

SAS LANGUAGE

YI-TING HWANG
DEPARTMENT OF STATISTICS
NATIONAL TAIPEI UNIVERSITY

SAS LANGUAGE

- 資料集選項
- 格式 (FORMAT)
- 函數 (FUNCTION)
- 語法 (STATEMENT)
- 整併資料集 -- BY 分群過程
- 產生隨機樣本

資料集選項

函數與運算

簡單數學運算

- 算術運算子
 - + (加), - (減), * (乘), / (除), ** (次方)
- 語法
 - target = expression
- 範例：

r1=x+y;

r2=x-y;

r3=x*y;

r4=x/y;

r5=x**2;

範例 - 中風

- 中風病人極有可能會罹患姿態性低血壓。
- 研究者收集一些中風病人的血壓資料。
- 運用可調整的床收集平躺，傾斜，與直立時的血壓。

檔案：statpackch4d1.xlsx

範例 - 中風資料

變數	說明
ID	編號
Age	年齡
Sex	性別
Examdate	檢查日期
CVAdate	中風發生日期
S0SBP	收縮壓 U1, U3, U5, U7, U9, U11
S0DBP	舒張壓 U1, U3, U5, U7, U9, U11
HUT_SS	暈眩
GRBED	臥床 (yes/no)
PH	姿態性低血壓 (yes/no)
RISK_FACTO	風險因子
DRUGS	用藥

範例

```
data ch4d1a;  
set ch4d1;  
dbp=(s0sbp+u1sbp)/2;  
diffp=u11sbp-s0sbp;  
run;  
proc print  
  data=ch4d1a(obs=10);  
var id s0sbp u1sbp  
  dbp diffp;  
run;
```

Obs	ID	S0SBP	U1SBP	dbp	diffp
1	1	125	125	125.0	-7
2	2	165	87	126.0	-37
3	3	178	157	167.5	-37
4	4	228	144	186.0	-76
5	5	122	122	122.0	-11
6	6	143	145	144.0	-6
7	7	169	148	158.5	32
8	8	133	107	120.0	-18
9	9	150	148	149.0	-18
10	10	191	189	190.0	-9

函數

- 語法
 - `targetvar=funcname(arg1, arg2, ...)`
 - `targetvar=funcname(of arg1 arg2 ...)`
- 範例
 - `s1=sum(x, y, z);`
 - `s2=sum(of x y z);`

數學函數

名稱	說明
GAMMA (arg)	Gamma 函數 (arg-1) !
LGAMMA (arg)	Gamma 函數的自然對數值
LOG (arg)	自然對數值
LOG10 (arg)	以 10 為底的對數值
EXP (power)	自然指數值
SQRT (arg)	開根號
ABS (arg)	絕對值
SIGN (argument)	判定數值的符號
MOD (num, div)	被除數 (Num) 除以除數 (div) 的餘數 (除數不可為 0)

特殊函數

函數	說明
<code>N(of arglist)</code>	計算沒有遺失值的數值變數個數
<code>NMISS(of arglist)</code>	計算有遺失值的數值變數個數
<code>CMISS(of arglist)</code>	計算有遺失值的字符串變數個數
<code>MISSING(numeric- character-expression)</code>	判定變數觀察值是否遺失

統計函數

函數	說明
MEAN (of arglist)	變數平均數
SUM (of arglist)	變數總和
RANGE (of arglist)	變數範圍，須包含至少二個以上的變數
STD (of arglist)	變數標準差，須包含至少二個以上的變數
STDERR (of arglist)	變數標準誤，須包含至少二個以上的變數
VAR (of arglist)	變數變異數，須包含至少二個以上的變數
MIN (of arglist)	變數最小值
MAX (of arglist)	變數最大值

範例

```
data ch4d1a;  
set ch4d1;  
msbp=mean (s0sbp, u1sbp)  
ssbp=std (of s0sbp  
u1sbp u3sbp u5sbp  
u7sbp u9sbp u11sbp);  
miss_measure=nmiss (of  
s0sbp u1sbp u3sbp  
u5sbp u7sbp u9sbp  
u11sbp);  
Run;
```

Obs	ID	S0SBP	U1SBP	msbp	ssbp	miss_measure
1	1	125	125	125.0	5.2190	0
2	2	165	87	126.0	22.8432	0
3	3	178	157	167.5	11.7149	0
4	4	228	144	186.0	28.6689	0
5	5	122	122	122.0	6.7788	0
6	6	143	145	144.0	7.9042	0
7	7	169	148	158.5	19.7821	0
8	8	133	107	120.0	9.4239	0
9	9	150	148	149.0	10.6904	0
10	10	191	189	190.0	10.5017	0

範例 - 高中學生電腦網路 態度的調查

1. 年 級 : (1) 高一 (2) 高二 (3) 高三
2. 學校性質 : (1) 公立 (2) 私立
3. 性 別 : (1) 男 (2) 女

檔案: statpackch4d2.sas7bdat

範例 – 高中學生電腦網路 態度

	非常同意	同意	沒意見	不同意	非常不同意
1. 與別人談論電腦網路時，我常常覺得有壓力 (Q28)	<input type="checkbox"/>				
2. 我認為電腦網路讓人們的生活更便利	<input type="checkbox"/>				
3. 我對電腦網路沒興趣	<input type="checkbox"/>				
4. 我相信透過練習，一定可以學好電腦網路	<input type="checkbox"/>				
5. 一旦接觸網路就會令我感到焦慮	<input type="checkbox"/>				
6. 電腦網路對我的生活而言，是一樣可有可無的東西	<input type="checkbox"/>				
7. 電腦網路可以應用到很多的領域上，如:教育、醫學、交通等等	<input type="checkbox"/>				
8. 我覺得電腦網路很複雜，令我不知所措	<input type="checkbox"/>				
9. 我覺得電腦網路在未來會是必要的工具	<input type="checkbox"/>				
10. 使用電腦網路，可以協助人們獲得大量的資訊 (Q37)	<input type="checkbox"/>				

範例

```
Data ch4d2;  
  set sas.statpackch4d2;  
  newvar11=newq28+newq29;  
  newvar12=sum(of newq28 newq29);  
  nmiss=nmiss(of newq28-newq37);  
  missstat=missing(newq28);  
  newrange=range(of newq28 newq29);  
run;
```

作一樣的動作，
但會因遺失值
而有不同的產
出值

範例

Obs	newq28	newq29	newvar11	newvar12	nmiss	missstat	newrange
1	5	1	6	6	0	0	4
2	5	1	6	6	0	0	4
3	5	2	7	7	0	0	3
4	4	2	6	6	0	0	2
5	3	2	5	5	0	0	1
6	2	3	5	5	1	0	1
7	5	3	8	8	0	0	2
8	3	2	5	5	0	0	1
9	3	5	8	8	1	0	2
10	4	4	8	8	0	0	0
11	4	2	6	6	0	0	2
12	.	1	.	1	1	1	0
13	4	1	5	5	0	0	3
14	2	2	4	4	0	0	0
15	4	2	6	6	1	0	2
16	4	2	6	6	0	0	2
17	4	3	7	7	0	0	1
18	4	3	7	7	0	0	1
19	2	2	4	4	0	0	0
20	.	5	.	5	1	1	0
21	3	4	7	7	0	0	1
22	3	2	5	5	0	0	1
23	2	2	4	4	0	0	0
24	3	1	4	4	1	0	2
25	4	1	5	5	0	0	3

資料集選項

語法 (STATEMENT)

(17)

簡單數學運算

- 比較運算子

- $=$ (EQ; 等於),
- \neq (NE; 不等於),
- $>$ (GT; 大於),
- $<$ (LT; 小於),
- \geq (GE; 大於等於),
- \leq (LE; 小於等於),
- \in (在集合中)

- 邏輯運算子

- $\&$ (AND), \mid (OR), \neg (NOT)

條件式

- 目的：產生虛擬變數、重新分組變數等動作
- 語法

Condition: IF condition [THEN action];
[ELSE action;]

