

SAS講義(一)

一、 認識SAS

二、 如何撰寫SAS程式讀取外部資料檔？

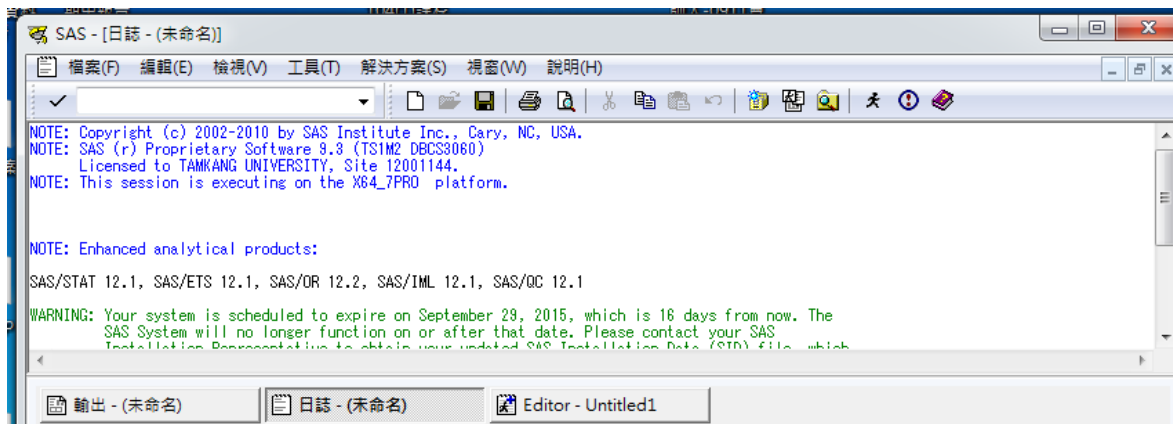
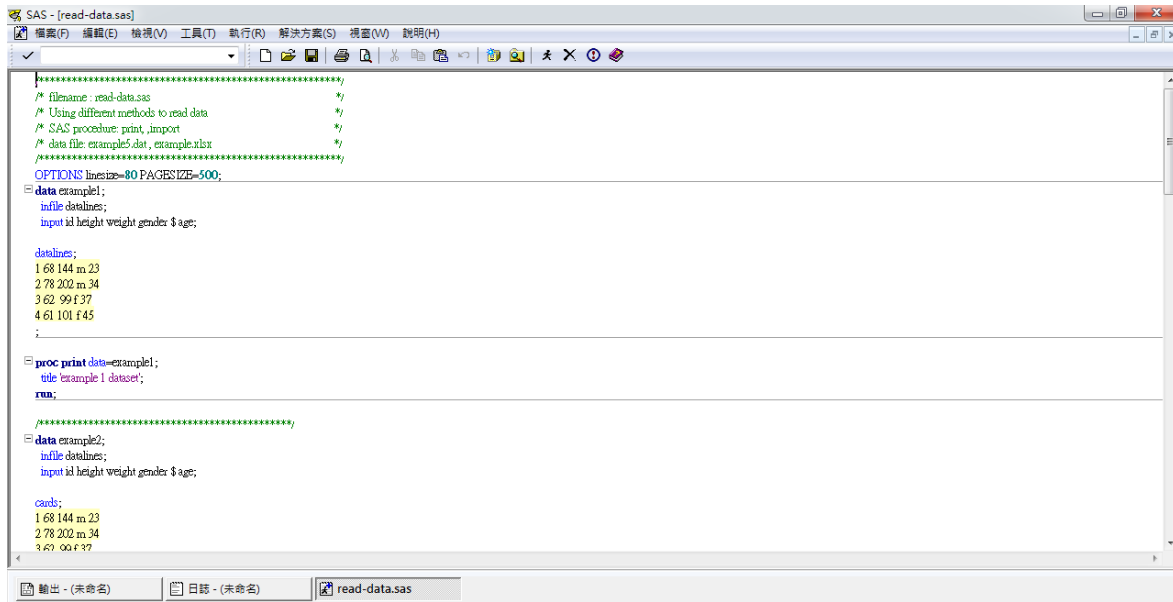
三、 如何準備SAS程式要讀取外部資料的文字檔（ASCII）？

四、 Data and Code Book（資料建檔與編碼）

一、認識SAS(Statistical Analysis System): *.sas、*.log、*.lst之檔案

為ASCII文字檔：

- SAS程式：**filename.sas (*.sas)**
- Log：**filename.log (*.log)**
- Output：**filename.lst (*.lst)**
filename.mht (*.mht) (Web page archive file format)



```

SAS - [日誌 - (未命名)]
檔案(F) 編輯(E) 檢視(V) 工具(T) 解決方案(S) 視窗(W) 說明(H)
WARNING: to obtain more information on your warning period.
1  OPTIONS linesize=80 PAGESIZE=500;
2  data example1;
3  infile datalines;
4  input id height weight gender $ age;
5  datalines;
NOTE: The data set WORK.EXAMPLE1 has 4 observations and 5 variables.
NOTE: DATA statement used (Total process time):
      real time    0.18 seconds
      cpu time     0.12 seconds

11 ;
12
13 proc print data=example1;
NOTE: 正在寫入 HTML Body (本文) 檔案: sashtml.htm
14 title 'example 1 dataset';
15 run;
NOTE: There were 4 observations read from the data set WORK.EXAMPLE1.
NOTE: PROCEDURE PRINT used (Total process time):
      real time    0.38 seconds
      cpu time     0.78 seconds

16 /*****
17 data example2;
18   infile datalines;
19   input id height weight gender $ age;
20   cards;
NOTE: The data set WORK.EXAMPLE2 has 4 observations and 5 variables.
NOTE: DATA statement used (Total process time):
      real time    0.01 seconds
      cpu time     0.01 seconds

26 ;
27
28 proc print data=example2;
29   title 'example 2 dataset';
30 run;

```

輸出 - (未命名) | 日誌 - (未命名) | read-data.sas * | Results Viewer - sasht...

