256 各字元 (characters)	無
欄位 (position)	1-n	NA

範例 – 變數包含的訊息

3	#	Variable	Type	Len	Pos	4	Label
	5	AdminLabor	Num	8	32		Administration: Labor
	6	AdminSupplies	Num	8	40		Administration: Supplies
	9	AdminTotal	Num	8	64		Administration: Total
	7	AdminUtilities	Num	8	48		Administration: Utilities
	3	ServicesFire	Num	8	16		Services: Fire
	2	ServicesPolice	Num	8	8		Services: Police
	8	ServicesTotal	Num	8	56		Services: Total
	4	ServicesWater_Sewer	Num	8	24		Services: Water & Sewer
10	Total		Num	8	72		Total Outlays
	1	Year	Num	8	0		

資料編碼

- 除文字（字串），所有資料要盡量變成數值型態
- 變數格式
 - 數值
 - 字串
- 編制編碼簿
 - 變數名稱
 - 類別選項的數值對應

資料輸入方式

- 欄 (Column)
 - 問項
 - 每一欄只包含一個訊息 (變數)
- 列 (Row)
 - 觀察值
 - 一筆觀察值可以有數列的資料

範例 -- 老人生活狀況調查

id	q1	q2	q3	q4	q5	q6_1	q6_2	q8
9	1	2	4	5	1	2	0	3
10	2	2	4	4	2	3	1	1
14	3	2	4	2	1	5	2	2
47	1	2	4	4	1	1	2	2
48	3	1	4	6	1	1	2	2
54	5	1	4	6	1	1	1	1
60	6	2	3	3	1	0	1	2
63	5	1	4	4	1	1	2	2
74	1	1	4	4	3	1	1	1
83	1	2	4	4	1	1	1	2

範例 – 學生體適能調查

編號 ID	姓名 NAME	出生年月日 BIRTHY BIRTHM		身高 HEIGHT	體重 WEIGHT	身體 質量 指數 BMI	坐姿體前 彎 FORWARD	一分 鐘仰 臥起 坐 UP
1	A	70	12	161	56.5	21.80	17	28
2	B	71	1	159	48	18.99	31	36
3	C	71	10	171	60	50.52	48	27
4	D	71	6	169	63	22.06	26	27
5	E	71	7	159	50	19.78	34	49
6	F	70	9	153	52	22.21	43	22
7	G	71	2	156	55	22.60	27	31
8	H	71	5	158.5	51	20.30	34	23

範例 – 心肌梗塞列聯表

用藥	Myocardial Infarction		
	死亡	發病	沒有發病
安慰劑	18	171	10845
阿司匹靈	5	99	10933

- 定義變數
 - 治療組別 (TRT)
 - 疾病狀態 (Myocardial Infarction; MI)
 - 人數 (COUNT)

範例 – 大專生對百貨公司及大型購物中心之消費情境問卷

1. 請問你是否去過大型購物中心 (Shopping Mall) 或百貨公司？ ☐1. 有。(請續達第二題) ☐2. 沒有。未曾去過大型購物中心或百貨公司的原因為何？ 【單選，請跳答第九題】 ☐ (1) 交通不便 ☐ (2) 沒有消費能力 ☐ (3) 沒有消費需求

2. 你最常去的大型購物中心或百貨公司為哪幾間？【可複選，最多三項】 ☐1. 新光三越 ☐2. SOGO ☐3. 中興百貨 ☐4. 遠東百貨 ☐5. 明德春天 ☐6. 衣蝶 ☐7. 明曜百貨 ☐8. 來來百貨 ☐9. 先施百貨 ☐10. 高島 ☐11. 大亞百貨 ☐12. 微風廣場 ☐13. 京華城 ☐其他_____

3. 你常去的原因為何？【可複選，最多三項】

☐1. 交通便利 ☐2. 服務完善 ☐3. 種類齊全 ☐4. 環境舒適
☐5. 會員優惠 ☐6. 動線流暢 ☐7. 贈品豐富 ☐8. 符合個人需求
☐9. 朋友相陪 ☐10. 無聊閒晃 ☐11. 商圈大吸引人 ☐12. 產品符合時尚 ☐13. 特賣活動與折扣多
☐14. 其他_____

SAS 簡介

DATA STEP

目的

- 產生 **SAS** 資料集 (**data sets**)
 - 由已存在的原始資料 (**external files**) 產生
 - 由已存在的 **SAS**資料集來產生新的 **SAS** 資料集
- 計算新變數
- 擷取資料
- 檔案管理

DATA STEP 必備語法

1. 輸入 DATA
2. 在 DATA 後設定資料集名稱
 - 設定新資料集名稱
 - 使用原有資料集名稱
 - 使用 _NULL_ (空的資料集)
3. 設定資料集
4. 完成所有以上步驟需輸入 RUN
5. 每一步動作的最後需輸入 “;”

輸入資料的方式

- 直接由 **DATA** 步驟輸入資料
 - 利用 **INPUT** 定義變數名稱及屬性
 - 輸入資料前須輸入 **CARDS** 或 **DATALINES**
 - 利用 **INFILE** 定義資料分隔方式等
- 外在檔案 (**external file**)
 - 文字檔
 - 利用 **INFILE** 定義外在檔案的位置
 - 再以 **INPUT** 指令定義變數名稱及屬性
 - 其他
 - **IMPORT Wizard**
 - **PROC IMPORT**

資料輸入

一般文字資料檔

文字資料檔

- 直接利用 **DATA STEP** 輸入資料
 - 利用 **INPUT** 來定義每一個欄位的變數名稱及變數的屬性
 - 完成 **DATA** 所有程式的撰寫後，輸入 **CARDS** 或 **DATALINES**
 - 在 **CARDS** 之後輸入原始資料
- 外在文字檔
 - 利用 **INFILE** 設定資料檔所在目錄及檔名及資料分隔方式
 - 利用 **INPUT** 來定義每個欄位的變數名稱及變數的屬性（格式）