- 任何 SAS 的邏輯判定式
- Action：
 - 指派動作： INPUT; IF; INFILE 或其他指令
 - 只有分號也可
- ELSE：
 - 若使用，需馬上接著 ELSE 或 IF
 - 可以有許多 ELSE

範例

- if $x < 20$;
 - 保留 $x < 20$ 的觀察值
- If state in ('MD', 'TX');
 - 保留 MD 及 TX 兩州的資料
- if $x < 20$ then delete;
 - 刪除 $x < 20$ 的觀察值
- If age ne agecheck then delete;
 - 刪除 age ne agecheck 的觀察值
 - Age 及 agecheck 為變數

範例

- ```
if (status='OK'
 and type=3)
then
count=count+1;
```
- 滿足 status=OK 及 type=3 則變數 count 加 1
- If x=0 then  
 if y ne 0  
 then  
 condition=1;  
 else  
 condition=2;
- 滿足 x=0 之下，再利用變數 y 進行判定變數 condition 的值

# 範例 - 血壓狀態

- 世界衛生組織發布的最新血壓標準
- 收縮壓低於120mmHg以及舒張壓低於80mmHg叫做血壓正常。
- 收縮壓在120~140 mmHg 或舒張壓在80~90 mmHg 屬於高血壓前期。
- 收縮壓  $> 140$ mmHg或舒張壓  $> 90$ mmHg，即稱為高血壓。

```
if s0sbp<120 and s0dbp<80
then status='血壓正常';
else if 120<=s0sbp<140 or
 80<=s0dbp<90
then status='高血壓前期';
```

| Obs | ID | S0SBP | S0DBP | status |
|-----|----|-------|-------|--------|
| 1   | 1  | 125   | 73    | 高血壓前期  |
| 2   | 2  | 165   | 95    | 高血壓    |
| 3   | 3  | 178   | 136   | 高血壓    |

# 範例

```
data ch4d1b;
set ch4d1;
if s0sbp<=120 then
 status_sbp=0;
else if s0sbp<=140
 then status_sbp=1;
else if s0sbp>140 then
 status_sbp=2;
if s0dbp<=80 then
 status_dbp=0;
else if s0dbp<=90 then
 status_dbp=1;
else if s0dbp>90 then
 status_dbp=2;
run;
```

產生異常  
的指標

| Obs | ID | S0SBP | status_sbp | S0DBP | status_dbp |
|-----|----|-------|------------|-------|------------|
| 1   | 1  | 125   | 1          | 73    | 0          |
| 2   | 2  | 165   | 2          | 95    | 2          |
| 3   | 3  | 178   | 2          | 136   | 2          |
| 4   | 4  | 228   | 2          | 91    | 2          |
| 5   | 5  | 122   | 1          | 93    | 2          |
| 6   | 6  | 143   | 2          | 76    | 0          |
| 7   | 7  | 169   | 2          | 85    | 1          |
| 8   | 8  | 133   | 1          | 87    | 1          |
| 9   | 9  | 150   | 2          | 72    | 0          |
| 10  | 10 | 191   | 2          | 91    | 2          |

# DO END 語法

- 利用 SAS 的指令將多個指令視為同一單位
- 語法：

```
DO ;
 statement ;
END ;
```

# 範例 - 學生健康檢查

| 生化指標 | 正常範圍                    | 生化指標 | 正常範圍                           |
|------|-------------------------|------|--------------------------------|
| SBP  | 110-140                 | HCT  | M : 40% - 50%<br>F : 35% - 45% |
| DBP  | 70-90                   | MCV  | 80-100                         |
| WBC  | 4-11                    | MCH  | 27-34                          |
| RBC  | M: 4.5-6.1;<br>F: 4-5.5 |      |                                |
| HB   | M: 13.8-18;<br>F: 12-16 |      |                                |

資料集 : statpackch4d3.xlsx

# 範例

```
if years>5 then
do;
 months=years*12;
 group=1;
end;
else
do;
 yrsleft=5-years;
 group=0;
end;
```

```
if sex='女' then do;
 if 4<=rbc<=5.5 then srbc=0;
 else if (rbc<4 | rbc>5.5)
 then srbc=1;
 if 13.8<=hb<=18 then
 sthb=1;
 else sthb=0;
end;
else if sex='男' then do;
 if 4.5<=rbc<=6.1 then
 srbc=0;
 else if (rbc<4.5 | rbc>6.1)
 then srbc=1;
 if 12<=hb<=16 then sthb=1;
 else sthb=0;
end;
```

# 定義群組

- 目的：對一序列的變數作相同的分析或計算時，可定義群組變數，使得這些變數的計算可以以迴圈方式執行，省去重複撰寫程式可能導致的錯誤
- 語法：
  - **ARRAY** name{dim} [ \$ ] [len]  
[elements] [(st\_values)] ;

# 範例 - ARRAY

- 數值陣列

- ARRAY **test** test1-test20;
- ARRAY **new**{8} new1 new2 new3 new4 new5 new6 new7 new8 (0,0,0,0,0,0,0,0);

- 字串陣列

- ARRAY **wkday**{7} \$ 9 mo tu we th fr sa su ('Monday','Tuesday','Wednesday','Thursday','Friday','Saturday','Sunday');

# 疊代 DO 指令

- 目的：若一序列的變數作相同的分析或計算時，可定義群組變數，再利用疊代 DO 指令來執行
- 語法：
  - DO index-variable=specification-1 <, . . . specification-n>;  
    . . . more SAS statements . . .  
END;
  - Index-variable: sas 會依指標變數來執行
  - Specification: 定義迴路的起始及終止點
    - start <TO stop> <BY increment>

# 疊代方式 -- 固定指標

- do month='JAN', 'FEB', 'MAR';
- do count=2, 3, 5, 7, 11, 13, 17;
- do i=5;
- do i=var1, var2, var3;
- do i='01JAN2001'd,  
'25FEB2001'd, '18APR2001'd;

# 疊代方式- 變動指標

- 設定指標的起始及終止點
  - do i=1 to 10;
  - do i=1 to exit;
  - do i=1 to x-5;
  - do i=1 to k-1,  
k+1 to n;
  - do i=k+1 to n-1;
- 設定指標增加量
  - do i=n to 1 by -1;
  - do i=.1 to .9 by .1,
  - do i=1 to 10 by 1,
  - do i=20 to 100 by 10;
  - do count=2 to 8 by 2;

# 範例

```
array class class1-class9;
count=0;
do cv=0.1 to 0.9 by 0.1;
 count=count+1;
 if prob_rs_0<=cv then
 class{count}=2;
 else class{count}=1;
end;
```

# 範例

```
array sbp s0sbp u1sbp u3sbp u5sbp
u7sbp u9sbp u11sbp;
array dsbp ds0sbp du1sbp du3sbp
du5sbp du7sbp du9sbp du11sbp;
do i=1 to 7;
 if sbp{i}<=120 then dsbp{i}=0;
 else if sbp{i}<=140 then dsbp{i}=1;
 else if sbp{i}>140 then dsbp{i}=2;
end;
```

# 範例

| Obs | ID | S0SBP | U1SBP | U3SBP | ds0sbp | du1sbp | du3sbp |
|-----|----|-------|-------|-------|--------|--------|--------|
| 1   | 1  | 125   | 125   | 113   | 1      | 1      | 0      |
| 2   | 2  | 165   | 87    | 127   | 2      | 0      | 1      |
| 3   | 3  | 178   | 157   | 156   | 2      | 2      | 2      |
| 4   | 4  | 228   | 144   | 156   | 2      | 2      | 2      |
| 5   | 5  | 122   | 122   | 106   | 1      | 1      | 0      |

# 範例

```
array sbp s0sbp u1sbp u3sbp u5sbp
u7sbp u9sbp u11sbp;
array dsbp ds0sbp du1sbp du3sbp
du5sbp du7sbp du9sbp du11sbp;
do i=1 to 7;
 if sbp{i}<=120 then dsbp{i}=0;
 else if sbp{i}<=140 then dsbp{i}=1;
 else if sbp{i}>140 then dsbp{i}=2;
end;
```