SAS - [輸出 - (未命名)]

example 1 dataset
2015年08月13日 星期日 下午08時32分26秒

Obs	id	height	weight	gender	age
1	1	88	144	m	23
2	2	78	202	m	34
3	3	82	99	f	37
4	4	61	101	f	45

輸出 - (未命名) | 日誌 - (未命名) | read-data.sas | Results Viewer - SAS ...

SAS - [Results Viewer - SAS Output]

example 1 dataset

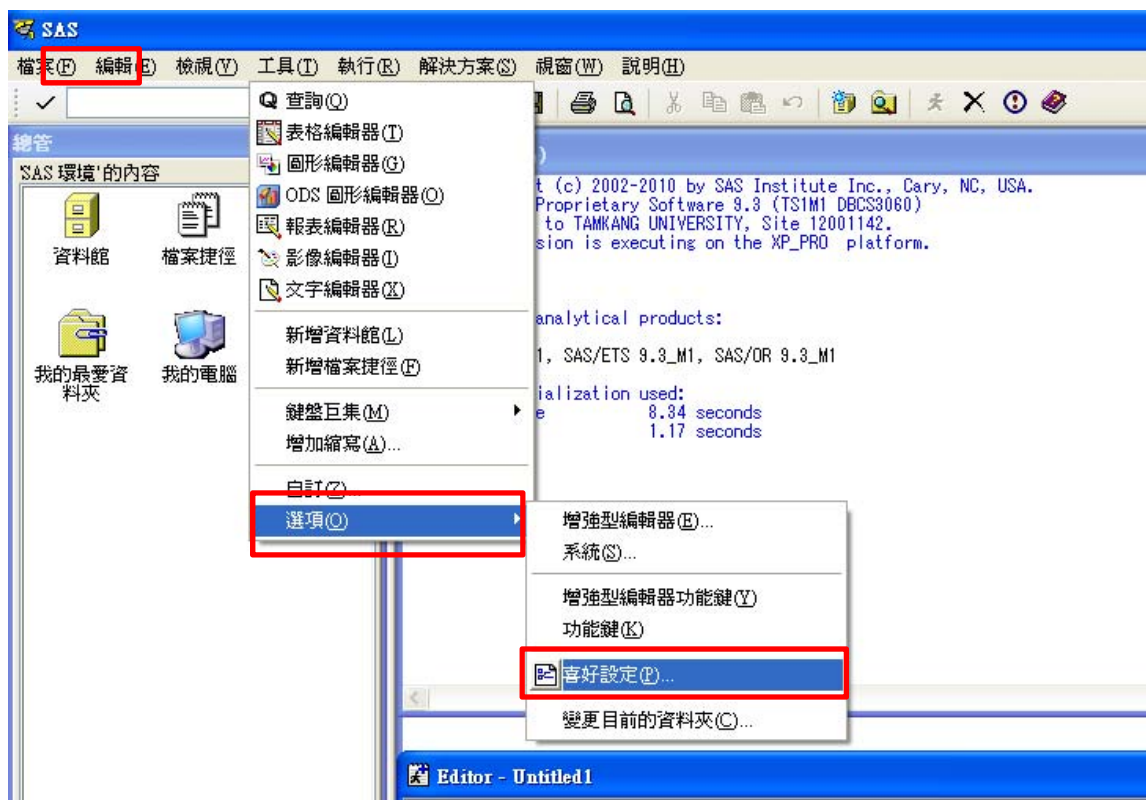
Obs	id	height	weight	gender	age
1	1	68	144	m	23
2	2	78	202	m	34
3	3	62	99	f	37
4	4	61	101	f	45

輸出 - (未命名) | 日誌 - (未命名) | read-data.sas | Results Viewer - SAS ...

