



臨床試驗常用統計分析方法

SAS統計軟體操作 (1)

醫學研究部 基礎醫學科 生統小組：陳韻仔 博士

授課日期：112年5月16日

The Role of Statistic in Clinical Trials

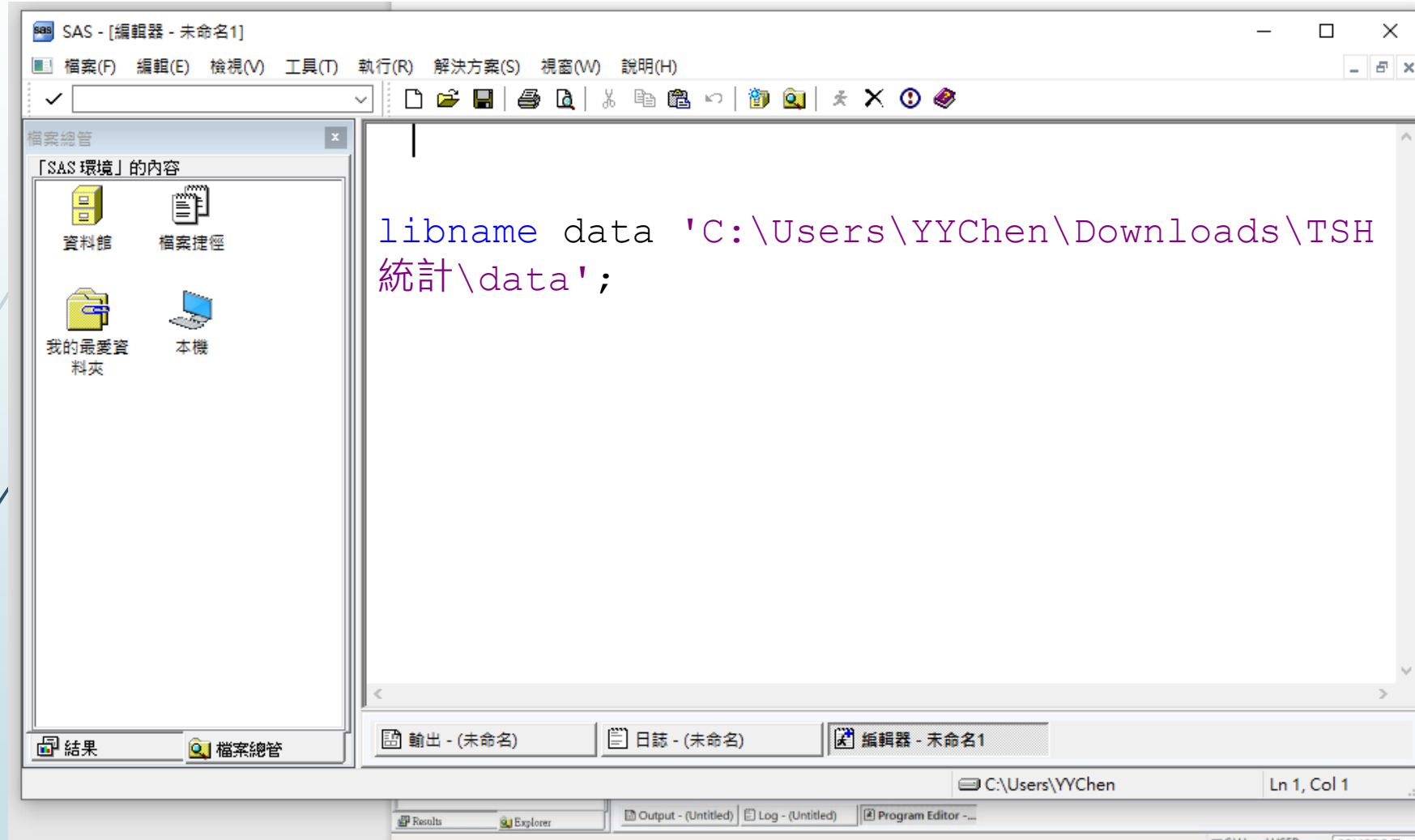
- **Sample size estimation**
- **Randomized assignment of treatment**
- **Analyze results (hypothesis testing and effect estimation)**
- **Control for confounding**

SAS在臨床試驗的重要性

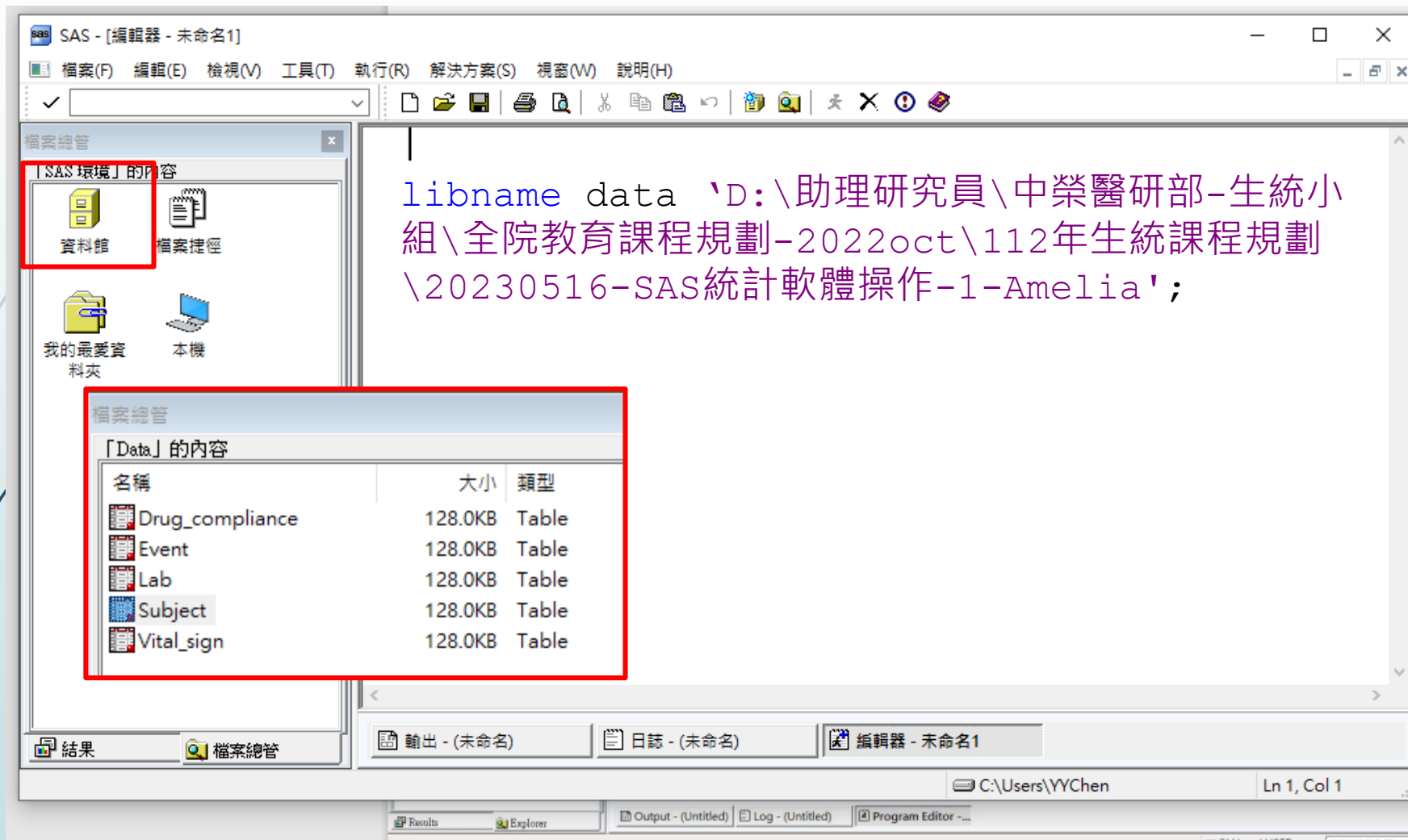
- ▶ **數據管理**：臨床試驗需要收集大量數據，包括受試者基本信息、治療方案、生命體徵、實驗室數據等等。SAS可以幫助研究人員有效地管理這些數據
- ▶ **數據分析**：豐富的統計分析功能，如描述性統計分析、存活分析、多變量分析、幫助研究人員從數據中提取有價值的信息，評估新藥的安全性和有效性
- ▶ **數據視覺化呈現**：將數據以圖表、表格等形式呈現，使數據更加直觀和易於理解，幫助研究人員發現數據中的規律和趨勢

