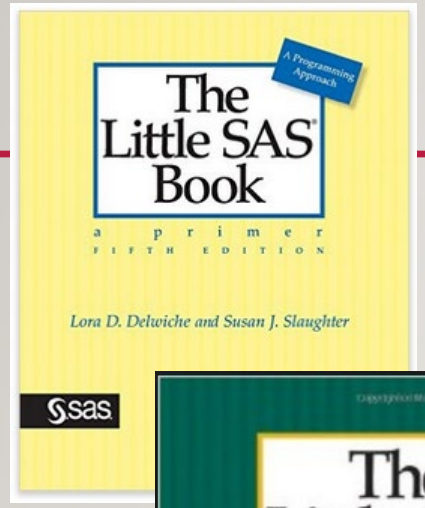
SAS PROGRAMMING : INTRODUCTION

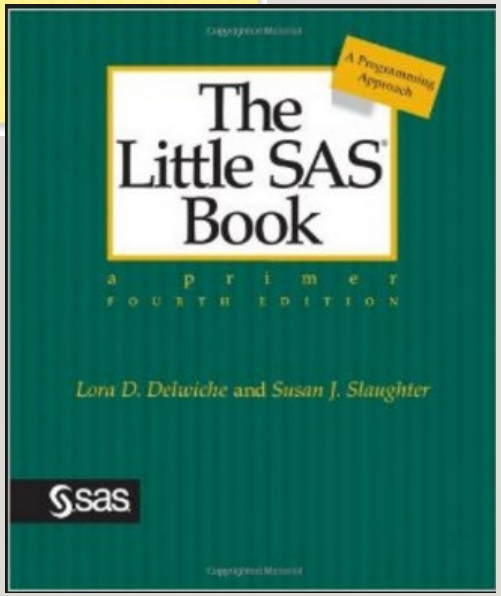
2018/11/16

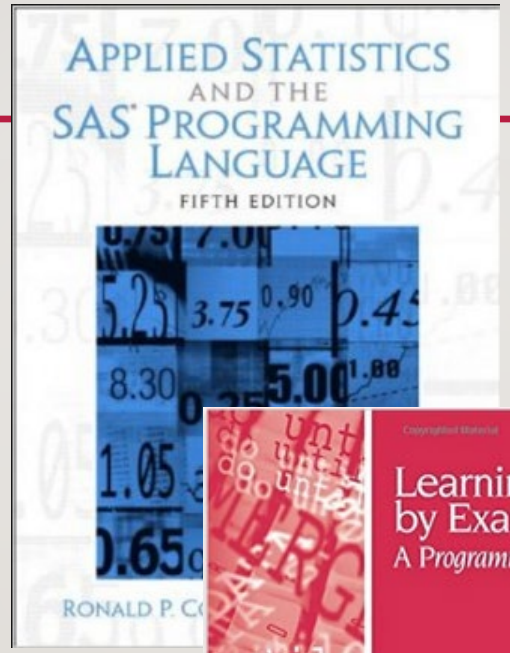
成大臨藥所

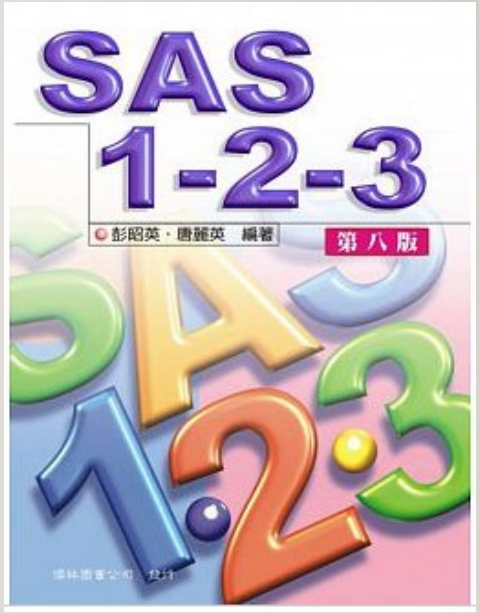
陳儒伶

SAS 基礎語法參考書目

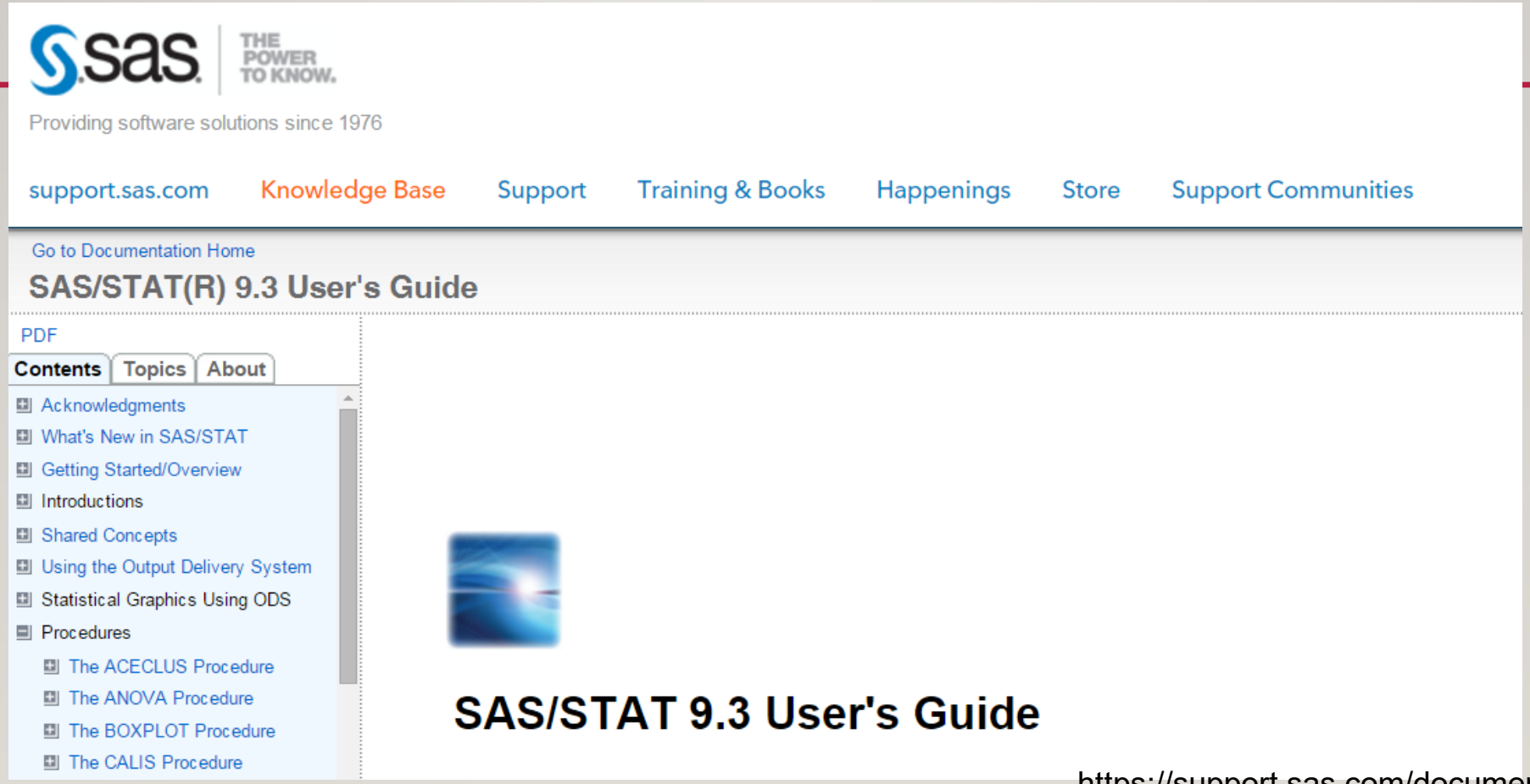








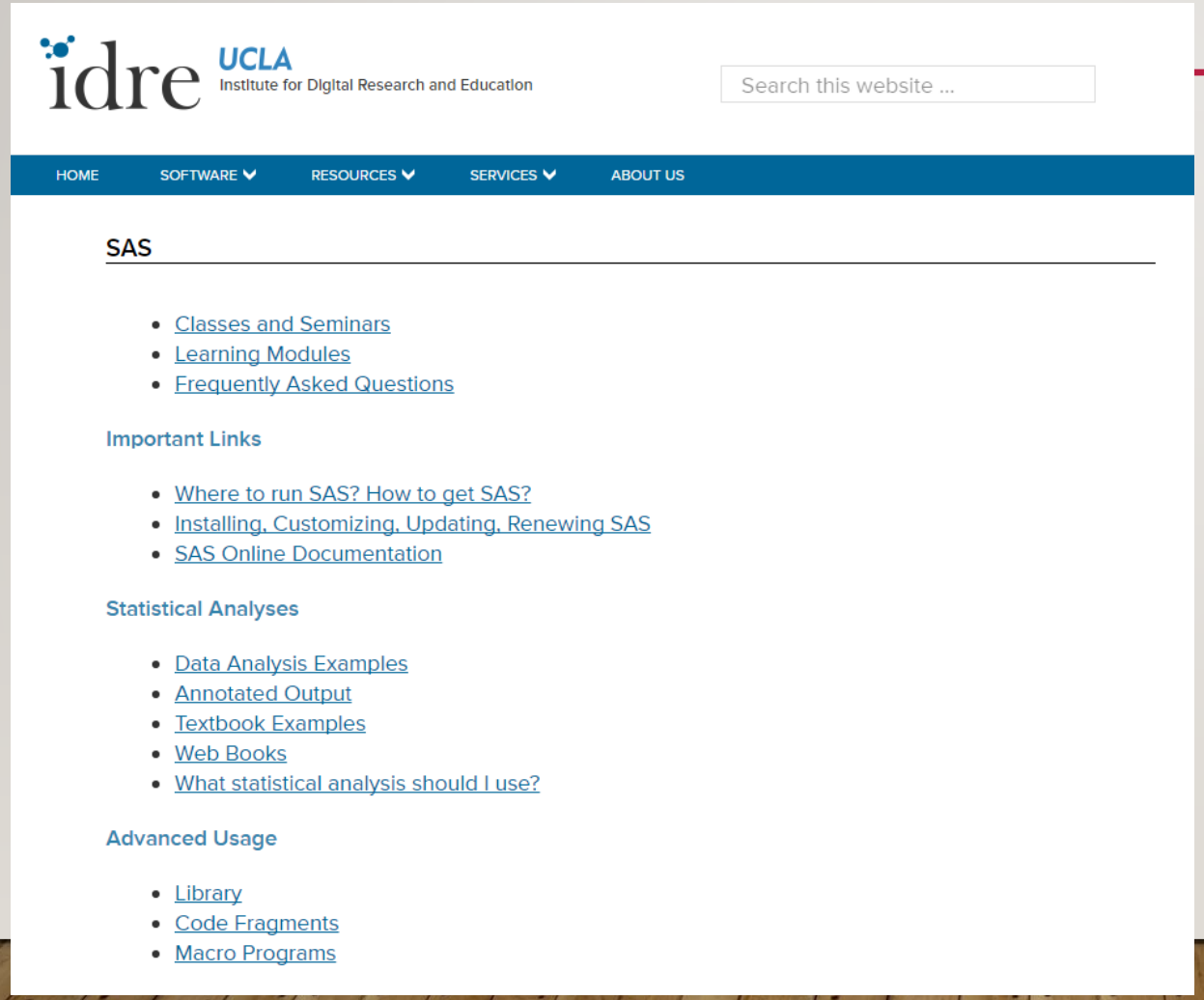
3 SAS ONLINE USER'S GUIDE



https://support.sas.com/documentation/cdl/en/statug/63962/HTML/default/viewer.htm#titlepage.htm