# 範例

```
array dbp s0dbp u1dbp u3dbp u5dbp u7dbp
u9dbp u11dbp;
length s0s u1s u3s u5s u7s u9s u11s $12;
array status $ s0s u1s u3s u5s u7s u9s
u11s;
do i=1 to dim(status);
 if sbp{i}<120 and dbp{i}<80 then
 status{i}='血壓正常';
 else if 120<=sbp{i}<140 or 80<=s0dbp<90
 then status{i}='高血壓前期';
 else status{i}='高血壓';
end;
```

# 範例

| Obs | ID | S0SBP | U1SBP | U3SBP | ds0sbp | du1sbp | du3sbp |
|-----|----|-------|-------|-------|--------|--------|--------|
| 1   | 1  | 125   | 125   | 113   | 1      | 1      | 0      |
| 2   | 2  | 165   | 87    | 127   | 2      | 0      | 1      |
| 3   | 3  | 178   | 157   | 156   | 2      | 2      | 2      |
| 4   | 4  | 228   | 144   | 156   | 2      | 2      | 2      |
| 5   | 5  | 122   | 122   | 106   | 1      | 1      | 0      |

# 範例 - 計數

```
count=0;
do i=1 to dim(status);
 if status{i}='高血壓'
 then count=count+1;
end;
```

| Obs | ID | s0s   | u1s   | u3s   | u5s   | u7s   | u9s   | u11s  | count |
|-----|----|-------|-------|-------|-------|-------|-------|-------|-------|
| 1   | 1  | 高血壓前期 | 高血壓前期 | 血壓正常  | 高血壓   | 高血壓   | 血壓正常  | 高血壓   | 3     |
| 2   | 2  | 高血壓   | 血壓正常  | 高血壓前期 | 高血壓前期 | 高血壓前期 | 高血壓前期 | 高血壓前期 | 1     |
| 3   | 3  | 高血壓   | 7     |
| 4   | 4  | 高血壓   | 7     |
| 5   | 5  | 高血壓前期 | 高血壓前期 | 高血壓   | 高血壓   | 高血壓   | 高血壓   | 高血壓   | 5     |

資料集選項

特殊函數

( 39 )

# 函數處理字符串變數

| 名稱                                             | 說明           |
|------------------------------------------------|--------------|
| INDEX (source, excerpt)                        | 尋找字符串中某些特殊字元 |
| LEFT (arg)                                     | 字符串向左對齊      |
| LENGTH (arg)                                   | 計算字符串長度      |
| RIGHT (arg)                                    | 字符串向右對齊      |
| SCAN (string, count<, charlist <, modifiers>>) | 以特殊字元將字符串分割  |

# 處理字串變數

| 名稱                                                                                                  | 說明             |
|-----------------------------------------------------------------------------------------------------|----------------|
| COMPRESS (source<, characters-to-remove)                                                            | 刪除或保留字串中某些特殊字元 |
| SUBSTR (argument, position<, n>)                                                                    | 擷取字串部份字元       |
| TRIM (argument)                                                                                     | 除去多餘空格         |
| <b>TRANWRD( <u>source-expression</u>, <u>target-expression</u>, <u>replacement-expression</u> )</b> | 置換或刪除部份字串      |
| <b>TRANSLATE( <u>expression</u>, <u>to-characters</u>, <u>from-characters</u> )</b>                 | 置換特殊字元         |

Carriage return 重置 '0D'x

Line feed 换行 '0A'x

Tab 09'x

# Compress 修飾詞 (Modifiers)

| 修飾詞 | 說明     |
|-----|--------|
| k   | 保留     |
| i   | 忽略大小寫  |
| d   | 刪除數字   |
| a   | 刪除字母   |
| p   | 刪除標點符號 |

- 保留數字

```
Keep_digits = compress(my_string, , 'dk');
```

- 保留 a, A, b, B, 2, 4, 6

```
comb_ab246 =compress(my_string,'ab246', 'ik');
```

- 刪除字母, 2, 4, 6, -

```
comb_mod = compress(my_string,'246-', 'a');
```

# 範例 -- 退化性膝關節炎

- 一般嚴重有退化性膝關節炎的老年人都會做膝蓋替換手術 (total knee replacement surgery; TKRS) 。
- 感染是常見的副作用，而糖尿病則以之為重要的危險因子。
- 資料集： statpackch4d4.sas7bdat

# 範例

| 變數      | 說明                                      |
|---------|-----------------------------------------|
| No      | 案例編號                                    |
| Age     | 年齡                                      |
| Sex     | F:女; M: 男                               |
| DMtype  | NI: Type II (NIDDM) ; ID: Type I (IDDM) |
| preopBS | 手術前的血壓                                  |
| postBS  | 手術後的血壓                                  |
| Side    | 手術邊                                     |
| PREKS   | 手術前的膝功能                                 |
| POSTKS  | 手術後的膝功能                                 |
| INFECT  | 手術感染 N: 沒有; P: 有; MO: 月份                |

# 範例

```
data ch4d4;
 set sas.statpackch4d4;
 dm=substr(dmtype,1,2);
 type=compress(dmtype);
 flag=index(type,'YR');
 yeardm=input(substr(type,3,flag-
 3),8.);
 flag1=index(preopbs,'/');
 lenpreopbs=length(preopbs);
run;
```

# 範例

substr

index

length

| Obs | dm       | type | flag   | yeardm | PREOPBS | flag1   | lenpreopbs | preopbs1 | preopbs2 |     |
|-----|----------|------|--------|--------|---------|---------|------------|----------|----------|-----|
| 1   | NI 10 YR | NI   | NI10YR | 5      | 10      | 120/160 | 4          | 7        | 120      | 160 |
| 2   | NI 11 YR | NI   | NI11YR | 5      | 11      | 100/150 | 4          | 7        | 100      | 150 |
| 3   | NI 4 YR  | NI   | NI4YR  | 4      | 4       | 150/200 | 4          | 7        | 150      | 200 |
| 4   | NI 8YR   | NI   | NI8YR  | 4      | 8       | 150/200 | 4          | 7        | 150      | 200 |
| 5   | NI 3YR   | NI   | NI3YR  | 4      | 3       | 85/110  | 3          | 6        | 85       | 110 |
| 6   | NI 1YR   | NI   | NI1YR  | 4      | 1       | 120/150 | 4          | 7        | 120      | 150 |
| 7   | NI 1YR   | NI   | NI1YR  | 4      | 1       | 120/150 | 4          | 7        | 120      | 150 |
| 8   | ID 35YR  | ID   | ID35YR | 5      | 35      | 200/250 | 4          | 7        | 200      | 250 |
| 9   | NI 5YR   | NI   | NI5YR  | 4      | 5       | 130/180 | 4          | 7        | 130      | 180 |
| 10  | NI 5YR   | NI   | NI5YR  | 4      | 5       | 130/180 | 4          | 7        | 130      | 180 |
| 11  | NI 4YR   | NI   | NI4YR  | 4      | 4       | 150/200 | 4          | 7        | 150      | 200 |
| 12  | NI 8YR   | NI   | NI8YR  | 4      | 8       | 200/270 | 4          | 7        | 200      | 270 |
| 13  | NI 25YR  | NI   | NI25YR | 5      | 25      | 160/200 | 4          | 7        | 160      | 200 |
| 14  | NI 25YR  | NI   | NI25YR | 5      | 25      | 160/200 | 4          | 7        | 160      | 200 |
| 15  | NI 20YR  | NI   | NI20YR | 5      | 20      | 180/230 | 4          | 7        | 180      | 230 |

Compress

# 函數 -- 合併字符串變數

| 名稱                                        | 說明                 |
|-------------------------------------------|--------------------|
| CAT(item1, item2, ...)                    | 合併時會字串起始與最後的空格     |
| CATS(item1, item2, ...)                   | 合併時會刪除中間字串起始與最後的空格 |
| CATT(item1, item2, ...)                   | 合併時會刪除中間字串最後的空格    |
| CATX(delimiter,<br>item-1 [, ... item-n]) | 合併字串，並在字串中間加入分隔符號  |

# 範例

```
data ch4d5;
informat x y z a $25. result $120.;
x=' The 2012 Olym';
y='pic Arts Festi';
z=' val included works by D ';
a='ale Chihuly. ';
result=cat(x,y,z,a);
result1=cats(x,y,z,a);
result2=catt(x,y,z,a);
put result;
put result1;
put result2;
```

The 2012 Olym                   pic Arts Festi                           val included works by Dale Chihuly.  
The 2012 Olympic Arts Festival included works by Dale Chihuly.  
The 2012 Olympic Arts Festi val included works by Dale Chihuly.