SAS 9.4 起始畫面

4



設定SAS永久檔位置



The screenshot shows the SAS software interface. On the left, the '檔案總管' (File Explorer) window displays the 'SAS 環境' (SAS Environment) folder, which contains a '資料庫' (Database) folder. Below it, a detailed view of the 'Data' folder is shown, listing several tables:

名稱	大小	類型
Drug_compliance	128.0KB	Table
Event	128.0KB	Table
Lab	128.0KB	Table
Subject	128.0KB	Table
Vital_sign	128.0KB	Table

In the main editor window, the following SAS code is displayed:

```
libname data 'D:\助理研究員\中榮醫研部-生統小組\全院教育課程規劃-2022oct\112年生統課程規劃\20230516-SAS統計軟體操作-1-Amelia';
```

Clinical Trial Data Management using SAS

- **Clinical Trial Data Management using SAS mostly for the following purposes:**
 - **Designing a protocol and beginning a study : Sample size**
 - Recruitment of Patients and Investigators : Randomization
 - Health Data Management : Data Management using SAS
 - Analyzing Data

Designing a Protocol and Beginning a Study : Sample Size

- ▶ 範例：想驗證新藥AAA的療效
 - ▶ 執行一個第三期、隨機、雙盲、多中心試驗，比較新藥AAA與安慰劑兩組在降血壓的療效上有無差異？
- ▶ 主要療效：8週後血壓 v.s. 基礎血壓質的差異
- ▶ 統計方法：連續型變數的T-test
- ▶ 虛無假說：新藥AAA與安慰劑兩組在降血壓的效果**無差異**
- ▶ 對立假說：新藥AAA與安慰劑兩組在降血壓的效果**有差異**

Designing a Protocol and Beginning a Study : Sample Size

8

虛無假說：新藥AAA與安慰劑兩組在降血壓的效果**無差異**

對立假說：新藥AAA與安慰劑兩組在降血壓的效果**有差異**

		實際狀況	
		新藥 AAA 沒有降血壓的效果	新藥 AAA 是有降血壓的效果(8 週後至少降 10 mmHg)
t-test 檢定後的決定	新藥 AAA 沒有降血壓的效果	正確決定	型 II 誤差 β (type II error)
	新藥 AAA 是有降血壓的效果(8 週後至少降 10 mmHg)	型 I 誤差 α (type I error)	正確決定 $(1-\beta)$ (power)

檢定力 (Power)：正確拒絕「錯誤的虛無假設」的機率

在樞紐性第三期臨床試驗相關法規規定，
整體型一誤差應控制在雙尾5%或以下；
整體型二誤差應控制在20%或以下；
檢定力 > 80%

Designing a Protocol and Beginning a Study : Sample Size

9

計算連續型變數要考慮的因素

1. 試驗目標	以安慰劑為對照之較優性試驗
2. 主要評估指標	8週後的血壓值與基礎值血壓值之差異
3. 虛無假說 (H0)	$\mu_{\text{Placebo}} = \mu_{\text{Drug AAA}}$
4. 對立假說 (Ha)	$\mu_{\text{Placebo}} \neq \mu_{\text{Drug AAA}}$
5. 統計檢定量	t-test
6. 型一誤差機率	$\alpha = 0.05$ (雙尾)
7. 型二誤差機率 (檢定力)	$\beta = 0.20$ ($1 - \beta = 0.80$)
8. 預期療效大小 (d)	3 mmHg
9. 標準差 (σ)	8 mmHg
10. 受試者隨機分派比 (AAA : placebo)	1 : 1 (m : 1)
11. 預期中途退出率(p)	p = 10%

$$N = \frac{\left(1 + \frac{1}{m}\right) * (Z_{\alpha} + Z_{\beta})^2 \sigma^2}{d^2}$$

考慮預期失訪率之樣本數

$$\frac{N}{(1-p)^2}$$

人工樣本數計算

$$\frac{(1+1) * (1.96 + 0.845)^2 * 8^2}{3^2} \approx 112, \quad \frac{112}{(1-0.1)^2} = 138$$

Designing a Protocol and Beginning a Study : Sample Size

10

► 利用 SAS 計算 連續型變數 T-test 差異的方法

```
dm "log;clear;output;clear;program;recall";  
title 'Clinical trials - two independent  
sample: contiuous variables';  
PROC POWER;  
TWOSAMPLEMEANS TEST=DIFF  
MEANDIFF=3  
STDDEV=8  
APLHA=0.05  
POWER=0.8  
NPERGROUP=.;  
run;
```

Clinical trials - two independent sample: contiuous variables

POWER 程序
平均差的雙樣本 t 檢定

固定案例元素	
分布	常態
方法	精進
Alpha	0.05
平均差	3
標準差	8
名目次方	0.8
側邊數目	2
Null 差異	0

計算 每個群組的 N	
實際次方	每個群組的 N
0.801	113

Designing a Protocol and Beginning a Study : Sample Size

11

► 計算 類別型變數 要考慮的因素

1. 試驗目標	以安慰劑為對照之較優性試驗
2. 主要評估指標	8 週後的血壓值 < 90 mmHg 的比率
3. 虛無假說 (H0)	$p_{1(\text{placebo})} = p_{2(\text{AAA})}$
4. 對立假說 (Ha)	$p_{1(\text{placebo})} \neq p_{2(\text{AAA})}$
5. 統計檢定量	Z-test
6. 型一誤差機率	$\alpha = 0.05$ (雙尾)
7. 型二誤差機率 (檢定力)	$\beta = 0.20$ ($1 - \beta = 0.80$)
8. 預期之療效大小(d)	10%
9. p_1 各組 < 90 mmHg 的比例	$p_1 = 10\%$
10. 受試者隨機分派比 (AAA : placebo)	1 : 1 (m : 1)
11. 預期中途退出率 (p)	$p = 5\%$

$$P_2 - P_1 = 20\% - 10\% = 10\%$$

$$P_2 = 10\% + 10\% = 20\%$$

類別變數計算各組樣本數之公式

$$N = \frac{\{Z_\alpha \sqrt{2\bar{p}(1-\bar{p})} + Z_\beta \sqrt{p_1(1-p_1) + p_2(1-p_2)}\}^2}{(p_1 - p_2)^2}, \text{ where } \bar{p} = \frac{p_1 + p_2}{2}$$