4 好用的網站-有例子可以練習

http://stats.idre.ucla.edu/sas/



包含STATA, SAS, SPSS

Past Classes and Workshops Available Online

Stata

- Introduction to Stata 13/14
- Regression with Stata
- Logistic Regression with Stata
- Beyond Binary Logistic Regression with Stata
- Multiple Imputation in Stata 14
- Introduction to Survey Data Analysis
- Applied Survey Data Analysis
- Survey Data Analysis
- Survival Analysis Using Stata
- Introduction to Programming in Stata

SAS

- Introduction to SAS 9.4
- Programming Basics in SAS 9.4
- Analyzing and Visualizing Interactions in SAS 9.4
- Regression with SAS
- Logistic Regression in SAS
- Repeated Measures Analysis in SAS
- Applied Survey Data Analysis using SAS 9.4
- Multiple Imputation in SAS 9.4
- Survival Analysis Using SAS
- Using Arrays in SAS
- Introduction to SAS Macro Language

SPSS

- Introduction to Regression with SPSS (Version 23)
- Introduction to SPSS version 22 (point-and-click and syntax)
- Introduction to SPSS Syntax, Part1 (using SPSS version 21)
- Introduction to SPSS Syntax, Part 2 (using SPSS version 21)
- Regression with SPSS
- Repeated Measures Analysis in SPSS
- Mediation and Moderation using the SPSS Process Macro

2.1 Import Wizard, Libnames and Proc import

We will start with inputting an Excel file into SAS first through the SAS Import Wizard.

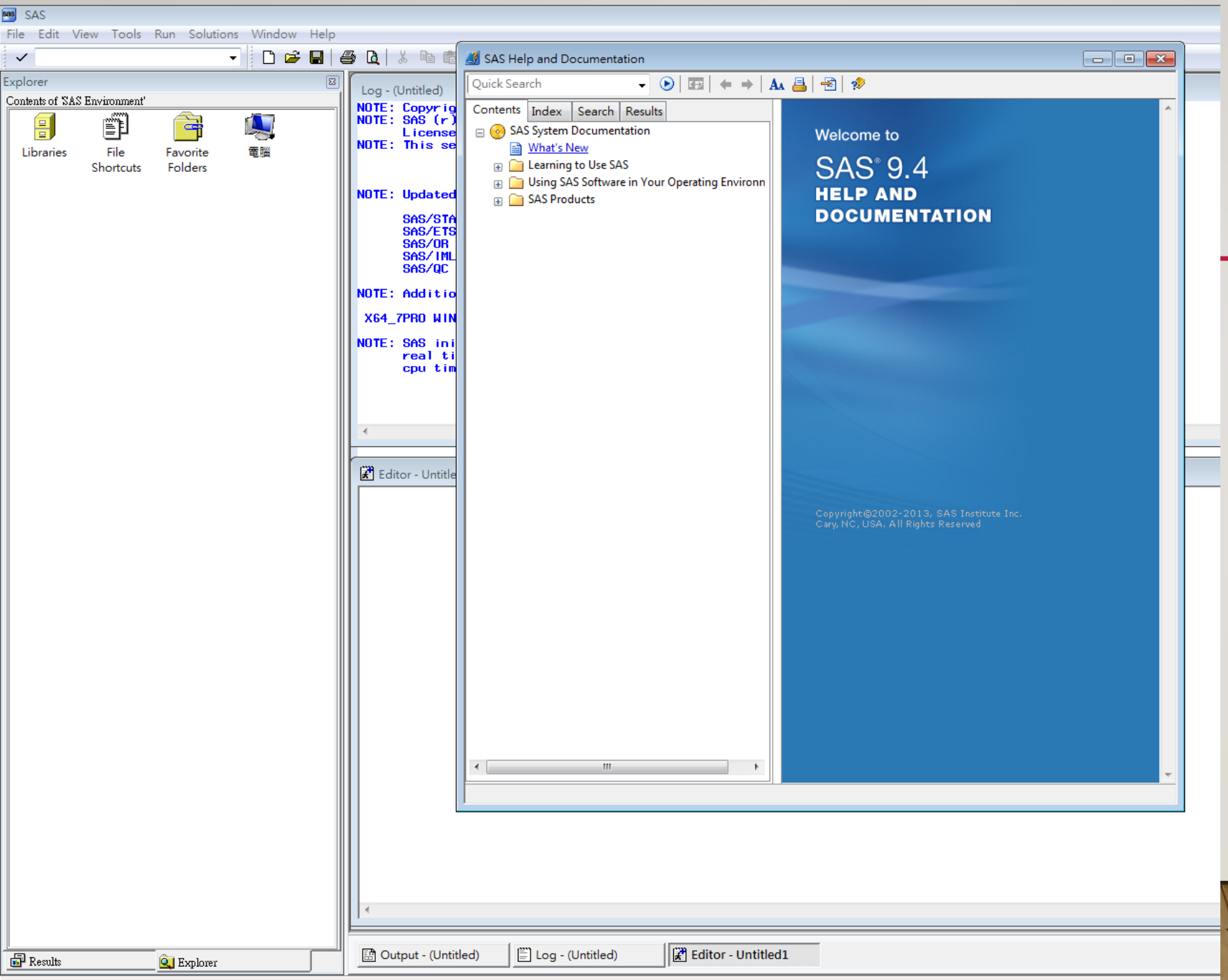
- **File**
  Import Data
  Choose Excel .xls format (this is the default)
  Click on Next
  Click on Browse to select a file: c:sas_datahs0.xls
  The default option is to read variable names from the first line, leave as it is.
  Click on Next
  Enter a name (**hs0**) for the data set
  Click on Finish

Below is the SAS syntax to import the same excel file.

```
proc import datafile="c:sas_datahs0.xls" dbms = xls out=hs0;
run;
```

Another way is to use the **libname** statement, which will be reintroduced in a later unit.

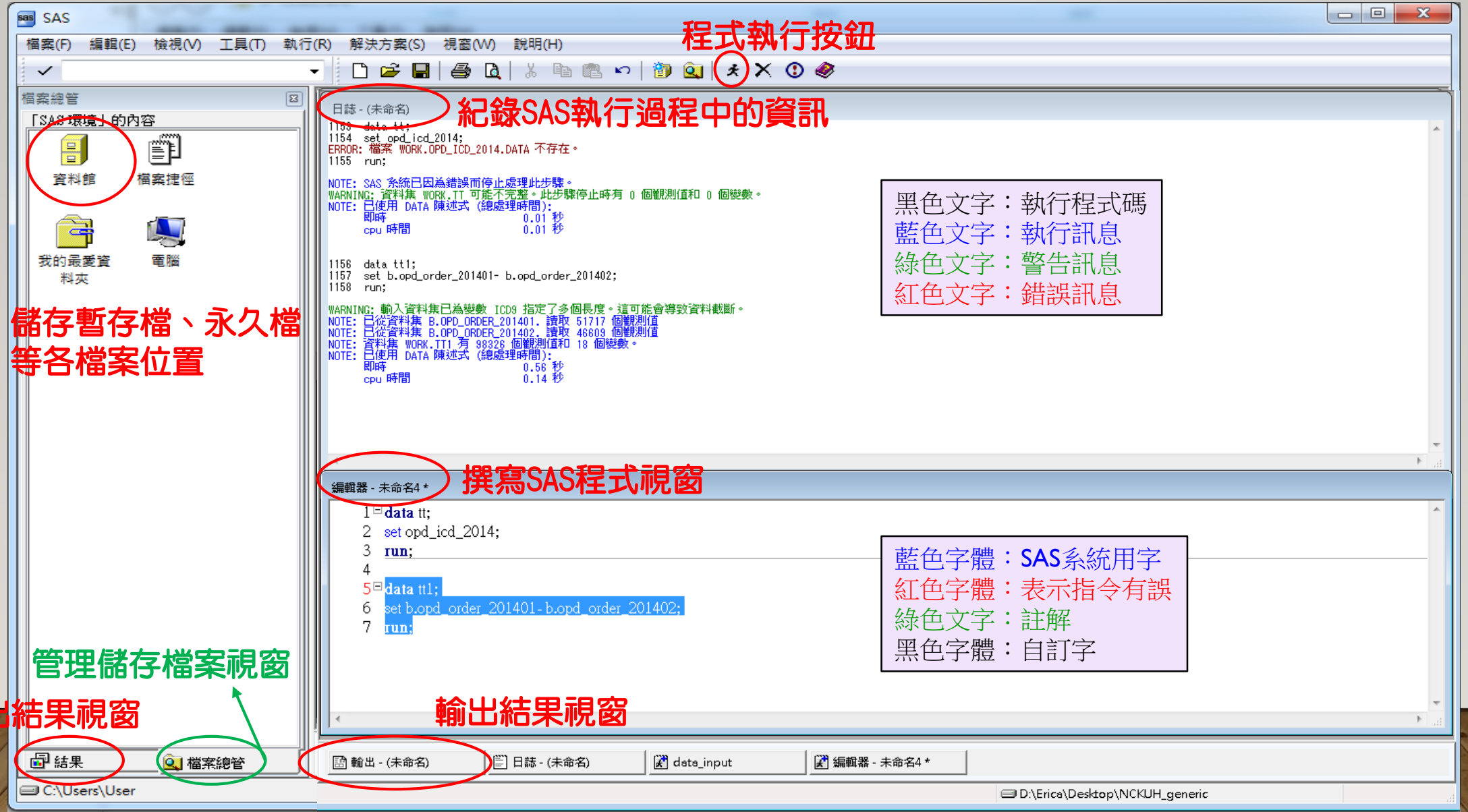
```
libname xlsdata 'c:sas_datahs0.xls';
proc print data = xlsdata."hs0$"n (obs=10);
run;
libname xlsdata clear;
```