INPUT 語法

- 設定
 - 變數名稱
 - 變數格式
 - 變數欄位

- 語法

INPUT *<pointer-control> variable <\$> < @@>;*

- 說明
 - ***pointer-control***: 移動游標至設定變數欄位
 - ***@@***: 設定游標停置，持續在同一組紀錄 (**record**) 讀取資料
 - ***\$***: 字串變數的變數屬性 (格式)，預設為數值格式
 - 變數後面可以定義變數資料所在欄位
 - 固定格式 — 變數位於固定欄位
 - 自由格式 --變數沒有固定欄位

設定輸入變數格式 (INFORMAT)

- 定義輸入變數的格式
- 2 種主要的類型
 - 字串：`$INFORMATw.`
 - 數值：
 - 數字：`INFORMATw.d`
 - 日期：`mmddyy10.`

範例

ID	Transaction date	Transaction amount
124325	08/10/2003	1250.03
7	08/11/2003	12500.02
114565	08/11/2003	5.11

```
data transact;  
input id  
      tran_date  
      amount  
      $6.  
      mmddyy10.  
      8.2;
```

變數

INFORMAT

範例 – 變數輸出格式

紀錄
天數

- 藉由設定 **FORMAT** 才可換成日期呈現方式

Obs	id	tran_date	amount
1	124325	15562	0.03
2	7	15563	0.03
3	114565	15563	0.03

範例

```
data transact;  
  input id  
        tran_date  
        amount  
        $6.  
        mmddyy10.  
        comma10.2;
```



ID	Transaction date	Transaction amount
124325	08/10/2003	\$1,250.03
7	08/11/2003	\$12,500.02
114565	08/11/2003	5.11

範例 – 輸出格式

金額

- 藉由 **INFORMAT** 的設定刪除 \$ 與 ,
- 藉由設定 **FORMAT** 數據呈現方式才可換成有逗號

Obs	id	tran_date	amount
1	124325	15562	1250.03
2	7	15563	12500.02
3	114565	15563	5.11

常用 INFORMAT/FORMAT

型態	形式	說明	範圍	預設值
字串	\$w.			8
數值	BESTw.	SAS 自選最好的數值格式		8
	COMMAw.d	包含逗號的數值格式		8
	DOLLARw.d	包含\$, 逗號, 小數點的數值格式		8
	Ew.	科學記號		8
	w.d	標準數值格式		8
	DATEw.	將日期表示成 ddmmmyy 或 ddmmmyyyy	5-9	7
	DDMMYYw.	將日期表示成 ddmmyyyy 或 ddmmyy	2-10	8
	MMDDYYw.	將日期表示成 mmddyyyy 或 mmddyy	2-8	8
	YYMMDDw.	將日期表示成 yymmdd 或 yyyyymmdd	2-8	8

範例 – INFORMAT / FORMAT

形式	格式設定	呈現方式
字串	\$8.	abvcdefh
數值	3.0	123
	DDMMYY8.	01/08/15
	MMDDYY8.	08/01/15
	YYMMDD10.	2015/08/01
	DATE9.	01AUG2015
	Comma10.	12,345,667

輸入變數的方式 -- 自由格式

- 不需設定變數所在欄位
- 預設以空格或記錄結束分隔變數與變數輸入值
- 若非以空格作為分隔，則可以利用 **INFILE** 中的 **DELIMITER** 選項來修改
- 範例

```
input name $ age;
```

```
-----+-----1-----+-----2-----+
```

```
Peterson 21
```

```
Morgan 17
```

範例 – 心肌梗塞列聯表

	Myocardial Infarction		
	死亡	發病	沒有發病
安慰劑	18	171	10845
阿司匹靈	5	99	10933

- 定義變數
 - 治療組別 (TRT)
 - 疾病狀態 (Myocardial Infarction; MI)
 - 人數 (COUNT)

範例 - 自由格式

```
Data death;  
  input trt $ mi $ count;  
cards;  
P FA 18  
P NFA 171  
P NA 10845  
A FA 5  
A NFA 99  
A NA 10933  
run;
```

輸入變數的方式 -- 固定格式

- 需指定變數數值所在資料欄位
- 可指定所需讀取資料的欄位即可
- 資料變數很多或是有字串變數，最好用固定格式
- 範例

```
input name $ 1-8 age 11-12;
```

```
-----+-----1-----+-----2-----+
```

```
Peterson    21
```

```
Morgan      17
```

範例 - 固定格式

```
data death;  
  input trt 1-1 mi 2-2 count 3-7;  
cards;  
01      18  
02     171  
0310845  
11        5  
12       99  
1310933  
run;
```

範例

- 欲了解纖維中棉花的比例與纖維的強度關聯
- 考慮五種不同比例的棉花含量 (15%, 20%, 25%, 30%, 35)

棉花含量	抗拉強度 (磅/英吋 ²)				
15	7	7	15	11	9
20	12	17	12	18	18
25	14	18	18	19	19
30	19	25	22	19	23
35	7	10	11	15	11

INFILE 語法

- 目的
 - 設定文字檔的路徑及檔名
 - 設定資料分隔符號
 - 設定其他資料相關選項
- 語法

INFILE *file-specification* <options>;

範例 -- 心肌梗塞列聯表

	Myocardial Infarction		
	死亡	發病	沒有發病
安慰劑	18	171	10845
阿司匹靈	5	99	10933

變數	欄位	說明
TRT	1-1	治療組別
MI	3-3	疾病狀態
Count	5-9	人數

資料檔：statpackch1d1.dat

範例 - 固定格式

```
data death;  
  infile 'statpackch1d1.dat';  
  input trt 1-1 mi 3-3 count  
        5-9;  
run;
```

動態指標 (POINTER CONTROL)

指令的作用	相對指令	絕對指令
欄位控制 (pointer controls)	$+n$	$@n$
記錄指標控制 (multiple records)	$@$	N/A
觀察值指標控制 (multiple observations)	$@@$	N/A
列控制 (line pointer controls)	$/$	$\#n$

範例 – 欄位指標

```
input name $ 1-8 age 11-12 @20  
gender $;
```

```
-----+-----1-----+-----2-----+
```

```
Peterson      21              M
```

```
Morgan        17              F
```