考慮預期失訪率之樣本數

$$\frac{N}{(1-p)^2}$$

人工樣本數計算

$$\frac{\{1.96\sqrt{2*0.15*0.85} + 0.845\sqrt{0.1*0.9 + 0.2*0.8}\}^2}{(0.1-0.2)^2} = 200, \quad \frac{200}{(1-0.05)^2} = 222$$

Designing a Protocol and Beginning a Study : Sample Size

► 利用 SAS 計算 類別型變數 比例差異的方法

```

title 'Cclinical trials - two independent
sample: proportion difference ';
PROC POWER;
TWO SAMPLE FREQ TEST=PCHI
GROUP PROPORTIONS = (0.1 0.2)
NULL PROPORTION DIFF = 0
ALPHA=0.05
POWER=0.8
NPERGROUP=.;
run;

```

Cclinical trials - two independent sample: proportion difference

POWER 程序
兩個比例的 Pearson 卡方檢定

固定案例元素	
分布	漸近常態
方法	常態近似
Null 比例差異	0
Alpha	0.05
群組 1 比例	0.1
群組 2 比例	0.2
名目次方	0.8
側邊數目	2

計算 每個群組的 N	
實際次方	每個群組的 N
0.800	199

Designing a Protocol and Beginning a Study : Sample Size

13

► 利用 SAS 計算 兩組相對風險 的方法

```
title 'Clinical trials - two independent  
sample: relative risk';  
PROC POWER;  
TWO SAMPLE FREQ TEST=PCHI  
RELATIVERISK=0.6667  
REFPROPORTION=0.6  
ALPHA=0.05  
POWER=0.8  
NPERGROUP=.;  
run;
```

REFPROPORTION: 參考組 事件數 的比例

Clinical trials - two independent sample: relative risk

POWER 程序
兩個比例的 Pearson 卡方檢定

固定案例元素	
分布	漸近常態
方法	常態近似
Alpha	0.05
參考 (群組 1) 比例	0.6
相對風險	0.6667
名目次方	0.8
側邊數目	2
Null 相對風險	1

計算 每個群組的 N	
實際次方	每個群組的 N
0.800	97

Clinical Trial Data Management using SAS

- **Clinical Trial Data Management using SAS mostly for the following purposes:**
 - Designing a protocol and beginning a study : Sample size
 - **Recruitment of Patients and Investigators : Randomization**
 - Health Data Management : Data Management using SAS
 - Analyzing Data

Methods of Randomization

- ▶ **Fixed randomization (固定隨機分派)**：是一種基於機率的隨機分派方法，它通過事先固定好的隨機分派表或隨機數生成器來實現分派。
 - ▶ **Allocation ratio (分派比例)**：是一種最常見的隨機分派方法，它是在各處理組之間按照一定比例進行分配的，通常是1：1或2：1等等。
 - ▶ **Allocation strata (分派層次)**：是一種在特定層次內進行分派的方法，通常是基於某些相關的特性，例如性別、年齡、病情等，以達到分組均衡的效果。這種方法通常用於比較不平衡的樣本組合。
 - ▶ **Block size (隨機區組設計)**：是一種在隨機分派中用於控制樣本均衡性的方法，它通常將樣本按照順序分為若干組，每一組中包含一定數量的樣本，以保持各處理組之間樣本均衡性。Moses-Oakford算法可以用來生成固定或可變的分組大小，以更好地控制樣本均衡性。

Methods of Randomization

► 利用 SAS 進行 Block Randomization

```
dm "log; clear; output;clear; program; recall;";
title "Generation of Random Codes for Parallel Groups";
option nodate ls=78 ps=120 nodate;
data one;
do block=1 to 6; /*有2種藥; 6種block組合的隨機排列要如何做*/
do sub=1 to 4;
input tmt @;
output;
end;
end;
cards;
1 1 2 2 2 2 1 1 1 2 1 2 1 2 2 1 2 1 2 1 2 1 1 2 ;
data one; set one;
do center=1 to 4;
output;
end;
proc sort data=one;
by center block sub;
proc plan ordered;
factors center=4 nblock=6;
treatments block=6 random;
output out=perblock;
run;
```

Generation of Random Codes for Parallel Groups

PLAN 程序

標繪圖因子			
因子	選取	層級	順序
center	4	4	Ordered
nblock	6	6	Ordered

處理因子

因子	選取	層級	順序
block	6	6	Random

center	nblock						block					
1	1	2	3	4	5	6	2	4	6	1	3	5
2	1	2	3	4	5	6	3	2	5	1	6	4
3	1	2	3	4	5	6	2	1	5	6	4	3
4	1	2	3	4	5	6	4	3	1	6	2	5

Methods of Randomization

利用 SAS 進行 Block Randomization

```
proc sort data=perblock; by center block;
data perblock; merge perblock one; by center block;
proc sort data=perblock; by center nblock sub;
data perblock; set perblock;
sub=_n_;
subject=(14*100000)+center*1000+(sub-(center-1)*24);
proc format;
value tmtf 1='Active Drug'
          2='Placebo      ';
value centerf 1='J.Smith,MD'
              2='M.Dole,MD '
              3='A.Hope,MD '
              4='C.Price,MD';
proc sort data=perblock; by center;
proc print data=perblock split='*'; page by center; by center;
id subject;
var tmt;
label subject='Subject*Number*_____ '
        tmt='Treatment*Assignment*_____ '
format tmt tmtf. center centerf.;
title1 '';
title2 'Example of Permuted-Block Randomization by Random Selection '
title3 'of Blocks for Four Centers and a Block Size of Four '
title4 'Program:[DRUGXXX.PXXX014.SAS.NTUJPL]PERBLK.SAS '
title5 'Random Codes for Drug XXX, Protocol XXX-014 '
title6 'Double-blind, Randomized, Placebo-Control, Two Parallel Groups';
run;
```