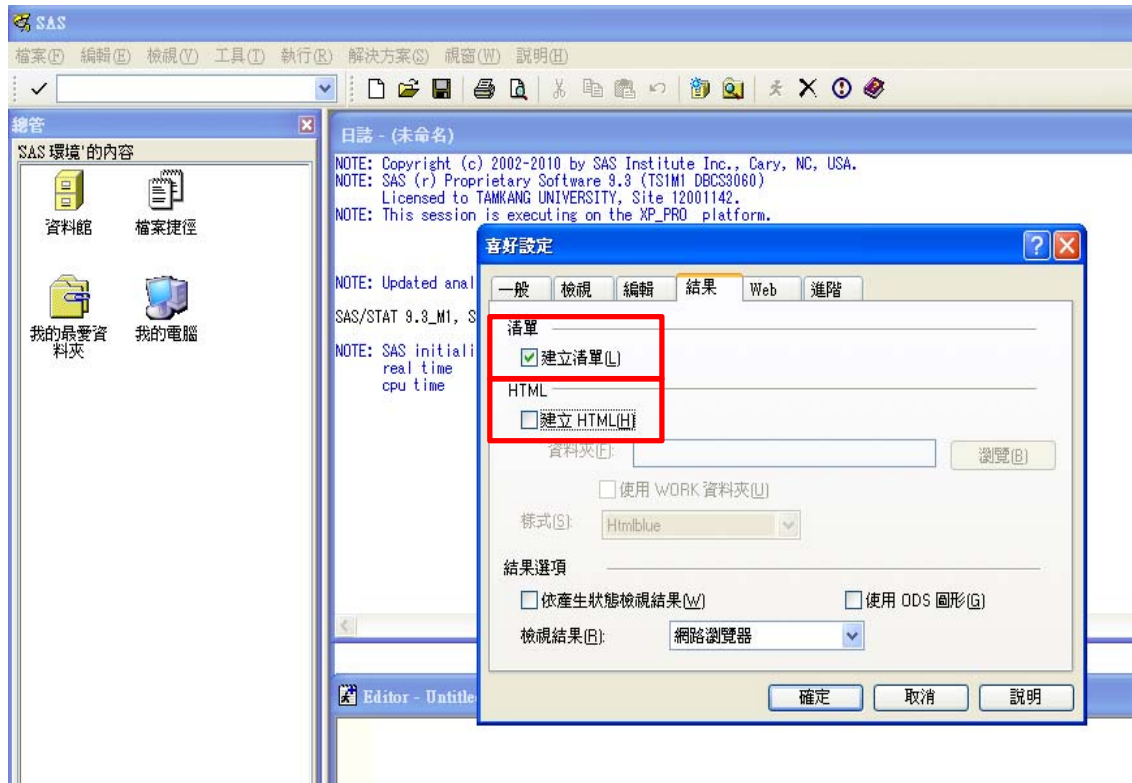
完成 | C:\Users\iris

若無輸出檔，請依以下步驟改SAS輸出設定

1. 打開「SAS」→「工具」→「選項」→「喜好設定」



2. 喜好設定的表單→「結果」→勾選「建立清單」，將「建立 HTML」的選項取消→確定即可



二、如何撰寫SAS程式讀取外部資料檔？

SAS程式 - 外部讀檔

```
options linesize=150 pagesize=500;
data data;
infile'C:\Documents and Settings\POLICE\桌面\101(1)資料 982 運統成績.dat' dlm='';
input ID class $ midtermexam finalreport finalexam hw1 hw2 presence end;
/*把在 word 剪下的第一列變數名稱貼上，但變數間 “,” 改為 “空一格” */
```

```
/**列出所讀取資料**/
```

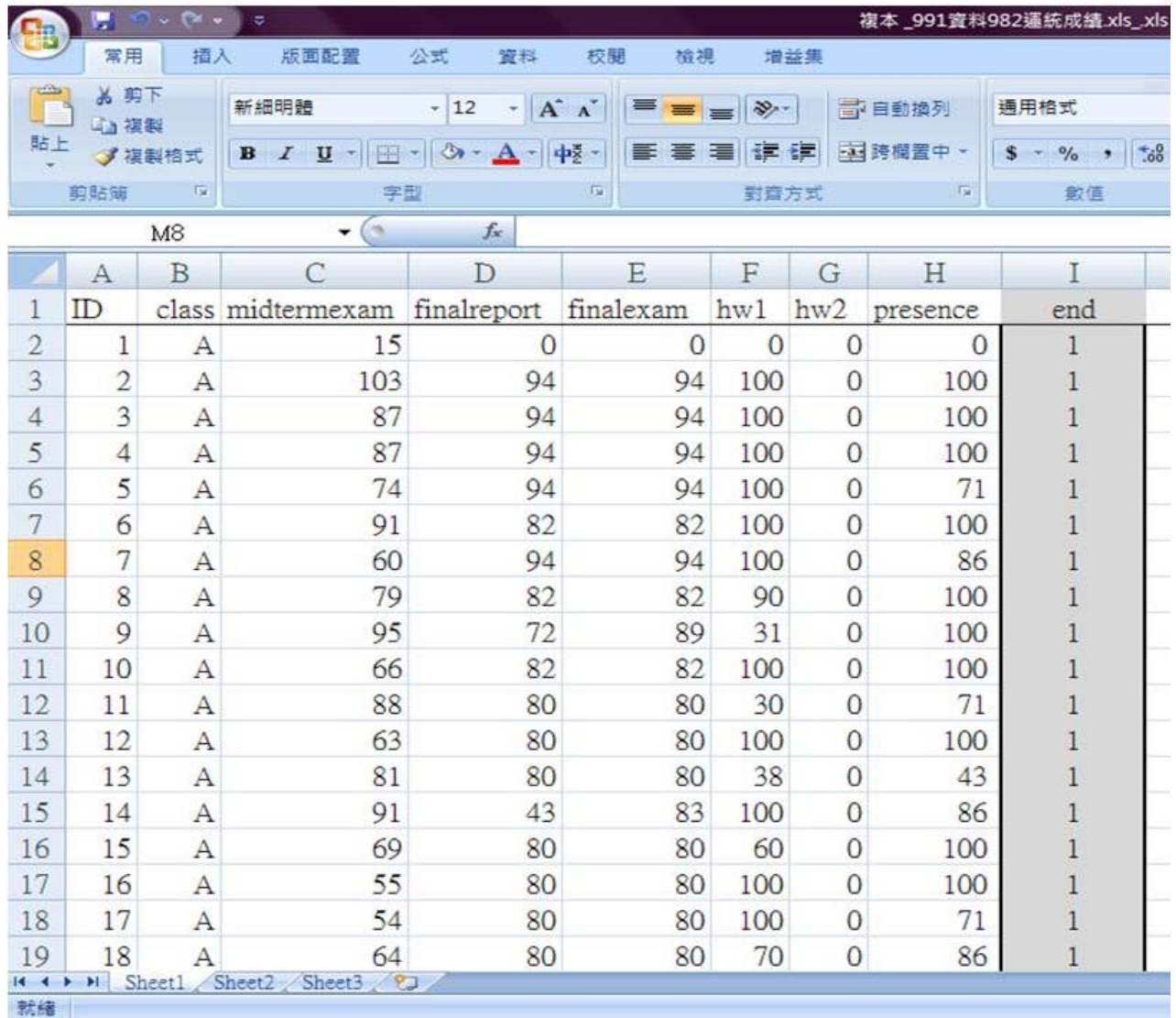
```
proc print data=data;
var ID class midtermexam finalreport finalexam hw1 hw2 presence end;
run;
```

```
proc freq data=data;
tables ID class midtermexam finalreport finalexam hw1 hw2 presence end;
run;
```