SAS軟體內建的文件

- 包含語法及例子

SAS環境介面


SAS
檔案(F) 編輯(E) 檢視(V) 工具(T) 執行(R) 解決方案(S) 視窗(W) 說明(H)
程式執行按鈕
檔案總管
「SAS環境」的內容
資料館
檔案捷徑
我的最愛資料夾
電腦
儲存暫存檔、永久檔等各檔案位置
日誌 - (未命名)
紀錄SAS執行過程中的資訊
1153 data tt;
1154 set opd_icd_2014;
ERROR: 檔案 WORK.OPD_ICD_2014.DATA 不存在。
1155 run;
NOTE: SAS 系統已因為錯誤而停止處理此步驟。
WARNING: 資料集 WORK.TT 可能不完整。此步驟停止時有 0 個觀測值和 0 個變數。
NOTE: 已使用 DATA 陳述式 (總處理時間):
即時 0.01 秒
cpu 時間 0.01 秒
1156 data tt1;
1157 set b.opd_order_201401- b.opd_order_201402;
1158 run;
WARNING: 輸入資料集已為變數 ICD9 指定了多個長度。這可能會導致資料截斷。
NOTE: 已從資料集 B.OPD_ORDER_201401. 讀取 51717 個觀測值
NOTE: 已從資料集 B.OPD_ORDER_201402. 讀取 46609 個觀測值
NOTE: 資料集 WORK.TT1 有 98326 個觀測值和 18 個變數。
NOTE: 已使用 DATA 陳述式 (總處理時間):
即時 0.56 秒
cpu 時間 0.14 秒
黑色文字：執行程式碼
藍色文字：執行訊息
綠色文字：警告訊息
紅色文字：錯誤訊息
編輯器 - 未命名4 *
撰寫SAS程式視窗
1 data tt;
2 set opd_icd_2014;
3 run;
4
5 data tt1;
6 set b.opd_order_201401-b.opd_order_201402;
7 run;
藍色字體：SAS系統用字
紅色字體：表示指令有誤
綠色文字：註解
黑色字體：自訂字
管理儲存檔案視窗
管理輸出結果視窗
輸出結果視窗
結果
檔案總管
輸出 - (未命名)
日誌 - (未命名)
data_input
編輯器 - 未命名4 *
C:\Users\User
D:\Erica\Desktop\NCKUH_generic

9 SAS程式

- SAS程式的基本語法主要包含**兩大程式步驟**：
  - **DATA** step：用來建立或修改SAS資料。
  - **PROC** step：處理(或分析)SAS資料。Ex: **proc** sort, **proc** univariate, **proc** means…
- SAS程式撰寫特色：
  - 程式的結尾，皆以**分號 (;)** 做結尾
  - 不區分英文**大小寫**
  - 空白列數目沒有限制
  - **資料集、欄位名稱**的命名方式，要注意**只能用英文**，**不可以用數字作開頭**，**文字之間不可以有空白**，，且**只能**允許使用“_”
    - **1year　(錯誤：數字作開頭)**
    - **Diagnosis_date (正確：使用底線連接，且沒有空白)**

10 SAS基礎教學

- 資料集存放位置
- 資料匯入
- 基本語法 - 常用函數介紹

11 資料集存放位置 – 永久檔 (LIBNAME)

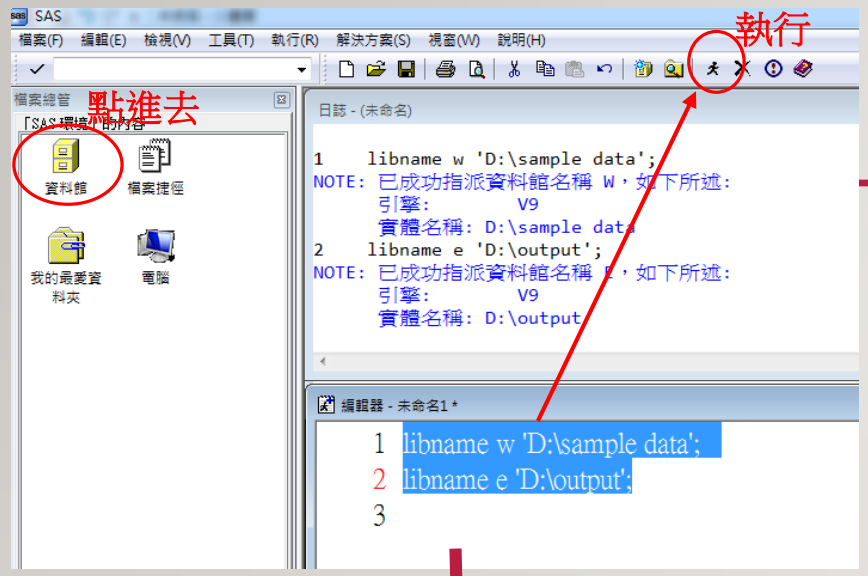
LIBNAME w ‘D:\sample data’ ;

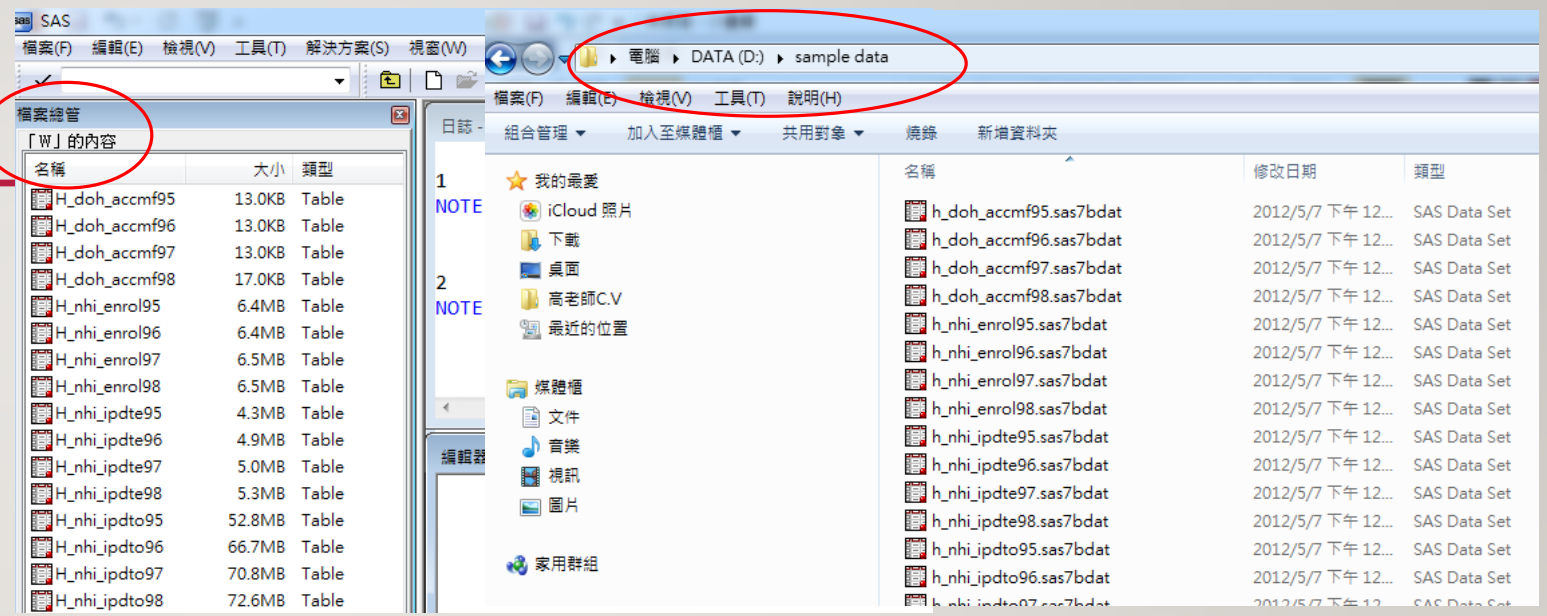
連結原始資料檔案
建立一個**資料館 (LIBNAME)**，**指定資料存取的路徑**。
例：建立資料館名稱為 “**w**”，指定存取在 “D:\sample data” 資料夾中的資料

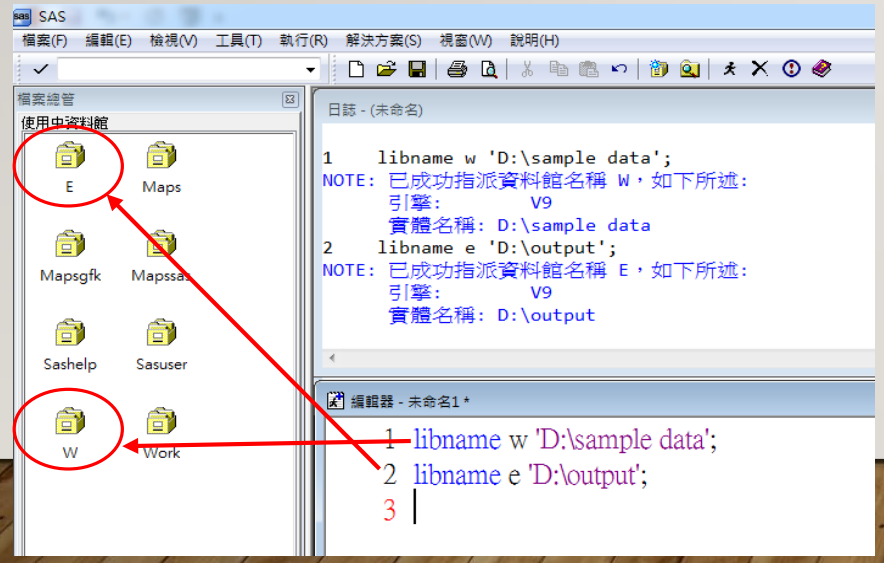
LIBNAME e ‘D:\output’ ;