範例 – 列記錄

變數	欄位	記錄 (record)
name	1–8	1
age	11–12	1
gender	20	1
region	1–5	2
race	8–16	2
weight	1–3	3
height	6–8	3

資料檔：statpackch1d2.txt

範例

```
/* 利用絕對記錄*/
```

```
Input #1 name 1-8 age 11-12 gender 20  
      #2 region 1-5 race 8-16  
      #3 weight 1-3 height 6-8;
```

```
/* 利用相對記錄*/
```

```
Input name 1-8 age 11-12 gender 20  
      / region 1-5 race 8-16  
      / weight 1-3 height 6-8;
```

```
/* 僅讀取第 1,3 記錄*/
```

```
Input #1 name 1-8 age 11-12 gender 20  
      #3 weight 1-3 height 6-8;
```

範例 – SAS 程式

```
data death;  
  infile 'statpackch1d2.txt';  
  Input  
    #1 Name 1-8 age 11-12  
      gender 20  
    #2 region 1-5 race 8-16  
    #3 weight 1-3 height 6-8;  
run;
```

範例 – 觀察值動態指標控制

```
data death;  
  input trt mi count @@;  
cards;  
0 1 18 0 2 171 0 3 10845  
1 1 5 1 2 99 1 3 10933  
run;
```

範例 – 設定日期格式方式

變數	欄位	格式
員工編號	7-14	字串
地址	16-82	字串
起聘日	83-91	Date9.

```
data death;  
  infile 'statpackch1d3.txt';  
  input name 7-14  
         address 16-82  
         @83 inithired date9.;  
run;
```

資料檔：statpackch1d3.txt

範例

變數	欄位	記錄	格式
編號	1-4	1	字串
姓	9-19	1	字串
名	20-29	1	字串
城市	30-42	1	字串
州別	43-44	1	字串
性別	1	2	字串
工作代碼	9-11	2	字串
薪水	20-29	2	數值
出生年月日	30-36	2	日期 (date7.)
起聘日	43-49	2	日期 (date7.)
住家電話	54-65	2	字串

資料檔：statpackch1d4.dat

範例 -- SAS 程式

```
data ch1d4;  
Infile "statpackch1d4.dat";  
input #1 id $ 1-4 lastname $ 9-19 firstname $ 20-29  
        city $ 30-42 state $ 43-44  
        #2 sex $ 1-1 jobcode $9-11 @20 salary  
        @30 bdy date7. @43 hiredate date7.  
        homephone $ 54-65;  
  
run;
```

選項設定 (OPTIONS)

指令	說明	預設值
CARDS	直接輸入資料設定資料輸入方式	
DELIMITER= <i>delimiter(s)</i>	設定變數分隔符號	空格
DSD (delimiter-sensitive data)	字串以引號方式分隔	
FIRSTOBS= <i>record-number</i>	設定第一個觀察值的起始列	1
N=	設定每一筆觀察值的紀錄數 (records)	1
OBS= <i>record-number</i>	設定輸入觀察值個數	
MISSOVER	避免 INPUT 指令往下一筆記錄區 尋找遺失的變數的輸入值	FLOWOVER
TRUNCOVER	避免 INPUT 指令在最後一個變數 的觀察值長度與格式設定長度不一 致，而忽略該筆觀察值	

範例 – 分隔字元設定

變數	順序	格式
編號	1	字串
分數一	2	數值
分數二	3	數值
分數三	4	數值
分數四	5	數值

```
data death;  
    infile 'statpackch1d5.dat'  
        delimiter=',';  
    Input id $ grade1 grade2  
        grade3 grade4;  
  
run;
```

檔案：statpackch1d5.dat
分隔符號：”，”

範例 – 設定起始觀察值所在列及 預定讀取觀察值個數

- 檔案說明 – 底特律兇殺案
- 變數說明
 - FTP – 每 100,000 人中全職警察占的比率
 - UEMP – 失業率
 - MAN – 每千人中在製造業的人數
 - LIC – 每 100,000 人有手槍執照的比率
 - GR – 每 100,000 人擁有手槍的比率
 - CLEAR – 有破案的凶殺案比例
 - WM – 白人的比例
 - NMAN – 每千人中在非製造業的比率
 - GOV – 每千人中在政府工作的人數
 - HE – 平均時薪
 - WE – 平均周薪
 - HOM – 每 100,000 人兇殺案占的比率
 - ACC – 每 100,000 人意外死亡占的比率
 - ASR – 每 100,000 人襲擊案占的比率
- 起始欄位 – 40 列
- 每一個觀察值有二列資料
- 只讀取 10 個觀察值

檔案：statpackch1d6.txt

範例 – SAS 程式

```
data death;  
  infile 'statpackch1d6.txt'  
        firstobs=40 obs=20;  
  input ftp uemp man lic gr clear wn  
        nman gov he we hom acc asr;  
  
run;
```

範例 – 讀取字串的問題

變數	欄位	格式
顧客編號	7-14	字串
居住州	15-16	字串
郵遞區號	17-21	字串
居住國家	22-41	字串
電話	42-53	字串
名字	54-107	字串

檔案：statpackch1d7.txt

```
data death;  
    infile 'statpackch1d7.txt'  
        truncover;  
    input customer 7-14  
        state $ 15-16  
        zipcode 17-21  
        country $ 22-41  
        tel_no $ 42-53  
        name $ 54-107;  
  
run;
```

範例 – 欄位遺失的問題與讀日期

變數	說明	欄位	格式
Order_num	訂單編號	7-11	字串
Cust_id	顧客編號	12-19	字串
Amount	總金額	20-32	數值
region	國家或州	33-41	字串
prepay	訂金	42-52	數值
Emp_id	負責職員編號	53-58	字串
Bill_date	帳單寄出日期	59-67	Date9.
Due_date	繳交日期	68-76	Date9.