Example of Permuted-Block Randomization by Random Selection of Blocks for Four Centers and a Block Size of Four
Program:[DRUGXXX.PXXX014.SAS.NTUJPL]PERBLK.SAS
Random Codes for Drug XXX, Protocol XXX-014
Double-blind, Randomized, Placebo-Control, Two Parallel Groups

center=J.Smith,MD

Subject Number	Treatment Assignment
1401001	Placebo
1401002	Placebo
1401003	Active Drug
1401004	Active Drug
1401005	Active Drug
1401006	Placebo
1401007	Placebo

center=M.Dole,MD

Subject Number	Treatment Assignment
1402001	Active Drug
1402002	Placebo
1402003	Active Drug
1402004	Placebo
1402005	Placebo
1402006	Placebo
1402007	Active Drug
1402008	Active Drug
1402009	Placebo
1402010	Active Drug
1402011	Placebo
1402012	Active Drug

Methods of Randomization

利用 SAS 進行 Block Randomization

```
proc sort data=perblock; by center;  
proc freq data=perblock;  
tables center*tmt/nopercent nocol norow;  
format tmt tmtf.;  
run;
```

Example of Permuted-Block Randomization by Random Selection
of Blocks for Four Centers and a Block Size of Four
Program:[DRUGXXX.PXXX014.SAS.NTUJPL]PERBLK.SAS
Random Codes for Drug XXX, Protocol XXX-014
Double-blind, Randomized, Placebo-Control, Two Parallel Groups

FREQ 程序

次軸	Table of center by tmt		
	tmt		
center	Active Drug	Placebo	總計
1	12	12	24
2	12	12	24
3	12	12	24
4	12	12	24
總計	48	48	96

Methods of Randomization

利用 SAS 進行 Block Randomization

```
proc freq data=perblock; by center;  
tables nblock*tmt/nopercent nocol norow;  
format tmt tmtf.;  
run;
```

Example of Permuted-Block Randomization by Random Selection
of Blocks for Four Centers and a Block Size of Four
Program:[DRUGXXX.PXXX014.SAS.NTUJPL]PERBLK.SAS
Random Codes for Drug XXX, Protocol XXX-014
Double-blind, Randomized, Placebo-Control, Two Parallel Groups

FREQ 程序

center=4

次數	Table of nblock by tmt		
	tmt		
nblock	Active Drug	Placebo	總計
1	2	2	4
2	2	2	4
3	2	2	4
4	2	2	4
5	2	2	4
6	2	2	4
總計	12	12	24

Clinical Trial Data Management using SAS

- **Clinical Trial Data Management using SAS mostly for the following purposes:**
 - Designing a protocol and beginning a study : Sample size
 - Recruitment of Patients and Investigators : Randomization
 - **Health Data Management : Data Management using SAS**
 - Analyzing Data

Data Management using SAS: 保留重要分析變數

21

```
/*清理資料*/
```

只有data.subject檔案有Group

```
data data.subject; set data.subject;
keep Center_no Screening_no Ran_no Group Visit1-Visit6 Gender Age; *保留必要的變數;
*drop 刪除不必要的變數;
if Center_no=. then delete;
if Group=. then delete; run;

data data.Drug_compliance; set data.Drug_compliance;
keep Center_no Screening_no Ran_no Visit Visit_date taken_no returned_no lost_no period_day
compliance_r;
if Center_no=. then delete; run;

data data.event; set data.event;
keep Center_no Screening_no Ran_no Visit start_visit start_visit_date end_visit end_visit_date
total_no_record AF_total;
if Center_no=. then delete; run;

data data.vital_sign; set data.vital_sign;
keep Center_no Screening_no Ran_no Visit sbp dbp;
if Center_no=. then delete; run;

data data.lab; set data.lab;
keep Center_no Screening_no Ran_no Visit hemoglobin WBC RBC AST ALT creatinine;
if Center_no=. then delete; run;
```

Data Management using SAS

➡ 利用 SAS 進行 資料管理

短資料格式

適合治療前後 (配對) 差異比較

ID	Group	SBP_Visit_1	SBP_Visit_2	SBP_Visit_3	DBP_Visit_1	DBP_Visit_2	DBP_Visit_3
S101	1	98	120	99	62	68	63
S102	2	141	130	168	91	78	82

長資料格式

適合重複測量分析

ID	Group	Visit	SBP	DBP	AST	Creatinine	AF_event
S101	1	1	98	62	43	70	0
S101	1	2	120	68	65	89	1
S101	1	3	99	63	51	67	0
S102	2	1	141	91	55	111	1
S102	2	2	130	78	90	99	1
S102	2	3	168	82	41	83	1

Data Management using SAS

➔ 利用 SAS 進行 資料管理

短資料格式：Data.subject_A

	Center_no	Screening_no	Ran_no	Group	Visit1	Visit2	Visit3	Visit4	Visit5	Visit6	Gender	Age	
1	1	S101	F-R101		2	02SEP2019	02SEP2019	09SEP2019	23SEP2019	28NOV2019	10FEB2020	1	70.024657534
2	1	S102	R-R101		1	21FEB2020	21FEB2020	*****	*****	*****	06AUG2020	1	64.084931507
3	1	S103	R-R102		1	09JUL2020	09JUL2020	16JUL2020	27JUL2020	09OCT2020	28DEC2020	1	64.44109589
4	2	S201	F-R201		2	10JUL2019	10JUL2019	18JUL2019	06AUG2019	01OCT2019	26DEC2019	1	37.260273973
5	2	S202	F-R202		1	28AUG2019	28AUG2019	10SEP2019	24SEP2019	.	26NOV2019	1	66.728767123
6	2	S203	F-R203		1	29NOV2019	29NOV2019	10DEC2019	24DEC2019	*****	09JUN2020	1	44.391780822
7	2	S204	R-R201		1	03JAN2020	03JAN2020	14JAN2020	04FEB2020	21APR2020	14JUL2020	0	68.726027397
8	3	S301	R-R301		2	01OCT2019	01OCT2019	08OCT2019	22OCT2019	31DEC2019	*****	0	67.260273973
9	3	S303	F-R301		2	15OCT2019	15OCT2019	22OCT2019	05NOV2019	.	14JAN2020	1	71.671232877
10	3	S304	R-R302		1	01NOV2019	01NOV2019	08NOV2019	22NOV2019	31JAN2020	24APR2020	1	68.205479452

短資料格式：Data.subject_B

	Center_no	Screening_no	Ran_no	Group	Visit1	Visit2	Visit3	Visit4	Visit5	Visit6	Gender	Age	
12	3	S306	F-R303		2	22NOV2019	22NOV2019	29NOV2019	13DEC2019	21FEB2020	*****	0	70.205479452
13	3	S307	R-R303		2	10DEC2019	10DEC2019	17DEC2019	31DEC2019	*****	02JUN2020	1	48.315068493
14	3	S308	R-R304		1	13DEC2019	13DEC2019	20DEC2019	03JAN2020	*****	05JUN2020	1	75.638356164
15	3	S309	R-R305		2	14JAN2020	14JAN2020	21JAN2020	04FEB2020	14A			
16	3	S310	R-R306		2	04FEB2020	04FEB2020	12FEB2020	26FEB2020	***			
17	3	S311	R-R307		1	20FEB2020	20FEB2020	27FEB2020	*****	***			
18	3	S312	F-R304		1	27FEB2020	27FEB2020	*****	*****	***			
19	3	S313	R-R308		1	*****	*****	*****	01APR2020	10J			
20	3	S314	R-R309		2	03APR2020	03APR2020	10APR2020	30APR2020	03J			