三、 如何準備SAS程式要讀取外部資料的文字檔（ASCII）？

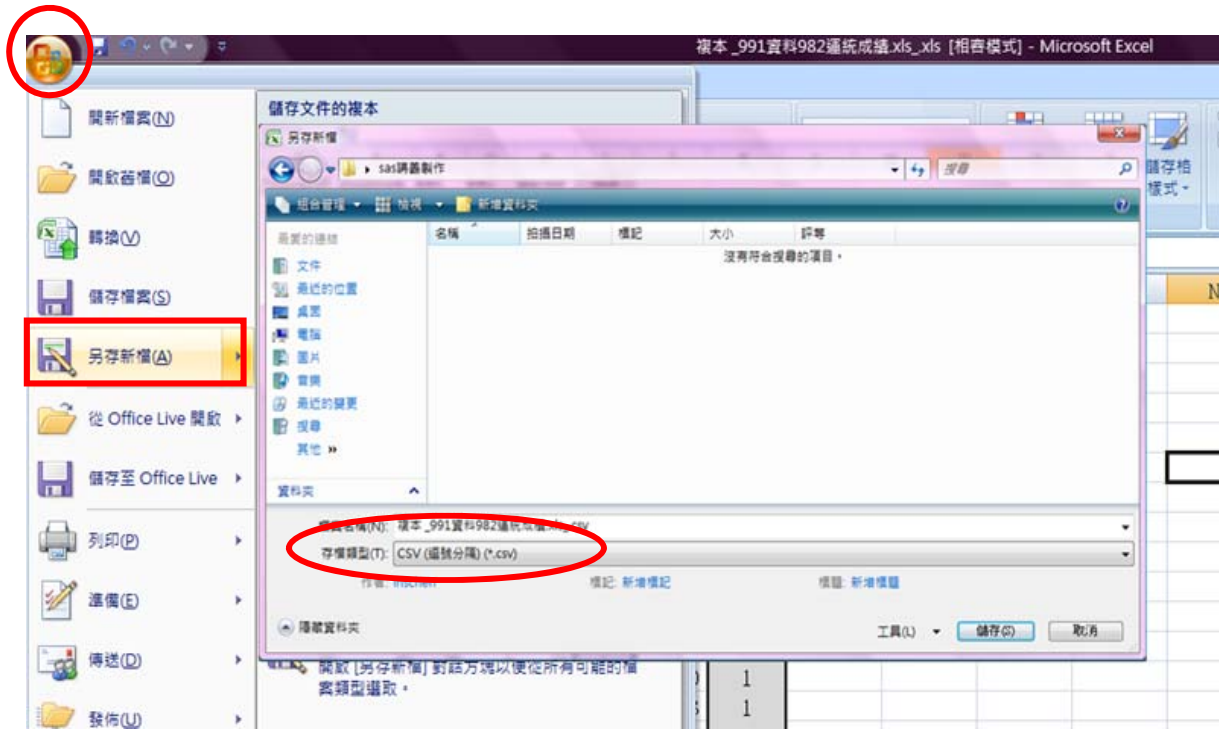
常見資料存在 Excel 資料檔中：轉換 Excel 資料檔為 SAS 可讀資料檔(ASCII code)，撰寫 SAS 程式讀此外部資料檔。