檔案：statpackch1d8.txt

範例 – 程式

```
data ch1d8;  
infile "statpackch1d8.txt" missover;  
input order_num $ 7-11  
      cust_id $12-19  
      amount 20-32 region $ 33-41  
      prepay 42-52 emp_id $ 53-58  
      @59 bill_date date9.  
      @68 due_date date9.;  
  
run;
```

SAS 簡介

PROCEDURE

PROCEDURE

- PROC 執行特殊分析工作
- SAS 的優點
 - 選項 (options) 及指令 (statements) 語法設定方式一致

PROC PRINT

- 目的 -- 列印出 **SAS** 資料集中的變數及觀察值
 - 可利用選項設定想要列印的報表形式
- 語法

```
PROC PRINT [DATA=new];  
[VAR var1- var10;]  
RUN;
```

範例

ID	Transaction date	Transaction amount
124325	08/10/2003	\$1,250.03
7	08/11/2003	\$12,500.02
114565	08/11/2003	5.11

Obs	id	tran_date	amount
1	124325	15562	1250.03
2	7	15563	12500.02
3	114565	15563	5.11

範例

```
proc print data=transact;  
var id tran_date amount;  
format  
    tran_date date10.  
    amount comma10.2;  
run;
```

Obs	id	tran_date	amount
1	124325	10AUG2002	1,250.03
2	7	11AUG2002	12,500.02
3	114565	11AUG2002	5.11

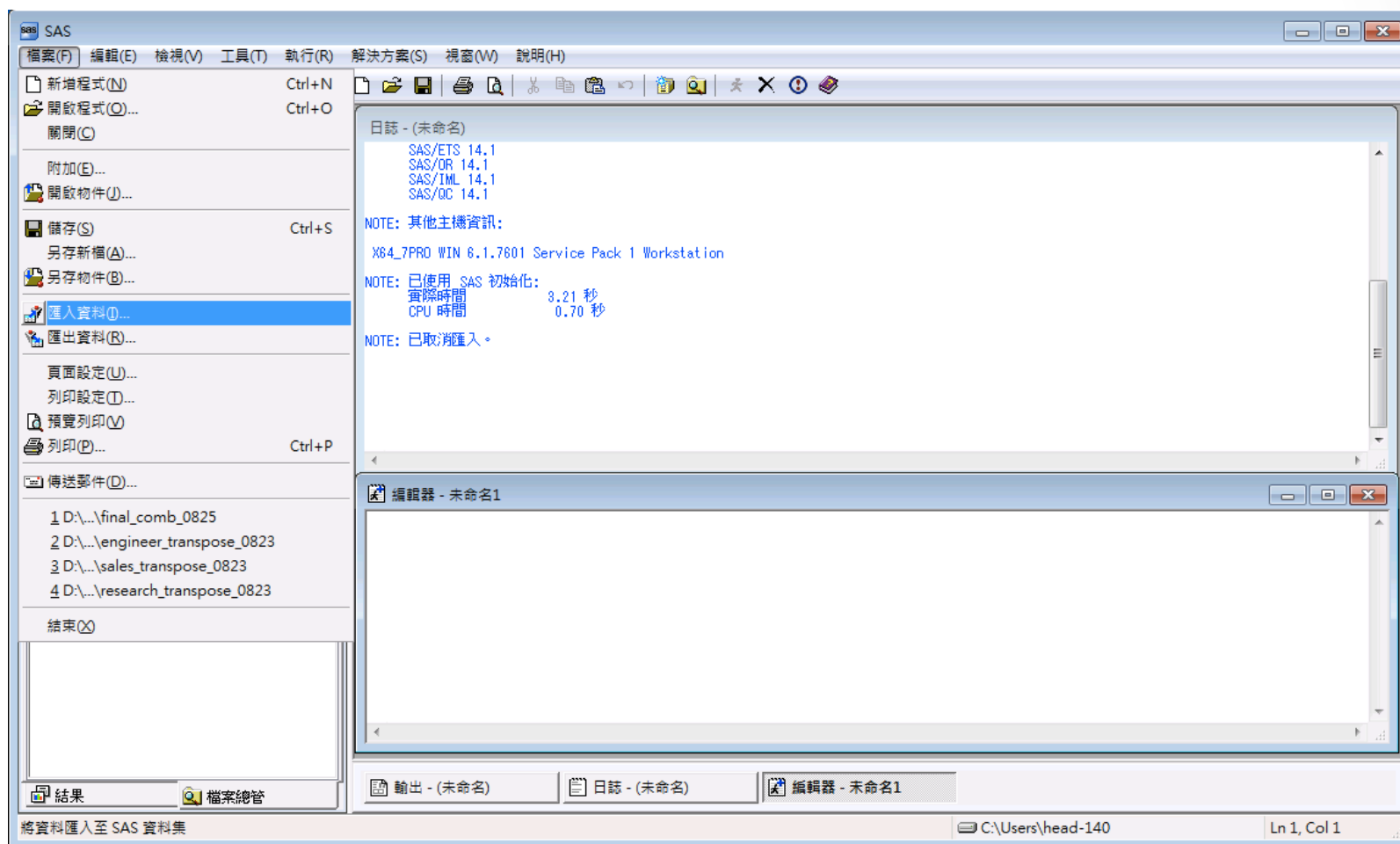
資料輸入 (續)