# 範例

```
data ch4d6;
informat x1 x2 x3 $4. result $15.;
x1 = 'var1';
x2 = 'var2';
x3 = 'var3';
result = catx(',', ' ', of x1-x3);
run;
```

| Obs | x1   | x2   | x3   | result         |
|-----|------|------|------|----------------|
| 1   | var1 | var2 | var3 | var1,var2,var3 |

# 函數 - 計數函數

| 名稱                                           | 修飾詞<br>(modifier)       | 說明                                         |
|----------------------------------------------|-------------------------|--------------------------------------------|
| COUNT(string,<br>substring [,<br>modifiers]) | i: 忽略大小<br>寫<br>t: 忽略空格 | 計數<br>substring 出<br>現在 string 的<br>次數     |
| COUNTC(string,<br>charlist [,<br>modifiers]) |                         | 計數 charlist<br>出現或沒有出現<br>在 string 的次<br>數 |
| COUNTW(string[,ch<br>ars[, modifiers]])      |                         | 計數 chars 出<br>現次數                          |

# Countc 修飾詞

| 修飾詞   | 說明           | 修飾詞 | 說明                   |
|-------|--------------|-----|----------------------|
| blank | 預設忽略空白       | i/I | 統計時忽略大小寫             |
| a/A   | 統計字母 (大小寫都算) | l/L | 統計小寫                 |
| b/B   | 統計時從右到左      | n/N | 統計字母、數字、底線           |
| c/C   | 統計特殊符號       | s/S | 統計特殊分隔符號，例如：空格、tab 等 |
| d/D   | 統計數字         | t/T | 統計時刪除字串後面的空白         |
| f/F   | 統計字母與底線      | u/U | 統計大寫字母               |
|       |              | v/V | 統計沒有出現的字元數           |
|       |              | x/X | 統計 16 進位數字 (包含小數點)   |

# Countw 修飾詞

| 修飾詞   | 說明      | 修飾詞 | 說明                       |
|-------|---------|-----|--------------------------|
| blank | 預設忽略空白  | i/I | 統計字母時，忽略大小寫              |
| a/A   | 統計字母    | l/L | 統計小寫字母                   |
| b/B   | 統計時從右到左 | n/N | 統計字母、數字、底線               |
| c/C   | 統計特殊符號  | s/S | 統計特殊分隔符號，例如：<br>空格、tab 等 |
| d/D   | 數字      | t/T | 統計時刪除字串後面的空<br>白         |
| f/F   | 統計字母與底線 | u/U | 統計大寫字母                   |
|       |         | x/X | 統計 16 進位數字 (包含小<br>數點)   |

預設統計的特殊字元：blank ! \$ % & ( ) \* + , - . / ; < ^ |

# 範例

```
data ch4d7;
informat xyz $50.;
xyz='This is a thistle? Yes, this is a
thistle.';
howmanythis1=count(xyz, 'this');
howmanythis2=count(xyz, 'this', 'i');
run;
```

| Obs | xyz                                        | howmanythis1 | howmanythis2 |
|-----|--------------------------------------------|--------------|--------------|
| 1   | This is a thistle? Yes, this is a thistle. | 3            | 4            |

# 範例

- 統計字串出現的所有字母與數字

```
count_a_d_chars = countc(string,, 'ad');
```

- 統計字串的空格

```
count_spaces = countc(string,, 'st');
```

- 統計字母 a 但忽略大小寫

```
count_a_ignore_case =
countc(string, 'a', 'i');
```

- 統計字串所有非 'ab' 的其他字元

```
count_not_ab = countc(string, 'ab', 'vt');
```

- 統計字串所有非字母與非數字的字元

```
count_not_a_not_n = countc(string,, 'adtv');
```

# 範例

```
data example;
input string $60.;
datalines;
It's a pleasant day today
I_am yet to receive payment
352+20+2=374
Send to my address xyz_a@gmail
;
proc print;
run;
```

# 範例

```
data newdata;
set example;
count_words=cou
ntw(string);
run;
```

| string                         | count_words |
|--------------------------------|-------------|
| It's a pleasant day today      | 5           |
| I_am yet to receive payment    | 5           |
| 352+20+2=374                   | 3           |
| Send to my address xyz_a@gmail | 5           |

By default SAS consider the followings as default delimiters:  
blank ! \$ % & ( ) \* + , - . / ; < ^ |

# 範例

```
data newdata;
set example;
count_words=countw(string, ' ') ;
count_words_=countw(string, '_') ;
run;
```

| string                         | count_words | count_words_ |
|--------------------------------|-------------|--------------|
| It's a pleasant day today      | 5           | 1            |
| I_am yet to receive payment    | 5           | 2            |
| 352+20+2=374                   | 1           | 1            |
| Send to my address xyz_a@gmail | 5           | 2            |

# 範例

```
data newdata;
set example;
count_words_digits=countw(string
, , 'd');
run;
```

| string                         | count_words_digits |
|--------------------------------|--------------------|
| It's a pleasant day today      | 1                  |
| I_am yet to receive payment    | 1                  |
| 352+20+2=374                   | 4                  |
| Send to my address xyz_a@gmail | 1                  |

# 範例 - 學生健康行為

| 問項   | 說明                                                                                                |
|------|---------------------------------------------------------------------------------------------------|
| 健康編號 |                                                                                                   |
| 學制   | 分進修學士班 (0) 及碩士班 (1)                                                                               |
| 系所   | 公共事務學院：不動產與城鄉環境學系，公共行政暨政策學系，財政學系<br>商學院：會計學系，企業管理學系，合作經濟學系，統計學系<br>法律學院：法學系<br>社會科學學院：社會工作學系，經濟學系 |
| 家族病史 | 心臟病, 糖尿病, 其他                                                                                      |
| 過去病史 | 過敏                                                                                                |
| 吸菸習慣 | 吸菸數量                                                                                              |

資料集：statpackch4d8.xls

# 處理數值小數點方式

| 名稱          | 說明     |
|-------------|--------|
| CEIL (arg)  | 無條件進位  |
| FLOOR (arg) | 無條件捨去  |
| INT (arg)   | 輸出整數部分 |
| ROUND (arg) | 四捨五入   |

# 範例 -- 雇員資料

| 變數   | 欄位    | 格式     |
|------|-------|--------|
| 員工編號 | 7-12  | 字串     |
| 起聘日期 | 13-21 | Date9. |
| 起薪   | 22-29 | 數值     |
| 部門   | 30-35 | 字串     |
| 工作代號 | 37-39 | 字串     |
| 性別   | 40    | 字串     |