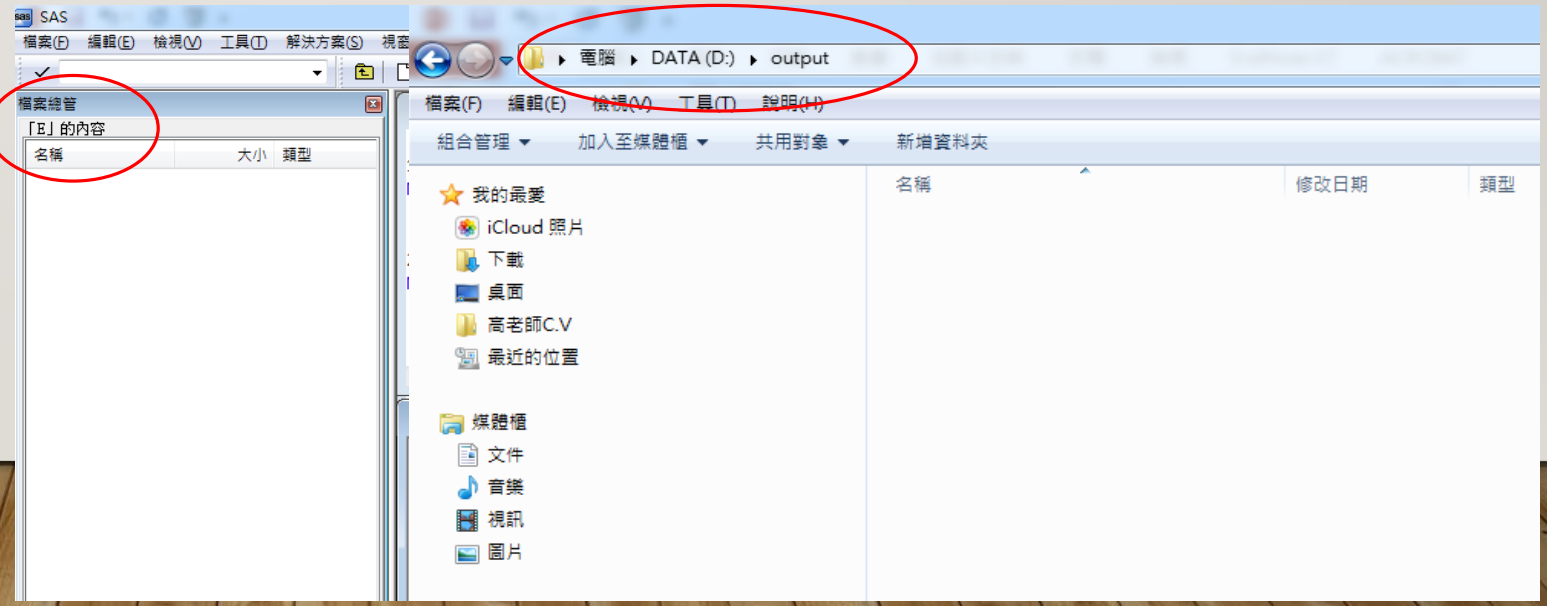
建立欲存放之永久檔儲存路徑
建立一個**資料館 (LIBNAME)**，資料館名稱為 “**e**”，**指定資料存取的路徑**在 “D:\output” 的”output”資料夾中

資料集存放位置－永久檔


SAS
檔案(F) 編輯(E) 檢視(V) 工具(T) 執行(R) 解決方案(S) 視窗(W) 說明(H)
執行
檔案總管
點進去
資料館
檔案捷徑
我的最愛資料夾
電腦
日誌 - (未命名)
1 libname w 'D:\sample data';
NOTE: 已成功指派資料館名稱 W，如下所述:
引擎: V9
實體名稱: D:\sample data
2 libname e 'D:\output';
NOTE: 已成功指派資料館名稱 E，如下所述:
引擎: V9
實體名稱: D:\output
編輯器 - 未命名1 *
1 libname w 'D:\sample data';
2 libname e 'D:\output';
3


SAS
檔案總管
「W」的內容
名稱 大小 類型
H_doh_accmf95 13.0KB Table
H_doh_accmf96 13.0KB Table
H_doh_accmf97 13.0KB Table
H_doh_accmf98 17.0KB Table
H_nhi_enrol95 6.4MB Table
H_nhi_enrol96 6.4MB Table
H_nhi_enrol97 6.5MB Table
H_nhi_enrol98 6.5MB Table
H_nhi_ipdte95 4.3MB Table
H_nhi_ipdte96 4.9MB Table
H_nhi_ipdte97 5.0MB Table
H_nhi_ipdte98 5.3MB Table
H_nhi_ipdto95 52.8MB Table
H_nhi_ipdto96 66.7MB Table
H_nhi_ipdto97 70.8MB Table
H_nhi_ipdto98 72.6MB Table
電腦 ▸ DATA (D:) ▸ sample data
檔案(F) 編輯(E) 檢視(V) 工具(T) 說明(H)
組合管理 加入至媒體櫃 共用對象 燒錄 新增資料夾
我的最愛
iCloud 照片
下載
桌面
高老師C.V
最近的位置
媒體櫃
文件
音樂
視訊
圖片
家用群組
名稱 修改日期 類型
h_doh_accmf95.sas7bdat 2012/5/7 下午 12... SAS Data Set
h_doh_accmf96.sas7bdat 2012/5/7 下午 12... SAS Data Set
h_doh_accmf97.sas7bdat 2012/5/7 下午 12... SAS Data Set
h_doh_accmf98.sas7bdat 2012/5/7 下午 12... SAS Data Set
h_nhi_enrol95.sas7bdat 2012/5/7 下午 12... SAS Data Set
h_nhi_enrol96.sas7bdat 2012/5/7 下午 12... SAS Data Set
h_nhi_enrol97.sas7bdat 2012/5/7 下午 12... SAS Data Set
h_nhi_enrol98.sas7bdat 2012/5/7 下午 12... SAS Data Set
h_nhi_ipdte95.sas7bdat 2012/5/7 下午 12... SAS Data Set
h_nhi_ipdte96.sas7bdat 2012/5/7 下午 12... SAS Data Set
h_nhi_ipdte97.sas7bdat 2012/5/7 下午 12... SAS Data Set
h_nhi_ipdte98.sas7bdat 2012/5/7 下午 12... SAS Data Set
h_nhi_ipdto95.sas7bdat 2012/5/7 下午 12... SAS Data Set
h_nhi_ipdto96.sas7bdat 2012/5/7 下午 12... SAS Data Set


SAS
檔案總管
使用中資料館
E
Maps
Mapsgfk
Mapssas
Sashelp
Sasuser
W
Work
日誌 - (未命名)
1 libname w 'D:\sample data';
NOTE: 已成功指派資料館名稱 W，如下所述:
引擎: V9
實體名稱: D:\sample data
2 libname e 'D:\output';
NOTE: 已成功指派資料館名稱 E，如下所述:
引擎: V9
實體名稱: D:\output
編輯器 - 未命名1 *
1 libname w 'D:\sample data';
2 libname e 'D:\output';
3


SAS
檔案總管
「E」的內容
名稱 大小 類型
電腦 ▸ DATA (D:) ▸ output
檔案(F) 編輯(E) 檢視(V) 工具(T) 說明(H)
組合管理 加入至媒體櫃 共用對象 新增資料夾
我的最愛
iCloud 照片
下載
桌面
高老師C.V
最近的位置
媒體櫃
文件
音樂
視訊
圖片
名稱 修改日期 類型

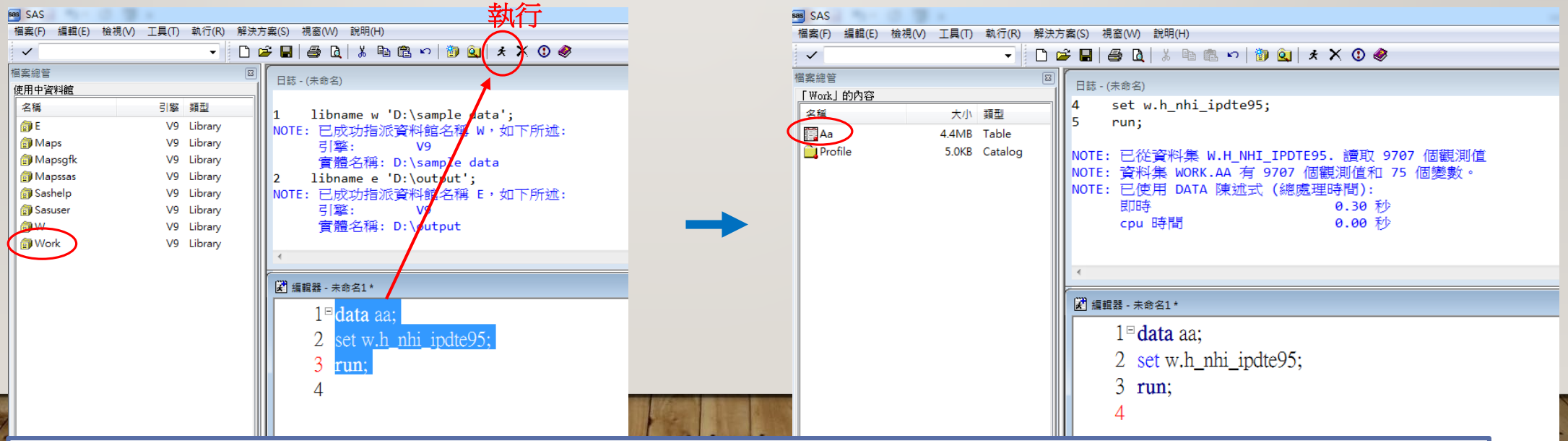
13 資料集存放位置 – 暫存檔

```
DATA aa;

set w.h_nhi_ipdte95;

RUN;
```

建立暫存資料集**aa**，複製W資料館的“**h_nhi_ipdte95** ”資料集後，存放到暫存檔**Work**資料館中



(**Note**): **Work資料館**是一個**資料暫存**的地方，如果將SAS關閉後， Work資料館裡的資料會消失

14 資料匯入 – LIST INPUT

建立資料 “cars1”

- **Reading Raw Data with the “INPUT” Statement**

宣告變數為 “文字”

```
DATA cars1;
   INPUT make $ model $ mpg weight price;
   CARDS;
   AMC Concord 22 2930 4099
   AMC Pacer   17 3350 4749
   AMC Spirit  22 2640 3799
   Buick Century 20 3250 4816
   Buick Electra 15 4080 7827
;
RUN;
```