資料非文字檔

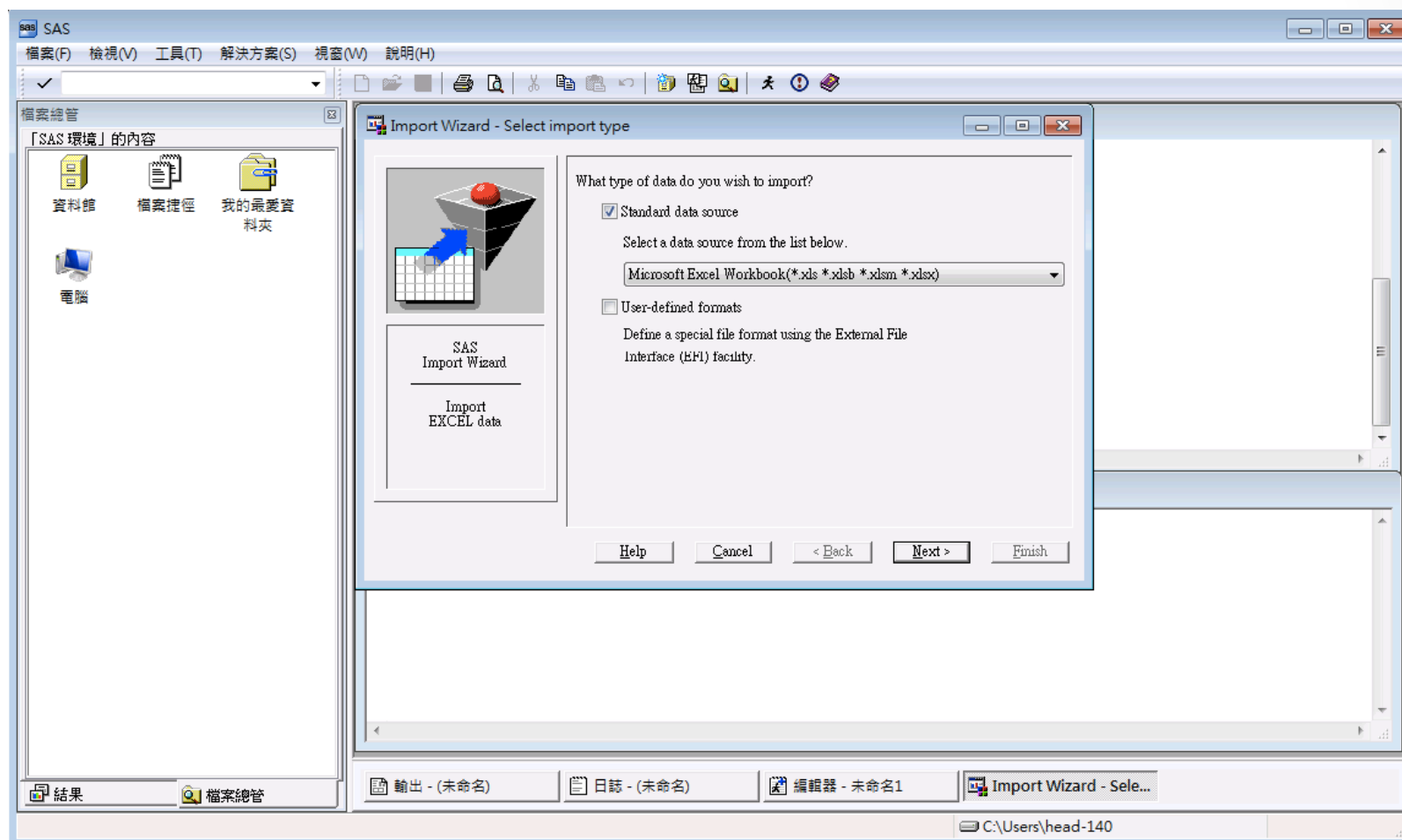
輸入非文字檔的資料

- 文字檔可利用 `INFILE` 及 `INPUT` 來產生 `SAS` 資料集
- 非文字檔則可利用 `IMPORT WIZARD`
- 資料檔：`statpackch1d9.xls`