資料集： statpackch4d9.txt

# 範例

```
data ch4d9;
infile "statpackch4d9.txt";
input id $ 7-12
@13 init_hire date9.
salary 22-29 dept $ 30-35
jobcode $ 37-39 sex $ 40;
salary1=ceil(salary);
salary2=floor(salary);
salary3=int(salary);
salary4=round(salary);
run;
```

| Obs | id     | salary   | salary1 | salary2 | salary3 | salary4 |
|-----|--------|----------|---------|---------|---------|---------|
| 1   | 119012 | 42340.58 | 42341   | 42340   | 42340   | 42341   |
| 2   | 120591 | 31000.55 | 31001   | 31000   | 31000   | 31001   |
| 3   | 127845 | 75320.34 | 75321   | 75320   | 75320   | 75320   |
| 4   | 129540 | 56123.34 | 56124   | 56123   | 56123   | 56123   |
| 5   | 135673 | 46322.58 | 46323   | 46322   | 46322   | 46323   |
| 6   | 212916 | 52345.58 | 52346   | 52345   | 52345   | 52346   |
| 7   | 216382 | 34004.65 | 34005   | 34004   | 34004   | 34005   |
| 8   | 234967 | 17000.00 | 17000   | 17000   | 17000   | 17000   |
| 9   | 237642 | 43200.34 | 43201   | 43200   | 43200   | 43200   |
| 10  | 239185 | 57920.66 | 57921   | 57920   | 57920   | 57921   |

# 變更變數屬性函數

| 名稱                                  | 說明              |
|-------------------------------------|-----------------|
| PUT (source, format.)               | 將變數格式轉換成字串格式    |
| INPUT (source, <?   ???> informat.) | 將變數格式轉換成預期的數值格式 |

# 回傳變數數值函數

| 名稱                   | 說明                                  |
|----------------------|-------------------------------------|
| VVALUE (var)         | 回傳變數數值，或格式化的數值。但設定的輸入僅可以是變數或是陣列表示式  |
| VVALUEX (expression) | 回傳變數標籤。設定的輸入僅可以是變數、陣列表示式、文字         |
| VLABEL (var)         | 回傳變數標籤。但設定的輸入僅可以是變數或是陣列表示式          |
| VLABELX (expression) | 回傳變數標籤，或格式化的數值。設定的輸入僅可以是變數、陣列表示式、文字 |

# 範例

```
data ch4d9;
infile "statpackch4d9.txt";
input id $ 7-12 init_hire $ 13-21
salary 22-29 dept $ 30-35 njobcode
37-39 sex $ 40;
d_hire=input(init_hire,date9.);
jobcode=put(njobcode,3.0);
Run;
```

# 範例

| Obs | id     | init_hire | d_hire     | njobcode | jobcode |
|-----|--------|-----------|------------|----------|---------|
| 1   | 119012 | 01JUL1973 | 07/01/1973 | 602      | 602     |
| 2   | 120591 | 05DEC1985 | 12/05/1985 | 602      | 602     |
| 3   | 127845 | 16JAN1972 | 01/16/1972 | 204      | 204     |
| 4   | 129540 | 01AUG1987 | 08/01/1987 | 204      | 204     |
| 5   | 135673 | 15JUL1989 | 07/15/1989 | 602      | 602     |
| 6   | 212916 | 15FEB1958 | 02/15/1958 | 602      | 602     |
| 7   | 216382 | 15JUN1990 | 06/15/1990 | 602      | 602     |
| 8   | 234967 | 19DEC1993 | 12/19/1993 | 602      | 602     |
| 9   | 237642 | 01NOV1981 | 11/01/1981 | 602      | 602     |
| 10  | 239185 | 07MAY1986 | 05/07/1986 | 602      | 602     |

# 其他函數

| 名稱                | 說明                  |
|-------------------|---------------------|
| DIF $< n >$ (arg) | 輸出現在的觀察值與前面第n筆觀察值的差 |
| LAG $< n >$ (arg) | 輸出前面第n筆觀察值          |

# 範例 -

| 變數   | 格式      |
|------|---------|
| X    | 數值      |
| Sale | Comma9. |

檔名: statpackch4d9.dat

| Obs | x | sale    | z | d | fmtsale |
|-----|---|---------|---|---|---------|
| 1   | 1 | 2115153 | . | . | 2115153 |
| 2   | 2 | 3325253 | . | . | 3325253 |
| 3   | 6 | 4435353 | 1 | . | 4435353 |
| 4   | 4 | 1545453 | 2 | 3 | 1545453 |
| 5   | 7 | 1555553 | 6 | 5 | 1555553 |

```
Data ch4d10;
Infile
'statpackch4d10.dat';
input x sale comma9.;
Z=lag2(x);
D=dif3(x);
fmtsale=input(sale,f20.
.);
run;
```

# 日期函數

| 名稱                     | 說明              |
|------------------------|-----------------|
| DATE ()                | 輸出現在日期          |
| DAY (date)             | 輸出日             |
| MDY (month, day, year) | 將日、月、年轉成 SAS 日期 |
| MONTH (date)           | 輸出月份            |
| TODAY ()               | 輸出現在日期          |
| YEAR (date)            | 輸出年             |

# 範例

```
data ch4d11;
 input id birthdt date7.;
 today=today();
 ageday=today-birthdt;
 ageyears=ageday/365.25;
 cards;
 2145 01dec69;
 run;
```

# 範例

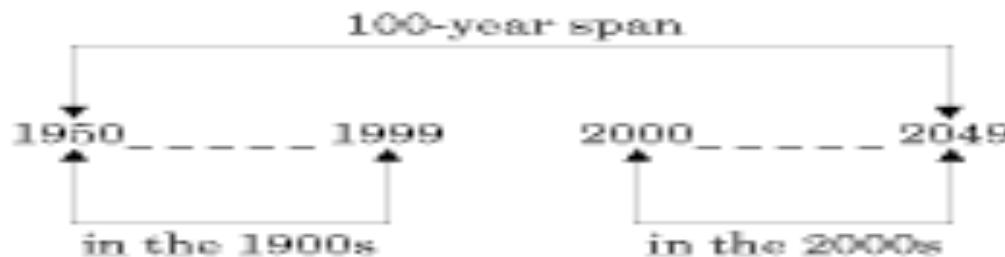
```
data ch4d9;
infile "statpackch4d9.txt";
input id $ 7-12 init_hire $ 13-21
salary 22-29 dept $ 30-35
njobcode 37-39 sex $ 40;
d_hire=input(init_hire,date9.);
d_year=year(d_hire);
d_month=month(d_hire);
duration=int((mdy(12,31,2000) -
d_hire)/365.25);
run;
```

# 範例

| Obs | id     | d_hire     | d_year | d_month | duration |
|-----|--------|------------|--------|---------|----------|
| 1   | 119012 | 07/01/1973 | 1973   | 7       | 27       |
| 2   | 120591 | 12/05/1985 | 1985   | 12      | 15       |
| 3   | 127845 | 01/16/1972 | 1972   | 1       | 28       |
| 4   | 129540 | 08/01/1987 | 1987   | 8       | 13       |
| 5   | 135673 | 07/15/1989 | 1989   | 7       | 11       |
| 6   | 212916 | 02/15/1958 | 1958   | 2       | 42       |
| 7   | 216382 | 06/15/1990 | 1990   | 6       | 10       |
| 8   | 234967 | 12/19/1993 | 1993   | 12      | 7        |
| 9   | 237642 | 11/01/1981 | 1981   | 11      | 19       |
| 10  | 239185 | 05/07/1986 | 1986   | 5       | 14       |

# SAS 日期計算方式

- SAS 以 1920 年當成時間原點
- 若 SAS 令 1950 年 1 月 1 號為 0 天，而 1950 年 12 月 31 號為 365 天



- 若要知道日期變數的真實日期，則需給 SAS 變數適當的格式
- 當日期的西元年只給二位時，如 11，則 SAS 會自動判斷為 1911，但若是指西元 2011 年時，則需以可用下面方式調整

```
options yearcutoff= 1920;
```

排序資料

資料集選項

# 名詞定義

- BY-group processing
  - 處理一個或多個以上的資料含有一個以上的相同變數的過程
  - 常用於 SET 或 MERGE多個以上的資料
- BY variable:
  - 資料裡共同有的變數名稱
- BY value:
  - 經由排序之後的原始資料值或格式化後的值
- BY group:
  - 包含有相同變數的所有觀察值

# 自動產生指標變數

- FIRST.variable:
  - 為 BY group的第一個觀察值的指標變數
- LAST.variable:
  - 為 BY group的最後一個觀察值的指標變數
- Variable 為出現在 BY 指令後面的變數
- 當資料處理完後，此二變數會自動刪除