先sort 照相同id排序，再直向 合併不同檔案：set

短資料格式：Data.subject

```
proc sort data=data.subject_A;  
by Center_no Screening_no ; run; *短資料格式: Visit  
1-Visit 6;  
  
proc sort data=data.subject_B;  
by Center_no Screening_no ; run; *短資料格式: Visit  
1-Visit 6;  
  
data data.subject;  
set data.subject_A (in=a) data.subject_B (in=b) ;  
by Center_no Screening_no ; run;
```

Data Management using SAS

➔ 利用 SAS 進行 資料管理

短資料格式 : Data.subject

	Center_no	Screening_no	Ran_no	Group	Visit1	Visit2	Visit3	Visit4	Visit5	Visit6	Gender	Age
1	1	S101	F-R101	2	02SEP2019	02SEP2019	09SEP2019	23SEP2019	28NOV2019	10FEB2020	1	70.024657534
2	1	S102	R-R101	1	21FEB2020	21FEB2020	*****	*****	*****	06AUG2020	1	64.084931507
3	1	S103	R-R102	1	09JUL2020	09JUL2020	16JUL2020	27JUL2020	09OCT2020	28DEC2020	1	64.44109589
4	2	S201	F-R201	2	10JUL2019	10JUL2019	18JUL2019	06AUG2019	01OCT2019	26DEC2019	1	37.260273973
5	2	S202	F-R202	1	28AUG2019	28AUG2019	10SEP2019	24SEP2019	.	26NOV2019	1	66.728767123
6	2	S203	F-R203	1	29NOV2019	29NOV2019	10DEC2019	24DEC2019	*****	09JUN2020	1	44.391780822
7	2	S204	R-R201	1	03JAN2020	03JAN2020	14JAN2020	04FEB2020	21APR2020	14JUL2020	0	68.726027397
8	3	S301	R-R301	2	01OCT2019	01OCT2019	08OCT2019	22OCT2019	31DEC2019	*****	0	67.260273973
9	3	S303	F-R301	2	15OCT2019	15OCT2019	22OCT2019	05NOV2019	.	14JAN2020	1	71.671232877
10	3	S304	R-R302	1	01NOV2019	01NOV2019	08NOV2019	22NOV2019	31JAN2020	24APR2020	1	68.205479452

橫向 合併不同檔案 : merge

長資料格式 : Data.vital_sign

	Center_no	Screening_no	Ran_no	Visit	SBP	DBP
1	1	S101	F-R101	1	138	79
2	1	S101	F-R101	2	138	79
3	1	S101	F-R101	3	144	74
4	1	S101	F-R101	4	140	68
5	1	S101	F-R101	5	121	67
6	1	S101	F-R101	6	126	74
7	1	S102	R-R101	1	133	88
8	1	S102	R-R101	2	133	88
9	1	S102	R-R101	3	121	69
10	1	S102	R-R101	4	133	89
11	1	S102	R-R101	5	125	80
12	1	S102	R-R101	6	127	81

長資料格式 : Data.lab

	Center_no	Screening_no	Ran_no	Visit	Hemoglobin	RBC	WBC	Creatinine
1	1	S101	F-R101	1	15.8	5.31	5440	1.15
2	1	S101	F-R101	3	15.6	5.29	5510	1.06
3	1	S101	F-R101	6	14.7	5.06	5050	1.24
4	1	S102	R-R101	1	13.3	4.54	5260	1.21
5	1	S102	R-R101	3	12.7	4.4	6070	1.14
6	1	S102	R-R101	6	13.6	4.54	6.34	1.1
7	1	S103	R-R102	1	16.8	5.41	9.09	1.33
8	1	S103	R-R102	3	16.3	5.01	8.93	1.2
9	1	S103	R-R102	6	12.8	3.94	8.36	1.23
10	2	S201	F-R201	1	16.2	4.98	5200	1.22
11	2	S201	F-R201	3	15.5	4.72	9100	1.1
12	2	S201	F-R201	6	15.6	4.6	7600	1.2

先sort 照相同id排序，再橫向 合併不同檔案：merge

```
proc sort data=data.subject;
by Center_no Screening_no ; run; *短資料格式: Visit 1-Visit 6;

proc sort data=data.Drug_compliance;
by Center_no Screening_no Visit; run; *長資料格式→有2次資料: Visit 3, Visit 5;

proc sort data=data.event;
by Center_no Screening_no Visit; run; *長資料格式→有4次資料: Visit 1, Visit 3, Visit 4, Visit
6;

proc sort data=data.Vital_sign; *長資料格式→有六次資料: Visit 1-Visit 6;
by Center_no Screening_no Visit; run;

proc sort data=data.lab;
by Center_no Screening_no Visit; run; *長資料格式→有3次資料: Visit 1, Visit 3, Visit 6;

data data.all;
merge data.Vital_sign (in=a) data.Drug_compliance (in=b) data.event (in=c) data.lab
(in=d) ;
by Screening_no visit ; run;

data data.final;
merge data.subject (in=a) data.all (in=b) ;
by Screening_no ; run;
```