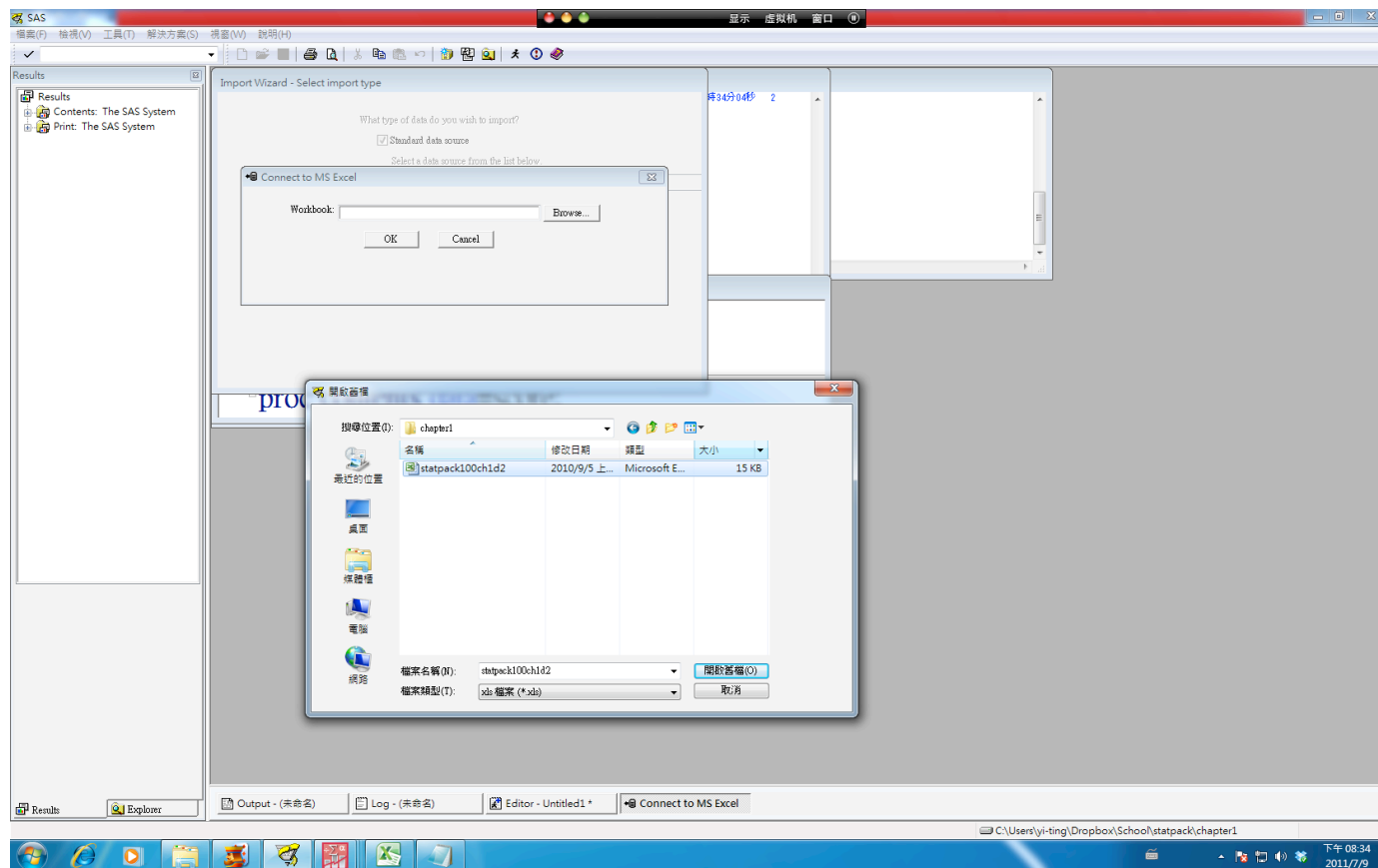
選擇匯入資料



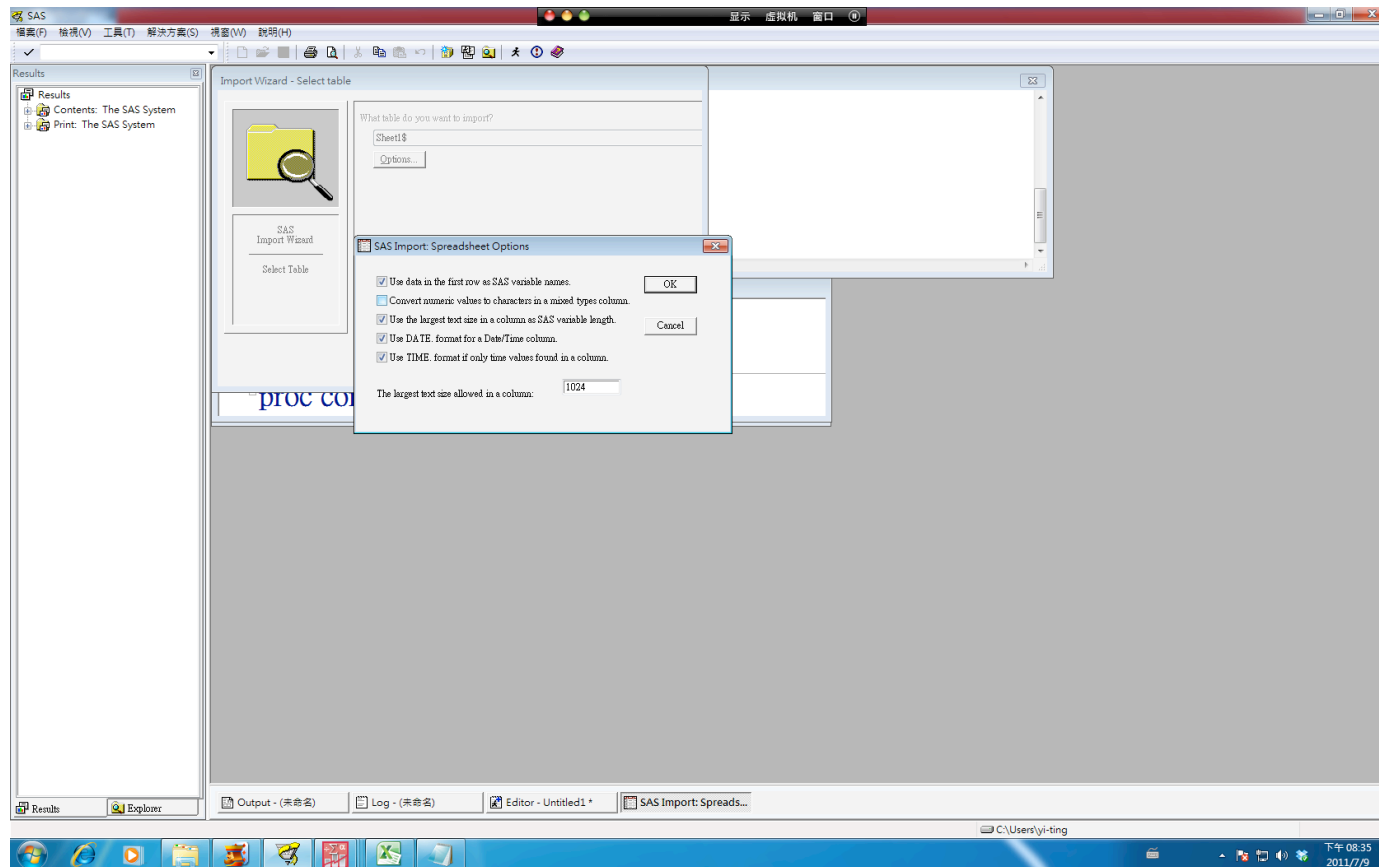
選擇檔案型態



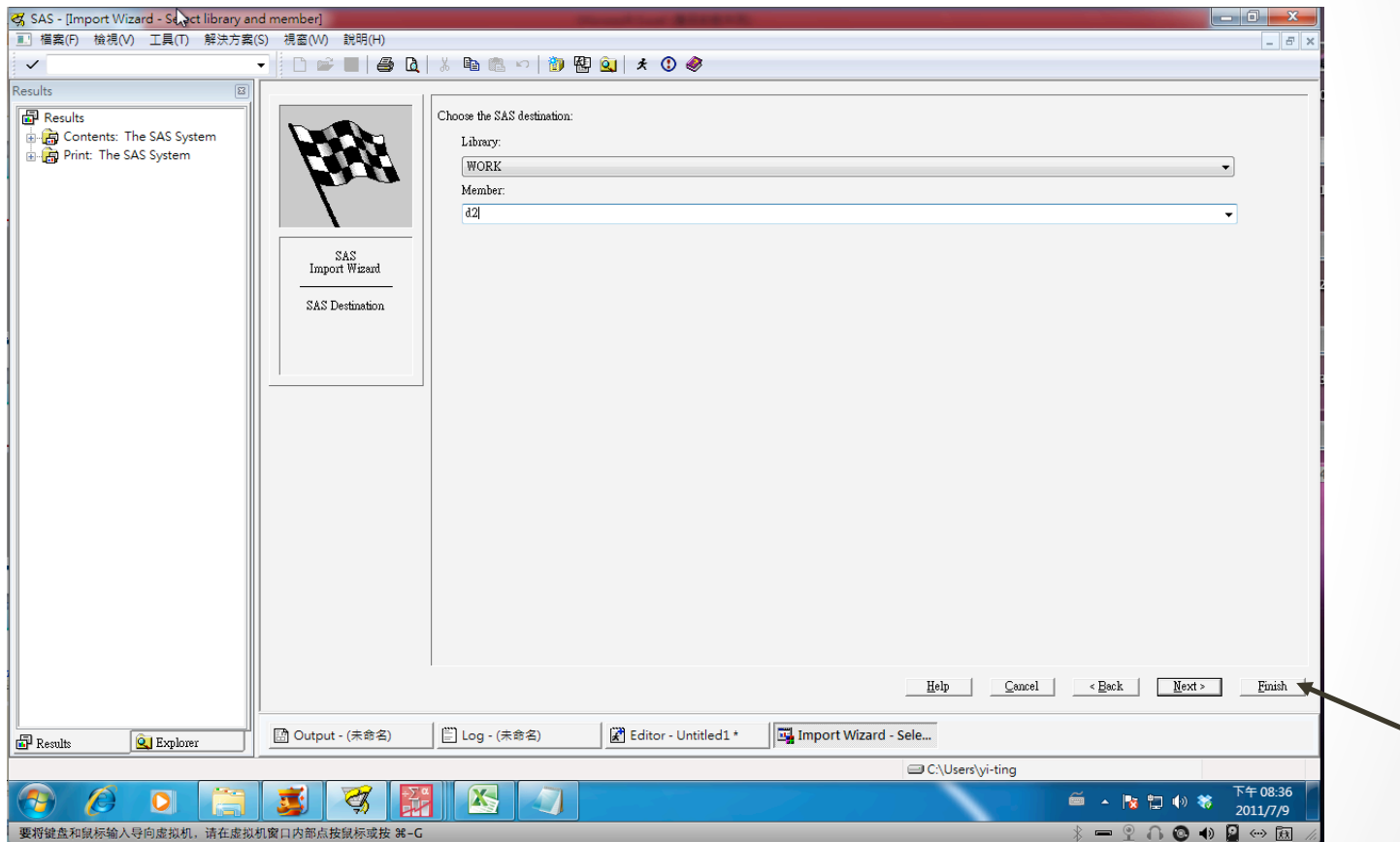
選擇目標檔案



選擇目標表單



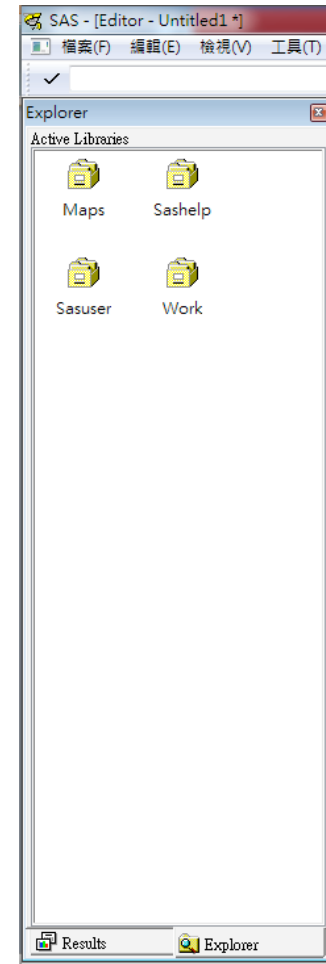
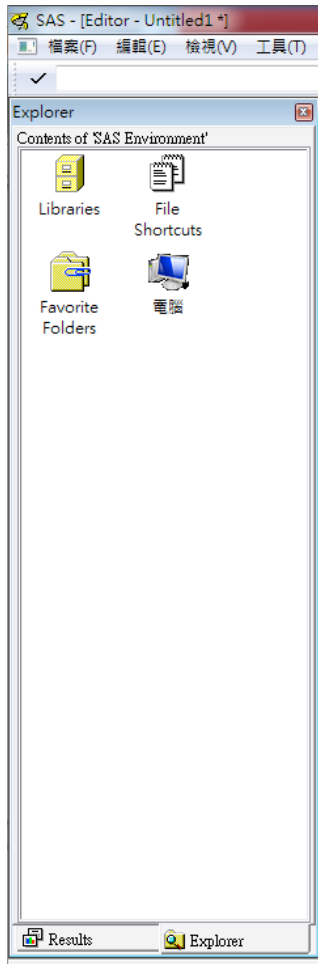
設定 SAS 資料集名稱



附註