# BY-GROUP 過程語法

- BY variable(s);
- BY <DESCENDING> variable(s)  
<NOTSORTED> <GROUPFORMAT>;
  - variable: 資料排序依據的變數名稱
  - GROUPFORMAT: 依格式化後的數值排序
  - DESCENDING: 依反向方式排序
  - NOTSORTED: 資料沒有依字母或是數值大小排序

# 範例 - 依單一變數排序

BY variable

| ZipCode | State | City     | Street     |          |
|---------|-------|----------|------------|----------|
| 33133   | FL    | Miami    | Rice st    | BY group |
| 33133   | FL    | Miami    | Thomas Ave |          |
| 33133   | FL    | Miami    | Surrey Dr  |          |
| 33133   | FL    | Miami    | Trade Ave  |          |
| 33146   | FL    | Miami    | Nervia St  | BY group |
| 33146   | FL    | Miami    | Corsica St |          |
| 33801   | FL    | Lakeland | French Ave | BY group |
| 33809   | FL    | Lakeland | Egret Dr   | BY Group |
| 85730   | AZ    | Tucson   | Domenic Ln | BY group |
| 85730   | AZ    | Tucson   | Gleeson Pl |          |

# 範例 - 依二個變數排序

| BY variables |          |            |         |
|--------------|----------|------------|---------|
| State        | City     | street     | zipCode |
| AZ           | Tucson   | Domenic Ln | 85730   |
| AZ           | Tucson   | Gleeson Pl | 85730   |
| FL           | Lakeland | French Ave | 33801   |
| FL           | Lakeland | Egret Dr   | 33809   |
| FL           | Miami    | Nervia St  | 33146   |
| FL           | Miami    | Rice st    | 33133   |
| FL           | Miami    | Corsica st | 33146   |
| FL           | Miami    | Thomas Ave | 33133   |
| FL           | Miami    | Surrey Dr  | 33133   |
| FL           | Miami    | Trade Ave  | 33133   |

# 範例 - 依 STATE CITY ZIPCODE 排序

| Observations in Four BY Groups |        |         |            | Corresponding FIRST. and LAST. Values |             |             |            |                |               |
|--------------------------------|--------|---------|------------|---------------------------------------|-------------|-------------|------------|----------------|---------------|
| State                          | City   | ZipCode | Street     | FIRST. State                          | LAST. State | FIRST. City | LAST. City | FIRST. ZipCode | LAST. ZipCode |
| AZ                             | Tucson | 85730   | Glen PI    | 1                                     | 1           | 1           | 1          | 1              | 1             |
| FL                             | Miami  | 33133   | Rice St    | 1                                     | 0           | 1           | 0          | 1              | 0             |
| FL                             | Miami  | 33133   | Tom Ave    | 0                                     | 0           | 0           | 0          | 0              | 0             |
| FL                             | Miami  | 33133   | Surrey Dr  | 0                                     | 0           | 0           | 0          | 0              | 1             |
| FL                             | Miami  | 33146   | Nervia St  | 0                                     | 0           | 0           | 0          | 1              | 0             |
| FL                             | Miami  | 33146   | Corsica St | 0                                     | 1           | 0           | 1          | 0              | 1             |
| OH                             | Miami  | 45056   | Myrtle St  | 1                                     | 1           | 1           | 1          | 1              | 1             |

# 範例 - 依 CITY STATE ZIPCODE 排序

| Observations in Four BY Groups |       |         |            | Corresponding FIRST. and LAST. Values |            |              |             |                |               |
|--------------------------------|-------|---------|------------|---------------------------------------|------------|--------------|-------------|----------------|---------------|
| City                           | State | ZipCode | Street     | FIRST. City                           | LAST. City | FIRST. State | LAST. State | FIRST. ZipCode | LAST. ZipCode |
| Miami                          | FL    | 33133   | Rice St    | 1                                     | 0          | 1            | 0           | 1              | 0             |
| Miami                          | FL    | 33133   | Tom Ave    | 0                                     | 0          | 0            | 0           | 0              | 0             |
| Miami                          | FL    | 33133   | Surrey Dr  | 0                                     | 0          | 0            | 0           | 0              | 1             |
| Miami                          | FL    | 33146   | Nervia St  | 0                                     | 0          | 0            | 0           | 1              | 0             |
| Miami                          | FL    | 33146   | Corsica St | 0                                     | 0          | 0            | 1           | 0              | 1             |
| Miami                          | OH    | 45056   | Myrtle St  | 0                                     | 1          | 1            | 1           | 1              | 1             |
| Tucson                         | AZ    | 85730   | Glen Pl    | 1                                     | 1          | 1            | 1           | 1              | 1             |

# 讀取 SAS 資料集

| 指令    | 功能                     |
|-------|------------------------|
| SET   | 輸入一個或多個 SAS 資料集        |
| MERGE | 整合多個 SAS 資料集裡的變數到一個觀察值 |

# 整合多個資料 (增加觀察值)

DATA1

| Year |
|------|
| 1991 |
| 1992 |
| 1993 |
| 1994 |
| 1995 |

COMBINED

DATA2

+

| Year |
|------|
| 1992 |
| 1993 |
| 1994 |
| 1995 |
| 1996 |

=

| Year |
|------|
| 1991 |
| 1992 |
| 1993 |
| 1994 |
| 1995 |
| 1996 |
| 1991 |
| 1992 |
| 1993 |
| 1994 |
| 1995 |
| 1996 |
| 1991 |
| 1992 |
| 1993 |
| 1994 |
| 1995 |
| 1996 |

```
data combined;
 set data1 data2;
 by Year;
run;
```

DATA1

| Year |
|------|
| 1991 |
| 1992 |
| 1993 |
| 1994 |
| 1995 |

DATA2

| Year |
|------|
| 1991 |
| 1992 |
| 1993 |
| 1994 |
| 1995 |

COMBINED

+

| Year |
|------|
| 1991 |
| 1992 |
| 1993 |
| 1994 |
| 1995 |

=

| Year |
|------|
| 1991 |
| 1992 |
| 1993 |
| 1994 |
| 1995 |
| 1991 |
| 1992 |
| 1993 |
| 1994 |
| 1995 |

```
data combined;
 set data1 data2;
run;
```

# 範例 - 增加觀察值

資料集: `animal`

| common | animal | number |
|--------|--------|--------|
| A      | Ant    | 5      |
| B      | Bird   |        |
| C      | Cat    | 17     |
| D      | Dog    | 9      |
| E      | Eagle  |        |
| F      | Frog   | 76     |

資料集二: `plant`

| common | plant    | number |
|--------|----------|--------|
| G      | Grape    | 69     |
| H      | Hazelnut | 55     |
| I      | Indigo   |        |
| J      | Jicama   | 14     |
| K      | Kale     | 5      |
| L      | Lentil   | 77     |

# 範例- 增加觀察值

## 合併語法

```
data concatenation;
 set animal plant;
run;
proc print
 data=concatenation;
 var Common
 Animal Plant
 Number;
 title 'Data Set
 CONCATENATION';
run;
```

## 結果

| Obs | common | animal | number | plant    |
|-----|--------|--------|--------|----------|
| 1   | A      | Ant    | 5      |          |
| 2   | B      | Bird   | .      |          |
| 3   | C      | Cat    | 17     |          |
| 4   | D      | Dog    | 9      |          |
| 5   | E      | Eagle  | .      |          |
| 6   | F      | Frog   | 76     |          |
| 7   | G      |        | 69     | Grape    |
| 8   | H      |        | 55     | Hazelnut |
| 9   | I      |        | .      | Indigo   |
| 10  | J      |        | 14     | Jicama   |
| 11  | K      |        | 5      | Kale     |
| 12  | L      |        | 77     | Lentil   |