Data Management using SAS

➡ 利用 SAS 進行 資料管理

長資料格式：Data.final

先sort 照相同id排序，再橫向 合併不同檔案：merge

	Center_no	Screening_no	Ran_no	Group	Visit1	Visit2	Visit3	Visit4	Visit5	Visit6	Gender	Age	Visit	SBP	DBP	Visit_date	Taken_no	Returned_no	Lost_no	Period_day	
1	1	S101	F-R101		2	02SEP2019	02SEP2019	09SEP2019	23SEP2019	28NOV2019	10FEB2020	1	70.024657534	1	138	79	
2	1	S101	F-R101		2	02SEP2019	02SEP2019	09SEP2019	23SEP2019	28NOV2019	10FEB2020	1	70.024657534	2	138	79	
3	1	S101	F-R101		2	02SEP2019	02SEP2019	09SEP2019	23SEP2019	28NOV2019	10FEB2020	1	70.024657534	3	144	74	09SEP2019	200	84	10	80
4	1	S101	F-R101		2	02SEP2019	02SEP2019	09SEP2019	23SEP2019	28NOV2019	10FEB2020	1	70.024657534	4	140	68	
5	1	S101	F-R101		2	02SEP2019	02SEP2019	09SEP2019	23SEP2019	28NOV2019	10FEB2020	1	70.024657534	5	121	67	28NOV2019	212	82	0	74
6	1	S101	F-R101		2	02SEP2019	02SEP2019	09SEP2019	23SEP2019	28NOV2019	10FEB2020	1	70.024657534	6	126	74	
7	1	S102	R-R101		1	21FEB2020	21FEB2020	*****	*****	*****	06AUG2020	1	64.084931507	1	133	88	
8	1	S102	R-R101		1	21FEB2020	21FEB2020	*****	*****	*****	06AUG2020	1	64.084931507	2	133	88	
9	1	S102	R-R101		1	21FEB2020	21FEB2020	*****	*****	*****	06AUG2020	1	64.084931507	3	121	69	*****	203	77	14	72
10	1	S102	R-R101		1	21FEB2020	21FEB2020	*****	*****	*****	06AUG2020	1	64.084931507	4	133	89	
11	1	S102	R-R101		1	21FEB2020	21FEB2020	*****	*****	*****	06AUG2020	1	64.084931507	5	125	80	*****	252	42	0	84
12	1	S102	R-R101		1	21FEB2020	21FEB2020	*****	*****	*****	06AUG2020	1	64.084931507	6	127	81	
13	1	S103	R-R102		1	09JUL2020	09JUL2020	16JUL2020	27JUL2020	09OCT2020	28DEC2020	1	64.44109589	1	138	78	
14	1	S103	R-R102		1	09JUL2020	09JUL2020	16JUL2020	27JUL2020	09OCT2020	28DEC2020	1	64.44109589	2	138	78	
15	1	S103	R-R102		1	09JUL2020	09JUL2020	16JUL2020	27JUL2020	09OCT2020	28DEC2020	1	64.44109589	3	118	69	16JUL2020	252	42	0	85
16	1	S103	R-R102		1	09JUL2020	09JUL2020	16JUL2020	27JUL2020	09OCT2020	28DEC2020	1	64.44109589	4	122	68	
17	1	S103	R-R102		1	09JUL2020	09JUL2020	16JUL2020	27JUL2020	09OCT2020	28DEC2020	1	64.44109589	5	140	75	09OCT2020	229	65	0	80
18	1	S103	R-R102		1	09JUL2020	09JUL2020	16JUL2020	27JUL2020	09OCT2020	28DEC2020	1	64.44109589	6	158	86	

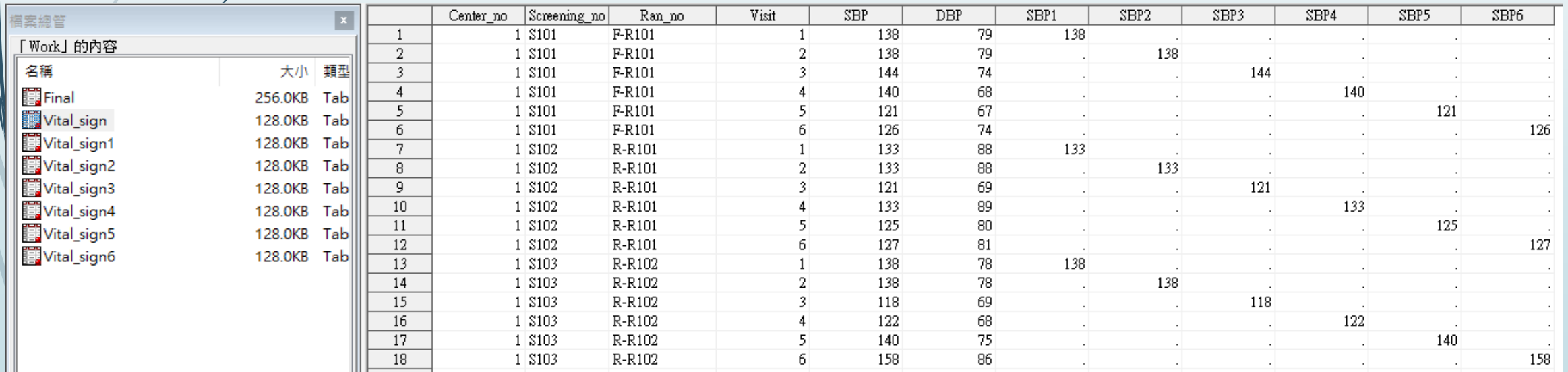
Data Management using SAS

► 利用 SAS 進行 資料管理

長資料格式：Vital_sign

/*長資料改成短資料格式*/

```
data vital_sign; set data.vital_sign;
SBP1=.; if visit=1 then SBP1=SBP;
SBP2=.; if visit=2 then SBP2=SBP;
SBP3=.; if visit=3 then SBP3=SBP;
SBP4=.; if visit=4 then SBP4=SBP;
SBP5=.; if visit=5 then SBP5=SBP;
SBP6=.; if visit=6 then SBP6=SBP; run;
```



	Center_no	Screening_no	Ran_no	Visit	SBP	DBP	SBP1	SBP2	SBP3	SBP4	SBP5	SBP6
1	1	\$101	F-R101	1	138	79	138
2	1	\$101	F-R101	2	138	79	.	138
3	1	\$101	F-R101	3	144	74	.	.	144	.	.	.
4	1	\$101	F-R101	4	140	68	.	.	.	140	.	.
5	1	\$101	F-R101	5	121	67	121	.
6	1	\$101	F-R101	6	126	74	126
7	1	\$102	R-R101	1	133	88	133
8	1	\$102	R-R101	2	133	88	.	133
9	1	\$102	R-R101	3	121	69	.	.	121	.	.	.
10	1	\$102	R-R101	4	133	89	.	.	.	133	.	.
11	1	\$102	R-R101	5	125	80	125	.
12	1	\$102	R-R101	6	127	81	127
13	1	\$103	R-R102	1	138	78	138
14	1	\$103	R-R102	2	138	78	.	138
15	1	\$103	R-R102	3	118	69	.	.	118	.	.	.
16	1	\$103	R-R102	4	122	68	.	.	.	122	.	.
17	1	\$103	R-R102	5	140	75	140	.
18	1	\$103	R-R102	6	158	86	158

Data Management using SAS

► 利用 SAS 進行 資料管理

短資料格式：Vital_sign1 – Vital_sign6

SBP1, SBP2.....SBP5, SBP6

名稱	大小	類型	Center_no	Screening_no	SBP1
Final	256.0KB	Tab	1	S101	138
Vital_sign	128.0KB	Tab	2	S102	133
Vital_sign1	128.0KB	Tab	3	S103	138
Vital_sign2	128.0KB	Tab	4	S201	148
Vital_sign3	128.0KB	Tab	5	S202	138
Vital_sign4	128.0KB	Tab	6	S203	121
Vital_sign5	128.0KB	Tab	7	S204	120
Vital_sign6	128.0KB	Tab	8	S301	180
			9	S302	131
			10	S303	123
			11	S304	135
			12	S305	122
			13	S306	119
			14	S307	124
			15	S308	128
			16	S309	153
			17	S310	127
			18	S311	125