輸入資料，也可以使用datalines

- 資料需以”空格”做間隔
- 資料屬文字變相且超過 8 bytes，需設定長度(Length)
- 不能有missing value
- 資料需符合標準格式

variable

	make	model	mpg	weight	price
1	AMC	Concord	22	2930	4099
2	AMC	Pacer	17	3350	4749
3	AMC	Spirit	22	2640	3799
4	Buick	Century	20	3250	4816
5	Buick	Electra	15	4080	7827

observation

15 資料匯入 – COLUMN INPUT

```
DATA cars2;
 INPUT make $ 1-5 model $ 6-12 mpg 13-14 weight 15-18 price 19-22;
 CARDS;
AMC  Concord2229304099
AMC  Pacer  1733504749
AMC  Spirit 2226403799
BuickCentury2032504816
BuickElectra1540807827
;
RUN;
```

- 資料需排列整齊
- 可以有missing value
- 資料需符合標準格式

16 資料匯入 – FROM EXTERNAL FILE

檔案位置　檔案名稱（含附檔名）

```
DATA cars3;
 INFILE "D:\input_data\cars3.txt";
 INPUT make $ model $ mpg weight price;
RUN;
```

以Tab為分隔符號

定義符號

```
DATA cars4;
 INFILE "D:\input_data\cars4.txt" DELIMITER='09'x ;
 INPUT make $ model $ mpg weight price;
RUN;
```

以 ， 為分隔符號

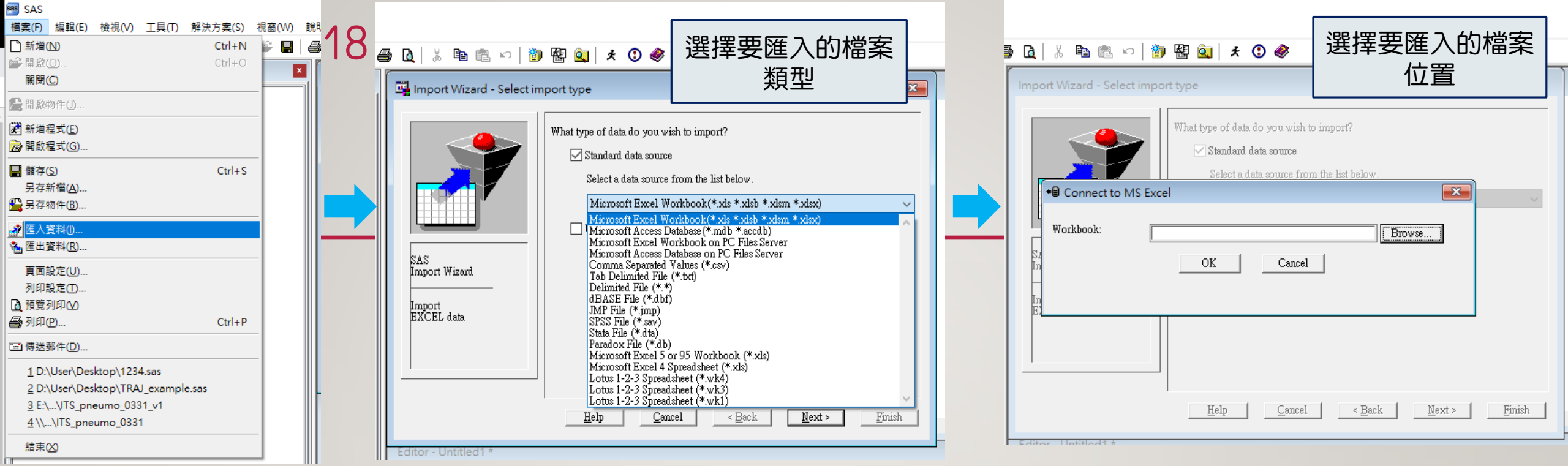
```
DATA cars5;
 INFILE "D:\input_data\cars5.txt" delimiter=',';
 INPUT make $ model $ mpg weight price;
RUN;
```

17 讀取健保藥品資料原始檔

```
data  data_name;
infile “data source" missover lrecl=1857;
input
nhi1 $ 1-2
nhi2 $ 4-13
nhi3 $ 15-16
nhi4 $ 18-27
nhi5 $ 29-37 …
nhi40 $ 1850-1857
;
run;
```

指定讀取檔案最大的長度範圍，內建256（範圍1-32767）

- 避免檔案中有遺漏值或是空白
- 不符合指定欄位大小的值就以”missing value”取代


18
檔案(F) 編輯(E) 檢視(V) 工具(T) 解決方案(S) 視窗(W) 說明
新增(N) Ctrl+N
開啟(O)... Ctrl+O
關閉(C)
開啟物件(J)...
新增程式(E)
開啟程式(G)...
儲存(S) Ctrl+S
另存新檔(A)...
另存物件(B)...
匯入資料(I)...
匯出資料(R)...
頁面設定(U)...
列印設定(T)...
預覽列印(V)
列印(P)... Ctrl+P
傳送郵件(D)...
1 D:\User\Desktop\1234.sas
2 D:\User\Desktop\TRAJ_example.sas
3 E:\...\ITS_pneumo_0331_v1
4 \\...\ITS_pneumo_0331
結束(X)
選擇要匯入的檔案類型
Import Wizard - Select import type
What type of data do you wish to import?
Standard data source
Select a data source from the list below.
Microsoft Excel Workbook(*.xls *.xlsb *.xlsm *.xlsx)
Microsoft Access Database(*.mdb *.accdb)
Microsoft Excel Workbook on PC Files Server
Microsoft Access Database on PC Files Server
Comma Separated Values (*.csv)
Tab Delimited File (*.txt)
Delimited File (*.*)
dBASE File (*.dbf)
JMP File (*.jmp)
SPSS File (*.sav)
Stata File (*.dta)
Paradox File (*.db)
Microsoft Excel 5 or 95 Workbook (*.xls)
Microsoft Excel 4 Spreadsheet (*.xls)
Lotus 1-2-3 Spreadsheet (*.wk4)
Lotus 1-2-3 Spreadsheet (*.wk3)
Lotus 1-2-3 Spreadsheet (*.wk1)
SAS Import Wizard
Import EXCEL data
Help Cancel < Back Next > Finish
Editor - Untitled1 *
選擇要匯入的檔案位置
Import Wizard - Select import type
Connect to MS Excel
Workbook:
Browse...
OK
Cancel
Import Wizard - Select table
What table do you want to import?
工作表1$
Options...
SAS Import Wizard
Select Table
選擇要匯入的工作表
Import Wizard - Select library and member
Choose the SAS destination:
Library:
WORK
Member:
read_data
SAS Import Wizard
SAS Destination
資料匯入要存放的位置及名稱
Import Wizard - Create SAS statements
The Import Wizard can create a file containing PROC IMPORT statements that can be used in SAS programs to import this data again.
If you want these statements to be generated, enter the filename where they should be saved:
D:\input_data\import.sas
Browse...
Replace file if it exists.
SAS Import Wizard
Select file
選擇資料匯入的SAS code存放位置

EXERCISE 1

- 建立資料館，將cd_2001 excel檔讀入成永久檔
- 練習將cd資料夾中的2002、2003年度資料匯入SAS

20 HOMEWORK

- 將OO資料夾中2001-2003年資料匯入成SAS檔
- 匯入最新版健保藥品資料檔

https://www.nhi.gov.tw/Content_List.aspx?n=238507DCFE832EAE&topn=3FC7D09599D25979&upn=E4B08D1A7687A588

下載藥品檔及欄位說明

21 基本語法 -在SAS中建立資料集

```
data aa;
set a.cd_2001;
run;
```

開始**資料步驟**，建立資料集 “aa”

讀取a資料館的“cd_2001 ”資料集，存入資料集“ aa ”中

run;做為此**資料步驟**的**結尾**

22 了解資料內容

proc contents data=a.cd_2001;**run**;

Data Set Name	A.CD_2001	觀測值	14280
成員類型	DATA	變數	37
引擎	V9	Indexes	0
已建立	2018/11/07 14:33:01	Observation Length	298
上次修改時間	2018/11/07 14:33:01	Deleted Observations	0
Protection		壓縮的	NO
Data Set Type		Sorted	NO
標籤			
Data Representation	WINDOWS_64		
編碼	ms-950 Traditional Chinese (PCMS)		

按字母排序的變數與屬性清單

#	變數	類型	長度
1	var1	字元	6
2	var2	字元	1
3	var3	字元	34
4	var4	字元	8
5	var5	字元	2
6	var6	字元	6
7	var7	字元	2
8	var8	字元	2
9	var9	字元	2
10	var10	字元	2
11	var11	字元	2
12	var12	字元	8

按字母排序的變數與屬性清單

#	變數	類型	長度	格式	輸入格式	標籤
21	AD_CLICK	數字	8			AD_CLICK
20	AD_EXPO	數字	8			AD_EXPO
22	AD_TIME	數字	8			AD_TIME
4	AGE	數字	8			AGE
15	AMOUNT	數字	8	DOLLAR21.2	DOLLAR21.2	AMOUNT
9	ANNU_INCOME	數字	8	DOLLAR21.2	DOLLAR21.2	ANNU_INCOME
16	CARD_TYPE	字元	10	$10.	$10.	CARD_TYPE
12	CHILDREN	數字	8			CHILDREN
6	CITY	字元	6	$6.	$6.	CITY
11	COMMUTE	數字	8			COMMUTE
2	COUNT_TEMP	數字	8			COUNT_TEMP
13	DATE	字元	20	$20.	$20.	DATE
7	DISTRICT	字元	8	$8.	$8.	DISTRICT
8	EDUCATION	字元	10	$10.	$10.	EDUCATION
1	ID	字元	12	$12.	$12.	ID
5	MARITAL_STATUS	數字	8			MARITAL_STATUS
14	MCC_CAT	字元	8	$8.	$8.	MCC_CAT
18	MOBILE_PAYMENT	數字	8			MOBILE_PAYMENT
17	PAYBY	字元	10	$10.	$10.	PAYBY
10	PRIVATE_VEHICLE	數字	8			PRIVATE_VEHICLE
19	QUANTITY	數字	8			QUANTITY
3	SEX	字元	2	$2.	$2.	SEX