# 整合多個資料 (增加變數)

One-to-one

| <b>DATA1</b> |      | <b>DATA2</b> |      | <b>COMBINED</b> |      |
|--------------|------|--------------|------|-----------------|------|
| VarX         | VarY | VarX         | VarY | VarX            | VarY |
| X1           | Y1   |              |      | X1              | Y1   |
| X2           | Y2   |              |      | X2              | Y2   |
| X3           | Y3   |              |      | X3              | Y3   |
| X4           | Y4   |              |      | X4              | Y4   |
| X5           | Y5   |              |      | X5              | Y5   |

合併二個資料集

| <b>DATA1</b> |      | <b>DATA2</b> |      | <b>COMBINED</b> |      |      |
|--------------|------|--------------|------|-----------------|------|------|
| Var          | VarX | Var          | VarY | Var             | VarX | VarY |
| 1991         | X1   | 1991         | Y1   | 1991            | X1   | Y1   |
| 1992         | X2   | 1991         | Y2   | 1991            | X1   | Y2   |
| 1993         | X3   | 1993         | Y3   | 1992            | X2   | *    |
| 1994         | X4   | 1994         | Y4   | 1993            | X3   | Y3   |
| 1995         | X5   | 1995         | Y5   | 1994            | X4   | Y4   |
|              |      |              |      | 1995            | X5   | Y5   |

```
data combined;
 set data1;
 set data2;
run;

data combined;
 merge data1 data2;
run;
```

```
data combined;
 merge data1 data2;
 by Year;
run;
```

# 範例 – 增加新變數 (一對一)

資料集: `animal1`

| common | animal |
|--------|--------|
| A      | Ant    |
| B      | Bird   |
| C      | Cat    |
| D      | Dog    |
| E      | Eagle  |
| F      | Frog   |

資料集: `plant1`

| common | plant    |
|--------|----------|
| A      | Apple    |
| B      | Banana   |
| C      | Coconut  |
| D      | Dewberry |
| E      | Eggplant |
| G      | Fig      |

# 範例 - 增加新變數 (一對一)

```
data combined;
 merge animal1 plant1;
run;

proc print data=combined;
title 'Data Set COMBINED';
run;
```

| Obs | common | animal | plant    |
|-----|--------|--------|----------|
| 1   | a      | Ant    | Apple    |
| 2   | b      | Bird   | Banana   |
| 3   | c      | Cat    | Coconut  |
| 4   | d      | Dog    | Dewberry |
| 5   | e      | Eagle  | Eggplant |
| 6   | g      | Frog   | Fig      |

# 範例 -- 增加新變數 (一對多)

資料集: `animal121`

| common | animal |
|--------|--------|
| A      | Ant    |
| B      | Bird   |
| C      | Cat    |
| D      | Dog    |
| E      | Eagle  |
| F      | Frog   |

資料集: `plant1`

| common | Plant    |
|--------|----------|
| A      | Apple    |
| B      | Banana   |
| C      | Coconut  |
| C      | Celery   |
| D      | Dewberry |
| E      | Eggplant |

# 範例 - 增加新變數 (一對多)

資料

```
data merge2;
 merge anima21
 plant1;
run;
```

輸出

| Obs | common | animal | plant    |
|-----|--------|--------|----------|
| 1   | a      | Ant    | Apple    |
| 2   | b      | Bird   | Banana   |
| 3   | c      | Cat    | Coconut  |
| 4   | c      | Dog    | Celery   |
| 5   | d      | Eagle  | Dewberry |
| 6   | e      | Frog   | Eggplant |

# 範例 -- 增加新變數 (配對)

資料集: **animal2**

| common | Animal |
|--------|--------|
| A      | Ant    |
| B      | Bird   |
| C      | Cat    |
| D      | Dog    |
| E      | Eagle  |
| F      | Frog   |

資料集: **plant2**

| common | plant    |
|--------|----------|
| A      | Apple    |
| B      | Banana   |
| C      | Coconut  |
| D      | Dewberry |
| E      | Eggplant |
| G      | Fig      |

# 範例 - 增加新變數 (配對)

資料

| Obs | common | animal | plant    |
|-----|--------|--------|----------|
| 1   | a      | Ant    | Apple    |
| 2   | b      | Bird   | Banana   |
| 3   | c      | Cat    | Coconut  |
| 4   | d      | Dog    | Dewberry |
| 5   | e      | Eagle  | Eggplant |
| 6   | f      | Frog   |          |
| 7   | g      |        | Fig      |

程式

```
data combined;
merge animal2
plant2;
by common;
run;
proc print
data=combined;
title 'Data Set
COMBINED';
run;
```

# IN

- 目的： 產生指標變數來說明目前資料對現在觀察值的貢獻情形
- 語法：  
 $IN=variable$
- 說明：
  - IN 用在 SET 及 MERGE 中
  - IN 若與 BY 合用時，若資料有配對到 IN 之值會設為 1。

# 增加新變數 (配對)

資料集: **animal**

| common | animal |
|--------|--------|
| A      | Ant    |
| B      | Bird   |
| C      | Cat    |
| D      | Dog    |
| E      | Eagle  |
| F      | Frog   |

資料集: **plant**

| common | plant    |
|--------|----------|
| A      | Apple    |
| B      | Banana   |
| C      | Coconut  |
| D      | Dewberry |
| E      | Eggplant |
| G      | Fig      |

# 範例

```
data combined;
merge animal1 (in=in1) plant2
 (in=in2);
by common;
flag1=in1;
flag2=in2;
run;
```

| Obs | common | animal | plant    | flag1 | flag2 |
|-----|--------|--------|----------|-------|-------|
| 1   | a      | Ant    | Apple    | 1     | 1     |
| 2   | b      | Bird   | Banana   | 1     | 1     |
| 3   | c      | Cat    | Coconut  | 1     | 1     |
| 4   | d      | Dog    | Dewberry | 1     | 1     |
| 5   | e      | Eagle  | Eggplant | 1     | 1     |
| 6   | f      | Frog   |          | 1     | 0     |
| 7   | g      |        | Fig      | 0     | 1     |

資料集選項

常用選項

# 常用選項

| 指令     | 作用            |
|--------|---------------|
| DROP   | 刪除資料集中設定變數    |
| IN     | 產生觀察值存在資料集的指標 |
| KEEP   | 保留資料集中設定變數    |
| OUTPUT | 輸出觀察值         |
| RENAME | 更新變數名稱        |
| WHERE  | 保留滿足條件的觀察值    |

# DROP 刪除變數

- 語法：
  - `DROP=variable-1 <...variable-n>`
  - `DROP varlists;`
- 說明
  - DATA 步驟可以有二種使用方式
  - PROC 步驟只可使用 `DROP=`

# 範例

```
data plan1;
 set payroll (drop=salary
 gender);
run;

data plan1p;
 set payroll;
 drop salary gender;
run;
```

# KEEP 保留變數

- 語法：
  - KEEP=variable-1 <...variable-n>
  - KEEP varlists;
- 說明
  - DATA 步驟可以有二種使用方式
  - PROC 步驟只可使用 KEEP=

# 範例

```
data combined;
 set combined (keep=common animal
 plant);
run;

data combined;
 set combined;
 keep common animal plant;
run;
```

# RENAME

- 目的： 重新命名變數名稱
- 語法：
  - RENAME= (old-name-1=new-name-1  
<...old-name-n=new-name-n>)
  - RENAME old-name-1=new-name-1  
<...old-name-n=new-name-n> ;
- 說明
  - DATA 步驟可以有二種使用方式
  - PROC 步驟只可使用 DROP=

# 範例

```
data combined;
 set combined
 (rename=(common=id)) ;
run;

data combined;
 set combined;
 rename common=id;
run;
```

# WHERE

- 目的：選擇滿足條件的觀察值
  - 邏輯字元： AND (&) , OR (|) , NE (^=) , EQ (=) , LE (<=) , GE (>=) , LT (<) , GT (>)
- 語法：
  - WHERE=(where-expression-1<logical-operator where-expression-n>)
  - WHERE where-expression-1<logical-operator where-expression-n>;