名稱	大小	類型	Center_no	Screening_no	SBP6
Final	256.0KB	Tab	1	S101	126
Vital_sign	128.0KB	Tab	2	S102	127
Vital_sign1	128.0KB	Tab	3	S103	158
Vital_sign2	128.0KB	Tab	4	S201	125
Vital_sign3	128.0KB	Tab	5	S202	119
Vital_sign4	128.0KB	Tab	6	S203	116
Vital_sign5	128.0KB	Tab	7	S204	130
Vital_sign6	128.0KB	Tab	8	S301	135
			9	S303	123
			10	S304	126
			11	S305	133
			12	S306	135
			13	S307	129
			14	S308	111
			15	S309	132
			16	S310	129
			17	S311	125
			18	S312	116

```
data vital_sign1; set vital_sign; if visit=1; keep Center_no Screening_no sbp1; run;
data vital_sign2; set vital_sign; if visit=2; keep Center_no Screening_no sbp2; run;
data vital_sign3; set vital_sign; if visit=3; keep Center_no Screening_no sbp3; run;
data vital_sign4; set vital_sign; if visit=4; keep Center_no Screening_no sbp4; run;
data vital_sign5; set vital_sign; if visit=5; keep Center_no Screening_no sbp5; run;
data vital_sign6; set vital_sign; if visit=6; keep Center_no Screening_no sbp6; run;
```

```
data data.vital_sign_m; merge vital_sign1 vital_sign2 vital_sign3 vital_sign4 vital_sign5
vital_sign6;
by Screening_no; run;
```

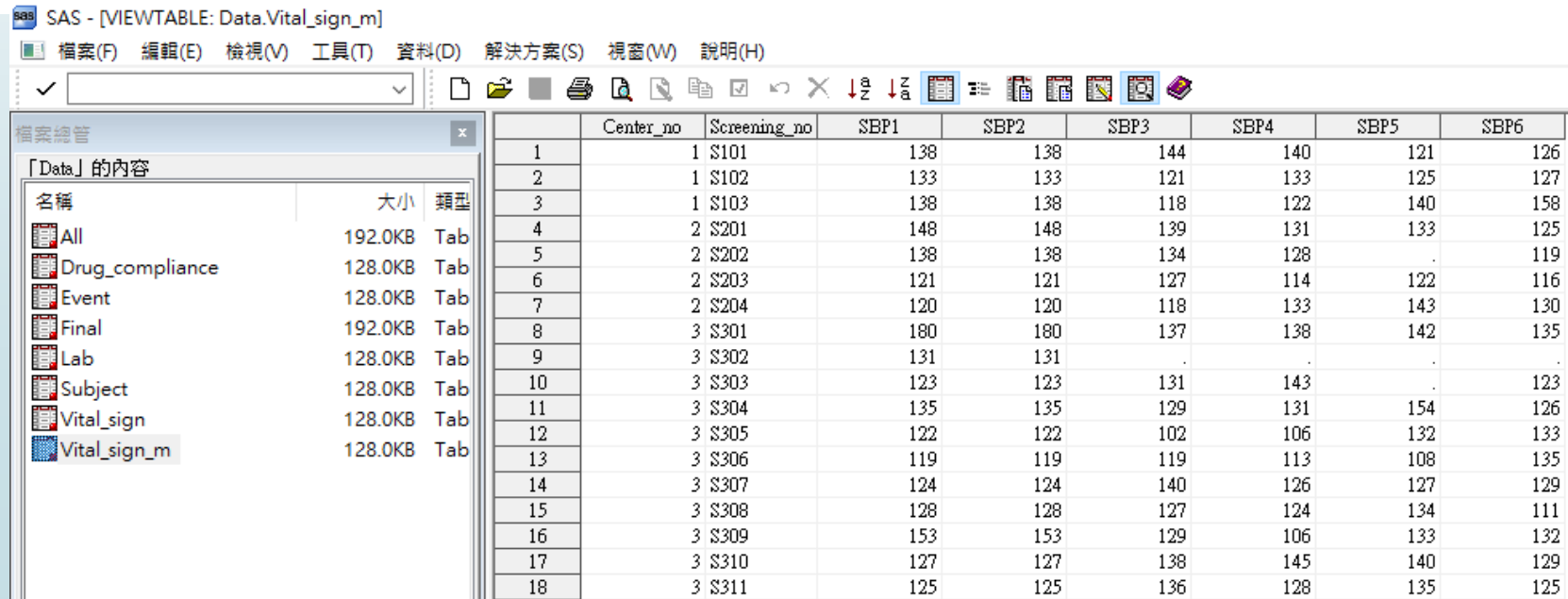
Data Management using SAS

► 利用 SAS 進行 資料管理

短資料格式：data.Vital_sign_m

```
data vital_sign1; set vital_sign; if visit=1; keep Center_no Screening_no sbp1; run;
data vital_sign2; set vital_sign; if visit=2; keep Center_no Screening_no sbp2; run;
data vital_sign3; set vital_sign; if visit=3; keep Center_no Screening_no sbp3; run;
data vital_sign4; set vital_sign; if visit=4; keep Center_no Screening_no sbp4; run;
data vital_sign5; set vital_sign; if visit=5; keep Center_no Screening_no sbp5; run;
data vital_sign6; set vital_sign; if visit=6; keep Center_no Screening_no sbp6; run;
```

```
data data.vital_sign_m; merge vital_sign1 vital_sign2 vital_sign3 vital_sign4 vital_sign5
vital_sign6;
by Screening_no; run;
```



SAS - [VIEWTABLE: Data.Vital_sign_m]

檔案(F) 編輯(E) 檢視(V) 工具(T) 資料(D) 解決方案(S) 視窗(W) 說明(H)

檔案總管

[Data] 的內容

	Center_no	Screening_no	SBP1	SBP2	SBP3	SBP4	SBP5	SBP6
1	1	S101	138	138	144	140	121	126
2	1	S102	133	133	121	133	125	127
3	1	S103	138	138	118	122	140	158
4	2	S201	148	148	139	131	133	125
5	2	S202	138	138	134	128	.	119
6	2	S203	121	121	127	114	122	116
7	2	S204	120	120	118	133	143	130
8	3	S301	180	180	137	138	142	135
9	3	S302	131	131
10	3	S303	123	123	131	143	.	123
11	3	S304	135	135	129	131	154	126
12	3	S305	122	122	102	106	132	133
13	3	S306	119	119	119	113	108	135
14	3	S307	124	124	140	126	127	129
15	3	S308	128	128	127	124	134	111
16	3	S309	153	153	129	106	133	132
17	3	S310	127	127	138	145	140	129
18	3	S311	125	125	136	128	135	125

Clinical Trial Data Management using SAS

- **Clinical Trial Data Management using SAS mostly for the following purposes:**
 - Designing a protocol and beginning a study : Sample size
 - Recruitment of Patients and Investigators : Randomization
 - Health Data Management : Data Management using SAS
 - **Analyzing Data**

Analyzing Data using SAS

► 利用 SAS 進行 資料計算

```
data data.vital_sign_m; set data.vital_sign_m;
mean_SBP = mean (SBP1, SBP2, SBP3, SBP4, SBP5, SBP6);
Avg_SBP=(SBP1+SBP2+SBP3+SBP4+SBP5+SBP6) / 6;
min_SBP = min (SBP1, SBP2, SBP3, SBP4, SBP5, SBP6);
max_SBP = max (SBP1, SBP2, SBP3, SBP4, SBP5, SBP6);
SBP_sqrt=sqrt (mean_SBP) ;
SBP_square=Avg_SBP**2 ;
SBP_exp=exp (mean_SBP) ;
run;
```