步驟 1：多加一欄為「end」，並將內容皆輸入為 1。（此動作是避免讀取大型資料時，最後一欄位有遺漏值，另可 check 資料是否正確讀到。）

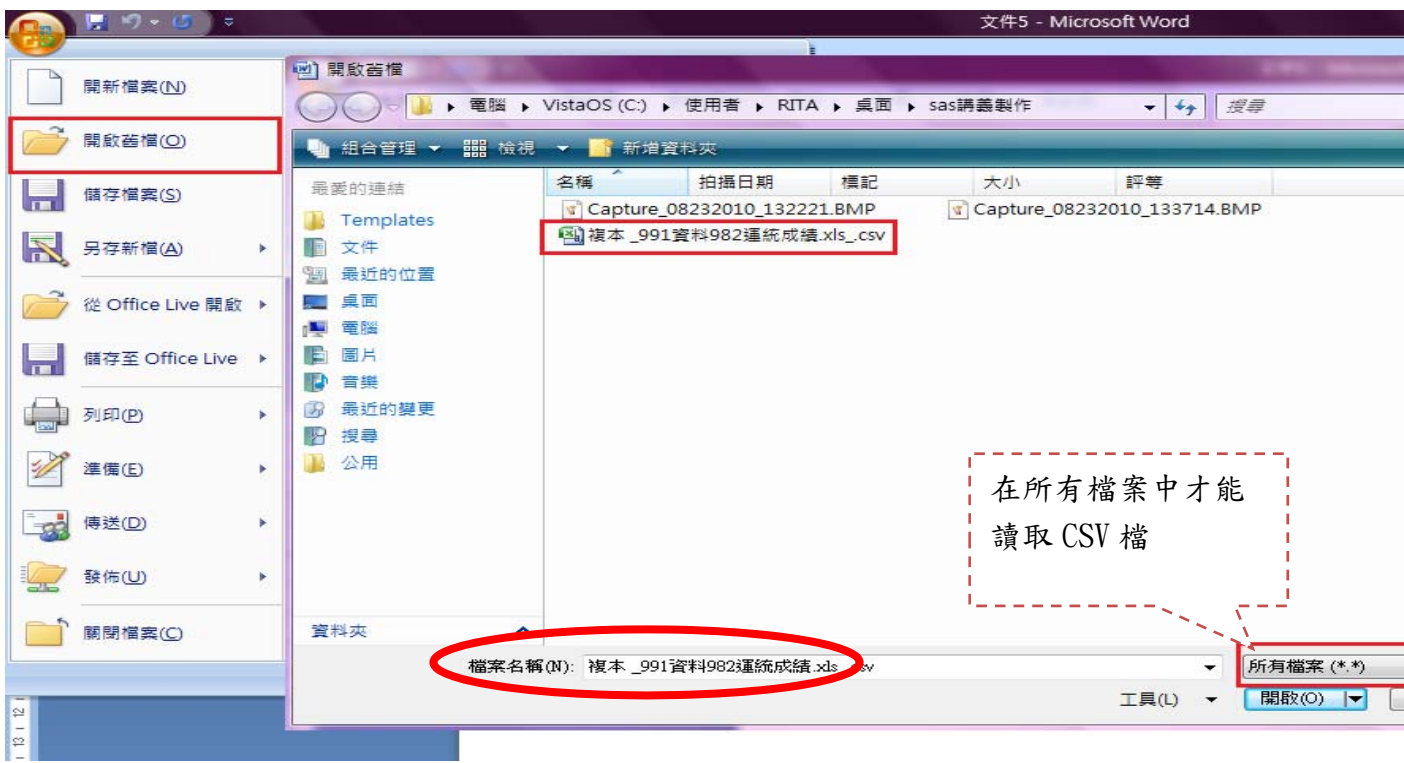


	A	B	C	D	E	F	G	H	I
1	ID	class	midtermexam	finalreport	finalexam	hw1	hw2	presence	end
2	1	A	15	0	0	0	0	0	1
3	2	A	103	94	94	100	0	100	1
4	3	A	87	94	94	100	0	100	1
5	4	A	87	94	94	100	0	100	1
6	5	A	74	94	94	100	0	71	1
7	6	A	91	82	82	100	0	100	1
8	7	A	60	94	94	100	0	86	1
9	8	A	79	82	82	90	0	100	1
10	9	A	95	72	89	31	0	100	1
11	10	A	66	82	82	100	0	100	1
12	11	A	88	80	80	30	0	71	1
13	12	A	63	80	80	100	0	100	1
14	13	A	81	80	80	38	0	43	1
15	14	A	91	43	83	100	0	86	1
16	15	A	69	80	80	60	0	100	1
17	16	A	55	80	80	100	0	100	1
18	17	A	54	80	80	100	0	71	1
19	18	A	64	80	80	70	0	86	1