# 範例

```
data combined;
 set combined
 (where=(common='A'));
run;

data whizmo;
 set sales;
 where common='A';
run;
```

# OUTPUT 指令

- 目的：強迫 SAS 輸出一個新的觀察值或新的資料集
  - SAS 會立刻將新的觀察值輸出
- 語法：

OUTPUT<data-set-name(s)>;

# 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 <sup>2</sup> ) |    |    |    |    |  |
|------|---------------------------|----|----|----|----|--|
| 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 |  |

# 範例 --

```
data ch4d12;
do cotton = 15 to 35 by 5;
 do rep = 1 to 5;
 input strength @@;
 output;
 end;
end;
cards;
7 7 15 11 9
12 17 12 18 18
14 18 18 19 19
19 25 22 19 23
7 10 11 15 11
run;
```

# 範例

```
data ch4d13;
infile 'statpackch4d13.txt'
firstobs=32;
input id group $ lead0
lead1 lead4 lead6;
run;
proc print data=t1c (obs=10);
run;
```

# 範例

| Obs | id | group | lead0 | lead1 | lead4 | lead6 |
|-----|----|-------|-------|-------|-------|-------|
| 1   | 1  | P     | 30.8  | 26.9  | 25.8  | 23.8  |
| 2   | 2  | A     | 26.5  | 14.8  | 19.5  | 21.0  |
| 3   | 3  | A     | 25.8  | 23.0  | 19.1  | 23.2  |
| 4   | 4  | P     | 24.7  | 24.5  | 22.0  | 22.5  |
| 5   | 5  | A     | 20.4  | 2.8   | 3.2   | 9.4   |
| 6   | 6  | A     | 20.4  | 5.4   | 4.5   | 11.9  |
| 7   | 7  | P     | 28.6  | 20.8  | 19.2  | 18.4  |
| 8   | 8  | P     | 33.7  | 31.6  | 28.5  | 25.1  |
| 9   | 9  | P     | 19.7  | 14.9  | 15.3  | 14.7  |
| 10  | 10 | P     | 31.1  | 31.2  | 29.2  | 30.1  |

# 範例

## 產生多筆觀察值

```
Data ch4d13a;
 set ch4d13;
 y=lead0; time=0; output;
 y=lead1; time=1; output;
 y=lead4; time=4; output;
 y=lead6; time=6; output;
 drop lead0 lead1 lead4
 lead6;

 run;

 proc print data=ch4d13a
 (obs=20);

 run;
```

## 輸出

| obs | id | group | y    | time |
|-----|----|-------|------|------|
| 1   | 1  | P     | 30.8 | 0    |
| 2   | 1  | P     | 26.9 | 1    |
| 3   | 1  | P     | 25.8 | 4    |
| 4   | 1  | P     | 23.8 | 6    |
| 5   | 2  | A     | 26.5 | 0    |
| 6   | 2  | A     | 14.8 | 1    |
| 7   | 2  | A     | 19.5 | 4    |
| 8   | 2  | A     | 21.0 | 6    |
| 9   | 3  | A     | 25.8 | 0    |
| 10  | 3  | A     | 23.0 | 1    |
| 11  | 3  | A     | 19.1 | 4    |
| 12  | 3  | A     | 23.2 | 6    |
| 13  | 4  | P     | 24.7 | 0    |
| 14  | 4  | P     | 24.5 | 1    |
| 15  | 4  | P     | 22.0 | 4    |
| 16  | 4  | P     | 22.5 | 6    |
| 17  | 5  | A     | 20.4 | 0    |
| 18  | 5  | A     | 2.8  | 1    |
| 19  | 5  | A     | 3.2  | 4    |
| 20  | 5  | A     | 9.4  | 6    |

# 範例

## 產生多筆資料集

```
data tlca tlcp;
 set ch4d13;
 if group='A' then
 output tlca;
 else if group='P'
 then output tlcp;
run;
proc print data=tlca
 (obs=5);
run;
proc print data=tlcp
 (obs=5);
run;
```

## 輸出

| Obs | id | group | lead0 | lead1 | lead4 | lead6 |
|-----|----|-------|-------|-------|-------|-------|
| 1   | 2  | A     | 26.5  | 14.8  | 19.5  | 21.0  |
| 2   | 3  | A     | 25.8  | 23.0  | 19.1  | 23.2  |
| 3   | 5  | A     | 20.4  | 2.8   | 3.2   | 9.4   |
| 4   | 6  | A     | 20.4  | 5.4   | 4.5   | 11.9  |
| 5   | 12 | A     | 24.8  | 23.1  | 24.6  | 30.9  |

| Obs | id | group | lead0 | lead1 | lead4 | lead6 |
|-----|----|-------|-------|-------|-------|-------|
| 1   | 1  | P     | 30.8  | 26.9  | 25.8  | 23.8  |
| 2   | 4  | P     | 24.7  | 24.5  | 22.0  | 22.5  |
| 3   | 7  | P     | 28.6  | 20.8  | 19.2  | 18.4  |
| 4   | 8  | P     | 33.7  | 31.6  | 28.5  | 25.1  |
| 5   | 9  | P     | 19.7  | 14.9  | 15.3  | 14.7  |

資料集選項

產生隨機樣本

(113)

# 離散型分配函數

| 指令                                    | 說明    |
|---------------------------------------|-------|
| RANBIN (seed, n, p)                   | 二項 分配 |
| RANPOI (seed, m)                      | 波松分配  |
| RANTBL (seed, $p_1, \dots, p_k$ )     | 離散分配  |
| RAND ('dist', parm-1, ...,<br>parm-k) | 特定分配  |

# 連續型分配函數

| 指令                  | 說明               |
|---------------------|------------------|
| RANNOR (seed)       | 標準常態分配           |
| NORMAL (seed)       | 標準常態分配           |
| RANCAU (seed)       | 柯西分配             |
| RANEXP (seed)       | 標準指數分配           |
| RANGAM (seed, $a$ ) | 尺度為 1 的 GAMMA 分配 |
| RANUNI (seed)       | 均匀分配 $(0, 1)$    |
| UNIFORM (seed)      | 均匀分配 $(0, 1)$    |

# 說明 -- SEED

- 作用：產生依序列隨機亂數的起始點
- SEED 是一個數值常數或是整數的表示式
- 只要使用相同的 SEED, SAS 會產生相同序列
- 若  $SEED < 0$ , SAS 使用現在的日期開始
- 範圍： $< 2^{31} - 1$

# 範例

```
data ch4d14;
rpoin=ranpoi(100,1);
rtbl=rantbl(100,0.2,0.2,0.3,0
.10,0.15,0.05);
rnor=rannor(100);
runif=uniform(100);
ppoin=poisson(1,1);
pnorm=probnorm(1.96);
run;
```

# 範例

- 一般嚴重有退化性膝關節炎的老年人都會做膝蓋替換手術 (total knee replacement surgery; TKRS) 。
- 感染是常見的副作用，而糖尿病則以之為重要的危險因子。
- 資料集： statpackch4d4.sas7bdat
- 將病人隨機分二組

```
Data ch4d15;
 set statpackch4d4;
 group= (uniform(123)<=0.5);
run;
```

# 範例 -- 產生多個變數

```
data ch4d14;
do n=1 to 100;
 Normal_x=10*rannor(53124)+50;
 Exponential_x=ranexp(18746363);
 output; /* 命令 SAS 輸出新觀察值 */
end;
drop n;
label
 Normal_x='Normal Random Variable'
 Exponential_x='Exponential Random
 Variable';
run;
```

# 範例 - 輸出

※

| Obs | Normal_x | Exponential_x |
|-----|----------|---------------|
| 1   | 53.1482  | 1.32262       |
| 2   | 53.3818  | 0.11661       |
| 3   | 35.4927  | 1.35538       |
| 4   | 34.2061  | 2.32706       |
| 5   | 56.1399  | 0.06677       |
| 6   | 52.6867  | 0.19636       |
| 7   | 59.1349  | 0.12954       |
| 8   | 43.7926  | 0.55627       |
| 9   | 55.1741  | 0.77812       |
| 10  | 60.4088  | 0.78040       |