短資料格式 : data.Vital_sign_m

SBP1	SBP2	SBP3	SBP4	SBP5	SBP6	mean_SBP	Avg_SBP	min_SBP	max_SBP	SBP_sqrt	SBP_square	SBP_exp
138	138	144	140	121	126	134.5	134.5	121	144	11.597413505	18090.25	2.5858767E58
133	133	121	133	125	127	128.666666667	128.666666667	121	133	11.343133018	16555.111111	7.572222E55
138	138	118	122	140	158	135.666666667	135.666666667	118	158	11.647603473	18405.444444	8.3039497E58
148	148	139	131	133	125	137.333333333	137.333333333	125	148	11.718930554	18860.444444	4.3965179E59
138	138	134	128	.	119	131.4	.	119	138	11.462983905	.	1.1649168E57
121	121	127	114	122	116	120.166666667	120.166666667	114	127	10.962055768	14440.027778	1.5407077E52
120	120	118	133	143	130	127.333333333	127.333333333	118	143	11.284207253	16213.777778	1.996016E55
180	180	137	138	142	135	152	152	135	180	12.328828006	23104	1.0298198E66
131	131	131	.	131	131	11.445523142	.	7.8086711E56
123	123	131	143	.	123	128.6	.	123	143	11.340194002	.	7.0838666E55

Analyzing Data using SAS

長資料格式：data.final

利用 SAS 進行 描述性統計分析：連續型變數

```
PROC MEANS DATA=data.final
MAXDEC=2  N  MEAN  STD  MEDIAN Q1 Q3
/*MEDIAN Q1 Q3  MIN MAX  N VAR SKEWNESS KURTOSIS T PROBT CLM /*CLM 樣本信賴區間*/;
VAR  SBP DBP Creatinine  AST ALT  ;
CLASS  visit  ;
/*BY  gender  ; */ RUN;
```

Visit	觀測值數目	變數	標籤	N	平均值	標準差	中位數	下四分位數	上四分位數
1	76	SBP	SBP	76	132.80	17.30	130.00	121.50	140.00
		DBP	DBP	76	80.20	11.26	79.00	73.50	89.00
		Creatinine	Creatinine	75	1.05	0.32	0.99	0.86	1.15
		AST	AST	76	22.71	10.44	20.00	17.00	23.00
		ALT	ALT	76	24.86	20.70	20.00	15.00	25.50
2	73	SBP	SBP	73	132.77	17.84	130.00	121.00	140.00
		DBP	DBP	73	79.85	11.58	79.00	73.00	89.00
		Creatinine	Creatinine	0
		AST	AST	0
		ALT	ALT	0
3	71	SBP	SBP	71	128.68	15.69	127.00	120.00	138.00
		DBP	DBP	71	77.55	10.45	78.00	70.00	86.00
		Creatinine	Creatinine	71	1.03	0.29	1.00	0.87	1.18
		AST	AST	71	22.15	9.33	20.00	17.00	23.00
		ALT	ALT	71	23.28	16.33	20.00	15.00	26.00

Visit	觀測值數目	變數	標籤	N	平均值	標準差	中位數	下四分位數	上四分位數
4	71	SBP	SBP	69	125.58	16.29	125.00	114.00	135.00
		DBP	DBP	69	78.03	14.18	76.00	70.00	82.00
		Creatinine	Creatinine	0
		AST	AST	0
		ALT	ALT	0
5	71	SBP	SBP	64	128.08	14.53	130.00	119.50	138.50
		DBP	DBP	64	77.83	11.09	79.00	68.50	85.50
		Creatinine	Creatinine	0
		AST	AST	0
		ALT	ALT	0
6	71	SBP	SBP	71	128.90	15.30	128.00	119.00	138.00
		DBP	DBP	71	77.32	9.99	79.00	70.00	85.00
		Creatinine	Creatinine	68	1.03	0.28	1.00	0.89	1.12
		AST	AST	68	21.10	7.72	20.00	17.00	23.00
		ALT	ALT	68	21.54	10.77	19.00	15.00	27.00

Analyzing Data using SAS

- 利用 SAS 進行 描述性統計分析：Student's t test

```

/*Student's t test*/
data lab; set data.final;
if visit=1; keep Group SBP gender; run;

PROC TTEST DATA=lab ;
CLASS  group; VAR  SBP  ; RUN;
    
```

TTEST 程序

變數: SBP (SBP)

Group	N	平均值	標準差	標準誤差	最小值	最大值
1	37	134.9	17.1188	2.8143	107.0	174.0
2	34	130.6	18.6664	3.2013	94.0000	180.0
Diff (1-2)		4.3871	17.8757	4.2467		

Group	方法	平均值	95% CL 平均值	標準差	95% CL 標準差
1		134.9	129.2 140.7	17.1188	13.9212 22.2366
2		130.6	124.0 137.1	18.6664	15.0559 24.5702
Diff (1-2)	集區	4.3871	-4.0848 12.8591	17.8757	15.3269 21.4491
Diff (1-2)	Satterthwaite	4.3871	-4.1207 12.8949		

方法	變異數	自由度	t 值	Pr > t
集區	均等	69	1.03	0.3052
Satterthwaite	不均等	67.023	1.03	0.3071

變異數相等性				
方法	分子自由度	分母自由度	F 值	Pr > F
Folded F	33	36	1.19	0.6104

Analyzing Data using SAS

► 利用 SAS 進行 描述性統計分析：
類別型變數 + Chi-square test

```
/*類別型變數 + Chi-square test*/  
data lab; set data.final;  
if visit=1; keep Group SBP gender; run;
```

```
PROC FREQ data=lab;  
table gender*group / chisq expected exact ; run;
```

SAS 系統

FREQ 程序

次數 預期 百分比 列百分比 欄百分比	Table of Gender by Group			
	Gender(Gender)	Group(Group)		總計
		1	2	
0		7	8	15
		7.8169	7.1831	
		9.86	11.27	21.13
		46.67	53.33	
		18.92	23.53	
1		30	26	56
		29.183	26.817	
		42.25	36.62	78.87
		53.57	46.43	
		81.08	76.47	
總計		37	34	71
		52.11	47.89	100.00

次數遺漏 = 5

表格 Group-Gender*s 的統計值

統計值	自由度	值	機率
卡方	1	0.2260	0.6345
擬度比卡方	1	0.2259	0.6346
連續性調整卡方	1	0.0340	0.8537
Mantel-Haenszel 卡方	1	0.2228	0.6369
Phi 係數		-0.0564	
列聯係數		0.0563	
Cramer V		-0.0564	

Fisher 精確檢定

儲存格 (1,1) 次數 (F)	7
左邊 Pr <= F	0.4262
右邊 Pr >= F	0.7782
表格機率 (P)	0.2044
雙邊 Pr <= P	0.7729



SAS統計軟體基礎操作訓練 (1)



Thank you for listening