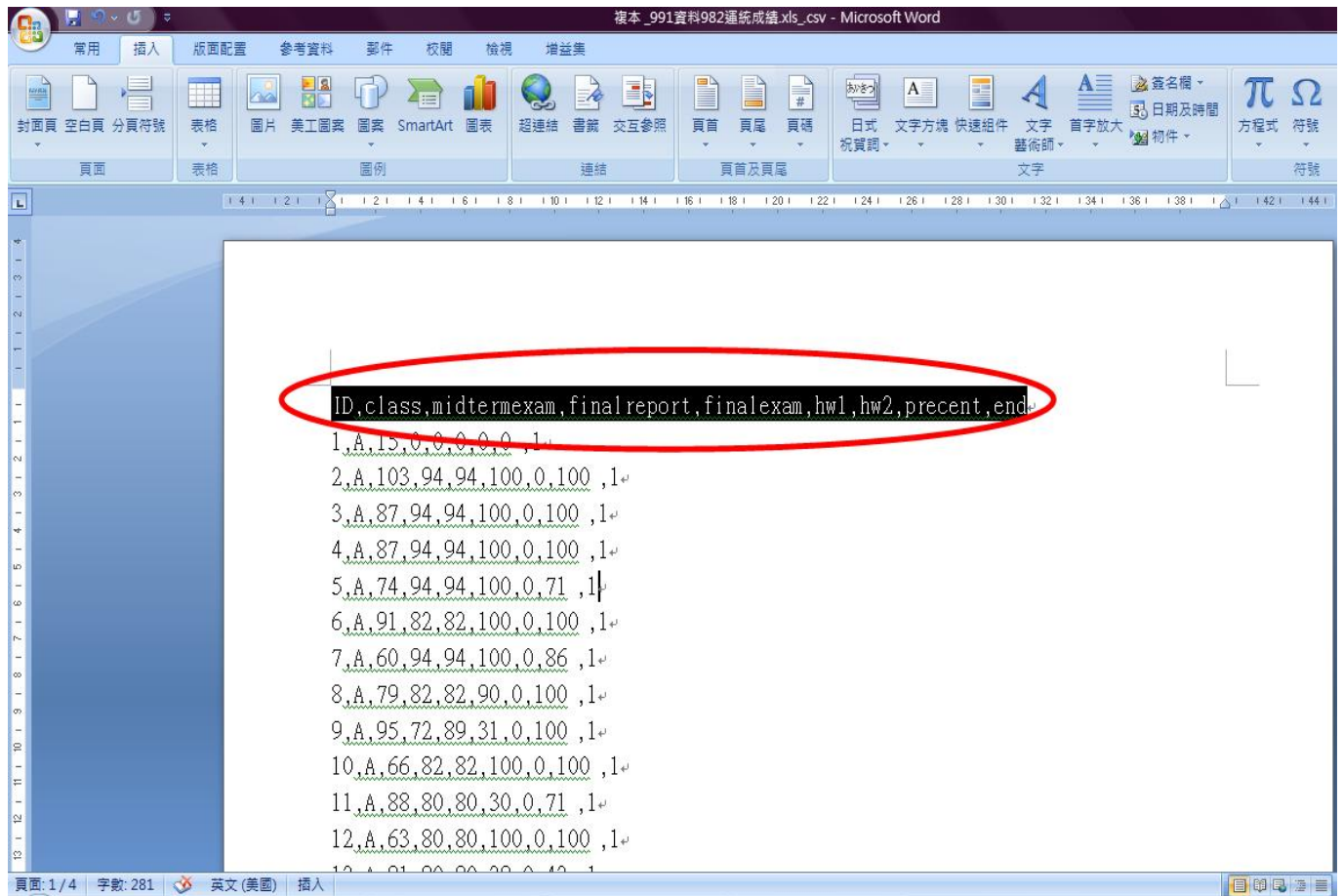
步驟 2：Office 按鈕→另存新檔→存檔類型:CSV (逗號分隔)(* .csv)



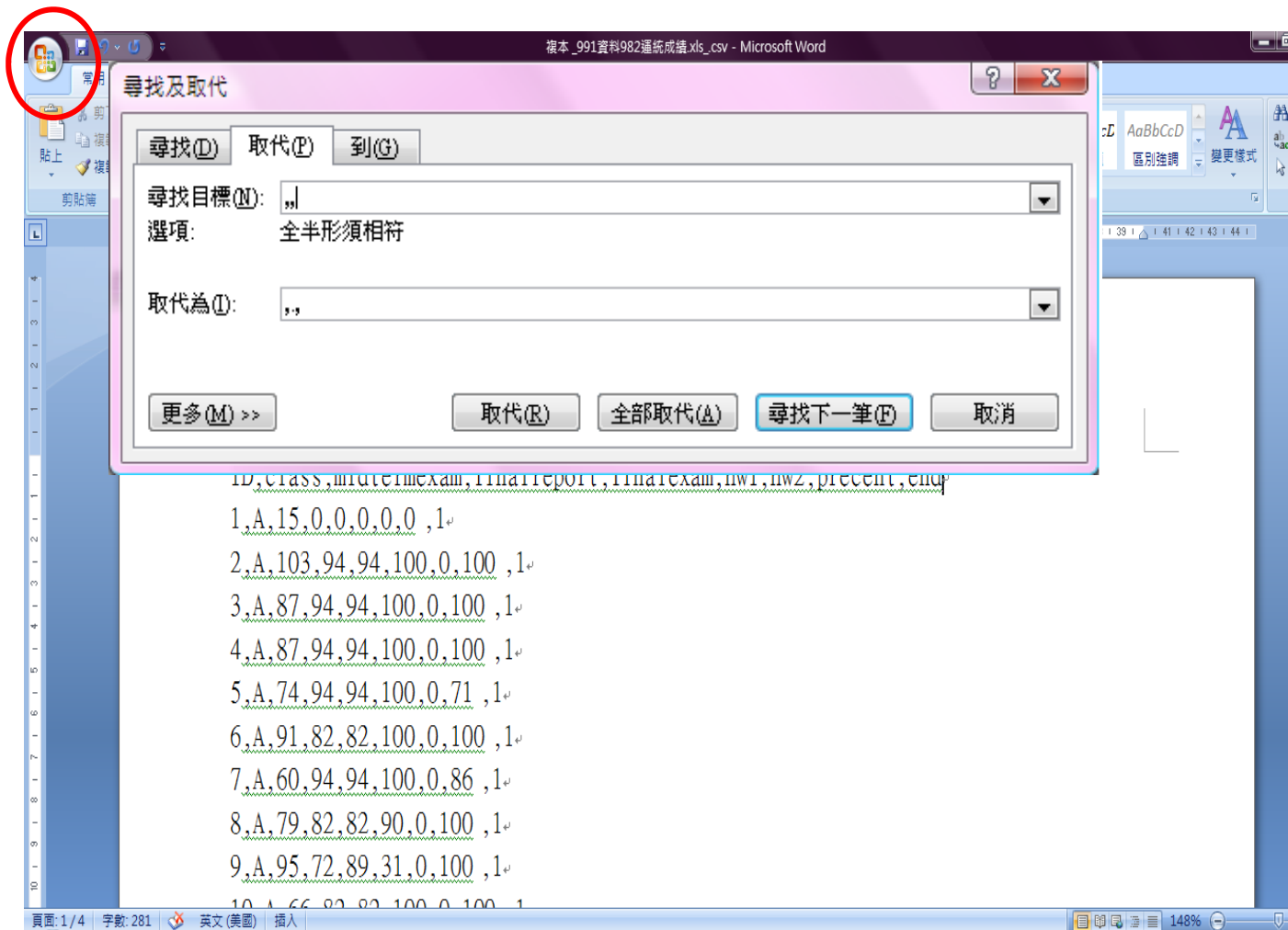
步驟 3：在 word→office 按鈕→開啟舊檔→開啟 __.csv(先前存的 CSV 檔案)