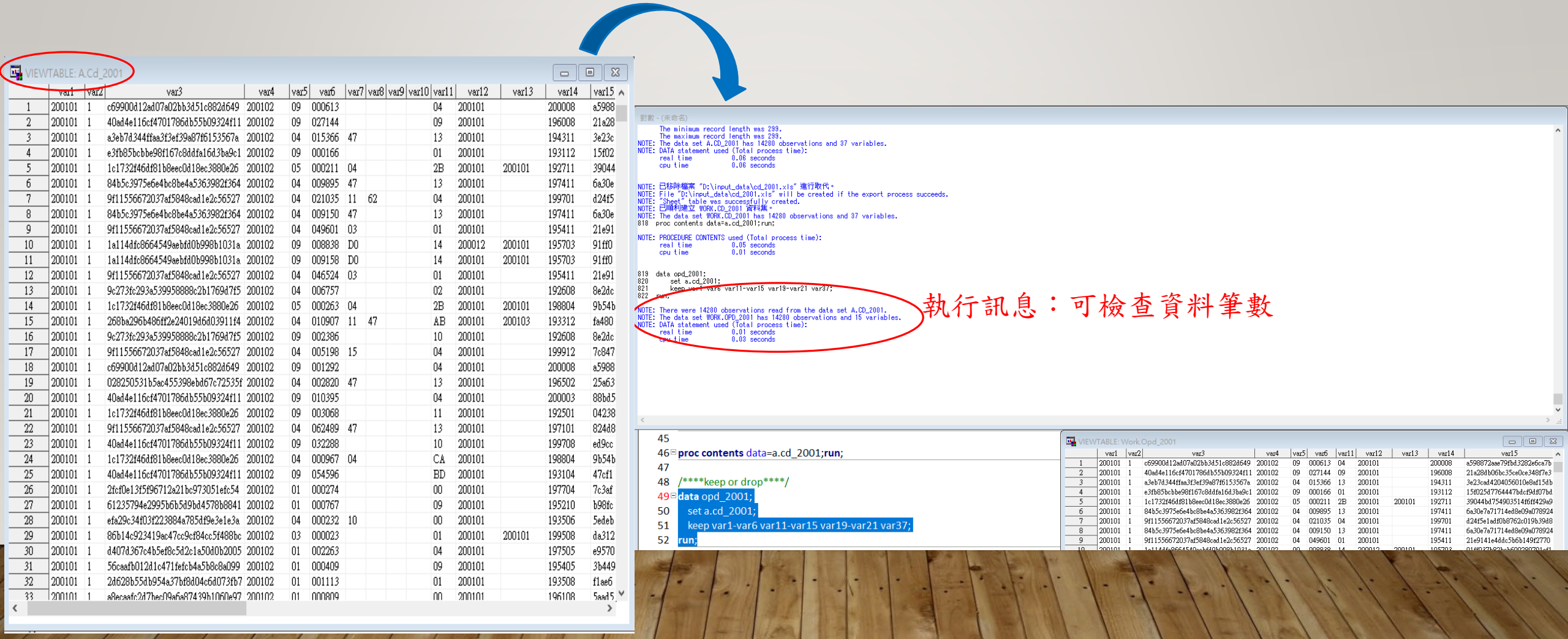
23 基本語法－保留需要分析欄位 (KEEP)

資料集“opd_2001”中，只**保留**資料集“a.cd_2001”中的**var1~var6、var11~var15、var19~var21、var37**這15個欄位

```
data opd_2001;
  set a.cd_2001;
  keep var1-var6 var11-var15 var19-var21 var37;
run;

data aa (keep=var1-var6 var11-var15 var19-var21 var37);
  set a.cd_2001;
run;
```

24 基本語法 – 保留需要分析欄位 (KEEP)



25 基本語法 – 去除不需要的欄位 (DROP)

```
data aa1 ;
   set a.cd_2001;
   drop var7-var10 var16-var18 var22-var36;
run;
```

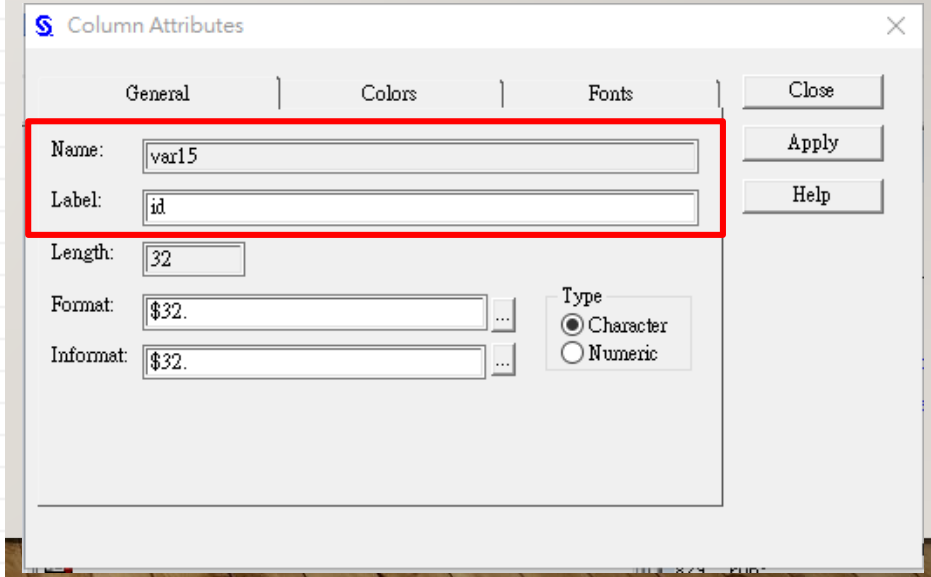
26 基本語法 – 註解欄位名稱 (LABEL)

- 給variable (變數)加上標籤，能清楚地處理資料

```
data aa2;
    set opd_2001;
    label var15=id;
run;
```

- 但程式撰寫時仍要用var15此欄位名稱

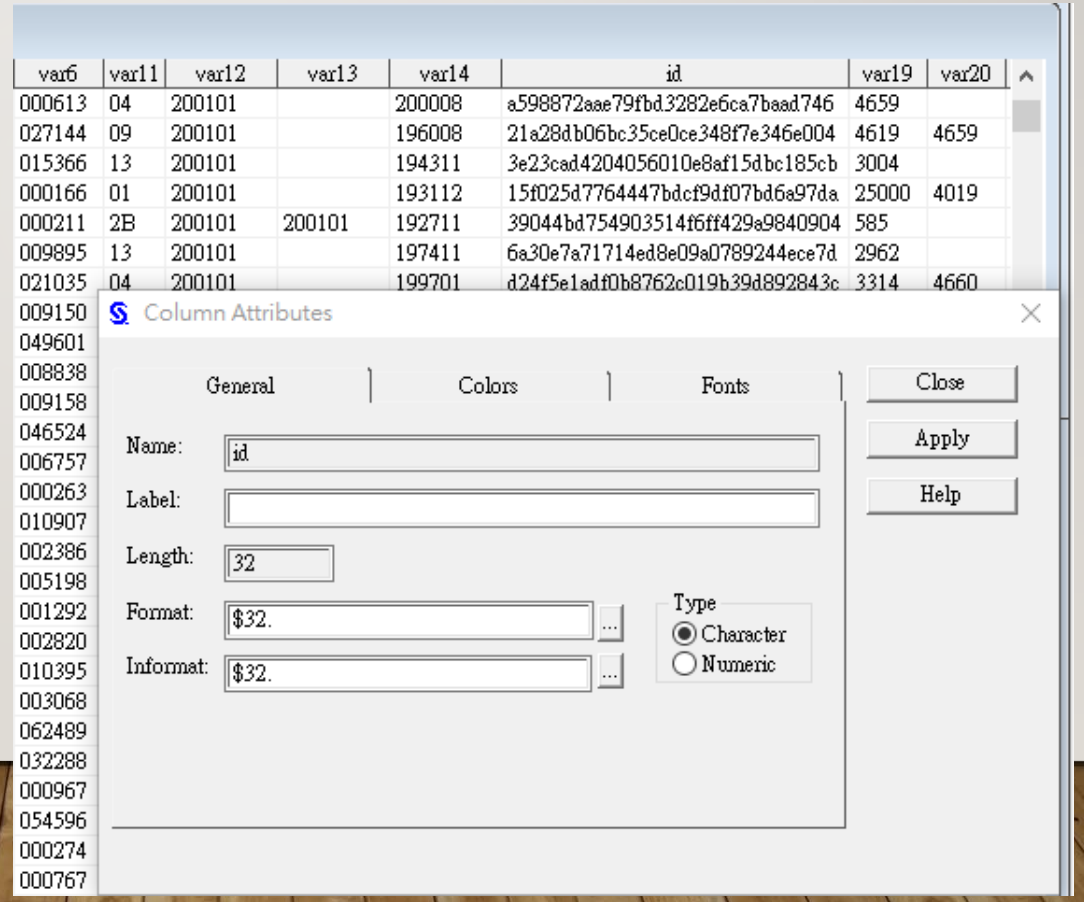
id
a598872aae79fbd3282e6ca7baad746
21a28db06bc35ce0ce348f7e346e004
3e23cad4204056010e8af15dbc185cb
15f025d7764447bdcf9df07bd6a97da
39044bd754903514f6ff429a9840904
6a30e7a71714ed8e09a0789244ece7d
d24f5e1adf0b8762c019b39d892843c
6a30e7a71714ed8e09a0789244ece7d
21e9141e4ddc5b6b149f277036c41c0
91ff037b82beb600289701af13c245e
91ff037b82beb600289701af13c245e
21e9141e4ddc5b6b149f277036c41c0
8e2dc3244ada07997cd22ff1f279e305
9b54b7354e946f0cf2c974e5be7890d
fa480b7d2b302ea5e2b7d5611f46926
8e2dc3244ada07997cd22ff1f279e305
7c8473f0cc2ba8b0e3ee7ef050fede51
a598872aae79fbd3282e6ca7baad746
25a63d35308899269253bf0e700056
88bd51e60b2c6999a53c527e3a0e361
0423860316612b33afedd13701181af
824d8d2c3294185d2a747374e71d1a
ed9ccf729e248e67b084f1ad7128ba1e
9b54b7354e946f0cf2c974e5be7890d



27 基本語法 – 改變欄位名稱 (RENAME)

- 改變variable (變數)名稱，以便於資料連結

```
data aa3;
    set opd_2001;
    rename var15=id;
run;
```



EXERCISE 2

- 將cd檔只留下var1-var6、var11-var15、var19-var21、var37等欄位
- 重新命名：身分證字號欄位為id，就醫日期為func_ym、出生日期為birth_ym