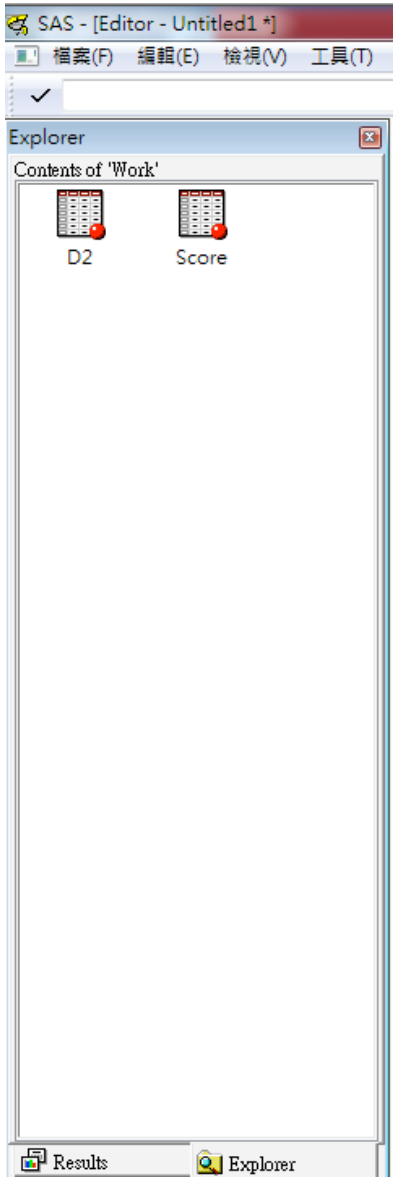
- 利用 **SAS EXPLORE** 來輸入資料集
 - **Libraries** (資料館) 為 **SAS** 資料檔的目錄設定
 - 預設的 **library** 為 **Work**
 - 所有在 **Work** 的資料檔均為暫存檔
 - **SAS** 資料檔的命名方式包含二部分
 - 第一部分：**library**
 - 第二部分：檔名
 - 範例：**work.score** 或 **score**
- 可利用 **SET** 來讀取檔案