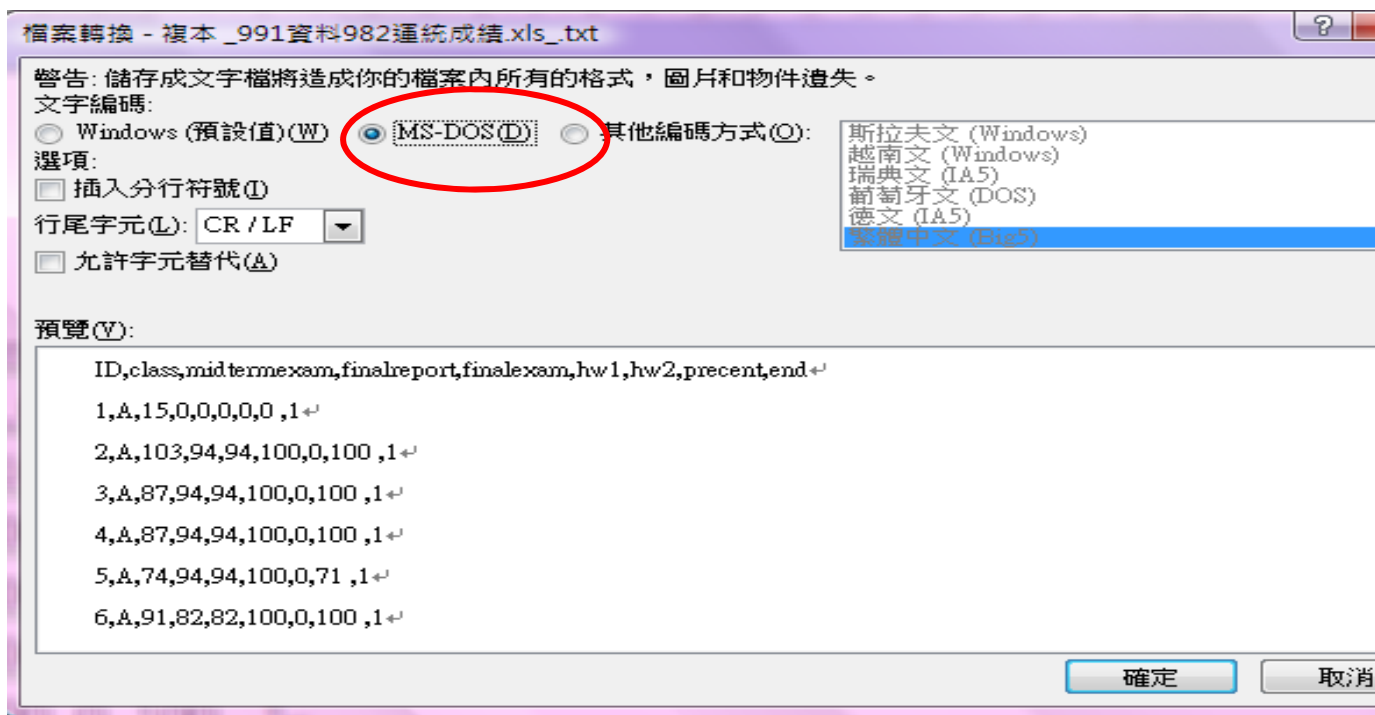
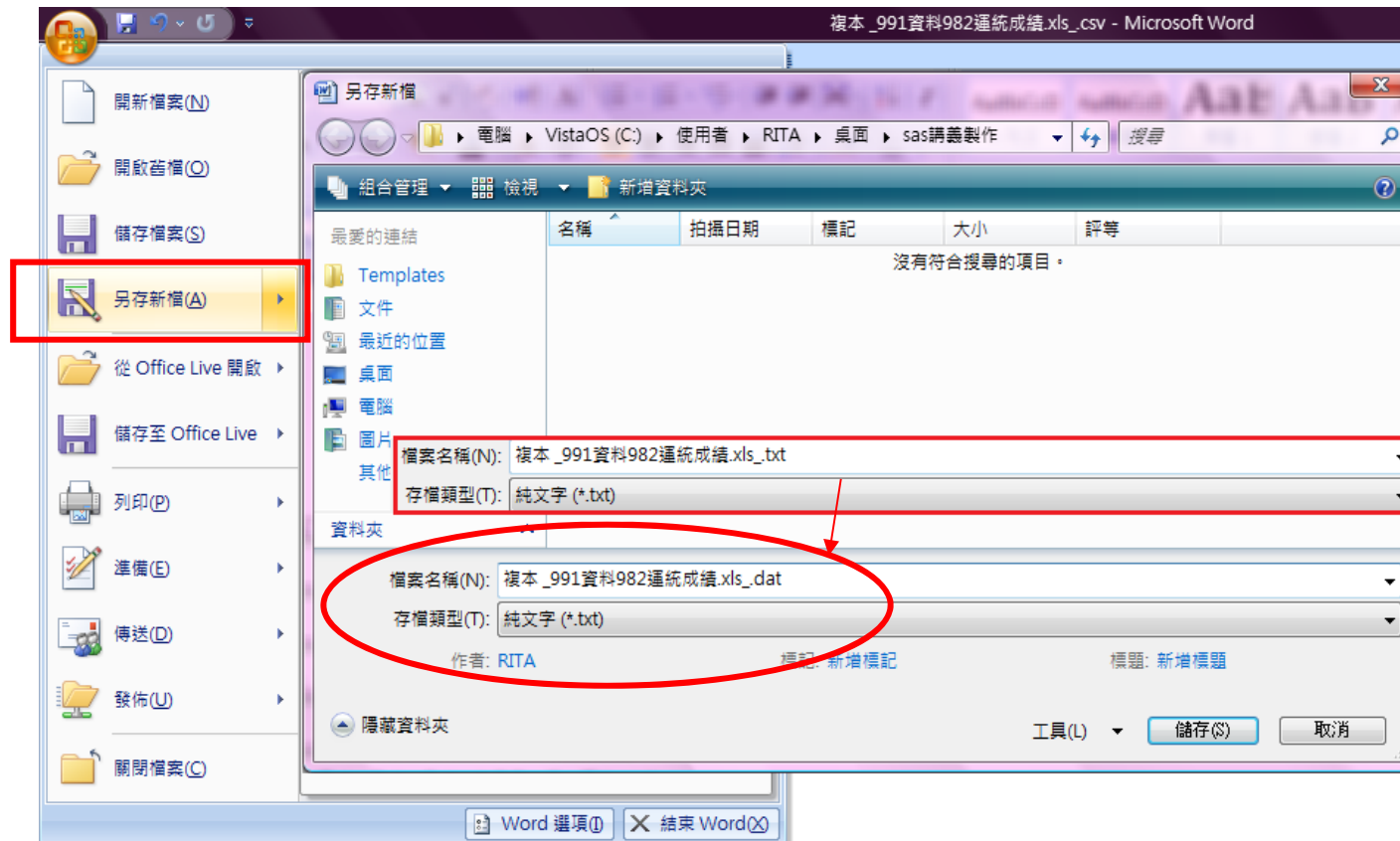
步驟 4：將第一列的變數名稱剪下(以利使用 SAS 時可以直接貼上，無須再重新輸入)