29 基本語法 –資料集依指定欄位排序

```
proc sort data= aa3;
by  id func_ym;
run;

proc sort data= aa3 ;
by id descending func_ym;
run;
```

宣告SAS對資料集 “aa3” 進行**排序處理**

依**id,就醫日期**兩欄位進行排序 (由小至大)

run;做為此**資料步驟**的**結尾**

將**就醫日期**由大至小排序

30 基本語法 –資料集依指定欄位排序

2. 同一id內，日期由先至後排序

1. Id由小至大排序

func_ym	var13	birth_ym	id
200101		198504	00a2e11374554409ff1a778dfad2ed0c
200102		198504	00a2e11374554409ff1a778dfad2ed0c
200102		198504	00a2e11374554409ff1a778dfad2ed0c
200102		198504	00a2e11374554409ff1a778dfad2ed0c
200102		198504	00a2e11374554409ff1a778dfad2ed0c
200103		198504	00a2e11374554409ff1a778dfad2ed0c
200103		198504	00a2e11374554409ff1a778dfad2ed0c
200103		198504	00a2e11374554409ff1a778dfad2ed0c
200103		198504	00a2e11374554409ff1a778dfad2ed0c
200109		198504	00a2e11374554409ff1a778dfad2ed0c
200109		198504	00a2e11374554409ff1a778dfad2ed0c
200109		198504	00a2e11374554409ff1a778dfad2ed0c
200111		198504	00a2e11374554409ff1a778dfad2ed0c
200101	200101	196407	011758f7dcd53031537f058819d427
200101	200101	196407	011758f7dcd53031537f058819d427
200101		196811	01c7dabf8b8c3b82120e43bb6b5937b
200105		196811	01c7dabf8b8c3b82120e43bb6b5937b
200105		196811	01c7dabf8b8c3b82120e43bb6b5937b
200106		196811	01c7dabf8b8c3b82120e43bb6b5937b
200107		196811	01c7dabf8b8c3b82120e43bb6b5937b
200107		196811	01c7dabf8b8c3b82120e43bb6b5937b
200107		196811	01c7dabf8b8c3b82120e43bb6b5937b
200107		196811	01c7dabf8b8c3b82120e43bb6b5937b

func_ym	var13	birth_ym	id
200111		198504	00a2e11374554409ff1a778dfad2ed0c
200109		198504	00a2e11374554409ff1a778dfad2ed0c
200109		198504	00a2e11374554409ff1a778dfad2ed0c
200109		198504	00a2e11374554409ff1a778dfad2ed0c
200103		198504	00a2e11374554409ff1a778dfad2ed0c
200103		198504	00a2e11374554409ff1a778dfad2ed0c
200103		198504	00a2e11374554409ff1a778dfad2ed0c
200103		198504	00a2e11374554409ff1a778dfad2ed0c
200102		198504	00a2e11374554409ff1a778dfad2ed0c
200102		198504	00a2e11374554409ff1a778dfad2ed0c
200102		198504	00a2e11374554409ff1a778dfad2ed0c
200102		198504	00a2e11374554409ff1a778dfad2ed0c
200101		198504	00a2e11374554409ff1a778dfad2ed0c
200101	200101	196407	011758f7dcd53031537f058819d427
200101	200101	196407	011758f7dcd53031537f058819d427
200107		196811	
200107		196811	
200107		196811	01c7dabf8b8c3b82120e43bb6b5937b
200107		196811	01c7dabf8b8c3b82120e43bb6b5937b
200106		196811	01c7dabf8b8c3b82120e43bb6b5937b
200105		196811	01c7dabf8b8c3b82120e43bb6b5937b
200105		196811	01c7dabf8b8c3b82120e43bb6b5937b
200101		196811	01c7dabf8b8c3b82120e43bb6b5937b

同一id內，日期由後至前排序

31 基本語法 – NODUPKEY

```
proc sort data=aa3 nodupkey;
   by id;
run;
```

1. 先將資料集”aa”中資料，依id由小至大排序

2. 排序後資料集，每一個id只取第一筆資料

```
proc sort data=aa3 out=bb nodupkey;
by id;
run;
```

將此段程式執行結果，另存至資料集”bb”

```
NOTE: There were 14280 observations read from the data set WORK.AA3.
NOTE: 13434 observations with duplicate key values were deleted.
NOTE: The data set WORK.BB has 846 observations and 15 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.01 seconds
      cpu time            0.00 seconds
```

32 基本語法 – 擷取功能函數 SUBSTR()

```
DATA cc;
SET bb;
   year = substr(left(func_ym), 1, 4);
   month = substr(left(func_ym), 5, 2);
   day = '15'
RUN;
```

- 語法：substr(變項欄位，起始位置，長度)
- left()：將指定欄位內文字歸左，避免擷取到空白值

擷取年度，由欄位左邊第1個位置開始，一共擷取4碼

擷取月份，由欄位左邊第5個位置開始，一共擷取2碼

擷取日期，由於就醫日期僅提供年月，故假設為15號

33 基本語法 – 日期轉換函數 MDY(MM, DD, YY)

```
DATA cc;
SET bb;
  year = substr(left(func_ym), 1, 4);
  month = substr(left(func_ym), 5, 2);
  day = '15';
  func_date = mdy(month, day, year);
  func_date1 = mdy(substr(left(func_ym),
  5, 2), '15',  substr(left(func_ym), 1, 4));
RUN;
```

- 語法：mdy(月份， 日期， 年度)

實務運用上常用的寫法!!!

34 基本語法 – 日期轉換函數 MDY(MM, DD, YY)

func_ym	var13	birth_ym	year	month	day	func_date	func_date
200101		198504	2001	01	15	14990	14990
200101	200101	196407	2001	01	15	14990	14990
200101		196811	2001	01	15	14990	14990
200105		195910	2001	05	15	15110	15110
200104		197906	2001	04	15	15080	15080
200103		195905	2001	03	15	15049	15049
200101		196611	2001	01	15	14990	14990
200012	200101	192501	2000	12	15	14959	14959
200101		197409	2001	01	15	14990	14990
200104		199602	2001	04	15	15080	15080
200101		195308	2001	01	15	14990	14990
200101		199802	2001	01	15	14990	14990
200103		197107	2001	03	15	15049	15049
200110		194909	2001	10	15	15263	15263
200103		191506	2001	03	15	15049	15049
200102		192604	2001	02	15	15021	15021
200101		195110	2001	01	15	14990	14990
200101	200101	195402	2001	01	15	14990	14990
200106		197502	2001	06	15	15141	15141
200102		199206	2001	02	15	15021	15021
200012	200101	194705	2000	12	15	14959	14959
200101		197608	2001	01	15	14990	14990
200112		197712	2001	12	15	15324	15324
200101		197910	2001	01	15	14990	14990
200105		195403	2001	05	15	15110	15110

(**Note**): 透過mdy()函數，會將日期轉換為距離1960年1月1日的天數。因此，buy_date及buy_date1內數字代表購買日期距離1960年1月1日的天數。

35 基本語法－轉換日期格式

```
DATA cc;
SET bb;
   year = substr(left(func_ym), 1, 4);
   month = substr(left(func_ym), 5, 2);
   day = '15';
   func_date = mdy(month, day, year);
   func_date1 = mdy(substr(left(func_ym), 5, 2), '15',
   substr(left(func_ym), 1, 4));
   format func_date mmddyy8. ;
   format func_date1 yymmdd10. ;
   drop year month day;
RUN;
```

將func_date中天數轉換成以mmddyy8.的格式呈現

將func_date1中天數轉換成以yymmdd10.的格式呈現

36 基本語法－轉換日期格式

func_ym	var13	birth_ym	year	month	day	func_date	func_date1
200101		198504	2001	01	15	01/15/01	2001-01-15
200101	200101	196407	2001	01	15	01/15/01	2001-01-15
200101		196811	2001	01	15	01/15/01	2001-01-15
200105		195910	2001	05	15	05/15/01	2001-05-15
200104		197906	2001	04	15	04/15/01	2001-04-15
200103		195905	2001	03	15	03/15/01	2001-03-15
200101		196611	2001	01	15	01/15/01	2001-01-15
200012	200101	192501	2000	12	15	12/15/00	2000-12-15
200101		197409	2001	01	15	01/15/01	2001-01-15
200104		199602	2001	04	15	04/15/01	2001-04-15
200101		195308	2001	01	15	01/15/01	2001-01-15
200101		199802	2001	01	15	01/15/01	2001-01-15
200103		197107	2001	03	15	03/15/01	2001-03-15
200110		194909	2001	10	15	10/15/01	2001-10-15
200103		191506	2001	03	15	03/15/01	2001-03-15
200102		192604	2001	02	15	02/15/01	2001-02-15
200101		195110	2001	01	15	01/15/01	2001-01-15
200101	200101	195402	2001	01	15	01/15/01	2001-01-15
200106		197502	2001	06	15	06/15/01	2001-06-15
200102		199206	2001	02	15	02/15/01	2001-02-15
200012	200101	194705	2000	12	15	12/15/00	2000-12-15
200101		197608	2001	01	15	01/15/01	2001-01-15
200112		197712	2001	12	15	12/15/01	2001-12-15
200101		197910	2001	01	15	01/15/01	2001-01-15
200105		195403	2001	05	15	05/15/01	2001-05-15
200103		196707	2001	03	15	03/15/01	2001-03-15
200101		198112	2001	01	15	01/15/01	2001-01-15
200103		198605	2001	03	15	03/15/01	2001-03-15
200101		199505	2001	01	15	01/15/01	2001-01-15
200102	200102	198703	2001	02	15	02/15/01	2001-02-15

其他日期格式：
http://documentation.sas.com/?docsetId=lrcon&docsetTarget=p1wj0wt2ebe2a0n11v4lem9hdc0v.htm&docsetVersion=9.4&locale=en

37 基本語法 - 四則運算

- 計算就醫年齡

```
DATA dd;
SET cc;
birth_yr = substr(left(birth_ym), 1, 4) *1;
AGE = year(func_date) - birth_yr;
RUN;
```

- **語法：** + - * /
  加減乘除四則運算
- **語法：** *1
  將文字格式轉換成數字格式，方便後續運算
- **語法：** year() month() day()
  運算函數，得出日期欄位的年、月、日

38 SAS OPERATORS

Symbol	Definition
**	exponentiation
*	multiplication
/	division
+	addition
-	subtraction

Symbol	Definition
=	equal to
^=	not equal to
~=	not equal to
>	greater than
<	less than
>= , => (註)	greater than or equal to
<= , =< (註)	less than or equal to