SAS EXPLORE



選擇 work

資料呈現在 EXPLORER



The screenshot shows the SAS VIEWTABLE window titled "SAS - [VIEWTABLE: Work.D2]". The menu bar includes "檔案(F)", "編輯(E)", "檢視(V)", "工具(T)", "資料(D)", "解決方案(S)", "搜尋(V)", and "說明(H)". The Explorer pane on the left shows "Contents of 'Work'" with "D2" and "Score" icons. The main area displays a data table with the following columns: ID, NAME, BIRTHY, BIRTHM, HEIGHT, WEIGHT, BMI, FORWARD, and UP.

	ID	NAME	BIRTHY	BIRTHM	HEIGHT	WEIGHT	BMI	FORWARD	UP
1	1	A	70	12	161	56.5	21.8	17	28
2	2	B	71	1	159	48	18.99	31	36
3	3	C	71	10	171	60	50.52	48	27
4	4	D	71	6	169	63	22.06	26	27
5	5	E	71	7	159	50	19.78	34	49
6	6	F	70	9	153	52	22.21	43	22
7	7	G	71	2	156	55	22.6	27	31
8	8	H	71	5	158.5	51	20.3	34	23

變數及變數標籤

資料集的選項

SAS 簡介 資料集

讀取已存在的 SAS 資料集

指令	功能
SET	輸入一個或多個 SAS 資料集

```
DATA new;  
    SET old;  
run;
```

儲存永久檔案的方式

- 利用 **LIBRARY** 來設定
 - 檔案位置 (**File location**)
 - 預儲存的 **SAS** 資料集的版本 (**Version**)
 - 資料檔的形態 (**Type**)
- 語法：

```
LIBNAME fileref version/type  
        'directory';
```

範例

```
libname mi 'c:\';  
data mi.d1; /* 輸出永久資料到 c:\ */  
set d1;  
run;  
data temp;  
set mi.d1; /* 讀取永久資料到 SAS */  
run;
```

設定變數標籤

- 設定變數標籤
- 語法

LABEL

variable-1='label-1'

. . .

<variable-n='label-n'>;

範例

加入變數標籤

```
data death;  
infile  
    'statpack104fch1d1.dat';  
input trt $ mi $ count;  
label  
    trt='Treatment'  
    mi='Myocardial  
    Infarction';  
run;
```

移除變數標籤

```
data death;  
infile  
    'statpack104fch1d1.dat'  
    ;  
input trt mi count;  
label  
    trt=' '  
    mi=' ' ;  
run;
```

自定變數的數值標籤

- 設定變數數值標籤
- 語法

```
FORMAT variables <$>format<w>.<d> ;
```


PROC FORMAT

- 目的: 客製化數值格式
- 語法:

```
PROC FORMAT <option(s)>;
```

```
VALUE <$>name <(format-option(s))>  
    value-range-set(s);
```

範例

變數	欄位	格式
姓名	1-17	字串
員工代碼	19-22	字串
起薪	24-28	Dollar6.
區域代碼	30-32	字串
起聘日期	34-40	Date7.

字串	標籤
BR1	Birmingham UK
BR2	Plymouth UK
BR3	York UK
BR5	INCORRECT CODE
US1	Denver USA
US2	Miami USA

檔案：statpackch1d10.txt

範例 – 定義新格式 (字串)

```
proc format;  
  value $city  
    'BR1'='Birmingham UK'  
    'BR2'='Plymouth UK'  
    'BR3'='York UK'  
    'US1'='Denver USA'  
    'US2'='Miami USA'  
    other='INCORRECT CODE';  
run;
```

範例

```
proc print data=staff noobs label;  
label salary='Salary in U.S. Dollars';  
format site $city.;  
title 'STAFF with a Format for the  
Variables';  
title2 'Salary and Site';  
run;
```

範例

Name	Id Number	Salary in U.S. Dollars	Site	Hire Date
Capalleti, Jimmy	2355	21163	Birmingham UK	30JAN79
Chen, Len	5889	20976	Birmingham UK	18JUN76
Davis, Brad	3878	19571	Plymouth UK	20MAR84
Leung, Brenda	4409	34321	Plymouth UK	18SEP74
Martinez, Maria	3985	49056	Miami USA	10JAN93
Orfali, Philip	0740	50092	Miami USA	16FEB83
Patel, Mary	2398	35182	York UK	02FEB90
Smith, Robert	5162	40100	INCORRECT CODE	15APR86
Sorrell, Joseph	4421	38760	Denver USA	19JUN93
Zook, Carla	7385	22988	York UK	18DEC91

範例 – 心肌梗塞列聯表

	Myocardial Infarction		
	死亡	發病	沒有發病
安慰劑 Placebo	18	171	10845
阿司匹 靈 Aspirin	5	99	10933

- 定義變數
 - 治療組別 (TRT)
 - 疾病狀態 (Myocardial Infarction; MI)
 - 人數 (COUNT)

範例 – 定義新數值格式

定義數值格式（標籤）

```
proc format;  
value trtf  
    0='Placebo'  
    1='Aspirin';  
value mif  
    1='Fatal attack'  
    2='Nonfatal attack'  
    3='No attack';  
run;
```

SAS 程式

```
data death;  
    input trt mi  
    count;  
    format trt trtf.  
    mi mif.;  
cards;  
0 1 18  
0 2 171  
0 3 10845  
1 1 5  
1 2 99  
1 3 10933  
run;
```