步驟5：補上遺漏值（SAS 內定遺漏值是“.”），資料欄位間以逗號隔開：常用→取代→尋找目標：“,”；取代為：“,.”→至少按2次「全部取代」。（取代動作是避免因讀遺漏值(即空隔)而發生問題，且至少按2次以確保完成取代動作。)



步驟6：office按鈕→另存新檔→檔案名稱：*.dat(即txt改為dat)；存檔類型：純文字檔→選擇儲存MS-DOS



四、Data and Code Book (資料建檔與編碼)

1	ID	class	midtermexam	finalreport	finalexam	hw1	hw2	presence	end
2	1	A	15	0	0	0	0	0	1
3	2	A	103	94	94	100	0	100	1
4	3	A	87	94	94	100	0	100	1
5	4	A	87	94	94	100	0	100	1
6	5	A	74	94	94	100	0	71	1
7	6	A	91	82	82	100	0	100	1
8	7	A	60	94	94	100	0	86	1
9	8	A	79	82	82	90	0	100	1
10	9	A	95	72	89	31	0	100	1
11	10	A	66	82	82	100	0	100	1
12	11	A	88	80	80	30	0	71	1
13	12	A	63	80	80	100	0	100	1
14	13	A	81	80	80	38	0	43	1
15	14	A	91	43	83	100	0	86	1
16	15	A	69	80	80	60	0	100	1
17	16	A	55	80	80	100	0	100	1
18	17	A	54	80	80	100	0	71	1
19	18	A	64	80	80	70	0	86	1
20	19	A	66	80	80	60	0	100	1
21	20	A	68	80	80	50	0	86	1

編碼的時候，先判斷變數為連續型或離散型。

變數	變數說明	資料類型 (離散型的類別和數值型的單位)
ID	學生編號	
class	班別名稱	離散型
		A 班
		B 班
midtermexam	期中考成績	分
finalreport	期末報告	分
finalexam	期末考成績	分
hw1	作業 1	分
		離散型
		+1：佳
hw2	作業 2	0
		-1：缺
presence	實習課出席率	百分比



Q:複選題如何 coding ?