註：also accepted for compatibility with previous releases of SAS. It is not supported in WHERE clauses or in PROC SQL

EXERCISE 3

- 將就醫日期與出生日期轉換成日期格式。
- 利用就醫日期與出生日期計算年齡。

TO BE CONTINUED……

41 基本語法 - FORMAT

```
proc format;
    value agegr
    low-<19="children"
     19-60="adult"
    61-high="elder"
    ;
run;
```

```
data ee;
    set dd;
    agegr=age;
    format agegr agegr.;
run;
```

AGE	agegr
16	children
37	adult
33	adult
42	adult
22	adult
42	adult
35	adult
75	elder
27	adult
5	children
48	adult
3	children
30	adult
52	adult
86	elder
75	elder
50	adult
47	adult
26	adult
9	children
53	adult

42 基本語法 – IF … THEN…

- 將欄位內資料進行分組，例：將年齡分成<=18, 19-65, >65 歲三組
- 將性別重新編碼

```
data ee1;
set dd;
   if age <=18 then age_group=1;
   else if 18 < age <=65 then age_group=2;
   else if age>65 then age_group=3;
   else age_group=999;
   if var37='F' then sex=0; else sex=1;
run;
```

條件設定應互斥且周延

var37	func_date	func_date1	birth_yr	AGE	age_group	sex
M	01/15/01	2001-01-15	1985	16	1	1
M	01/15/01	2001-01-15	1964	37	2	1
M	01/15/01	2001-01-15	1968	33	2	1
M	05/15/01	2001-05-15	1959	42	2	1
M	04/15/01	2001-04-15	1979	22	2	1
F	03/15/01	2001-03-15	1959	42	2	0
F	01/15/01	2001-01-15	1966	35	2	0
M	12/15/00	2000-12-15	1925	75	3	1
M	01/15/01	2001-01-15	1974	27	2	1
M	04/15/01	2001-04-15	1996	5	1	1
M	01/15/01	2001-01-15	1953	48	2	1
F	01/15/01	2001-01-15	1998	3	1	0
M	03/15/01	2001-03-15	1971	30	2	1
M	10/15/01	2001-10-15	1949	52	2	1
F	03/15/01	2001-03-15	1915	86	3	0
F	02/15/01	2001-02-15	1926	75	3	0
M	01/15/01	2001-01-15	1951	50	2	1
M	01/15/01	2001-01-15	1954	47	2	1
M	06/15/01	2001-06-15	1975	26	2	1
M	02/15/01	2001-02-15	1992	9	1	1
M	12/15/00	2000-12-15	1947	53	2	1
F	01/15/01	2001-01-15	1976	25	2	0

43 基本語法 – IF … THEN…

- 條件限制，例：只分析男性病患

```
data ff;
    set ee1;
    if sex=1 then output;
run;
```

var37	func_date	func_date1	birth_yr	AGE	age_group	sex
M	01/15/01	2001-01-15	1985	16	1	1
M	01/15/01	2001-01-15	1964	37	2	1
M	01/15/01	2001-01-15	1968	33	2	1
M	05/15/01	2001-05-15	1959	42	2	1
M	04/15/01	2001-04-15	1979	22	2	1
M	12/15/00	2000-12-15	1925	75	3	1
M	01/15/01	2001-01-15	1974	27	2	1
M	04/15/01	2001-04-15	1996	5	1	1
M	01/15/01	2001-01-15	1953	48	2	1
M	03/15/01	2001-03-15	1971	30	2	1
M	10/15/01	2001-10-15	1949	52	2	1
M	01/15/01	2001-01-15	1951	50	2	1
M	01/15/01	2001-01-15	1954	47	2	1
M	06/15/01	2001-06-15	1975	26	2	1
M	02/15/01	2001-02-15	1992	9	1	1
M	12/15/00	2000-12-15	1947	53	2	1
M	12/15/01	2001-12-15	1977	24	2	1
M	05/15/01	2001-05-15	1954	47	2	1
M	03/15/01	2001-03-15	1967	34	2	1
M	01/15/01	2001-01-15	1981	20	2	1
M	03/15/01	2001-03-15	1986	15	1	1
M	01/15/01	2001-01-15	1995	6	1	1

44 基本語法 – IF … THEN…

- 配合substr function 進行條件篩選。篩選有HTN診斷 (ICD9:401-405)之病人

```
data htn;
set a.cd_2001;
if substr(left(var19),1,3) in ('401','402','403','404','405') or
  substr(left(var20),1,3) in ('401','402','403','404','405') or
  substr(left(var21),1,3) in ('401','402','403','404','405') then output;
run;
```

- 篩選有DM診斷 (ICD9:250)之病人

```
data dm;
set a.cd_2001;
if substr(left(var19),1,3) ='250' or substr(left(var20),1,3) ='250' or substr(left(var21),1,3) ='250'
then output;
run;
```

EXERCISE 4

- 將年齡分成<=18, 19-30, 31-50, 51-70, >70 歲五組
- 產生新的變數，將性別重新編碼：male=1, female=0
- 利用cd_2001檔擷選具有HTN診斷的病人

46 資料合併 SET

- 將資料做縱向連結 (增加observation)

```
data car6;
set cars3 cars4;
run;
```

VIEWTABLE: Work.Cars3

	make	model	mpg	weight	price
1	AMC	Concord	22	2930	4099
2	AMC	Pacer	17	3350	4749
3	AMC	Spirit	22	2640	3799

VIEWTABLE: Work.Cars4

	make	model	mpg	weight	price
1	AMC	Concord	22	2930	4099
2	AMC	Pacer	17	3350	4749
3	AMC	Spirit	22	2640	3799
4	Buick	Century	20	3250	4816
5	Buick	Electra	15	4080	7827

VIEWTABLE: Work.Car6

	make	model	mpg	weight	price
1	AMC	Concord	22	2930	4099
2	AMC	Pacer	17	3350	4749
3	AMC	Spirit	22	2640	3799
4	AMC	Concord	22	2930	4099
5	AMC	Pacer	17	3350	4749
6	AMC	Spirit	22	2640	3799
7	Buick	Century	20	3250	4816
8	Buick	Electra	15	4080	7827

➢ 欄位名稱與格式需相同

	make	model	mpg	weight	price	var1	var2	var3	var4
1	AMC	Concord	22	2930	4099				
2	AMC	Pacer	17	3350	4749				
3	AMC	Spirit	22	2640	3799				
4			.	.	.	200101	1	0b64808157449e660c2b767d8bf2cce	200102
5			.	.	.	200101	1	7ead6cd0b8d12f97fc285ae5f8c68dc2	200102
6			.	.	.	200101	1	a78800b3ace1fdd28af8505332f5d41f	200102
7			.	.	.	200105	1	1a71b6fddc9c366d9d48f5e356eae18	200106
8			.	.	.	200104	1	6f0db153524d2b1babda6030b0fa42b	200105

```
1061  data test2;
1062  set cc test1;
ERROR: Variable func_ym has been defined as both character and numeric.
1063  run;
```

47 資料合併 MERGE

- 將資料做橫向連結 (增加變數variable)

```
proc sort data=a.cd_2001; by var1-var6; run;
proc sort data=oo_2001; by var1-var6; run;

data cdoo_2001;
    merge a.cd_2001( in=x) oo_2001 (in=y);
    by var1-var6;
    if x and y;
run;
```

注意：使用MERGE時，要記得先將資料依據"**鍵值**"排序(PROC SORT)。

x,y 可以自訂，但不可與dataset中的variable名稱相同

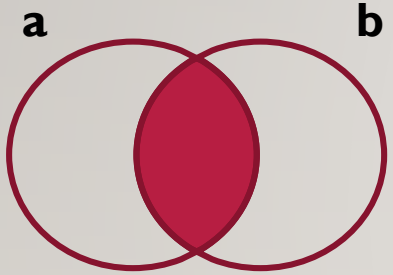
依據"**鍵值**"連結

兩個資料表間取"**交集**"

48 資料合併

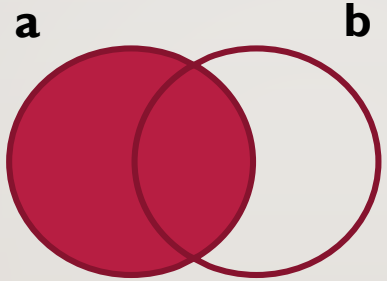
SAS語法為例：
DATA xxx ; MERGE a b;
If…..

交集



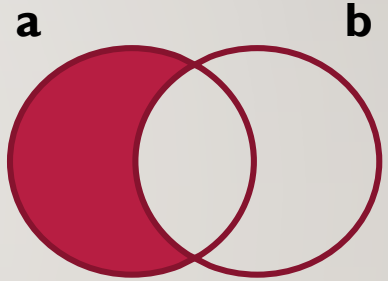
if a and b;

$a \cup (a \cap b)$



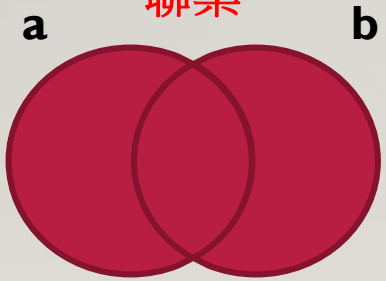
if a;

$a \cap b'$



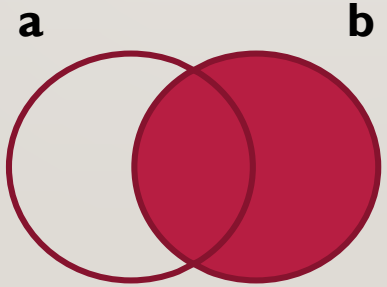
if a and not b;

聯集



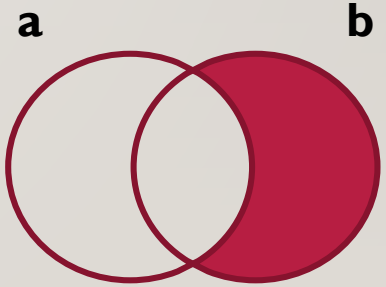
無需用 if

$b \cup (a \cap b)$



if b;

$a' \cap b$



if b and not a;

49 資料轉置 - TRANSPOSE

有by就要先sort

```
proc sort data=drug_2001; by var15 var12 ; run;

proc transpose data=drug_2001 out=atc_2001 prefix=atc;
    var nhi40;
    by var15 var12;
run;
```

欲轉置的variable

依照哪些欄位轉置

轉置後的欄位名稱